

**Implementación y análisis de un método automático de detección del contorno superior de
la lengua en secuencias de imágenes de ultrasonido**

Carlos Alfredo Lozano Herrera, Janna Pamela Gómez Reyes

Trabajo de grado para optar al título de Ingeniero Electrónico

Director

Franklin Alexander Sepulveda Sepulveda

PhD en Ingeniería

Universidad Industrial de Santander

Facultad de Ingenierías Físico-Mecánicas

Escuela de Ingenierías Eléctrica, Electrónica y de Telecomunicaciones

Bucaramanga

2017

Dedicatoria

“A ella”

Carlos Alfredo Lozano Herrera.

A Dios por llenarme de su gracia para cumplir cada uno de mis sueños.

A mis padres Mario y Teresa, quienes con su amor y dedicación son parte fundamental en cada uno de mis pasos.

A mis hermanas Nadia, Paola, Narly, Angélica y mi hermano Mario, ellos son la alegría de mi vida.

A mis amigos Carlos Lozano, Oscar Jaimes, Alejandra Jaramillo, Cristian Esparragoza, gracias por acompañarme en esta etapa que culmina.

Janna Pamela Gómez Reyes.

Agradecimientos

A Dios por no dejarme desistir en este desafiante, pero tan gratificante camino...

*A mi férreo padre Gustavo Lozano y mi asidua hermana Carolina Lozano por darme un motivo
más por el cual seguir adelante.*

*A mis tíos Beatriz, Oscar, Benjamín, Miriam, Dilcia, Tatiana, Verena, Alcira, Gustavo, Maruja y
Alfredo por su constante apoyo.*

A mi abuela Elodia por su inextinguible amor y a mis primos por los buenos momentos vividos.

*A mis amigos Janna Pamela, Luis Eduardo, Oscar Eduardo, William Gilberto, Elkin Horacio,
Tina, y Nicolás por hacer mi experiencia en esta ciudad algo inolvidable.*

*A los maestros Franklin Sepulveda, Martha Menjura y Tamás Gábor Csapó por compartir su
conocimiento y permitir el desarrollo de este trabajo.*

A todas las personas que tuve la oportunidad de conocer en este camino, gracias.

Carlos Alfredo Lozano Herrera

A Dios por ser la luz de mis días.

A mis padres por su apoyo y entrega.

A mis amigos por su solidaridad y cariño.

Al profesor Franklin Alexander Sepúlveda por ser nuestro guía y mentor en este proyecto, a la profesora Martha Menjura por su aporte a nuestro proceso de investigación.

A todos aquellos que por algún motivo formaron parte de mi vida en estos años, gracias a cada uno por su aporte.

Janna Pamela Gómez Reyes

Contenido

	Pág.
Introducción	15
1. Planteamiento y definición del problema	16
1.1. Objetivos	19
1.1.1. Objetivo general.....	19
1.1.2. Objetivos específicos	19
2. Proceso de adquisición de datos	19
2.1. Escáner ultrasonido Pi 7.5 MHZ.....	20
2.2. Software Seemore	21
2.3. EdgeTrak.....	22
2.4. Base de datos imágenes de ultrasonido.....	23
2.4.1. Creación de frase ajustada a los fonemas del castellano de Colombia	23
2.4.2. Adquisición imágenes de ultrasonido	25
2.5. Ruido <i>Speckle</i>	26
3. Método automático para la detección del contorno de la lengua.....	27
3.1. Filtros de difusión	27
3.1.1. Difusión anisotrópica	27
3.1.2. Filtro de difusión anisotrópica reductor de speckle (DsFsrاد)	29
3.2 Umbralización de una imagen	30

3.3. Dilatación de una imagen.....	31
3.4. Detector de bordes Canny	32
3.5 Métodos de evaluación cuantitativa.....	34
4. Resultados.....	35
4.1. Extracción de parámetros.....	35
4.2. Reducción ruido <i>Speckle</i>	36
4.3 Modelado	41
4.4. <i>Edgetrak</i> y MSD	42
4.5. Tiempo de ejecución del código (python)	45
5. Conclusiones	47
6. Recomendaciones	48
7. Trabajo Futuro	49
Citas Bibliográficas.....	50
Referencias Bibliográficas	56
Apéndices.....	58

Lista de Figuras

	Pág.
<i>Figura 1.</i> Escáner ultrasonido patología del habla PI 7.5 MHz [12]	21
<i>Figura 2.</i> Pantalla del monitor del ordenador usando <i>Seemore</i>	22
<i>Figura 3.</i> Constitución del tracto vocal [29].....	24
<i>Figura 4.</i> Histograma transcripción fonética	25
<i>Figura 5.</i> Imagen original con contorno extraído para cada hablante.	36
<i>Figura 6.</i> Porcentaje de SNR mejorado.....	37
<i>Figura 7.</i> Porcentaje promedio de SNR mejorado.....	38
<i>Figura 8.</i> Imagen original e imagen después de aplicar DsFsrad hablante 1.	38
<i>Figura 9.</i> Imagen original e imagen después de aplicar DsFsrad hablante 2.	39
<i>Figura 10.</i> Imagen original e imagen después de aplicar DsFsrad hablante 3.	39
<i>Figura 11.</i> Imagen original e imagen después de aplicar DsFsrad hablante 4.	39
<i>Figura 12.</i> Imagen original e imagen después de aplicar DsFsrad hablante 5.	40
<i>Figura 13.</i> Imágenes resultado de la segmentación.....	41
<i>Figura 14.</i> Regresión polinomial.....	42
<i>Figura 15.</i> Imagen original con 3 contornos hablante 1.....	43
<i>Figura 16.</i> Valor MSD promedio para los 5 hablantes.....	45

Lista de Tablas

	Pág.
Tabla 1. <i>Información hablantes</i>	26
Tabla 2 <i>Características señal a ruido hablante 1</i>	40
Tabla 3 <i>Valor MSD hablante 1</i>	44
Tabla 4 <i>Tiempo de ejecución algoritmo implementado.</i>	46

Lista de Apéndices

	Pág.
Apéndice A – Carta consentimiento hablantes	58
Apéndice B – Base de datos imágenes de ultrasonido.....	58
Apéndice C – Tablas valores SNR 5 hablantes	58
Apéndice D – Imágenes con contornos extraídos por EdgeTrak, de forma manual e implementando algoritmo presentado en este proyecto.....	63
Apéndice E – Tablas con valores MSD para cada uno de los 5 hablantes.	66

Resumen

Título: Implementación y análisis de un método automático de detección del contorno superior de la lengua en secuencias de imágenes de ultrasonido *

AUTORES: CARLOS ALFREDO LOZANO HERRERA
JANNA PAMELA GÓMEZ REYES **.

Palabras Clave: imágenes de ultrasonido, ruido *speckle*, *EdgeTrak*, detección de contorno, difusión anisotrópica, operador Canny.

Descripción:

La lengua es un órgano fundamental en la producción del habla, sin embargo, la ubicación, forma y la naturaleza de la misma han dificultado el proceso de medición de su dinámica. Existen varias técnicas de medición del movimiento de la lengua, pero estas técnicas son usualmente invasivas y costosas. Por su parte, el ultrasonido permite la visualización del contorno superior de la lengua, es un método no-invasivo y menos costoso comparado con otros.

En este trabajo se implementa un método automático de extracción del contorno superior de la lengua, el cual es producto de la combinación de varios métodos desarrollados en el estado del arte. En contraste a otros métodos, el presentado en este trabajo no requiere de inicialización manual. El procedimiento consiste, primero que todo, en una fase de suavizado y reducción de ruido *speckle* haciendo uso de un filtro de difusión anisotrópica *DsFsradi*; luego se realiza una segmentación a la imagen obtenida; después, se obtiene el contorno mediante el uso del operador *Canny*; y finalmente, el contorno extraído se representa mediante unos pocos coeficientes obtenidos por regresión polinomial.

Los resultados se compararon con el software *EdgeTrak* sobre un conjunto de registros tomados con el dispositivo de ultrasonido portátil. Con los registros se formó una base de datos compuesta por 24 videos de 17 segundos de duración en promedio para 5 hablantes que reportan no poseer alguna alteración del habla. De estos registros se extrajeron un total de 2400 imágenes con resolución de 992x699 píxeles. La frase pronunciada se seleccionó por recomendación de una lingüista. El dispositivo corresponde a una sonda de ultrasonido USB PI 7.5MHz de Interson.

* Proyecto de grado

** Facultad de Ingenierías Físico-mecánicas. Escuela de Ingenierías Eléctrica, Electrónica y de Telecomunicaciones.
Director: Franklin Alexander Sepulveda Sepulveda, Dr.

Abstract

Title: Implementation and analysis of an automatic method for detection of the upper contour of the tongue in sequences of ultrasound images.*

Authors: Carlos Alfredo Lozano Herrera
Janna Pamela Gómez Reyes**

Key words: ultrasound images, speckle noise, EdgeTrak, contour detection, anisotropic diffusion, Canny operator.

Description:

The tongue is a fundamental part in speech production, however, the location, form and the nature of this organ have made difficult the process of measurement of its dynamics. There are several techniques for measuring tongue movement, but these techniques are usually invasive and costly. In contrast, ultrasound allows the visualization of the upper contour of the tongue, is a non-invasive method and less expensive compared to others.

In this work, an automatic method of extraction of the upper contour of the tongue is implemented, which is the product of the combination of several methods developed in the state of the art. In contrast to other methods, the one presented in this work does not require manual initialization. The procedure consists, first, in a smoothing and speckle noise reduction phase making use of an anisotropic diffusion filter DsFsrاد; then a segmentation is done to the obtained image; later the contour is obtained by using the Canny operator; and finally, the extracted contour is represented by a few coefficients obtained by polynomial regression.

Results were compared with the EdgeTrak software on a set of records taken with the portable ultrasound device. With the records, a database was formed consisting of 24 videos of average length 17 seconds for 5 speakers who report not having any alteration of the speech. From these records, a total of 2400 images with a resolution of 992x699 pixels were extracted. Pronounced phrase was selected on the advice of a linguist. The device corresponds to an Interson PI 7.5MHz USB ultrasound probe.

* Project of grade

** Physico-mechanical engineering faculty. school of electrical, electronics and telecommunications engineering. Supervisor: Franklin Alexander Sepulveda Sepulveda Dr.

Introducción

La forma de la lengua es parte fundamental de la producción y control del habla, y entender su movimiento nos permite comprender los trastornos del lenguaje, entre otras cosas. Este órgano posee gran flexibilidad, tiene la habilidad de adquirir diversas formas para producir los diferentes fonemas y es muy importante durante el proceso de articulación.

Diversas técnicas se han utilizado para estudiar la dinámica de la lengua y la forma del tracto vocal, entre las que se mencionan cineradiografía de rayos X [1], articulografía electromagnética [2] y videos de resonancia magnética [3]. En el caso de la cineradiografía, la exposición a altos niveles de radiación no permitió que su uso se extendiera, aun así, algunos estudios se realizaron usando este tipo de datos [4]. En contraste, el articulógrafo electromagnético no tiene problemas de radiación ionizante lo cual permite tomar muestras durante tiempos relativamente largos [5]. Este dispositivo rastrea sensores basados en bobinas inmersas en un campo magnético que rodea toda la zona de la cabeza. Estas bobinas se adhieren al órgano objetivo [6]. Aunque este dispositivo benefició a las áreas encargadas del estudio del habla; el incómodo proceso de adecuación de los receptores y coste del sistema limitan hoy en día el uso de este método. Adicionalmente el articulógrafo electromagnético no permite observar todo el tracto vocal, solo algunos puntos. Finalmente se tiene la técnica de videos de resonancia magnética, mediante la cual si se puede observar la forma del tracto vocal en su totalidad; sin embargo, debido a su alto costo, la generación de ruido de los equipos utilizados y la tasa de muestreo baja, han hecho que esta técnica sea usada sólo con fines científicos y experimentales.

A modo de método alternativo se tiene el ultrasonido, el cual es un método no-invasivo y notablemente menos costoso que los métodos mencionados anteriormente. Pero a raíz de que los dispositivos de ultrasonido pueden detectar sólo una parte del tracto vocal, se requieren métodos adicionales que permitan estimar aquellas otras zonas no mostradas. No obstante, luego de ser procesadas, las imágenes de ultrasonido podrían ser una parte constitutiva de un sistema de bio-realimentación articulatoria. En consecuencia, este trabajo se enfoca en la implementación de un método para la extracción del contorno de la lengua en imágenes de US.

Se aplica un conjunto de técnicas de procesamiento de imágenes [7, 8, 9, 10,11], con el fin de facilitar la detección del contorno de la parte superior de la lengua en secuencias de imágenes de ultrasonido en dos dimensiones, sin inicialización manual. Las imágenes procesadas han sido obtenidas a través de una sonda de ondas de ultrasonido [12].

Se propone realizar una comparación con el método *EdgeTrak* [13], desarrollado al interior de la escuela de salud oral de la Universidad de Maryland y el departamento de computación y ciencias informáticas de la Universidad de Delaware en los Estados Unidos de América. Debido a que es un método que ofrece calidad en los resultados obtenidos, el mismo se ha convertido en un referente para las investigaciones relacionadas con la producción y desarrollo del habla [7, 14]. Por tal motivo los resultados del presente trabajo serán comparados con este sistema.

