

Operador de Energía de Teager para la Detección de Hipernasalidad en Niños con Labio y Paladar Hendido

Juan R. Orozco-Arroyave¹
Jonny A. Uribe²
Jesús F. Vargas-Bonilla³

Resumen

El labio y/o paladar hendido (LPH) es una malformación, que tiene orígenes de tipo genético y ambiental. En Colombia, 6 de cada 10000 niños nacen con esta malformación, mientras en el resto del mundo la proporción se encuentra en 1 de cada 10000. El LPH trae consigo patologías en el habla tales como: hipernasalidad, hiponasalidad, golpe glótico, entre otras. De todas estas patologías, la hipernasalidad es la más recurrente en pacientes con LPH, apareciendo aproximadamente en el 90% de los casos. En este trabajo se hace un análisis, basado en resultados experimentales, del desempeño del Operador de Energía de Teager (TEO, por las siglas en inglés de Teager Energy Operator), para la detección de hipernasalidad en pacientes con LPH. Se analiza una versión generalizada del TEO con el fin de validar su capacidad discriminante en la detección de hipernasalidad, aplicándolo sobre una base de datos con registros de voz reales, de niños con LPH y niños control. Los resultados obtenidos comprueban que el TEO posee gran capacidad discriminante, y puede aportar información relevante en el proceso de detección de hipernasalidad.

-
- 1 Grupo de Investigación en Telecomunicaciones Aplicadas, GITA, Universidad de Antioquia, Medellín-Colombia, rafa.rozco@gmail.com
 - 2 Grupo de Investigación en Telecomunicaciones Aplicadas, GITA, Universidad de Antioquia, Medellín-Colombia, sirjoy.ur@gmail.com
 - 3 Grupo de Electrónica de Potencia, Automatización y Robótica, GEPAR, Universidad de Antioquia, jfvargas@udea.edu.co

Fecha de recepción: 16 de Agosto de 2010
Fecha de aceptación: 09 de Enero de 2011

Palabras clave

Hipernasalidad, Operador de energía de Teager (TEO), Labio y/o Paladar Hendido (LPH).

Abstract

The Cleft Lip and Palate (CLP) is a multi-factorial malformation that has genetic and environmental origins. In Colombia, 6 of 10000 children born with this malformation, while in the rest of the world, the proportion is about 1 of each 10000. The CLP originates some kinds of pathologies in the speech such as hypernasality, hyponasality, glottal coup, among others. From all of these pathologies, the hypernasality is the most frequent in CLP patients, appearing approximately, in 90% of the cases. In this work, based on experimental results, a performance analysis of the Teager Energy Operator (TEO) is made in the framework of the hypernasality detection on CLP patients. A generalized version of TEO is applied to validate its discrimination capacity over a dataset containing real voice registers of children with CLP and healthy children. The results showed that TEO has discrimination capacity and can contribute with important information in the hypernasality detection issue.

Keywords

Hypernasality, Teager Energy Operator (TEO), Cleft Lip and Palate (CLP).

1. INTRODUCCIÓN

Los pacientes con LPH presentan problemas de emisión vocal y resonancia, tales como: hipernasalidad, hiponasalidad, golpe glótico, entre otros (Castellanos, 2005). Debido a que el 90% de los pacientes con LPH son hipernasales, es de especial interés científico estudiar esta patología (Castellanos, et al. 2006). Desde la década de 1970, existen estudios sobre análisis acústico de voces patológicas (Fujimura & Lindqvist, 1971). Estas investigaciones han orientado a los investigadores hacia el análisis del espectro de la voz, particularmente en las bajas frecuencias.

Algunas alteraciones de las señales en el tiempo pueden ser estudiadas mediante características acústicas; una de las más estudiadas por los investigadores es el período fundamental de la señal vocal, conocida en inglés como Pitch. Mediante esta característica se busca medir el período de vibración de las cuerdas vocales; cuando esta medida se aleja de valores previamente identificados como normales, es posible inferir que el tracto vocal objeto de estudio está sufriendo algún tipo de patología (Kasuya et al., 1983). Otra característica es el Jitter, que representa la variación que el pitch sufre a lo largo del tiempo. Así mismo, la variación de amplitud en el Pitch, constituye el denominado Shimmer, que es otro de los patrones importantes en la identificación de patologías de voz (Kasuya et al., 1983).

Además del análisis efectuado en el dominio temporal, es posible apoyarse en los modelos clásicos de producción del habla para proponer métricas que permitan dar cuenta de patologías como la hipernasalidad. En general, se han encontrado métricas que toman en cuenta el ruido turbulento producido por el tracto vocal en condiciones patológicas. Es así como se han planteado la relación de armónicos a ruido (HNR, por las siglas en inglés de Harmonic to Noise Excitation Ratio) (Yumoto, 1982), la Energía de Ruido Normalizada (NNE, por las siglas en inglés de Normalized Noise Energy), formulada inicialmente en (Kasuya et al., 1986) y la relación de excitación glotal a ruido (GNE, por las siglas en inglés de Glottal to Noise Excitation Ratio) (Michaelis et al., 1997), como métricas útiles en la detección de patologías en la voz.

