

---

# Herramienta de diagnóstico y evaluación para voces alaríngeas

---

---

Alumno: Larrea Murga, Ainhoa

Directora: Navas Cordón, Eva

Curso: 2015-2016

Fecha: 30 de septiembre de 2016

---



## Índice

<b>Resumen</b>	<b>iv</b>
<b>Laburpena</b>	<b>v</b>
<b>Abstract</b>	<b>vi</b>
<b>Lista de tablas</b>	<b>vii</b>
<b>Lista de figuras</b>	<b>ix</b>
<b>Lista de acrónimos</b>	<b>xi</b>
<b>Memoria</b>	<b>12</b>
<b>1.- Objetivos y alcance</b>	<b>12</b>
<b>2.- Beneficios del proyecto</b>	<b>15</b>
<b>3.- Estado del arte</b>	<b>17</b>
3.1.- Entrenamiento de la voz esofágica .....	17
3.2.- Problemas asociados con la voz esofágica o erigmofónica.....	18
3.3.- Evaluación de la voz esofágica .....	18
<i>Prótesis fonatoria</i> .....	19
3.4.- Métodos de evaluación objetiva.....	19
<i>Fonetograma</i> .....	20
<i>Análisis espectrográfico</i> .....	21
<b>4.- Análisis de alternativas</b>	<b>24</b>
4.1.- Adquisición de la base de datos.....	24
4.1.1.- <i>Grabación de la base de datos</i> .....	24
4.1.2.- <i>Grabaciones existentes</i> .....	24
4.1.3.- <i>Comparativa</i> .....	24
4.2.- Tipos de grabaciones.....	25
4.3.- Parámetros a analizar.....	25

4.4.- Herramienta software para programar .....	27
4.4.1.- <i>Matlab</i> .....	27
4.4.2.- <i>GNU Octave</i> .....	27
4.4.3.- <i>Genius</i> .....	27
4.4.4.- <i>Comparativa</i> .....	28
<b>5.- Análisis de riesgos</b>	<b>29</b>
<b>Metodología</b>	<b>31</b>
<b>6.- Descripción de tareas. Gantt</b>	<b>31</b>
6.1.- Equipo de trabajo .....	31
6.2.- Plan de trabajo .....	31
6.3.- Descripción de paquetes de trabajo .....	32
6.4.- Diagrama Gantt .....	34
<b>7.- Descripción de la solución</b>	<b>35</b>
7.1.- Creación de la base de datos .....	35
7.1.1.- <i>Búsqueda de bases de datos existentes</i> .....	35
7.1.2.- <i>Edición de las bases de datos</i> .....	37
7.1.3.- <i>Base de datos resultante</i> .....	38
7.2.- Creación de la herramienta.....	39
7.2.1.- <i>La periodicidad</i> .....	39
7.2.2.- <i>La regularidad de la fonación</i> .....	43
7.2.3.- <i>El ruido en frecuencia alta</i> .....	47
7.2.4.- <i>La voz rota o quebrada</i> .....	49
7.2.5.- <i>La velocidad del habla</i> .....	52
7.3.- Representación de los resultados .....	55
<b>8.- Aspectos económicos: costes</b>	<b>58</b>
8.1.- Precios unitarios.....	58
8.1.1.- <i>Precios unitarios por mano de obra</i> .....	58

---

8.1.2.- Precios unitarios por recurso material .....	58
8.2.- Resumen económico de recursos humanos .....	59
8.3.- Resumen económico de materiales .....	60
8.4.- Resumen económico final .....	61
<b>9.- Conclusiones</b> .....	<b>62</b>
<b>10.- Bibliografía</b> .....	<b>63</b>
<b>11.- Anexo I: Normativa aplicable. Protección de datos</b> .....	<b>65</b>
<b>12.- Anexo II: Resultados</b> .....	<b>66</b>
<i>Resultados de la periodicidad</i> .....	66
<i>Resultados de la regularidad de la fonación</i> .....	68
<i>Resultados del ruido de frecuencia alta</i> .....	70
<i>Resultados de la voz rota</i> .....	72
<i>Resultados de la velocidad del habla</i> .....	74

## Resumen

Tras una laringectomía total, la calidad de la voz del paciente se degrada considerablemente, obteniendo como resultado lo que se denomina "voz alaríngea". Los pacientes de este tipo de cirugías necesitan asistir a un logopeda para la rehabilitación de la voz, aprendiendo y entrenando la forma de hablar de nuevo.

En este proyecto se presenta una herramienta para la evaluación de voces alaríngeas. Para ello se realiza un análisis de la voz a estudiar evaluando ciertos parámetros característicos de este tipo de voces. En base a los parámetros a analizar, se ha diseñado la herramienta para posteriormente desarrollarla y finalmente comprobar su correcto funcionamiento con voces alaríngeas reales. La herramienta también representa los resultados para conseguir una visión rápida e intuitiva de las fortalezas y defectos de la voz analizada.

**Palabras clave:** laringectomía total, rehabilitación de la voz, voz alaríngea, análisis de voz.

## Laburpena

Laringektomi total baten ondoren, gaixoaren ahots kalitatea zeharo degradatzen da, eta honen emaitza laringegabeko ahotsa da. Ahotsaren birgaitetarako, kirurgia mota hau pairatu duten pazienteek terapeutara joateko beharra dute, ahotsa entrenatuz berriro hitz egiten ikasi ahal izateko.

Proiektu honetan laringegabeko ahotsak ebaluatzeko tresna bat aurkezten da. Bertan, ahots analisi bat egiten da ahots horien zenbait parametro karakteristiko ebaluatuz. Tresna, analisi parametroak oinarritzat hartuz, diseinatu eta ondoren garatu egin da eta azkenik, honen funtzionamendu egokia egiaztatzeko, benetako laringegabeko ahotsekin probak egin dira. Pazientearen ahotsaren indarguneak eta ahuleziak modu azkar eta intuitiboan ezagutu ahal izateko tresnak eskaintzen duen beste funtzionalitatetako bat emaitzen adierazpena da.

