Revista Cubana de Ciencias Informáticas
Vol. 15, No. 1, Enero-Marzo, 2021
ISSN: 2227-1899 | RNPS: 2301
http://rcci.uci.cu
Pág. 70-88

Tipo de artículo: Artículo original
Temática: Bioinformática
Recibido: 10/12/2020 | Aceptado: 10/01/2021

Modelación bio-inspirada del sistema auditivo para el procesamiento del habla.

Bio-inspired modeling of the auditory system for speech processing.

Viviana Abad Peraza[1*] https://orcid.org/0000-0002-9000-7155
Ernesto Arturo Martínez Rams [1] https://orcid.org/0000-0001-8364-1265

[1]Universidad de Oriente, Departamento de Telecomunicaciones, Santiago de Cuba. Avenida de Las Américas s/n, Santiago de Cuba, Cuba. {mviviana, eamr@uo.edu.cu}.

RESUMEN

En el presente trabajo se implementa un modelo bio-inspirado del sistema auditivo periférico y central para el procesamiento de sonidos correspondientes al lenguaje humano. En el mismo se concluye que la representación espectral del modelo de fonación humano corresponde con las características tonotópicas de la cóclea, del sistema auditivo periférico. Se observó cómo el proceso de inhibición lateral, del sistema auditivo central, agudiza la selectividad espectral para obtener un perfil más definido de las trayectorias de los formantes del habla. Con la modelación de las neuronas básicas se lograron detectar los formantes y las características dinámicas del habla, así como la detección de ráfagas de ruido en estas.

Palabras clave: sistema auditivo; camino auditivo; modelo bio-inspirado del sistema.

Revista Cubana de Ciencias Informáticas
Vol. 15, No. 1, Enero-Marzo, 2021
ISSN: 2227-1899 | RNPS: 2301
http://rcci.uci.cu
Pág. 70-88

ABSTRACT

In the present work, a bio-inspired model of the peripheral and central auditory system for the processing of sounds corresponding to the human language is implemented. In this work, is concluded that the spectral representation of the human phonation model corresponds to the tonotopics characteristic in the cochlea, of the peripheral auditory system. Was showed how the lateral inhibition process in the central auditory system increases the spectral selectivity to obtain a more defined profile of the speech formants trajectories. Also, was possible to detect the formants and dynamic characteristics of speech with basic neurons simulations, as well as the noise burst detection in them by modeling.

Keywords: auditory system; auditory pathways; bio-inspired model of the auditory system.

Introducción

El sistema auditivo es el conjunto de órganos que hacen posible el sentido del oído (KandelER, 2013; Ortega-Garcia, 2000). La función del sistema auditivo es transformar las variaciones de presión de las ondas sonoras en impulsos neurales. La información generada se transmite al cerebro para la asignación de significados. Por otra parte, la audición es el sentido que le permite a los órganos captar el sonido del ambiente. El sistema auditivo se divide en tres partes fundamentales: periférico; central; y superior o corteza auditiva.

El proceso de la audición implica un proceso fisiológico y otro psicológico. En el proceso fisiológico es donde se capta el sonido, se codifica y se envía al cerebro. Los órganos que participan en este proceso conforman el sistema auditivo periférico. En el proceso psicológico es donde se interpretan estos sonidos, se reconocen y se dotan de significado. Los órganos que permiten la percepción del sonido conforman el llamado sistema auditivo central. El sistema auditivo periférico se conforma por el oído externo, medio e interno (Alper, 2017; de Nava, 2019; Hixon, 2018; Thomassin, 2016). Este sistema realiza la captura, conducción, modificación, ampliación y descomposición espectral de las ondas sonoras que llegan al pabellón auditivo. Estas ondas se

convierten en impulsos neurales en las fibras aferentes de la porción coclear del nervio auditivo. Siendo estas fibras las que codificarán espacio-temporalmente las principales características de las señales percibidas para su transmisión al sistema nervioso central. El sistema auditivo central se compone por el núcleo coclear, el núcleo olivar superior, el colículo inferior y el cuerpo medial geniculado (KandelER- JesselTM 2013).

La neurona es la célula básica encargada de transmitir la información en forma de impulso nervioso por una serie de neuronas, una después de otra. La sinapsis es el punto de unión de dos neuronas y controla la transmisión de señales. Esta ejerce una acción selectiva inhibitoria, bloqueando las señales que les llegan, o excitatorias permitiendo el paso de estas (Antonietti et al. 2018; Kandel et al. 2013). Entre las principales neuronas que intervienen en cada uno de estos sistemas se encuentran las primary-like (Pl), onset (On), chopper (Ch), y pauser (Pb), que actúan a nivel del núcleo coclear (NC). Dentro de las agrupaciones de neuronas están las neuronas de inhibición lateral (NIL), las de CF (frecuencia característica), las FM (frecuencia modulada) y NB (ráfaga de ruido, noise burst) (Gómez-Vilda, 2009). Las fibras o NIL están en toda la superficie del NC, principalmente en los NC dorsales (Hernández-Zamora, 2014; PX, 2011). Las neuronas CF se encuentran en el núcleo olivar superior, las FM en el colículo inferior y las NB en el cuerpo medial geniculado (Gómez-Vilda, 2011). La inhibición lateral (IL) se refiere a la inhibición que tienen entre sí las neuronas vecinas de las vías auditivas. Cada neurona del camino neural tiene sinapsis excitatoria con su correspondiente zona tonotópica de la cóclea y sinapsis inhibitoria con las zonas tonotópicas vecinas. La inhibición lateral a este nivel permite agudizar la selectividad espectral y en consecuencia obtener un perfil más definido de las trayectorias de los formantes del habla produciendo estimaciones nítidas de los picos espectrales. La selectividad de frecuencia está dada por el factor de calidad que relaciona la frecuencia central o característica de la neurona IL con el ancho de banda de un estímulo por encima de un umbral de 10 dB (Hernández-Zamora, 2014; PX, 2011).

Una vez que la información ha sido procesada en la cóclea y se ha realizado la transducción mecánico-neural, todavía debe ser procesada en varios centros neurales antes de llegar a la corteza superior (córtex). Esta información primero discurre por las fibras nerviosas hasta el núcleo coclear. Luego pasa al núcleo olivar superior, desde donde asciende hasta el colículo inferior y al cuerpo geniculado medial del tálamo. Por último, llega al lóbulo temporal de la corteza superior (córtex), donde es procesada en último término antes de pasar a los centros del lenguaje (área de Wernicke y área de Broca) (Gómez Vilda, 2009; Ferrández Vicente, 1998).

