



Características acústicas en la identificación de la hipernasalidad de niños

Omar Castrillón Gómez

Germán Castellanos Domínguez

Genaro Daza Santacoloma

Universidad Nacional de Colombia. Grupo de Control y Procesamiento Digital de Señales.

Resumen

Se presenta el análisis de las diferentes características acústicas, y su influencia en la identificación automática de hipernasalidad. La metodología de selección efectiva de características, incluye el preproceso del espacio inicial de observaciones y está basada en el análisis de independencia estadística. De forma paralela, se propone la síntesis de una característica de diagnóstico especializado, basada en el análisis de la emisión acústica de la voz hipernasal. Como resultado se obtiene que, aunque las características acústicas permiten diferenciar con la suficiente precisión la patología, la característica propuesta con un nivel de complejidad computacional menor, no requiere muestras para entrenamiento y permite diferenciar los grados de compromiso de resonancia de la patología.

Palabras Claves: Características acústicas, hipernasalidad, resonancia, patología.

Introducción

En el tratamiento de niños con Labio y Paladar Hendido (LPH) corregido, se pueden presentar problemas de resonancia y emisión vocal, tales como: Hiponasalidad e Hipernasalidad. Sin embargo, de acuerdo al informe presentado en [1], se indica que es más frecuente encontrar casos con hipernasalidad (~90%), mientras la hiponasalidad ocurre en un (~10%). El interés en la detección de la hipernasalidad está relacionado con que su presencia indica problemas anatómico, neurológico y del sistema nervioso periférico [2]. La presencia de hipernasalidad, entendida como el escape de aire nasal y articulaciones compensatorias, conlleva a la baja inteligibilidad de la voz, la cual ocasiona un deterioro de la comunicación con su entorno que se manifiesta en cambios de actitud interpersonal y de comportamiento.

En el trastorno de aprendizaje velofaríngeo, la distorsión en la producción acústica conlleva a la voz nasalizada, además, por cuanto la pérdida de aire o escape nasal es masiva, se ven alterados los mecanismos articulatorios. El paciente no puede emitir de manera clara e inteligible y reemplaza su esfínter velopalatino por articulaciones glotales que

permiten una articulación más clara: /p/, /t/ /k/, /b/, /d/, /g/ provienen de golpes de glotis, mientras que los sonidos /ch/, /s/, /t/, /j/ son sustituidos por soplos roncós [12]. A pesar de que el paladar duro haya sido reparado quirúrgicamente, puede no proveer aún la competencia velofaríngea para una producción de habla normal. Incluso, si el paladar es potencialmente adecuado después de la cirugía, las conductas de habla previas pueden haber llevado al desarrollo de errores de articulación compensatoria o compensaciones fisiológicas cuyo fin es aproximarse a la inteligibilidad, aumentando así el número de patrones patológicos en el habla. Como resultado se tiene que las articulaciones compensatorias, generalmente, persisten después del manejo postónico o posquirúrgico que ha previsto un cierre adecuado y deben ser corregidos antes de elevar el funcionamiento del esfínter velofaríngeo mediante terapia del lenguaje.

En los últimos años ha crecido el interés por el análisis acústico de voz (AAV) como un método alternativo para el diagnóstico y tratamiento en la identificación de desórdenes funcionales en voces de niños [3,4,5]. Este tipo de análisis demuestra grandes ventajas sobre los métodos tradicionales debido a su naturaleza no invasiva y a su potencial

para proveer una medida cuantitativa acerca del estado clínico del funcionamiento de la laringe y el tracto vocal. Las características acústicas o parámetros objetivos son frecuentemente usadas para representar la voz patológica sobre vocales sostenidas [6,7,8]. Sin embargo, estos vectores están limitados en su robusticidad debido a la dificultad de estimación en condiciones reales de perturbaciones con estructura no estacionaria [9]. Aunque se han realizado numerosos análisis de efectividad de los diferentes tipos de características propuestas para la evaluación objetiva de desórdenes de voz [6,13], de ellos no se puede tomar un conjunto estándar de parámetros para la identificación de la hipernasalidad, debido a que cada desorden afecta de diferente manera diversos aspectos de la emisión de voz.

