Método de Línea de Transmisión aplicado a la Acústica del Tracto Vocal a través de un Modelo 3D Reconstruido

Alexandre S. Brandão⁽¹⁾, Edson Cataldo⁽²⁾, Fabiana R. Leta⁽¹⁾

(1) Universidad Federal Fluminense, Departamento de Ingeniería Mecánica y Programa de Posgrado en Ingeniería Mecánica. R. Passo da Pátria, 156, bloco D, sala 302, CEP: 24210-240, São Domingos, Niterói, RJ - Brasil. (e-mail: abrand@operamail.com, fabiana@ic.uff.br)

(2) Universidad Federal Fluminense, Departamento de Matemática Aplicada. Programa de Posgrado en Ingeniería de Telecomunicaciones. R. Mário Santos Braga, S/N, CEP: 24.020-140, Valonguinho, Niterói, RJ - Brasil. (e-mail: ecataldo@im.uff.br)

Recibido Ago. 03, 2011; Aceptado Oct. 04, 2011; Versión Final recibida Oct. 11, 2011

Resumen

Se ha simulado la propagación acústica en el tracto vocal humano usando el Método de La Línea de Transmisión (TLM). Se construyeron mallas tridimensionales a partir de segmentaciones de imágenes de resonancia magnética. Inicialmente se construyen modelos simples de uno y dos tubos y luego se simula la misma señal en un modelo de tracto vocal humano cuya forma representa la producción de la vocal abierta /a/, simulando también la propagación de una señal glotal. Para el modelo del tracto vocal, las señales de voz generadas mediante las simulaciones fueron comparadas con muestras de voz del mismo sujeto a quien se extrajo la malla. Se propone un factor de compensación de la dispersión numérica dependiente de la topología para mejorar la determinación de los formantes. Se concluye que el factor de compensación mejora la eficacia y exactitud del método TLM.

Palabras clave: Método de La Línea de Transmisión, Dispersión numérica, Producción de voz

Transmission Line Method applied to vocal tract acoustics through a 3D reconstructed model

Abstract

The acoustic propagation in the human vocal tract using the Transmission Line Matrix method (TLM) has been simulated. Tri-dimensional meshes are built from segmented magnetic resonance images. Initially, simple models of one and two tubes were constructed to then simulate the same signal in a human vocal tract model whose shape represents the production of the open vowel /a/, simulating at the same time a human glottal signal. For the vocal tract model, the voice signals generated by the simulations were compared with voice samples from the same subject from whom the mesh was extracted. A compensation factor for the topology-dependent numerical dispersion for improving formant determination is proposed. It is concluded that the proposed compensation factor improves the effectiveness and accuracy of the TLM method.

Keywords: Transmission Line Method, Numerical Dispersion, Voice Production

INTRODUCCIÓN

La comprensión de la producción de la voz humana como un proceso mecánico, debido a su complejidad y flexibilidad, sigue siendo un tema abierto y muchos investigadores han estudiado este tema y técnicas relacionadas durante los años (Fant, 1970; Cataldo et al., 2006, 2008; Martínez et. al, 2010). Este artículo analiza la propagación del sonido a través del tracto vocal (la parte que va desde la glotis hasta la boca), el encargado de filtrar y amplificar la señal de la glotis, originando los sonidos sonoros. En general, los modelos de tubo se utiliza para relacionar las configuraciones de la forma del tracto vocal a las frecuencias de los formantes observado, pero, en este documento, además, imágenes de resonancia magnética (MRI) de un tracto vocal humano son extraídos con el individuo manteniendo al mismo tiempo una forma de la vocal /a/ fijo teniendo en cuenta la lengua portuguesa de Brasil (equivalente al portugués europeo para la vocal /a/ abierta). Entonces, una malla TLM tridimensional para TLM se construye sobre los voxels segmentados de la secuencia de resonancia magnética y una simulación TLM se realiza para evaluar sus frecuencias de resonancia. La señal de la vocal /a/ resultante de la simulación TLM se compara (en el rango de 0 a 10.000 Hz) con los registrados de la persona en posición supina y en posición vertical.

Clásicamente, el método de elementos finitos (FEM) se ha utilizado para evaluar las frecuencias de resonancia del tracto vocal (llamadas formantes de la voz). Por otro lado, el Método de los Elementos de Contorno (BEM) puede acomodar fácilmente las fronteras de geometría compleja, como el tracto vocal. Además, el BEM puede modelar las regiones con mayor precisión que el FEM, porque todas las aproximaciones consideradas se limitan a la superficie del modelo. Una de las limitaciones típicas de las formulaciones de BEM es dar lugar a matrices totalmente pobladas, lo que aumenta los requerimientos de almacenamiento y la complejidad computacional. Esta limitación se puede resolver con el uso del enfoque *fast multipole* BEM (Bapat et al., 2009). Teniendo en cuenta los problemas acústicos, sea en modelos FEM (Arenas et al., 2006) o BEM, la formulación matemática es basada en la ecuación de Helmholtz de modo que se debe calcular la solución que indica el comportamiento del campo acústico para cada frecuencia inspeccionada.

