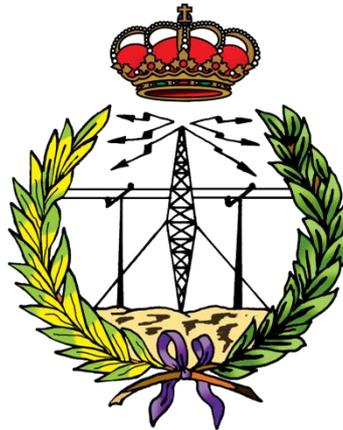


**ESCUELA UNIVERSITARIA DE INGENIERÍA TÉCNICA
DE TELECOMUNICACIÓN**



PROYECTO FIN DE CARRERA

Mejora de la interfaz del simulador de voz Vox

Cristina Iglesias Marugán

Septiembre 2012

*Son muchas las manos y corazones
que contribuyen al éxito de una persona.*

W.D.

Resumen

A partir de un simulador de vocales denominado Vox, programado en MATLAB, desarrollado originalmente en la Universidad Técnica de Aquisgrán por Malte Kob [1] y mejorado en el Departamento de ICS de la EUITT [2], se pueden generar voces sintéticas.

La principal limitación del simulador es que sólo puede generar *vocales* sintéticas, además la simulación se realiza a partir de parámetros anatómicos y fisiológicos fijos. La estructura actual del programa dificulta la modificación rápida de cualquiera de los parámetros básicos de la misma, circunstancia que podría mejorar mediante una interfaz gráfica.

El proyecto consistirá, por un lado, en completar el simulador haciendo posible también la síntesis a partir de los parámetros anatómicos de hombres, mujeres y niños; y por otro, en el diseño e implementación de una interfaz gráfica de usuario que nos permita seleccionar los diferentes parámetros físicos para la simulación y recoger los resultados de la misma de manera más sencilla.

Abstract

Starting from a vowels simulator called Vox, programmed in MATLAB, originally developed in the Technical college of Aquisgrán by Malte Kob [1] and improved in the ICS Department of the EUITT [2], with this programme you can generate synthetic voices.

The main limitation of the simulator is that it only can generate synthetic vowels; moreover the simulation is made from anatomical and physiological fixed parameters. The current structure of the programme complicates the quick modification of any of the basic parameters of it, circumstance that could be improved through a graphic interface.

On the one hand, the project consists in completing the simulator doing the synthesis possible, from the anatomical woman, men and children parameters; on the other hand, the design and implementation of a graphic user interface, that allow us to select different physical parameters to the simulation and gather the results of it in a simple way.

Índice de Contenidos

ÍNDICE DE FIGURAS.....	5
ÍNDICE DE TABLAS	7
LISTA DE ACRÓNIMOS	8
1. INTRODUCCIÓN	10
2. MARCO TECNOLÓGICO	14
2.1. Producción de la voz humana.....	16
2.1.1. Anatomía de la producción vocal.....	16
2.1.2. Fisiología de la fonación	20
2.2. Simulador Vox	22
2.2.1. Modelo de las cuerdas vocales (CV).....	24
2.2.2. Modelo del tracto vocal (TV).....	26
2.2.3. Unión entre ambos modelos.....	28
2.3. GUI en MATLAB®	29
2.3.1. Construcción interactiva de una GUI.....	29
3. DESCRIPCIÓN DE LA SOLUCIÓN PROPUESTA.....	34
3.1. Implementación de la interfaz	36
3.2. Modificaciones en Vox	41
3.3. Parametrizaciones.....	44
3.3.1. Parámetros de las cuerdas vocales	44
3.3.2. Parámetros del tracto vocal	46
3.3.3. Parámetros de simulación	47
4. RESULTADOS	50
4.1. Almacenamiento de variables.....	52
4.2. Simulación según el locutor elegido.....	54
4.3. Frecuencias fundamentales.....	61
4.4. Variación de otros parámetros.....	63
5. CONCLUSIONES Y LINEAS FUTURAS.....	70
6. REFERENCIAS BIBLIOGRÁFICAS	74

Índice de Figuras

Figura 1. Aparato fonador humano	16
Figura 2. Estructura histológica CV [5]	17
Figura 3. Diferentes esquemas usados para la estructura en capas de las CV [5].....	18
Figura 4. Esquema fases ciclo glótico	21
Figura 5. Funcionamiento simulador Vox	23
Figura 6. Vista en sección (arriba) y lateral (abajo) del modelo de CV de Kob [1] .	24
Figura 7. Disposición de las CV del modelo.....	24
Figura 8. Tracto vocal modelado [7].....	26
Figura 9. Concatenación de cilindros. Ondas acústicas reflejadas y transmitidas ...	27
Figura 10. Unión entre el modelo de las cuerdas vocales y el del tracto vocal [2]...	28
Figura 11. Acceder a GUIDE en MATLAB	30
Figura 12. Ventana GUIDE Quick Start	30
Figura 13. Ventana de diseño en GUIDE.....	31
Figura 14. Barra de herramientas en GUIDE.....	31
Figura 15. Ventana interfaz de usuario inicio	36
Figura 16. Panel “Parámetros cuerdas vocales”	37
Figura 17. Dimensiones CV	38
Figura 18. Panel “Parámetros tracto vocal”	38
Figura 19. Panel “Parámetros simulación”	39
Figura 20. Panel “Simulación”	40
Figura 21. Mensaje mostrado para salir del programa.....	40
Figura 22. Mensajes ejemplo de error	40
Figura 23. Estructura inicial del código	41
Figura 24. Estructura final del código	42
Figura 25. Obtención de las x secciones del TV (ej.: $x=22$)	43
Figura 26. Comprobación del correcto almacenamiento de las variables (valores) .	52
Figura 27. Editor de variables para las estructuras obtenidas	53
Figura 28. Simulación H según l_g (hombre).....	55
Figura 29. Zoom del flujo glótico según l_g (hombre).....	55

Figura 30. Señal de voz simulada según l_g media (hombre)	56
Figura 31. Simulación l_g min, altura y grosor máximas ambas capas para N=13 (izquierda) y N=17 (derecha)	57
Figura 32. Flujo glótico según l_g (mujer)	58
Figura 33. Señal de voz simulada según l_g media (mujer)	59
Figura 34. Flujo glótico según l_g (niño) para dimensiones tabla 9	59
Figura 35. Señal de voz simulada según l_g media (niño) para dimensiones tabla 9..	60
Figura 36. Flujo glótico según l_g (niño) para dimensiones tabla10	60
Figura 37. Señal voz simulada según l_g media (niño) para dimensiones tabla 10	61
Figura 38. Flujo glotal obtenido para una mujer promedio	62
Figura 39. Flujo glotal obtenido para un niño promedio	62
Figura 40. Señal de voz y flujo glótico para varias vocales.....	64
Figura 41. Comparación flujo glótico variando el n° de masas de las CV	65
Figura 42. Comparación flujo glótico variando altura y grosor de las CV	66
Figura 43. Comparación flujo glótico variando el n° de cilindros del TV	67
Figura 44. Comparación flujo glótico variando la presión subglótica.....	68

Índice de Tablas

Tabla 1. Propiedades físicas y geométricas de las CV del modelo	25
Tabla 2. Propiedades físicas y geométricas de las CV del modelo más actualizado .	25
Tabla 3. Descripción de los componentes de la GUI	32
Tabla 4. Parámetros relacionados con las cuerdas vocales	37
Tabla 5. Parámetros relacionados con el tracto vocal	38
Tabla 6. Parámetros relacionados con la simulación	39
Tabla 7. N° de masas según locutor.....	44
Tabla 8. Longitud total de la CV según locutor	44
Tabla 9. Altura y grosor de la cubierta y el cuerpo según locutor (1).....	45
Tabla 10. Altura y grosor de la cubierta y el cuerpo según locutor (2).....	46
Tabla 11. Rango de la altura y grosor de la cubierta y el cuerpo según locutor	46
Tabla 12. N° de cilindros según locutor	46
Tabla 13. Longitud total del tracto vocal para las distintas vocales según locutor	47
Tabla 14. Longitud de cada cilindro del tracto para las distintas vocales según locutor.....	47
Tabla 15. Presión subglótica según locutor.....	47
Tabla 16. Resultado del almacenamiento en eCuerdas	53
Tabla 17. Resultado del almacenamiento en eSimul.....	54
Tabla 18. Resultado del almacenamiento en eTracto.....	54
Tabla 19. Tiempo transcurrido de la simulación en función del n° cilindros TV	67

Lista de Acrónimos

Acrónimo	Descripción
TV	Tracto Vocal
CV	Cuerda(s) Vocal(es)
TN	Tracto Nasal
LPC	Lineal Predictive Coding
GUI	Graphical User Interface
GUIDE	Graphical User Interface Development Environment

1. INTRODUCCIÓN

En este primer capítulo se realiza una breve descripción del proyecto y su estructura, se justifica su necesidad y se citan los diferentes objetivos que se buscan con su realización.

La investigación acerca de la voz (síntesis, reconocimiento de voz...) se ha intensificado considerablemente en las últimas décadas. Los modelos físicos de alta calidad para la síntesis de voz se han visto incrementados coincidiendo con el gran desarrollo computacional.

Un modelado satisfactorio del habla basado en la fisiología de un hablante, en el que se basa este proyecto fin de carrera, aun no está disponible, aunque sí en desarrollo debido a las grandes ventajas que supondría: síntesis de voz individual basada en la fisiología real, aumento de la naturalidad del habla sintética como consecuencia de la utilización de un conjunto de parámetros basados en la física para la transmisión de la voz y obtención de ciertos diagnósticos a través de la comparación de voces grabadas. Para que estos objetivos se puedan alcanzar, la investigación debe centrarse en varias etapas: medidas de los parámetros anatómicos estáticos y dinámicos, correspondencia de tales parámetros con modelos matemáticos, desarrollo e implementación de estos modelos y desarrollo de métodos de medida para validar los resultados.

Se dispone de un simulador de voz en Matlab, denominado Vox, creado y descrito originalmente por Malte Kob en [1] en el que se permite la generación, única y exclusivamente, de vocales para unos valores anatómicos fijos distribuidos por el código. De ahí, surge la necesidad y conveniencia de poder modificarlos fácilmente, a través de una *interfaz de usuario* básica, para simular las diferentes vocales en función de los parámetros deseados por el usuario. Modificación tanto de parámetros anatómicos de las cuerdas vocales y el tracto vocal en función del locutor como otros parámetros propios de la simulación (duración, frecuencia de salida y presión subglótica).

Por lo que el proyecto consistirá, por un lado, en completar el simulador haciendo posible también la síntesis a partir de los parámetros de hombres/mujeres/niños; y por otro, en el diseño e implementación de una interfaz gráfica que permita seleccionar los diferentes parámetros de la simulación.

❖ Objetivos

- Completar el simulador existente con modelos de cuerdas vocales (CV) y tracto vocal (TV) adicionales para hombres, mujeres y niños con medidas normales.
- Establecer los márgenes normales de los parámetros anatómicos y fisiológicos modelados en Vox (cuerdas vocales y tracto vocal).
- Desarrollar una interfaz gráfica que permita fijar valores adecuados para la simulación y recoger resultados.

- Examinar las posibilidades de mejoras o ampliación de este proyecto así como indicar posibles direcciones para futuras investigaciones.

❖ Metodología de trabajo

Para lograr los objetivos propuestos se sigue una serie de pasos detallados a continuación:

- Estudio del simulador existente desarrollado en el entorno de programación de Matlab.
- Recopilación de información necesaria.
- Ampliación y mejora del código existente.
- Implementación de la interfaz gráfica de usuario para el simulador Vox.

❖ Estructura principal de la memoria

El trabajo recogido en este proyecto se organiza de la siguiente manera:

En el capítulo 2 se presenta la producción del habla humana, estudiando la anatomía y fisiología de los principales órganos involucrados en la misma; el simulador Vox y el modulo GUIDE de Matlab. El capítulo 3 detalla la creación de la interfaz gráfica, las modificaciones realizadas en el código del simulador y el estudio de las diferentes parametrizaciones posibles según el sexo y la edad del locutor. El siguiente, capítulo 4, muestra diversas pruebas y los resultados obtenidos de las mismas. Por último, en el capítulo 5 se listan las conclusiones finales acontecidas tras el desarrollo del trabajo y las posibles líneas futuras a seguir.

2. MARCO TECNOLÓGICO

En este capítulo se presenta una descripción básica tanto de la anatomía, apartado 2.1.1, como de la fisiología de fonación, apartado 2.1.2, para producción de la voz humana (2.1). Una vez comprendidas estas bases, se presenta en el punto 2.2 el simulador VOX creado por Malte Kob quien define un modelado tanto para las cuerdas vocales (2.2.1) como para del tracto vocal (2.2.2) siendo la unión entre ambos modelos la descrita en el apartado 2.2.3. Por último, se realiza una descripción de la herramienta GUIDE de MATLAB (2.3) la cual nos permitirá crear de modo interactivo una interfaz de usuario de la forma detallada en el en punto 2.3.1.

2.1. Producción de la voz humana

2.1.1. Anatomía de la producción vocal

Los sonidos que intervienen en la comunicación humana son producidos en el aparato fonador.

❖ Aparato fonador

Son varios los sistemas anatómicos, representados y diferenciados en la figura 1, implicados en la producción de la voz humana [3]:

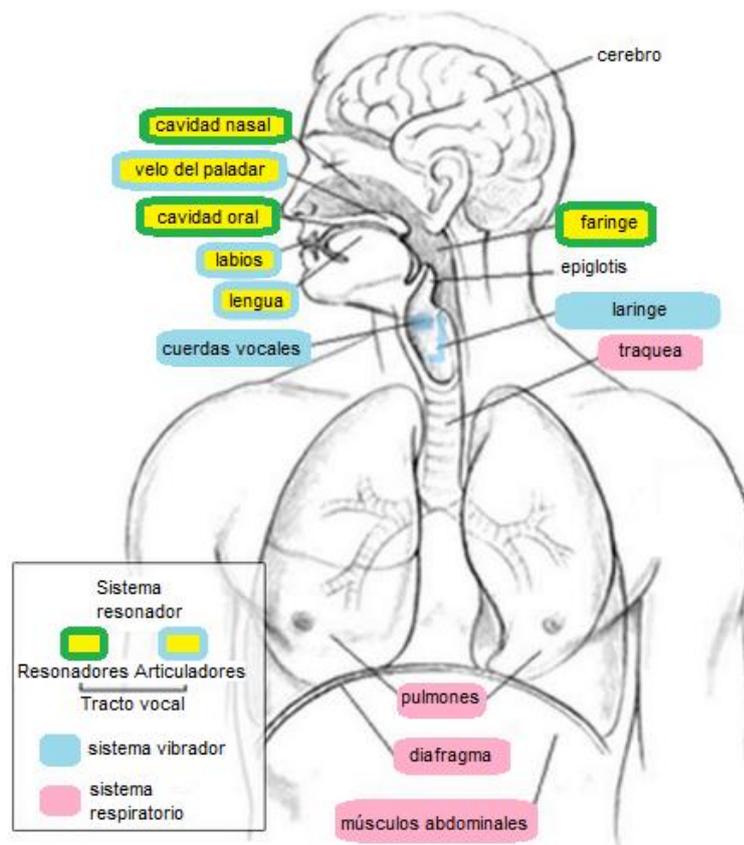


Figura 1. Aparato fonador humano

- El **sistema respiratorio**, formado por las estructuras infragloticas, generador de flujo para la excitación de las cuerdas vocales.

Constituido por la tráquea, los pulmones, el diafragma y los músculos del abdomen. Los pulmones, ubicados en la caja torácica, expelen aire y lo impulsan a través de la tráquea (órgano de carácter cartilaginoso y membranoso que va desde la laringe a los bronquios) para que alcance a las cuerdas vocales, a las que hace vibrar.

- El **sistema vibrador**, ubicado en la laringe, encargado de producir variaciones en la presión de aire.

Las cuerdas vocales, pliegues vocales, constituyen el componente principal de la laringe. Éstas están conformadas por dos músculos recubiertos por una mucosa, y se encuentran sujetas a unos cartílagos que sirven de soporte. Es el flujo de aire pulmonar a su paso por la glotis, espacio comprendido entre las cuerdas vocales, el que determina la vibración de las cuerdas vocales.

- El **sistema resonador**, integrado por las cavidades supraglóticas, amplifica la señal de voz al pasar por el tracto vocal y nasal.

Hace función de resonador, amplificador y modulador del sonido. Este sistema está constituido por el tracto vocal, cavidad comprendida entre la glotis y los labios, que a su vez está compuesto por la faringe, cavidad oral y nasal y el paladar. La faringe es un órgano muscular y membranoso que conecta la nariz y la boca con la laringe y el esófago respectivamente, formando parte por tanto del aparato respiratorio como del digestivo. Se continúa con la cavidad oral que abre un espacio previo a la faringe. La cavidad nasal, separada de la oral a través del paladar duro y blanco, se localiza por encima de la oral.

❖ Estructura de la laringe y de las cuerdas vocales

La laringe, órgano principal para la producción de la voz, se encuentra situado entre los órganos efectores (pulmones y tráquea) y las cavidades resonadoras (faringe, cavidad oral y cavidad nasal). La estructura de la laringe está formada por un esqueleto compuesto por una serie de cartílagos unidos entre sí por articulaciones, ligamentos, membranas y músculos.

Las cuerdas vocales se encuentran en la parte más estrecha de la vía aérea. La estructura histológica de la CV, figura 2, está constituida por cinco capas [4].

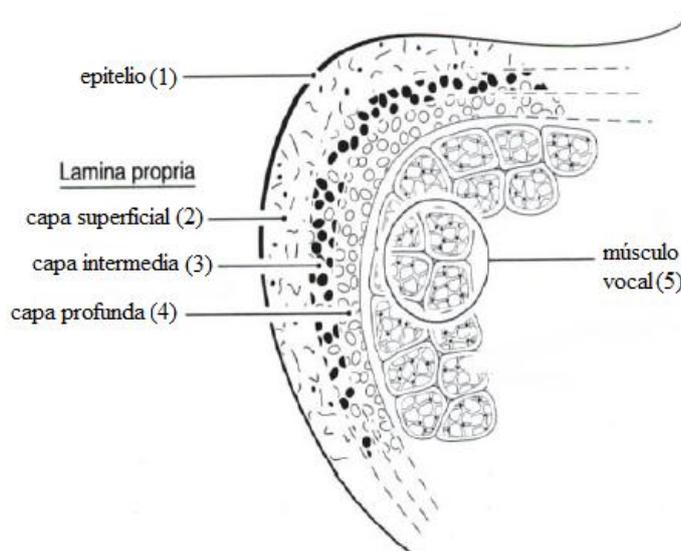


Figura 2. Estructura histológica CV [6]

El epitelio protege a los pliegues vocales del continuo rozamiento durante la fonación, es de tipo escamoso. La lámina propia está compuesta por: una capa superficial (espacio de Reinke) formada por un tejido conjuntivo laxo que permite a la mucosa realizar sus deslizamientos durante la fonación, una capa intermedia formada por fibras elásticas y una profunda formada por fibras de colágeno. El músculo vocal es la porción más profunda del pliegue vocal [5].

Existen diferentes esquemas de estructuración de las cuerdas vocales en capas, figura 3, recogidos en [6]. En la estructura de tres capas, la *mucosa* está compuesta por el epitelio y la capa superficial de la lámina propia, el *ligamento* por la capa intermedia y profunda de la lámina propia, y el *músculo* compuesto por el músculo tiroaritenoides o vocal. En la estructura de dos capas, el *cuerpo* es equivalente a la capa profunda de la lámina propia y el músculo, y el término *cubierta* describe la combinación del epitelio, la capa superficial e intermedia de la lámina propia.

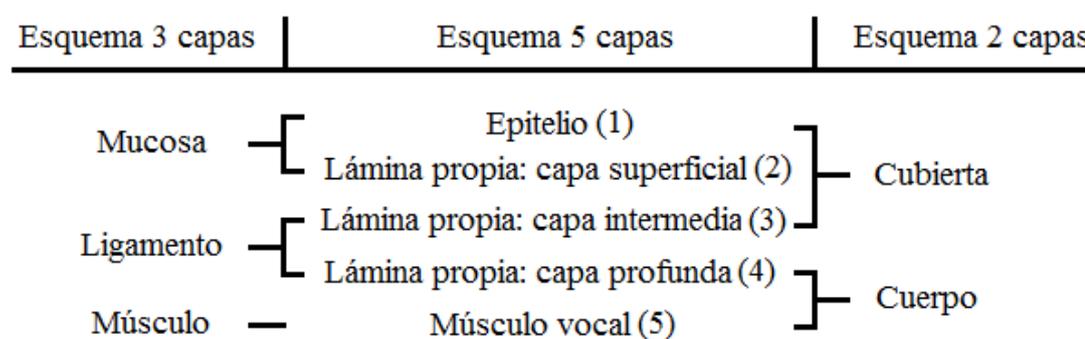


Figura 3. Diferentes esquemas usados para la estructura en capas de las CV [6].