En el presente trabajo se analiza la efectividad estadística de las diferentes características acústicas en la identificación automática de la hipernasalidad. Las características acústicas reflejan parte de la información contenida en el análisis perceptual, en parte debido a que su estimación es derivada, directamente o indirectamente, del comportamiento de las cuerdas vocales. Por este motivo, es conveniente el empleo de técnicas de análisis multivariado en la determinación de la efectividad de las características de voz. La efectividad se estudia empleando técnicas de análisis multivariado MANOVA en un enfoque heurístico de selección de características para el entrenamiento de los clasificadores. Adicionalmente, se analiza el empleo de las transformada wavelet para la estimación de calidad de emisión vocal.

Materiales y Métodos

Nasalización y emisión nasal

La Nasalización se define como la comunicación que existe entre la cavidad nasal y el resto del tracto vocal; mientras que emisión nasal se refiere al escape anormal de aire por la ruta nasal. Este escape anormal reduce la presión intraoral causando distorsión en las consonantes. Cuando el escape de aire resulta en un resoplo audible, la emisión nasal es más obstruiva y el habla es seriamente afectada. La nasalidad, comúnmente llamada hipernasalidad, se refiere a la baja calidad de voz, que resulta de la adición inapropiada del sistema de resonancia nasal al tracto vocal. En contraste a la emisión nasal, la nasalidad no envuelve grandes flujos de aire nasal, por lo que no hay cambios significativos en la presión de aire intraoral. Para la identificación de esta patología se han propuesto principalmente dos métodos, el primero basado en el modelamiento de la señal (diagnóstico especializado) y el segundo fundamentado en técnicas de reconocimiento de patrones.

Característica de diagnóstico especializado

Considerando que la voz normal está compuesta por resonancias en varias frecuencias o formantes F_k , se propone el siguiente modelo acústico [2]:

$$S_n(w) = \sum_{k=1}^K F_k(w) \quad (1)$$

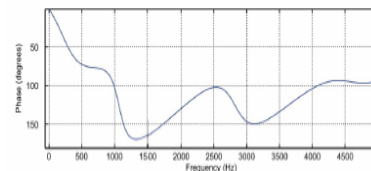
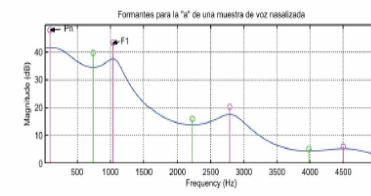
En contraste a la voz normal, la voz nasalizada es la aparición de los antifonantes F y formantes nasales F :

$$S_h(w) = \sum_{k=1}^K F_k(w) - \sum_{l=1}^L \hat{F}_l(w) + \sum_{m=1}^M \hat{F}_m(w) \quad (2)$$

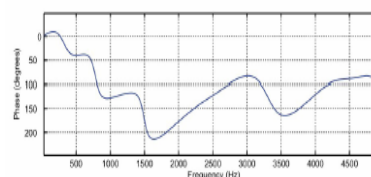
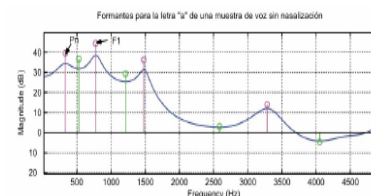
Se ha sugerido que la intensidad en la reducción del primer formante es un índice primario de nasalidad. En [2], se filtran los formantes superiores, de tal manera que la voz normal filtrada tendrá un solo tiene una componente, mientras la voz nasalizada corresponderá a una señal con múltiples componentes, los cuales pueden ser estimados a través del operador de potencia instantánea de Teager.

En forma general, en los pacientes con LPH, la nasalidad se manifiesta en una insuficiencia del puerto velofaríngeo, acoplado la cavidad nasal a sonidos vocálicos, lo cual da lugar a la aparición de una resonancia adicional en la característica amplitud-frecuencia del tracto vocal, y una disminución notable de los formantes F_1 y F_2 [15] como se puede ver en la Figuras 1(a) y 1(b).

Figura 1: Energía del Pitch para la distinción del golpe glótico.