En el núcleo coclear distintas neuronas se encargan de procesar diferentes características de la señal. Las células primary-like transmiten con rigurosidad la información que les llega, las onset detectan el inicio de la señal, las chopper se disparan regularmente y de forma proporcional a la intensidad del estímulo y las células pauser retardan la información. El siguiente centro que procesa la información es el núcleo olivar superior, que posee dos subdivisiones principales, la zona lateral que detectará las diferencias en intensidad de la información que llega a ambos oídos y el centro medial que posee células sensibles a la diferencia interaural temporal. La funcionalidad del núcleo olivar será, por lo tanto, localizar el origen de la fuente sonora, utilizando las diferencias en intensidad si los estímulos poseen frecuencias elevadas, y la diferencia temporal si éstas son inferiores. La información de salida del núcleo coclear y del complejo olivar alcanza el colículo inferior. Este último se encarga de codificar simultáneamente la información sobre los componentes del estímulo y su localización en el espacio. Esta funcionalidad consistente en la detección de los componentes FM podría venir determinada por una red neuronal que confrontará la información con distintos retardos para pasar de un código temporal a un código espacial. El núcleo geniculado lateral del tálamo es el siguiente centro de proceso. En su división ventral se han localizado neuronas organizadas según su frecuencia, con diferentes latencias y sensibles a componentes ruidosos, pudiéndose almacenar así información retardada. En el núcleo medial se ha detectado plasticidad sináptica, que permitiría el etiquetado de estímulos con significado perceptual (Musiek, 2018). La información, por último, alcanza la corteza auditiva (CA). En la zona primaria de la CA existe una representación tonotópica conjugada con bandas excitatorio - excitatorias y excitatorio - inhibitorias y con una organización por intensidades creciente. Además, se han detectado neuronas CF, que caracterizan a los fonemas estáticos, como las vocales, y que consisten en componentes de frecuencia característica constante. Así mismo, células FM, que responden ante determinadas transiciones en frecuencia precisas en un sentido dado, y células NB que detectan estímulos ruidosos centrados en una frecuencia determinada y con un cierto ancho de banda. Es decir, en esta estructura se detectan los tres componentes del habla conjuntamente con la temporalidad deseada. Es por ello que el objetivo del presente trabajo es implementar un modelo bio-inspirado del sistema auditivo, periférico y central, y analizar su comportamiento ante estímulos sonoros.

Revista Cubana de Ciencias Informáticas
Vol. 15, No. 1, Enero-Marzo, 2021
ISSN: 2227-1899 | RNPS: 2301
http://rcci.uci.cu
Pág. 70-88

Métodos o Metodología Computacional

Descripción del modelo auditivo bio-inspirado

El modelo bio-inspirado que se implementa en el presente trabajo se muestra en la figura 1. El mismo está basado en el modelo propuesto en (Gómez-Vilda, 2009).

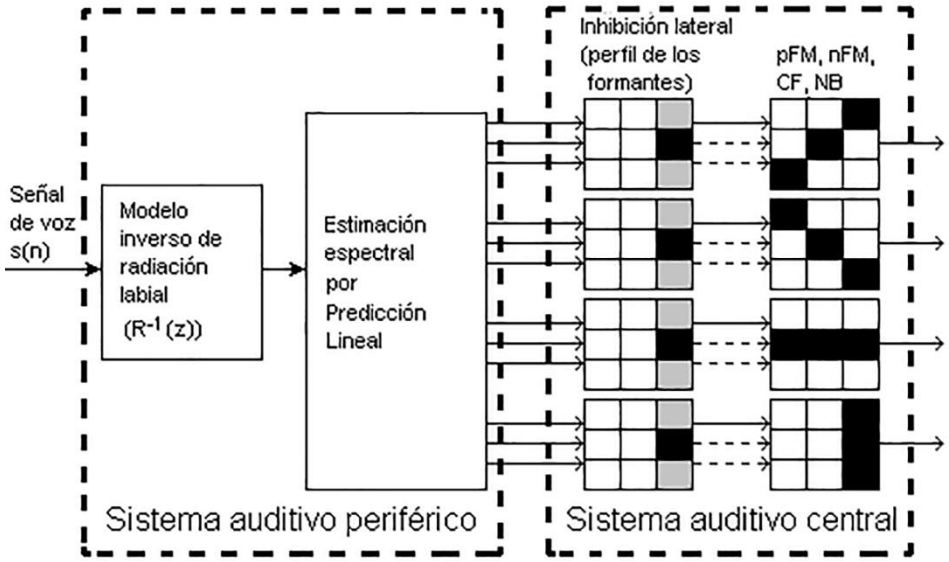

Fig. 1 – Modelo bio-inspirado del sistema auditivo.

Modelo del sistema periférico

El modelo bio-inspirado del sistema auditivo periférico se encuentra conformado por el bloque del modelo inverso de radiación labial y por el bloque de estimación espectral, figura 1 (Gómez Vilda, 2009). Este último bloque se implementa teniendo en cuenta el algoritmo de la codificación predictiva lineal (Linear Predictive Coding, LPC). El modelo inverso de radiación labial junto con el modelo de la LPC, corresponden a un modelo de fonación cuya representación en frecuencia es equivalente a la descomposición espectral del sonido

en la cóclea. En este caso la envolvente espectral resultante coincide con la energía del espectro de la señal del habla.

El efecto de radiación de los labios se modela como un filtro de respuesta a impulso finita (FIR) en celosía, el cual se representa mediante la siguiente función de transferencia:

$$H_1(z) = R^{-1}(z) = 1 - r_f z^{-1}$$

Esta función cancela el polo de primer orden introducido por los efectos de la radiación en los labios. Siendo r_f el primer coeficiente de reflexión del filtro. Para el cálculo de este coeficiente de reflexión, se utiliza el algoritmo de *Burg*, en el cual se minimiza la suma de los valores cuadrático medio de la energía residual hacia delante y hacia atrás, asegurando la estabilidad del filtro de síntesis.

La técnica de LPC, consiste en estimar el valor actual de una señal x(n) como una combinación lineal de las muestras pasadas y presentes a la entrada. La función de transferencia *H(z)* del filtro predictor en función del filtro de error de predicción *A(z)* y de la ganancia *G* del filtro predictor es H(z)=G/A(z).