Valores normativos de las dimensiones de las cuerdas vocales

La longitud total de cada cuerda vocal, dimensión anteroposterior de la glotis, de un humano (sin especificar sexo) varía entre 9 y 25 mm [7], resultado obtenido tras el estudio morfológico de cincuenta laringes humanas; 10 y 24 mm para [8]; 10 y 15 mm [9] e incluso entre 10 y 16 mm para Titze en [10].

En la infancia se produce la mutación de la voz debido al crecimiento de la laringe. Las cuerdas vocales de un niño crecen de 4 a 11 mm, o un 60% en longitud, mientras que las de una niña crecen de 1.5 a 4 mm, o un 34% en longitud [11].

En adultos la longitud de la cuerda vocal para la mujer oscila entre los 13 y los 17 mm mientras que para el hombre se encuentra entre los 17 y los 24 mm [12]; en [13] se muestran medidas bastante similares siendo de entre 12.5 y 17.5 mm para mujer y de 17.5 a 25 mm para hombre; mientras que para Hirano en [14] la longitud promedio en mujeres es de 11 a 15 mm y de 17 a 21 mm en hombres.

La información acerca de las dimensiones de las cuerdas vocales resulta escasa, tanto que los científicos desarrolladores de modelos físicos y matemáticos de las cuerdas vocales se ven, en ocasiones, forzados a crear sus modelos en base a informaciones incompletas y de manera un tanto arbitraria, tal y como se indica en [15].

Obtener valores normativos tanto de la altura como del grosor de la cuerda vocal resulta complicado ya que, además de la escasez de información cuantitativa, la interpretación de la misma supone una dificultad añadida. Cada autor utiliza su propia terminología y agrupa las capas de la cuerda vocal según la necesidad de las medidas a realizar para su estudio.

En [6] y [16] se especifica el grosor de las cinco capas que conforman la cuerda vocal, en cambio no se especifica la altura de cada una. Siendo para [6] los grosores del epitelio de entre 0.05 y 1 mm, la lámina superficial de 0.5 mm, la lámina intermedia y profunda de entre 1 y 2 mm y el músculo 7-8 mm. En cambio en [16] el epitelio se define de 0.05 mm, la lámina superficial e intermedia de 0.3 mm cada una, la lámina profunda de 0.5 mm y el músculo de 1.15 mm. Comparando las medidas de ambas publicaciones, no queda claro el valor del músculo aritenoides ya que ambos valores son muy dispares.

En diferentes tesis y estudios sobre el modelo de cuerda vocal para la síntesis de voz se dan valores exactos tanto del grosor como de la altura de la cuerda vocal, diferenciando entre el cuerpo y la cubierta. En [17] y [1] la altura de la cubierta es de 0.5 mm y la del cuerpo 2.5 mm mientras que el grosor de la cubierta es de 1 mm y el del cuerpo es de 2.5 mm. En cambio en [2] las dimensiones dadas son de 1 y 2.7 mm para altura de la cubierta y el cuerpo y de 1.7 mm y 2.2 mm respectivamente para el grosor. Por otro lado, se establece que la altura total (cubierta más cuerpo) se encuentra entre los 3 mm [18] y los 4.5 mm [19]. Respecto al grosor no se ha encontrado por la literatura ningún margen establecido a priori.

La madurez histológica de las cuerdas vocales se produce a partir de los 11 años, previo a la misma, la cuerda vocal aun no está compuesta por las cinco capas conocidas [20]. Existe información nula acerca del ancho y largo de las mismas para niños, únicamente G^a Tapia en [21] incide en que el grosor de la mucosa respecto de la longitud total de la cuerda vocal es de 5/10 en niños y de 1/10 en adultos. Para hombres y mujeres adultos no se encuentran diferencias estadísticamente significativas en anchura y grosor de la cuerda vocal según el estudio descrito en [22].

❖ Estructura del tracto vocal

El tracto vocal es la cavidad que se extiende desde el estrecho pasaje de la laringe a los labios. Durante la fonación, la geometría del TV cambia intencionadamente, como consecuencia de la contracción o relajación de los músculos, para producir el sonido deseado. El músculo más importante para los cambios geométricos del TV es la lengua. Ésta permite variaciones rápidas del diámetro del tracto vocal en un amplio rango. Otros órganos competentes son la mandíbula, velo, dientes y labios.

Valores normativos de las dimensiones del tracto vocal

El tracto vocal tiene una longitud variable de 17 a 20 cm para el hombre, 14.5 a 17 cm para la mujer y de 7 a 10 cm para niños [23]. Su sección transversal puede variar de 0 a 20 cm² debido al movimiento de los órganos articuladores del tracto (mandíbula inferior, lengua, labios y velo del paladar) [3].

2.1.2. Fisiología de la fonación

La producción de un sonido mediante el aparato fonador humano está originada por el flujo de aire pulmonar espirado que se ve sometido a compresiones y rarefacciones por acción de la vibración de las cuerdas vocales. Los resonadores tanto orales como nasales juegan su papel fundamental proporcionando volumen y timbre a la voz.

El conocimiento de los fenómenos que acontecen en fisiología laríngea se debe a varios autores pues de sus investigaciones derivan las principales teorías que nos han permitido conocer y explicar la producción vocal. El primer autor que aporta unos fundamentos sólidos sobre la fisiología vocal, en el siglo XIX, es Johannes Peter Müller cuyas investigaciones dieron lugar a la emisión de la *Teoría Mioelástica* de la fonación, actualmente conocida como *Teoría Fuente-Filtro*. Según esta teoría existe una fuente generadora de sonido o fuente de excitación glótica, que al atravesar las estructuras supraglóticas actúan de resonadores antes de alcanzar el aire atmosférico y hacerse audible, confiriéndole nuevos matices al sonido original, al actuar de filtro.

Esta teoría fue completada y difundida, años más tarde, por J. Van den Berg [24] siendo la teoría más aceptada de la vibración de las cuerdas vocales: *Teoría Mioelástica-Aerodinámica*. Divide el proceso de producción vocal en tres acontecimientos:

- Fase 1. Durante la inspiración las cuerdas vocales son abducidas hacia la posición intermedia o lateral. Cuando comienza la espiración, los músculos extrínsecos aductores hacen que las cuerdas vocales se aproximen entre sí. Las cuerdas vocales contactan en la línea media, lo que combinado con el inicio de la espiración genera un aumento rapidísimo de la presión subglótica.
- Fase 2. La presión subglótica del aire que viene de los pulmones produce una resistencia en las cuerdas vocales cerradas. Eventualmente esta presión se hace mayor que la presión que las mantiene cerradas, por lo que se abren de forma momentánea para dejar salir el aire, liberándose parte de la presión.
- Fase 3. Al fluir el aire rápidamente por la laringe, decrece la presión subglótica y las cuerdas vocales vuelven a juntarse hasta la oclusión completa. El descenso de la presión subglótica provoca, a la altura de las cuerdas vocales abducidas un movimiento de aspiración hacia la línea media debido al principio de Bernoulli. Cada vez que las cuerdas vocales se abren producen una columna de aire que provoca cambios de presión y hace vibrar la mucosa vocal.

Otra contribución fundamental es la de M. Hirano [14], que apoyando la teoría mioelástica-aerodinámica, considera la cuerda vocal como una estructura formada por dos masas de propiedades mecánicas diferentes, el cuerpo (estructura rígida) y la cubierta (estructura elástica) dando así lugar a la *Teoría cuerpo-cubierta*.

Todo el proceso anterior descrito se repite de forma cíclica dando lugar al **ciclo vibratorio** de las cuerdas vocales o ciclo fonatorio (figura 4) formado por fase abierta y una fase cerrada.

La fase abierta, a su vez, se divide en:

- Etapa de apertura: la presión subglótica separa las cuerdas vocales hasta un punto limitado por su grado de elasticidad (I).

- Etapa de cierre: las cuerdas se aproximan hasta su total contacto (VII).

Durante la fase cerrada (VII a IX), la presión subglótica aumenta progresivamente a su vez que se reduce la presión intercordal y el grado de oclusión. La separación de las CV comienza por su borde inferior, avanzando hacia arriba, hasta abrirse lo suficiente como para producir la salida de aire (II), comenzando así un nuevo ciclo.

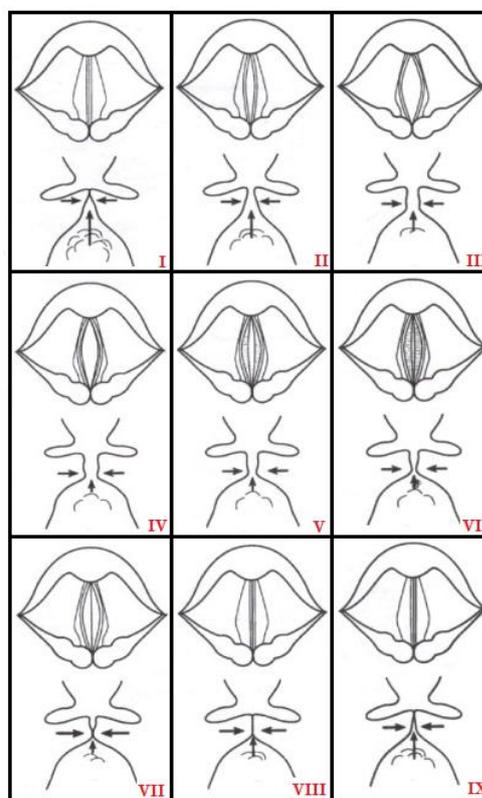


Figura 4. Esquema fases ciclo glótico.

Por lo tanto y en resumen, los cambios de presión provocados al paso del aire por los pliegues vocales producen el sonido laríngeo.

❖ Acción del tracto vocal

El tracto vocal realiza un proceso de filtrado natural. Actúa como un filtro paso banda concentrando la energía alrededor de las frecuencias de resonancia (frecuencia natural de vibración de un objeto). Estas concentraciones de energía sonora que se dan alrededor de unas frecuencias determinadas se denominan formantes. Los formantes permiten distinguir los sonidos del habla humana, sobre todo vocales y otros sonidos sonoros. Normalmente sólo los dos primeros formantes (F1 y F2), determinados por la posición de la lengua, son necesarios para caracterizar una vocal, sin embargo, en ciertos casos pueden ser necesarios más. Los formantes posteriores determinan otras propiedades acústicas como el timbre.

Debido a las alteraciones que el sonido sufre por su paso por el tracto vocal, al favorecer ciertas bandas de frecuencia en detrimento de otras como consecuencia de sus propiedades de filtrado en un momento dado, es evidente concluir que el sonido laríngeo es diferente al emitido por los labios.

2.2. Simulador Vox

Previo a la descripción particular del simulador Vox, se realiza una pequeña introducción acerca de la síntesis de voz y una exposición sobre los diferentes tipos de sintetizadores.

Se entiende como síntesis de voz a la producción artificial del habla humana. Existen varias técnicas que permiten realizar el proceso de sintetización de la voz, siendo las más destacadas [25]:

- Síntesis concatenativa

Se basa en el encadenamiento de los segmentos de la voz pregrabada, es decir, de las unidades extraídas del habla humana. Para ello se necesita disponer de una amplia base de datos.

- Síntesis por formantes

Los valores de las resonancias del tracto vocal (formantes), la fuente de excitación periódica (con tono) o aperiódica (ruido) y el nivel energético son parametrizados. Estos parámetros se actualizan en tiempo real para una generación de voz continua. Se puede trabajar de acuerdo a la forma de instalación de los filtros (serie o paralelo) que dan lugar a cada una de las resonancias.

- Síntesis de predicción lineal

Esta técnica parte de un tratamiento temporal de la señal acústica con ciertos parámetros que permiten ahorrar la redundancia de información que se da en segmentos próximos de la voz. Aparte de constituir una buena herramienta para la parametrización de la señal de voz, al mismo tiempo, por un proceso inverso, posibilita la regeneración de la señal acústica previamente parametrizada por un algoritmo LPC.

- Síntesis articulatoria

Se realiza una analogía entre parámetros relativos a los órganos articulatorios y sus movimientos con parámetros circuitales. Son modelos físicos basados en los mecanismos del aparato fonador.

La síntesis articulatoria resulta menos popular que otras técnicas debido a su compleja y costosa implementación; a la obtención dificultosa de los parámetros involucrados y al tiempo excesivo de cálculo. En cambio, al tratarse de una síntesis individual basada en la fisiología real se podría conseguir un aumento de la naturalidad del habla sintética y la obtención de diferentes diagnósticos a través de la comparación de voces grabadas.

❖ Vox

El simulador Vox está basado en una tesis doctoral sobre el estudio de la voz cantada [1], que define un modelado físico de las partes del órgano de voz relevantes para la producción de voz. Este trabajo se centraba concretamente en la síntesis de vocales en vez de la síntesis de voz.

Se realiza un modelado tanto del movimiento de las cuerdas vocales como del tracto vocal (síntesis articulatoria), quedando totalmente excluido el modelado del tracto nasal (TN) lo cual hace que el simulador programado se aleje, en parte, de la naturalidad de voz humana.

El simulador, programado en su totalidad en lenguaje de Matlab, trabaja en tiempo continuo. Su funcionamiento se puede organizar de acuerdo a la figura 5:

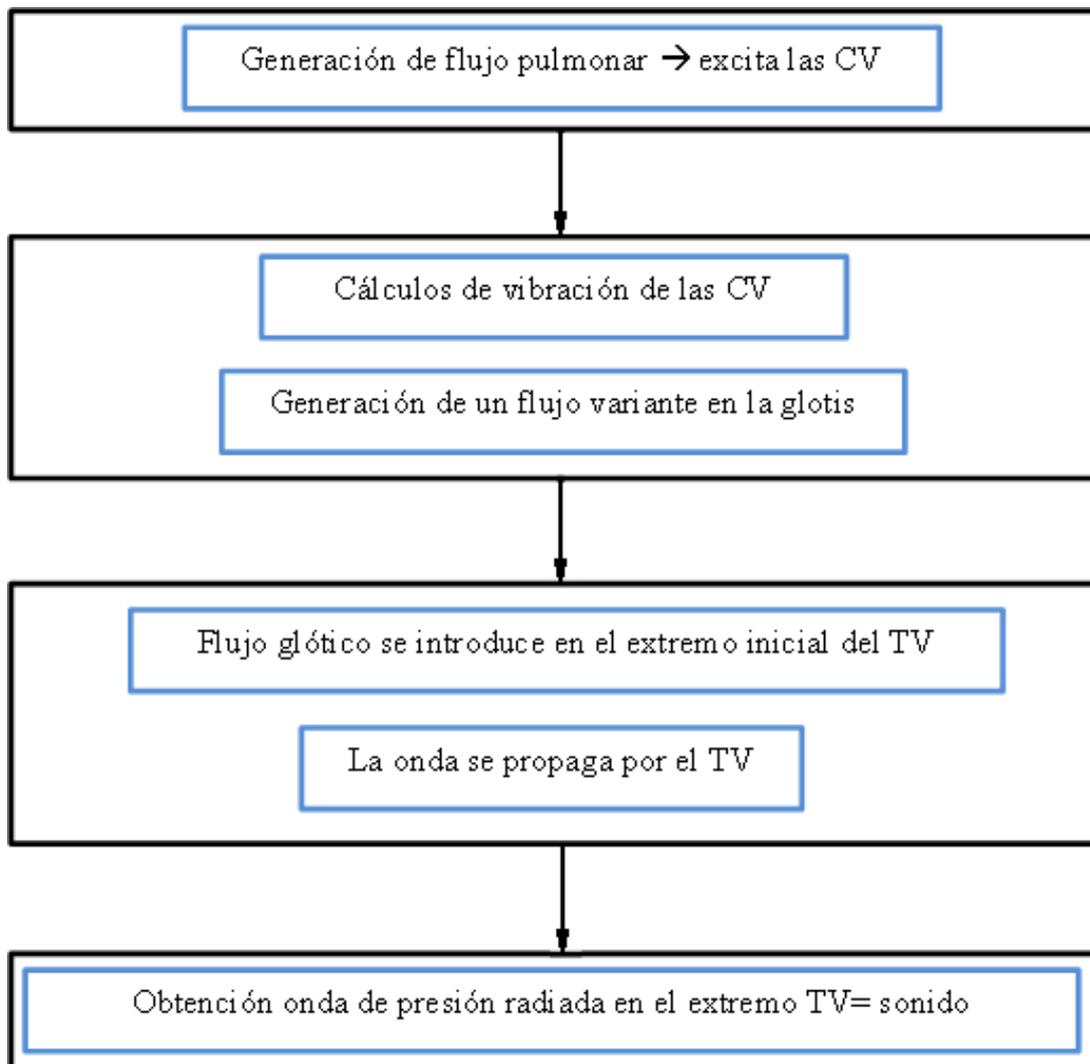


Figura 5. Funcionamiento simulador Vox.

2.2.1. Modelo de las cuerdas vocales (CV)

El modelo utilizado por M. Kob, figura 6, está basado en el modelo de 16 masas desarrollado por I.R. Titze [17] pero incluye algunas de las modificaciones publicadas más recientemente, como es la de B. Story e I.R. Titze [16].

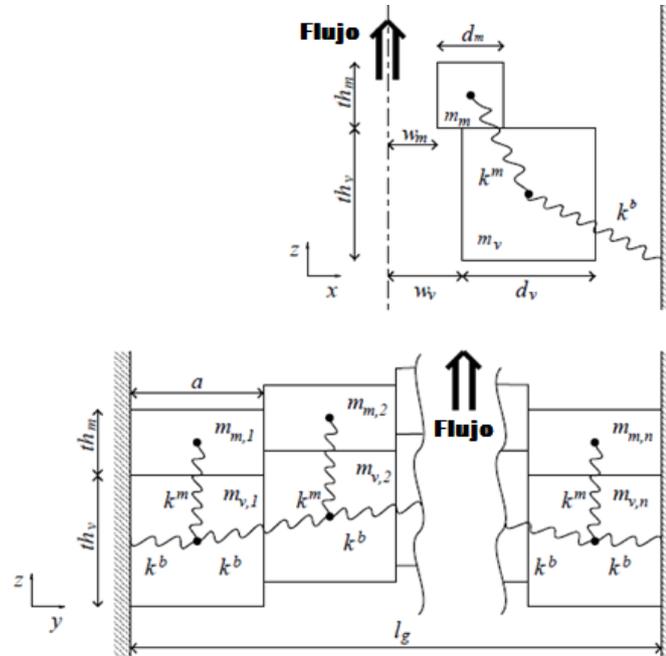


Figura 6. Vista en sección (arriba) y lateral (abajo) del modelo de Kob [1].

Cada cuerda vocal está dividida en n segmentos iguales de forma paralela. A su vez, cada segmento está dividido en dos partes independientes: una que representa el cuerpo de las CV o *musculus vocalis* (subíndice v) y otra que representa la cubierta de las CV o *mucosa* (subíndice m). Cada una de las masas del cuerpo m_v está conectada a la frontera por un resorte con rigidez k^b y amortiguamiento D^b , y las pequeñas masas m_m por un resorte de rigidez k^m y amortiguamiento D^m .

La disposición de las CV simuladas según el modelo está representada en la figura 7 (con n igual a 15).

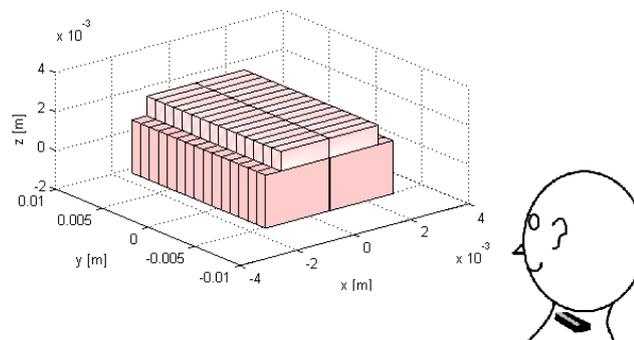


Figura 7. Disposición de las CV del modelo.

▪ Propiedades de las cuerdas vocales

En la tabla 1 se listan los valores físicos y geométricos de los principales parámetros de las cuerdas vocales del modelo obtenidos de [26] y [27].

Tabla 1. Propiedades físicas y geométricas de las CV del modelo.

Capa	Longitud [mm]	Grosor [mm]	Altura [mm]	Densidad [mg/mm ³]	Masa [mg]
Cuerpo	$l_g=14$	$d_v=2.5$	$th_v=2.5$	$\rho_v=1.04$	$m_{voc.,total}=91$
Ligamento vocal	$l_g=14$	$d_l=1.0$	$th_l=1.0$	$\rho_l=1.04$	$m_{lig.,total}=91$
Cubierta	$l_g=14$	$d_m=1.0$	$th_m=0.5$	$\rho_m=1.02$	$m_{m,total}=7$

Estos valores se van a ver modificados, tabla 2, en la versión más actualizada del simulador Vox, sobre la que se ha trabajado, atendiendo a [2].