(a) Energía del Pitch (caso patológico)



(b) Energía del Pitch (caso normal)

Características acústicas

Se agrupan en dos categorías de acuerdo a las propiedades acústicas que deben medir: las basadas en el ruido aditivo, entre ellas están, HNR (*Harmonic to Noise Ratio*) que corresponde al promedio de la componente de ruido de la emisión vocal; GNE (*Glottal Noise Excitation*) definida como la estimación del ruido y basada en la presunción de que los pulsos glóticos resultantes de la colisión de los pliegues vocales conllevan a una excitación sincrónica en las diferentes bandas de frecuencia y NEP (*Normalized Error prediction*), que puede expresarse como la relación de las medias geométrica y aritmética del modelo espectral.

Otras características acústicas se asocian al ruido de modulación en frecuencia, entre ellas, el Pitch o periodo fundamental de la señal y el Jitter definido como el porcentaje de variación promedio entre dos valores de Pitch consecutivos. De otra parte, se tienen las características asociadas a modelos paramétricos de generación de voz, entre ellos, los coeficientes cepstrum derivados del análisis de predicción lineal (LPC *Linear Prediction Coefficients*), los coeficientes cepstrum calculados sobre la escala ponderada de frecuencias (MFCC *Mel-Frequency Cepstrum Coefficients*) y los coeficientes RASTA (*Relative Spectral Transform*) [14].

Marco experimental

Base de datos

Corresponde a una muestra, que evalúa 90 niños que corresponde a las clases normal y con hipernasalidad (45 pacientes por cada clase), valorados por el especialista. Cada grabación está conformada por cinco palabras seleccionadas: /coco/, /gato/, /jugo/, /mano/ y /papá/. Las señales fueron tomadas en condiciones de bajo nivel de ruido ambiental usando un micrófono dinámico unidireccional (cardioid). El rango de todas las señales de voz adquiridas está entre (-1,1).

Espacio Inicial de Características

El conjunto total de características acústicas de voz consideradas son: Pitch (F_0), Jitter, Jitter porcentual (J_p), Coeficiente de Perturbación de Tono, NEP, GNE, HNR, Energía, Cruces por cero, Cepstrales, LPC y MFCC.

Dada una característica x , a la cual le corresponde el vector $\mathbf{x}$ formado por las observaciones $\{x_i; i=1, \dots, n\}$ de medida para cada una de las k clases, se realiza la estimación de los momentos en calidad de valores de descripción de la estructura de aleatoriedad de las distribuciones empíricas. Los momentos de análisis por cada característica x son los siguientes:

1. *Parámetros de posición.* El momento inicial de primer

orden (valor medio $\bar{x}$) se estima a efectos de ser removido de la observación, esto es, $x - \bar{x}$, en la medida en que se considere que este no tiene información y pueda generar un sesgo inapropiado en la interpretación de los resultados. En la práctica existen otras estimaciones del valor medio, por ejemplo la mediana.

2. *Parámetros de escala.* Se consideran los siguientes: la varianza, el valor cuadrático medio, el coeficiente de variación $\bar{x}/s_x$, el valor pico a pico y la desviación de mediana absoluta $\text{med}|x_i - \bar{x}|$. Además, se consideran los momentos centralizados de orden $l=3, \dots, 6$.

3. *Parámetros de forma.* Corresponden a los coeficientes de asimetría (oblicuidad) y exceso (curtosis). Además de los anteriores momentos, se pueden considerar los cumulantes de orden $m=1, \dots, 4$.

En total se analizan 128 variables de representación por cada palabra. Por lo tanto, el espacio inicial de características de los clasificadores se incrementa hasta 640 características de voz.

Preproceso de datos

Tiene como objetivo la disminución de la influencia, y en lo posible eliminación de los errores de medida ocasionados, entre otras razones, por errores sistemáticos en el registro de las señales, fallas ocasionales en los dispositivos de medida, etc., así como el control de la homogeneidad de las propiedades estadísticas de las diferentes características del fenómeno aleatorio en análisis. El preproceso de los datos consiste del análisis de registros anómalos de cada característica y la verificación de normalidad de cada una de ellas.

