

ADAPTACIÓN DEL DIAGRAMA DE BLOQUES DE UN ALGORITMO PARA EXTRACCIÓN DE LA FRECUENCIA FUNDAMENTAL EN LLANTO INFANTIL

Ing. Daniel I. Escobedo Beceiro, profesor e investigador auxiliar, daniele@fie.uo.edu.cu

Ing. Lic. Frank Sanabria Macias, profesor asistente

Dr. C. Sergio D. Cano Ortiz, profesor titular

Dr. C. Enrique J. Marañón Reyes, profesor titular

Centro de Estudios de Neurociencias, Procesamiento de Imágenes y Señales (CENPIS)

Grupo de Procesamiento de Voz. Facultad de Ingeniería Eléctrica, Universidad de Oriente.

Teléfono: (53-22) 646141. Ave. Las Américas s/n. CP 90900. Santiago de Cuba.

RESUMEN

Se presentan adecuaciones al diagrama de bloques de un algoritmo clásico *SIFT: Simple Inverse Filter Tracking* usado para estimar la Frecuencia Fundamental (Fo) del habla. Las adecuaciones permiten obtener Fo en la señal del llanto provocado en recién nacidos. Estas vocalizaciones tienen Fo promedio (normalmente entre 400 y 600 hertz) superior al rango del habla. El procedimiento empleado cambia la frecuencia de corte del filtro inicial del diagrama del algoritmo clásico adaptándose a los requerimientos del llanto. El algoritmo clásico decide automáticamente las tramas sordas y sonoras de donde obtiene Fo; las adecuaciones aprovechan la selección manual de las tramas sonoras para obtener Fo. Esta aparente desventaja hace más eficaz la selección de tramas sonoras. En el trabajo, el enventanado de la señal no tiene solapamiento, tampoco diezmado, evitando submuestreo de la señal y su afectación espectral. Estas modificaciones junto a técnicas de corta duración hacen efectiva la adaptación.

Palabras Claves: frecuencia fundamental en llanto infantil, estimación de tono, algoritmo SIFT.

ABSTRACT

The paper shows some adaptation to SIFT (Simple Inverse Filter Tracking) block diagram. SIFT is a classic algorithm used for speech pitch (Fo) estimation. The adaptation allows us the pitch estimation in pain cry signal from neonates. These cry vocalizations have a normal pitch between 400 and 600 hertz. Cry pitch range is more than speech. The implementation of this paper changes the cut off frequency from prefiltering in classic block diagram according to cry signal. Classic algorithm makes automatic decision about the unvoiced and voiced frames where the pitch is estimated. In the paper, the pitch is estimated from voiced frames through manual procedure. That apparent disadvantage allows an efficient selection of voiced frames. The procedure used considers signal windowing without overlap. The procedure also exclude decimate avoiding downsampling and affectation of signal spectrum. The modification achieved together short duration technique make an efficient adaptation for cry pitch estimation.

Key Words: *fundamental frequency in infant cry, pitch estimation, SIFT algorithm.*

INTRODUCCIÓN

Los Sistemas Subglotal y Supraglotal, y la Laringe intervienen en la producción de voz. El Sistema Subglotal está compuesto por los pulmones y la musculatura respiratoria asociada; la tráquea es la que conduce el aire de los pulmones a la laringe, en este sistema se genera el flujo de aire para producir la voz. La Laringe convierte el flujo cuasi-estacionario de aire de los pulmones en una serie periódica o cuasi-periódica de golpes de aire, usando en ello las cuerdas vocales, este golpe de aire es el **Pulso de Glotis**, cuya frecuencia es la **Frecuencia Fundamental** o **Tono** (F_0), dado en hertz (Hz) [1], [2]. El rango de valores de F_0 en hombres comprende desde 80 Hz hasta alrededor de 250 Hz, y en mujeres y niños, de 200 Hz ó 250 Hz hasta valores superiores. El Sistema Supraglotal lo forman faringe, boca, los músculos de la región palatina y del maxilar inferior, lengua, labios y, además de esta vía oral, la vía nasal. En la producción de la voz las cavidades supralaríngeas se comportan como un filtro variable de frecuencias capaz de transformar la energía acústica. Los sonidos del lenguaje o **fonemas** generalmente pueden agruparse en **sonoros** y **sordos**. En la fonación de sonidos sonoros, las cuerdas vocales se mueven en respuesta a una alternación rápida de fuerzas aerostáticas y aerodinámicas generadas por el flujo de aire pulmonar. En la formación del tren de impulsos acústicos que sucede en la laringe, influye la tensión de los ligamentos vocálicos, la movilidad y masa de éstos y su longitud, así como la presión ejercida sobre dichas cuerdas desde los pulmones. En lenguaje normal, un sonido sonoro se origina en la laringe, y es el resultado de una abertura y oclusión regulares de la glotis por medio de las cuerdas vocales.