1. Planteamiento y definición del problema

Mostrar la posición y el movimiento de los articuladores es considerado un medio efectivo para guiar al usuario en la producción correcta de fonemas. Esta tarea se puede llevar a cabo mediante

animaciones audiovisuales del tracto vocal [15]. Por otra parte, la realimentación articulatoria es un componente importante en el entrenamiento de la pronunciación, pero debe cumplir los siguientes requerimientos: 1) debe ser fácil de interpretar, 2) debe determinar si la señal de voz fue correctamente generada en términos tanto cuantitativos como cualitativos, 3) debe estar en capacidad de mostrar las acciones que el usuario debería seguir con el fin de mejorar la pronunciación. Los sistemas de bio-realimentación articulatoria cumplen estas tres condiciones y es especialmente importante en el cumplimiento de la tercera condición.

Una forma de obtener una realimentación visual del mecanismo del tracto vocal es mediante el uso de inversión articulatoria; sin embargo, aunque se ha desarrollado una considerable cantidad de métodos de inversión articulatoria [16], los mismos aún solo han sido probados para poblaciones de tamaño muy reducido [17], y por tanto se desconoce su real efectividad dado que este tipo de sistemas son notablemente afectados por la variabilidad entre-hablantes.

Con el fin de atacar el problema de alta variabilidad entre-hablantes de la inversión articulatoria, se pueden utilizar métodos que permitan obtener, mediante información proveída por instrumentos de medición, la posición de algunos articuladores del tracto vocal. Dentro de los aparatos que pueden utilizarse para el seguimiento del tracto vocal se pueden mencionar: video de rayos-X [18], rayos X de microcontactos [19], ultrasonido [20], articulógrafo electromagnético (EMA) [21] y videos de resonancia magnética [22]. Los sistemas de video de rayos-X y rayos X de microcontactos ya no se usan hoy en día por cuestiones de seguridad. Dentro de las posibilidades restantes, en el presente trabajo se propone el uso de ultrasonido por ser el más susceptible de masificarse debido a su reducido precio respecto a los sistemas EMA y de MRI.

En lo referente a la técnica de ultrasonido, al revisar trabajos recientes se observa que generalmente los métodos descritos usan modelos adaptivos y flexibles para facilitar la interpretación de imágenes, por lo general, estos modelos presentan características o parámetros ajustables que permiten controlar diferentes variables y elementos de la imagen, tales como el contorno, o la forma, por mencionar algunos, permitiendo obtener una aproximación más certera acerca de la imagen a estudiar. En particular, en [23] se expone un método que extrae y hace seguimiento de la superficie de la lengua de manera automática, el cual incluye el uso de contornos deformables. El usuario previamente entrega un modelo inicial del contorno para así ofrecer un marco de referencia en el proceso de detección del contorno de las siguientes imágenes. En contraste, en [7] se propone otro método, pero sin la necesidad de un marco de imagen inicial. Desde el punto de vista del criterio de suma de distancias, su desempeño es similar con los resultados entregados por el comúnmente difundido software *EdgeTrak*; pero a diferencia de este último, el propuesto en [7] no requiere patrones de referencia para dar inicio al algoritmo.

En el presente trabajo se implementa un método alternativo y ajustado al tipo de dispositivo de ultrasonido PI 7:5 MHz de Interson [12], el cual ha sido diseñado para propósitos de terapia del habla. El grupo CEMOS de la Universidad Industrial de Santander dispone de este aparato. Para ello se propone realizar un proceso de reducción de ruido *speckle*, el cual es típicamente encontrado en imágenes de ultrasonido. Luego han de aplicarse técnicas de segmentación que faciliten la extracción del contorno de la lengua mediante un proceso de detección de bordes. Después se realiza una regresión polinomial sobre el contorno extraído. Los resultados se comparan con aquellos obtenidos mediante métodos recientes (*EdgeTrak*) a través del uso de criterios de comparación expuestos en trabajos recientes, por ejemplo, en [24].

1.1. Objetivos

1.1.1. Objetivo general Implementar un método de extracción de contornos 2D sin inicialización manual, sobre secuencias de imágenes de ultrasonido del perímetro superior de la lengua.

1.1.2. Objetivos específicos

- Extraer parámetros de imágenes de ultrasonido que permitan definir el contorno de la lengua, así como reducir el efecto del ruido speckle presente en este tipo de imágenes.
- Realizar el modelado del contorno superior de la lengua a partir de los parámetros extraídos de las imágenes de ultrasonido.
- Comparar los resultados obtenidos luego de la implementación del método sin inicialización manual con aquellos obtenidos a partir de un sistema referencia de extracción del contorno de la lengua.

2. Proceso de adquisición de datos

El presente capítulo describe el proceso de formulación y toma de muestras de las imágenes de ultrasonido que serán posteriormente tratadas. Para ello se requiere acondicionar el transductor [12], una sonda intercostal de frecuencia 7.5 MHz con un ángulo de visión de 90° y de

profundidad ajustable. Esta sonda tiene una superficie curvada que la hace ideal para obtener imágenes sagitales de la lengua. Se describirán a continuación este dispositivo y demás elementos empleados para el fin mencionado.

2.1. Escáner ultrasonido Pi 7.5 MHZ

Las ondas de ultrasonido son producidas por un transductor, ver *Figura 1.*, el cual está formado por elementos activos hechos de materiales especiales de cristal cerámico llamados piezoeléctricos, esos materiales producen ondas sonoras cuando un campo eléctrico pasa a través de ellos, al igual que forma un campo eléctrico cuando recibe una onda sonora [25].

Al utilizar el transductor el ajuste de 90° permite normalmente una vista completa de la lengua de un hablante. Se debe tener en cuenta que el hueso hioides es un factor limitante en la dirección posterior, y la sombra de la mandíbula limita la vista en la dirección anterior, lo cual dificulta la visualización de la totalidad de la lengua.

Las aplicaciones del PI 7.5 MHz son principalmente para patologías del habla, sus especificaciones son [12]:

- Rango de profundidad: 3 cm – 10 cm.
- Frecuencia del pulso: 7.5 MHz, 12, 15, 24 MHz.
- Ángulo de escaneado: 90 grados.
- Cuadros por segundo: 12 fps – 15 fps.
- Tamaño: 17 cm y 3.5 cm.
- Peso: 120 gramos.
- Zona Focal: 0.5 cm – 4 cm.

- Punto Focal: 2 cm.
- Cable: USB 2.0, 2 m (5 m opt.)
- Windows 7,8.



Figura 1. Escáner ultrasonido patología del habla PI 7.5 MHz [12]

2.2. Software Seemore

El software asociado al equipo “Escáner ultrasonido USB patología del habla PI 7.5 MHz” de la empresa Interson es *Seemore*, el cual es compatible con el sistema operativo Windows, este permite visualizar en tiempo real la lectura que realiza el escáner, posibilita la adquisición de imágenes y/o video del movimiento y posición de la lengua y almacena la información del paciente.

Al abrir la ventana en el lado izquierdo se encuentra la imagen de ultrasonido e información sobre el paciente, en la parte derecha de la pantalla se encuentra una ventana de interfaz del usuario que contiene varias pestañas que permiten establecer preferencias a realizar en el examen de ultrasonido, lo anterior se puede observar en la *Figura 2*.

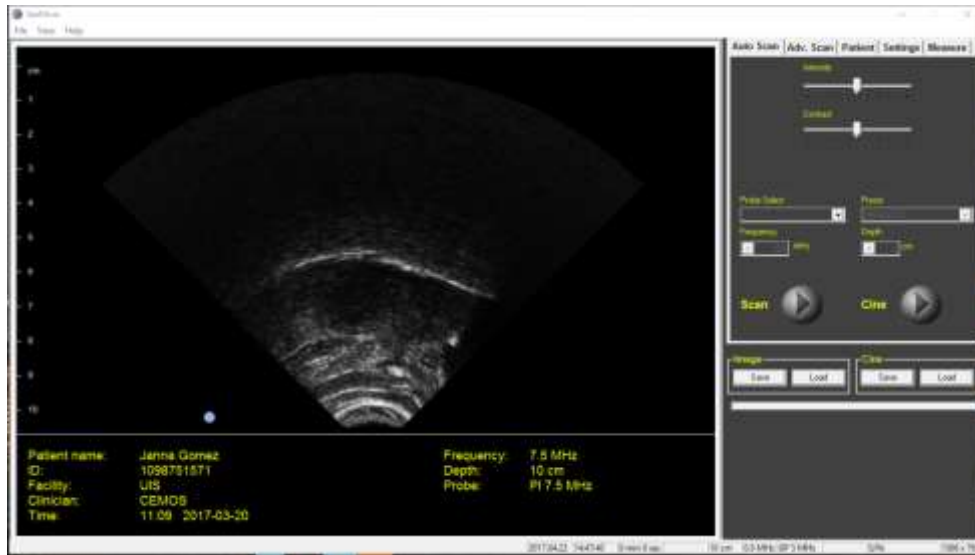


Figura 2. Pantalla del monitor del ordenador usando *Seemore*.

2.3. EdgeTrak

EdgeTrak es un sistema que puede rastrear la superficie de la lengua a través de una secuencia de imágenes de ultrasonido 2D. Dado que está siendo utilizado actualmente por científicos de varias instituciones, la retroalimentación de ellos y la validación indican que el sistema es eficiente y preciso para la investigación del habla y las aplicaciones relacionadas, por esta razón se decide utilizar este método como punto de comparación.

Las imágenes de ultrasonido se encuentran en escala de grises [26] así que pueden ser tratadas por el sistema mencionado. Este extrae el contorno abierto del borde inferior de una imagen de degradado a blanco y negro. El contorno sirve como inicialización para la imagen siguiente en la secuencia, la ubicación del contorno es optimizada a través de programación dinámica utilizando un nuevo modelo de contorno deformable basado en modelos de contorno activos (*snakes*) [27].

Para tratar el ruido y los bordes no relacionados en las imágenes, la información de la región se aplica a contornos abiertos y la intensidad en las regiones locales se incorpora en los gradientes del borde como energía externa.

La definición de energía para *Snakes* es:

$$E_{Total} = \alpha E_{int} + \beta E_{ext} \quad (1)$$

Donde E_{int} es la energía interna, E_{ext} la energía externa, α y β son parámetros de ajuste. E_{int} controla la forma del contorno y está relacionada estrictamente a las propiedades geométricas del contorno. E_{ext} fija el contorno a la imagen y define las características de interés de la imagen. La energía externa es usualmente definida como el negativo del gradiente de la imagen.

2.4. Base de datos imágenes de ultrasonido

Con la necesidad planteada del estudio de imágenes de ultrasonido que muestren la posición de la lengua y debido a que no se encuentra una base de datos de fácil acceso, se genera un archivo que reúne las imágenes adquiridas bajo los parámetros que se exponen en la Sección 2.4.1., necesarios para su posterior tratamiento.

2.4.1. Creación de frase ajustada a los fonemas del castellano de Colombia El tracto vocal está constituido por la cavidad oral, nasal, la faringe y la laringe. Dentro de estas cavidades, ver *Figura 3*, se encuentran los órganos de articulación [28].

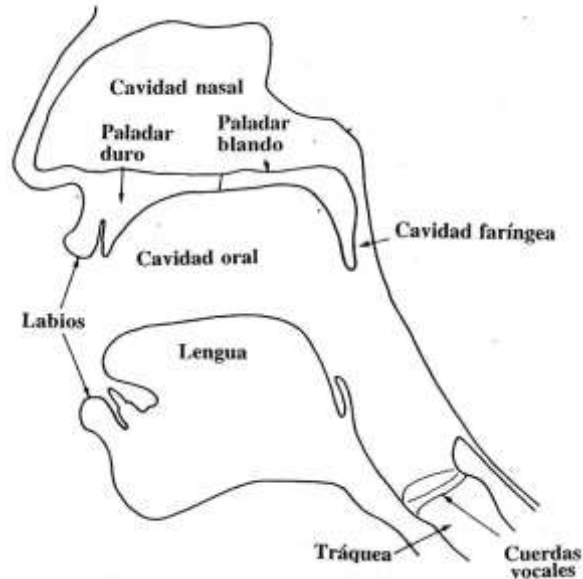


Figura 3. Constitución del tracto vocal [29].

La lengua está dentro de la clasificación de los órganos articulatorios activos lo cual quiere decir, que adopta diferentes posiciones para ayudar en las diversas configuraciones que presenta el tracto vocal como filtro acústico para el sonido que produce la laringe.

El rasgo de articulación lateral designa la forma y posicionamiento de la lengua respecto al tracto vocal, los segmentos de modo de articulación lateral se producen cuando el centro de la lengua se eleva hasta tocar la parte superior de la boca, y por tanto se bloquea el paso central del aire y se le fuerza a un flujo lateral a lo largo de los lados de la lengua [30].