Selección efectiva de características

La metodología propuesta de reducción de los espacios iniciales, está basada en el análisis de *relevancia fuerte*, que estudia la correlación entre las características. La metodología de reducción contempla la combinación de las técnicas heurísticas (DFCC) [14], para la generación de características, y de las técnicas estadísticas de análisis multivariado, para la evaluación de las mismas. Particularmente, como función de costo intrínseca de la técnica heurística se emplea el análisis de varianza de dimensión múltiple (MANOVA), a efecto de encontrar un grupo de características que genere una mayor discriminancia entre clases.

Clasificación

El clasificador que se emplea en el trabajo es bayesiano, se utilizan 5 clasificadores de éste tipo, cada uno de ellos con el fin de analizar el error entre clases (hipernasal y control) para cada una de las palabras previamente mencionadas. Además, se realizó validación cruzada para ob-

servar la variación en los parámetros del clasificador y su capacidad de generalización.

Resultados y discusión

Análisis de características acústicas

La efectividad de las características acústicas se mide de acuerdo al rendimiento de clasificación, el clasificador empleado en este trabajo es bayesiano. A continuación se presentan los resultados para cada una de las etapas.

El procedimiento de detección de valores anómalos se realiza por cada una de las características x , permite hacer clara la calidad de medición de cada registro de las bioseñales. En el caso de las características acústicas los resultados se muestran en la Tabla 1. Aunque la reducción en promedio para el conjunto de palabras es ~25 al 30% el clasificador no converge, debido a que la dimensión del espacio de entrenamiento es desmesurado y por tanto las matrices de covarianza tienden a ser singulares.

Tabla 1

Palabra de análisis	Dimensión espacio reducido [%]	Rendimiento de clasificación
Coco	70.3	No converge
Gato	71.9	No converge
Jugo	61.7	No converge
Mano	71.1	No converge
Papá	68	No converge

En cuanto a la verificación de normalidad, el juicio sobre la estructura gaussiana de los datos se puede realizar a partir de la respectiva prueba de hipótesis. En el trabajo se emplea la prueba de Kolmogorov-Smirnov.

Tabla 2

Palabra de análisis	Dimensión espacio reducido [%]	Rendimiento de clasificación
Coco	61.7	No converge
Gato	64.8	No converge
Jugo	53.1	No converge
Mano	64.8	No converge
Papá	60.9	No converge

Si la prueba de verificación de la distribución da como resultado el rechazo de la hipótesis de normalidad, entonces se debe tomar las medidas para transformar la observación de tal manera que pueda cumplirse aceptarse la hipótesis sobre la normalidad de los datos, particularmente se empleó la técnica de *Box-Cox*. Los resultados de reducción de dimensión y rendimiento de clasificación para la prueba de verificación de normalidad se muestran

en la Tabla 2.

La reducción de dimensión en término porcentuales obtenida en paso es similar a la que se consiguió con la eliminación de características anómalas, sin embargo no es suficiente para la convergencia del clasificador, debido a que la dimensión del espacio de características sigue siendo muy elevada para el cálculo de las matrices de covarianza bases para el clasificador bayesiano. Además, otro problema que se aprecia del espacio de variables está en la diferencia considerable de rangos de medida (10^{-3} , 10^6 , 10^{15}) que aún exhiben las características acústicas. El principal propósito de la selección efectiva es encontrar aquellas características que permiten una mayor discriminancia entre clases, que en este caso se realiza a través del uso de una técnica heurística de crecimiento tipo filtro con función de costo multivariada (MANOVA). El procedimiento de selección efectiva de características se realiza para dos casos de análisis (Tabla 3.): a) Sin aplicar preproceso a los datos, b) Luego de aplicar las etapas de preproceso de los datos. Los respectivos resultados muestran que aunque el porcentaje de reducción pueda considerarse de un mismo orden para ambos casos de análisis, el rendimiento del clasificador es sustancialmente diferente y mucho mejor en el caso en que se utiliza el preproceso de datos.