Los sonidos sordos (no vocálicos o no sonoros, en los que no puede estimarse F_0), son sonidos constituidos por ruido o turbulencias, los que se producen con una cantidad de aire espirado mucho mayor que los sonidos vocálicos, estos sonidos sordos se alcanzan cuando hay exigencia de un nivel de aire turbulento para su producción. Esto ocurre en ciertos sonidos o fonemas de una lengua, algunos consonánticos, dichos sonidos se producen mayormente en el tracto supralaríngeo que es donde se transforma sustancialmente el flujo de aire glotal proveniente de los pulmones y del tracto laríngeo, para así dar forma a los sonidos del habla, agregándoseles, en este Sistema Supralaríngeo la mayoría de las características que hacen a los sonidos entendibles como fonemas de una lengua, donde coexisten sonidos sonoros, sordos y de sonoridad intermedia. Así, con la **articulación** como actividad coordinada del cerebro, demás sistemas componentes funcionales y los músculos del Sistema Supraglotal, es que en sincronización, se producen los sonidos del habla, con sus atributos acústicos, para la comunicación.

Llanto Infantil

El llanto provocado ante un estímulo de dolor (pellizco en el calcáneo en este trabajo) de un niño recién nacido tiene, por su naturaleza, diferencias significativas en sus características acústicas en comparación con dichas características en el lenguaje articulado. El neonato emite prevocalizaciones o vocalizaciones con Frecuencia Fundamental en el rango de 400 Hz a 600 Hz [3-7] en los llantos de niños normales, y generalmente por encima de 650 Hz en casos patológicos, sobre todo en Asfixia y daños en el Sistema Nervioso Central, entre otros. También es característico del llanto, aún en casos normales, mucho más en patológicos, la presencia de desplazamientos ascendentes del tono [3], [5], [8], variaciones de duración e intensidad, entre otros atributos cuantitativos y cualitativos del llanto que se alteran ante la presencia de estados patológicos, atributos que no son tratados en este trabajo.

En nuestro trabajo el fundamento para proponer las adecuaciones al diagrama de SIFT clásico para usarlo en el llanto infantil se sustenta en la teoría acústica de voz y lenguaje (Fant, 1960; Stevens, 1964; Flanagan, 1972) [9-12], y en la analogía entre el **Modelo General de Producción de Voz** y el **Modelo Físioacústico del Llanto Infantil de H. Golub** [13-15]. En ello debe recordarse que dos factores básicos están presentes en la producción de voz y lenguaje: el punto o zona de origen del sonido y el filtro [1], o sea, la salida de voz y lenguaje puede representarse por el producto de la Fuente por el Filtro. Es decir, una Fuente con excitación sonora o aleatoria (ruido) y un Filtro lineal polo-cero variable en el tiempo en el que puede simularse el comportamiento del tracto vocal. Así, a través de la **Teoría de Predicción Lineal (LPC: Linear Predictive Code)** [9], [12], [16], [17], puede obtenerse un modelo que conserve la correspondencia entre los modelos de Producción de Voz y Lenguaje y Físioacústico del Llanto Infantil. Este modelo es el de un Codificador LPC como el de la figura 1.

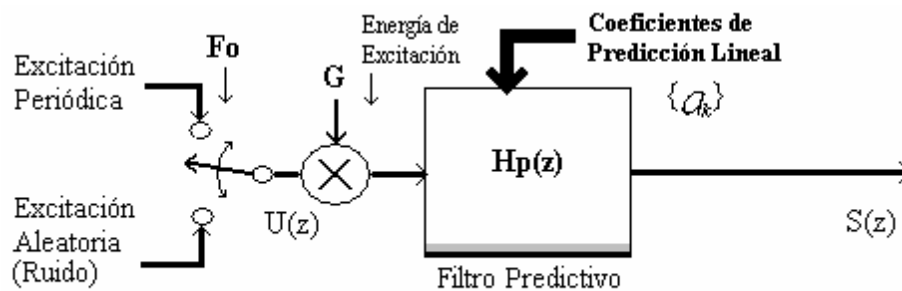


Figura 1. Modelo de un Codificador LPC

La función de transferencia del filtro digital de dicho modelo puede expresarse por:

$$\frac{S(z)}{U(z)} = \frac{G}{1 - \sum_{k=1}^p a_k z^{-k}} \quad (1)$$

Donde G es la ganancia de la Energía de Excitación. Este Modelo de Codificador LPC es derivado del modelo general de producción de voz donde el filtro lineal (Filtro Predictivo) $H_p(z)$ es todo polo, de orden p (orden del predictor), cuyos coeficientes $\{a_k\}$ varían en el tiempo. Los parámetros de este filtro lineal se asumen constantes dentro de una ventana de corta duración, muchos de estos parámetros se obtienen aplicando **Técnicas de Corta Duración** [9], [18], [19]; dicho filtro, es capaz de modelar la función de transferencia del tracto vocal, y si p es de valor alto, también modela la forma del pulso de glotis y los efectos de la radiación labial.

El cálculo de los Coeficientes LPC en el Modelo de Codificador presentado usando el **Método de Autocorrelación** [9], [12], [16], [20], permite obtener el **Tono** de la señal de llanto infantil del diagrama de bloques de nuestro trabajo utilizando el filtrado inverso y la detección del primer pico del Período Fundamental (T_0) en la Función de Autocorrelación de la Onda Residual [21-23].

Algoritmo SIFT

La figura 2 presenta el diagrama de bloques del algoritmo SIFT (Markel, 1972; Veprek y Scordilis, 2002) [22], [23]. En el diagrama la señal objeto de estimación del tono, es decir la señal de habla, pasa por un filtro digital (*FIR: Finite Impulse Response*) pasa bajos de frecuencia de corte en 800 Hz, además, dicha señal es diezmada, o sea, submuestreada por un factor igual a 5. Luego sigue un análisis de predicción lineal de cuarto orden en tramas o segmentos de señal de 32 milisegundos (ms). Cada trama es enventanada con una ventana del tipo Hamming y preenfatzada. Después del filtrado inverso en el diagrama y aplicar autocorrelación normalizada se obtiene la función de autocorrelación de la onda residual, así, si la señal rebasa determinado

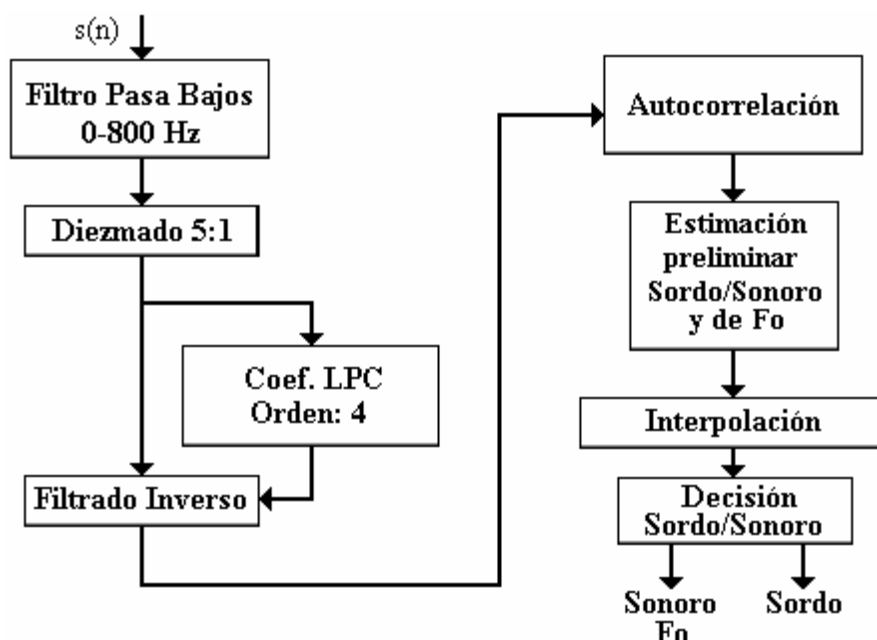


Figura 2. Diagrama de Bloques del Algoritmo SIFT

valor umbral de intensidad esta trama o segmento es considerada sonora, también se producen comparaciones respecto a un valor umbral a los segmentos anteriores y posteriores al segmento de análisis y si se rebasa el valor umbral en éstos son tramas sonoras, si no es así son tramas sordas. La estimación de la Frecuencia Fundamental es sólo en las tramas sonoras. Para incrementar la resolución en la estimación de tono, la función de autocorrelación en cada trama sonora fue interpolada en la vecindad del pico preliminarmente calculado usando una interpolación parabólica. De esta forma, el inverso del período fundamental, dado la periodicidad por el primer pico normalizado de la Función de Autocorrelación de la Onda Residual, nos da el valor de la Frecuencia Fundamental.