Tabla 2. Propiedades físicas y geométricas de las CV del modelo más actualizado.

Capa	Longitud [mm]	Grosor [mm]	Altura [mm]	Densidad [mg/mm ³]	Masa [mg]
Cuerpo	$l_g=14$	$d_v=2.2$	$th_v=2.7$	$\rho_v=1.04$	$m_{voc.,total}=86$
Ligamento vocal*	$l_g=14$	$d_l=2.2$	$th_l=2.2$	$\rho_l=0$	$m_{lig.,total}$
Cubierta	$l_g=14$	$d_m=1.7$	$th_m=1.0$	$\rho_m=1.02$	$m_{m,total}=24$

* En esta versión, el ligamento vocal se supone incrustado dentro del cuerpo de las cuerdas vocales y, por lo tanto, sus dimensiones no afectan a las dimensiones de los elementos de las CV, sino que sólo se consideran para el cálculo de las fuerzas elásticas. Para ser coherente, en este modelo se supone $\rho_l = 0$ mg/mm³, por lo tanto, la masa del ligamento es parte de la masa corporal (86 mg).

▪ Movimiento de las cuerdas vocales

En conjunto, cada masa está sometida a fuerzas elásticas que las unen a las masas contiguas pertenecientes a un mismo tejido, a fuerzas de resorte que actúan como enlaces entre los tejidos (masas superiores e inferiores del modelo) y entre éstos y los cartílagos laríngeos, a fuerzas de contacto, amortiguación, tensión activa y presiones aerodinámicas. Todas estas fuerzas se añaden por separado en las direcciones x y z para cada masa (el movimiento en la dirección y no está permitido en este modelo) para obtener las ecuaciones correspondientes a la segunda ley de Newton. Finalmente, se obtienen un total de $8 \cdot n$ ecuaciones diferenciales que son solucionadas numéricamente a través del método de Runge-Kutta.

2.2.2. Modelo del tracto vocal (TV)

Siguiendo la propuesta de programación de [28], el tracto vocal, figura 8, se ha simulado según el modelo de Kelly-Lochbaum [29].

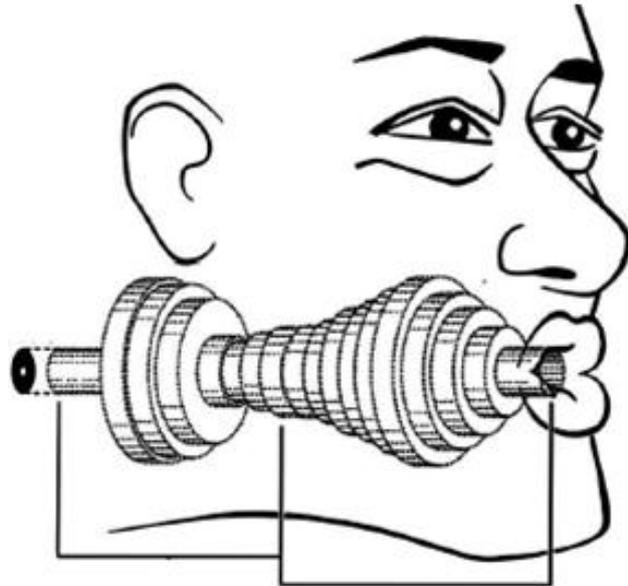


Figura 8. Tracto vocal modelado [7].

El modelo de tracto vocal es una concatenación de M cilindros huecos de misma longitud, pero de diferentes secciones. La sección del último cilindro corresponde con la apertura de los labios (A_M) y la del primer cilindro con la sección del volumen supra-glótico (A_1), siendo A_0 la apertura de la glotis.

Es al final de esta concatenación de cilindros donde la radiación de la onda acústica aparece.

En el modelo programado se ha elegido $M=44$, es decir, 44 cilindros huecos cuyas longitudes y secciones, dependientes de la vocal a simular, se encuentran recogidas en [30].

▪ Frecuencia de muestreo

La frecuencia de muestreo del modelo debe ser el inverso del tiempo que le cuesta a la onda propagarse a lo largo de un cilindro, supuestos todos de la misma longitud:

$$F_s = \frac{c}{l} = \frac{1}{T_s}$$

Siendo c la velocidad del sonido a través del aire y l la longitud de cada cilindro. Para el modelo glotal el periodo de muestreo se ha hecho igual a $2 \cdot T_s$ con el fin de obtener $F_s \approx 16$ kHz [2].

▪ Presiones

Tomando de inicio todas las presiones como nulas y atendiendo a las ecuaciones descritas en [1], se calculan todas las presiones participantes en el proceso.

Como anteriormente se ha anticipado, el flujo glotal se introduce en el extremo pertinente de una línea de concatenación de cilindros que simulan, según el modelo de guía de ondas planas, el tracto vocal. En cada transición, las ondas acústicas se reflejan y se transmiten (figura 9) de acuerdo con el coeficiente de reflexión propuesto por Kelly y Lochbaum.

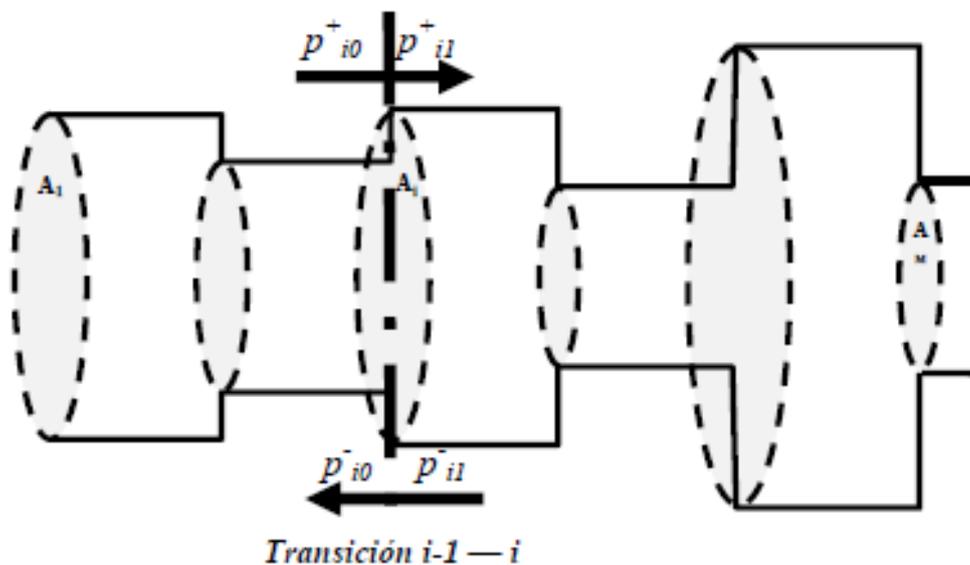


Figura 9. Concatenación de cilindros. Ondas acústicas reflejadas y transmitidas.

Se produce un caso excepcional al llegar a la última transición, ya que se asume que no entra ninguna onda acústica a los labios.

Por otra parte, el coeficiente de reflexión de un cilindro abierto es sin memoria por lo que para la radiación de presión en los labios, se asume el modelo de reflexiones en el dominio del tiempo de [31].

Según este modelo, las reflexiones (y las transmisiones) de deben modelar como una convolución de la onda incidente con la función de reflexión. Esta función es un impulso multiplicado por una constante en el caso de las discontinuidades entre cilindros, pero en el caso de la radiación producida al final de un cilindro abierto la duración de la respuesta no es nula. Para no necesitar un sistema de memoria infinita, se usa una estrategia similar a la de [31], considerando que el período de muestreo T_s es pequeño.

2.2.3. Unión entre ambos modelos

En la figura 10 se puede visualizar la unión entre el modelo de las cuerdas vocales y el modelo del tracto vocal explicados anteriormente.

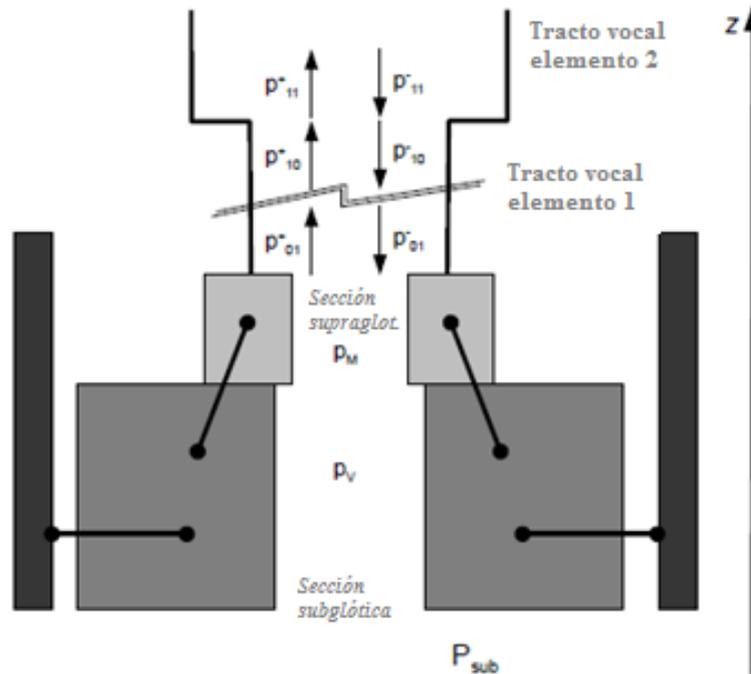


Figura 10. Unión entre el modelo de las cuerdas vocales y el del tracto vocal [18].

Las ondas de presión ascendentes que avanzan en sentido positivo por el modelo simulado de tracto vocal se representan como p^+ mientras que las ondas de presión descendentes como p^- .

En la notación p^{\pm}_{ij} , i se refiere al número de la unión, siendo $i=0$ la transición glotis-tracto, mientras que $j=0$ hace referencia al lado superior de la unión y $j=1$ al lado inferior.

2.3. GUI en MATLAB®

Matrix Laboratory, comúnmente conocido como MATLAB, es un entorno de programación para el desarrollo de algoritmos, el análisis de datos, la visualización y el cálculo numérico.

MATLAB dispone de un modulo denominado GUIDE (Graphical User Interface Development Environment) que permite crear de modo interactivo una interfaz gráfica de usuario (GUI).

Aun cuando sería posible escribir un archivo .m que contenga todos los comandos para elaborar una GUI, es mucho mas fácil utilizar el GUIDE porque permite hacerlo interactivamente.

Las principales ventajas de usar GUIDE descritas en [32] son:

- Composiciones simples: el diseño gráfico es intuitivo para el desarrollo de la GUI y proporciona estructuras perfectamente organizadas para diseños más complejos.
- Generación automática de código: GUIDE generará automáticamente los archivos de extensiones .fig y .m, el primero de los cuales contiene la descripción completa de la GUI y el segundo contiene las funciones que ejecutan y controlan la misma.
- Control directo sobre las propiedades de los objetos: se proporciona una visión completa de las propiedades disponibles permitiendo modificar el comportamiento de los componentes de la GUI de manera eficiente.

2.3.1. Construcción interactiva de una GUI

La elaboración de una GUI incluye dos tareas básicas como son:

- Diseño de la GUI: distribución de los componentes de la interfaz.
- Programación de la GUI: programación de cada componente.

1) Diseño de una GUI

Se dispone de varias opciones diferentes para acceder a la herramienta GUIDE en MATLAB (figura 11), accediendo a través del menú “File”, del botón “Start”, del acceso directo, del acceso directo de “MATLAB Toolbar” o simplemente introduciendo `>>guide` en la ventana de comandos.

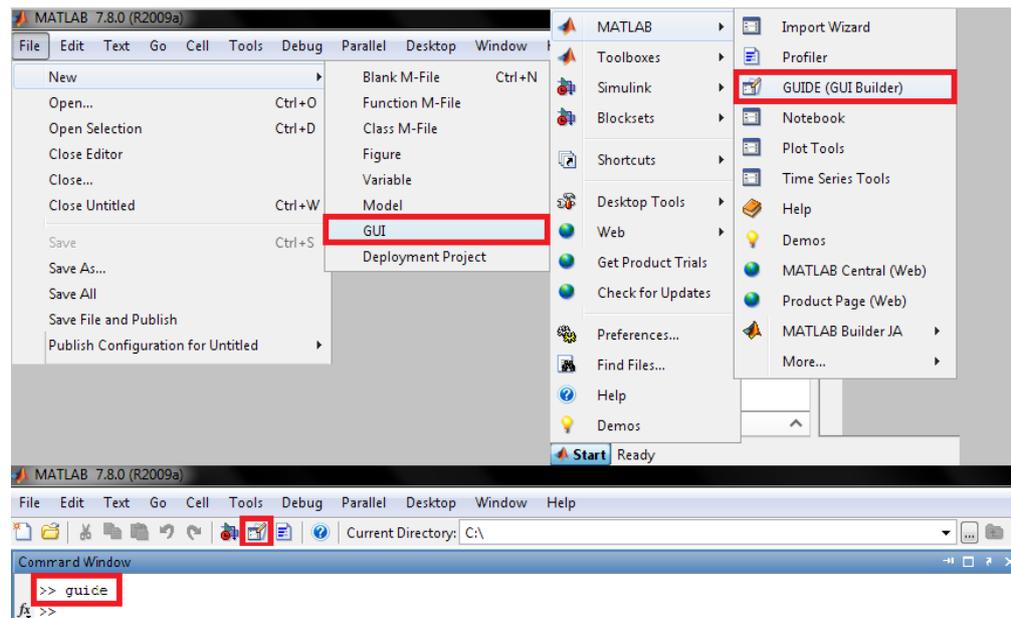


Figura 11. Acceder a GUIDE en MATLAB.

Tras ejecutar cualquiera de las acciones que inician la herramienta aparecerá una ventana de selección denominada *GUIDE Quick Start* (figura 12) que ofrece la posibilidad de escoger entre crear una nueva GUI en blanco o abrir uno existente como ejemplo.

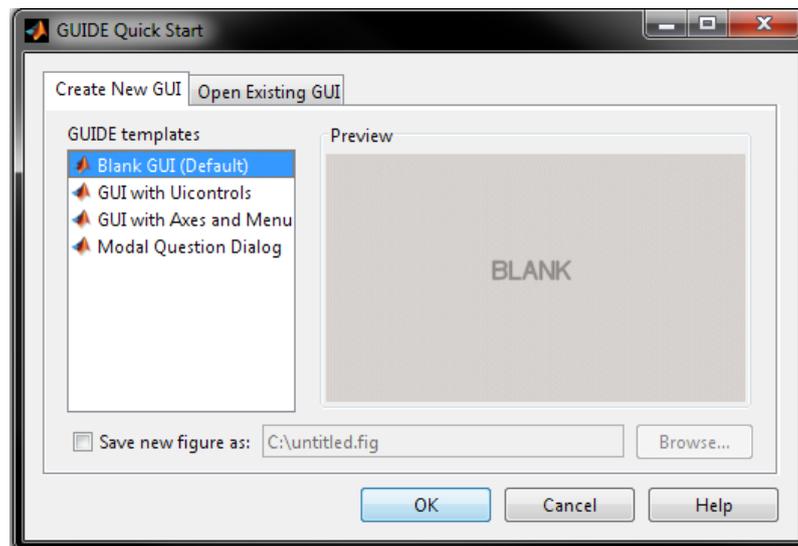


Figura 12. Ventana GUIDE Quick Start.

Para diseñar el aspecto visual de una GUI, completamente desde cero, se selecciona la opción *Blank GUI (Default)*. Acto seguido aparece la pantalla principal de GUIDE (figura 13).

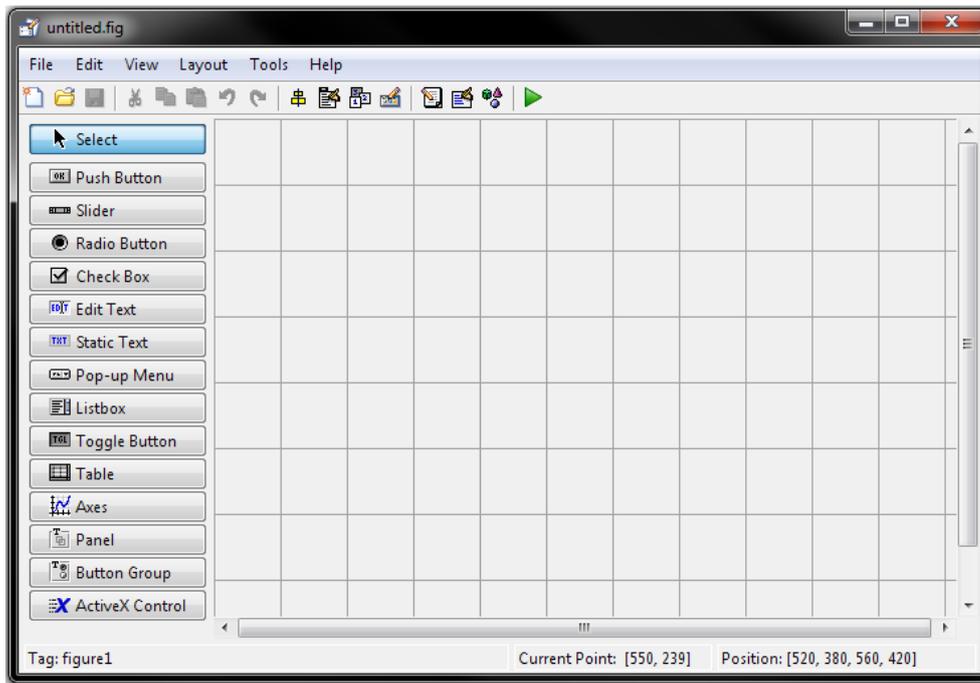


Figura 13. Ventana de diseño en GUIDE.

En esta ventana se pueden distinguir tres secciones fundamentales de gran importancia:

- Barra de herramientas. Compuesta tanto de las funciones básicas (crear nueva figura, abrir, guardar, etc...) y las funciones necesarias para diseñar una GUI, figura 14, detalladas a continuación.

	Alineación de objetos
	Editor de menús
	Editor de orden de tabulación
	Editor de barra de herramienta
	Editor de ficheros M
	Inspector de Propiedades
	Buscador de objetos
	Ejecutar

Figura 14. Barra de herramientas en GUIDE.

- Alineación de objetos: sobre los objetos seleccionados permite realizar diferentes tipos de alineado.
- Editor de menús: permite diseñar una estructura de menús desplegables.

- Editor de orden de tabulación: permite configurar el orden de desplazamiento del foco sobre los objetos de la GUI al pulsar la tecla de tabulación.
- Editor de barra de herramienta: permite crear una barra de herramientas con las funcionalidades deseadas en la interfaz.
- Editor de ficheros M: abre el archivo .m donde se encuentra el código fuente de la interfaz.
- Inspector de propiedades: posibilita la modificación de las propiedades de los objetos creados.
- Buscador de objetos: muestra la estructura jerárquica de los objetos.
- Ejecutar: guarda y ejecuta el código fuente asociado a la GUI.
- Área de diseño. Zona donde se ubican los objetos.
- Paleta de componentes. Muestra todos los objetos disponibles para la realización de la interfaz.

GUIDE dispone de un amplia paleta de componentes para realizar la interfaz gráfica. Para emplazar estos objetos se debe arrastrar desde la paleta hasta el área de diseño. En la tabla 3 se muestran los distintos objetos.

Tabla 3. Descripción de los componentes de la GUI.

Control	Descripción
Push Button	Invoca un evento inmediatamente
Slider	Usado para representar un rango de valores
Radio Button	Indica una opción que puede ser seleccionada
Check Box	Indica el estado de una opción o atributo
Edit Text	Caja para editar texto
Static Text	Muestra una cadena de caracteres en una caja
Pop-up Menu	Provee una lista de opciones
List Box	Muestra una lista deslizable
Toggle Button	Permite elegir entre dos estados: on/off
Table	Inserta una tabla
Axes	Muestra imagines o gráficos
Panel	Agrupar botones como un grupo
Button Group	Permite exclusividad de selección con los Radio Button
ActiveX Component	Aplicaciones gráficas

Las propiedades de estos objetos se gestionan a través de la opción *Property Inspector* a la cual se accede al hacer click con el botón derecho sobre el objeto en sí. Se modelan tanto las propiedades físicas (tamaño, tipo de fuente, posición, etc...) como las propiedades relativas a aspectos de programación (tag, callback, etc...).

2) Programación de una GUI

Una vez concluido el diseño de la interfaz, GUIDE genera automáticamente un archivo .m que, en resumen, contiene:

- Código de inicialización de la GUI.
- Código para implementar tareas previas a la visualización en pantalla de la GUI.
- Código de los *callbacks* (funciones que responden al evento generado por una acción del usuario).