Modelo bio-inspirado para la simulación del sistema auditivo central

El modelo bio-inspirado del sistema auditivo central está constituido por el bloque de inhibición lateral y por cada una de las plantillas que simulan las neuronas de CF, FM y NB.

Proceso de inhibición lateral

En el proceso de inhibición lateral de la figura 1 se emplean máscaras o plantillas reticulares para procesar el espectrograma X(m,n) como una imagen, siendo esta la salida del estimador espectral (Gómez-Vilda, 2011):

$$\widetilde{X_{LI}}(m,n) = u\left(\sum_{i=-r}^{r} w_{LI}(i)X(m+i,n) - \vartheta_{LI}(m,n)\right)$$

Revista Cubana de Ciencias Informáticas
Vol. 15, No. 1, Enero-Marzo, 2021
ISSN: 2227-1899 | RNPS: 2301
http://rcci.uci.cu
Pág. 70-88

Donde n y m representan el tiempo y la frecuencia en la que se encuentra indexada la señal. w_{LI} es una máscara con un patrón específico y un conjunto de pesos. Para máscaras de orden 3x3 (r=1), los pesos son w_{-1}= w_1=$-1/2$ y w_0=1. $u(.)$ es una función de activación no lineal (paso unitario o sigmoide) que se activa cuando el valor resultante supera un umbral específico $\vartheta_{LI}(m,n)$. Para las pruebas experimentales del presente trabajo se consideró con valor nulo el umbral.

Para producir resultados imparciales, en las plantillas de las máscaras reticulares, el peso asociado a cada cuadro negro es fijado en $+1/s_b$ (peso excitatorio) y el peso asociado a cada cuadro blanco es fijado a $-1/s_w$ (peso inhibitorio), siendo s_b y s_w el número de cuadros blancos y negros encontrados en la máscara 3x3, respectivamente (Gómez Vilda, 2011).

A cada salida del proceso de inhibición lateral se le aplican cuatro plantillas o máscaras reticulares diferentes para la detección de pendientes positivas y negativas de la FM (pFM y nFM), la frecuencia característica (CF) y el ruido Burst (NB).

Proceso de seguimiento dinámico

El seguimiento dinámico de los formantes se realiza al reconocer los segmentos donde los dos primeros formantes permanecen relativamente estables. Esta actividad se registra utilizando las máscaras de pFM y de nFM, lo cual es similar a tratar el espectrograma como una imagen auditiva, según la siguiente ecuación:

$$X_{DF} \pm (m,n) = u\left[\sum_{p=-P}^{P}\sum_{q=0}^{Q} w_{DF} \pm (p,q) X_{LI}(m+p, n-q)\right]$$

La matriz de pesos $w_{DF} \pm (p,q)$ es un histograma en forma de campana desplazado en el índice de tiempo (q) que sigue un ancho de banda ascendente o descendente en la frecuencia para implementar las máscaras de seguimiento de FM, que también se conocen como campos receptivos de FM. Los valores prácticos para p y q son cuatro (4) y ocho (8), respectivamente, resultando en una máscara de 9x9. El siguiente paso es la

Revista Cubana de Ciencias Informáticas
Vol. 15, No. 1, Enero-Marzo, 2021
ISSN: 2227-1899 | RNPS: 2301
http://rcci.uci.cu
Pág. 70-88

separación de las salidas del campo receptivo ascendente y descendente en la primera y segunda bandas de formantes de la forma siguiente:

$$X_{BF}\pm(\varphi)=u\left[\sum_{i=-\beta_\varphi}^{\beta_\varphi} w_{BF}\pm(i,\varphi)\,X_{DF}\pm(\gamma_\varphi+i)\right];\;\; w_{BF}(i,\varphi)=\Gamma\left(\xi_i\middle|\mu_\varphi,\sigma_\varphi\right)=\frac{1}{\sigma_\varphi\sqrt{2\pi}}e^{\frac{(\xi_\varphi-\mu_\varphi)^2}{2\sigma_\varphi{}^2}}$$

donde φ es el índice del formante, γ_φ y β_φ son los índices de la frecuencia central y la mitad del ancho de banda. Los pesos de la suma $w_{BF\pm}$ se seleccionan para reproducir la probabilidad de salida del formante de acuerdo con la función de densidad gaussiana, siendo μ_φ y σ_φ la media de la banda y la desviación estándar, $-\beta_\varphi\Omega\le\xi_i\le\beta_\varphi\Omega;\;\; \xi_i=i\Omega;\;\; \mu_\varphi=\gamma_\varphi\Omega;\; \sigma_\varphi=\beta_\varphi\Omega.$

El siguiente proceso en este modelo es la asignación dinámica de los formantes, debido a que la actividad de los campos receptivos de FM ascendentes (pFM) y descendentes (nFM) puede ser ambigua. Esto se debe al carácter ruidoso de los picos neuronales, es decir, que no será infrecuente que los campos receptivos muestren actividad para un mismo formante, lo que resulta en ambigüedad. En el presente trabajo este resultado se precisa mediante la comparación entre las probabilidades de los campos receptivos pFM y nFM. De esta forma, el que muestre mayor probabilidad de ocurrencia en cada instante de tiempo será el que se represente.

Resultados y discusión

Modelo del sistema periférico

Modelo inverso de radiación labial

En la figura 2a se representa un segmento de voz de la vocal /a/ de un locutor masculino. En la figura 2b se muestra la señal de salida del filtro eliminador del efecto de radiación de los labios. Comparando ambas señales se puede apreciar como la señal de salida mantiene la misma forma de onda, aunque resaltando el

Revista Cubana de Ciencias Informáticas
Vol. 15, No. 1, Enero-Marzo, 2021
ISSN: 2227-1899 | RNPS: 2301
http://rcci.uci.cu
Pág. 70-88

contenido de altas frecuencias lo cual se observa en los cambios bruscos que se manifiestan en esta señal. La señal resultante constituye la entrada al bloque de estimación espectral por predicción lineal.