Con el fin de fomentar las diferentes posiciones que adquiere la lengua al pronunciar fonemas con rasgos de modo articulación lateral, se crea una frase balanceada a partir de consonantes oclusivas y consonantes palatales que están descritas por la posición del cuerpo de la lengua. La frase construida es la siguiente:

“Al niño Nicolás Calle le gusta correr durante la noche en la jungla, una tarde se encontró con el loro José y un gato saltó y tiró un coco”

Se realiza una transcripción fonética a la frase balanceada con el recurso en línea [31], los resultados obtenidos son:

/al 'niño niko 'las 'kafe le 'ɣusta ko 'rer ðu 'rante la 'notʃe en la, 'xunɣla 'una 'tarðe se eŋkon 'tro kon el 'loro xo 'se i uŋ 'gato sal 'to i ti 'ro un 'koko/

Se construye un histograma, ver *Figura 4*, el cual expone la frecuencia de cada fonema dentro de la frase balanceada, los picos están presentes en las vocales abiertas a, e, o.

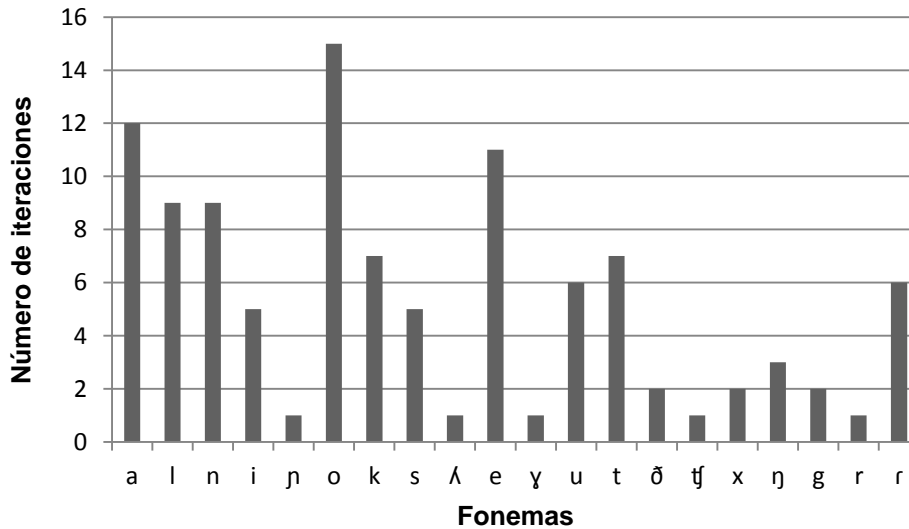


Figura 4. Histograma transcripción fonética

2.4.2. Adquisición imágenes de ultrasonido Para las muestras se ubica el escáner de ultrasonido en contacto con la parte inferior de la mandíbula, se registran los datos personales del hablante en el software *Seemore* [32], luego se procede a ajustar la profundidad del escáner a 10 cm y los *frames* a 520, todo esto para poder almacenar por completo en un video la pronunciación de la frase balanceada la cual tiene una duración promedio de 17 segundos.

Cada uno de los cinco hablantes activa el botón que da inicio y fin a la grabación de las muestras. El video se almacena en formato .cine, seguidamente desde el mismo software se procede a seccionar el video y extraer 100 imágenes secuenciales de este en formato .JPG. A partir de estas imágenes se seleccionan de forma aleatoria entre 4 y 10 secuencias de 5 imágenes, esto para cada uno de los hablantes. Los hablantes se clasifican de la siguiente forma:

Tabla 1.
Información hablantes

Nombre	Género	Edad	Región geográfica de origen	Número de videos	Total de imágenes
Hablante 1	F	23	Norte	5	500
Hablante 2	M	22	Oriente	6	600
Hablante 3	F	20	Oriente	4	400
Hablante 4	F	19	Oriente	4	400
Hablante 5	F	20	Oriente	5	500
Total					2400

2.5. Ruido *Speckle*

El ruido *speckle* [33] se presenta frecuentemente en las imágenes de ultrasonido. Este ruido degrada significativamente la calidad de la imagen que se construye a partir de las ondas de altas frecuencias, dificulta la discriminación de detalles finos en las imágenes durante un examen de diagnóstico y también afecta el procesamiento (segmentación y detección de bordes).

El fenómeno de degradación de una imagen de ultrasonido responsable de generar el ruido *speckle* es el fenómeno de difusión, por el cual los objetos que tienen tamaño similar a la longitud de onda de ultrasonido oscilan emitiendo en todas las direcciones la onda incidente. Esta transmisión es la responsable de generar el ruido [33].

Al hablar del ruido *speckle* se refiere a la descripción del patrón granular presente en las imágenes de ultrasonido, es considerado un ruido multiplicativo, y para describir sus características se usa un análisis estocástico.

3. Método automático para la detección del contorno de la lengua

Para la detección del contorno se aplican un conjunto de técnicas de procesamiento de imágenes entre las cuales está el filtrado, la segmentación y detección de bordes. En esta sección se explican cada una de las técnicas empleadas en el método aplicado en el presente proyecto.

3.1. Filtros de difusión

Los filtros de difusión eliminan el ruido de una imagen a través de la solución de una ecuación diferencial parcial (PDE). El suavizado de una imagen depende de los bordes y sus direcciones. La difusión anisotrópica es una técnica no lineal que ha resultado ser muy eficiente para reducir los niveles del ruido en una imagen y mejorar el contraste simultáneamente [34]. Esta técnica realiza el suavizado en regiones homogéneas de la imagen, sin requerir información alguna de la energía espectral de la imagen, y reteniendo los bordes.

3.1.1. Difusión anisotrópica. Perona y Malik [35] propusieron la siguiente PDE para suavizar una imagen en el dominio continuo:

$$\begin{cases} \frac{\partial I}{\partial t} = \text{div}[c(|\nabla I|) \cdot \nabla I] \\ I(t=0) = I_0 \end{cases} \quad (2)$$

Donde ∇ es el operador del gradiente, div el operador de divergencia, $\|\cdot\|$ denota la magnitud, $c(x)$ el coeficiente de difusión, e I_0 la imagen original. Los autores [35] sugirieron dos coeficientes de difusión:

$$c(x) = \frac{1}{1 + \left(\frac{x}{k}\right)^2} \quad (3)$$

Y

$$c(x) = e^{-\left(\frac{x}{k}\right)^2} \quad (4)$$

Donde k es un parámetro de la magnitud del borde.

En el método de difusión anisotrópica, la magnitud del gradiente detecta un borde o la frontera de una imagen como un paso discontinuo en intensidad. Si $|\nabla I| \gg k$, entonces $c(|\nabla I|) \rightarrow 0$, y se tiene un filtro pasa-todo. Si, por el contrario, $|\nabla I| \ll k$, entonces $c(|\nabla I|) \rightarrow 1$, y se tendría difusión isotrópica (filtro Gaussiano) [36].

Una formulación discreta de la ecuación 2 está dada por:

$$I_s^{t+\Delta t} = I_s^t + \frac{\Delta t}{|\bar{\eta}_s|} \sum_{p \in \bar{\eta}_s} c(|\nabla I_{s,p}^t|) \nabla I_{s,p}^t \quad (5)$$

Donde I_s^t es la imagen mostrada en el dominio discreto, s denota la posición del pixel en una grilla discreta de dos dimensiones (2-D), y Δt es el tamaño del paso de la variable tiempo, $\bar{\eta}_s$ representa la vecindad espacial del pixel s , $|\bar{\eta}_s|$ es el número de pixeles en la ventana, y

$$\nabla I_{s,p}^t = \nabla I_p^t - \nabla I_s^t, \forall p \in \bar{\eta}_s \quad (6)$$

Dentro de las ventajas de la difusión anisotrópica se pueden destacar el suavizado intra-región, y la preservación de los bordes. La difusión anisotrópica se desempeña bien en imágenes corruptas por ruido aditivo [37].

El artículo [36], resalta una ecuación diferencial parcial PDE enfocada a la reducción de ruido *speckle*, llamada DsFsrad, Sección 3.1.2. El enfoque de reducción de ruido basado en la PDE permite la generación de una imagen escala-espacio independientemente al tamaño de la ventana del filtro. DsFsrad no simplemente preserva los bordes, también los mejora, prohibiendo la difusión a través de los bordes y permitiendo la difusión alrededor de estos. DsFsrad es adaptativo y no altera el desempeño en regiones homogéneas o en regiones cercanas a los bordes. La nueva técnica de difusión está basada en el mismo enfoque de MMSE (*minimum mean square error*) al igual que los filtros de Lee y Frost [38, 39, 40, 41]. De esta forma DsFsrad es una extensión sensible a bordes de un filtro adaptativo convencional de ruido *speckle*, de la misma forma en que el filtro de difusión anisotrópica expuesto por Perona y Malik [35] es una extensión sensible a bordes de un filtro promedio. En este sentido se extienden las aplicaciones de la difusión anisotrópica a radares y ultrasonido médico, en los cuales el ruido multiplicativo espacialmente correlacionado está presente.

3.1.2. Filtro de difusión anisotrópica reductor de speckle (DsFsrad) Para la reducción de ruido *speckle* se utiliza un método de difusión anisotrópica para suavizar el moteado presente en las imágenes US. Dada una imagen de intensidad $I_o(x, y)$ con potencia finita y sin valores cero sobre el soporte de la imagen Ω , la imagen de salida $I(x, y; t)$ se desarrolla de acuerdo con la siguiente PDE propuesta por [36]:

$$\begin{cases} \frac{\partial I(x, y; t)}{\partial t} = \text{div}[c(q)\nabla I(x, y; t)] \\ I(x, y; 0) = I_o(x, y), \frac{\partial I(x, y; t)}{\partial t} |_{\partial\Omega} = 0 \end{cases} \quad (7)$$

Donde $\partial\Omega$ denota el borde de Ω , \vec{n} , es la exterior normal a $\partial\Omega$, y

$$c(q) = \frac{1}{1 + \frac{[q^2(x,y;t) - q_0^2(t)]}{[q_0^2(t)(1+q_0^2(t))]} \quad (8)$$

$$q(x, y; t) = \sqrt{\frac{\left(\frac{1}{2}\right)\left(\frac{|\nabla I|}{I}\right)^2 - \left(\frac{1}{4}\right)\left(\frac{|\nabla^2 I|}{I}\right)^2}{\left[1 + \left(\frac{1}{4}\right)\left(\frac{|\nabla^2 I|}{I}\right)^2\right]^2}} \quad (9)$$

$$q_0(t) = \frac{\sqrt{\text{var}[z(t)]}}{\bar{z}(t)} \quad (10)$$

$q_0(t)$ es la función escala *speckle* y $\text{var}[z(t)]$ y $\bar{z}(t)$ son la varianza de intensidad y la media sobre un área homogénea en t , respectivamente.

En el DsFsrاد propuesto por [39] e implementado en este proyecto, el coeficiente de variación instantáneo $q(x, y; t)$ sirve como detector de borde en imágenes moteadas. La función exhibe altos valores en los bordes o en características de alto contraste y produce valores bajos en regiones homogéneas.

La modificación refleja el soporte de la difusión isotrópica en regiones homogéneas de la imagen donde $q(x, y; t)$ fluctúa alrededor de $q_0(t)$. Similar al parámetro k en (3) y (4), la función de escala *speckle* $q_0(t)$ controla efectivamente la cantidad de suavizado aplicada a la imagen por DsFsrاد.

3.2 Umbralización de una imagen

La umbralización es una técnica de segmentación ampliamente utilizada, este método se utiliza cuando existe una clara diferencia entre los objetos a extraer respecto del fondo de la escena.

Debido a que los principios del método son la similitud entre píxeles pertenecientes a un objeto y sus diferencias respecto al resto de la imagen [8], la imagen a tratar debe tener una escena que posea un fondo uniforme.

Se ha derivado un método para seleccionar automáticamente un umbral a partir de un histograma de niveles de gris desde el punto de vista del análisis discriminante. Un umbral óptimo (o conjunto de umbrales) se selecciona por el criterio discriminante; es decir, maximizando la medida discriminante, q , (o la medida de separabilidad de las clases resultantes en niveles de grises).

El algoritmo aplicado, algoritmo de Otsu [9] trata de encontrar un valor umbral (z) que minimice la varianza dentro de clase, así lo muestra la siguiente relación:

$$\sigma_w^2(z) = q_1(z)\sigma_1^2(z) + q_2(z)\sigma_2^2(z) \quad (11)$$

$$q_1(z) = \sum_{i=1}^t P(i) \quad \& \quad q_2(z) = \sum_{i=t+1}^I P(i) \quad (12)$$

$$\mu_1(z) = \sum_{i=1}^t \frac{iP(i)}{q_1(z)} \quad \& \quad \mu_2(z) = \sum_{i=z+1}^I \frac{iP(i)}{q_2(z)} \quad (13)$$

$$\sigma_1^2(z) = \sum_{i=1}^z [i - \mu_1(z)]^2 \frac{P(i)}{q_1(z)} \quad \& \quad \sigma_2^2(z) = \sum_{i=z+1}^I [i - \mu_2(z)]^2 \frac{P(i)}{q_2(z)} \quad (14)$$

Donde, $q_1(z)$ y $q_2(z)$ son segmentos definidos a partir de un umbral z , $\mu_1(z)$ y $\mu_2(z)$ son las intensidades medias propias de cada segmento, $\sigma_1^2(z)$ y $\sigma_2^2(z)$ son las varianzas dentro de cada segmento, $\sigma_w^2(z)$ es la varianza entre segmentos.