El problema de la detección automática de patologías en la voz también ha sido tratado en el dominio espectral, particularmente en la detección de hipernasalidad los autores han enfocado sus esfuerzos en el análisis de la región de baja frecuencia del espectro vocal. Glass & Zue (1985), concentraron su búsqueda en dicha región, formulando seis características acústicas potencialmente útiles para la detección de nasalización en el idioma Inglés.

Las características son: El centro de masa en las bajas frecuencias del espectro de la voz (0 a 100 Hz); la desviación estándar alrededor del centro de masa; el porcentaje de tiempo máximo en que hay un polo extra en las bajas frecuencias; el porcentaje de tiempo mínimo en que hay un polo extra en las bajas frecuencias; el máximo valor de profundidad del valle que existe entre el primer polo y el polo extra debido a la nasalización; y el mínimo valor de la diferencia promedio entre el primer polo y el polo extra.

Apoyándose en el trabajo de Glass, Chen (1996; 1997) buscó detectar nasalización en el francés y en el inglés, encontrando de forma recurrente dos polos adicionales en el espectro, uno antes y otro después del primer formante para vocablos nasales, por lo cual formuló las diferencias entre la amplitud del primer formante A_1 , y las amplitudes de los polos extra P_0 y P_1 , como dos métricas de nasalidad en la voz.

Por su parte, Vijayalakshmi & Ramasubba (2005) exploraron el espectro de la voz buscando detectar hipernasalidad. Apoyándose en la teoría propuesta en Murthy et al. (1989), utilizaron las funciones de retraso de grupo para mejorar la resolución espectral, detectando un pico adicional situado en 250 Hz para voces hipernasales. Posteriormente, tomando como base el resultado anterior, Vijayalakshmi et al. (2007) probaron la capacidad de detección de hipernasalidad de su técnica usando funciones de retraso de grupo de banda limitada y obtienen resultados de hasta el 88,7% de acierto, en la vocal /i/.

Recientemente, Vijayalakshmi et al. (2009) presentaron una técnica para detección de hipernasalidad basada en la modificación de los polos del espectro de predicción lineal (LP, por las siglas en inglés de Linear Prediction) de la voz. El método consiste en calcular el espectro LP con 28 coeficientes, con el fin de poder identificar los picos adicionales debidos a las componentes de

nasalización. Una vez detectado el pico más alto, éste es aplanado mediante la modificación de la magnitud del polo correspondiente en el diagrama de polos y ceros, luego se procede a generar otra señal de voz sintetizada, y finalmente la señal original y la sintetizada son comparadas mediante el coeficiente de correlación. Cuando el coeficiente es mayor que 0,65 se estima que la muestra es normal, pues el espectro con polo aplanado es muy similar al espectro original, de lo contrario se estima que la muestra es hipernasal, pues la modificación del espectro genera una diferencia sustancial con respecto al original, lo cual hace que ambas señales estén poco correlacionadas.

Otra de las técnicas que pueden ser aplicadas para la detección de hipernasalidad, es utilizar operadores matemáticos para modificar las señales bajo estudio, ofreciendo mejores prestaciones en el dominio transformado. Cairns et al. (1994) proponen una técnica de detección de hipernasalidad basada en el operador de energía de Teager (TEO). La aplicación del Operador de Energía de Teager (TEO), para la detección de hipernasalidad en la voz, permite tener en cuenta diferentes componentes en frecuencia que aparecen en el espectro de voz hipernasal, las cuales pueden ser separadas utilizando dicho operador. Cairns et al. (1996) hace la comparación entre los perfiles de energía de Teager de voces con hipernasalidad simulada limitadas en banda, usando filtros pasabaja y pasabanda, y posteriormente aplicando la correlación como medida de diferencia entre ambos perfiles; de igual forma lo hace para las señales de voz sana.

Sus resultados indican que existe mayor nivel de correlación entre los perfiles filtrados provenientes de voces sanas que entre los perfiles filtrados provenientes de voces hipernasales. El resultado obtenido por Cairns, valida el hecho de que en el espectro de la señal hipernasal aparecen picos y valles extra, debidos al exceso de nasalización en la señal de voz. Para su experimento, Cairns trabajó con 11 personas sanas, y para el registro de las voces hipernasales, estas mismas personas simularon la patología. Adicionalmente, para las pruebas de clasificación entre vocales sanas e hipernasales, utilizó sólo la vocal /a/ y la /i/. Para tener certeza acerca de la existencia de la patología en los registros simulados, éstos fueron evaluados usando un Nasometro de Kay Elemetrics.