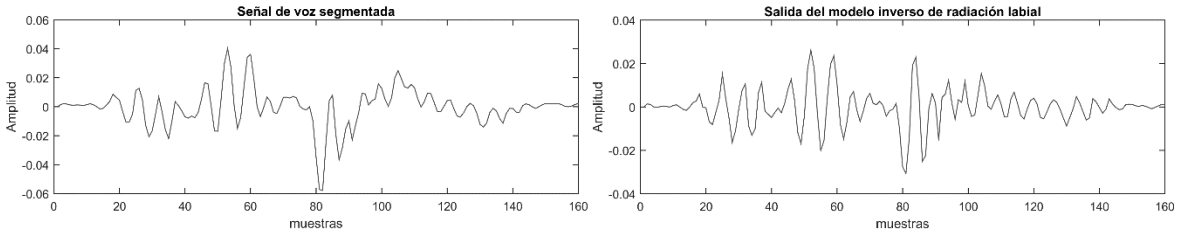

Fig. 2 – a) Segmento de señal de voz presentada al filtro eliminador del efecto de radiación de los labios; b) Señal de salida del filtro.

Modelo de estimación espectral de la cóclea

Este proceso se refiere a la extracción de las características de la voz en cada ventana y se simula teniendo en cuenta el modelo de filtro todo - polos. Un aspecto importante en el análisis de la LPC es la determinación del orden del predictor *p*. Su orden mínimo corresponde con el tiempo requerido para que la onda del sonido viaje dos veces desde la glotis hasta los labios, $t=2 \cdot l/v$, siendo *l* la longitud del tracto vocal (≈17 cm) y *v* la velocidad del sonido (340 m/s). El tiempo resulta ser de 1 ms, lo que significa que se requiere de una memoria de 1 ms. El orden mínimo, para la frecuencia de muestreo f_s=8 kHz, resulta ser $p_{mín}=t \cdot f_s=8$. Para el presente trabajo los mejores resultados se lograron p=14, órdenes mayores no mejoraron de manera apreciable los resultados.

En la figura 3 se pueden observar los diferentes cocleogramas para las cinco vocales de un locutor masculino. Se puede notar en los cocleogramas que las bandas más oscuras corresponden con la región de los formantes de la voz, principalmente los dos primeros F_1 y F_2. En los cocleogramas correspondientes a las vocales /o/ y /u/ estos formantes se encuentran notablemente solapados, debido a que ambos poseen regiones espectrales que se solapan y ambas de baja energía. Para la vocal /a/ aunque existe un ligero solapamiento entre la región de estos formantes se pueden observar más definidos. En el caso de las vocales /e/ e /i/ la separación entre estos primeros formantes es mucho mayor y por ello los cocleogramas obtenidos son más nítidos con respecto a los anteriores.

Revista Cubana de Ciencias Informáticas
Vol. 15, No. 1, Enero-Marzo, 2021
ISSN: 2227-1899 | RNPS: 2301
http://rcci.uci.cu
Pág. 70-88

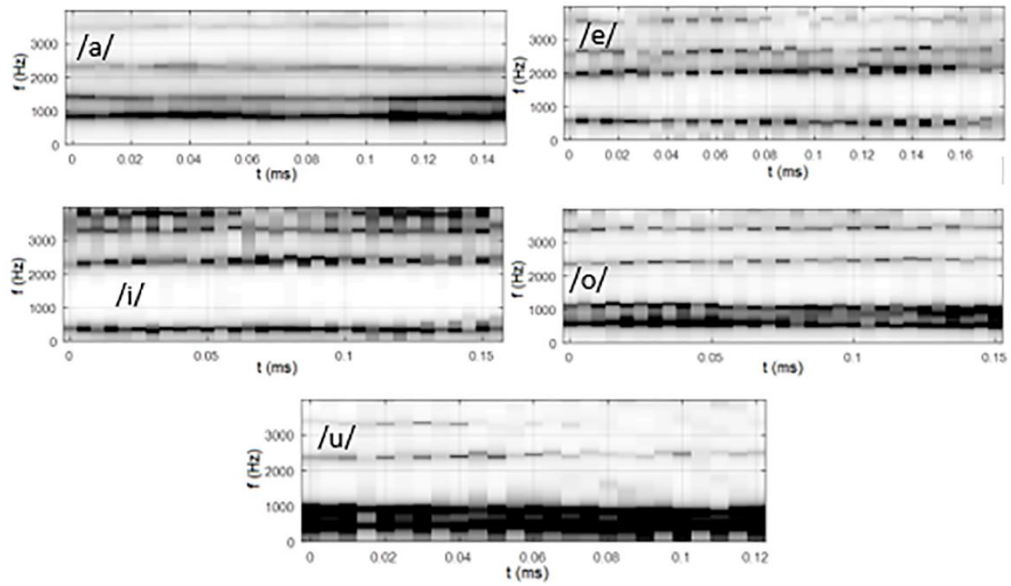

Fig. 3 – Cocleogramas de las vocales desde la /a/ hasta la /u/ para un locutor masculino.

En estas imágenes también se puede observar que el formante F_1, tiene una mayor frecuencia en la vocal /a/, próximo a 1 000 Hz; y una menor frecuencia en las vocales /i/ y /u/, en los 400 Hz y 550 Hz aproximadamente, respectivamente. F_2 es mayor en la /i/ (≈2 100 Hz) y luego en la /e/ (≈2 000 Hz), por lo que este formante no muestra gran diferencia en estas vocales.

Modelo bio-inspirado para la simulación del sistema auditivo central

Inhibición lateral

En la figura 4 se muestra el resultado del proceso de la inhibición a la señal resultante de los cocleogramas obtenidos en el apartado anterior. En esta se puede observar una mejor definición de la región de los formantes por lo que es posible determinar con más facilidad las regiones de frecuencias de los mismos.

Revista Cubana de Ciencias Informáticas
Vol. 15, No. 1, Enero-Marzo, 2021
ISSN: 2227-1899 | RNPS: 2301
http://rcci.uci.cu
Pág. 70-88