Otra alternativa es utilizar el método TLM, que se aplicó originalmente para simular los campos electromagnéticos en geometrías complejas (Johns y Beurle, 1971). Por otro lado, el método TLM también se ha aplicado a diferentes problemas mecánicos y su adecuación a la acústica ya ha sido demostrada, por ejemplo en (Katsamanis y Maragos, 2008), donde los autores han investigado el potencial de la correcta aplicación del TLM 3D para simulaciones en el tracto vocal para la síntesis de las consonantes fricativas. La geometría del tracto vocal se determinó a partir de resonancia magnética 3D. El método TLM produce una señal de salida en el dominio del tiempo en lugar de un campo de presión estática por cada componente de frecuencia, que es proporcionada por los métodos FE y BE. La señal de salida en el dominio del tiempo se puede convertir fácilmente en el dominio de frecuencia utilizando el algoritmo de la transformada rápida de Fourier (FFT). Por otra parte, la respuesta en frecuencia de un determinado modelo también se puede obtener mediante la aplicación de la técnica de codificación de predicción lineal (LPC) a la señal de salida en el dominio del tiempo.

Otras ventajas de TLM son la aplicación más fácil y el hecho de que las condiciones de contorno se especifican en los nodos deseados, simplemente mediante el establecimiento de los valores para los coeficientes de transmisión y reflexión. TLM también se conoce como el método de guía de ondas digitales (DWM) entre algunos investigadores. Su principal limitación es el error de dispersión, ya que la velocidad de propagación de una onda depende de su frecuencia y la dirección de viaje. Esto conduce a errores de propagación de las ondas y la consecuente diferencia en los valores que se esperaban para las frecuencias de resonancia. El grado de error de dispersión es altamente dependiente de la topología de las mallas (Fontana y Rocchesso, 2001; Campos y Howard, 2005). Murphy et al. (2007) mencionan que la minimización de la dispersión se puede lograr mediante el uso de topologías de malla interpoladas o dodecaédricas. Por otra parte, la construcción de mallas de rejilla uniforme es más fácil y menos costosa computacionalmente. Otros enfoques han tratado de resolver el problema de la dispersión numérica con las técnicas de transformaciones en la escala de frecuencias (*frequency warping*),

que se utilizan para corregir los formantes desajustados (Savioja y Välimäki, 2003). La dispersión numérica en función de la dirección es muy difícil de ser compensado por el efecto cambia dependiendo de las distancias entre los nodos de origen y destino (Speed, 2008).

Síntesis de voz articulatoria implica el cálculo de la función del área del tracto vocal, lo que se puede hacer: (i) directamente a partir de datos de RM (Clément et al., 2007); (ii) a partir de perfiles mediosagital 2D de RM o; (iii) a partir de modelos 3D reconstruidos de RM (Ventura et al., 2009). Esta función es entonces utilizada para configurar un modelo 1D TLM o DWM para sintetizar el habla. En esto, las simulaciones se ejecutan directamente en las mallas TLM 3D extraídos de las secuencias de resonancia magnética. Ambos métodos tienen sus ventajas y aplicaciones. Mientras que los modelos de malla 3D TLM pueden determinar con mayor precisión los formantes, no pueden proporcionar en tiempo real la síntesis en un ordenador común. Por otro lado, los modelos de articulación pueden ofrecer la síntesis de voz con calidad suficiente, pero los formantes no son tan precisos como en los modelos 3D TLM. También en este trabajo, un factor de compensación se propone que se elimina la dispersión numérica dependiente de la topología en mallas de red 3D uniformes TLM, proporcionando la mejora en la determinación de los formantes por el método TLM.

EL MÉTODO DE LA LÍNEA DE TRANSMISIÓN (TLM)

El TLM se aplicará para simular la propagación del sonido, descrito por la ecuación de onda acústica, teniendo en cuenta el dominio como una red de líneas de transmisión, donde los pulsos acústicos se propagan. Una malla 3D TLM se compone de puntos y líneas ortogonales (Fig. 1). Los valores nodales representan la presión $p(r,t)$, la solución de la ecuación de onda acústica, en una posición dada $r(x,y,z)$, en el instante t . En cada paso de tiempo, cada nodo recibe impulsos de la presión de incidentes y envía los impulsos dispersos, generado con la llamada matriz de dispersión, lo que presenta la relación entre los pulsos de entrada y salida.

Fig. 1: The 3D TLM. (a) La propagación de pulsos en el instante t . (b) la propagación de pulsos en el instante $t+\Delta t$.

Varios modelos de dispersión se pueden encontrar en la literatura. En este trabajo, el modelo utilizado será el nodo 3D de derivación sin pérdidas, ya que tiene mucho éxito en la solución de los problemas acústicos.