Algoritmo SIFT con adecuaciones

La figura 3 muestra el diagrama de bloques del algoritmo SIFT con algunas adecuaciones para estimar la Frecuencia Fundamental en el llanto infantil, provocado ante dolor, en niños de 0 a 28 días de nacidos.

La señal de llanto llega al filtro FIR tipo I (simétrico, sin distorsión de fase), pasa bajos de frecuencia de corte igual a 3500 Hz. El valor de la frecuencia de corte responde a contener en el espectro el rango de frecuencias de F_0 del llanto para casos normales y patológicos, así como sus desplazamientos de frecuencia [8] y también los deslizamientos (*Gliding*) [3], [5], que son vistos principalmente en casos patológicos y son los incrementos del tono hasta valores de 600 Hz en 100 ms. Este filtro pasa bajos preserva la riqueza espectral para la extracción de F_0 en las fonaciones.

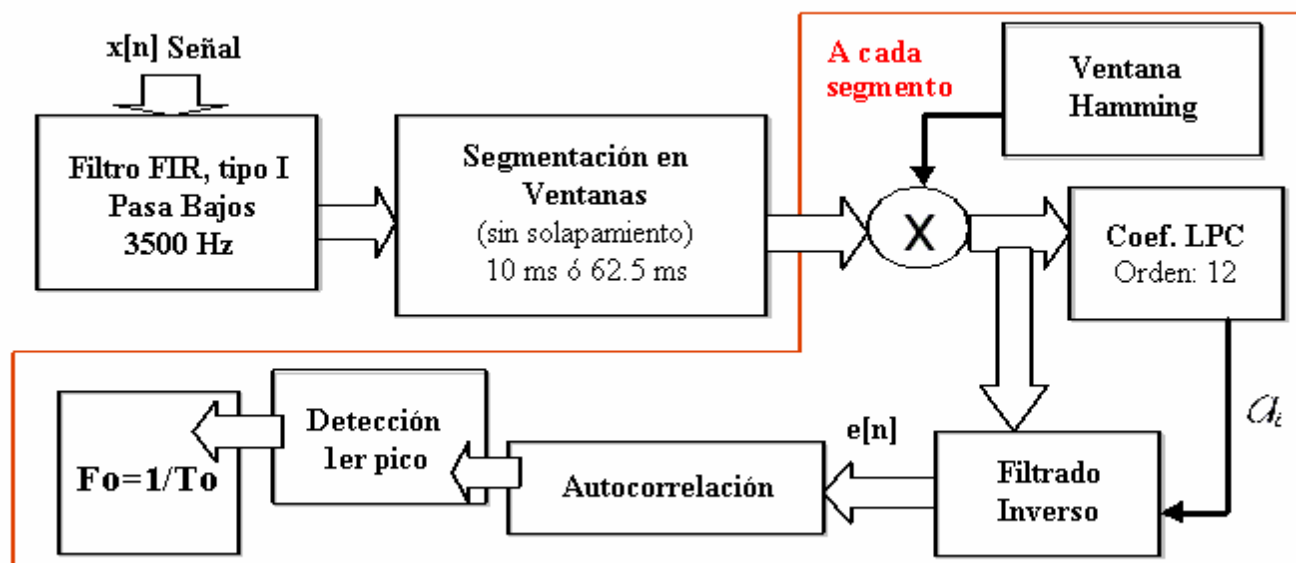


Figura 3. Diagrama de Bloques del Algoritmo SIFT con adecuaciones

En nuestro diagrama de bloques no hay diezmado, o sea no hay submuestreo ya que no hay problemas de costo computacional, en la actualidad las computadoras garantizan rapidez y potencia de cálculo. Sigue en el

diagrama la segmentación manual. En el bloque de segmentación existen dos ventanas de diferente duración (10 ms ó 62.5 ms) que pueden usarse indistintamente; entre las ventanas no hay solapamiento, todo lo antes explicado garantiza que haya una menor afectación espectral a la señal de llanto objeto de investigación, afectación que si se produce en una mayor medida en el algoritmo original.