Al aplicar un umbral, z , la imagen en escala de grises, $f(x,y)$, quedará binarizada; etiquetando con '1' los pixeles correspondientes al objeto y con '0' aquellos que son del fondo. En palabras simples, el método Otsu calcula automáticamente un valor umbral del histograma de imagen para una imagen bimodal [9].

3.3. Dilatación de una imagen

A partir de la imagen binaria se puede observar que existen pixeles que deberían pertenecer al objeto propio de la imagen pero que por diferentes motivos no fueron identificados. También

aparecen píxeles que se sabe que no deberían haber aparecido y por tanto se hace necesario incorporar los píxeles al objeto al que pertenecen. Para esto se tiene en cuenta el siguiente concepto: si alguno de los píxeles vecinos al píxel en estudio pertenece al objeto entonces el píxel de estudio también pertenece al objeto, esta operación lógica se conoce como dilatación.

El resultado de la dilatación es el conjunto de elementos tal que al menos algún elemento del conjunto estructurante B está contenido en el conjunto X , cuando B se desplaza sobre el conjunto X [10]:

$$X \oplus B = \{x | X \cap B_x \neq \emptyset\} \quad (15)$$

B_x se define como la traslación de B por x .

Esta operación representa un crecimiento progresivo del conjunto X . Al pasar el elemento estructurante dentro del conjunto, éste no se modificará. Sin embargo, en la frontera del conjunto X , al desplazar a B , el conjunto resultado se expansionará. La aplicación iterada de este operador haría degradar la imagen, haciendo coincidir el conjunto dilatado con la imagen. La dilatación es una transformación extensiva:

$$X \subseteq X \oplus B \quad (16)$$

En el algoritmo, un elemento de píxel es '1' si al menos un píxel debajo del núcleo es '1'. Por lo tanto, aumenta la región blanca en la imagen o el tamaño de los objetos de primer plano.

3.4. Detector de bordes Canny

El detector de bordes Canny es un operador de detección de bordes que utiliza un algoritmo de múltiples etapas para detectar una amplia gama de bordes en las imágenes, es conocido como el detector óptimo de bordes.

Fue desarrollado por John F. Canny en 1986 [11] basándose en unos criterios durante el desarrollo de su trabajo, el primer criterio fue la baja tasa de error. Es importante que los bordes que ocurren en las imágenes no se pierdan y que no haya respuestas a los no-bordes. El segundo criterio es que los puntos del borde estén bien localizados, en otras palabras, la distancia entre los píxeles de borde, tal como se encuentran por el detector y el borde real, debe ser mínima. El tercer criterio es tener una sola respuesta a un solo borde, este surgió debido a que los dos primeros criterios no fueron lo suficientemente sustanciales como para eliminar completamente la posibilidad de respuestas múltiples a un solo borde.

El detector de bordes contiene una secuencia de etapas [42], y la primera es suavizar la imagen para eliminar el ruido. Luego encuentra el gradiente de la imagen para resaltar las regiones con altas derivadas espaciales.

$$Edge_Gradient (G) = \sqrt{G_x^2 + G_y^2} \quad (17)$$

$$Angle (\theta) = \tan^{-1}\left(\frac{G_y}{G_x}\right) \quad (18)$$

G_x es la primera derivada en dirección vertical, G_y es la primera derivada en dirección horizontal.

El algoritmo sigue pistas a lo largo de estas regiones y suprime cualquier píxel que no esté al máximo (supresión no máxima). La matriz de gradiente se reduce ahora más por histéresis. La histéresis se utiliza para rastrear a lo largo de los píxeles restantes que no se han suprimido. La histéresis utiliza dos umbrales y si la magnitud está por debajo del primer umbral, se pone a cero (se hace un “no borde”). Si la magnitud está por encima del umbral alto, se hace un borde. Y si la magnitud está entre los 2 umbrales, entonces se pone a cero a menos que haya una trayectoria de este píxel con un píxel de gradiente sobre el umbral alto.

3.5 Métodos de evaluación cuantitativa

- Relación señal a ruido (SNR): Una señal se define como toda información significativa y representativa para construir el mensaje que se desea comunicar, el ruido es aquel dato o conjunto de datos que acompañan a la señal dificultando su transmisión y almacenamiento.

La relación señal a ruido es una de tantas formas de evaluación cuantitativa que da una medida de cuanto ha corrompido el ruido a una señal, para el caso de las imágenes de ultrasonido tratadas en este proyecto, se utiliza el método propuesto por Burckhardt [43], el cual halla una relación entre la media \bar{x} y la desviación estándar σ_x en regiones homogéneas.

$$SNR = \frac{\bar{x}}{\sigma_x} \quad (19)$$

- Relación pico señal a ruido (PSNR): La relación pico señal a ruido está definida como la relación entre la máxima energía de una señal y el ruido que afecta a su representación verídica expresada en escala logarítmica.

Para poder expresar esta medida cuantitativa de la calidad del procesamiento de la imagen, se halla la relación de Error Cuadrático Medio (MSE) [37], este es utilizado para determinar el cambio entre la imagen procesada f y la imagen original g ; $M \times N$ es el tamaño de la imagen y g^2_{max} es la intensidad máxima en la imagen original.

$$PSNR = -10 \log \frac{MSE}{g^2_{max}} \quad (20)$$

$$MSE = \frac{1}{MN} \sum_{i=1}^M \sum_{j=1}^N (g(i,j) - f(i,j))^2 \quad (21)$$

- Media de la suma de las distancias (MSD): Para realizar la evaluación del contorno extraído a través del método sin inicialización manual implementado en el presente trabajo se

utiliza la media de la suma de las distancias (MSD) [44], la cual compara cada pixel de una curva dada con el pixel más cercano en la curva con la cual es comparado. La comparación se realiza respecto a EdgeTrak y respecto al trazado de referencia obtenido manualmente. La Ecuación 22, provee una evaluación en pixeles de la media de la distancia desde un contorno U a un contorno V. Los contornos se definen como un conjunto de coordenadas (x, y): U es un conjunto de puntos en dos dimensiones (u_1, \dots, u_n) y V es un conjunto de puntos en dos dimensiones (v_1, \dots, v_m) .

$$MSD(U, V) = \frac{1}{m+n} (\sum_{i=1}^m \min_j |v_i - u_j| + \sum_{i=1}^n \min_j |u_i - v_j|) \quad (22)$$

4. Resultados

4.1. Extracción de parámetros

Se extrae el contorno para cada una de las imágenes estudiadas a través del método implementado en este proyecto, basado en un conjunto de técnicas de procesamiento de imágenes [7, 8, 9, 10, 11]. En la *Figura 5* se puede observar cómo se resalta con una línea de color cian el contorno resultante al aplicar el método implementado en el presente proyecto para cada uno de los 5 hablantes.

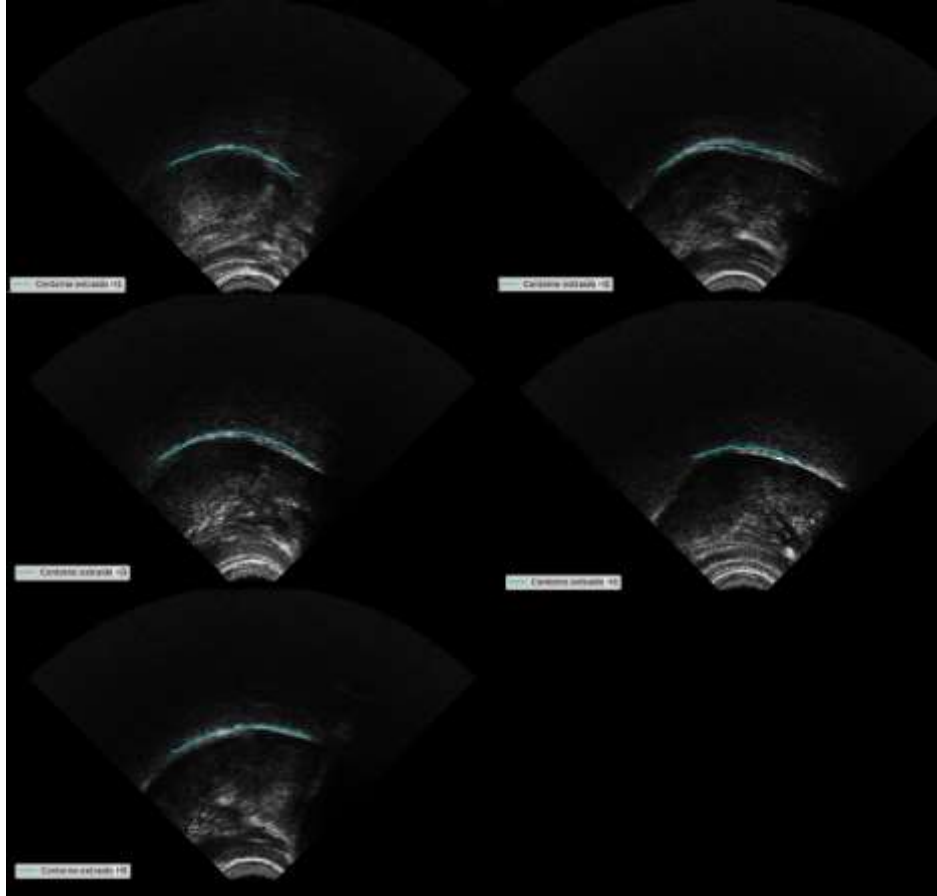


Figura 5. Imagen original con contorno extraído para cada hablante.

4.2. Reducción ruido *Speckle*

Las imágenes de ultrasonido presentan baja relación señal a ruido [24]. A modo de ejemplo ver los datos de la primera columna de la Tabla 2. Las imágenes de US cuentan con ruido *speckle*, el cual oculta las estructuras propias de la imagen. Con el fin de contrarrestar este efecto disminuyendo el ruido presente en estas se aplica un filtro de difusión anisotrópica, ver sección 3.1. Las tablas con los resultados de los 5 hablantes se encuentran en el ANEXO C.

La Figura 6, muestra el rango de mejora que se obtiene en este proyecto, para las 3 secuencias conformadas por 5 imágenes cada una, totalizando 15 imágenes para cada uno de los hablantes.

Se observa que se obtuvo una mejora de hasta el 32% en la SNR de la imagen filtrada con respecto a la SNR de la imagen original, lo cual se asemeja a los resultados obtenidos por métodos similares en la literatura [37].

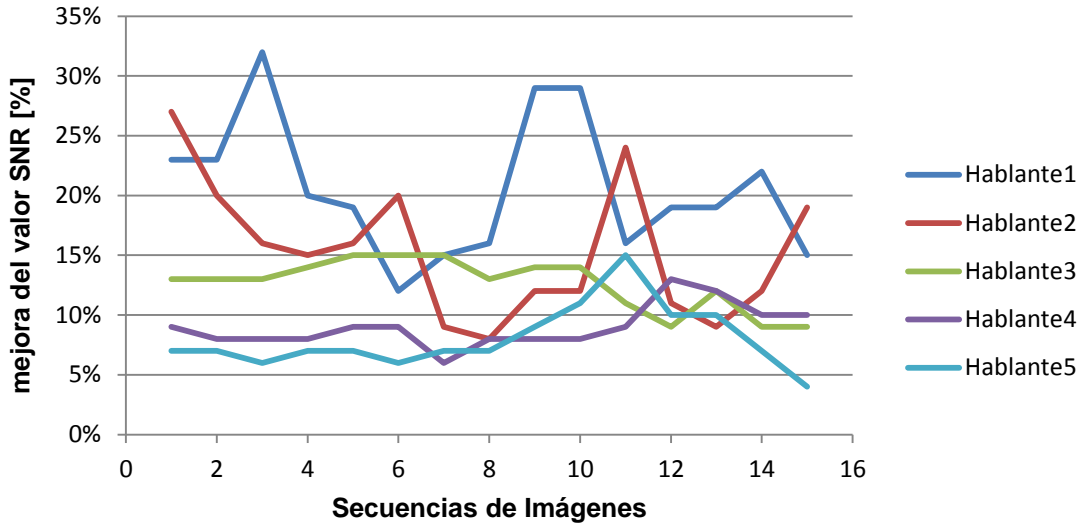


Figura 6. Porcentaje de SNR mejorado.

La Figura 7. Muestra el porcentaje promedio de mejora de SNR alcanzado por cada uno de los 5 hablantes. También permite observar el porcentaje promedio total de mejora el cual equivale a un 13%, este resultado se obtiene de las 75 imágenes tratadas en el presente proyecto.