La relación entre los coeficientes de transmisión (τ) y de reflexión (ρ) en las puertas de un nodo 3D de derivación sin pérdidas y la matriz de dispersión, que se indican a continuación, se derivan en (Cogan et al., 2006). La Eq. (1) muestra la relación entre (τ) y (ρ):

$$\tau = (1 - \rho) \frac{Z_T}{Z} \quad (1)$$

donde Z es la impedancia acústica de una única línea de transmisión y Z_T es la impedancia en paralelo de las líneas conectadas a una dada línea. La matriz de dispersión está dada por Eq. (2).

$$\begin{bmatrix} V_1^O \\ V_2^O \\ V_3^O \\ V_4^O \\ V_5^O \\ V_6^O \end{bmatrix} = \begin{bmatrix} \rho & \tau & \tau & \tau & \tau & \tau \\ \tau & \rho & \tau & \tau & \tau & \tau \\ \tau & \tau & \rho & \tau & \tau & \tau \\ \tau & \tau & \tau & \rho & \tau & \tau \\ \tau & \tau & \tau & \tau & \rho & \tau \\ \tau & \tau & \tau & \tau & \tau & \rho \end{bmatrix} \begin{bmatrix} V_1^I \\ V_2^I \\ V_3^I \\ V_4^I \\ V_5^I \\ V_6^I \end{bmatrix} \quad (2)$$

donde los subíndices representan los números de las puertas del nodo para los pulsos de presión de entrada V^I y salida V^O .

El método más utilizado para minimizar la dispersión numérica dependiente de frecuencia es establecer el paso de discretización espacial (ΔL) mucho menor que la longitud de onda (λ), de acuerdo con la Eq. (3).

$$\frac{\Delta x}{\lambda} \leq 0.1 \quad (3)$$

que asegura que la diferencia de velocidad de propagación entre las ondas de diferentes longitudes de onda será pequeña. Los valores considerados para ΔL son 1 mm para las mallas de los modelos de tubos y 0.968498 mm para la malla del tracto vocal. Estos valores se explican en la sección siguiente. Por lo tanto, de acuerdo con la Eq. (3), los valores de frecuencia máxima que se pueden viajar a través de la malla son 34.310 Hz y 35.425 Hz para los modelos de tubo y las mallas del tracto vocal, respectivamente.

Una vez que ΔL se define, la discretización temporal, Δt , puede ser definido por la Eq. (4):

$$\Delta t = \frac{\Delta x}{cD} \quad (4)$$

donde c es la velocidad de propagación del sonido en el aire. El factor de D es necesario para compensar la dispersión numérica dependiente de la topología, como se mencionó antes. Por otro lado, un valor global de este factor es muy difícil de encontrar, debido a que el efecto acumulado cambia dependiendo de la distancia entre los nodos de origen y destino.

EXTRACCIÓN DE MALLA

Las mallas TLM se construyen sobre los voxels (elementos volumétricos de imagen), que tienen la etiqueta de objeto (o escala de grises) de los volúmenes segmentados MRI. En consecuencia, la malla TLM extraída va a coincidir con el objeto segmentado la medida de lo posible, dependiendo únicamente del tamaño de los voxels. La imagen 3D debe ser isótropa, con el fin de obtener una malla uniforme. El proceso se aplicará y se discutirán, en primer lugar, a las mallas modelo de tubo y luego a la malla del tracto vocal.

Mallas de los Modelos de Tubos

El proceso de obtención de las mallas del tubo se realiza en cuatro pasos: (i) La secuencia de imágenes para la generación de la malla del tubo se construye a partir de un solo dibujo (en este caso, el editor de imágenes Kolourpaint (Dang, 2003) se utiliza para dibujar la primera rebanada del tubo), (ii) Esta única imagen se guarda en el formato DICOM con el editor de imágenes Gimp (Mattis y Kimball, 2001), (iii) Varias copias de esta imagen DICOM se crean generando una

secuencia de imágenes que se abre y se adjuntará a una imagen de volumen con la longitud del tubo deseado. Todo este proceso se lleva a cabo en un programa de código abierto desarrollado en esta investigación, que se llama ModaVox, y, finalmente, (iv) Las rebanadas de la imagen de volumen son segmentadas, cada una con un único disco, añadiendo en una imagen segmentada 3D a partir de la cual se extrae la malla TLM. Para construir un modelo de malla de dos tubos con secciones de diferentes diámetros, dos secuencia de imágenes se debe crear, como se describió anteriormente. Para un mejor modelado del efecto de la radiación, las mallas de tubo tienen terminaciones con 7 píxeles de largo y 4 píxeles mayor que el diámetro de la sección del tubo anterior. Para todos los modelos de tubo de la discretización del espacio ΔL es igual a 1 mm.