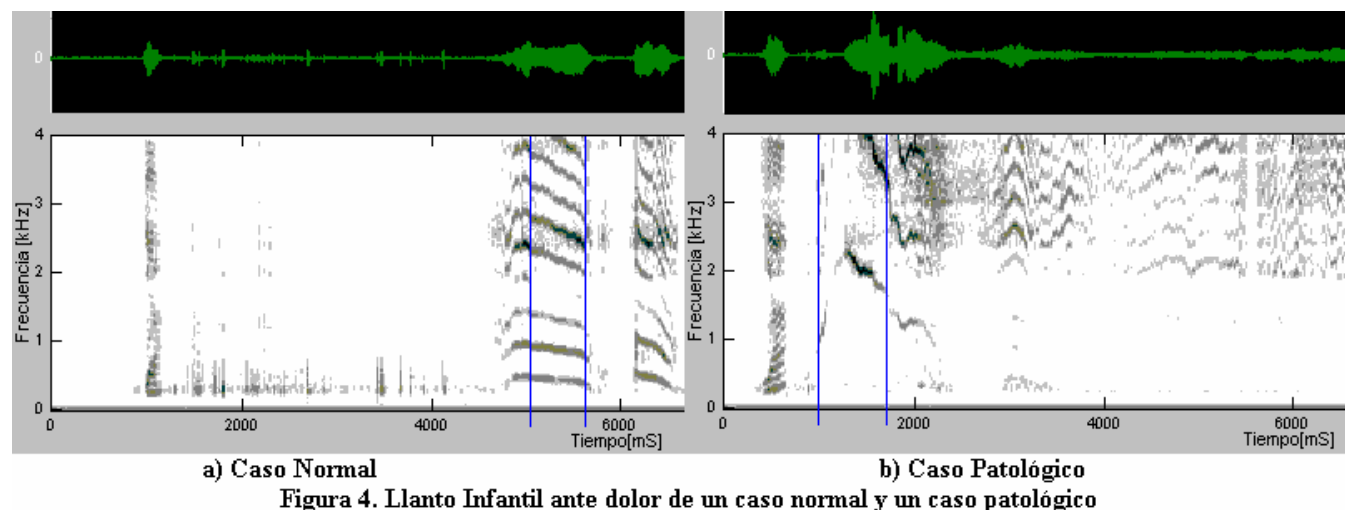
Las ventanas de distinta duración responden a experiencias realizadas. En el algoritmo original, para el habla, con 32 ms de ventana se obtienen buenos resultados en la estimación de F_0 [22], [23]. Por su parte, como el llanto infantil es tan cambiante, con variaciones de frecuencia superiores al rango del habla, se tomó ventana de 62.5 ms la cual proporciona la posibilidad en la exploración manual de apreciar mayor cantidad de períodos de la señal y adicionalmente ver mejor los umbrales de amplitud para, en conjunto, considerar hasta donde la sonoridad de una trama. Se han obtenido buenos resultados en la estimación del tono de llanto con ventana de 62.5 ms. La ventana de 10 ms ha sido probada con buenos resultados en la extracción de F_0 en el llanto, no obstante visualizarse en ella menor número de períodos, si bien es una ventana eficaz requiere mayor exigencia en la selección de tramas sonoras. Aunque el proceso de selección del conjunto de tramas sonoras en una señal es manual y éste sea un procedimiento de selección que requiere de alguna experiencia, lo que hace al algoritmo formar parte de una herramienta especializada, esta selección de las zonas sonoras de una señal de llanto se convierte en un proceso eficaz, procesándose sólo las tramas sonoras de donde se extrae F_0 y desechándose las tramas sordas. Todo lo anterior sustituye lo que en el algoritmo original es la selección automática de tramas sonoras y sordas, ésta es una de las partes sensibles que puede introducir errores en la estimación.