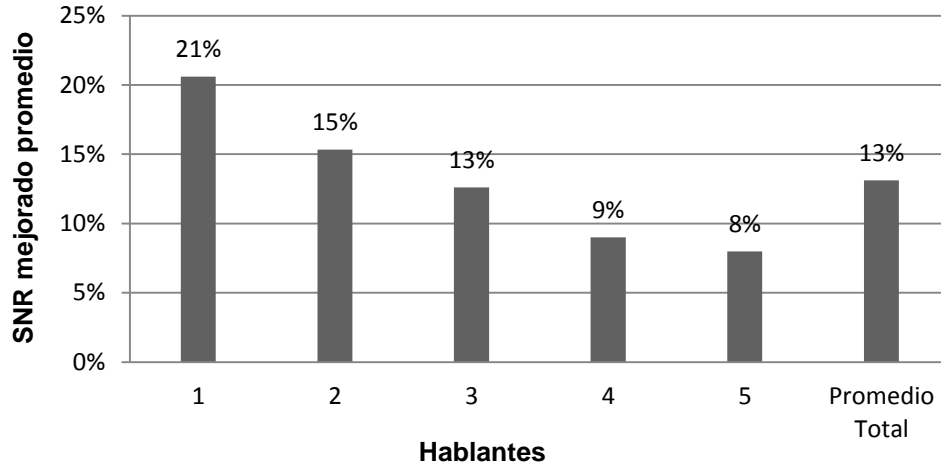


Figura 7. Porcentaje promedio de SNR mejorado.

En las Figuras 8, 9, 10, 11, y 12, se observa la región de interés de la imagen original y la imagen después de aplicar el filtro DsFsrad, el cual reduce el ruido granular presente en la imagen de ultrasonido.



Figura 8. Imagen original e imagen después de aplicar DsFsrad hablante 1.



Figura 9. Imagen original e imagen después de aplicar DsFsrاد hablante 2.



Figura 10. Imagen original e imagen después de aplicar DsFsrاد hablante 3.



Figura 11. Imagen original e imagen después de aplicar DsFsrاد hablante 4.



Figura 12. Imagen original e imagen después de aplicar DsFsrاد hablante 5.

Se registran los valores obtenidos a partir de los métodos cuantitativos de evaluación para cada una de las imágenes pertenecientes a las tres secuencias de cada hablante, tal como expone la Tabla 2 para el caso del hablante 1; y de este modo estimar el incremento del parámetro SNR luego de ser filtrada la imagen original a través de DsFsrاد.

Tabla 2

Características señal a ruido hablante 1

Secuencias	Imagen Muestra	SNR	SNR	SNR	MSE	PSNR	
		(original)	(Despeckle) N= 150 Lambda =0.025	Mejorado	(error cuadrático medio)		
Hablaante 1	Secuencia_1	H1S1_001	1,4059	1,7355	23%	5,6537	8,95
		H1S1_002	1,2723	1,5644	23%	5,4042	9,3419
		H1S1_003	1,3545	1,7915	32%	5,4097	9,3331
		H1S1_004	1,2298	1,4785	20%	5,6645	8,9678
		H1S1_005	1,2484	1,4915	19%	5,7645	8,8158
	Secuencia_6	H1S6_001	1,1224	1,253	12%	5,8909	8,6445
		H1S6_002	1,186	1,3622	15%	5,8746	8,6687
		H1S6_003	1,171	1,3537	16%	5,8132	8,7598
		H1S6_004	1,5202	1,9613	29%	5,6078	9,0724
		H1S6_005	1,6523	2,1296	29%	5,2951	9,5707

Secuencias	Imagen Muestra	SNR	SNR	SNR	MSE	PSNR
		(original)	(Despeckle) N= 150 Lambda =0.025	Mejorado	(error cuadrático medio)	
Secuencia_9	H1S9_001	1,1605	1,3504	16%	5,5583	9,1493
	H1S9_002	1,2223	1,4567	19%	5,5767	9,1035
	H1S9_003	1,2255	1,46	19%	5,4883	9,2423
	H1S9_004	1,2536	1,5353	22%	5,6154	9,0607
	H1S9_005	1,1377	1,306	15%	5,543	9,1734
Promedio		1,27749333	1,54864	21%	5,61066	9,05692667
Varianza		0,0220192	0,06321137		0,03026452	0,06906774
Desviación estándar		0,14838869	0,25141871		0,17396701	0,26280742

Luego de la etapa de filtrado se aplican los métodos de segmentación expuestos en la sección 3.2 y 3.3, obteniendo los resultados mostrados en la *Figura 13* a continuación.



Figura 13. Imágenes resultado de la segmentación.

4.3 Modelado

A través de una regresión polinomial con la función *poly.fit* [45] de la librería NumPy de Python, se realizó el modelado del contorno extraído (color morado), obteniendo los coeficientes que

modelan el contorno extraído y posteriormente se grafica la nueva función (color cian), tal como se muestra en la *Figura 14*.

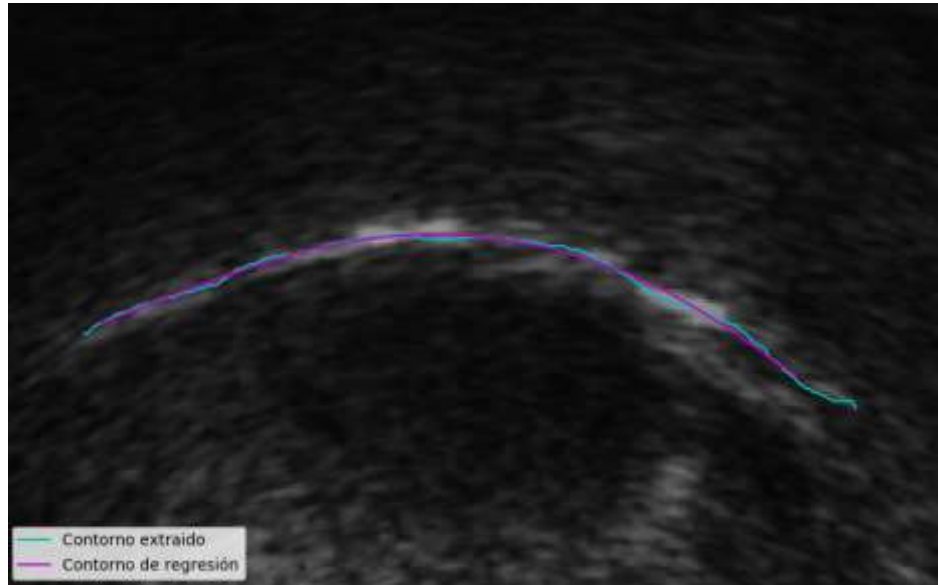


Figura 14. Regresión polinomial.

4.4. Edgetrak y MSD

Se incluyen las imágenes de una secuencia por hablante donde se observa el contorno hallado por *EdgeTrak* (verde), el contorno hallado por el algoritmo implementado en el proyecto luego del modelado (azul) y el trazado del contorno manual(amarillo), este último fue realizado por los autores de este proyecto y sirve como referencia para hallar el valor MSD y realizar la comparación entre distancia de pixeles entre métodos.

Un pixel es igual a 0,2040 mm para el caso del sistema de US adquirido por el grupo CEMOS de la Universidad Industrial de Santander, los valores de MSD obtenidos en la implementación de este proyecto se asemejan a los encontrados en el estado del arte [7].

En la *Figura 15* se observa las cinco imágenes originales pertenecientes a una secuencia por hablante, las imágenes de los 5 hablantes se encuentran en el Anexo D. Los contornos sobrepuestos en la imagen permiten el análisis de MSD y así obtener un valor cuantitativo como resultado de su comparación.

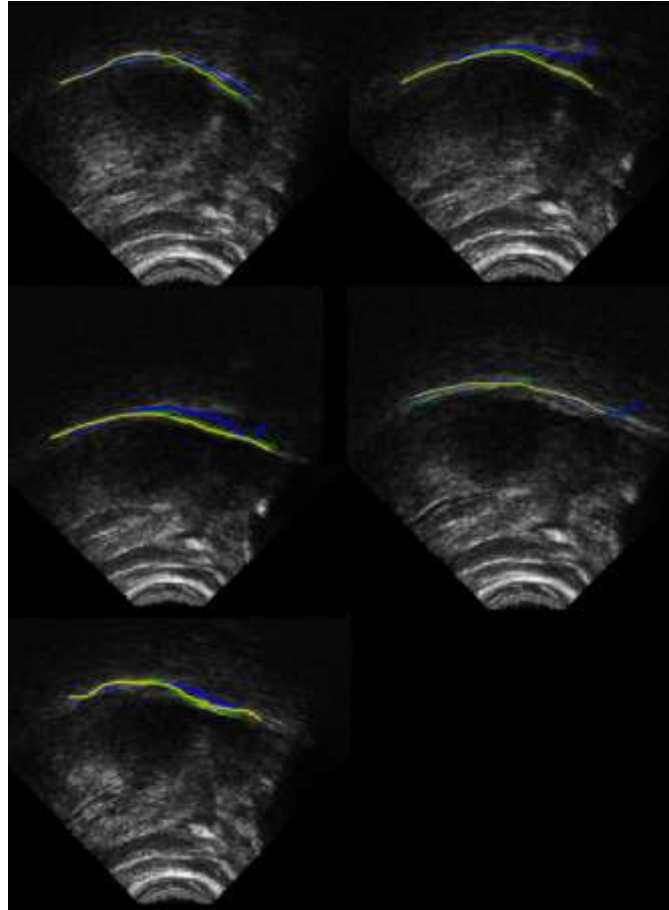


Figura 15. Imagen original con 3 contornos hablante 1.

En la Tabla 3, se encuentran registrados los valores de MSD *EdgeTrak*, el cual es resultado de comparar el contorno de trazado manual y el contorno generado por *EdgeTrak*; el valor de MSD Regresión, es el resultado de comparar el contorno de trazado manual y el contorno trazado por

la regresión polinomial que se encuentra el algoritmo implementado, estos para el hablante 1 (ver Anexo E para información de todos los hablantes).

Tabla 3
Valor MSD hablante 1

	Secuencias	Imagen muestra	MSD <i>EdgeTrak</i>	MSD Regresión	
Hablante 1	Secuencia_1	H1S1_001	1,8571	1,9000	
		H1S1_002	1,9429	1,9429	
		H1S1_003	1,6386	1,6988	
		H1S1_004	1,8500	1,7000	
		H1S1_005	1,8714	1,8714	
	Secuencia_6	H1S6_001	1,8857	0,5143	
		H1S6_002	1,9286	1,9143	
		H1S6_003	1,7143	1,7286	
		H1S6_004	1,8857	0,5143	
		H1S6_005	1,6386	1,6988	
	Secuencia_9	H1S9_001	1,8000	1,8286	
		H1S9_002	1,9286	0,5000	
		H1S9_003	1,7857	1,7514	
		H1S9_004	1,8143	1,7286	
		H1S9_005	1,8857	1,7000	
		Promedio		1,82847567	1,53279133
		Varianza		0,00957703	0,28768247
		Desviación estándar		0,0978623	0,5363604

En la *Figura 16*. Se tienen los valores resultantes promedio de MSD para cada uno de los 5 hablantes.

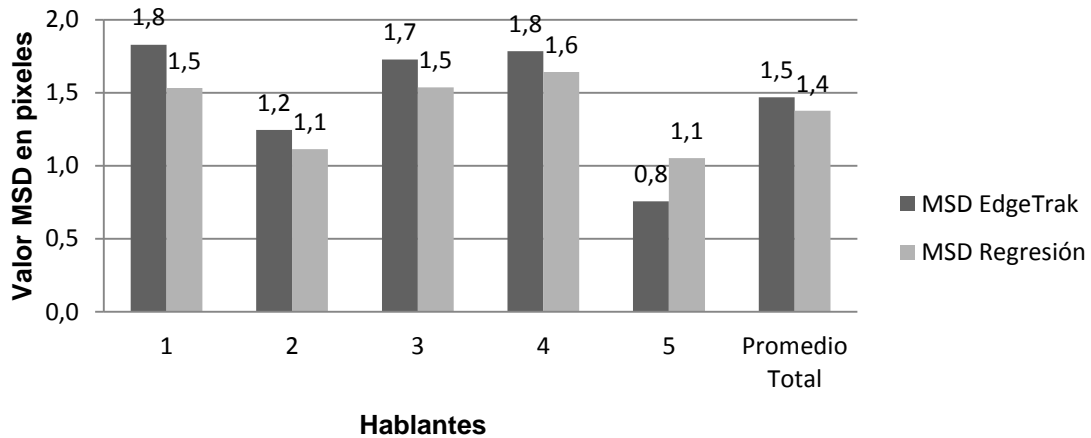


Figura 16. Valor MSD promedio para los 5 hablantes.

4.5. Tiempo de ejecución del código (python)

Se realizó un cálculo de los tiempos de ejecución de cada una de las secciones que se encuentran dentro del código implementado en el proyecto, se tomaron 5 mediciones y finalmente se promediaron estos valores, tal como se puede observar en la Tabla 4.