**Hitz gakoak:** laringektomi betea, ahotsaren birgaitzea, alaringe ahotsa, ahots analisia.

---

## Abstract

After a total laryngectomy, the quality of the patient's voice is considerably degraded, getting what is known as "alaryngeal voice". Therefore, this kind of patients requires a speech therapist for voice rehabilitation, as well as to learn how to talk again.

Thus, the objective of this project is to create a tool in order to evaluate the alaryngeal voices. In this sense, to study certain voice characteristic parameters a voice analysis is required. Based on those parameters, the tool is designed to be developed and tested with real alaryngeal voices. Moreover, the tool also measures the results of the strengths and weaknesses of the voice analyzed.

**Index Terms:** total laryngectomy, voice rehabilitation, alaryngeal voice, voice analysis.

## Lista de tablas

Tabla 1. Matriz probabilidad frente a impacto .....	30
Tabla 2. Equipo de trabajo .....	31
Tabla 3. Plan de trabajo .....	31
Tabla 4. Descripción de unidades de tiempo de trabajo .....	32
Tabla 5. PT.0: Gestión del proyecto .....	32
Tabla 6. PT.1: Estudio de características de las voces esofágicas .....	32
Tabla 7. PT.2: Búsqueda de bases de datos existentes .....	32
Tabla 8. PT.3: Estudio y decisión de los parámetros a analizar .....	33
Tabla 9. PT.4: Creación de la herramienta.....	33
Tabla 10. PT.5: Mejora de la herramienta .....	33
Tabla 11. Resumen de las características de la base de datos creada .....	38
Tabla 12. Precios unitarios por mano de obra.....	58
Tabla 13. Material fungible .....	58
Tabla 14. Material amortizable.....	58
Tabla 15. Coste de recursos humanos por paquete de trabajo .....	59
Tabla 16. Resumen económico de recursos humanos .....	59
Tabla 17. Resumen económico de materiales.....	60

---

Tabla 18 Resumen del Presupuesto .....	61
Tabla 19. Valores orientativos de la periodicidad .....	67
Tabla 20. Valores orientativos de la regularidad .....	69
Tabla 21. Valores orientativos del ruido de alta frecuencia .....	71
Tabla 22. Valores orientativos de la voz rota .....	73
Tabla 23. Valores orientativos de la velocidad del habla .....	75

## Lista de figuras

Figura 1. Traqueotomía .....	12
Figura 2. Fases para realizar el sonido en pacientes laringectomizados.....	17
Figura 3. Resultado del fonetograma informatizado.....	20
Figura 4. Resultado del fonetograma de una voz sana.....	21
Figura 5. Espectrograma de una voz sana .....	22
Figura 6. Espectrograma de una voz esofágica.....	22
Figura 7. Ejemplo del resultado de la escala A4S [7] .....	26
Figura 8. Diagrama de Gantt.....	34
Figura 9. Edición de las grabaciones .....	37
Figura 10. Periodicidad del fonema a y fonema s .....	40
Figura 11. Espectro y formantes del fonema a.....	41
Figura 12. Evolución de la F0 de una mujer sana .....	42
Figura 13. Evolución de la MVF de una mujer sana.....	42
Figura 14. Evolución de la MVF en una mujer con voz esofágica.....	42
Figura 15. Espectrograma del CGD y forma de onda de una voz sana.....	44
Figura 16 Espectrograma del CGD y forma de onda de una voz esofágica .....	44
Figura 17. Espectro del CGD y la forma de onda de una voz sana diciendo “asa”	46

Figura 18. Espectro del CGD y la forma de onda de una voz esofágica diciendo “sun” .....	46
Figura 19. LTAS de una voz sana (azul) y una voz esofágica (verde) .....	48
Figura 20. Detección de voz rota en voz sana .....	50
Figura 21. Detección de voz rota en voz esofágica.....	50
Figura 22. SR de una voz esofágica .....	53
Figura 23. SR de una voz sana.....	54
Figura 24. Evaluación de una voz esofágica.....	56
Figura 25. Tabla proporcionada por la herramienta .....	56
Figura 26. Evaluación de tres señales de la misma voz esofágica .....	57
Figura 27. Evaluación de tres señales de la misma voz sana .....	57
Figura 28. Histograma de la periodicidad .....	67
Figura 29. Histograma de la regularidad.....	68
Figura 30. Histograma del ruido de frecuencia alta .....	70
Figura 31. Histograma de la voz rota .....	72
Figura 32. Histograma de la velocidad del habla.....	74

---

## Lista de acrónimos

**SPL** Sound pressure level

**A4S** Automatic Acoustic Assessment of Alaryngeal Speech

**SRH** Summation of the Residual Harmonics

**MVF** Maximum Voices Frequency

**CGD** Chirp Group Delay

**HF** High frequency

**LP** Linear Prediction

**VAD** Voice activity detection

**SNR** Signal to Noise ratio

## Memoria

### 1.- Objetivos y alcance

La voz humana se forma cuando el aire producido por los pulmones pasa por la laringe donde un esqueleto cartilaginoso permite el paso del aire. En este esqueleto hay unas paredes internas con dos repliegues llamados cuerdas vocales. Éstas, para producir la voz, se mueven haciendo mayor o menor el espacio existente entre ellas. En el momento en el que el aire pasa a través de las cuerdas vocales, éstas se ponen en vibración, produciendo de esta forma el sonido.