Malla del Tracto Vocal

El proceso de generación de la malla del tracto vocal se realiza en cuatro pasos: (i) La secuencia de RM sagital en 3D se extrae de un sujeto para la forma de la vocal /a/, con un tiempo de adquisición de 18 min 39 seg. El sujeto emite el /a/ periódicamente, ayudando a sí mismo para mantener el tracto vocal en una posición fija durante la extracción de la secuencia, la cual estuvo compuesta inicialmente por 186 rebanadas de 512 X 512 píxeles. La secuencia fue adquirida utilizando una máquina General Electric, modelo GE Medical Systems HDxt Signa, con una fuerza de campo magnético de 1.5 Tesla, utilizando los siguientes parámetros: T1 imagen de resonancia magnética, secuencia de exploración (GR), bobina de cabeza de 8 canales (8HRBRAIN), tiempo de eco de 4,744 ms, tiempo de repetición de 10,584 ms, campo de visión (FOV) 48 X 48 cm, espaciamiento en píxeles (0.9375, 0.9375) mm y 1 mm de espesor de rebanada con ningún hueco, por lo tanto, una imagen anisótropa; (ii) Las dimensiones del volumen de interés (VOI) se pasan, a través de la interfaz del ModaVox, para un objeto de la clase vtkExtractVOI de la VTK de la biblioteca (Schroeder et al., 2002), que extrae el VOI, lo que reduce la imagen del volumen original, que es sobremuestreada para eliminar la anisotropía. Por lo tanto, las dimensiones finales del voxel fueron (0.968498, 0.968498, 0.968498) mm, (iii) En esta nueva imagen isotrópica en 3D, la segmentación se realiza a través de redes neuronales, eliminación de islas y algunos ajustes manuales (todas estas operaciones se realizaron con el programa ModaVox); La parte de los dientes fue segmentado de forma manual. (iv) Por último, la malla TLM uniforme se extrae de la imagen en 3D segmentada.

DETALLES DE LAS SIMULACIONES

Atributos de las Mallas

Cada nodo de la malla tiene la presión y los valores límite asociados. Los valores de presión son las soluciones de la correspondiente ecuación de onda acústica, en cada nodo, en cada momento. Los valores de contorno definen las condiciones reflectantes, que se van a aplicar, con el establecimiento de los valores específicos de los parámetros ρ y τ en la Eq. (2). Los valores de contorno y sus correspondientes significados son los siguientes: (i) Valor 0 - nodos interiores ($\rho=-2/3$ y $\tau=1/3$); (ii) Valor 1 - condición reflexiva ($\rho=1$ y $\tau=0$); (iii) Valor 2 - nodos de entrada (También nodos interiores), (iv) Valor 3 - condición de espacio libre ($\rho=-1$ y $\tau=0$).

Señales de Entrada

Una de las señales de entrada representando la variación de presión es formado por 200 componentes de frecuencia, que van desde 50 a 10.000 Hz, cada uno con amplitud unitaria, dado por la Eq. (5):

$$entrada = \sum_{n=1}^{200} \cos(2\pi f_n k \Delta t) \quad (5)$$

donde k es el número del paso de la iteración y Δt es el intervalo de tiempo de discretización TLM dado por la Eq. (4). Esta señal se utiliza como entrada tanto en las mallas de concatenación de tubos como en la malla del tracto vocal. La otra señal de entrada utilizada fue la señal glotal, que

Señales de Salida

TLM APLICADO A LOS MODELOS DE TUBO

Primeramente, el TLM se aplicó a los modelos de concatenación de tubos para comparar los resultados con los que están disponibles en la literatura. Las simulaciones se ejecutan en mallas de modelos de tubo único, que se ilustran en la Fig. 2. Las simulaciones se realizaron considerando $c=343.1$ m/s como la velocidad del sonido en el aire, con $D=1$ en la Eq. (4), y las funciones de respuesta en frecuencia se muestran en la Fig. 3.

Fig. 3: Funciones de respuesta en frecuencia calculadas a partir de simulaciones en las mallas de tubos, considerando $c=343.1$ m/s, $d = 24$ mm de longitud y $L=170$ mm de largo. (a) Tubo abierto. (b) Tubo cerrado.

En el caso del tubo abierto, las frecuencias de resonancia obtenidas fueron: $f_1= 550.67$ Hz, $f_2= 1101.34$ Hz, etc. Por otro lado, los valores deben ser de $f_1=1009.12$ Hz, $f_2=2018.23$ Hz, etc., de acuerdo con la Eq. (6), bien conocida de la literatura:

$$f_n = nc / 2L \quad (6)$$

donde $c=343.1$ m/s es la velocidad del sonido en el aire y L es la longitud del tubo.