Tabla 3

	espacio sin pre-proceso	Rend sin pre-proceso	espacio con pre-proceso	Rend con pre-proceso
Coco	3.9	83.3	3.1	99.9
Gato	9.3	88.9	3.1	99.9
Jugo	7	91.1	3.1	99.9
Mano	6.3	93.3	3.1	99.9
Papá	3.9	91.1	3.1	99.9

El análisis de la Tabla 4, la cual muestra las características seleccionadas para cada una de las palabras, da como resultado que las características efectivas están relacionadas directamente con los siguientes parámetros acústicos: GNE el cual participa en todas las palabras y definida como una forma de estimación del ruido, HNR que corresponde al promedio de la componente de ruido de la emisión vocal, el NEP que puede expresarse como la relación de las medias geométrica y aritmética del modelo espectral, estas características se encuentran asociadas a las mediciones de ruido aditivo, lo cual puede relacionarse con las componentes adicionales en caso de hipernasalidad descritas en el modelo matemático mostrado en (2). Otro tipo de características presentes son las características asociadas a modelos paramétricos de generación de voz (LPC y Coeficientes Cepstrales), particularmente a los modelos ARMA, los cuales también son excitados por ruido. En este caso los coeficientes de predicción lineal (LPC) que mayor influencia tienen son los primeros, por ser los de mayor amplitud y por tanto menos sensibles a la variabilidad de medida.

Tabla 4

Palabra de análisis	Características efectivas seleccionadas
Coco	Promedio del GNE, mediana del GNE, desviación estándar de los coeficientes Cepstrales, 3er momento centralizado de los coeficientes Cepstrales.
Gato	Media de la matriz de covarianza del GNE, desviación estándar del HNR, kurtosis de los coeficientes cepstrales, el sexto coeficiente cepstral
Jugo	NEP, máximo GNE, LPC 1, LPC 2
Mano	Desviación estándar del GNE, varianza del GNE, LPC 3, LPC 5
Papá	Máximo GNE, desviación estándar del GNE, varianza de los coeficientes cep-

Los resultados del estudio de concordancia entre los evaluadores y el sistema de reconocimiento automático se presentan en la tabla 5, de la cual se puede ver que el desempeño del sistema es altamente significativo. Las pruebas se hicieron mediante el procedimiento de Jackknife para un nivel de confianza del 95%.

Tabla 5

Concordancia entre especialistas (fonoaudiólogos) y el sistema de reconocimiento automático [%]	
	Sistema de reconocimiento automático
Especialista 1	98
Especialista 2	97.7
Especialista 3	86.2
Especialista 4	88.7

Característica de diagnóstico especializado

En este caso los resultados de la evaluación se presentan en la tabla 6. El estudio de concordancia realizado en igualdad de condiciones mostró una efectividad del 69%

Tabla 6

	Sin nasalidad [%]	Nasalidad [%]		
		Leve	Moderada	Severa
Pacientes control	3.1	30.1	59.8	7
Pacientes patológicos	2.1	65.1	28.1	4.7

Conclusiones

Las características acústicas permiten diferenciar con la suficiente precisión la patología. Uno de los principales problemas en el agrupamiento de las características de voz para el análisis de hipernasalidad, es que muchos de ellas son sensibles a varias propiedades acústicas. Esta mutua dependencia puede ser una de las razones para su

interpretación sea aparentemente contradictoria en los diferentes resultados encontrados en la literatura. Por lo tanto, se deben identificar los patrones de articulación de fonemas velares (detección de posturas labiales) y emisión vocal (análisis acústico) con el suficiente poder discriminante para su reconocimiento.

En el caso de la detección de hipernasalidad es preferible el diseño de características de diagnóstico especializado, que reflejen la naturaleza concreta de irregularidad en el estado funcional de la voz. En el caso particular se propone una característica de diagnóstico especializado que ofrece un nivel de detección similar a las demás características acústicas tratadas, pero con un nivel de complejidad computacional menor (se calcula una única característica), no requiere muestras para entrenamiento y además permite diferenciar los grados de compromiso de resonancia de la patología.

La detección de los grados en la hipernasalidad es importante ya que tanto en el diagnóstico como en el tratamiento y cirugía, es necesario para el especialista clínico disponer de un sistema taxonómico de referencia que pueda servir para relacionar los diferentes grados de compromiso de resonancia respecto a los patrones velofaríngeos y que a su vez que permita generar diversos tipos de alternativas de tratamiento y excluir otras.