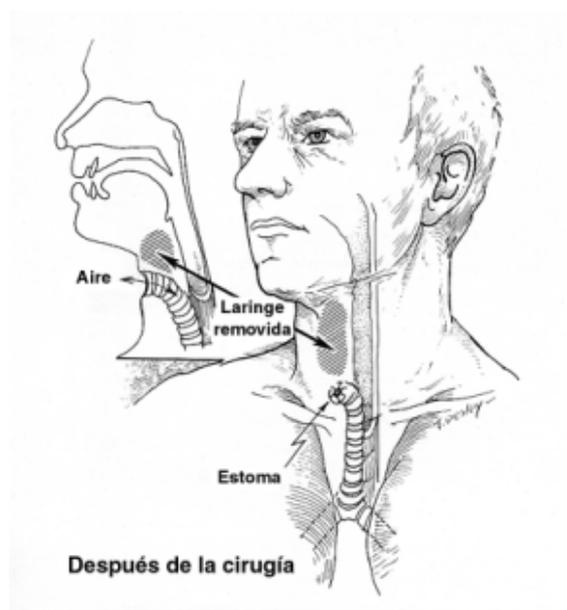


Figura 1. Traqueotomía

Las voces alaríngeas son las voces de pacientes que han sido sometidos a una laringectomía total. Este tipo de cirugía se dirige a todas las formas de cáncer rechazados por la cirugía parcial o radioterapia. Básicamente supone una extirpación de la laringe, órgano encargado de la fonación, perdiendo así las cuerdas vocales. En esta cirugía, la tráquea se une a un hueco, llamado estoma,

que se hace en la piel frente al cuello para que el paciente respire, en lugar de hacerlo a través de la boca o de la nariz, como se puede ver en la figura 1.

Por ello es necesario una reeducación de la voz, trabajando estrechamente junto con los terapeutas del lenguaje.

Hay tres condiciones fundamentales hacia las que puede derivar un laringectomizado:

1. Colocarle una prótesis fonatoria en el propio acto quirúrgico.
2. Proveerle de una prótesis electrónica externa.
3. Enseñarle una pseudo voz esofágica por medio de ejercicios foniátricos.

El objetivo principal de este proyecto es el de realizar una herramienta sencilla con la que analizar las voces esofágicas de los pacientes y evaluarlas de forma objetiva. Esta herramienta permitirá analizar las fortalezas y defectos de la voz del paciente y hacer un seguimiento de su evolución.

Para llegar al objetivo final, se deben realizar unos pasos intermedios que constituyen los objetivos secundarios.

1. Revisión bibliográfica: Como primer objetivo secundario se analizará la literatura sobre las características propias de las voces de tipo alaríngeas, características que las diferencian de las voces sanas.
2. Obtención de la base de datos: Para el análisis de las características de las voces alaríngeas, se buscará una base de datos con voces de pacientes laringectomizados y de personas sanas. Idealmente se buscará una base de datos con voces tanto masculinas como

femeninas, de distintas edades y donde tanto los pacientes sin laringe como las personas sanas digan el mismo texto. De esta forma se podrá realizar un análisis más exacto comparando frases idénticas. En el caso de no encontrar una base de datos con el mismo contenido, se utilizarán grabaciones de duraciones similares.

3. Selección y estudio de los parámetros para la evaluación de voces alaríngeas: Se estudiarán las opciones analizadas en el punto 1 y se decidirán los parámetros claves del estudio para la realización de la herramienta. Una vez decididos los parámetros a estudiar, se realizará la investigación pertinente de cada uno de ellos para el análisis de las voces alaríngeas.
4. Desarrollo de la herramienta de análisis de voz alaríngea: Teniendo los parámetros a estudiar decididos y la forma en la que se analizarán, se procederá a la realización de la herramienta utilizando una herramienta software de análisis matemático. Con una herramienta preliminar terminada, se comenzará con la realización de distintas pruebas y mejoras de la misma.

## **2.- Beneficios del proyecto**

En esta sección se describen los beneficios técnicos, económicos y sociales que se presentan durante la realización del presente proyecto. Los beneficios que se pretenden obtener mediante el desarrollo de este TFM se centran en el ámbito de la mejora de la evaluación objetiva de las voces alaríngeas.

El beneficio principal de la realización de la herramienta resultante de este proyecto es que ayuda a evaluar las capacidades orales del paciente laringectomizado, midiendo de forma objetiva distintos factores de la voz. Gracias a ella el logopeda evalúa al paciente de una forma rápida y más exacta, y por lo tanto, se consigue que ésta facilite el trabajo del terapeuta. Es esencial conseguir un método para analizar las voces con estas características ya que, hasta ahora, la evaluación se realizaba de manera subjetiva en base a la experiencia del terapeuta, de tal forma que el trabajo del logopeda era costoso.

Con este proyecto se analizan distintos aspectos de la voz pudiendo realizar una evaluación más amplia y así detectando correctamente las debilidades del habla del paciente. Con los resultados obtenidos de la herramienta, el logopeda también puede realizar una comparación de la voz actual con voces del mismo paciente de sesiones anteriores. Por lo que, otro de los beneficios de la realización de este proyecto es que permite hacer una comparativa y hacer así un seguimiento de la evolución del paciente. De esta forma se facilita la detección de aspectos que no mejoran en el habla del paciente y el logopeda puede actuar en base a ello.

De todo ello se podría deducir que esta herramienta podría llegar a ser fundamental en el trabajo del profesional sanitario en relación al supuesto tratado, de forma que haría posible poder atender de una manera más rápida y eficiente a un mayor número de pacientes sin perder por ello la calidad y el trato personalizado. Económicamente, al reducir tiempo y agilizar el proceso, se

ahorrarían costos y a su vez se tratarían más pacientes, lo que aumentaría los ingresos por este concepto obteniendo un beneficio económico para el terapeuta o para el centro hospitalario.