En el caso del tubo cerrado, los valores obtenidos para las frecuencias de resonancia en las simulaciones, fueron de $f_1=275.335$ Hz, $f_2=826.005$ Hz, etc. Por otro lado, los valores deben ser de $f_1=504.55$ Hz, $f_2=1513.67$ Hz, etc., de acuerdo con la Eq. (7), que también es bien conocida de la literatura:

$$f_n = (2n - 1)c / 4L \quad (7)$$

En los dos casos presentados anteriormente, teniendo en cuenta los valores de las frecuencias de resonancia obtenidos de las simulaciones (ver Fig. 3) y las ecuaciones correspondientes para los tubos abierto y cerrado, los dos modelos de tubo único parecen tener una longitud de $L=311.529$ mm, teniendo en cuenta $c=343.1$ m/s.

Los tubos parecen ser más largos debido a que la velocidad de propagación del sonido es menor de lo que debería ser. Por lo tanto, un factor de compensación de la velocidad se propone: usando la Eq. (6) con $L=0.17$ m y la primera frecuencia de resonancia ($f_1= 550.67$ Hz) para la simulación de tubo abierto (ver Fig. 3), es posible obtener la aparente velocidad del sonido, la cual está dada por $c=187.2278$ m/s. Del mismo modo, para la malla de tubo cerrado, usando la Eq. (7) con $L=0.17$ m y la primera frecuencia de resonancia ($f_1=275.335$ Hz) para la simulación de tubo cerrado (ver Fig. 3) da $c = 187.2278$ m/s para la velocidad aparente del sonido. Por lo tanto, los modelos de tubo único proporcionan una forma eficiente de determinar un valor global para la velocidad aparente del sonido en las simulaciones en que se consideran mallas TLM 3D de rejilla uniforme. La relación entre las velocidades real (343.1 m/s) y aparente (187.2278 m/s) del sonido en la malla TLM 3D uniforme da el factor de compensación de la velocidad (1.8325) para las mallas TLM uniforme, lo que se debe utilizar en vez de $D=1$ en la Eq. (4).

Entonces, la velocidad del sonido se puede cambiar a $c=343.1 \times 1.8325 = 628.7307$ m/s, con el mantenimiento de los 170 mm de longitud de los tubos. O, alternatively, debido a la linealidad de la Eq. (4), es posible construir un nuevo tubo de malla considerando $L=170 / 1.8325 = 92.77 \cong 93$ mm de largo, con el mantenimiento de $c=343.1$ m/s. A pesar de que las dos alternativas funcionan, es más sencillo aumentar el valor de la velocidad que cambiar la estructura de las mallas. Además, las simulaciones con tubos mas cortos ($L=93$ mm) también han traído las frecuencias de resonancia a los valores correctos, confirmando que el factor de compensación es correcto.

Considerando entonces un tubo con 170 mm de largo, 24 mm de diámetro y la compensación de la velocidad de propagación del sonido ($c=343.1 \times 1.8325 = 628.7307$ m/s), los valores de las frecuencias de resonancia son compatibles con los obtenidos analíticamente, incluso colocando el nodo de origen cercano a las paredes del tubo (Fig. 4).

Fig. 4: Funciones de respuesta en frecuencia calculadas a partir de simulaciones en las mallas de tubo único, considerando $c=628.7307$ m/s, $L=170$ mm y $d=24$ mm. (a) Tubo abierto. (b). Tubo cerrado.

Modelo de Dos Tubos de un tracto vocal para la Simulación de la vocal /a/

El TLM y el factor de compensación discutido antes se aplican a un modelo de dos tubos utilizados para la generación de un sonido de la vocal /a/ (Fig. 5).

Fig. 5: Modelo de dos tubos correspondiente a la generación de la vocal /a/.

Para el modelo de dos tubos, los valores analíticos de las frecuencias de resonancia se estimaron mediante el cálculo del módulo de impedancia acústica $Z(z)$ en el punto de transición entre las dos secciones, dado por la Eq. (8):

$$Z(z) = \frac{\rho_a c \cos\left(\frac{\omega L_1}{c}\right) \sin\left(\frac{\omega L_2}{c}\right)}{A_1 \sin\left(\frac{\omega L_1}{c}\right) \sin\left(\frac{\omega L_2}{c}\right) - A_2 \cos\left(\frac{\omega L_1}{c}\right) \cos\left(\frac{\omega L_2}{c}\right)} \quad (8)$$

donde L_1 , L_2 , A_1 y A_2 son, respectivamente, las longitudes y secciones transversales de las dos secciones de tubos, $\rho_a = 1.204 \text{ kg/m}^3$ es la densidad del aire y $c = 343.1 \text{ m/s}$ es la velocidad de propagación del sonido en el aire. Teniendo en cuenta $L_1 = 90 \text{ mm}$, $L_2 = 80 \text{ mm}$, $A_1 = 100 \text{ mm}^2$, $A_2 = 700 \text{ mm}^2$ y $\omega = 2\pi f$ en la Eq. (8), los valores analíticos del módulo de impedancia acústica se pueden trazar en el rango de frecuencia deseada para la comparación con los resultados de la simulación.