Por otro lado, Pruthi (2004) y Pruthi & Espy-Wilson (2005; 2007) analizaron el espectro de la voz con el fin de detectar nasalización en el idioma Inglés, y trabaja con registros de 630 personas sanas que pronunciaron 10 frases cada una. El trabajo de Pruthi, mejoró lo planteado por Cairns en varios aspectos: en vez de utilizar filtros pasabajas y pasabanda, usó dos filtros pasabanda centrados en el primer formante, uno de ellos con banda estrecha (100 Hz), y otro con mayor ancho de banda (1000 Hz). El primer formante fue calculado usando un rastreador basado en el algoritmo ESPS (Talkin, 1987), diferente a la técnica empleada por Cairns.

En el presente artículo, se busca validar los resultados obtenidos por Cairns pero utilizando una base de datos más amplia, con registros de voces hipernasales reales, provenientes de voces de niños con LPH. Adicionalmente, se implementan versiones alternativas del TEO, a partir de la generalización presentada en Kvedalen (2003) y de lo presentado en Ying (1993). Para comprobar la capacidad discriminante de las diferentes versiones del TEO, se implementó un clasificador cuadrático basado en la covarianza de las características, obteniendo tasas de acierto en la clasificación de 93,81%. Comprobando que el TEO posee gran capacidad discriminante para la detección de hipernasalidad.

El resto del artículo está organizado así: en el marco teórico se presentan algunos conceptos empleados en los experimentos realizados. En la sección de metodología, se describen tanto la base de datos empleada como el clasificador implementado. Posteriormente se presentan la discusión y los resultados obtenidos. Finalmente, aparecen las conclusiones más importantes derivadas de este trabajo.

2. MARCO TEÓRICO

2.1 Formantes Vocálicos e Hipernasalidad

Los sonidos que son generados por la vibración de las cuerdas vocales, están caracterizados por un alto contenido armónico; a partir del espectro de dichos sonidos, es posible identificar resonancias, las cuales aparecen como picos en la envolvente del espec-

tro y reciben el nombre de formantes vocálicos, también conocidos como formantes orales, cuya posición, amplitud y ancho de banda se ven alterados debido a la hipernasalidad. De acuerdo con Pruthi & Espy-Wilson (2007), el primero de los formantes puede estar rodeado de picos adicionales debidos a la nasalización excesiva de la voz, formando así espectros multicomponente.

2.2 Perfiles de Energía Teager

Dada una señal $x(n)$, su perfil de energía Teager se define, de acuerdo con (Kaiser, 1990), como en (1):

$$\psi_d[x(n)] = x(n)^2 - x(n+1)x(n-1) \tag{1}$$

Una de las características de este operador es su sensibilidad a entradas multicomponente. Sea una señal compuesta tal que $x(n) = s(n) + g(n)$. Su perfil está dado por (2):

$$\begin{aligned} \psi(x(n)) &= \psi[s(n) + g(n)]^2 - [s(n+1) + g(n+1)][s(n-1) + g(n-1)] \\ &= \psi[s(n)] + \psi[g(n)] + \psi_{cross}[s(n), g(n)] + \psi_{cross}[g(n), s(n)] \end{aligned} \tag{2}$$

Donde $\psi_{cross}[g(n), s(n)] = g(n)s(n) - g(n+1)s(n-1)$. Esto significa que el operador de energía no obedece al principio de superposición, y añade términos extra (ψ_{cross}) para señales multicomponente.

2.3 Correlación sobre los Perfiles de Energía de Teager

El CTEO (Correlation Teager Energy Operator), es una métrica propuesta en Cairns et al. (1996) para detectar hipernasalidad. Su razonamiento puede ser descrito de la siguiente forma: la señal de voz sana contiene sólo formantes orales (3),

$$S_{NORMAL} = \sum F(\omega) \tag{3}$$

Donde $F(\omega)$ representa los formantes orales. Por su parte, la voz hipernasal contiene formantes, anti formantes (valles en el espectro) y formantes nasales (4):

$$S_{NASAL} = \sum F(\omega) - \sum AF(\omega) + \sum NF(\omega) \quad (4)$$

Donde $AF(\omega)$ representa los anti formantes (el signo menos da cuenta de que los antifonantes aparecen como valles en el espectro), y $NF(\omega)$ son los formantes extra debidos a la hipernasalidad de la señal de voz. Si una voz sana se filtra con un filtro pasabajas apropiado, es posible extraer el primer formante, dado por F_1 en (5):

$$S_{NORMAL} - LPF(\omega) = F_1(\omega) \quad (5)$$

Mientras que al aplicar este mismo filtro a una señal hipernasal, aparecen términos adicionales a F_1 , tales como los antifonantes y los formantes nasales (6):

$$S_{NASAL} - LPF(\omega) = F_1(\omega) - \sum AF(\omega) + \sum NF(\omega) \quad (6)$$