Un beneficio técnico indirecto de este TFM es la del estudio de la literatura existente y la investigación que ha sido necesaria para su realización, consiguiendo nuevos métodos para el análisis de la señal de voz.

Otro beneficio derivado de este proyecto, es que, en un futuro, se podría realizar una aplicación para el móvil, de tal forma que, además de ser una herramienta para el logopeda, puede llegar a ser una forma de que el paciente pueda evaluarse a sí mismo sin necesidad de acudir al hospital. De ninguna de las formas podrá sustituir al trabajo que realiza el médico, pero puede ser una manera de que el paciente evalúe si los ejercicios que realiza en casa para la mejora del habla, están dando sus frutos o no. Simplemente como una herramienta de apoyo.

Por lo que, en resumen, este TFM aporta innumerables beneficios en el ámbito de la medicina a la hora de evaluar y de realizar un seguimiento al paciente. Ya que al hacerlo de forma más amplia, más exacta y objetiva, se consigue un análisis más completo y más rápido, pudiendo llegar a suponer una reducción de visitas del paciente al logopeda.

### 3.- Estado del arte

En este apartado se describe el estado del arte de este TFM. Para comenzar se explica cómo consiguen los pacientes reeducar las voces al perder la laringe y las cuerdas vocales, los problemas que conlleva la laringectomía total, y para finalizar, los métodos de evaluación existentes hasta la fecha, tanto métodos de medida subjetiva como objetiva.

#### 3.1.- Entrenamiento de la voz esofágica

En el caso de los pacientes laringectomizados, existen métodos para recuperar la capacidad del habla, como es, por ejemplo, la erigmofofonía, consiguiendo un tipo de voz conocida como voz esofágica. El objetivo para el paciente es encontrar una fuente de producción de sonido que le permita comunicarse. La erigmofofonía consiste en la introducción de aire en el esófago y luego permitir que sea expulsado de forma controlada produciendo un tono que se utiliza como fuente sonora y que será moldeado por el tracto supravocal al igual que en las personas sin la laringectomía [1]. Es decir, es una técnica que consiste en tragar aire hacia el esófago para posteriormente echarlo hacia afuera como se puede ver en la figura 2. Es algo parecido a eructar, pero controlando los sonidos que se realizan durante el mismo gracias al entrenamiento.



Figura 2. Fases para realizar el sonido en pacientes laringectomizados

### **3.2.- Problemas asociados con la voz esofágica o erigmofónica**

Cuando los pacientes están aprendiendo a hablar pueden adoptar ciertas posturas o hábitos que entorpecen la comunicación y deben ser corregidos para evitar que se hagan costumbre [1].

a. Excesivo ruido en el estoma: el movimiento de aire forzado para la comunicación hace que, sobre todo, durante la espiración haya un ruido fuerte y que sea menos inteligible el habla de laringectomizado.

b. Ruido al introducir aire hacia el aparato digestivo: al intentar una inyección rápida y fuerte de aire, se escucha un ruido al tragar.

c. Pausas al hablar: a medida que el paciente va adquiriendo facilidad para la erigmofonía, la reducción de la latencia entre la ingestión de aire y la fonación debe ir disminuyendo paulatinamente.

### **3.3.- Evaluación de la voz esofágica**

En el momento de evaluar si la rehabilitación del paciente es adecuada o no, se debe cuestionar si: el paciente utiliza de forma constante la voz para comunicarse, si alcanzó un desarrollo completo de la voz pero la utiliza de forma intermitente o si definitivamente no ha sido posible aprender a utilizar la voz como medio de comunicación.

En la evaluación perceptual o subjetiva de la voz erigmofónica, sólo se puede tener en cuenta el grado de la inteligibilidad de la misma, parámetro que se evalúa dependiendo de la experiencia del evaluador.

Los parámetros que se buscan a la hora de analizar la evolución de una voz esofágica son los siguientes [1] [2]:

- Para producir la fonación, la absorción de aire ha de ser muy rápida para que el tiempo de latencia entre la absorción y la fonación sea muy corto como se ha comentado anteriormente.
- Por cada toma de aire se ha de producir entre cuatro y nueve sílabas.
- Que las palabras por minuto lleguen a ser entre 80 y 130.
- Que la frecuencia fundamental se mantenga entre 52 y 80 hercios.
- Y que la intensidad media no sea inferior a los 15-18 dB sobre la voz normal.

### **Prótesis fonatoria**

Como se ha explicado anteriormente, existe la posibilidad de colocar una prótesis fonatoria a los pacientes laringectomizados. Los beneficios son que estos pacientes consiguen una mayor inteligibilidad que aquellos que no la usan. Las diferencias más importante entre la voz esofágica con y sin prótesis fonatoria son la velocidad el habla (*nº de sílabas / tiempo de fonación (s)*), el tiempo de fonación y la intensidad del habla. Las tres variables se relacionan con la inteligibilidad y están reducidas en el habla esofágica debido a la ausencia del elemento efector pulmonar, lo que ocasiona una voz entrecortada y pausada en el laringectomizado con voz esofágica [2].

### **3.4.- Métodos de evaluación objetiva**

El análisis objetivo de la voz incluye mediciones acústicas y aerodinámicas. Las mediciones acústicas consisten en la determinación de la frecuencia fundamental, el rango de frecuencia, el jitter que es básicamente analizar las oscilaciones de la frecuencia, intensidad, rango de intensidad, el shimmer que analiza las oscilaciones de intensidad y la relación de tonos armónicos y componentes ruidosos. Por otro lado, las mediciones aerodinámicas incluyen la determinación de la fonación y los tiempos de fricción [3].