Para construir el modelo de malla de dos tubos para generar una vocal /a/, se utilizaron las mismas dimensiones presentadas anteriormente. Sin embargo, debido a la aproximación de la sección de tubo de menor diámetro ($d_1 = 2(A_1/\pi)^{(1/2)}$) a $d_1 = 12 \text{ mm}$, la sección de tubo 2 de diámetro (d_2) se aproximó a los 32 mm (ver Fig. 5), de modo que fuera mantenida la relación entre los valores originales de d_2 y d_1 , y, por consiguiente, entre las áreas de las secciones transversales, tan inalterada cuan posible, es decir,

$$\frac{d_2}{d_1} = \frac{29.85410660}{11.28379167} = 2.6457 \cong 2.6666 = \frac{32}{12} \quad (9)$$

La Figura 6 muestra el módulo de la impedancia acústica, calculado por la Eq. (8) en escala logarítmica, y la función de respuesta en frecuencia obtenida a partir de la simulación.

Fig. 6: Modelo de malla de dos tubos para generar una vocal /a/. (a) Impedancia acústica dada por Eq. (8) en escala logarítmica. (b) Función de respuesta en frecuencia ($c=628.7307 \text{ m/s}$).

En este caso, se hace una comparación entre cantidades diferentes, ya que sólo los picos de resonancia se están comparando. Cabe señalar que los valores de las frecuencias

correspondientes a los picos son muy compatibles, a excepción del pico de resonancia entre 8000 y 9000 Hz (cuando los valores se sustituyen en la Eq. (8) tanto en el numerador como el denominador van a cero).

TLM APLICADO A UNA MALLA DE TRACTO VOCAL HUMANO

Para validar la metodología propuesta, el TLM se aplica a una malla 3D del tracto vocal reconstruida a partir de imágenes de resonancia magnética, cuya forma se ajusta a la de un tracto vocal humano, que corresponde a la producción de la vocal /a/. Para la malla del tracto vocal, que se muestra en la Fig. 7-(a), el valor considerado para ΔL es de 0.968498 mm debido a la eliminación de anisotropía y el valor considerado para la velocidad del sonido c es de 628.7307 m/s, debido al uso del factor de compensación de velocidad descrito anteriormente.

La función de respuesta en frecuencia obtenida a partir de la simulación TLM se compara con el obtenido a partir de una señal de voz humana de la vocal /a/ del mismo individuo de que se extrajo la secuencia de resonancia magnética. Dos muestras humanas de la vocal /a/ se registraron: una con el sujeto en posición supina, similar a la mantenida en el equipo de resonancia magnética, y la otra con el sujeto en posición vertical, ya que se comprobó que la posición del sujeto puede afectar a la forma del tracto vocal (Kitamura et al., 2005). Cabe recordar que las imágenes de RM se obtuvieron con el sujeto en posición supina. La Figura 7-(b), muestra la función de respuesta en frecuencia en tres casos: dos de ellos correspondientes a un tracto vocal humano y otro correspondiente a una simulación.

Fig. 7: (a) Malla del tracto vocal que corresponde a una forma de la vocal /a/. Vista que muestra los valores de contorno. (b) Las respuestas de frecuencia del tracto vocal para la vocal /a/ (Humana vs. Simulación TLM).

En la Fig. 7-(b), los cinco primeros formantes de la señal de salida de simulación TLM son muy cercanos a los de la señal de voz humana grabada en posición supina. La Tabla 1 muestra una comparación (con porcentaje de error) entre las señales de voz humana y TLM teniendo en cuenta la posición supina.

Tabla 1: Comparación de los formantes: vocal /a/ (señal de salida de TLM vs. Voz humana grabada en posición supina).

Formantes (Hz)	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10
Señal de salida TLM	562.5	1200	2358	3510	4198	4650	5258	5800	6868	7550
Voz humana (supino)	533	1143	2300	3517	4383	5650	6695	7885	8815	9815
Error (%)	5.53	4.98	2.52	0.19	4.22	17.69	21.46	26.44	22.08	23.07

En la Tabla 1 se ve que el error se mantiene por debajo de 5.55 % durante los primeros 5 formantes. Los formantes más bajos, debido a las longitudes de onda mayores, son más susceptibles a los efectos de difracción, mientras que los formantes superiores (por encima de 5

kHz), debido a las longitudes de onda más pequeñas, son más susceptibles al efecto de la reflexión. Por lo tanto, es más difícil para los formantes más altos ser los mismos entre distintas emisiones vocales, por mínimas diferencias en la forma del tracto vocal tienen un mayor efecto sobre estos componentes. En la Fig.7-(b), incluso las dos muestras de voz humana tienen formantes diferentes por encima de 5 kHz. Por otra parte, la radiación de sonido que ocurre en la abertura de los labios provoca un efecto de filtro pasa altas, que es más fuerte en la respuesta en frecuencia calculada a partir de la simulación TLM en la Fig. 7-(b). El mismo efecto ocurre también en los resultados de los modelos de tubo, porque como la señal de entrada de la Eq. (5) tiene una magnitud uniforme distribuida en sus componentes de frecuencia, no se puede compensar el efecto del filtro pasa-altas en la señal de salida. En una emisión de voz real, la señal glotal tiene una relación de reducción de armónicos de 12 dB/octava, lo que permite la compensación del efecto pasa-altas causado por la radiación de sonido en las aberturas de los labios (Fig. 8).