Se observan los valores encontrados del tiempo de ejecución en las secciones del código, el cual tiene un tiempo promedio total de ejecución de 117,7110 segundos; el punto crítico se encuentra en el filtro *despeckle* (DsFsrad) el cual ocupa un 95% del tiempo total de ejecución, debido a la cantidad de iteraciones seleccionadas para obtener resultados aceptables [37] de los métodos de evaluación cuantitativa descritos en la sección 3.5.

Para medir el tiempo de ejecución del algoritmo se utiliza un computador portátil con las siguientes especificaciones:

- Procesador: Intel® Core(™) i7-4510U CPU @ 2.00 GHz 2.60 GHz
- RAM: 8.00 GB (7.89 GB usables)
- Tipo de sistema: Sistema operativo de 64 bits, procesador basado en x64

- Windows 10 Pro 2016 Microsoft Corporation. All Rights reserved

Versión Python

- Python 3.6.0 (v3.6.0:41df79263a11, Dec 23 2016, 08:06:12) [MSC v.1900 64 bit (AMD64)] on win32

Tabla 4

Tiempo de ejecución algoritmo implementado.

Nº	Sección	Duración 1[Seg]	Duración 2 [Seg]	Duración 3 [Seg]	Duración 4 [Seg]	Duración 5 [Seg]	Duración prom [Seg]
1	Importación de Imágenes +recorte	0,0200	0,0300	0,0300	0,0300	0,0200	0,0260
2	Puntos de referencia	---	---	---	---	---	---
3	Filtro DsFsrad (150 niter)	112,4125	110,7082	111,9622	113,1248	111,5884	111,9592
4	ROI & SNR	117,8965	112,7696	115,8555	116,6051	115,8757	115,8005
5	RMSE	116,1181	116,5850	116,0947	116,3095	119,4575	116,9130
6	Umbralización	118,7884	115,0325	116,0264	118,9395	116,0181	116,9610
7	Dilatación & recorte ROI	117,6753	116,1346	116,7425	116,6566	117,8095	117,0037
8	Contornos & Canny	117,1782	116,0842	118,1482	117,3954	116,3948	117,0401
9	Máscara+ media de Canny	117,0963	117,1057	117,0911	117,0160	117,1049	117,0828
10	Regresión polinomial	117,3376	116,1954	117,1730	117,9944	117,3761	117,2153
11	MSD	118,0006	116,9051	117,2789	117,3926	117,6208	117,4396
12	Graficando	117,9387	117,6235	117,5774	117,5774	117,8379	117,7110
	Promedio						105,9229
	Varianza						1236,0811
	Desviación estándar						35,1579

5. Conclusiones

A partir del trabajo desarrollado y los resultados obtenidos, es posible exponer la siguiente conclusión general:

Se implementó un método automático para la extracción del contorno superior de la lengua en secuencias de imágenes de ultrasonido, con lo anterior, fue posible visualizar las variaciones que adopta el contorno de la lengua mientras el hablante pronuncia la frase establecida, esto sin la necesidad de una inicialización manual a partir de puntos de referencia.

De forma más específica:

Se extrajo el contorno de secuencias de imágenes de ultrasonido para cinco hablantes a partir de técnicas de procesamiento de imágenes como: filtrado, segmentación y detección de bordes, obteniendo los resultados expuestos en la sección 4.1. Se utilizó un filtro de difusión anisotrópica para reducir el ruido granular *speckle* presente con frecuencia en imágenes de ultrasonido. El método cuantitativo para evaluar esta reducción fue la relación señal a ruido SNR. Tal como expone la Tabla 2 y el Anexo C, se logró una mejora de hasta 32% en el valor de SNR para la imagen filtrada. Finalmente se obtiene un porcentaje de mejora promedio de 13% para las 75 imágenes tratadas. Los resultados obtenidos se asemejan a los obtenidos en el estado del arte [37].

Se logró modelar el contorno extraído de la parte superior de la lengua a través de unos pocos coeficientes obtenidos por medio de una regresión polinomial. La Figura 12 muestra el contorno modelado en contraste con el extraído por el método implementado. El contorno modelado se compara con el trazado manual de referencia a través del método cuantitativo de evaluación

MSD. El resultado promedio MSD obtenido es de 1,4 píxeles. Resultado inferior a los valores obtenidos en el estado del arte [7], lo que indica una menor distancia entre los contornos, y sustenta la implementación del método de extracción y su modelado.

Se comparó el contorno obtenido a través del método sin inicialización manual implementado, con el obtenido a través del método semi-automático usado con frecuencia en la literatura, *EdgeTrak*. Haciendo uso de la media de la suma de distancias MSD, se estimó la distancia entre el trazado manual de referencia en contraste con *EdgeTrak*, tal como lo consigna la *Figura 16* y el Anexo E, obteniendo un valor MSD promedio de 1,5 píxeles, lo cual indica que la distancia entre el contorno obtenido a través de *EdgeTrak* y el contorno de referencia, es superior a la distancia que existe entre el método implementando y el contorno generado por medio del trazado manual.

6. Recomendaciones

Para la adquisición de imágenes a través del escáner de ultrasonido, es importante contar con una estructura de soporte del escáner, de forma tal que aseguren la homogeneidad en las muestras con respecto al ángulo de visión del transductor.

Es importante que se utilice una cantidad considerable de gel conductor durante la adquisición de las imágenes y videos, esto permite una mejor visualización de las estructuras que se desean estudiar.

7. Trabajo Futuro

Se propone analizar los puntos críticos del tiempo de ejecución del algoritmo implementado, expuestos en la Tabla 12, con el fin de reducirlos y acercar su funcionamiento al tiempo real.

Del mismo modo se sugiere relacionar de forma más directa el contorno de la lengua con su respectivo fonema para cada hablante.

Citas Bibliográficas

- [1] Russell, G. O. (1928). The vowel: Its physiological mechanism as shown by x-ray. Columbus: Ohio State University Press.
- [2] Carstens Medizinelektronik GmbH, “ 3D Electromagnetic Articulograph”. Recuperado de <http://www.articulograph.de/>
- [3] Toutios, A; Narayanan, S; (2013). Articulatory synthesis of french connetced speech from EMA data, signal analysis & interpretation laboratory (SAIL), university of southern california, USA.
- [4] Maeda, S; Laprie, Y; (2013). Vowel and prosodic factor dependent variations of vocal-tract length. InterSpeech - 14th Annual Conference of the International Speech Communication Association - 2013, Aug 2013, Lyon, France.
- [5] Alan Wrench. “MOCHA-TIMIT”, The Centre for Speech Technology Research. [Online]. Available in: <http://www.cstr.ed.ac.uk/research/projects/artic/mocha.html>.
- [6] Schonle, P., Grabe, K., Wenig, P., Hohne, J., Schrader,J., & Conrad, B. (1987). Electromagnetic articulography: Use of alternating magnetic fields for tracking movements of multiple points inside and outside the vocal tract. *Brain and Language*, 31, 26–35.
- [7] Wans, H; Wang, S; Denby, B; Dang, J; (2015). Automatic tongue contour tracking in ultrasound sequences without manual initialization, Tianjin University, China, Universite

- Pierre et Marie Curie, Paris, France, Japan Advaced Institute of science and technology, Ishikawa, Japan. 16-19.
- [8] Visión artificial, Universidad Nacional de Quilmes, Ing. En automatización y control industrial, octubre de 2005 Segmentación por Umbralización, Método Otsu, recuperado de,
<http://iaci.unq.edu.ar/materias/vision/archivos/apuntes/Segmentaci%C3%B3n%20por%20umbralizaci%C3%B3n%20M%C3%A9todo%20de%20Otsu.pdf>
- [9] Alexander Mordvintsev & Abid K, OpenCV-python Documentation, Abril 2017, recuperado de <https://media.readthedocs.org/pdf/opencv-python-tutroals/latest/opencv-python-tutroals.pdf>
- [10] Operadores Morfológicos, procesamiento de imágenes, segmentación, recuperado de <http://alojamientos.us.es/gtocom/pid/tema5-1.pdf>
- [11] Canny, J., (1986) A Computational Approach To Edge Detection, IEEE Trans. Pattern Analysis and Machine Intelligence, 8(6):679–698
- [12] Seemore Imaging Canada, ULTRASONIC SCANNING APPARATUS PI 7.5 MHz, system, Recuperado de <http://seemore.ca/portable-ultrasound-products/pi-7-5-mhz-speech-language-pathology-99-5544-can/>
- [13] Min Li and Chandra Kambhamettu, Video/Image Modeling and Synthesis *EdgeTrak*, University of Delaware, Newark, 19716, Mauren Stone, Vocal Tract Visualization Lab, University of Maryland, Baltimore, 21201.
- [14] Csapó, T; Lulich, S; (2015). Error analysis of extracted tongue contours from 2d ultrasound images. In INTERSPEECH 2015, 16th Annual Conference of the International

- Speech Communication Association, Dresden, Germany, September 6-10, 2015, pages 2157–2161
- [15] Badin P., Elisei F., Bailly G., Tarabalka Y. (2008) An Audiovisual Talking Head for Augmented Speech Generation: Models and Animations Based on a Real Speaker's Articulatory Data. In: Perales F.J., Fisher R.B. (eds) Articulated Motion and Deformable Objects. AMDO 2008. Lecture Notes in Computer Science, vol 5098. Springer, Berlin, Heidelberg
- [16] Sepúlveda, D. M. Casas, G. Castellanos. (2012). Importancia de las frecuencias de resonancia del tracto vocal en la estimación de posiciones articulatorias, *Revista Ingeniería Biomédica*, ISSN 1909-9762. Volumen 6, number 11, 2012.
- [17] Afshan, A., and Kumar Ghosh, P., (2015). Improved subject-independent acoustic-to-articulatory inversion. *Speech Communication*, 66:1 – 16,.
- [18] Shinji Maeda (1990). *Speech Production and Speech Modelling*, chapter Compensatory articulation during speech: evidence from the analysis and synthesis of vocal-tract shapes using articulatory model, Kluwer Academic Publishers, pages 131–149,.
- [19] Chao Qin; Miguel, A; Carreira-Perpinan; (2010). Reconstructing the full tongue contour from EMA-X-ray microbeam. In *IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, pages 4190–4193.
- [20] D. Kouamé and et. al. (200). Ultrasound imaging: signal acquisition, new advanced processing for biomedical and industrial application. In *Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, pages 993–996.

- [21] A. Wrench. The MOCHA-TIMIT articulatory database. Technical report, Queen Margaret University College, 1999. Recuperado de www.cstr.ed.ac.uk/research/projects/artic/mocha.html.
- [22] Narayanan, S; Toutios, A; Ramanarayanan, V; Lammert, A; Kim, J; Lee, S; Nayak, K; Kim, Y; Zhu, Y; Goldstein, L; Byrd, D; Bresch, E; Ghosh, P; Katsamanis, A; Proctor, M; (2014). Real-time magnetic resonance imaging and electromagnetic articulography database for speech production research (tc). The Journal of the Acoustical Society of America, 136 (3).
- [23] Akgul, Y. S; Kambhamettu, C; Stone, M. (Oct 1999). Automatic extraction and tracking of the tongue contours. IEEE Transactions on Medical Imaging, 18(10):1035–1045.
- [24] Xu, K., Roussel, P., & Denby, B. (2016). A comparative study on the contour tracking algorithms in ultrasound tongue images with automatic re-initialization, 139(May), 154–160.
- [25] National Institute of Biomedical Imaging and Bioengineering, temas científicos, ultrasonido, (2006). Recuperado de <https://www.nibib.nih.gov/espanol/temas-cientificos/ultrasonido#1981>
- [26] Vargas, A; Amescua, L; Bernal, A; Pineda, C. (2008). Principios físicos básicos del ultrasonido, sonoanatomía del sistema musculoesquelético y artefactos ecográficos. Instituto Nacional de Cardiología Ignaci Chávez, México, Nov.Dic, 361-373
- [27] Kass, M., Witkin, A. & Terzopoulos, D. (1988). Snakes: Active contour models. Int J Comput Vision 1: 321.

- [28] Rasgos distintivos del tracto vocal, recuperado de, <http://hispanoteca.eu/Gram%C3%A1ticas/Gram%C3%A1tica%20espa%C3%B1ola/Ortograf%C3%ADa-%20RAE%202010-Letras%20y%20fonemas.htm>
- [29] Daniloff, R., Shuckers, G. y Feth, L. (1980). The physiology of speech and hearing. Englewood Cliffs, NJ: Prentice Hall
- [30] Educando a fondo, El fonema, rasgo distintivo, (Agosto 2013. Recuperado de <http://enmagoro.blogspot.com.co/2013/08/el-fonema-rasgos-distintivos.html>
- [31] Aplicación por Xavier López Morrás, 2004, Transcriptor fonético automático del español, recuperado de <http://www.aucel.com/pln/transbase.html>
- [32] Interson, Abril 2017, System Requirements, recuperado de <http://www.interson.com/system>
- [33] Hernández, N; (2010). Procesamiento digital de imágenes de ultrasonido, Instituto Politécnico Nacional, Escuela superior de Ingeniería Mecánica y Eléctrica
- [34] Christos P. Loizou, Constantinos S. Pattichis, Despeckle filtering algorithms and software for ultrasound imaging, Morgan & Claypool publisher, Synthesis Lectures on algorithms and software in engineering, Andreas Spanias, Series Editor.
- [35] Perona, P; Malik, J; (1990.). Scale-space and edge detection using anisotropic diffusion, IEEE Trans. Pattern Anal. Mach. Intell., vol. 12, no. 7, pp. 629–639, July.
- [36] Y. Yongjian and S.T. Acton, “Speckle reducing anisotropic diffusion,” IEEE Trans. Image Process., vol. 11, no. 11, pp. 1260–1270, November 2002.
- [37] Giraldo-Guzman, J; Porto-Solano, O; Cadena-Bonfanti, A; Contreras-Ortiz, S; Speckle Reduction in Echocardiography by Temporal Compounding and Anisotropic Diffusion Filtering, Universidad Tecnológica de Bolívar, Cartagena de Indias, Colombia.