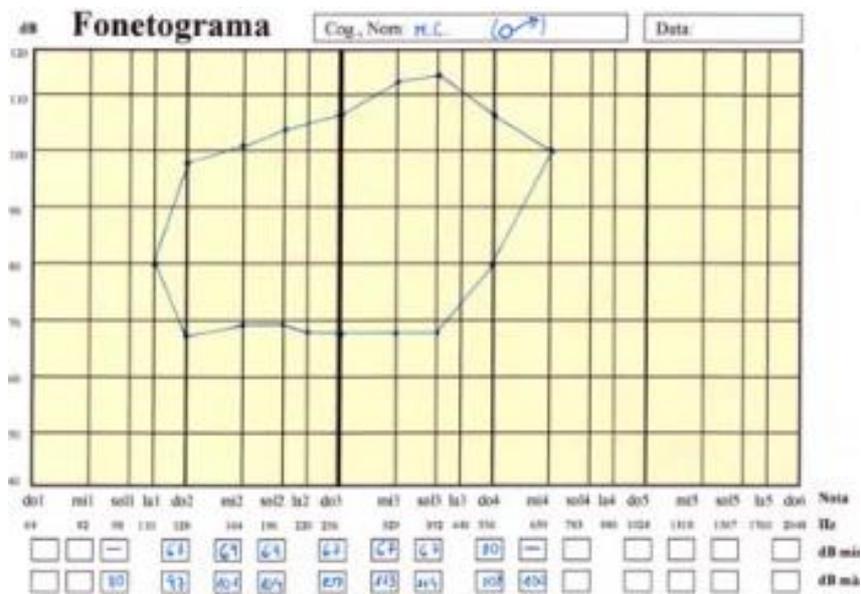


Figura 4. Resultado del fonotograma de una voz sana

Uno de los productos que utiliza esta técnica es, por ejemplo, el lingWAVES Phonetogram Plus. Este producto ofrece un registro del fonotograma estandarizado en 10 minutos. Gracias al lingWAVES Phonetogram Plus se consigue el rango de frecuencia fundamental y el rango del nivel de presión sonora de una voz, ya sea hablando o cantando [5].

### *Análisis espectrográfico*

En el caso del análisis espectrográfico, la energía de las voces esofágicas tiene una tendencia a que se concentre en todas las frecuencias, a diferencia de las voces sanas donde se concentra en áreas específicas. El ruido es un componente prácticamente constante que ensucia el trazado espectrográfico concentrándose fundamentalmente en las frecuencias más altas.

A continuación se analizan los siguientes espectrogramas, figuras 5 y 6. La parte superior de las figuras representa el registro en banda ancha, y en la parte inferior, la banda estrecha.

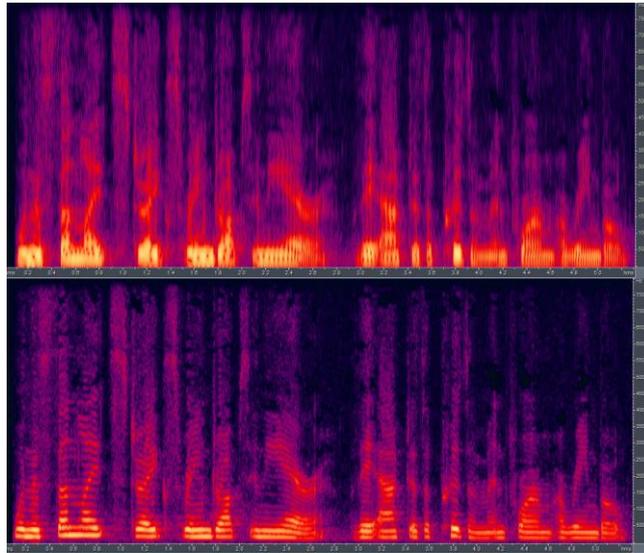


Figura 5. Espectrograma de una voz sana

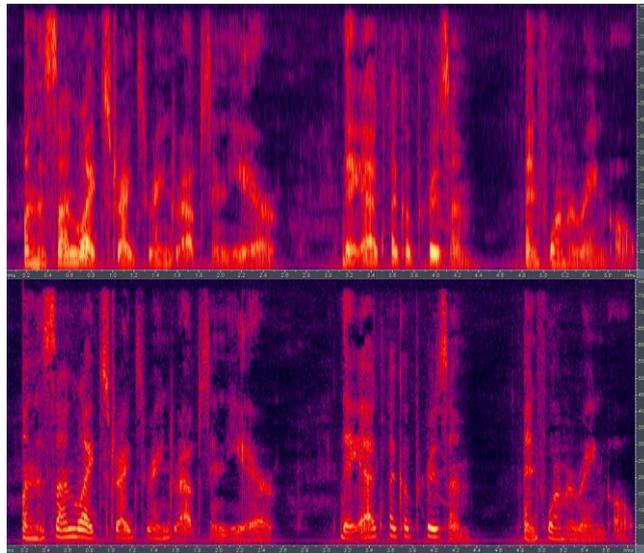


Figura 6. Espectrograma de una voz esofágica

En la figura 5, en el espectrograma de banda ancha, se pueden apreciar los espacios azules oscuros entre las estrías verticales resultantes de la vibración de las cuerdas vocales. En el caso de la banda estrecha, se ven las líneas horizontales, los armónicos, siendo el primero el fundamental.

En el caso de la figura 6, la diferencia entre los sonidos sordos y sonoros son menores, presentando pocas diferencias entre vocales y consonantes sordas. Analizando los armónicos, vemos que se encuentran más mezclados y menos nítidos que en el caso de la voz sana [6], ya que la esofágica es una voz menos periódica.