En nuestro diagrama, no se utiliza el enfatizado de la parte alta del espectro. La ventana Hamming usada en cada segmento de análisis donde es estimado el tono, satisface nuestras expectativas. En cada segmento, además del enventanado (multiplicación por la ventana Hamming), se calculan los Coeficientes LPC (orden 12). Luego sigue en el diagrama el bloque de Filtrado Inverso obteniéndose el error o residuo. Continúa el bloque de la Autocorrelación, la cual como procesamiento en el tiempo, es favorecida por la respuesta de fase lineal del filtro FIR del diagrama, aunque debe considerarse alguna distorsión de fase introducida por el filtro inverso. La detección del primer pico normalizado del período fundamental de la Función de Autocorrelación de la Onda Residual, empleando interpolación, permite estimar F_0 en el segmento sonoro y así sucesivamente en cada segmento de una fonación de llanto.

Estimación de F_0 en Llanto Infantil usando el Algoritmo SIFT con adecuaciones

Se utiliza el algoritmo SIFT con adecuaciones para estimar F_0 en llanto infantil y se compara el resultado con una herramienta profesional a estos efectos. La figura 4 presenta una composición de imágenes del software **Anavoz 1.0** (Analizador de Voz) [24], [25] de Oscilograma, en la parte superior (Gráfico de Amplitud *versus*

Tiempo) y Espectrograma, en la parte inferior, del llanto provocado por un estímulo de dolor de un caso normal y uno patológico, ambos casos de 10 días de nacidos.



El caso de la figura 4, inciso a, es un niño recién nacido a Término Normal de 3250 gramos de peso al nacer, de madre con edad gestacional de 40 semanas. Esta figura muestra algo más de 6 segundos de llanto después de pellizco en el calcáneo, y se puede apreciar en el Espectrograma un patrón característico de llanto de dolor con melodía mayormente descendente.

El caso patológico de la figura 4, inciso b, se refiere a una niña recién nacida Pretérmino, de 33 semanas de edad gestacional, es una niña de bajo peso al nacer igual a 1370 gramos, tiene una Depresión Respiratoria, es además pequeña para su edad gestacional, siendo un caso de Crecimiento Intrauterino Retardado. En esta figura e inciso, también con el mismo tipo de estímulo, se observa algo más de 6 segundos de llanto con melodía ascendente-descendente en el inicio de la primera fase espiratoria con notable incremento del tono. Al igual que la grabación de llanto del caso normal, en este caso también se aprecia una fonación en forma de columna al inicio de la grabación que es la palabra “ya” como indicador oral, dicha por el investigador que hace la grabación; este indicador permite estimar atributos asociados a la duración en el llanto.

Los Espectrogramas de la figura 4 al ser de Banda Estrecha (ancho de banda de los filtros de 45 Hz), permiten visualizar los armónicos en las regiones sonoras del llanto en cada caso y puede, según la escala de frecuencias, apreciarse el lugar geométrico del primer armónico, que es el armónico fundamental o F_0 . En la figura 4, las rectas paralelas de color azul limitan en cada caso (caso normal: 510.47 ms, caso patológico: 562.5 ms) la región de la primera fase espiratoria donde se estimó F_0 . Así, usando ventana Hamming para el cálculo de F_0 se tomaron tres segmentos de 62,5 ms en cada caso (segmentos inicial, intermedio y final). Los resultados de la estimación de F_0 usando el algoritmo SIFT con adecuaciones se presentan en la tabla 1, en dicha tabla se compara la estimación del **Anavoz 1.0** (usa algoritmo SIFT con adecuaciones) para el caso

normal con la estimación efectuada en el propio caso con una herramienta estándar de la Kay Elemetrics, el **DSP Sona-Graph, Model 5500 [26]**.

Caso Normal								
Segmentos seleccionados	Inicio y fin de segmento desde el centro del indicador oral (ms)			Anavoz 1.0 Fo (Hz)			DSP Sona-Graph Fo (Hz)	
inicial	4000 - 4062.5			473.9			470.83	
intermedio	4199.10 - 4261.60			444.0			450.0	
final	4447.97 - 4510.47			399.5			404.16	
Estadígrafo de Fo	media	mínimo	máximo	amplitud	S	S ²	CV (%)	No.Seg.
Anavoz 1.0	439.13	399.5	473.9	74.4	37.43	1401	8.52	3
DSP Sona-Graph	441.66	404.16	470.83	66.67	34.10	1162.81	7.72	3
Caso Patológico								
Segmentos seleccionados	Inicio y fin de segmento desde el origen del Espectrograma (ms)			Anavoz 1.0 Fo (Hz)				
inicial	1040 - 1102.5			1151				
intermedio	1290 - 1352.5			1933.4				
final	1540 - 1602.5			1444.4				
Estadígrafo de Fo	media	mínimo	máximo	amplitud	S	S ²	CV (%)	No.Seg.
Anavoz 1.0	1509.6	1151	1933.4	782.4	395.25	156222.5	26.18	3

Leyenda. S: Desviación Estándar; S²: Varianza; CV: Coeficiente de Variación; No. Seg: Número de Segmentos

Tabla 1. Resultados de la Estimación de Fo del Anavoz 1.0.