Agradecimientos

El actual trabajo se hace en el marco del proyecto “Acústica de labio y paladar hendido en la zona centro del país” financiado por la Universidad de Caldas y la Universidad Nacional de Colombia. De igual manera, se agradece la colaboración de la doctora Colombia Quintero, de las fonoaudiólogas Beatriz Mejía y Ana Maria Escandón, así como del epidemiólogo Arnoby Chacón, quienes colaboraron en la experticia de la base de datos de voz, empleada en el presente trabajo.

Bibliografía

1. G. Castellanos, F. Prieto, C. Quintero. 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. *Informe Colciencias*. Diciembre 2004.
2. D. A. Cairns, J. H. L. Hansen, and J. E. Riski. A noninvasive technique for detecting hipernasal speech using a nonlinear operator. *IEEE Transaction on Biomedical Engineering*, 1996.
3. G. Niedzielska. Acoustic analysis in the diagnosis of voice disorders in children. *International Journal in Pediatric Otorhinolaryngology*, 2000.
4. J. González, T. Cervera, and J. L. Miralles. Análisis acústico de la voz: Fiabilidad de un conjunto de parámetros multidimensionales. *Acta Otorrinolaringol*, 2002.

- 5. G. Niedzielska, E. Glijer, and A. Niedzielski.** Acoustic analysis of voice in children with noduli vocales. *International Journal in Pediatric Otorhinolaryngology*, 2001.
- 6. P. Yu, M. Ouaknine, J. Revis, and A. Giovanni.** Objective voice analysis for dysphonic patients: A multiparametric protocol including acoustic and aerodynamic measurements. *Journal of Voice*, 2001.
- 7. S. Hadjitodorov and P. Mitev.** A computer system for acoustic analysis of pathological voices and laryngeal diseases screening. *Medical Engineering and Physics*, 2002.
- 8. R. D. Kent, H. Vorperian, and J. Kent.** Voice dysfunction in dysarthria: application of the multi-dimensional voice program. *Journal of communication Disorders*, 2003.
- 9. M. Gupta and A. Gilbert.** Robust speech recognition using wavelet coefficients features. 2001.
- 10. I. Gavat, M. Zirra, and V. Enescu.** Pitch detection speech by dyadic wavelet transform, 1997.
- 11. Adriane Parraga.** APLICACIAO da Transformada Wavelet Packetna Análise e CLASSIFICACIAO de sinais de vozes patológicas, Ph.D. thesis, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2002.
- 12. A. Habbaby.** Enfoque integral del niño con fisura labiopalatina. Buenos Aires: *Médica Panamericana*, 2002.
- 13. Mathias Frohlich. and Dirk Michaelis.** Acoustic voice analysis by mean of the hoarseness diagram, *Journal of Speech, Language and hearing research*, vol. 3, no. 43, pp. 706, 2000.
- 14. G. Castellanos, O. D. Castrillón, and E. Guijarro.** Multivariate analysis techniques for effective feature selection in voice pathologies, CASEIB, 2004.
- 15. Baken R. J.** Clinical Measurement of Speech and Voice, 401 West A Street, Suite 325, San Diego, California 92101-7904, 1996.
- 16. Mathias Frohlich. and Dirk Michaelis.** Acoustic voice analysis by mean of the hoarseness diagram. *Journal of Speech, Language and hearing research*, vol. 3, no. 43, pp. 706, 2000.

Recibido el 24 de Febrero del 2006.
 Aceptado el 17 de Abril del 2006.
 Publicado (on-line) 1 de Diciembre del 2006.
<http://www.auditio.com/revista>

Contacto con los autores:

O. Castrillón Gómez.
 E-mail: odcg@Walla.com

G. Castellanos Domínguez.
 E-mail: gcastell@telesat.com.co

G. Daza Santacoloma.
 E-mail: genarodaza@ieee.org

Para citar este artículo:

O. Castrillón Gómez, G. Castellanos Domínguez, G. Daza Santacoloma (2006). Características acústicas en la identificación de hipernasalidad en niños. [en línea]. *Auditio: Revista electrónica de audiolología*. 1 Diciembre 2006, vol. 3(2), pp. 43-48. <<http://www.auditio.com/revista/pdf/vol3/2/030203.pdf>>