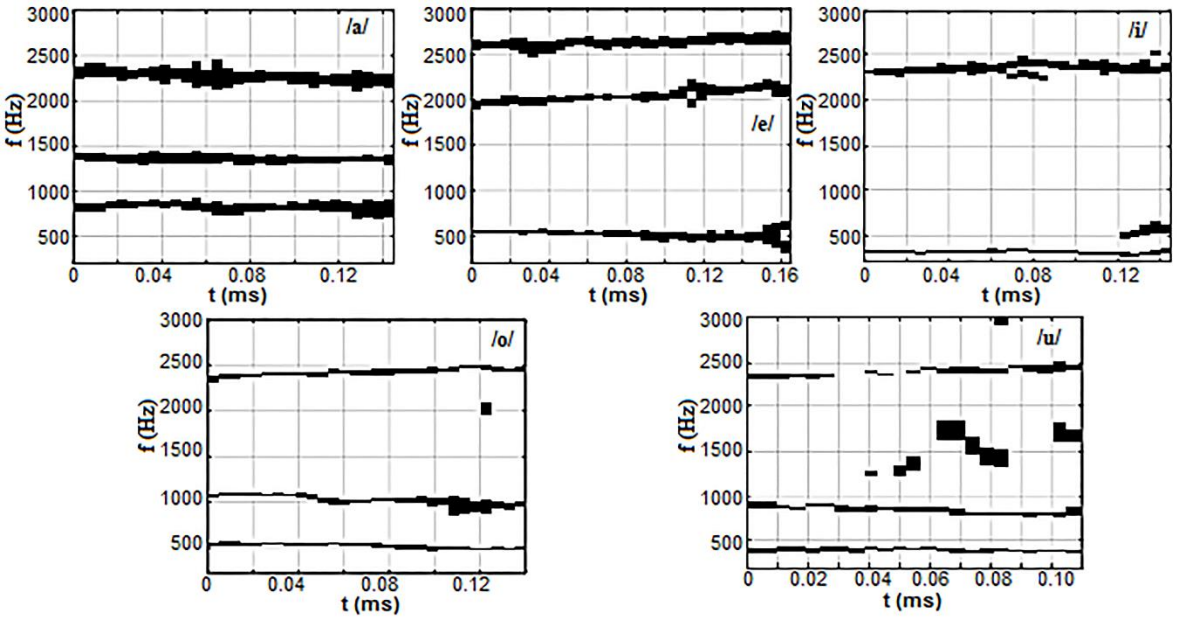

Fig. 4 – Proceso de inhibición lateral para las vocales de un locutor masculino en el idioma español, X_{LI}.

Neuronas básicas del habla.

En la figura 5 se muestra la respuesta de los cuatros neuronas básicas del habla (CF, pFM, nFM, NB) para la secuencia de la palabra /veíamos/ para un locutor masculino de idioma español de la base de datos española AHUMADA (Ortega García, 2000).

En esta se puede evidenciar el conjunto de todos los resultados expuestos anteriormente. Por ejemplo, se puede observar la transición del primer y segundo formante de la /e/ a la /i/, F_1/e/ → F_1/i/ y F_2/e/ → F_2/i/, notándose que F_2/e/ y F_2/i/ son mayores que F_2/a/ y F_2/o/. El formante F_2 comienza a ascender en la vocal /e/, desde aproximadamente los 0,1 ms, se hace mayor en la vocal /i/, hasta 0,25 ms y a partir de aquí desciende para la vocal /a/ hasta los 0,35 ms aproximadamente. Este mismo formante vuelve a aumentar en la consonante /m/ y desciende nuevamente en la vocal /o/, en el intervalo de 0,45 ms – 0,5 ms aproximadamente, donde asciende con la consonante /s/.

Revista Cubana de Ciencias Informáticas
Vol. 15, No. 1, Enero-Marzo, 2021
ISSN: 2227-1899 | RNPS: 2301
http://rcci.uci.cu
Pág. 70-88

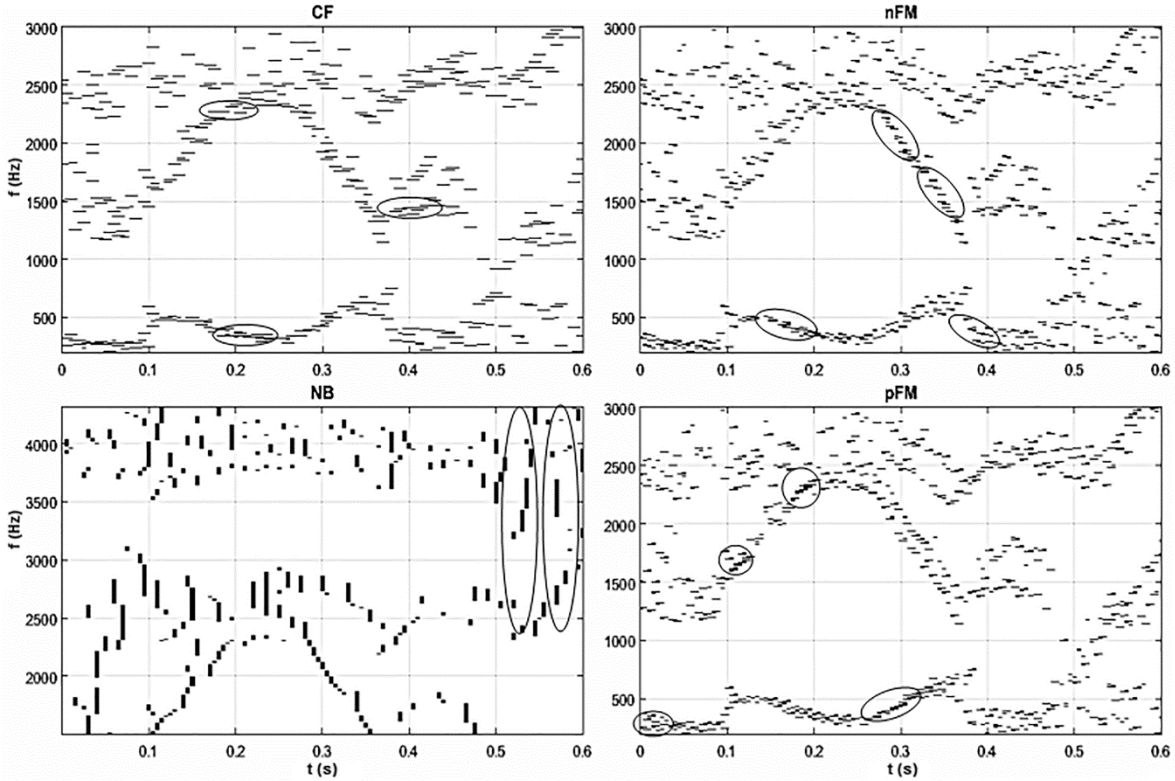

Fig. 5 – Respuesta de las neuronas básicas del habla para la palabra /veíamos/ de un locutor masculino.