Cada componente tiene sus propios *callbacks* al igual que propiedades. De este modo, el diseñador simplemente debe codificarlas asignando a cada componente la función deseada.

3. DESCRIPCIÓN DE LA SOLUCIÓN PROPUESTA

En este capítulo se presenta una descripción general de la interfaz de usuario implementada (3.1) la cual permite la selección de diferentes parámetros de la simulación además de posibilitar la síntesis de vocales a partir de los parámetros físicos tanto de un hombre como de una mujer o un niño. A continuación, en el apartado 3.2, se describen las principales modificaciones realizadas sobre el código base del simulador Vox. Y por último, en el apartado 3.3 se detallan las diferentes parametrizaciones escogidas en función del locutor seleccionado, tanto de los parámetros de las cuerdas vocales (3.3.1) como los del tracto vocal (3.3.2) y otros parámetros generales de la simulación (3.3.3).

3.1. Implementación de la interfaz

A lo largo de este apartado se detalla brevemente el proceso de diseño e implementación de la interfaz gráfica creada, utilizando la herramienta GUIDE de Matlab y siguiendo los pasos del apartado anterior, para la simulación de vocales. Esta interfaz permite una comunicación rápida e intuitiva entre el usuario y el ordenador.

Cabe destacar que la interfaz implementada no posibilita la modificación de absolutamente todos los parámetros inmersos en el proceso de simulación de vocales, habiéndose escogido sólo aquellos que a priori parecieron más interesantes.

La interfaz consta de una única ventana, figura 15, compuesta a su vez por cuatro bloques o paneles perfectamente diferenciables.

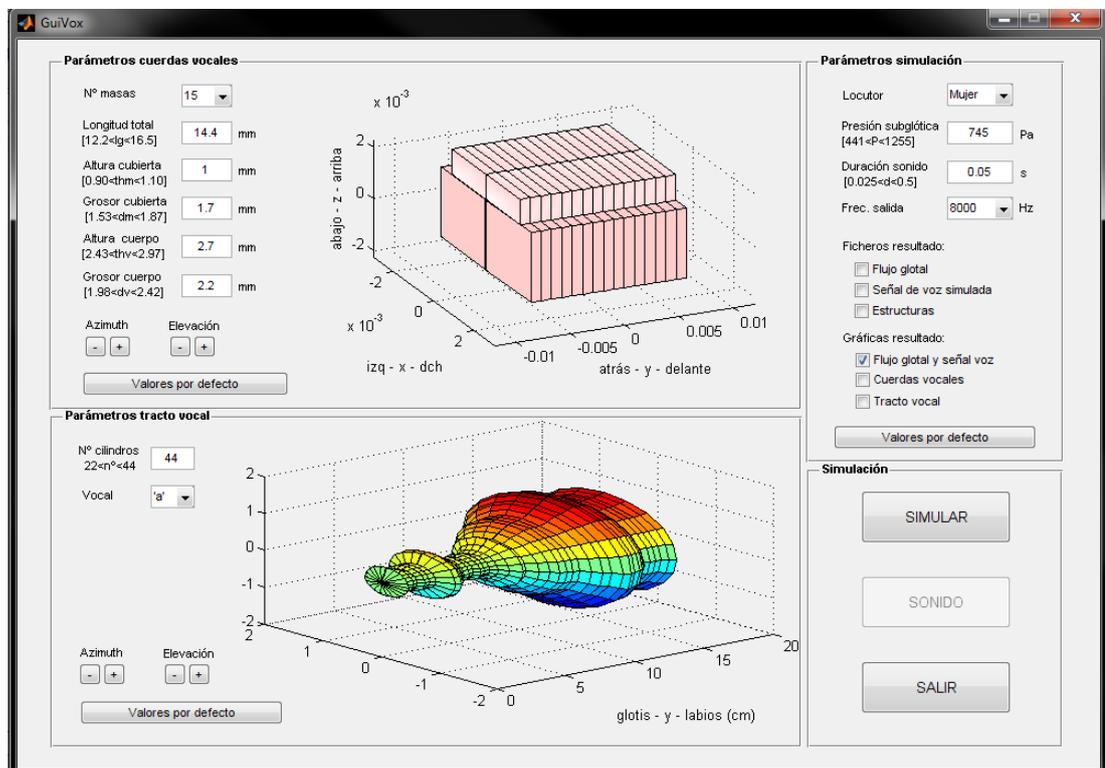


Figura 15. Ventana interfaz de usuario inicio.

Siendo los paneles mencionados:

- Parámetros cuerdas vocales.
- Parámetros tracto vocal.
- Parámetros simulación.
- Simulación.

- Parámetros cuerdas vocales.

Tal y como se puede observar en la figura 16 este bloque de la interfaz incluye, en su parte izquierda, las diferentes opciones disponibles para introducir los parámetros relacionados directamente con la forma y el tamaño de las cuerdas vocales modeladas y, en su parte derecha, una representación gráfica de las mismas atendiendo a estos parámetros.

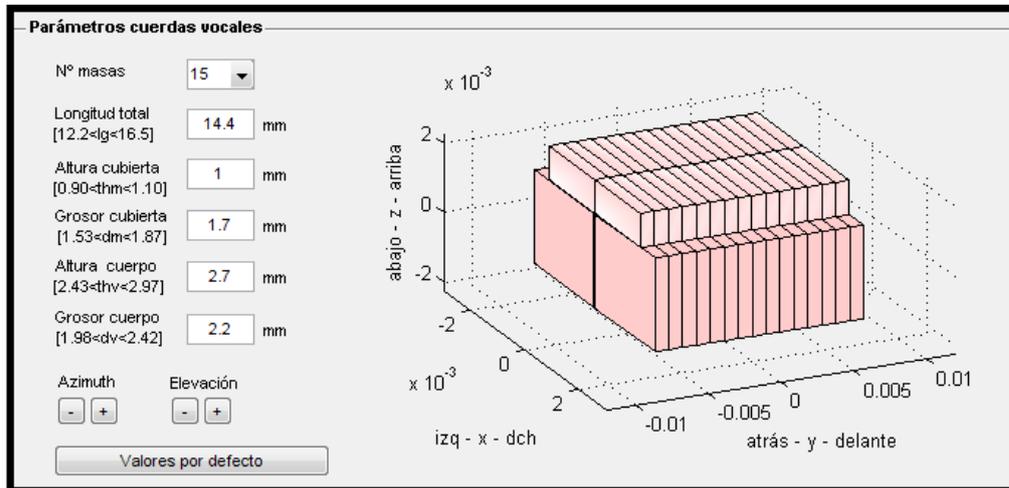


Figura 16. Panel “Parámetros cuerdas vocales”.

En la tabla 4 se enumeran y describen todos y cada uno de los parámetros modificables en este panel.

Tabla 4. Parámetros relacionados con las cuerdas vocales.

Parámetro	Descripción
Nº masas	Número total de elementos que conforman una cuerda vocal. Valores a elegir de 13 a 17.
Longitud total CV (l_g)	Dimensión antero-posterior de la glotis.
Altura cubierta (th_m)	Dimensión vertical de la cubierta (<i>mucosa</i>).
Grosor cubierta (d_m)	Profundidad lateral de la cubierta.
Altura cuerpo (th_v)	Dimensión vertical del cuerpo (<i>vocalis</i>).
Grosor cuerpo (d_v)	Profundidad lateral del cuerpo.
Azimuth	Giro respecto al eje Z de la representación gráfica
Elevación	Giro sobre el plano XY de la representación gráfica
Valores por defecto	Botón que recarga los valores predeterminados de cada uno de los parámetros descritos.

En la figura 17, se especifican varios de estos parámetros modificables por parte del usuario.

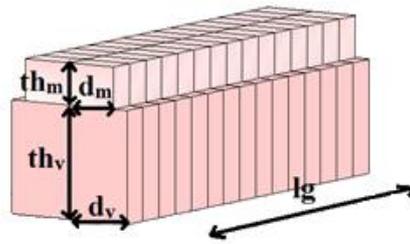


Figura 17. Dimensiones CV.

- Parámetros tracto vocal

El segundo bloque de la interfaz, figura 18, ofrece la posibilidad de modificar varios parámetros que afectan directamente a la forma y tamaño del tracto vocal modelado. Al igual que en el bloque precedente, en la parte izquierda se encuentran los parámetros modificables y en la derecha una representación gráfica aproximada del tracto vocal en cuanto a estos parámetros.

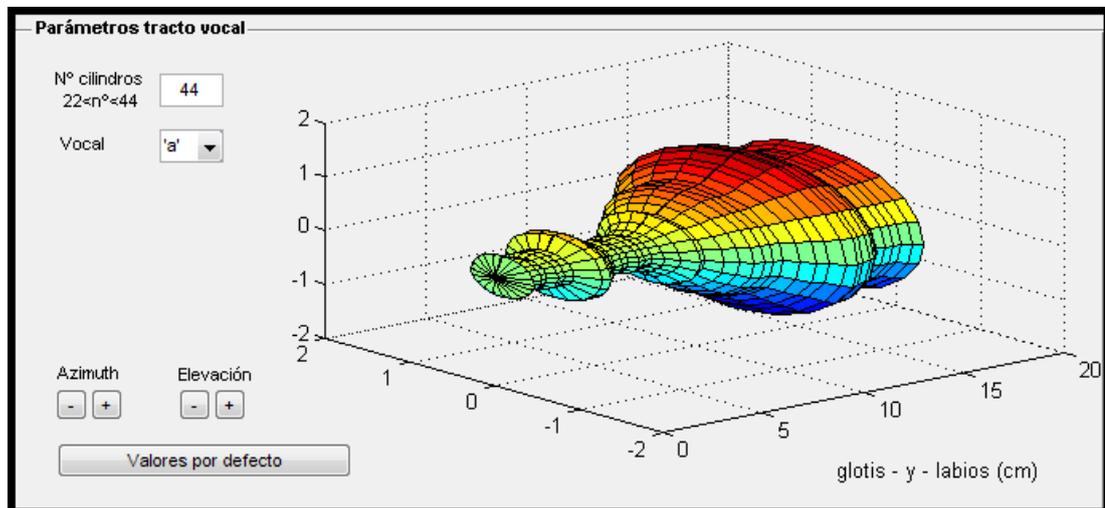


Figura 18. Panel "Parámetros tracto vocal".

En la tabla 5 se distinguen y describen los diferentes parámetros representados en este panel cuyos valores predeterminados se muestran en la anterior figura.

Tabla 5. Parámetros relacionados con el tracto vocal.

Parámetro	Descripción
Nº cilindros	Número total de cilindros que conformaran el modelo del tracto vocal.
Vocal	Vocal elegida para su simulación. Valores a elegir: /a/, /A/, /ae/, /E/, /i/, /I/, /o/, /O/, /u/, /U/.
Azimuth	Giro respecto al eje Z de la representación gráfica
Elevación	Giro sobre el plano XY de la representación gráfica
Valores por defecto	Botón que recarga los valores predeterminados de cada uno de los parámetros descritos.

Respecto a las representaciones gráficas tanto de las cuerdas vocales como del tracto vocal mostradas en la ventana principal de la interfaz, cabe indicar que las mismas se actualizan *automáticamente* al modificar el valor de cualquier parámetro.

- Parámetros simulación

El siguiente bloque, representado en la figura 19, permite tanto elegir el sexo y edad del locutor como modificar ciertos parámetros generales de la simulación; además de ofrecer al usuario la posibilidad de elegir qué ficheros desea guardar y qué gráficos mostrar con los resultados obtenidos de la simulación.

La selección desplegable del parámetro locutor (hombre, mujer, niño) produce una modificación automática en los parámetros descritos en los dos bloques anteriores, es decir, una variación de los parámetros físicos que describen las cuerdas vocales y el tracto vocal en función de la opción seleccionada. Los nuevos valores de estos parámetros aparecerán en sus campos correspondientes.

Figura 19. Panel “Parámetros simulación”.

En la tabla 6 se describen los diferentes parámetros representados en este panel.

Tabla 6. Parámetros relacionados con la simulación.

Parámetro	Descripción
Locutor	Sexo y/o edad del locutor simulado. Opción a elegir: hombre/mujer/niño/promedio.
Duración	Duración del sonido a simular.
Frec. salida	Frecuencia de muestreo del resultado. Valores a elegir de 8000 a 32000 Hz.
Presión subglótica	Presión subglótica.
Ficheros resultado	Elección de los archivos .mat a guardar tras la finalización de la simulación.
Gráficas resultado	Elección de las figuras a mostrar tras la finalización de la simulación.
Valores por defecto	Botón que se encarga de cargar los valores predeterminados de cada uno de los parámetros descritos.

- Simulación

Este último bloque, mostrado en la figura 20, es el más simple de todos. Se encarga de iniciar la simulación, mostrar el sonido obtenido y salir de la aplicación.

El botón “Simular” es el responsable de iniciar y ejecutar todo el proceso de simulación manteniendo, en cada caso, los valores deseados por el usuario de los parámetros anteriormente descritos.

El botón “Sonido” no aparecerá activo para el usuario hasta que la simulación haya finalizado completamente. Se encarga de lanzar el audio además de realizar una copia del mismo en un archivo .wav.

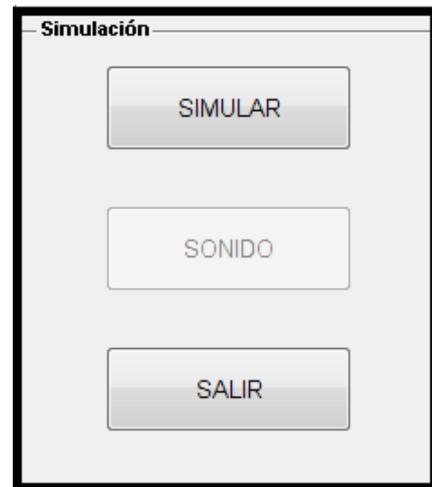


Figura 20. Panel “Simulación”.

Por último, el botón “Salir” muestra un mensaje en pantalla (figura 21) que da al usuario la opción de abandonar el simulador o en cambio seguir trabajando en él.

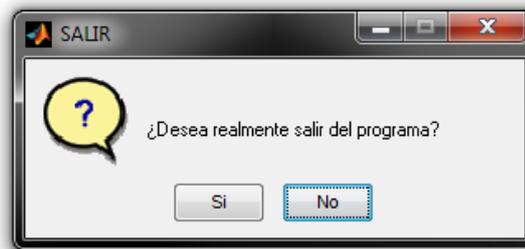


Figura 21. Mensaje mostrado para salir del programa.

- Comprobación de errores

En cada una de las cajas de texto editables que aparecen en la interfaz se evalúa si el valor introducido por el usuario es un valor numérico y si se encuentra entre ciertos valores coherentes, en caso contrario se muestra un mensaje de error para que el valor sea corregido (figura 22).



Figura 22. Mensajes ejemplo de error.

3.2. Modificaciones en Vox

Las principales modificaciones que se han realizado sobre el simulador de partida en Matlab son:

- 1. Limpieza de funciones sin uso.
- 2. Reorganización del código de la función principal (*FullVoxCoupled*) y otras funciones de inicialización (*VF_setup*, *SelectVocalTract*, *VT_InitMC*).
- 3. Creación de nuevas funciones.

Tras una inspección exhaustiva del funcionamiento del simulador de voz Vox, se llega a la conclusión de que, de partida, el número de funciones totales que conforman el mismo son más de las realmente necesarias. Por lo tanto se realiza una criba de las mismas, eliminando así aquellas funciones inservibles para el funcionamiento del simulador.

Tras eliminar dichas funciones, el código del programa queda bastante simplificado y es mostrado, a continuación, en forma de diagrama en la figura 23.

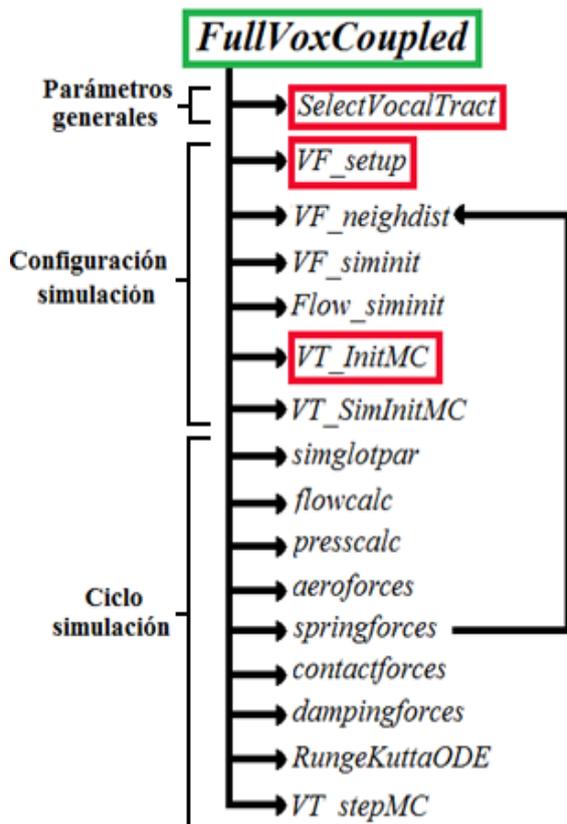


Figura 23. Estructura inicial del código.

Recuadrada en verde se encuentra la función principal, función que simula la producción de voz. Dicha función se ve modificada en cuanto a su nombre que pasa a ser **Vox** y en cuanto a su forma.

Por otro lado, las funciones con recuadro rojo han sido eliminadas, pero no sus contenidos que han sido incluidos también en algunas de las nuevas funciones creadas.

SelecVocalTract proporcionaba la geometría del tracto vocal según la vocal elegida, *VF_setup* creaba e inicializaba la estructura de las cuerdas vocales, mientras que *VT_InitMC* creaba e inicializaba la estructura del tracto vocal.

Sobre la figura 24 podemos comprobar cómo queda la estructuración del código resultante. Las funciones creadas se encuentran recuadradas en azul.

Tal y como se detalló en apartados anteriores, una vez concluido el diseño de la interfaz en GUIDE, se genera automáticamente un archivo .m, archivo que corresponde a la función **GuiVox**. Dicha función incluye numerosas subfunciones que corresponden a funciones de inicialización de la GUI y otras que recogen las variables introducidas o que responden al evento generado por una acción del usuario, siendo la más importante de ellas la función *pushbutton_simular_Callback*. Esta función inicia el proceso de simulación realizando la inicialización del tracto vocal (*inicializaTracto*), de las cuerdas vocales (*inicializaCuerdas*) y de la simulación (*inicializaSimul*); y realizando la llamada a la función principal del simulador: *Vox* (*FullVoxCoupled* modificada).

- *GuiVox*: recoge todas las funciones asociadas a los componentes creados en *GuiVox.fig*, a través de la herramienta GUIDE, además de las funciones de inicialización de la interfaz. Esta función es la principal en cuanto al correcto funcionamiento de la interfaz gráfica. Nada más ejecutarse se muestran los valores por defecto en todos los campos correspondientes y se crean dichas variables (globales) con esos valores. Según se vayan variando los parámetros por parte del usuario se irán modificando dichos valores iniciales y almacenando en sus variables.

pushbutton_simular_Callback: es una de las funciones de *GuiVox*. Se hace especial atención sobre la misma debido a que es la responsable de iniciar el proceso de simulación además se encargarse de mostrar en los ejes gráficos las CV y el TV.

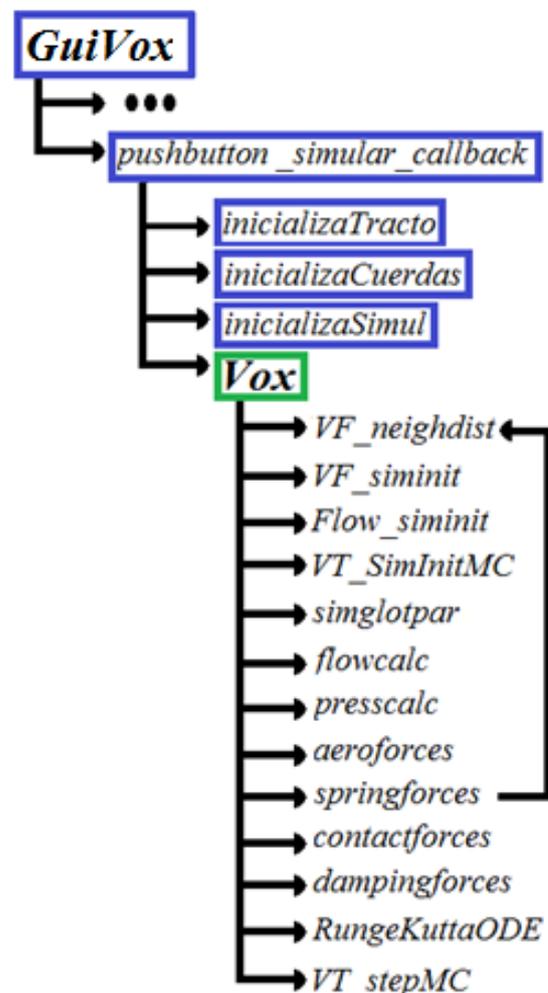


Figura 24. Estructura final del código.