Por lo que se puede concluir que en este estudio se demuestra que el análisis espectrográfico es muy útil a la hora de analizar estos dos tipos de voces. Ya que la separación entre las estrías verticales y los armónicos se ven afectados en el caso de las voces esofágicas, siendo complicado diferenciar los tipos de sonidos e identificar los armónicos. Al analizar las zonas donde deberían de aparecer las estrías y los armónicos, aparecen zonas de energías verticales y horizontales que se corresponden con el ruido de fricción del aire. También se observan frecuentes franjas verticales que son silencios previos a la expulsión del aire y a la explosión de salida del mismo.

## **4.- Análisis de alternativas**

En este apartado se muestran las diferentes alternativas que se han analizado relacionado con diversos aspectos de este TFM para poder cumplir con los requisitos del mismo.

### **4.1.- Adquisición de la base de datos**

Inicialmente se valoran dos alternativas diferentes para la adquisición de una base de datos de voces sanas y de voces esofágicas:

- Grabación de nuestra propia base de datos.
- Utilización de grabaciones ya existentes.

#### ***4.1.1.- Grabación de la base de datos***

Una de las opciones es diseñar nuestra propia base de datos. Decidir qué tipo de grabaciones interesa para el proyecto y grabarlas con la calidad y las especificaciones técnicas deseadas.

#### ***4.1.2.- Grabaciones existentes***

Otra opción es utilizar una base de datos o partes de distintas bases de datos del laboratorio o de internet. Después se modifican las grabaciones a nuestro gusto con un programa de edición de audio.

#### ***4.1.3.- Comparativa***

Aunque en un inicio se comenzó con el proyecto con la idea de grabar la base de datos en colaboración con el hospital de Basurto, por problemas administrativos se escogió la segunda opción, utilizar grabaciones de bases de datos ya existentes. De esta forma se agiliza la obtención de las voces.

## 4.2.- Tipos de grabaciones

Una vez se ha tomado la decisión de la utilización de grabaciones de bases de datos ya existentes, es necesario realizar un segundo análisis de alternativas para decidir qué tipo de grabaciones se van a utilizar.

Existen infinidad de opciones sobre las características de las grabaciones para utilizar en este proyecto. Se tendrán en cuenta las siguientes especificadas:

- Grabaciones tanto de hombres como de mujeres.
- Que la duración de las grabaciones de voces sanas y de voces esofágicas sean parecidas: frases de aproximadamente entre 8 y 10 segundos.
- Grabaciones en formato Windows PCM.
- Grabaciones en modo monofónico.
- Frecuencia de muestreo de 16 kHz.
- Resolución de 32 bits.

En el caso de que las grabaciones encontradas no cumplan estas especificaciones, se modificarán con un programa sencillo de edición de audio.

## 4.3.- Parámetros a analizar

Una de las decisiones más importantes a tomar en este TFM es elegir los parámetros que va a analizar la herramienta creada en este proyecto.

Existen muchos parámetros con lo que se podría hacer una buena y muy completa evaluación de la voz alaríngea, pero uno de los objetivos del proyecto es que la herramienta facilite el trabajo al logopeda, por lo que no es conveniente analizar muchas características y dificultar el diagnóstico con excesivos resultados.

Por lo tanto se van a analizar los cinco parámetros de la escala A4S propuesta por Thomas Drugman [7]: la periodicidad, la regularidad, el ruido en frecuencia alta, la voz rota y la velocidad del habla. Esta escala tiene cinco dimensiones normalizadas y es representada mediante un pentágono, permitiendo una visualización rápida e intuitiva como se puede ver en la figura 7. Esta figura es un ejemplo del resultado de la evaluación propuesta por Drugman de un paciente en diferentes etapas de la terapia con el logopeda.

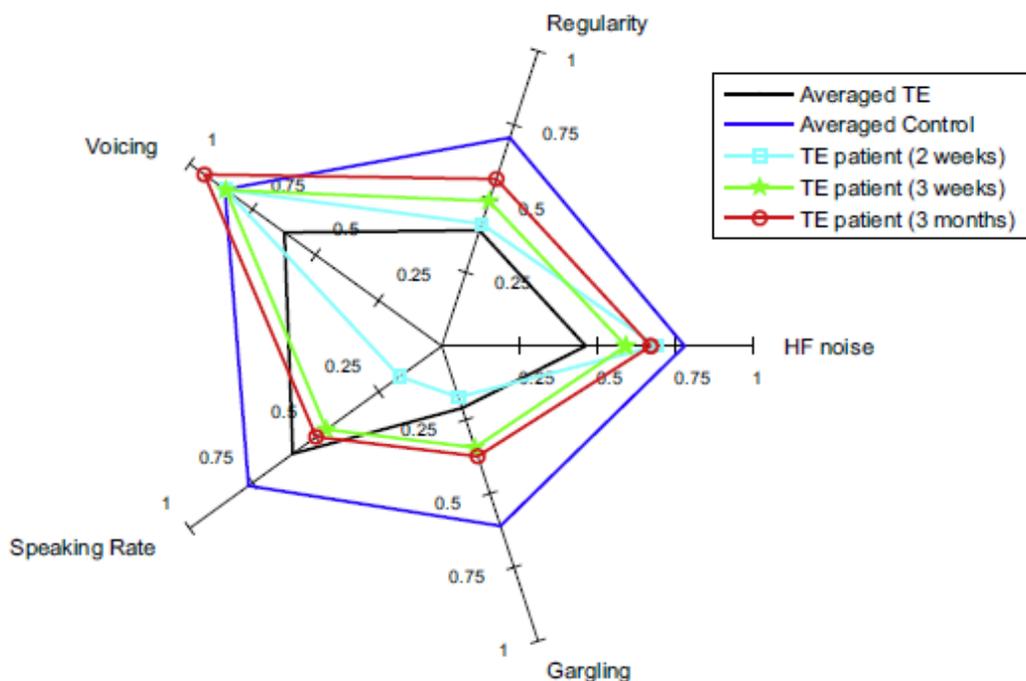


Figura 7. Ejemplo del resultado de la escala A4S [7]

El parámetro de la periodicidad será medido en base a la proporción de sonoridad (*voicing proportion*), la regularidad se medirá con el retardo de grupo del chirp (*Chirp Group Delay*), el ruido de alta frecuencia con la energía relativa en el espectro a largo plazo (*High frequency noise*), la voz rota la proporción de tramas de voz detectada como voz rota o quebrada (*creakiness*), y la velocidad del habla con la proporción de voz frente a silencio (*speaking rate*).