En el caso de F_1, F_1/e/ es mayor que F_1/i/, y por tanto alcanza su mayor valor en el intervalo de 0,1 ms - 0,2 ms y luego desciende en el intervalo de 0,2 ms - 0,27 ms. A partir de este último intervalo de tiempo F_1 aumenta, ya que F_1/a/ es mayor que F_1/e/ y F_1/i/ y así se mantiene hasta aproximadamente los 0,38 ms, donde vuelve a descender debido a la pronunciación de la vocal /o/. Cercano a los 0,5 ms hay una rápida transición en este formante, donde aumenta y vuelve a descender, lo cual viene dado por la posición de los órganos de articulación del habla para la pronunciación de esta consonante. Esta es una consonante fricativa sonora, en el cual el elemento principal que identifica este tipo de fonema es la aparición de bandas de energía en zonas muy bajas del espectro, alrededor de los 150 Hz, fruto de la vibración de las cuerdas vocales que alcanzará el exterior a pesar de la oclusión que produce el punto de constricción. Esta última afirmación se puede corroborar en esta figura, donde se puede notar que a la /v/ le corresponde frecuencias próximas a los 200 Hz. Fueron resaltados además algunas áreas donde se puede apreciar la actividad de cada una de estas neuronas.

En la figura 6 se muestra la detección de las neuronas básicas del habla para un locutor femenino de habla hispana. Esta señal fue grabada con un dispositivo móvil en presencia de ruido ambiente. En este caso, se realiza la detección de las neuronas básicas del habla en la secuencia /foto/. En esta secuencia se encuentra la vocal /o/, de sonido sonoro, la cual se encuentra por primera vez luego de la fricativa sorda /f/. En la segunda ocasión, luego de la plosiva u oclusiva sorda /t/. En la misma se puede observar que se detectan las neuronas CF características de la vocal /o/ alrededor de los 500 - 540 Hz y de los 770 - 830 Hz, para F_1/o/ y F_2/o/ respectivamente, lo cual coincide con el rango de los formantes para esta vocal. Instantes antes de la pronunciación de la /t/ en el que existe una oclusión (0,32 – 0,33 s), se observa ausencia de formantes. Luego se observa la activación de las neuronas nFM en F_2 por la pronunciada pendiente negativa debido a la inmediatez de la vocal /o/ luego del sonido plosivo /t/, tendiendo a alcanzar la frecuencia CF del segundo formante.

Seguimiento dinámico de los formantes.

Los campos receptivos que detectan las frecuencias ascendentes o descendentes de los formantes, mediante plantillas de máscaras, corresponden a la ecuación que transforma la entrada $X_{LI}(m,n)$ en las respectivas salidas $X_{DF}+(m)$ y $X_{DF}-(m)$. Siendo m el orden respectivo de las bandas de frecuencias que se están buscando, y el signo + o - se refiere al sentido positivo o negativo de la pendiente. En el caso específico que se muestra en las simulaciones, las dimensiones de las matrices de los pesos $w_{DF}\pm(p,q)$ son 9x9, lo que significa que la conectividad en frecuencia se extiende desde +4 hasta -4 neuronas vecinas.

Revista Cubana de Ciencias Informáticas
Vol. 15, No. 1, Enero-Marzo, 2021
ISSN: 2227-1899 | RNPS: 2301
http://rcci.uci.cu
Pág. 70-88

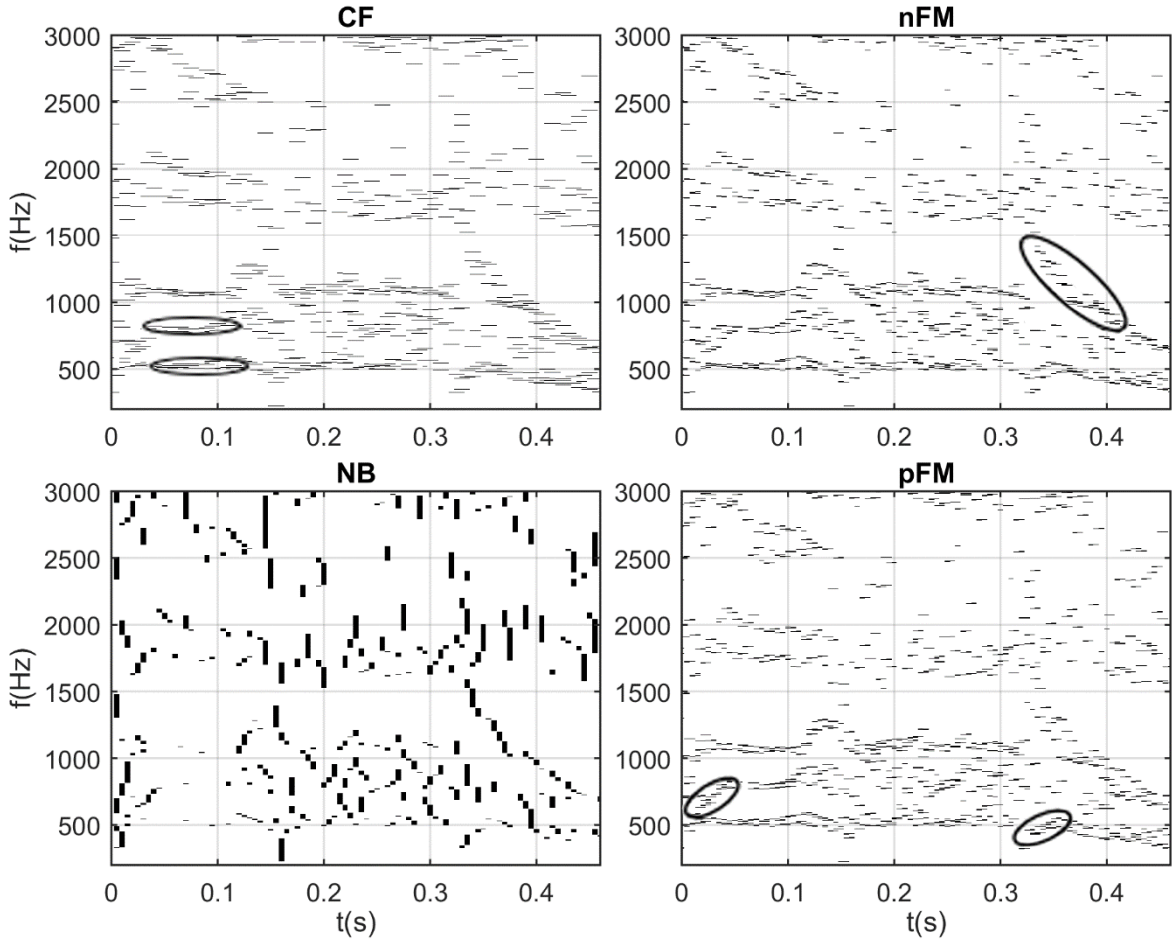

Fig. 6 – Respuesta de las neuronas básicas del habla para la palabra /foto/ de un locutor femenino.