Si en vez de utilizar un filtro pasabajas, usamos uno pasabanda alrededor de F_1 , aparecerá sólo el primer formante tanto para la voz normal (7) como para la voz hipernasal (8):

$$S_{NORMAL} - BPF(\omega) = F_1(\omega) \quad (7)$$

$$S_{NASAL} - BPF(\omega) = F_1(\omega) \quad (8)$$

Buscando explotar esta asimetría, el TEO es calculado sobre las señales filtradas, tanto pasabajas como pasabanda, para luego medir la correlación entre ambas respuestas. Dado que el filtrado pasabajas sobre la voz hipernasal dejará una señal multicomponente, su perfil de energía Teager tendrá una contribución debida a los términos cruzados (ψ_{cross}), es decir, los perfiles de energía de la señal pasabajas y pasabanda para la voz hipernasal serán diferentes, mientras que para la voz normal, éstos serán similares

tanto para la señal pasabaja como para la señal pasabanda. De esta forma, entre más bajo sea el valor de la correlación para los perfiles de Energía de las señales pasabajas y pasabanda, más diferentes serán las señales analizadas.

En el trabajo de Cairns se utiliza la función de correlación para cuantificar la diferencia entre los perfiles obtenidos, se emplean las voces de 11 personas sanas. Para obtener datos de muestras de voz hipernasales, las mismas 11 personas sanas simularon voces hipernasales. El nivel de nasalización de cada registro hipernasal simulado, fue validado usando un nasómetro de Kay Elemetrics. Finalmente, para el ejercicio de clasificación, Cairns utilizó sólo las vocales /a/ e /i/.

Por otra parte, Pruthi (2007), buscando detectar nasalización en el idioma Inglés, mejora el trabajo de Cairns eliminando la necesidad de efectuar sincronización de Pitch y utilizando dos filtros pasabanda en vez de uno pasabaja y uno pasabanda. En su trabajo, las frecuencias centrales de los filtros se encuentran alrededor del primer formante, cuya ubicación es conocida mediante un rastreador de formantes ESFS (Talkin, 1987). Uno de los filtros es de banda estrecha (100 Hz), mientras que el otro posee mayor ancho de banda (1000 Hz).

Para sus validaciones, Pruthi utiliza una base de datos extensa, compuesta por 6300 frases. Sus resultados muestran que la correlación de los perfiles de energía Teager contribuye en la clasificación de un sonido como nasal o normal. En el presente artículo, se aplica el concepto presentado por Cairns para la detección de hipernasalidad en registros de voz reales. Adicionalmente, se exploran otras implementaciones del TEO, las cuales se describen a continuación:

2.4 TEO Generalizado

De acuerdo con Kvedalen (2003), el TEO puede ser generalizado de tal forma que la expresión matemática es como en (9):

$$\psi_d[x(n)] = x(n)^{\frac{2}{m}} - x(n+1)x(n-1)^{\frac{1}{m}} \quad (9)$$

Donde m y $M \in Z$. Considerando esta expresión, se efectuaron diversas pruebas con el fin determinar cuáles son los valores de m y M más adecuados para efectos de detección de hipernasalidad.

2.5 TEO Basado en la FFT

Ying et al. (1993) plantean la posibilidad de calcular la energía de una señal de voz utilizando una forma modificada del Operador de Energía de Teager (TEO). La forma matemática propuesta por Ying et al. se muestra en (10):

$$\Psi_n = \left[\sum_{i=1}^{\Omega} i^2 S_n(i) \right]^{\frac{1}{2}} \quad (10)$$

Donde Ψ_n denota el TEO modificado, S_n es la densidad espectral de potencia de la n -ésima trama de voz, calculada usando la FFT, e i es el valor de la frecuencia en el dominio discreto.

3. METODOLOGÍA

3.1 Base de Datos

La base de datos utilizada para el desarrollo de este trabajo, fue la misma utilizada en Castellanos (2005). La cual fue suministrada por el grupo de Control y Procesamiento Digital de Señales (GC&PDS) de la Universidad Nacional de Colombia, sede Manizales. Está conformada por registros de voces provenientes de 110 niños sanos y 156 con LPH, los cuales fueron diagnosticados como hipernasales por un experto en Fonoaudiología.

Para este trabajo, fueron considerados los registros provenientes de la pronunciación sostenida de las cinco vocales del idioma Español. Los registros tuvieron una duración promedio de 300 ms, y fueron capturados en condiciones controladas: dentro de una cabina sono-amortiguada, utilizando un micrófono omnidireccional de alta ganancia, cableado profesional balanceado (conectores XLR) y una tarjeta de audio profesional. Todos los registros fueron

digitalizados con una frecuencia de muestreo de 48000 muestras por segundo, utilizando 16 bits en la cuantización.

3.2 Implementación

Considerando los buenos resultados obtenidos en Pruthi & Espy-Wilson (2007), usando filtros FIR pasabanda de orden 200, para las pruebas realizadas también fueron implementados filtros de este tipo.