- [38] Lee, J.S. (1981) “Speckle analysis and smoothing of synthetic aperture radar images,” *Comp. Graphics Image Process.*, vol. 17, pp. 24–32
- [39] Lee, J.S. (1980) “Digital image enhancement and noise filtering by using local statistics,” *IEEE Trans. Pattern Anal. Mach. Intell.*, PAMI-2, no. 2, pp. 165–168.
- [40] Lee, J.S. (1981) “Refined filtering of image noise using local statistics,” *Comput. Graphics Image Process*, vol. 15, pp. 380–389
- [41] Frost, V.S; Stiles, J.A; Shanmungan, K.S ; Holtzman, J.C ; (1982). “A model for radar images and its application for adaptive digital filtering of multiplicative noise,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 4, no. 2, pp. 157–165
- [42] Maini, R; Aggarwal, Dr.H. *Study and Comparison of various Image Edge Detection Techniques*, Punjabi university, Patiala, 147002, India.
- [43] Burckhardt, C.B. (1978). “Speckle in ultrasound b-mode scans,” *Sonics and Ultrasonics*, *IEEE Transactions on* 25(1), pp. 1–6
- [44] Jaumard-hakoun, A., Xu, K., Roussel-ragot, P., & Stone, M. (n.d.). Tongue contour extraction from ultrasound images based on deep neural network. Université Pierre et Marie Curie, Paris, France, Vocal Tract Visualization Lab, University of Maryland Dental School, Baltimore, USA, Present affiliation: Institut Langevin, ESPCI ParisTech, PSL Research University, Paris, France.
- [45] Función `poly.fit`, librería NumPy, Python. Recuperado de <https://docs.scipy.org/doc/numpy/reference/generated/numpy.polyfit.html>

Referencias Bibliográficas

- Abd-Elmoniem, K; A.-B. Youssef, a-B; Kadah, Y. (2002). Real-time speckle reduction and coherence enhancement in ultrasound imaging via nonlinear anisotropic diffusion,” IEEE Trans. Biomed. Eng., vol. 49, no. 9, pp. 997–1014, doi:10.1109/TBME.2002.1028423.
- Black, M;apiro, G; Marimont, D; Heeger, D; (1998). “Robust anisotropic diffusion,” IEEE Trans. Image Process., vol. 7, no. 3, pp. 421–432, doi:10.1109/83.661192.
- Burckhardt, C.B. (1978). “Speckle in ultrasound b-mode scans,” Sonics and Ultrasonics, IEEE Transactions on 25(1), pp. 1–6
- Canny, J., (1986) A Computational Approach To Edge Detection, IEEE Trans. Pattern Analysis and Machine Intelligence, 8(6):679–698
- Csapó, T; Lulich, S; (2015). Error analysis of extracted tongue contours from 2d ultrasound images. In INTERSPEECH 2015, 16th Annual Conference of the International Speech Communication Association, Dresden, Germany, September 6-10, 2015, pages 2157–2161.
- Frost, V.S; Stiles, J.A; Shanmungan, K.S ; Holtzman, J.C ; (1982). “A model for radar images and its application for adaptive digital filtering of multiplicative noise,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 4, no. 2, pp. 157–165
- Giraldo-Guzman, J; Porto-Solano, O; Cadena-Bonfanti, A; Contreras-Ortiz, S; Speckle Reduction in Echocardiography by Temporal Compounding and Anisotropic Diffusion Filtering, Universidad Tecnológica de Bolívar, Cartagena de Indias, Colombia.
- Hernández, N; (2010). Procesamiento digital de imágenes de ultrasonido, Instituto Politécnico Nacional, Escuela superior de Ingeniería Mecánica y Eléctrica

- Jaumard-hakoun, A., Xu, K., Roussel-ragot, P., & Stone, M. (n.d.). Tongue contour extraction from ultrasound images based on deep neural network. Université Pierre et Marie Curie, Paris, France, Vocal Tract Visualization Lab, University of Maryland Dental School, Baltimore, USA, Present affiliation: Institut Langevin, ESPCI ParisTech, PSL Research University, Paris, France.
- Lee, J.S. (1981) "Refined filtering of image noise using local statistics," *Comput. Graphics Image Process*, vol. 15, pp. 380–389
- Lee, J.S. (1980) "Digital image enhancement and noise filtering by using local statistics," *IEEE Trans. Pattern Anal. Mach. Intell.*, PAMI-2, no. 2, pp. 165–168.
- Lee, J.S. (1981) "Speckle analysis and smoothing of synthetic aperture radar images," *Comp. Graphics Image Process.*, vol. 17, pp. 24–32
- Loizou, C.P; Pattichis, C.S; Christodoulou, C.I; Istepanian, R.S.H; Pantziaris, M; Nicolaidis, A. (2005). Comparative evaluation of despeckle filtering in ultrasound imaging of the carotid artery," *IEEE Trans. Ultrason. Ferroelectr. Freq. Control*, vol. 52, no. 10, pp. 1653–1669
- Maini, R; Aggarwal, Dr.H. Study and Comparison of various Image Edge Detection Techniques, Punjabi university, Patiala, 147002, India.
- Rerona, P; Malik, J; (1990.). Scale-space and edge detection using anisotropic diffusion, *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 12, no. 7, pp. 629–639, July.
- Xu, K., Roussel, P., & Denby, B. (2016). A comparative study on the contour tracking algorithms in ultrasound tongue images with automatic re-initialization, 139(May), 154–160.

Apéndices

Apéndice A – Carta consentimiento hablantes

Apéndice B – Base de datos imágenes de ultrasonido

Apéndice C – Tablas valores SNR 5 hablantes

Tabla 5
Características señal a ruido hablante 1

Secuencias	Imagen Muestra	SNR	SNR	SNR	MSE	PSNR
		(original)	(Despeckle) N= 150 Lambda =0.025	Mejorado	(error cuadrático medio)	
Habla 1	H1S1_001	1,4059	1,7355	23%	5,6537	8,95
	H1S1_002	1,2723	1,5644	23%	5,4042	9,3419
	Secuencia_1 H1S1_003	1,3545	1,7915	32%	5,4097	9,3331
	H1S1_004	1,2298	1,4785	20%	5,6645	8,9678
	H1S1_005	1,2484	1,4915	19%	5,7645	8,8158
Habla 6	H1S6_001	1,1224	1,253	12%	5,8909	8,6445
	H1S6_002	1,186	1,3622	15%	5,8746	8,6687
	Secuencia_6 H1S6_003	1,171	1,3537	16%	5,8132	8,7598
	H1S6_004	1,5202	1,9613	29%	5,6078	9,0724
	H1S6_005	1,6523	2,1296	29%	5,2951	9,5707
Secuencia_9 H1S9_001	1,1605	1,3504	16%	5,5583	9,1493	

Secuencias	Imagen Muestra	SNR	SNR (Despeckle)	SNR	MSE	PSNR
		(original)	N= 150 Lambda =0.025	Mejorado	(error cuadrático medio)	
	H1S9_002	1,2223	1,4567	19%	5,5767	9,1035
	H1S9_003	1,2255	1,46	19%	5,4883	9,2423
	H1S9_004	1,2536	1,5353	22%	5,6154	9,0607
	H1S9_005	1,1377	1,306	15%	5,543	9,1734
Promedio		1,27749333	1,54864	21%	5,61066	9,05692667
Varianza		0,0220192	0,06321137		0,03026452	0,06906774
Desviación estandar		0,14838869	0,25141871		0,17396701	0,26280742

Tabla 6
Características señal a ruido hablante 2

Secuencias	Imagen Muestra	SNR	SNR (Despeckle)	SNR	MSE	PSNR
		(original)	N= 150 Lambda =0.025	Mejorado	(error cuadrático medio)	
Secuencia_5	H2S5_001	1,3732	1,7443	27%	4,6362	10,656
	H2S5_002	1,1138	1,3341	20%	4,7881	10,2705
	H2S5_003	1,0634	1,2329	16%	4,9078	10,1441
	H2S5_004	1,0654	1,2269	15%	4,9729	1,2269
	H2S5_005	1,1122	1,2936	16%	4,7275	10,4866
Secuencia_6	H2S6_001	1,1771	1,4159	20%	4,4033	11,051
	H2S6_002	0,9824	1,072	9%	4,4721	10,899
	H2S6_003	0,91827	0,9956	8%	4,5133	10,8369
	H2S6_004	1,027	1,1528	12%	4,7738	10,3319
	H2S6_005	1,0097	1,1279	12%	4,7411	10,4962
Secuencia_7	H2S7_001	1,2101	1,4956	24%	4,7567	10,4678
	H2S7_002	0,9579	1,062	11%	4,6431	10,573
	H2S7_003	0,9516	1,035	9%	4,6868	10,5791
	H2S7_004	1,09	1,223	12%	4,7758	10,2929
	H2S7_005	1,147	1,3673	19%	4,8084	10,3392
Promedio		1,079938	1,25192667	15%	4,70712667	9,91007333
Varianza		0,01388614	0,03990969		0,02396903	5,83285554

Secuencias	Imagen Muestra	SNR (original)	SNR (Despeckle) N= 150 Lambda =0.025	SNR Mejorado	MSE (error cuadrático medio)	PSNR
	Desviación estandar	0,11783948	0,1997741		0,15481936	2,41513054

Tabla 7
Características señal a ruido hablante 3

Secuencias	Imagen Muestra	SNR (original)	SNR (Despeckle) N= 150 Lambda =0.025	SNR Mejorado	MSE (error cuadrático medio)	PSNR	
Hablante 3	H3S1_001	1,0111	1,147	13%	5,8123	8,7441	
	H3S1_002	1	1,1322	13%	5,8722	8,6722	
	Secuencia_1	H3S1_003	0,9963	1,1269	13%	5,883	8,6391
		H3S1_004	1,0087	1,147	14%	5,7802	8,7922
		H3S1_005	1,0169	1,1676	15%	5,7472	8,8419
		H3S2_001	1,0174	1,1739	15%	5,7826	8,7886
		H3S2_002	1,0069	1,159	15%	5,6332	8,9987
	Secuencia_2	H3S2_003	0,9817	1,112	13%	5,6114	9,032
		H3S2_004	0,9959	1,1367	14%	5,6125	9,048
		H3S2_005	0,9958	1,1317	14%	5,5919	9,097
		H3S3_001	1,058	1,1782	11%	4,6147	10,7654
		H3S3_002	0,9733	1,062	9%	4,4986	10,9695
	Secuencia_3	H3S3_003	1,0612	1,1844	12%	4,6068	10,7631
		H3S3_004	1,038	1,1319	9%	4,7534	10,4737
		H3S3_005	1,019	1,1061	9%	4,8033	10,3831
Promedio		1,01201333	1,13977333	13%	5,37355333	9,46724	
Varianza		0,00062003	0,00101392		0,28863037	0,81020997	
Desviación estandar		0,02490043	0,03184208		0,53724331	0,90011664	

Tabla 8

Características señal a ruido hablante 4

Secuencias	Imagen Muestra	SNR	SNR	SNR Mejorado	MSE	PSNR
		(original)	(Despeckle) N= 150 Lambda =0.025		(error cuadrático medio)	
Secuencia_1	H4S1_001	0,8924	0,9703	9%	5,684	8,8688
	H4S1_002	0,8798	0,9537	8%	5,7282	8,819
	H4S1_003	0,8828	0,9563	8%	5,6385	8,9385
	H4S1_004	0,8729	0,9453	8%	5,8541	8,6301
	H4S1_005	0,8802	0,9558	9%	5,8128	8,6916
Secuencia_2	H4S2_001	0,8894	0,9692	9%	5,8654	8,6307
	H4S2_002	0,8927	0,9737	6%	5,7966	8,6983
	H4S2_003	0,8919	0,9668	8%	5,7525	8,6944
	H4S2_004	0,898	0,9672	8%	5,5821	9,0607
	H4S2_005	0,8878	0,961	8%	5,8077	8,7338
Secuencia_8	H4S8_001	0,9845	1,0752	9%	5,3131	9,507
	H4S8_002	1,1306	1,275	13%	5,3092	9,4087
	H4S8_003	1,068	1,1984	12%	5,4577	9,2564
	H4S8_004	1,0271	1,1314	10%	5,3993	9,2976
	H4S8_005	1,0437	1,1498	10%	5,2247	9,6
Promedio		0,94145333	1,02994	9%	5,61506	8,98904
Varianza		0,00727437	0,01158023		0,04824878	0,11554475
Desviación estandar		0,08528993	0,10761146		0,21965604	0,33991874