Como se aprecia, a modo de ejemplo, se ha estimado Fo en algo más de 500 ms de la primera espiración de llanto en cada caso de la figura 4, es decir, cerca de 9 segmentos de 62,5 ms fueron explorados por caso, seleccionándose tres valores de Fo, correspondientes a los segmentos referidos.

En la tabla 1, en el caso normal, al comparar la estimación que hace Anavoz 1.0 con la del DSP Sona-Graph puede verse que son muy parecidas. Si se comparan valores del estadígrafo como el valor medio, la Desviación Estándar y el Coeficiente de Variación de Fo, la diferencia es de 2.5 Hz, 3.3 y 0.8 % respectivamente, lo cual muestra cuán cerca están los valores de Fo estimados por Anavoz 1.0 al DSP Sona-Graph. También se aprecia en la tabla que como caso normal el valor medio de Fo se encuentra dentro del rango normal. La tabla 1, pone en evidencia que el valor medio de Fo para un caso notablemente patológico como el mostrado está muy por encima de 650 Hz, más del doble (1509.6 Hz), y que la variabilidad de Fo, refleja un Coeficiente de Variación del 26.18 %.

Pruebas realizadas con ventanas de 10 ms han aportado resultados comparables, previo análisis de las características de la señal y del contexto.

AGRADECIMIENTOS

Los autores agradecen la colaboración de los Servicios de Neonatología de los hospitales Materno Sur "Mariana Grajales" y "Juan Bruno Zayas", de Santiago de Cuba, y de los doctores Luís Regüeiferos, Leonardo Capdevila, Magalis Ruiz y Guillermo Delgado, colaboradores del Grupo de Procesamiento de Voz del CENPIS, de la Facultad de Ing. Eléctrica de la Universidad de Oriente.

CONCLUSIONES

Se ha adaptado un algoritmo SIFT de estimación de Fo en habla para la estimación de este parámetro en llanto infantil. La adaptación es debida a las diferencias existentes en características y atributos acústicos que tiene el lenguaje articulado en relación con el llanto, sobre todo en neonatos cuyo llanto es muy variable y de valor medio de Fo por encima del rango del habla, y como se ha podido apreciar en los casos patológicos, dicho rango de frecuencia es mucho más elevado.

El algoritmo SIFT con adecuaciones se ha implementado de tal manera que el mismo admite poder reajustar sus parámetros cuando las experiencias y la práctica así lo requieran.

Existe una diversidad de algoritmos para la estimación de Fo en el habla, no obstante casi siempre la extracción de este parámetro en determinadas señales, contextos y aplicaciones tiene sus exigencias, a veces prácticamente imposibles de cumplir a plenitud. En el caso del llanto, por sus características aún no resulta fácil obtener Fo con la misma precisión con la que se alcanza en ciertos contextos del habla, todo ello exige la adaptación de algoritmos existentes o la implementación de nuevos algoritmos que contribuyan a una mejor estimación de Fo lo cual aún no es un problema resuelto.

REFERENCIAS

- [1] **Quilis, A.** (1981), Fonética Acústica de la Lengua Española, Editorial Gredos, S. A, Madrid.
- [2] **Hirano, M.** (1981), Clinical Examination of Voice, Springer-Verlag/Wien, N. York.
- [3] **Wasz-Höckert, O., Lind, J., Vuorenkoski, V., Partanen, T., Valanne, E.** (1968), "The infant cry a spectrographic and auditory analysis". Clinics in Devel. Medicine. 29.
- [4] **Sirviö, P., Michelsson, K.** (1976), "Sound-spectrographic cry analysis of normal and abnormal newborn infants". Folia phoniatic. 28, 161-173.
- [5] **Wasz-Höckert, O., Michelsson, K., Lind, J.** (1985), "Twenty-five years of scandinavian cry research". In Lester, B. M., Boukydis, C. F. Z., (eds). Infant Crying: Theoretical and Research Perspectives. Plenum Publishing Corporation. N. York, 83-104.
- [6] **Michelsson, K., Michelsson, O.** (1999), "Phonation in the newborn, infant cry". International Journal of Pediatric Otorhinolaryngology. 49. Suppl. 1, 297-301.