A pesar de que Pruthi presenta todos sus resultados usando anchos de banda de $\omega_1 = 100 \text{ Hz}$ y $\omega_2 = 1000 \text{ Hz}$ para los filtros de banda estrecha y banda ancha, respectivamente, considerando que la hipernasalidad afecta el ancho de banda del primer formante, dichos valores fueron variados para analizar el comportamiento de las métricas bajo diferentes condiciones de ancho de banda. Es necesario resaltar que los resultados de Pruthi fueron obtenidos trabajando con voces sanas, y su objetivo era detectar nasalización en el idioma Inglés, no hipernasalidad. Con el objetivo de verificar la validez de la metodología propuesta por él, aplicándola a la detección de hipernasalidad, se hicieron pruebas estadísticas para diferentes valores de anchos de banda en los filtros. Los valores de los anchos de banda probados fueron: para ω_1 : 50 Hz, 100 Hz, 150 Hz, ..., 600 Hz y para ω_2 : 900 Hz, 1000 Hz y 1100 Hz.

Adicionalmente, se evaluó la capacidad de discriminación de dos variaciones del TEO; la primera variación corresponde al caso generalizado con valores de exponente $m = 1, 2, \dots, 5$, y retraso $M = 1, 2, \dots, 5$, y la segunda, cuando es calculado a partir de la densidad espectral de potencia ponderada (Ying, 1993). La prueba estadística fue hecha mediante el test de *Kruskal-Wallis*, definiendo la siguiente hipótesis nula: H_0 : El CTEO presenta la misma distribución para las clases normal e hipernasal.

En todas las pruebas realizadas, el valor del estadístico Chi-Cuadrado fue tan grande, que el correspondiente valor de la probabilidad (p) de aceptación de la hipótesis nula fue siempre cero. Debido a esto, se decidió utilizar los valores del estadístico como indicador de aumento o disminución en la capacidad discriminante de la métrica. Esto es posible teniendo en cuenta que los valores de

χ^2 y p son correspondientes, es decir, un valor alto de χ^2 produce un valor pequeño de p y viceversa (NIST, 2010).

Mediante la prueba estadística se comprobó que métricas calculadas provenían de dos poblaciones diferentes (voces sanas e hipernasales), y que por ende éstas seguían distribuciones diferentes. Posterior a la prueba estadística, se seleccionaron algunas de las métricas con mejores resultados (diferentes valores de m y M), dejando los anchos de banda fijos en los valores implementados por Pruthi ($\omega_1 = 50$ Hz y $\omega_2 = 900$ Hz) y se implementó un clasificador cuadrático, el cual se describe a continuación.

3.3 Clasificador Cuadrático

En general, con el fin de dividir el espacio de características en c regiones de decisión, y buscando minimizar la rata de errores, se define el siguiente conjunto de funciones discriminantes (Duda et al., 2001) dadas por (11):

$$g_i(X) = P(w_i|X) + \ln(P(w_i)) \quad (11)$$

Tomando distribuciones normales multivariadas para los datos, es decir, con densidad de probabilidad dada por (12):

$$p(X) = \frac{1}{(2\pi)^{d/2} |\Sigma|^{1/2}} e^{[(X-\mu)^T \Sigma^{-1} (X-\mu)]} \quad (12)$$

Donde X es un vector columna con d componentes, μ es un vector d - dimensional de medias, Σ es la matriz de covarianzas con dimensión $d \times d$, $|\Sigma|$ denota el determinante de la matriz y Σ^{-1} su inversa. $(X - \mu)^T$ es la transpuesta de $(X - \mu)$.

Es posible expandir la expresión de (11) obteniendo un conjunto de funciones discriminantes para datos con densidad normal, así (13):

$$g_i(X) = -\frac{1}{2}(X - \mu_i)^T \Sigma_i^{-1} (X - \mu_i) - \frac{d}{2} \ln(2\pi) - \frac{1}{2} \ln|\Sigma_i| + \ln P(w_i) \quad (13)$$

Cuando las matrices de covarianza se consideran iguales pero arbitrarias, para todas las clases, como en el clasificador implementado en este trabajo, las características serán clasificadas dentro de hiperelipsoides iguales, tales que el hiperelipsoide de la i - ésima clase estará centrado alrededor del vector de medias μ_i .

Dado que para el caso descrito $\Sigma_i = \Sigma$ es independiente de i , al igual que el término $\frac{d}{2} \ln(2\pi)$, éstos pueden ser ignorados en (13) y de esta forma obtener el conjunto de funciones discriminantes que fueron implementadas en el presente artículo (14):

$$g_i(X) = -\frac{1}{2}(X - \mu_i)^T \Sigma_i^{-1}(X - \mu_i) + \ln P(\omega_i) \quad (14)$$

El término $(X - \mu_i)^T \Sigma_i^{-1}(X - \mu_i)$ se define como distancia cuadrática de Mahalanobis (Duda et al., 2001), de tal forma que la regla de decisión para agrupar un vector de características en una clase u otra es: Un vector de características X pertenece a la clase c cuando la distancia de Mahalanobis entre el vector y la media de c es la menor.