## 4.4.- Herramienta software para programar

El último análisis de alternativas que se debe realizar es el del software que se utilizará para la programación de la herramienta de evaluación de la voz.

A continuación se listan las diferentes alternativas:

### **4.4.1.- Matlab**

Matlab es una herramienta de computación matemática que ofrece un entorno de desarrollo integrado con su propio lenguaje de programación. Este software suele ser usado para análisis de datos, manipulación de matrices o representación de funciones, también suele ser usado como vínculo de comunicación con otros programas y otros lenguajes, incluso en programación hardware.

### **4.4.2.- GNU Octave**

GNU Octave es un lenguaje de alto nivel, dedicado principalmente a programación y computación numérica. Permite solucionar numéricamente problemas tanto lineales como no lineales, así como cualquier tipo de operación numérica. Ofrece grandes herramientas gráficas para visualización y manipulación de datos y es usado normalmente a través de una línea de comandos interactiva. El lenguaje es extremadamente parecido a Matlab.

### **4.4.3.- Genius**

Se trata de un software de cálculo de propósito general, es similar en muchos aspectos a Matlab, Maple o Mathematica. Puede ser usada tanto como una simple calculadora como una herramienta de cálculo numérico. La sintaxis es intuitiva y está diseñado pensando en el lenguaje de las matemáticas.

#### *4.4.4.- Comparativa*

Puesto que todas ellas tienen características similares, se ha decidido utilizar la herramienta de cálculo software de Matlab por su simplicidad y porque cumple todas las necesidades que se requiere para la realización de este proyecto. Otra de las razones por las que se ha decidido utilizar este software es que hay ciertas funciones específicamente desarrolladas en abierto para analizar voces que sufren los problemas explicados anteriormente.

## 5.- Análisis de riesgos

En todo proyecto es de suma importancia analizar al comienzo los riesgos que existen.

Los factores que se han tenido en cuenta son:

- **Probabilidad de ocurrencia:** probabilidad de que el suceso ocurra.
- **Impacto:** efecto que ello conllevaría en el sistema y sus repercusiones.

Estos aspectos se medirán en tres escalas: baja, media y alta. Se reparte de la siguiente forma:

- **Baja/Bajo:** Son aquellos sucesos con poca probabilidad y/o bajo impacto.
- **Media/Medio:** Son aquellos sucesos con probabilidad media y/o medio impacto.
- **Alta/Alto:** Son aquellos sucesos con probabilidad alta y/o alto impacto.

A continuación se detallan los riesgos iniciales que hay que tener presentes para poder resolverlos o intentar evitarlos lo antes posible:

1. Mala elección de la base de datos. Puede ocurrir que las voces escogidas para comprobar el correcto funcionamiento de la herramienta desarrollada en este proyecto no sean representativas, es decir, no sean lo suficientemente variadas en cuanto a género, nivel de mejora... La probabilidad de que esto ocurra es media ya que se va a trabajar con voces ya existentes y sin poder escoger entre mucha variedad. Por otro lado el impacto es bajo ya que los resultados se exponen en función de la base de

datos completa que se utiliza, evaluando la voz en comparación con el resto. Por lo que a medida que aumente esta base de datos, los resultados se reescalarán y los resultados variarán a medida que se consiga una base de datos más completa.

2. Mala elección de los parámetros a analizar. Otro de los riesgos a tener en cuenta es que no se haya realizado una correcta elección de los parámetros para evaluar las voces. La probabilidad de que esto ocurra es baja ya que existen varios trabajos donde hablan sobre las características de este tipo de voces y se han escogido unos parámetros que evalúan dichas características. El impacto también sería bajo ya que aunque algún parámetro no sea de todo concluyente, se evalúan más parámetros por lo que se hace un análisis variado e independiente entre ellos.

Después de analizar cada uno de los riesgos de manera independiente, se muestran en la siguiente tabla para analizar la probabilidad e impacto de los diferentes riesgos de este proyecto enfrentados.

		Impacto		
		Bajo	Medio	Alto
Probabilidad	Baja	Riesgo 2		
	Media	Riesgo 1		
	Alta			

Tabla 1. Matriz probabilidad frente a impacto

En la tabla 1 se ve gráficamente la relación de los diferentes riesgos entre probabilidad e impacto.