- *inicializaTracto*: inicializa una estructura con los datos necesarios del tracto vocal, recogidos de la interfaz, para simular una vocal a partir de las secciones y longitudes de cada cilindro del TV para las distintas vocales descritas.

Dado el hecho de que se puede elegir en la interfaz el nº de cilindros (x) que conforman el TV, se realizan modificaciones en esta función. Por un lado, se calculan las secciones correspondientes a esos x cilindros y por otro, se aumenta la longitud de cada uno de ellos para mantener la longitud total constante. Se parte inicialmente de los valores de 44 secciones (dependientes de la vocal), por lo que para obtener los x valores necesarios ($x < 44$) se suaviza el contorno del TV original de 44 secciones por medio de un polinomio de alto grado (se ha elegido arbitrariamente un grado 8). Posteriormente se muestrea este contorno para obtener los x valores deseados.

Se tiene especial atención en que los valores obtenidos mediante este proceso no sean negativos (TV cerrado) ni sean menores que la mínima sección definida entre las 44 secciones iniciales. La figura 25 puede ayudar a comprender el proceso (ejemplo para la vocal /i/ siendo el nº de cilindros deseados 22).

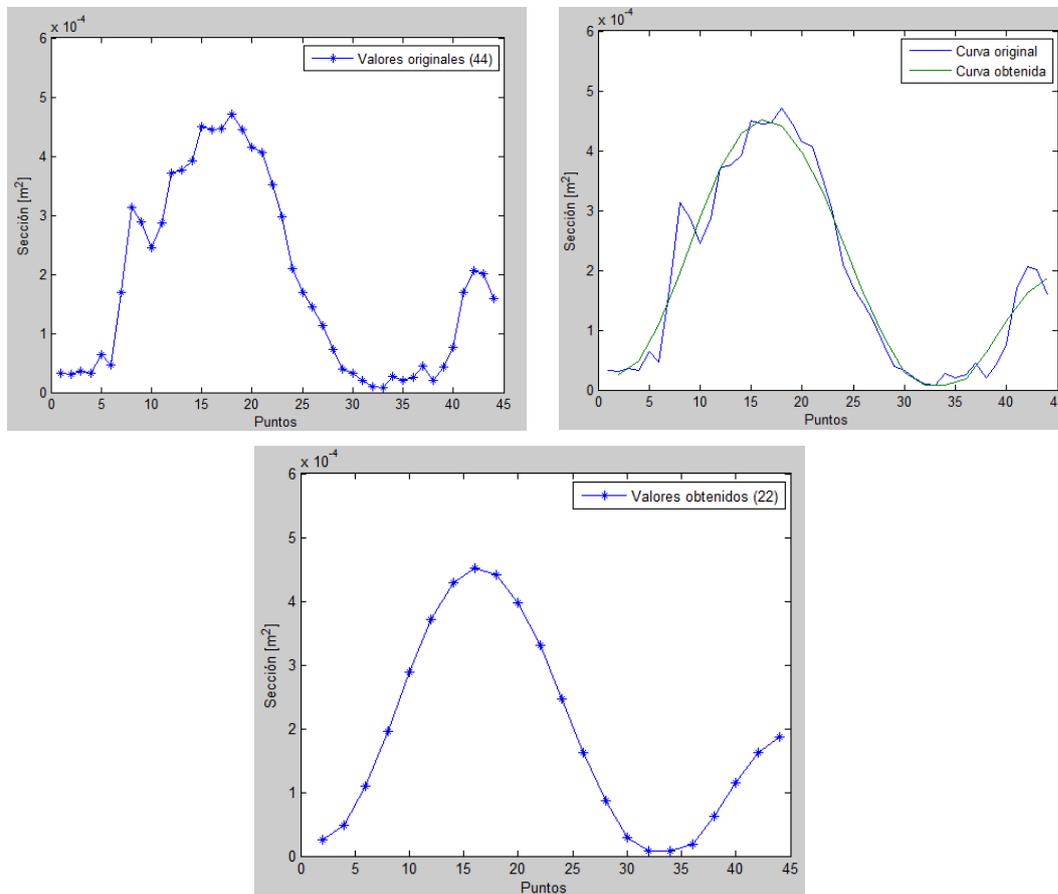


Figura 25. Obtención de las x secciones del TV (ej.: $x = 22$)

En la imagen superior izquierda, se muestran los 44 valores de las secciones del TV (*) para la vocal seleccionada mientras que en la superior derecha se realiza una comparativa entre la curva original generada con los 44 valores mencionados y la curva final obtenida a través del polinomio creado. Por último, en la imagen de abajo centrada, se muestran los 22 valores obtenidos (*)

- *inicializaCuerdas*: inicialización de los parámetros de configuración de las cuerdas vocales (nº elementos de la CV, dimensiones de las mismas, distribución de las masas, constantes de elasticidad, de amortiguamiento, constantes mecánicas del ligamento y otras geométricas además de la tensión activa). Parte de los datos provienen de la interfaz y la otra parte son datos fijos.
- *inicializaSimul*: inicializa una estructura con la información general necesaria para realizar una simulación en Vox (duración del sonido simulado, frecuencia de muestreo de simulación y de salida, presión subglótica, nombres de los ficheros resultantes...). Los datos se recogen de la interfaz.

3.3. Parametrizaciones

En este apartado se describen las diferentes parametrizaciones escogidas para la simulación artificial de una vocal en función del locutor escogido a partir de los valores normativos comentados en el apartado 2.1.

3.3.1. Parámetros de las cuerdas vocales

- **Número de masas**

Se ha decidido que el número de segmentos en que se divide la cuerda vocal siga siendo por defecto de 15 como en [2]. En la tabla 7 se observa el valor por defecto y el rango permitido al usuario seleccionar para el mismo. Se comprueba que este parámetro es independiente al tipo de locutor seleccionado.

Tabla 7. Nº de masas según locutor.

Locutor	Nº de masas	Rango nº de masas
Hombre	15	13-17
Mujer	15	13-17
Niño	15	13-17

- **Longitud total**

En función de los valores normativos encontrados en la literatura, se ha decidido que las longitudes totales de la cuerda vocal tanto para hombre como para mujer o niño sean las mostradas en la tabla 8.

Tabla 8. Longitud total de la CV según locutor.

Locutor	Longitud [mm]	Rango longitud [mm]
Hombre	19.5	17.2-23.3
Mujer	14.4	12.2-16.5
Niño	8.7	7.5-9.9

Para obtener el valor medio y el rango de la longitud de la cuerda vocal de los niños nos hemos basado en lo que afirma Kahane en [11], es decir, en que la longitud final de la CV aumenta hasta un 60% y un 34% respectivamente en chicos y chicas respecto a su longitud en la infancia. Para ello se han calculado los porcentajes citados, restado de la longitud normativa de los hombres y las mujeres, y posteriormente, realizado una media aritmética para obtener los valores del conjunto.

- **Altura y grosor de la cubierta y el cuerpo.**

A priori se realiza una primera estimación, tabla 9, de las dimensiones de la altura y el grosor de las dos partes que componen la cuerda vocal (cubierta y cuerpo) en función de la escasa información encontrada en la literatura sobre las mismas.

Tabla 9. Altura y grosor de la cubierta y el cuerpo según locutor (1)

Locutor	Altura cubierta Thm [mm]	Grosor cubierta dm [mm]	Altura cuerpo Thv [mm]	Altura cuerpo dv [mm]
Hombre	1	1.7	2.7	2.2
Mujer	1	1.7	2.7	2.2
Niño	3.7/4.3/4.9	3.7/4.3/4.9	2.7	2.2

En el modelo descrito en [2] la longitud de la cuerda vocal es de 14 mm, valor muy similar a la longitud media obtenida para una mujer adulta, por lo que los valores propuestos en [2] para la altura y el grosor de la cubierta y el cuerpo pueden pertenecer perfectamente a los de una mujer. Se toman, por consiguiente, estos valores como válidos para la mujer adulta.

Según un estudio acerca de los cambios histológicos de las cuerdas vocales en función de la edad [22], las diferencias en altura y grosor de la cuerda vocal no son significativas, por lo que se toman para el hombre adulto los valores de altura y grosor de la cubierta y el cuerpo idénticos a los de la mujer.

En esta primera estimación, los valores promedio de la altura y grosor de la cubierta se suponen, como indica [21], iguales a 5/10 de la longitud total de la cuerda vocal. Por lo que se obtiene para cada dimensión 3.7/4.3/4.9 mm para unas longitudes de CV de 7.5/8.7/9.9 mm, lo cual ya a simple vista parece un tanto exagerado. Los valores de altura y grosor del cuerpo se desconocen para los niños, por lo que se estiman iguales a los de los adultos. Tras realizar varias simulaciones de prueba (apartado 4.2) se concluye que las dimensiones escogidas para la cubierta no son válidas ya que arrojan resultados no coherentes. Como última opción, se prueba a realizar simulaciones con valores de altura y grosor, para ambas partes de la CV, idénticos a los de los adultos (tabla 10). Los resultados obtenidos son bastante más coherentes a los obtenidos con los valores de la tabla 9 para niños por lo que resultan definitivos.

Tabla 10. Altura y grosor de la cubierta y el cuerpo según locutor (2).

Locutor	Altura cubierta Thm [mm]	Grosor cubierta dm [mm]	Altura cuerpo Thv [mm]	Altura cuerpo dv [mm]
Hombre	1	1.7	2.7	2.2
Mujer	1	1.7	2.7	2.2
Niño	1	1.7	2.7	2.2

En la tabla 11 se define el margen de variación de estos parámetros que se ha fijado en un $\pm 10\%$ de los valores promedio elegidos.

Tabla 11. Rango de la altura y grosor de la cubierta y el cuerpo según locutor.

Locutor	Altura cubierta Thm [mm]	Grosor cubierta dm [mm]	Altura cuerpo Thv [mm]	Altura cuerpo dv [mm]
Hombre	0.90-1.10	1.53-1.87	2.43-2.97	1.98-2.42
Mujer	0.90-1.10	1.53-1.87	2.43-2.97	1.98-2.42
Niño	0.90-1.10	1.53-1.87	2.43-2.97	1.98-2.42

3.3.2. Parámetros del tracto vocal

- **Número de cilindros.**

Aparte de la vocal que se desee simular (que determina la forma del TV), el único parámetro modificable es el número de cilindros de los que consta. El valor por defecto utilizado es de 44 tubos de acuerdo con [30] y la variación permitida se muestra en la tabla 12, donde se observa que no depende del sexo y edad del locutor.

Tabla 12. N° de cilindros según locutor.

Locutor	N° de cilindros	Rango cilindros
Hombre	44	22-44
Mujer	44	22-44
Niño	44	22-44

Independientemente del número de cilindros, lo que sí varía en función del sexo, edad y la vocal simulada es la longitud total del tracto vocal (parámetro no visible en la interfaz). En el caso del sexo y la edad la variación es acorde con los valores normativos presentados en 2.1.1. En cuanto a las vocales, la posición de los labios más o menos prominentes hace que haya pequeñas diferencias en la longitud total para cada vocal.

Por todo ello, a partir de un array original (en Vox inicial) que contiene las longitudes de cada cilindro para cada vocal [30] se ha calculado la longitud total del TV (multiplicando por 44) para cada vocal. Acto seguido se ha obtenido el promedio

de dichas longitudes (16.99 cm) y en función del mismo, y en comparación con las medidas promedio del TV de un hombre (18.5 cm), mujer (15.75cm) y niño (8.5 cm), se han construido tres nuevos arrays con las longitudes totales del TV en función de la vocal (tabla 13). Posteriormente, los valores de los arrays resultantes han sido divididos entre 44 para obtener la longitud de cada cilindro en función de la vocal y locutor (tabla 14). Por último dichos arrays han sido agrupados en una matriz denominada mVocalTractSamplingDistAll en la función *inicializaTracto*.

Tabla 13. Longitud total del tracto vocal para las distintas vocales según locutor.

	Longitud tracto vocal [cm]									
	'i'	'I'	'E'	'ae'	'A'	'a'	'O'	'U'	'o'	'u'
Vox	16,10	16.41	15.93	16.32	16.98	17.29	17.69	17.38	17.07	18.04
Hombre	17.56	17.86	17.34	17.78	18.48	18.83	18.27	18.92	18.57	19.62
Mujer	14.92	15.22	14.78	15.14	15.75	16.02	16.41	16.10	15.84	16.72
Niño	8.05	8.23	7.96	8.18	8.49	9.67	8.84	8.71	8.54	9.02

Tabla 14. Longitud de cada cilindro del tracto para las distintas vocales según locutor.

	Longitud de cada cilindro [cm]									
	'i'	'I'	'E'	'ae'	'A'	'a'	'O'	'U'	'o'	'u'
Vox	0.366	0.373	0.362	0.371	0.386	0.393	0.402	0.395	0.388	0.410
Hombre	0.399	0.406	0.394	0.404	0.420	0.428	0.438	0.430	0.422	0.446
Mujer	0.339	0.346	0.336	0.344	0.358	0.364	0.373	0.366	0.360	0.380
Niño	0.183	0.187	0.181	0.186	0.193	0.197	0.201	0.198	0.194	0.205

3.3.3. Parámetros de simulación

▪ Presión subglótica

En la tabla 15 se presenta el valor medio y el rango de variación de la presión subglótica, en condiciones normales, para adultos y niños. Los resultados han sido obtenidos a partir de las medidas de 8 mujeres de entre 22 y 29 años [33], 7 hombres de entre 23 y 30 años [33] y un número indeterminado de niños de entre 8 y 12 años [34].

Tabla 15. Presión subglótica según locutor.

Locutor	P. subglótica media [Pa]	P. Subglótica Rango [Pa]
Hombre	745	373-1236
Mujer	745	441-1255
Niño	667	324-1344

Los valores posibles a introducir en el simulador son, por lo tanto, los mostrados en la tabla anterior.

- **Duración**

Hace referencia a la duración total del sonido a simular. Debido a la gran carga computacional que exige cada ciclo de simulación se ha escogido un margen de variación poco amplio, pero suficiente, de 0.05 a 0.5s.

- **Frecuencia de salida**

Corresponde a la frecuencia de muestreo del resultado obtenido. Los posibles valores a elegir varían desde los 8000 Hz (valor por defecto) hasta los 32000 Hz coincidiendo con las frecuencias de muestreo típicas utilizadas para audio. No se ha dado la posibilidad de alcanzar los 44100 Hz (estándar del CD-Audio) ya que para esta frecuencia de salida, en algunos casos concretos, se superaba la frecuencia de muestreo de simulación (determinada por la longitud de los cilindros en el modelo de guía de ondas).

4. RESULTADOS

En este capítulo se describen las pruebas a la que ha sido sometida la interfaz gráfica creada para comprobar su funcionamiento. En el apartado 4.1 se comprueba el correcto almacenamiento de las variables recogidas por la GUI. A continuación, en función del locutor seleccionado se realizan diferentes simulaciones variando la longitud total de las CV establecidas para caso (apartado 4.2). Se evalúan las frecuencias fundamentales obtenidas para ciertos locutores (4.3) y por último, para un único locutor determinado se realizan diferentes simulaciones de prueba variando otros parámetros de simulación en el apartado 4.4.

4.1. Almacenamiento de variables

En primer lugar y antes de realizar una comparativa entre los resultados obtenidos según la simulación de una voz femenina, masculina o infantil, se efectúa una comprobación acerca del correcto almacenamiento de los valores introducidos por el usuario por parte de la GUI.

Para ello se ejecuta una simulación de prueba en la que se modifican todos y cada uno de los parámetros variables de la interfaz para, posteriormente, a través del fichero resultado “estructuras”, corroborar que los valores escogidos (en este caso al azar) han sido almacenados correctamente en cada una de sus variables y que, por tanto, la simulación ha sido realizada en base a los mismos.

En esta simulación los valores introducidos han sido los mostrados en la figura 26. Para obtener el fichero que contiene las diferentes estructuras con los parámetros de configuración de las cuerdas vocales (eCuerdas), el tracto vocal (eTracto) y otros parámetros generales de la simulación (eSimul) se hace click en el checkbox correspondiente.

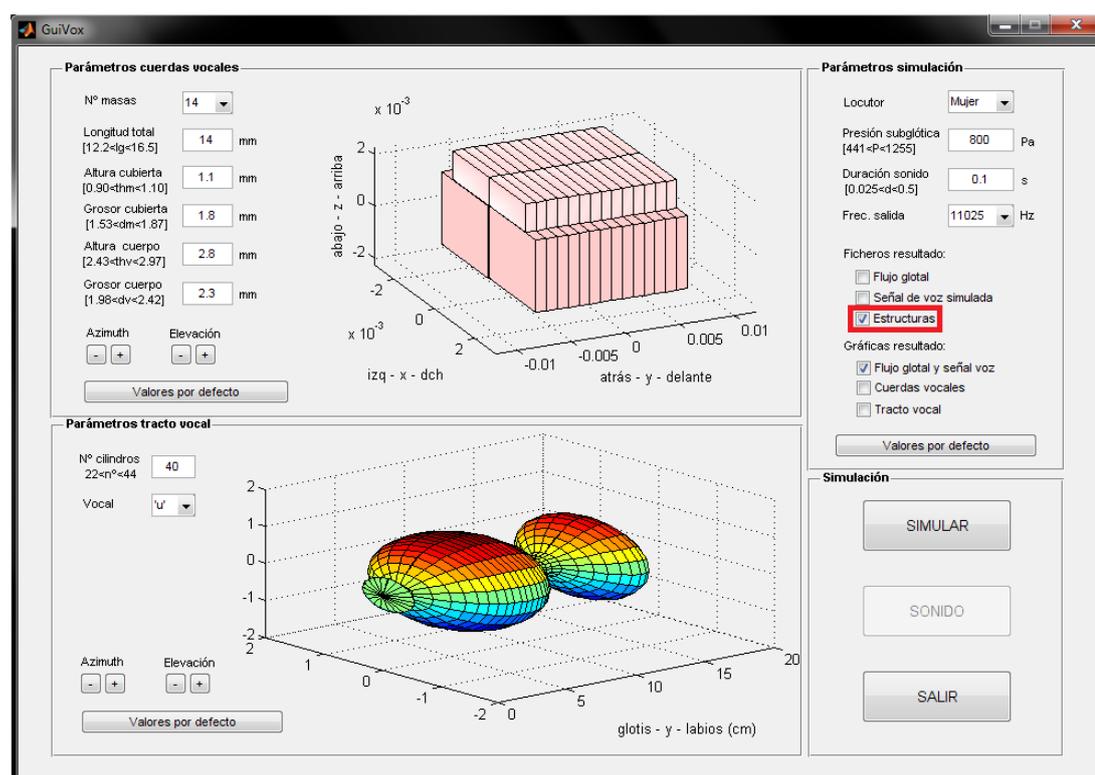


Figura 26. Comprobación del correcto almacenamiento de las variables (valores).

Una vez terminada la simulación, se obtiene un fichero denominado “Estructuras.mat”. A continuación, en la figura 27 se muestran los valores que fueron almacenados y se comprueba si son los correctos.