3.4 Resumen de las Pruebas e Implementaciones Realizadas

Las implementaciones efectuadas, para cada una de las técnicas presentadas en este artículo, son resumidas a continuación: Se consideraron las cinco vocales del castellano; el ancho de banda ω_1 fue modificado de 50 Hz hasta 600 Hz, en pasos de 50 Hz; el ancho de banda ω_2 fue probado para los valores de 900 Hz, 1000 Hz y 1100 Hz; el valor de m en (8) fue modificado para valores enteros consecutivos de 1 a 5; el valor de M en (8) fue modificado para valores enteros de 1 a 5; fueron implementadas dos versiones de TEO; y se eligieron los mejores resultados de la prueba de hipótesis para aplicarlos al clasificador cuadrático, considerando como características, los valores de las métricas por cada trama, la media y los cuartiles de dichos valores. Los resultados obtenidos son analizados en la siguiente sección

4. RESULTADOS Y DISCUSIÓN

Dentro de las pruebas realizadas, como se dijo anteriormente, fue implementado el TEO utilizado por Pruthi, donde $\omega_1 = 100 \text{ Hz}$, $\omega_2 = 1000 \text{ Hz}$, y de acuerdo con (9), $m = 1$ y $M = 1$. También se probó con el TEO formulado por Ying en (10), para diferentes valores de ancho de banda ω_1 y ω_2 ; adicionalmente, aplicando la forma generalizada del TEO, fueron variados tanto los anchos de banda como los valores del exponente m y el retraso M .

La Fig. 1, muestra cómo cambia el valor del estadístico Chi-Cuadrado del CTEO cuando se deja el exponente fijo en $m = 2$ y se aumenta el valor del retraso con $M = 1, 2, \dots, 5$. Los resultados mostrados, corresponden a valores diferentes de anchos de banda (ω_1, ω_2) para cada vocal. Esto es lógico si se considera que la estructura del primer formante es diferente para cada una de las vocales.

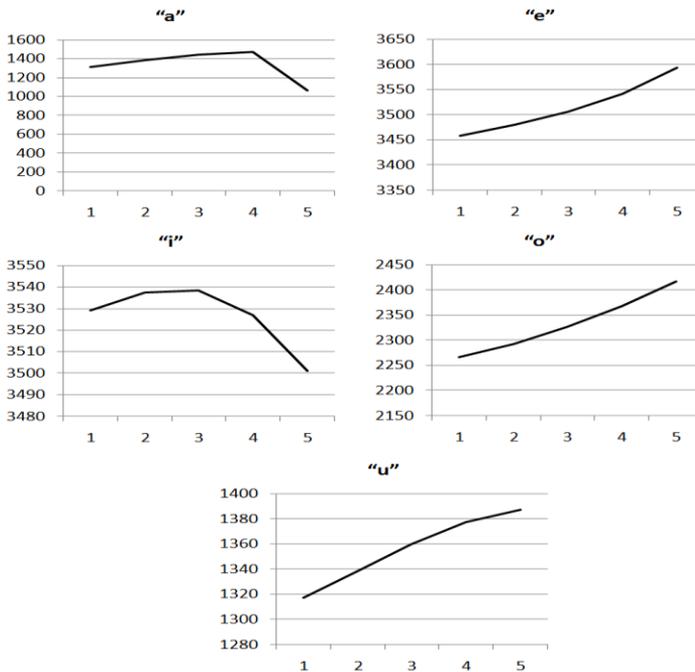


Fig. 1. Valores de Chi-Cuadrado para las vocales 'aeiou' cuando $m=2$ y $M=1, 2, \dots, 5$

La Tabla 1 especifica los anchos de banda implementados en los filtros para cada vocal presentada en la Fig. 1. Con el fin de analizar la incidencia del cambio tanto del exponente m , como del retraso M , se probaron diferentes combinaciones de valores para filtros con diferentes anchos de banda. Cada una de las líneas en la Fig. 2, muestra los valores más altos obtenidos del estadístico Chi-Cuadrado, y se resaltan los máximos valores para cada vocal. En el eje horizontal, se indican las combinaciones de exponente y retraso implementadas, de acuerdo con el formato: $CTEO_m_M$. Nótese que los primeros dos datos en el eje horizontal, corresponden con los obtenidos aplicando el método propuesto por Pruthi y el método propuesto por Ying, respectivamente.