Luego de este proceso se prosigue con la separación de las pendientes positivas y negativas en las bandas de salida de los dos primeros formantes, en la cual los pesos $w_{BF} \pm (i, \varphi)$ se obtienen por la función de densidad de probabilidades gaussiana, lo cual garantiza que la mayor probabilidad de activación de las pendientes (positivas o negativas) se encuentre centrada en el índice de la banda de frecuencia del formante con que se esté trabajando. El último paso dentro de esta simulación es la asignación dinámica de los campos receptivos de FM dentro de un mismo formante. En general todos los procesos incluidos en este epígrafe se han implementado utilizando el concepto de la inhibición lateral.

Revista Cubana de Ciencias Informáticas
Vol. 15, No. 1, Enero-Marzo, 2021
ISSN: 2227-1899 | RNPS: 2301
http://rcci.uci.cu
Pág. 70-88

Los tres procesos que se simulan dentro de este epígrafe incluyen una función de activación no lineal, que en estos casos fue escogida la función paso unitario, con lo cual se pueden reducir aleatoriedades en la detección de las pendientes en los formantes. En la figura 7 se muestra el resultado del seguimiento dinámico realizado a la palabra /veíamos/ de un locutor masculino.

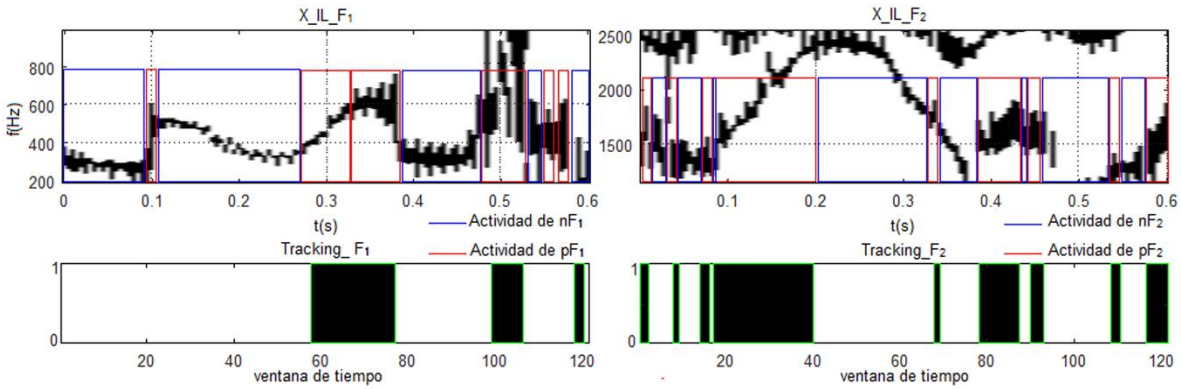

Fig. 7 – Seguimiento dinámico de los formantes para la palabra /veíamos/ de un locutor masculino.

El algoritmo implementado logra identificar las actividades de los campos receptivos de FM, en los que solo ha tenido confusión en tres cortos intervalos de tiempo en F_1, localizados en los 0,1 ms, 0,55 ms y 0,57 ms aproximadamente. Sin embargo, también puede notarse como positivo que esto no significa que las pendientes o transiciones que aparecen inesperadamente (en intervalos de tiempo muy cortos) no las detecte, porque se puede observar que en el caso del seguimiento de F_1, se pudo identificar la actividad de pF_1 en 0,59 ms aproximadamente, y la actividad de pF_2 en los 0,01, 0,035 y 0,08 ms, aproximadamente, así como, en los 0,335, 0,425 y 0,54 ms. Aunque, es cierto que estos casos pueden catalogarse como aislados debido a otros casos en los cuales ocurren estas mismas transiciones y no son detectados.

En el caso de la figura 8 se muestra el seguimiento dinámico de los formantes para la palabra /foto/ del mismo locutor femenino visto anteriormente. En esta secuencia, los dos primeros formantes se encuentran cercanos y en algunos intervalos se solapan. Sin embargo, la actividad de los campos receptivos de FM es mucho mayor con respecto al de la palabra /veíamos/, figura 7. Las pendientes, tanto positivas como negativas en los formantes son más notables y frecuentes. También se pueden notar algunos intervalos de tiempo en los cuales

Revista Cubana de Ciencias Informáticas
Vol. 15, No. 1, Enero-Marzo, 2021
ISSN: 2227-1899 | RNPS: 2301
http://rcci.uci.cu
Pág. 70-88

se encuentran ligeramente desfasados la actividad de algún campo receptivo con respecto a su activación. No obstante, como un resultado general, se puede decir que desde el punto de vista subjetivo se alcanza una precisión aceptable, debido a que detecta los episodios principales de ascenso y descenso de los formantes.

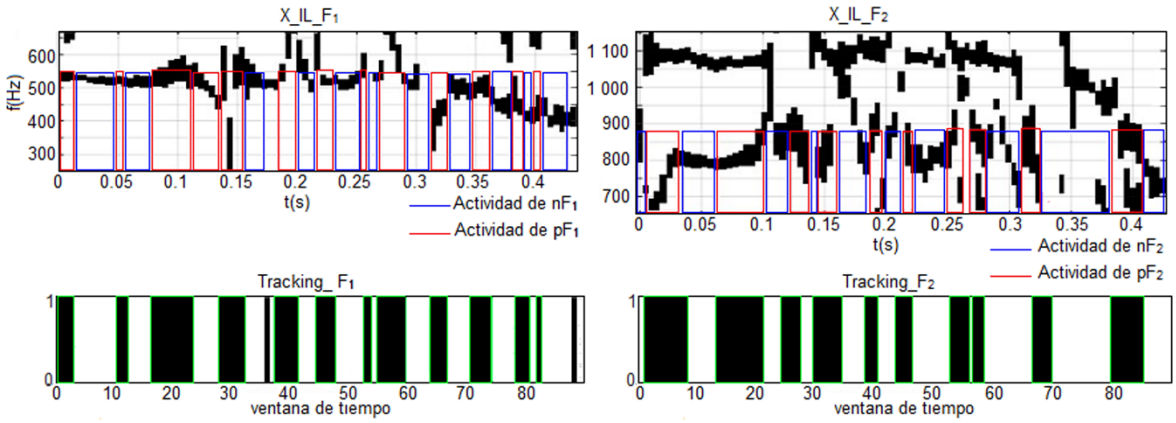

Fig. 8 – Seguimiento dinámico de los formantes para la palabra /foto/ de un locutor femenino.