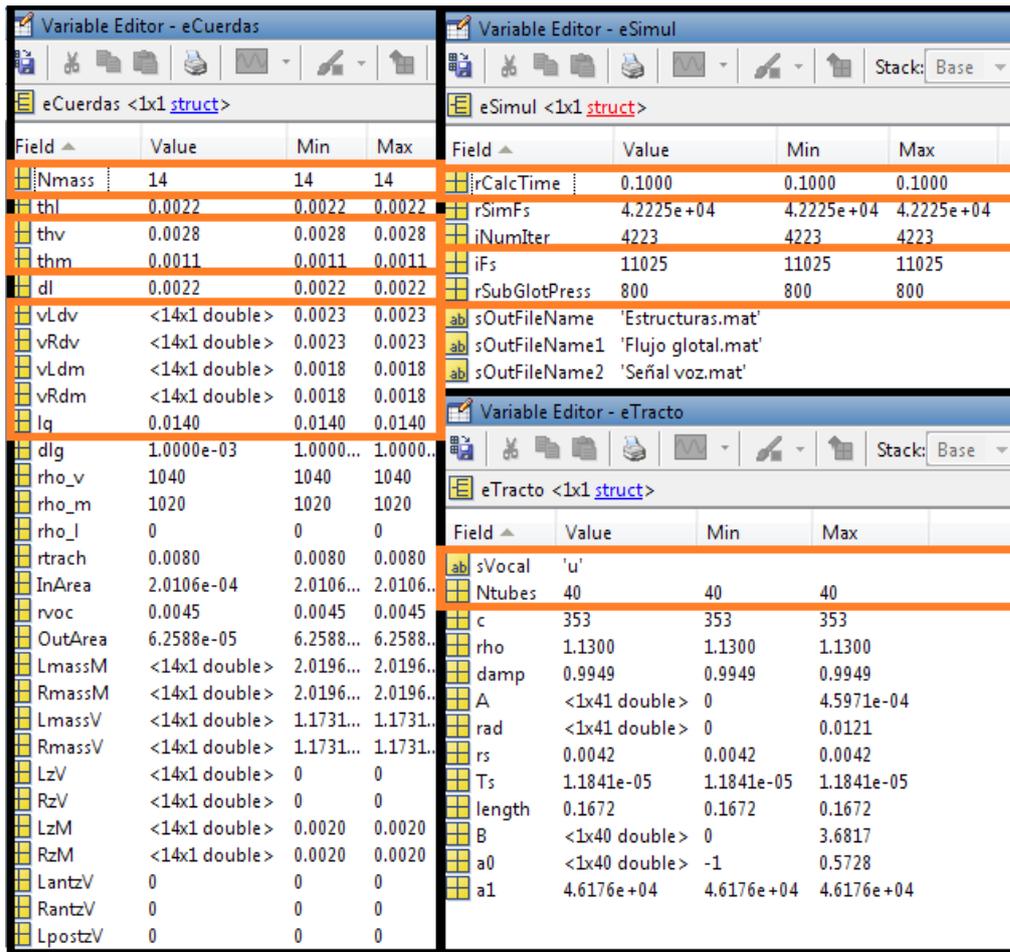


Figura 27. Editor de variables para las estructuras obtenidas.

La estructura eCuerdas está compuesta no sólo por las variables que se muestran en la figura anterior sino por más. En la tabla 16 comprobamos el resultado respecto a las variables que nos interesan.

Tabla 16. Resultado del almacenamiento en eCuerdas.

Variable	Correspondencia	Valor almacenado	Resultado
Nmass	Nº masas	14	<input checked="" type="checkbox"/>
lg	Longitud total	0.0140 m (14 mm)	<input checked="" type="checkbox"/>
thm	Altura cubierta	0.0011 m (1.1 mm)	<input checked="" type="checkbox"/>
vLdm/vRdm	Grosor cubierta (lado izq. y dcho.)	0.0018 m (1.8 mm)	<input checked="" type="checkbox"/>
thv	Altura cuerpo	0.0028 m (2.8 mm)	<input checked="" type="checkbox"/>
vLdv/vRdv	Grosor cuerpo (lado izq. y dcho.)	0.0023 m (2.3 mm)	<input checked="" type="checkbox"/>

En la tabla 17 se muestra un cuadro resumen que como en el caso anterior establece una correspondencia entre la variable y el campo de la GUI, su valor y el resultado del almacenamiento del mismo para eSimul.

Tabla 17. Resultado del almacenamiento en eSimul.

Variable	Correspondencia	Valor almacenado	Resultado
rSubGlottPress	Presión	800 Pa	<input checked="" type="checkbox"/>
rCalcTime	Duración	0.1 s	<input checked="" type="checkbox"/>
iFs	Frecuencia salida		<input checked="" type="checkbox"/>

Por último, en la tabla 18, se muestran los resultados obtenidos para eTracto.

Tabla 18. Resultado del almacenamiento en eTracto.

Variable	Correspondencia	Valor almacenado	Resultado
Ntubes	Nº cilindros	40	<input checked="" type="checkbox"/>
sVocal	Vocal	'u'	<input checked="" type="checkbox"/>

Por lo tanto, podemos concluir que el almacenamiento de los parámetros fijados a través de la interfaz se realiza correctamente y sin ningún tipo de problema.

4.2. Simulación según el locutor elegido

Para evaluar el funcionamiento del simulador se efectúan diferentes simulaciones para sacar conclusiones en función de las gráficas resultantes, para cada caso, de la señal de voz y del flujo glótico en función del tiempo. Las simulaciones realizadas en este apartado se realizan para una vocal /a/ de duración sostenida de 25 ms con los parámetros por defecto de cada locutor variando, exclusivamente, la longitud de las cuerdas vocales.

❖ Locutor masculino

Se realizan tres simulaciones diferentes variando la longitud de la cuerda vocal (l_g) masculina para su valor mínimo (17.2 mm), medio (19.5 mm) y máximo (23.3 mm) establecido. El resto de parámetros, tal y como se comenta, van a permanecer constantes en todos los casos y con sus valores por defecto para el hombre adulto.

En la figura 28 se muestran para los tres casos las gráficas resultantes.

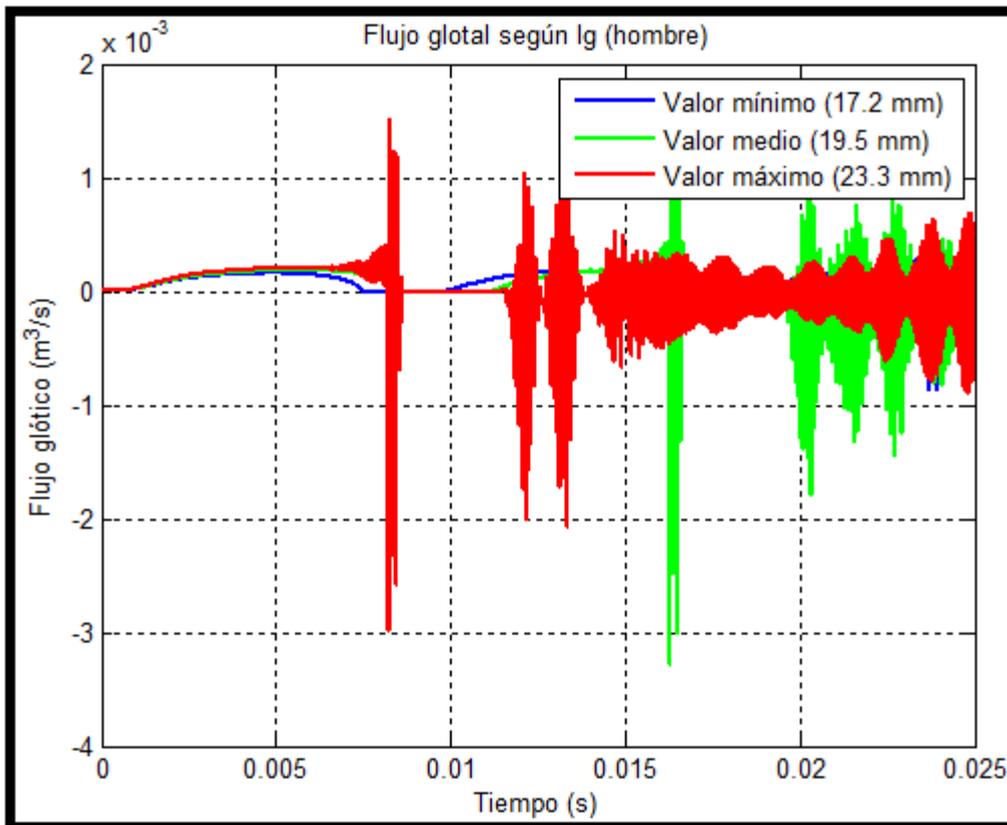


Figura 28. Flujo glótico según l_g (hombre).

Los resultados obtenidos no son coherentes para ninguna de las longitudes de la cuerda vocal. La forma de onda del flujo glótico para los tres casos no resulta reconocible. En la figura 29, se muestra un zoom sobre la anterior figura.

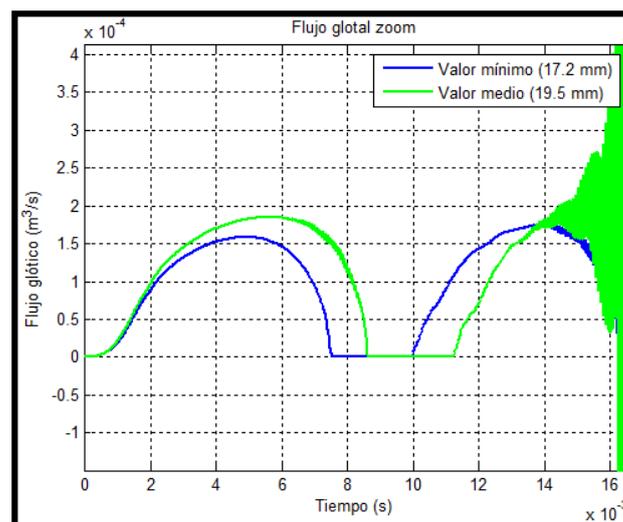


Figura 29. Zoom del flujo glótico según l_g (hombre).

Para la longitud media y mínima, en un principio, parece que el flujo obtenido toma una forma realista. Esto puede ser debido a que los parámetros seleccionados en el simulador hacen que la solución numérica de las ecuaciones diferenciales implicadas sea inestable. Este problema no implica que los valores de longitud elegidos sean incorrectos, sino que posiblemente deben ir acompañados de cambios en los valores de otros parámetros, como tensiones y rigideces que gobiernan el movimiento de las masas, para obtener resultados parecidos a los de la realidad.

Se comprueba además, en la figura 30, que la señal de voz simulada no corresponde a la de una vocal /a/.

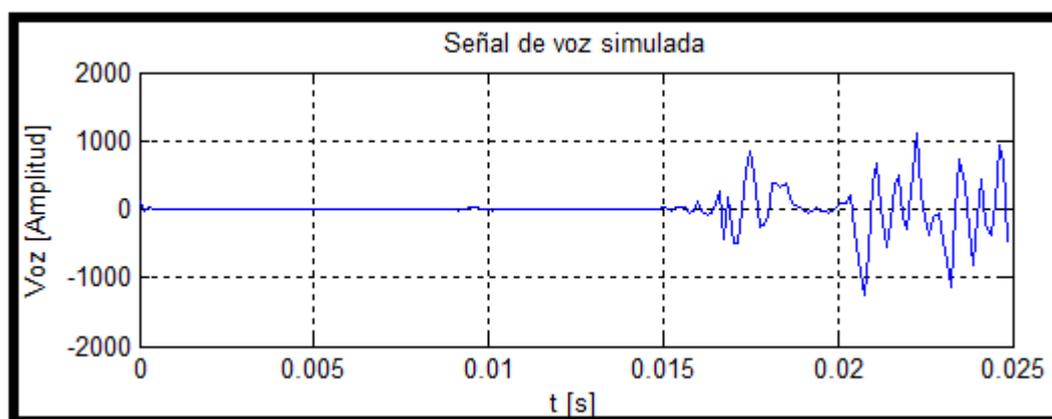


Figura 30. Señal de voz simulada según l_g media (hombre).

Dado los malos resultados obtenidos se realiza sobre el código inicial de Vox, del que se partió, una simulación de ejemplo para la longitud mínima de 17.2 mm con el fin de determinar si estos resultados son debidos a limitaciones propias del simulador inicial o son debidas a la mala parametrización de las dimensiones de la cuerda vocal masculina.

Se comprueba que las gráficas resultantes son prácticamente idénticas en ambas simulaciones por lo que estos malos resultados se deben a limitaciones, hasta ahora desconocidas, del simulador inicial del que partimos.

No obstante, vamos a realizar unas pequeñas variaciones sobre otros de los parámetros definidos en la GUI, dejando la longitud de la cuerda vocal mínima, para observar si los resultados siguen siendo similares a los obtenidos anteriormente.

- Variación del nº de masas que conforman la CV: se ha probado a simular con 13 y 17 respectivamente, obteniendo resultados similares a los anteriores.

- Variación de la altura y el grosor de la cubierta y el cuerpo hasta sus valores máximos: no hay cambios apreciables.

- Variación de la altura y el grosor de la cubierta y el cuerpo (valores máximos: thm 1.1 mm, dm 1.87 mm, thv 2.97 mm y dv 2.42 mm) y modificación del número de masas a 13 y 17 (figura 31).

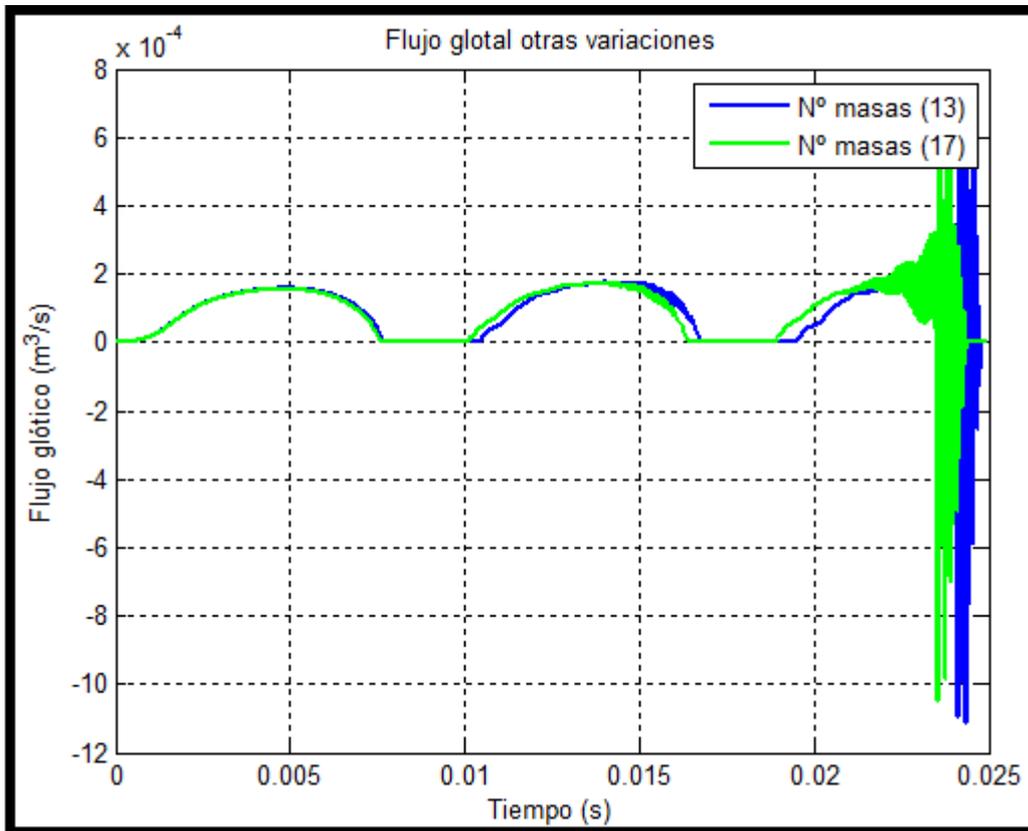


Figura 31. Simulación l_g min, altura y grosor máximas ambas capas para $N=13$ (izquierda) y $N=17$ (derecha).

Los resultados obtenidos siguen mostrando la misma tendencia que en las simulaciones anteriores, en las que al cabo de unos pocos pulsos glóticos se producen inestabilidades. Por este motivo, podemos concluir que el simulador Vox a partir de los parámetros anatómicos de un hombre no funciona correctamente.

❖ Locutor femenino

Al igual que en el punto anterior se realizan tres simulaciones diferentes (figura 32) variando la longitud de la cuerda vocal (l_g) femenina para su valor mínimo (12.2 mm), medio (14.4 mm) y máximo (16.5 mm) establecido. El resto de parámetros van a permanecer constantes en todos los casos y con sus valores por defecto para la mujer adulta.

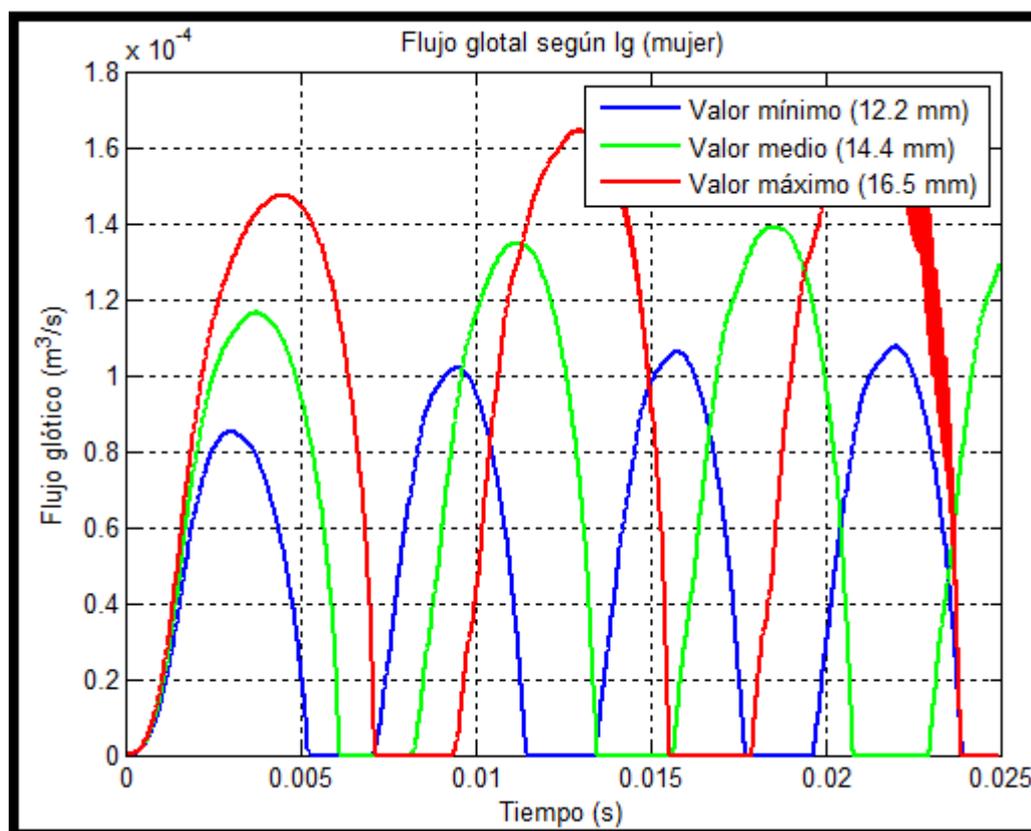


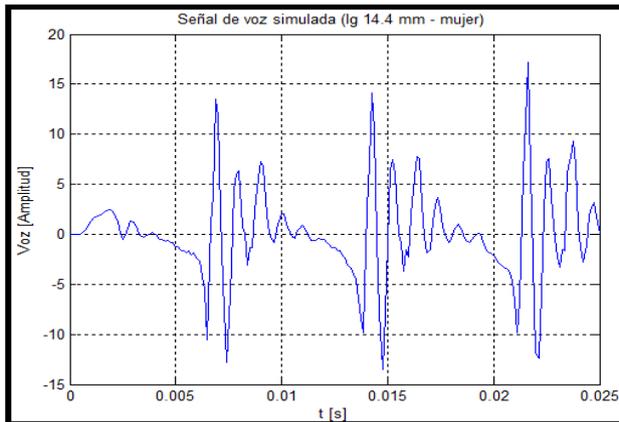
Figura 32. Flujo glótico según l_g (mujer).

Para una longitud de cuerda vocal máxima (16.5 mm) se comprueba que aunque la señal de flujo glótico tiene una forma y periodicidad adecuadas según se avanza en el tiempo se vuelve inestable tal y como ocurría en el caso del locutor masculino.

Realizando otras simulaciones de prueba, se establece que, para valores superiores a 15 mm de longitud de cuerda vocal, el simulador no responde acorde a lo esperado.

Tanto para una longitud mínima como media, los resultados son correctos. Los flujos glóticos obtenidos son proporcionales a la longitud de las cuerdas vocales.

En la figura 33 se muestra la señal de voz resultante para una longitud media de las CV de un locutor femenino (14.4 mm).



En este caso, para un locutor femenino, se comprueba que la señal de voz resultante sí tiene un aspecto parecido al de una vocal /a/.

Figura 33. Señal de voz simulada según l_g media (mujer).

❖ Locutor infantil

En primer lugar, se van a realizar varias simulaciones de prueba (figura 34) para los valores inicialmente propuestos en la tabla 9 del apartado 3.3.1 de las dimensiones de grosor y altura de las CV por defecto variando la longitud de la cuerda vocal (l_g) infantil para su valor mínimo (7.5 mm), medio (8.7 mm) y máximo (9.9 mm) establecido.