Tabla 9

Características señal a ruido hablante 5

Secuencias	Imagen Muestra	SNR	SNR	SNR Mejorado	MSE	PSNR
		(original)	(Despeckle) N= 150 Lambda =0.025		(error cuadrático medio)	
Secuencia_1	H5S1_001	0,7667	0,8214	7%	5,2573	8,0888
	H5S1_002	0,8386	0,8964	7%	5,1382	9,8148
	H5S1_003	0,8085	0,8578	6%	5,0988	9,8472
	H5S1_004	0,8337	0,8879	7%	4,9894	10,0529
	H5S1_005	0,8734	0,9372	7%	4,9096	10,193
Secuencia_2	H5S2_001	0,711	0,7513	6%	4,403	11,156
	H5S2_002	0,7229	0,7723	7%	4,3138	11,3166
	H5S2_003	0,7472	0,7997	7%	4,2741	11,397
	H5S2_004	0,7936	0,8656	9%	4,1444	11,6647
	H5S2_005	0,8328	0,9225	11%	4,0218	11,9081
Secuencia_4	H5S4_001	0,9724	1,1204	15%	4,2646	11,4334
	H5S4_002	0,9604	1,0607	10%	4,3753	11,2109
	H5S4_003	0,9549	1,0523	10%	4,4848	10,9963
	H5S4_004	0,7992	0,8521	7%	4,3572	11,2469
	H5S4_005	0,7366	0,7697	4%	4,1145	11,7618
Promedio		0,82346	0,89115333	8%	4,54312	10,8058933
Varianza		0,00729807	0,01252608		0,17208512	1,04386414
Desviación estandar		0,08542873	0,11191999		0,41483144	1,0216967

Apéndice D – Imágenes con contornos extraídos por EdgeTrak, de forma manual e implementando algoritmo presentado en este proyecto.

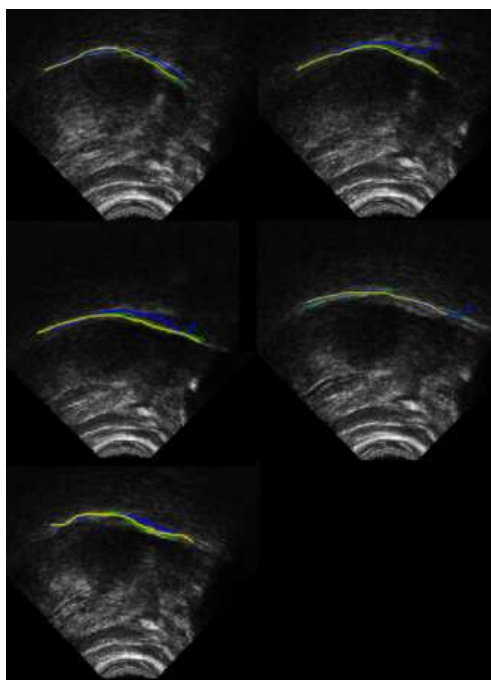


Figura 17. Imagen original con 3 contornos hablante 1.

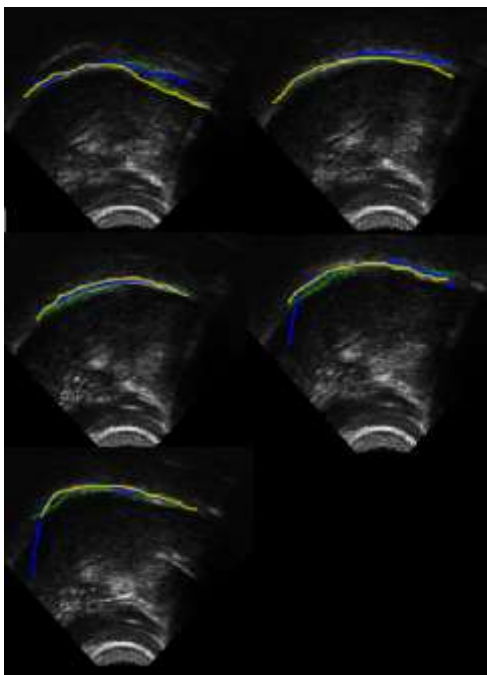


Figura 18. Imagen original con 3 contornos hablante 2.

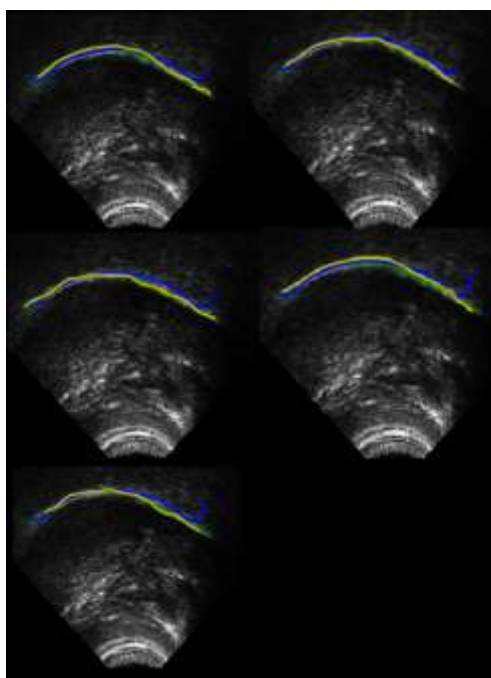


Figura 19. Imagen original con 3 contornos hablante 3.

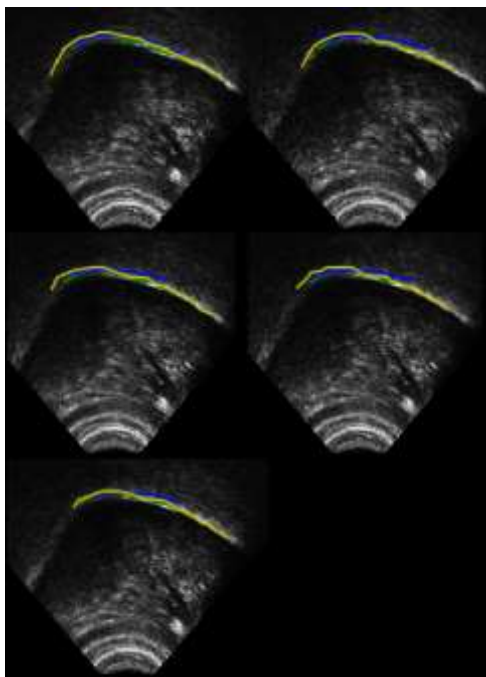


Figura 20. Imagen original con 3 contornos hablante 4.

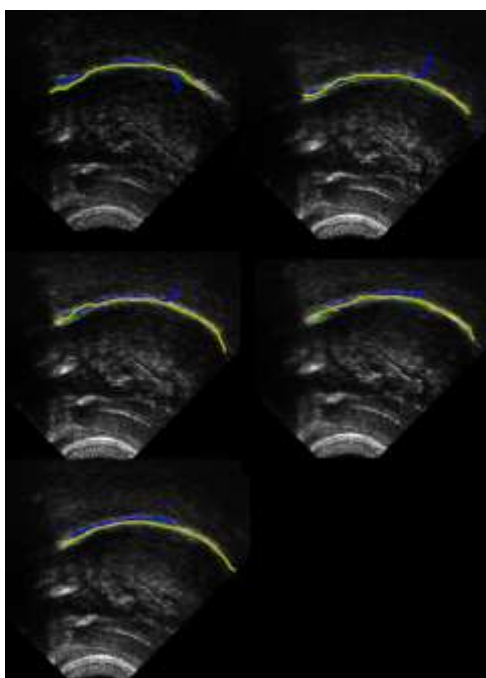


Figura 21. Imagen original con 3 contornos hablante 5.

Apéndice E – Tablas con valores MSD para cada uno de los 5 hablantes.

Tabla 10

Valor MSD hablante 1

Secuencias	Imagen muestra	MSD	MSD
		<i>EdgeTrak</i>	Regresión
Secuencia_1	H1S1_001	1,8571	1,9000
	H1S1_002	1,9429	1,9429
	H1S1_003	1,6386	1,6988
	H1S1_004	1,8500	1,7000
	H1S1_005	1,8714	1,8714
Secuencia_6	H1S6_001	1,8857	0,5143
	H1S6_002	1,9286	1,9143
	H1S6_003	1,7143	1,7286
	H1S6_004	1,8857	0,5143
	H1S6_005	1,6386	1,6988
Secuencia_9	H1S9_001	1,8000	1,8286
	H1S9_002	1,9286	0,5000
	H1S9_003	1,7857	1,7514
	H1S9_004	1,8143	1,7286
	H1S9_005	1,8857	1,7000
Promedio		1,82847567	1,53279133
Varianza		0,00957703	0,28768247
Desviación estándar		0,0978623	0,5363604

Tabla 11

Valor MSD hablante 2

Secuencias	Imagen	MSD	MSD
	muestra	<i>EdgeTrak</i>	Regresión
Secuencia_5	H2S5_001	1,7714	1,8000
	H2S5_002	1,8714	1,8857
	H2S5_003	1,8858	1,3000
	H2S5_004	1,8714	1,8714
	H2S5_005	1,8419	1,7857
Secuencia_6	H2S6_001	0,7714	0,5143
	H2S6_002	0,3143	0,3429
	H2S6_003	1,6364	1,5194
	H2S6_004	0,4500	1,8375
	H2S6_005	0,5432	0,5062
Secuencia_7	H2S7_001	1,9143	1,7286
	H2S7_002	0,5672	0,3284
	H2S7_003	0,9231	0,4308
	H2S7_004	1,7500	0,2647
	H2S7_005	0,5714	0,6032
Promedio		1,24554339	1,11457651
Varianza		0,42262413	0,47008457
Desviación estándar		0,65009548	0,68562714

Tabla 12

Valor MSD hablante 3

Secuencias	Imagen	MSD	MSD
	muestra	<i>EdgeTrak</i>	Regresión
Secuencia_1	H3S1_001	1,9286	1,4429
	H3S1_002	1,9286	1,8429
	H3S1_003	1,9714	1,4714
	H3S1_004	1,9286	1,7000
	H3S1_005	1,9000	1,4143
Secuencia_2	H3S2_001	1,9429	1,9429
	H3S2_002	1,8571	1,8286
	H3S2_003	1,8857	1,8143

Secuencias	Imagen muestra	MSD <i>EdgeTrak</i>	MSD Regresión
	H3S2_004	1,9143	1,4429
	H3S2_005	1,9714	1,8429
	H3S3_001	1,8429	1,4714
	H3S3_002	1,8571	0,6714
Secuencia_3	H3S3_003	0,6000	1,9429
	H3S3_004	0,5143	0,3143
	H3S3_005	1,8750	1,9167
Promedio		1,72785487	1,53730121
Varianza		0,22772738	0,22391302
Desviación estándar		0,4772079	0,47319449

Tabla 13

Valor MSD hablante 4

Secuencias	Imagen muestra	MSD <i>EdgeTrak</i>	MSD Regresión
	H4S1_001	1,8429	1,8429
	H4S1_002	1,4571	1,8000
Secuencia_1	H4S1_003	1,8000	1,7857
	H4S1_004	1,4286	1,7714
	H4S1_005	1,8000	1,7857
	H4S2_001	1,8857	1,8857
	H4S2_002	1,9714	1,9714
Secuencia_2	H4S2_003	1,8714	1,8714
	H4S2_004	1,8571	1,7286
	H4S2_005	1,9143	1,9143
	H4S8_001	1,8571	1,5000
	H4S8_002	1,8571	0,6714
Secuencia_8	H4S8_003	1,8143	0,8143
	H4S8_004	1,6714	1,6571
	H4S8_005	1,7429	1,6429
Promedio		1,78476114	1,64285667
Varianza		0,02418266	0,14807584
Desviación estándar		0,15550775	0,38480623

Tabla 14

Valor MSD hablante 5

Secuencias	Imagen	MSD	MSD
	muestra	<i>EdgeTrak</i>	Regresión
Secuencia_1	H5S1_001	0,5000	1,7286
	H5S1_002	0,4571	0,3857
	H5S1_003	0,4000	0,3857
	H5S1_004	0,3714	0,3571
	H5S1_005	0,3143	0,1429
Secuencia_2	H5S2_001	0,3429	1,3571
	H5S2_002	0,0857	1,4143
	H5S2_003	0,2714	0,3286
	H5S2_004	0,4857	0,4143
	H5S2_005	0,4286	0,4714
Secuencia_4	H5S4_001	1,9286	1,8714
	H5S4_002	1,9286	1,9143
	H5S4_003	1,9429	1,6143
	H5S4_004	1,4286	1,5571
	H5S4_005	0,4714	1,8571
Promedio		0,75714099	1,05333173
Varianza		0,45367292	0,48435726
Desviación estándar		0,67355246	0,6959578