Fig. 8: (a) Diagrama de bloques del modelo fuente-filtro de producción de voz. (Adaptado de (Fant, 1970)). (b) Efecto de la señal de prueba en el modelo fuente-filtro.

Para simular las paredes del tracto vocal de los tejidos blandos, la condición de frontera de paredes se cambia a una de no totalmente reflexiva. Como la mayoría del tracto vocal humano consiste en los tejidos blandos, el valor promedio de la impedancia acústica de los tejidos blandos humanos, que está dada por 1630000 Ns/m^3 (Hendee y Ritenour, 2002), se ha aplicado, indicando los valores de $\rho = 0.9974688987$ y $\tau = 0.0005062203$ para los coeficientes de reflexión y de transmisión en las paredes del tracto vocal. La Figura 9 compara el contenido de frecuencia de la señal de salida a partir de dos simulaciones TLM diferentes a la de la señal de voz humana grabada por el sujeto en posición supina.

Fig. 9: Comparación de FFTs. (a) La señal de salida para la simulación TLM con paredes de tejidos blandos y la señal glotal como entrada vs. La voz humana. (b) La señal de salida para la simulación TLM con paredes rígidas y la señal en la Eq. (5) como entrada vs. La voz humana.

Fant y Båvegård (1997) encontraron que la fosa piriforme introduce una depresión espectral alrededor de 5200 Hz en el espectro de voz. Este resultado también es confirmado por mediciones en seres humanos, en la obra de Dang y Honda (1997). Ahora, la misma depresión espectral (alrededor de 5200 Hz) se puede ver en el espectro de FFT de la señal de salida de simulación TLM en la Fig. 9-(a), que muestra que la condición reflexiva representando paredes hechas de tejido blando y el uso de la señal glotal como entrada han aumentado la calidad de la muestra de voz sintética. La Tabla 2 muestra el resumen de las simulaciones. El ordenador utilizado para realizar tiene un procesador de 2.66 GHz y 1GB de RAM.

Tabla 2: Resumen de las simulaciones presentadas.

Malla	Numero de Puntos	ΔL (mm)	c (m/s)	Tiempo de Simulación	Tiempo de Sonido (s)
Fig. 2-(a)	92929	1	343,1	8h 27'	0,2
Fig. 2-(b)	88274	1	343,1	7h 44'	0,2
Fig. 2-(a)	92929	1	628,7307	15h 34'	0,2
Fig. 2-(b)	88274	1	628,7307	14h 35'	0,2
Fig. 5	89666	1	628,7307	16h 10'	0,2
Fig. 7-(a)	148289	0.968	628,7307	27h 12'	0,2

CONCLUSIONES

El modelo numérico desarrollado en este trabajo se puede aplicar para el cálculo de las frecuencias de resonancia de conductos complejos, incluyendo el estudio de la propagación del sonido a través de una malla 3D del tracto vocal reconstruida a partir de imágenes de resonancia magnética, proporcionando excelentes resultados.

Las mallas TLM también pueden mostrar datos escalares, representando la distribución de presión en toda la malla en iteraciones individuales. Sin embargo, la secuencia de todos los valores de presión en el punto de salida, almacenados durante la simulación, proporciona información completa acerca de la frecuencia del modelo en estudio. En comparación con soluciones FEM o BEM, la distribución de la presión en la malla TLM en una sola iteración es relativa a todos los componentes de frecuencia de la señal que está siendo impulsado en la malla. En cambio, en FEM o BEM, una dada distribución de presión es relativa a una sola frecuencia y es proporcionada por la solución de la ecuación de Helmholtz.

El factor de compensación desarrollado se ha aplicado a simulaciones en diferentes modelos y los resultados demuestran que mejora la eficacia y la exactitud del método TLM. El uso de este factor permitió la correcta determinación de los formantes en las estructuras analizadas, eliminando la dispersión numérica dependiente de la topología en mallas TLM 3D de rejilla uniforme.

AGRADECIMIENTOS

Los autores agradecen al Dr. Augusto Alair SMD dos Santos, del Hospital das Clínicas de Niterói por la disposición de los equipos de RM de la clínica ProEcho. Este trabajo fue apoyado por FAPERJ, por CAPES y por CNPq.