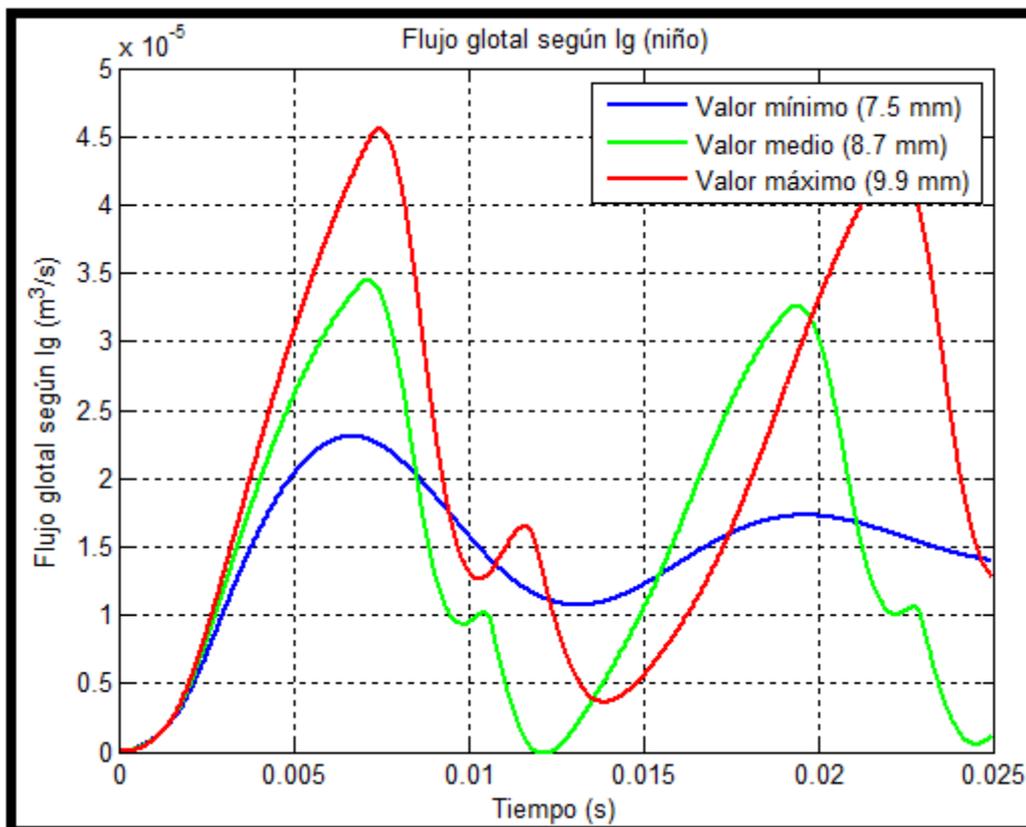
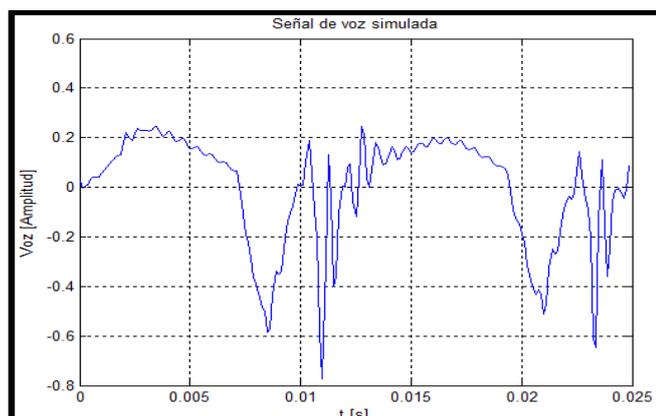


Figura 34. Flujo glótico según l_g (niño) para dimensiones tabla 9.

También se muestra a continuación la señal de voz resultante para la longitud media de las cuerdas vocales infantiles para estas dimensiones (figura 35).



Los resultados obtenidos tanto en la figura 34 como en la 35 muestran unas formas de ondas que no se corresponden con las esperadas. En consecuencia se prueba a realizar simulaciones con valores de altura y grosor, para ambas partes de la CV, idénticos a los de los adultos (tabla 10 del apartado 3.3.1).

Figura 35. Señal de voz simulada según l_g media (niño) para dimensiones tabla 9.

Se realizan, por tanto, tres simulaciones diferentes (figura 36) variando únicamente la longitud de la cuerda vocal (l_g) infantil para su valor mínimo (7.5 mm), medio (8.7 mm) y máximo (9.9 mm) establecido. El resto de parámetros van a permanecer constantes en todos los casos y con sus valores por defecto para un niño.

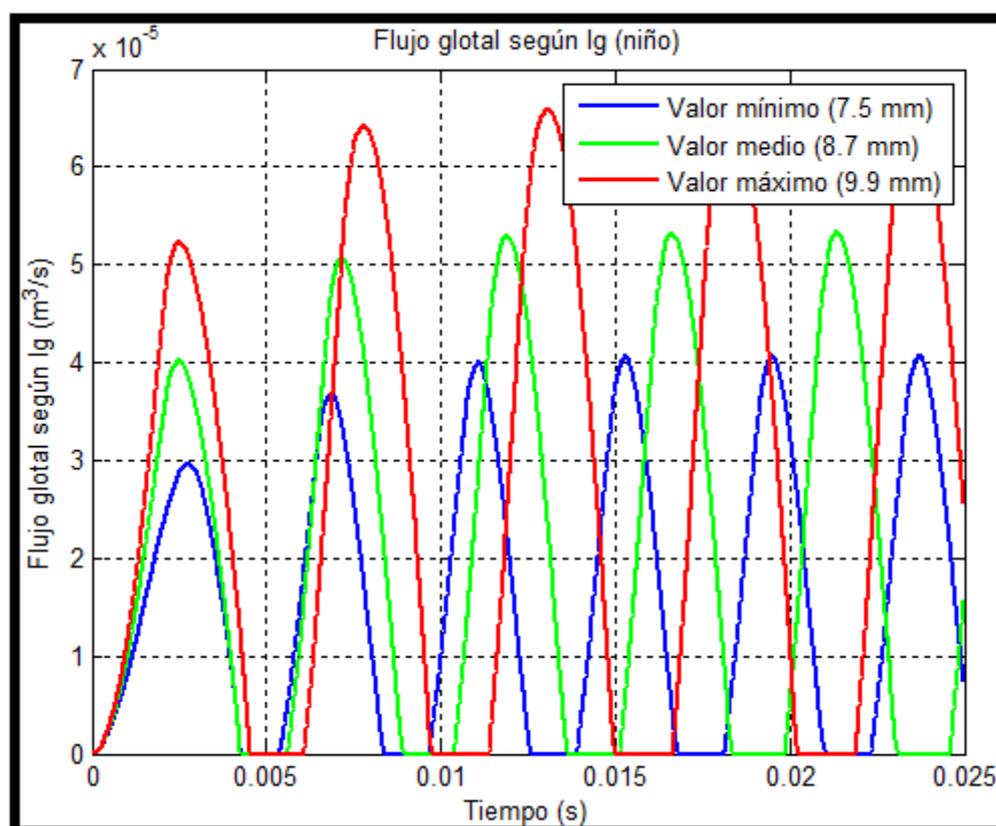


Figura 36. Flujo glótico según l_g (niño) para dimensiones tabla 10.

Los resultados obtenidos son coherentes, por lo que, se deja como valores por defecto de la altura y el grosor de las capas de la CV, los mostrados en la tabla 10. Se cumple igualmente que a mayor longitud de las cuerdas vocales, mayores son los flujos glóticos resultantes.

En la figura 37, se presenta la señal de voz resultante para la longitud media de las cuerdas vocales infantiles para las dimensiones escogidas.

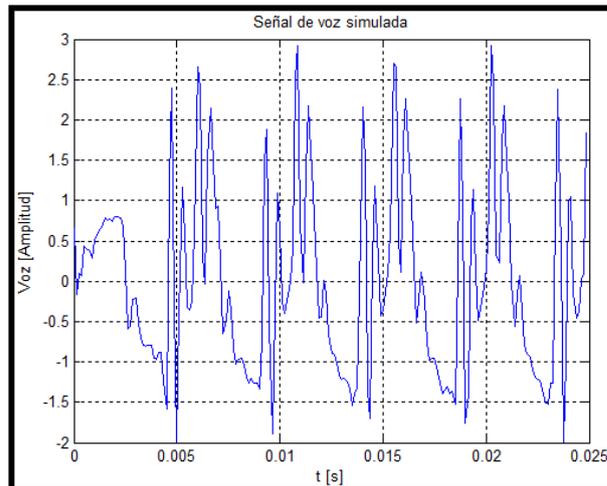


Figura 37. Señal voz simulada según l_g media (niño) para dimensiones tabla 10.

En este caso, la señal resultante es correcta, ya que es similar a la señal de voz de una vocal /a/.

4.3. Frecuencias fundamentales

La frecuencia fundamental (F_0) es el parámetro acústico más importante que caracteriza la voz de un locutor.

Se procede a obtener las frecuencias fundamentales para un locutor femenino y un locutor infantil a partir de las gráficas obtenidas del flujo glótico. Se excluye, de esta manera, a un locutor masculino debido al mal funcionamiento comprobado del simulador Vox a partir de los parámetros anatómicos reales de un hombre promedio.

Las simulaciones se van a realizar, en ambos casos, para una longitud media establecida de la CV correspondiente en cada caso y para una vocal /a/.

❖ Frecuencia fundamental para un locutor femenino (F_{o1})

A partir de la figura 38 se obtiene el periodo de la señal (T_1) para posteriormente obtener la frecuencia fundamental resultante de la simulación.

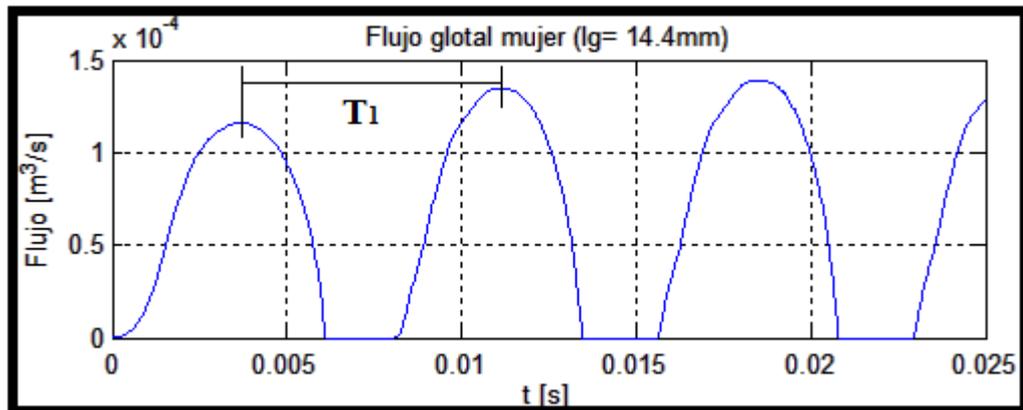


Figura 38. Flujo glotal obtenido para una mujer promedio.

Para este caso el valor obtenido de T_1 ha sido de 7.4 ms, por lo que:

$$F_{o1} = \frac{1}{T_1} = \frac{1}{0.0074 \text{ s}} = 135 \text{ Hz}$$

❖ Frecuencia fundamental para un locutor infantil (F_{o2}):

De manera equivalente, a partir de la figura 39, se calcula la frecuencia fundamental resultante.

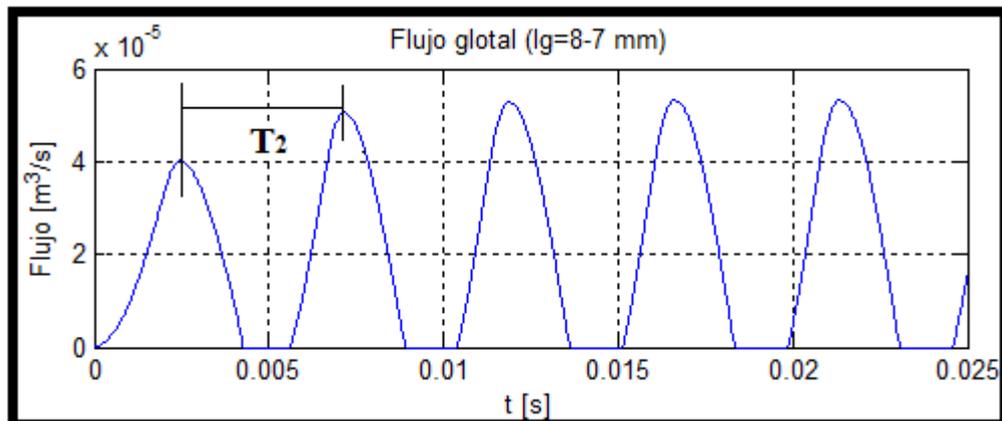


Figura 39. Flujo glotal obtenido para un niño promedio.

Para este caso el valor obtenido de T_2 ha sido de 4.6 ms, por lo que:

$$F_{o2} = \frac{1}{T_2} = \frac{1}{0.0046 \text{ s}} = 217 \text{ Hz}$$

Como se puede comprobar en ambos casos, las frecuencias fundamentales obtenidas no coinciden con las frecuencias fundamentales medias establecidas para una mujer o niño siendo éstas de 200 Hz y 325 Hz respectivamente [6]. Este hecho puede ser debido o bien al funcionamiento interno de Vox o bien a que hay que variar otros parámetros de las CV tales como las rigideces o las tensiones para obtener frecuencias más adecuadas. Sin embargo, los resultados obtenidos son coherentes ya que se cumple que a menor longitud de la cuerda vocal se obtiene una mayor frecuencia.

En este caso, se ha intentado predecir diferencias en la frecuencia fundamental en base a la longitud de la cuerda vocal únicamente, por lo que se asume que hay tensiones comunes en los tejidos laríngeos cuando probablemente no fuera el caso. Aun en base a esta suposición atrevida de igualdad de la tensión del tejido en estado relajado, la frecuencia fundamental se puede estimar de acuerdo a la siguiente proporción [6]:

$$\frac{Fo2}{Fo1} = \frac{Lg1}{Lg2}$$

donde los subíndices 1 y 2 hacen referencia a las dos longitudes y dos frecuencias comparadas.

Resultando en nuestro caso:

$$\frac{217 \text{ Hz}}{135 \text{ Hz}} = \frac{14.4 \text{ mm}}{8.7 \text{ mm}} \approx 1,6$$

Por lo que la relación descrita por Titze en [6] se cumple.

4.4. Variación de otros parámetros

Para no llenar el capítulo de gráficas repetitivas se ha elegido realizar el resto de pruebas exclusivamente para un locutor *femenino*.

Se van a realizar diferentes simulaciones:

- Simulación en función de la vocal a simular.
- Simulación en función del número de masas que conforman la CV.
- Simulación en función de la altura y grosor de la CV.
- Simulación en función del número de cilindros que conforman el TV.
- Simulación en función de la presión subglótica

❖ Simulación en función de la vocal a simular

Se realizan, a continuación, varias simulaciones en función de la vocal a sintetizar manteniendo los parámetros por defecto establecidos para una mujer adulta. En la figura 40 se muestra tanto la señal de voz como el flujo glótico obtenido para las vocales: /a/, /ae/, /i/, /o/ y /u/.

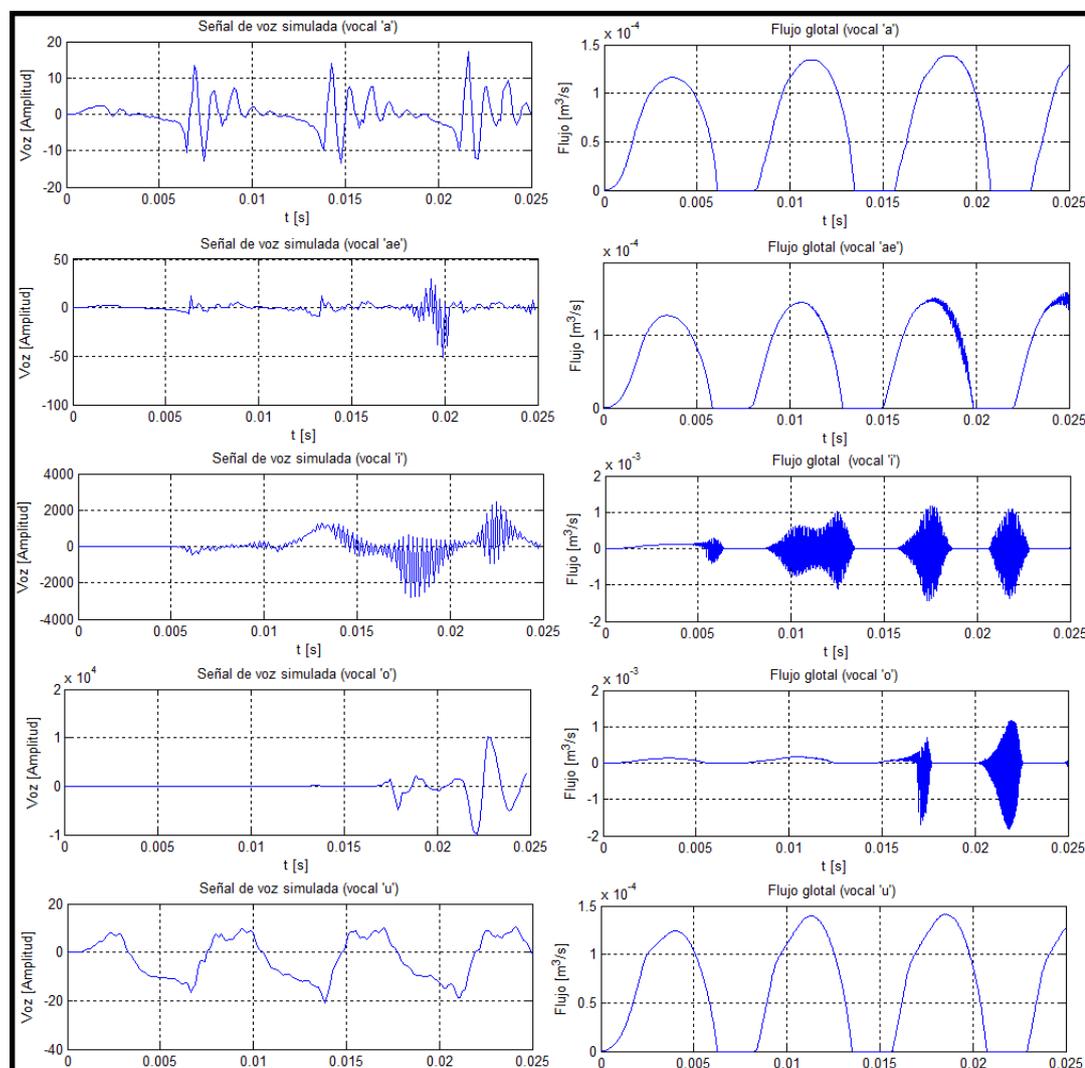


Figura 40. Señal de voz y flujo glótico para varias vocales.

Los resultados obtenidos son los deseados únicamente para la vocal /a/ y /u/ tanto por la forma de la señal de voz como por el flujo obtenido en cada caso; para /ae/ en cambio, la señal de voz no parece correcta y aunque su flujo glótico tiene una forma y periodicidad adecuadas según se avanza en el tiempo la señal del flujo glótico se vuelve inestable; para /i/ y /o/ tanto la señal de voz como el flujo glótico son erróneos. Estos malos resultados provienen del simulador inicial de partida (secciones y longitudes de los cilindros comunes).

Los sonidos obtenidos son difícilmente diferenciables además de tener muy poca naturalidad.

❖ Simulación en función del número de masas que conforman la CV.

Todas las simulaciones realizadas de aquí en adelante se realizan para la vocal /a/.

En la figura 41 se realiza una comparativa entre los flujos glóticos obtenidos según la variación del número de masas (segmentos que forman las cuerdas vocales) al valor mínimo (13), medio (15) y máximo (17) posibles.