Tabla 1. Valores de anchos de banda por vocal, implementados en los filtros utilizados para medir el CTEO mostrado en la Fig. 1

Vocal	BW1	BW2
A	600	1100
E	50	900
I	600	1100
O	50	900
U	600	1100

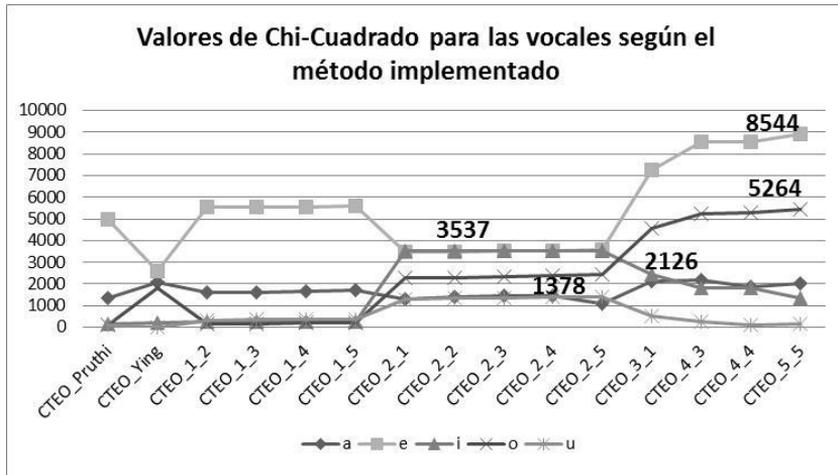


Fig. 2. Valores de Chi-Cuadrado para el CTEO con diferentes combinaciones de m y M

La Tabla 2 muestra cuáles fueron los valores de ancho de banda utilizados para obtener los máximos resaltados en la Fig. 2. Para la clasificación automática de voces sanas e hipernasales, se implementó un clasificador cuadrático, el cual considera poblaciones de diferente tamaño y diferente covarianza. Los resultados obtenidos, presentados en la Tabla 3, muestran el porcentaje de acierto en la clasificación y la desviación estándar obtenida luego de diez pruebas con cada combinación de exponente y retraso; dichas combinaciones fueron elegidas de acuerdo con la Tabla 2.

Tabla 2. Anchos de banda implementados para obtener los máximos valores de Chi-Cuadrado resaltados en la Fig. 2

Vocal	M	M	bw1	bw2	Chi-Cuadrado
a	4	3	25	450	2156
e	5	5	25	450	8919
i	2	3	300	550	3538
o	5	5	25	450	5434
u	2	5	300	550	1387

Tabla 3. Porcentajes de acierto en la clasificación de voces sanas e hipernasales, usando un clasificador cuadrático y las combinaciones de m y M de la Tabla 2

m	M	% de acierto	Desviación estándar
4	3	93,76%	0,58%
5	5	93,81%	0,61%
2	3	93,60%	0,54%
2	5	93,79%	0,54%
Pruthi		92,59%	0,51%
Ying		92,64%	0,61%

5. CONCLUSIONES

En el trabajo presentado, se evaluó la capacidad discriminante del TEO, aplicando la metodología propuesta por Cairns y complementada por Pruthi. En este caso, dicha metodología fue aplicada a la detección de hipernasalidad en voces reales de niños con LPH. Los resultados muestran que la capacidad discriminante del

CTEO puede aumentar a partir de la modificación de los parámetros de retraso M y de exponente m , en la definición generalizada de los perfiles de energía Teager (TEO). Los valores altos obtenidos para el estadístico Chi-cuadrado, permiten afirmar que las muestras provienen de dos distribuciones diferentes. Con base en esto, fue implementado un clasificador cuadrático para poblaciones con diferente covarianza y se obtuvieron porcentajes de acierto en la clasificación de alrededor del 93%.

Debido a que los resultados dados por Cairns habían sido obtenidos probando con voces hipernasales simuladas, y que los resultados de Pruthi habían sido obtenidos buscando nasalización y no hipernasalidad, era difícil afirmar de forma directa, que las metodologías propuestas por estos autores iban a tener buen desempeño en la clasificación de voces sanas e hipernasales reales. Los resultados presentados en este artículo, permiten sugerir que el CTEO es una métrica con gran poder de discriminación en voces con hipernasalidad y que debe ser tenida en cuenta en la implementación de sistemas multicomponente que busquen identificar dicha patología en la voz.

6. AGRADECIMIENTOS

Este trabajo es financiado por el centro de excelencia ARTICA, a través del proyecto N°1115-470-22055. Los autores agradecen a ARTICA, COLCIENCIAS, el Ministerio de TIC y la Clínica Noel de Medellín por su constante apoyo en el desarrollo de este proyecto. Así mismo, los autores agradecen al Comité para el Desarrollo de la Investigación (CODI), de la Universidad de Antioquia, por su apoyo a través del proyecto “Sistema de almacenamiento de historias fonoaudiológicas de pacientes con Labio y/o Paladar Hendido (LPH)”.