REFERENCIAS

Airas, M., Pulakka, H., Bäckström, T., Alku, P., *A toolkit for voice inverse filtering and parametrisation* In: Proceedings of the 9th European Conference on Speech Communication and Technology. pp. 2145–2148 (2005).

Arenas, J. P., Aguayo, J. L., Gerges, S. N. Y. y Pobrete, V. H., *Caracterización de la Impedancia Acústica de Entrada en Cámaras de Expansión Reactivas*. Información Tecnológica, vol.17, no.3, p.117-122 (2006).

- Bapat, M. S., Shen, L., Liu, Y. J., *Adaptive fast multipole boundary element method for three-dimensional half space acoustic wave problems*. Engineering Analysis with Boundary Elements 33 (8-9), 1113–1123 (2009).
- Campos, G. R., Howard, D. M., *On the computational efficiency of different waveguide mesh topologies for room acoustic simulation*. IEEE Trans. Speech Audio Processing 13 (5), 1063–1072 (2005).
- Cataldo, E., Leta, F. R., Lucero, J., Nicolato, L., *Synthesis of voiced sounds using low-dimensional models of the vocal cords and time-varying subglottal pressure*. Mechanics Research Communications 33 (2), 250–260 (2006).
- Cataldo, E., Sampaio, R., Lucero, J., Soize, C., *Modeling random uncertainties in voice production using a parametric approach*. Mechanics Research Communications 35 (7), 454–459 (2008).
- Clément, P. y otros 5 autores, *Vocal tract area function for vowels using three-dimensional magnetic resonance imaging. A preliminary study*. Journal of Voice 21 (5), 522–530 (2007).
- Cogan, D., O'Connor, W., Pulko, S., *Transmission Line Matrix in Computational Mechanics*. CRC Press, Taylor & Francis Group, Boca Raton, Florida, pp. 102–104 (2006).
- Dang, C., 2003. *Kolourpaint is a free, easy-to-use paint program for KDE* (en línea), <http://kolourpaint.sourceforge.net/> Acceso: 16 Octubre (2006).
- Dang, J., Honda, K., *Acoustic characteristics of the piriform fossa in models and humans*. Journal of the Acoustical Society of America 101 (1), 456–465 (1997).
- Fant, G., *The Acoustic Theory of Speech Production*, 2nd Edition. Mouton, The Hague, pp. 66 (1970).
- Fant, G., Båvegård, M., *Parametric model of VT area functions: vowels and consonants*. Speech, Music and Hearing - Quarterly Progress and Status Report 38 (1), 001–020 (1997).
- Fontana, F., Rocchesso, D., *Signal-theoretic characterization of waveguide mesh geometries for models of twodimensional wave propagation in elastic media*. IEEE Trans. Speech Audio Processing 9 (2), 152–161 (2001).
- Hendee, W. R., Ritenour, E. R., *Medical Imaging Physics*. Wiley-Liss, New York, pp. 312 (2002).
- Johns, P. B., Beurle, R. L., *Numerical solution of two-dimensional scattering problems using a transmission-line matrix*. In: Proceedings of IEE. Vol. 118. pp. 1203–1209 (1971).
- Katsamanis, A., Maragos, P., *A fricative synthesis investigations using the transmission line matrix method*. Journal of the Acoustical Society of America 123, 3741 (2008).
- Kitamura, T. y otros 9 autores, *Difference in vocal tract shape between up right and supine postures: Observations by an open-type MR scanner*. Acoustical Science and Technology 26, 465–468 (2005).
- Martínez, A., Pineda, Z. y Ramos, A. *Ocultamiento del Mensaje de Señales de Voz usando la Transformada de Ondita Haar Discreta*. Información Tecnológica, vol.21, no.3, p.135–140 (2010).
- Mattis, P., Kimball, S., 2001. *Gnu image manipulation program* (en línea), <http://www.gimp.org/> Acceso: 13 Mayo (2007).
- Murphy, D., Kelloniemi, A., Mullen, J., Shelley, S., *Acoustic modeling using the digital waveguide mesh*. IEEE Signal Processing Magazine 24 (2), 55–66 (2007).

- Savioja, L., Välimäki, V., *Interpolated rectangular 3-d digital waveguide mesh algorithms with frequency warping*. IEEE Trans. Speech Audio Processing 11 (6), 783–789 (2003).
- Schroeder, W., Martin, K., Lorensen, B., *The Visualization Toolkit: An Object-Oriented Approach to 3-D Graphics*, 3rd Edition. Kitware, Inc (2002).
- Speed, M. D. A., *Modelling sound propagation in the vocal tract with a three-dimensional digital waveguide mesh*. Tesis de Magister, University of York, Heslington, Chp. 2 (2008).
- Ventura, S. M. R., Freitas, D. R. S., Tavares, J. M. R. S., *Application of MRI and biomedical engineering in speech production study*. Computer Methods in Biomechanics and Biomedical Engineering 12 (6), 671–681 (2009).