Conclusiones

La representación espectral de la salida del modelo de fonación humano corresponde con las características espectrales de salida de la cóclea, lo cual permitió modelar el sistema auditivo periférico. Se modeló el proceso de inhibición lateral entre neuronas auditivas del núcleo coclear. En este se apreció como se define a la salida de este proceso las frecuencias características que conforman la señal del habla, los formantes. Con respecto a la modelación de las neuronas básicas del sistema auditivo central, CF, FM y NB se puede concluir que: Con la salida de las neuronas CF se puede detectar las características relativamente estables en los formantes y que estas neuronas responden a diferentes bandas estrechas de frecuencias centradas en una específica y están organizadas de forma tonotópica; Las neuronas de FM son especializadas en detectar cambios en la frecuencia y su función es crucial en la detección de características dinámicas del habla. En particular, las neuronas pFM y nFM permiten detectar las pendientes positivas y negativas de los formantes. Con estas se pueden detectar las regiones temporales o dinámicas correspondientes a los cambios de fonemas. Las

neuronas NB se pueden emplear para la detección de ráfagas de ruido en sonidos sordos o no sonoros, como una reacción del sistema auditivo ante estímulos de banda ancha. Con respecto a la modelación del seguimiento dinámico de los formantes se puede concluir que las neuronas pFM y nFM juegan un papel fundamental en este proceso. Con el modelo implementado se logró detectar los cambios positivos y negativos de las pendientes en una secuencia de voz.

Referencias

Ortega-Garcia, J., J. Gonzalez-Rodriguez and v. Marrero-Aguiar. Ahumada: A Large Speech Corpus In Spanish For Speaker Characterization And Identification. 2000. 2-3, Speech Communication, Vol. 31.

Thomassin, J.-M. And P. Barry. Anatomía Y Fisiología Del Oído Externo. 2016. 3, S.L. : Emc-Otorrinolaringología, Vol. 45.

Antonietti, A., C. Casellato, E. D'angelo And A. Pedrocchi. Bioinspired Adaptive Spiking Neural Network To Control Nao Robot In A Pavlovian Conditioning Task.. 2018. S.L. : 7th Ieee International Conference On Biomedical Robotics And Biomechatronics (Biorob).

PX, Joris, C. Bergevin, R. Kalluri, M. Mc Laughlin, Et Al. Frequency Selectivity In Old-World Monkeys Corroborates Sharp Cochlear Tuning In Humans. 2011. 42, S.L. : Proceedings Of The National Academy Of Sciences, Vol. 108.

Hernández-Zamora, E. And A. Poblano. La Vía Auditiva: Niveles De Integración De La Información Y Principales Neurotransmisores. 2014. 5, S.L. : Gaceta Médica De México, Vol. 150.

Gómez-Vilda, P., J. M. Ferrández-Vicente, V. Rodellar-Biarge, A. Álvarez-Marquina, Et Al. Neuromorphic Detection Of Speech Dynamics. 2011. 8, S.L. : Neurocomputing, Vol. 74.

Gómez-Vilda, P., J. M. Ferrández-Vicente, V. Rodellar-Biarge, R. Martínez-Olalla, Et Al. Neuromorphic Speech Processing: Objectives And Methods. 2011. S.L. : In Machine Audition: Principles, Algorithms And Systems. Igi Global.

Alper, C. M., M. Luntz, H. Takahashi, S. N. Ghadiali, Et Al. Panel 2: Anatomy (Eustachian Tube, Middle Ear, And Mastoid—Anatomy, Physiology, Pathophysiology, And Pathogenesis). 2017. 4, S.L. : Otolaryngology–Head And Neck Surgery, Vol. 156.

De Nava, A. S. L. And S. Lasrado. Physiology, Ear. 2019.

Hixon, T. J., G. Weismer And J. D. Hoit. Preclinical Speech Science: Anatomy, Physiology, Acoustics, And Perception.2018. Plural Publishing.

Kandeler, S. And S. Jesseltm Hudspethaj. Principles Of Neural Science 5th Ed. New York : Ny: Mcgraw-Hill Education, 2013.

Musiek, F. E. And J. A. Baran. The Auditory System: Anatomy, Physiology, And Clinical Correlates. 2018 S.L. : Plural Publishing.

Gómez-Vilda, P., J. M. Ferrández-Vicente, V. Rodellar-Biarge And R. Fernández-Baíllo. Time-Frequency Representations In Speech Perception. 2009. 4-6, S.L. : Neurocomputing, Vol. 72.

Vicente, Jose Manuel Ferrandez. Estudio Y Realización De Una Arquitectura Jerárquica Bioinspirada Para El Reconocimiento Del Habla. 1998. Madrid : Universidad Politécnica De Madrid.

Gómez-Vilda, P., V. Rodellar-Biarge, C. Muñoz, L. M. Mulasa, Et Al. Vowel-Consonant Speech Segmentation By Neuromorphic Units. 2011. Biology, Computation And Linguistics: New Interdisciplinary Paradigms.

Conflicto de interés

Los autores autorizan la distribución y uso de su artículo.

Contribuciones de los autores

1. Conceptualización: Ernesto Arturo Martínez Rams.
2. Curación de datos: Viviana Abad Peraza.
3. Análisis formal: Viviana Abad Peraza.

4. Adquisición de fondos: Ernesto Arturo Martínez Rams.
5. Investigación: Viviana Abad Peraza.
6. Metodología: Ernesto Arturo Martínez Rams.
7. Administración del proyecto: Ernesto Arturo Martínez Rams.
8. Recursos: Viviana Abad Peraza.
9. Software: Ernesto Arturo Martínez Rams.
10. Supervisión: Ernesto Arturo Martínez Rams.
11. Validación: Viviana Abad Peraza.
12. Visualización: Viviana Abad Peraza.
13. Redacción - borrador original: Viviana Abad Peraza.
14. Redacción – revisión y edición: Ernesto Arturo Martínez Rams.