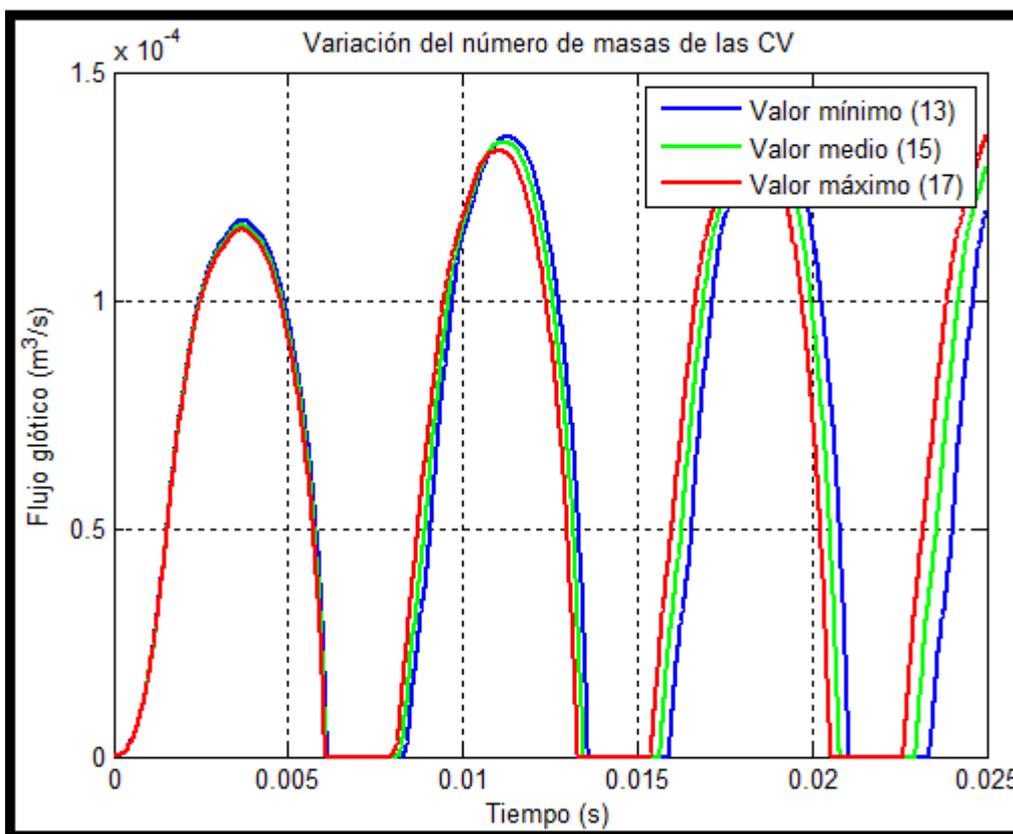


Figura 41. Comparación flujo glótico variando el nº de masas de las CV.

Se observa que los flujos glóticos obtenidos son equivalentes. Existe un pequeño retardo entre las diferentes señales que se ve incrementado según el paso del tiempo. Cuanto menor es el número de masas mayor es el retardo. De manera análoga, la diferencia entre los flujos glóticos va en aumento según transcurre el tiempo.

❖ Simulación en función de la altura y grosor de la CV

Se efectúa diferentes simulaciones modificando los valores de la altura y grosor tanto del cuerpo como de la cubierta a sus valores mínimos, medios y máximos (figura 42).

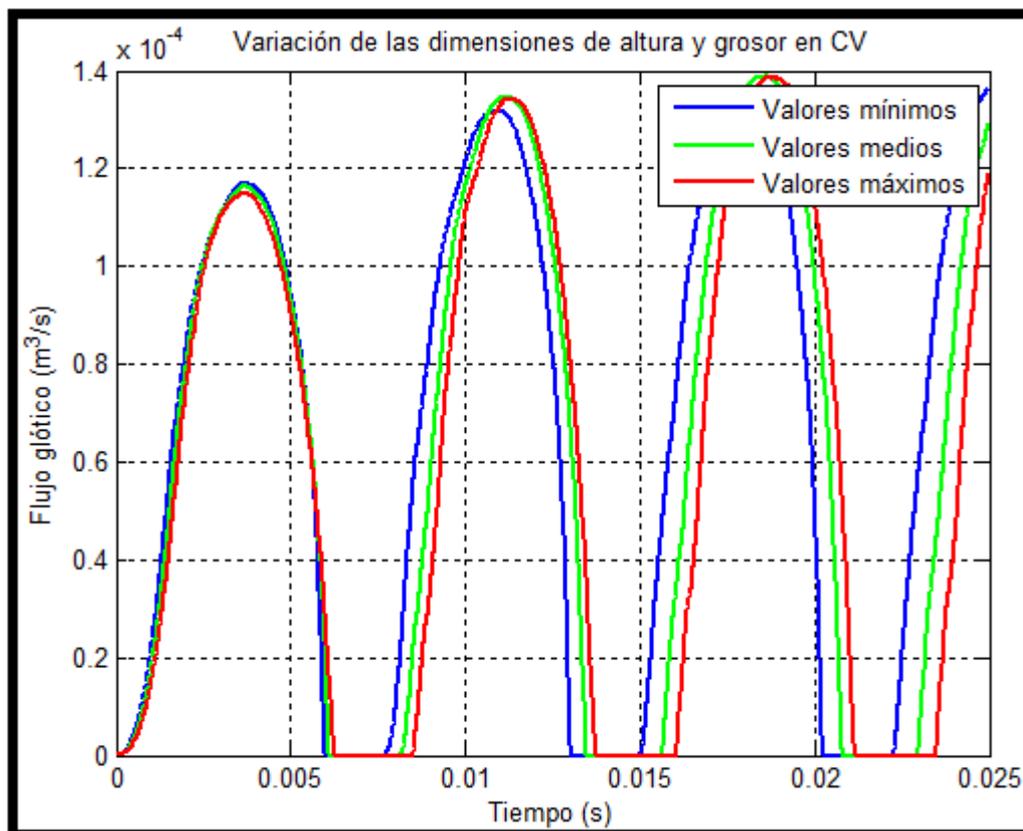


Figura 42. Comparación flujo glótico variando altura y grosor de las CV.

Las diferencias son pequeñas, aunque se observa igualmente que van en aumento en función del tiempo. En este caso a mayores valores de las dimensiones de la altura y grosor de ambas capas de la CV, mayor es el retardo. Esto es lógico puesto que mayores dimensiones implican mayor masa de las cuerdas, lo que implica una frecuencia fundamental más baja y un periodo mayor.

❖ Simulación en función del número de cilindros que conforman el TV.

Se realiza igualmente una comparativa entre los flujos glóticos obtenidos tras realizar diversas simulaciones variando el número total de cilindros que van a formar el tracto vocal (figura 43). Se prueba a simular con 22, 33 y 44 cilindros.

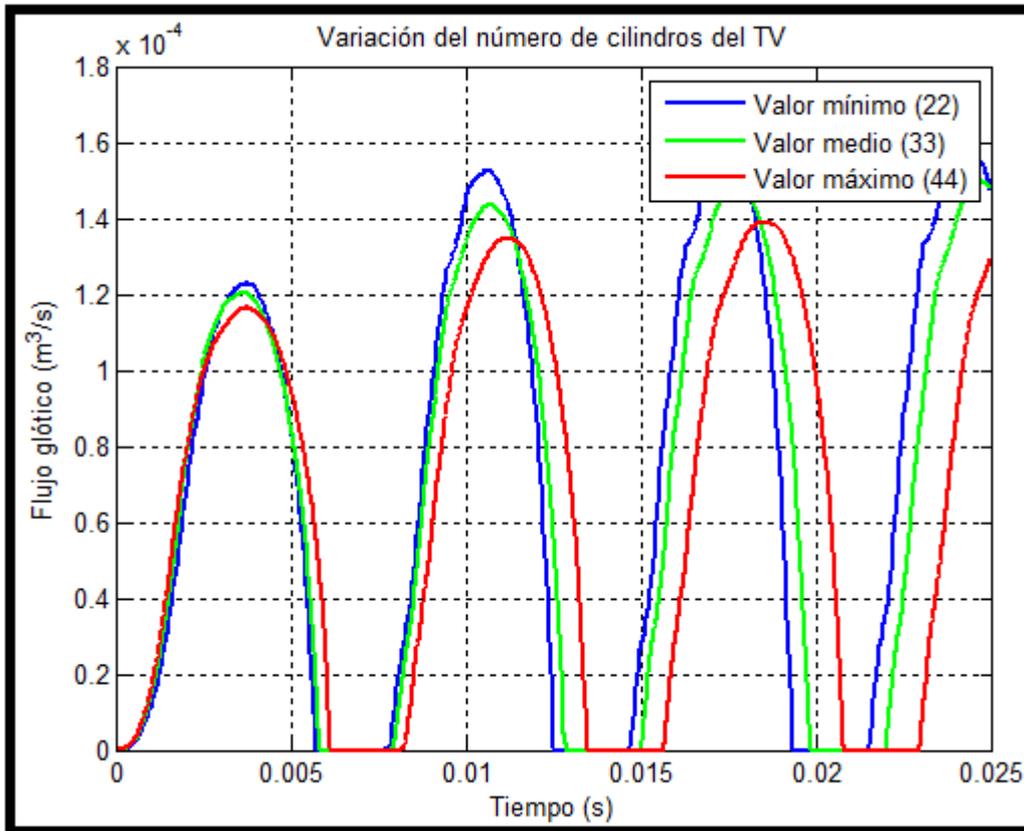


Figura 43. Comparación flujo glótico variando el nº de cilindros del TV.

Se comprueba que a menor número de cilindros, mayor es la degradación obtenida de las señales resultantes del flujo glótico. El objetivo que se pretende al simular con un número inferior de cilindros es que el tiempo de simulación sea menor por lo que de manera paralela se ha cronometrado el tiempo que tarda el simulador en realizar la síntesis para cada caso. Para una simulación de una vocal 'a' de duración de 25 ms se han obtenido los resultados mostrados en la tabla 19.

Nº cilindros	Tiempo transcurrido de la simulación [s]
22	6.731455
33	8.964026
44	12.187763

Tabla 19. Tiempo transcurrido de la simulación en función del nº cilindros TV.

Se comprueba de esta manera que compensa simular con un número inferior de cilindros asumiendo las degradaciones obtenidas en la señal resultante.

❖ Simulación en función de la presión subglótica

Acto seguido, en la figura 44, se realiza una comparativa de los flujos glóticos resultantes de las simulaciones realizadas según la variación de la presión subglótica al valor mínimo (441 Pa), medio (745 Pa) y máximo (1255 Pa) establecido.

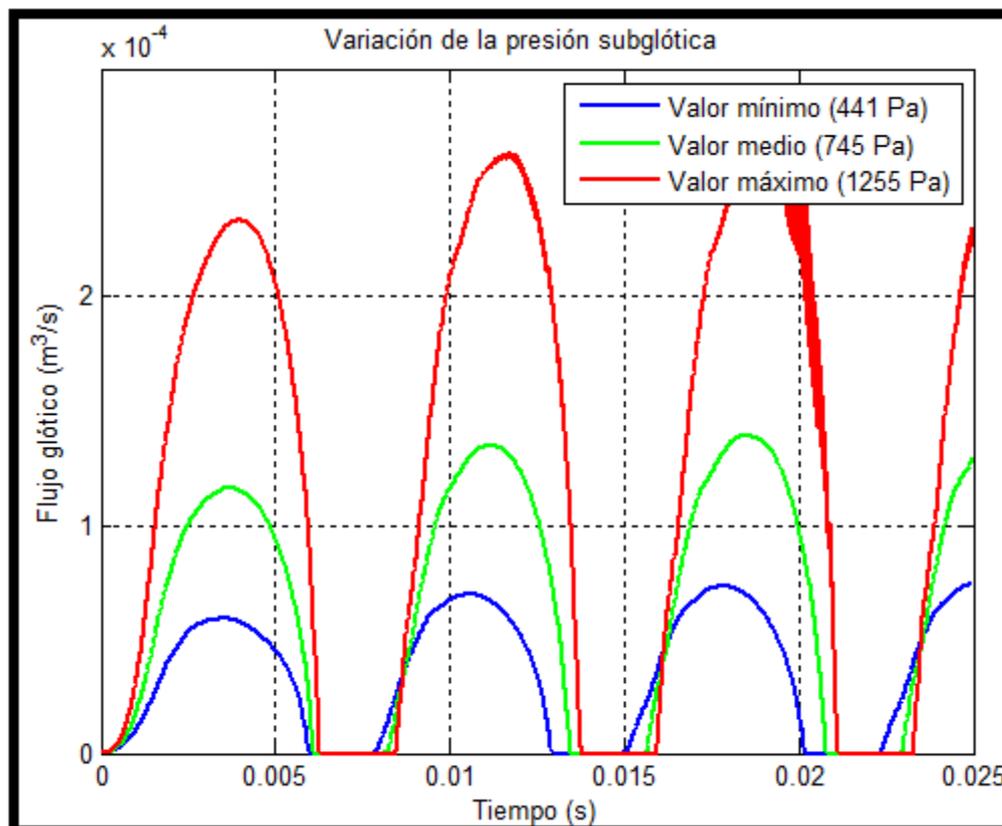


Figura 44. Comparación flujo glótico variando la presión subglótica.

Respecto a los flujos obtenidos para el valor de presión mínimo y medio simulado, los resultados son coherentes ya que a mayor presión subglótica mayor es el flujo glótico resultante. En cambio, para el valor de presión máximo establecido, aunque esta tendencia también se cumple, se comprueba que el sistema se vuelve inestable. Se realizaron otras pruebas, obteniendo la conclusión de que a presiones mayores de 1000 Pa, el simulador no responde de manera adecuada.

5. CONCLUSIONES Y LINEAS FUTURAS

En este capítulo se incluye una visión general del trabajo desarrollado: problema propuesto, solución planteada y resultados obtenidos. Todo ello permite obtener determinadas conclusiones sobre el proyecto. Finalmente, se proponen posibles líneas futuras de trabajo para completar y mejorar tanto el simulador Vox como su interfaz gráfica.

El estudio y comprensión del funcionamiento del simulador Vox en su totalidad han hecho posible la mejora del código del programa y la implementación de una interfaz de usuario sencilla e intuitiva, que permite modificar rápidamente algunos parámetros fundamentales.

Una de las principales limitaciones del simulador de partida, además de la imposibilidad de generar palabras (sólo vocales), es que realiza las simulaciones en base a unos parámetros anatómicos y fisiológicos fijos. Tras una minuciosa labor de investigación y búsqueda de información acerca de las dimensiones tanto de las cuerdas vocales como del tracto vocal, en función del sexo y de la edad de un locutor, se ha conseguido completar el simulador inicial haciendo posible también la síntesis de vocales a partir de los parámetros obtenidos.

Los resultados obtenidos atendiendo a estos parámetros reales de las cuerdas vocales y el tracto vocal no han sido, en algunos casos, los esperados ya sea debido a limitaciones del simulador de partida o a la no modificación de algunos de los parámetros de las cuerdas vocales tales como rigideces, constantes de amortiguamiento y tensión activa, cuyos valores normativos no están disponibles en la literatura científica revisada.

Por ello se propone, por un lado, realizar el ajuste de las parametrizaciones escogidas para un locutor masculino, femenino e infantil; y por otro, ampliar la interfaz de usuario haciendo posible la modificación del resto de parámetros de las cuerdas vocales que permanecieron fijos en el trabajo actual.

Una línea a seguir para obtener un mayor realismo de las simulaciones realizadas sería la inclusión de un modelo de tracto nasal acoplado al tracto vocal modelado en el simulador, permitiéndole así generar voz más natural por medio de la nasalización.

Por último, se podría estudiar la opción de añadir a la interfaz gráfica un nuevo modulo que permitiera simular voces patológicas con los trastornos más frecuentes como son los nódulos, pólipos, quistes, laringitis, edemas y cánceres laríngeos.

6. REFERENCIAS BIBLIOGRÁFICAS

- [1] M. Kob, *Physical Modeling of the Singing Voice*. Berlin: Logos Verlag, 2002.
- [2] R. Fraile, M. Kob, J. I. Godino-Llorente, N. Sáenz-Lechón, V. J. Osma-Ruiz, J. M. Gutiérrez-Arriola, *Physical simulation of laryngeal disorders using a multiple-mass vocal-fold model*. Biomedical Signal Processing & Control, Vol. 7(1), pp. 65-78, 2012.
- [3] A. J. Lara Peinado, *Corrección experimental de lesiones iatrogénicas de la cuerda vocal. Modelo experimental canino*. Tesis doctoral, Universidad de Alcalá, 2006.
- [4] M. Hirano, *Morphological structure on the vocal cord as a vibrator and its variations*, Folia Phoniatr. 26, pp. 89-94, 1974.
- [5] B. Torres, *Anatomía de la voz*, Ed. Paidotribo, 2008.
- [6] I.R. Titze, *Principles of voice production (Second Printing)*, NCVS, pp. 16-18, 2000.
- [7] M. M. Joshi, S. S. Hoshi y S. D. Joshi, *The morphological study of adult human larynx in a Western Indian population*, Journal of Laryngology and Voice, Vol I, 2011.
- [8] A. Hesham, A. Hamza, A. Gaafar, A. Ahmed, *Partial laser arytenoidectomy in the management of bilateral vocal fold immobility: A modification based on functional anatomical study of the cricoarytenoid joint*. Saudi Med J;26:1539-45, 2005.
- [9] M. S. Hahn, B. A. Teply, M. M. Stevens, S. M. Zeitels y R. Langer, *Collagen composite hydrogels for vocal fold lamina propria restoration*, Biomaterials 27, pp 1104-1109, 2006.
- [10] I. R. Titze, *The human instrument*, Sci. Am. 298 (1): 94-101, 2008.
- [11] J. C. Kahane, *Growth of the human prepubertal and pubertal larynx*, J Speech Hear 25: 226- 455, 1982.
- [12] D. G. Childers, *Speech processing and synthesis toolboxes*. John Wiley e hijos, New York, 2000.
- [13] I. R. Titze, *Principles of voice production*, Prentice Hall, 1994.
- [14] M. Hirano, *Phonosurgery, basic and clinical investigations*, Otología (fukuoka), 21:239-442, 1975.
- [15] P. Sidlof, J. G. Svec, J. Horacek, J. Vesely, I. Flepacek, R. Havlik, *Geometry of human vocal folds and glottal channel for mathematical and biomechanical modeling of voice production*, Journal Biomechanics 41, pp 985-995, 2008.
- [16] B. H. Story e I.R. Titze, *Voice simulation with a body-cover model of the vocal folds*, J. Acoust. Soc. Am. 97, no. 2, pp. 1249-1260, 1995.
- [17] I.R. Titze, *The Human Vocal Cords: A Mathematical Model, Part I*, Phonetica 28, pp. 129-170, 1973.

- [18] B. H. Story, *An overview of the physiology, physics and modeling of the sound source for vowels*, Acoust. Sci. & Tech. 23, 2002.
- [19] F. Alipour, D. A. Berry e I. R. Titze, *A finite-element model of vocal-fold vibration*, J. Acoust. Soc. Am 108 (6), 2000.
- [20] J.C. Torres Pazmiño y E. Marrugo, *Relación entre la longitud de la cuerda vocal y la composición histológica de la lámina propia en laringes pediátricas*, Rev. Fac. Med, Vol. 57 N°3, 2009.
- [21] R. García-Tapia e I. Cobeta, *Diagnóstico y tratamiento de los trastornos de la voz*, Editorial Garsi, Madrid, 1996.
- [22] J. A. Ximenes, P. H. Saldivado, D. Hiroshi y L. Ubirajara, *Histologic changes in human vocal folds correlated with aging: a histomorphometric study*, Ann Otol Rhinol Laryngol 112, 2003.
- [23] J. Sundberg, *The Science of the Singing Voice*, Northern Illinois University Press, Dekalb, Illinois, 1987.
- [24] J. Van Den Berg, *Myoelastic-aerodynamic theory of voice production*, Journal of Speech and Hearing Research I, pp. 227-244, 1958.
- [25] J. Martí Roca, *Situación actual de la síntesis de voz*, Estudios de Fonética Experimental IV, pp. 147-166, 1990.
- [26] T. Haji, K. Mori, K. Omori y N. Isshiki, *Mechanical Properties of the Vocal Fold*, Acta Otolaryngol 112, pp. 559-565, 1992.
- [27] K. Ishizaka y T. Kaneko, *On Equivalent Mechanical Constants of the Vocal Cords*, J. Acoust. Soc. Japan 24, no. 5, pp.312-313, 1968.
- [28] S. Mathur, *Variable-length vocal tract modeling for speech synthesis*. University of Arizona. Tesis disponible: <http://sal.shs.arizona.edu/~smathur/MSThesis.html>, 2003.
- [29] J. L. Kelly and C. C. Lochbaum, *Speech synthesis*, actas del 4th International Congress on Acoustic, 1962. Impreso en: J.L. Flanagan and L.R. Rabiner (Editors): *Speech Synthesis* (Dowden, Hutchinson & Ross, Stoudsburg), S. 127-130, pp. 1-4.
- [30] B. H. Story e I.R. Titze, *Parameterization of vocal tract area functions by empirical orthogonal modes*, Journal of Phonetics, vol. 26, pp. 223-260, Jul. 1998.
- [31] J. Martínez, J. Agulló y S. Cardona, *Conical bores. Part II: Multiconvolution*, Journal of the Acoustical Society of America, vol. 84, pp. 1620-1627, Nov. 1988.
- [32] S. T. Smith, *Matlab®: Advanced GUI Development*. Dog Ear Publishing, 2006.
- [33] J. V. Wilson y H. A. Leeper, *Changes in laryngeal airway resistance in young adult men and women as a function of vocal sound pressure level and syllable context*, Journal of voice, vol. 6, pp. 235-245, 1992.
- [34] A. Keilmann y C. A. Bader, *Development of aerodynamic aspects in children's voice*, 31 (2-3): 183-90, 1995.