- [7] **Michelsson, K., Eklund, K., Leppänen, P., Lyytinen, H.** (2002), "Cry characteristics of 172 healthy 1-to 7-day-old infants". *Folia Phoniatr. Logop.* 54, 190-200.
- [8] **Escobedo B, D. I., Cano O, S. D., Coello F, E., Regüíferos P, L., Capdevila B, L.** (2001), "Rising shift of pitch frequency in the infant cry of some pathologic cases". 2nd Int. Workshop on Models and Analysis of Vocal Emissions for Biomedical Applications. CD-ROM © University of Firenze, Italy.
- [9] **Furui, S.** (1989), *Digital Speech Processing, Synthesis and Recognition*, M. Dekker, Inc. N. York.
- [10] **Flanagan, J. L.** (1976), "Computers that talk and listen: man-machine communication by voice". *Proceedings of the IEEE.* 64, 405-415.
- [11] **Rabiner, L. R., Schafer, R. W.** (1976), "Digital techniques for computer voice response: implementations and applications". *Proceedings of the IEEE.* 64 (4), 416-433.
- [12] **Fernández, L. C., Martínez A, J. A., Alvarez M, J.** (1987), "Sistemas digitales de almacenamiento y síntesis de señal vocal en telefonía". *Mundo Electrónico* (171), 91-96.
- [13] **Golub, H. L., Corwin, M. J.** (1985), "A physioacoustic model of the infant cry". In Lester, B. M., Boukydis, C. F. Z., (eds). *Infant Crying: Theoretical and Research Perspectives*. N Y., P. Press, 59-82.
- [14] **Golub, H. L., Corwin, M. J.** (1982), "Infant cry: a clue to diagnosis". *Pediatrics.* 69 (2), 197-201.
- [15] **Pinyerd, B. J.** (1994), "Infant cries: physiology and assessment". *Neonatal Network.* 13 (4), 15-20.
- [16] **Makhoul, J.** (1975), "Linear prediction: a tutorial review". *Proc. of the IEEE.* 63, 561-580.
- [17] **Ribbun, B., Perkis, A., Paliwal, K. K., Ramstad, T.** (1991), "Performance study of stochastic speech coders". *Speech Communication.* 10 (3), 277-301.
- [18] **Cano O, S. D., Escobedo B, D. I., Socarrás R, M.** (1992), "Las técnicas de corta duración en el procesamiento digital de la voz". V Congreso Latinoamericano de Control Automático. 2, 1-7.
- [19] **Cano O, S. D., Escobedo B, D. I.** (2000), "Caracterización de los fonemas españoles basada en el mapa de la función energía en las bandas de alta y baja frecuencia". *TELEC '2000*, International Conference-Santiago de Cuba. CD-ROM. Universidad de Alcalá, España.
- [20] **McGonegal, C. A., Rabiner, L. R., Rosenberg, A. E.** (1977), "A subjective evaluation of pitch detection methods using LPC synthesized speech". *IEEE Transactions on Acoustics, Speech, and Signal Processing.* ASSP-25 (3), 221-229.
- [21] **Oppenheim, A., Shaffer, V.,** *Discrete-Time Signal Processing.*
- [22] **Markel, J.** (1972), "The SIFT algorithm for fundamental frequency estimation". *IEEE Transactions on Audio and Electroacoustics.* Au-20 (5), 367-377.
- [23] **Veprek, P., Scordilis, M. S.** (2002), "Analysis, enhancement and evaluation of five pitch determination techniques". *Speech Communication.* 37, 249-270.

- [24] Escobedo B, D. I., Sanabria M, F., Cano O, S. D., Marañón R, E. J. (2006),** Manual de Usuario Anavoz 1.0, Universidad de Oriente, Santiago de Cuba. (Registro de Software: 1846-2008).
- [25] Escobedo B, D. I., Sanabria M, F., Cano O, S. D., Marañón R, E. J. (2008),** "Analizador de voz, llanto infantil y habla usando Matlab". FIE'08, Conferencia Internacional, 5ta Edición. CD-ROM© Universidad de Oriente. Santiago de Cuba.
- [26] Kay Elemetrics Corp,** DSP Sona-Graph. Model 5500. Reference Guide, Pine Brook. NJ, USA.