7. REFERENCIAS

Cairns, D.A., Hansen, J.H., Riski, J.E., (1994); Detection of hypernasal speech using a nonlinear operator, Proceedings of IEEE Conference on Engineering in Medicine and Biology Society, 253-254.

- Cairns, D.A., Hansen, J.H., Riski, J.E., (1996); A noninvasive technique for detecting hypernasal speech using a nonlinear, IEEE Transactions on Biomedical Engineering, 43(1), 35-45.
- Castellanos, G., (2005); Análisis acústico de voz y de posturas labiales en pacientes de 5 a 15 años con labio y/o paladar hendido corregido en la zona centro del país, Reporte de actividades, COLCIENCIAS.
- Castellanos, G., Daza, G., Sanchez, L., Castrillon, O., Suarez, J., (2006); Acoustic speech analysis for hypernasality detection in children, Proc. 28th Annual Int. Conf. of the IEEE Engineering in Medicine and Biology Society EMBS '06, 5507-5510.
- Chen, M.Y., (1996); Acoustic correlates of nasality in speech, PhD Thesis, Harvard-MIT division of Health Sciences and Technology, USA.
- Chen, M.Y., (1997); Acoustic correlates of English and French nasalized vowels, Journal of the Acoustical Society of America, 102(4), 2360-2370.
- Duda, R.O., Hart, P.E., Stork, D.G., (2001); Pattern Classification, Editorial John Wiley & Sons. 2a Edición, Canada.
- Fujimura, O., Lindqvist, J., (1971); Sweep-tone measurements of the vocal tract characteristics, Journal of the Acoustical Society of America, 49(2), 541-558.
- Glass, J.R., Zue, V.W., (1985); Detection of nasalized vowels in American English, Proc. Int Acoustics, Speech, and Signal Processing ICASSP, 1569-1572.
- Kaiser, J.F., (1990); On a simple algorithm to calculate the 'energy' of a signal, Proc. Int Acoustics, Speech, and Signal Processing ICASSP, 381-384.
- Kasuya, H., Kobayashi, Y., Kobayashi, T., (1983); Characteristics of pitch period and amplitude perturbations in pathologic voice, IEEE International Conference on Acoustics, Speech, and Signal Processing ICASSP, 1372-1375.
- Kasuya, H., Ogawa, S., Kikuchi, Y., (1986); An adaptive comb filtering method as applied to acoustic analysis of pathological voice, IEEE International Conference on Acoustics, Speech, and Signal Processing ICASSP, 669-672.
- Kvedalen, E., (2003); Signal processing using the Teager Energy Operator and other nonlinear operators, Master Thesis, Department of Informatics, University of Oslo, Norway.

- Michaelis, D., Gramss, T., Strube, H.W., (1997); Glottal to Noise Excitation Ratio - a new measure for describing pathological voices, *Acta Acustica*, 83, 700-706.
- Murthy, H.A., Madhu, K.V., Yegnanarayana, B., (1989); Formant extraction from phase using weighted group delay function, *Electronics Letters*, 25(23).
- NIST, (2010); NIST/SEMATECH e-Handbook of Statistical Methods, Disponible on-line en: <http://www.itl.nist.gov/div898/handbook/>
- Pruthi, T. (2005); Analysis, vocal-tract modeling and automatic detection of vowel nasalization, PhD Thesis, University of Maryland, USA.
- Pruthi, T., Espy-Wilson, C., (2004); Acoustic parameters for automatic detection of nasal manner, *Journal of Speech Communication*, 43(3), 225-239.
- Pruhi, T., Espy-Wilson, C., (2007); Acoustic Parameters for the Automatic Detection of Vowel Nasalization, *INTERSPEECH*, 1925-1928.
- Talkin, D., (1987); Speech formant trajectory estimation using dynamic programming with modulated transition costs, *Journal of the Acoustical Society of America*, 82(S1).
- Vijayalakshmi, P., Ramasubba, M., (2005); The Analysis on Band-Limited Hypernasal Speech Using Group Delay Based Formant Extraction Technique, *Proc. Interspeech. Conf*, 665-668.
- Vijayalakshmi, P., Ramasubba, M., O'Shaghnessy, D., (2007); Acoustic analysis and detection of hypernasality using a group delay function, *IEEE Transactions on biomedical engineering*, 54(4).
- Vijayalakshmi, P., Nagarajan, T., Jayanthan, R.V., (2009); Selective pole modification-based technique for the analysis and detection of hypernasality, *Proc. Of TENCON, IEEE Region 10*, 1-5.
- Ying, G.S., Mitchell, C.D., Jamieson, L.H., (1993); Endpoint detection of isolated utterances based on a modified Teager energy measurement, *Proc. IEEE Int Acoustics, Speech, and Signal Processing ICASSP*, 732-735.
- Yumoto, E., Gould, W.J., Baer, T., (1982); Harmonics to Noise Ratio as hoarseness index of degree of hoarseness, *Journal of the Acoustical Society of America*, 71(6).