## Metodología

### 6.- Descripción de tareas. Gantt

#### 6.1.- Equipo de trabajo

<b>Id.</b>	<b>Nombre</b>	<b>Responsabilidad</b>
P1	Eva Navas Cordón	Director del proyecto
P2	Ainhoa Larrea Murga	Ingeniero Junior

Tabla 2. Equipo de trabajo

#### 6.2.- Plan de trabajo

<b>Paquete, Tarea o Hito</b>	<b>Duración</b>
PT.0: Gestión del proyecto	28 Semanas
T.0.1: Gestión y seguimiento del proyecto	26 Semanas
T.0.2: Elaboración de la documentación	4 Semanas
H.6: Fin del proyecto	
PT.1: Estudio de las características propias de las voces esofágicas	3 Semanas
T.1.1: Búsqueda y recopilación de la literatura existente	1 Semana
T.1.2: Lectura y análisis de las características de las voces esofágicas	3 Semanas
H.1: Características de las voces definidas	
PT.2: Búsqueda de bases de datos existentes	1 Semana
T.2.1: Búsqueda y recopilación de bases de datos	2 Días
T.2.2: Análisis y edición de las grabaciones adquiridas	3 Días
H.2: Base de datos completada	
PT.3: Estudio y decisión de parámetros a analizar	4 Semanas
T.3.1: Estudio de parámetros a analizar ya existentes	3 Semanas
T.3.2: Selección de parámetros a analizar	1 Semana
H.3: Parámetros para analizar con la herramienta decididos	
PT.4: Creación de la herramienta	15 Semanas
T.4.1: Creación de la herramienta preliminar	15 Semanas
H.4: Herramienta preliminar completada	
PT.5: Mejora de la herramienta	3 Semanas
T.5.1: Pruebas y mejoras de la herramienta	3 Semanas
H.5: Herramienta completada	

Tabla 3. Plan de trabajo

Unidad	Descripción
Proyecto	28 Semanas (21 ECTS)
Semana	5 Días
Día	4 - 8 Horas

Tabla 4. Descripción de unidades de tiempo de trabajo

### 6.3.- Descripción de paquetes de trabajo

PT.0: Gestión del proyecto	Inicio: 29-02-16	Fin: 27-09-16
<p>Para la correcta realización del proyecto es necesario realizar una gestión y un seguimiento del proyecto para posteriormente poder completar un informe del mismo.</p> <p><b>Recursos técnicos:</b> MH1, MH2 y MH3; MS1, MS2, MS3 y MS4</p>		
T.0.1: Gestión y seguimiento del proyecto		
Tiempo estimado: 26 Semanas	Responsable: P1 (50h)	
T.0.2: Elaboración de la documentación		
Tiempo estimado: 4 Semanas	Responsable: P2 (80h)	Revisión: P1(8h)

Tabla 5. PT.0: Gestión del proyecto

PT.1: Estudio de características de las voces esofágicas	Inicio: 29-02-16	Fin: 20-03-16
<p>Con el fin de obtener el resultado deseado a la hora de finalizar el proyecto, se realiza un estudio de las características propias de las voces esofágicas, analizando lo que las diferencia de las voces de las personas sanas.</p> <p><b>Recursos técnicos:</b> MH1, MH2 y MH3; MS1</p>		
T.1.1: Búsqueda y recopilación de la literatura existente		
Tiempo estimado: 1 Semana	Responsable: P2 (5h)	
T.1.2: Lectura y análisis de las características de las voces esofágicas		
Tiempo estimado: 3 Semanas	Responsable: P2 (60h)	Revisión: P1 (5h)

Tabla 6. PT.1: Estudio de características de las voces esofágicas

PT.2: Búsqueda de bases de datos existentes	Inicio: 21-03-16	Fin: 27-03-16
<p>Se buscan varias bases de datos de voces esofágicas y voces sanas creando una base de datos de las características deseadas. Se editan las grabaciones existentes para adecuarlas a las necesidades del proyecto.</p> <p><b>Recursos técnicos:</b> MH1 y MH3; MS1 y MS4</p>		
T.2.1: Búsqueda y recopilación de bases de datos existentes		
Tiempo estimado: 2 Días	Responsable: P2 (8h)	
T.2.2: Análisis y edición de las grabaciones adquiridas		
Tiempo estimado: 3 Días	Responsable: P2 (12h)	Revisión: P1 (2h)

Tabla 7. PT.2: Búsqueda de bases de datos existentes

<b>PT.3: Estudio y decisión de los parámetros a analizar</b>	Inicio: 28-03-16	Fin: 24-04-16
<p>Una vez estudiadas las características de las voces y las grabaciones, se procede a investigar los parámetros que se analizan este tipo de voces y a estudiarlos. Después se decide cuáles son los parámetros a analizar en este proyecto para crear una herramienta en base a ello.</p> <p><b>Recursos técnicos:</b> MH1 y MH2; MS1, MS2 y MS3</p>		
T.3.1: Estudio de parámetros a analizar ya existentes		
Tiempo estimado: 4 Semanas		Responsable: P2 (60h)
T.3.2: Decisión de parámetros a analizar		
Tiempo estimado: 1 Semana		Responsable: P2 (16h)      Revisión: P1 (8h)
Tabla 8. PT.3: Estudio y decisión de los parámetros a analizar		

<b>PT.4: Creación de la herramienta</b>	Inicio: 25-03-16	Fin: 7-08-16
<p>La mayor carga de trabajo en este proyecto reside en la realización de la herramienta preliminar mediante el software de análisis matemático.</p> <p><b>Recursos técnicos:</b> MH1; MS1, MS2 y MS3</p>		
T.4.1: Creación de la herramienta preliminar		
Tiempo estimado: 15 Semana		Responsable: P2 (300h)      Revisión: P1 (15h)
Tabla 9. PT.4: Creación de la herramienta		

<b>PT.5: Mejora de la herramienta</b>	Inicio: 8-08-16	Fin: 25-09-16
<p>Una vez conseguida una herramienta preliminar, se realizan las pruebas y mejoras de las herramientas pertinentes para cumplir con los objetivos del proyecto.</p> <p><b>Recursos técnicos:</b> MH1; MS1, MS2 y MS3</p>		
T.5.1: Pruebas y mejoras de la herramienta preliminar		
Tiempo estimado: 3 Semana		Responsable: P2 (60h)      Revisión: P1 (10h)
Tabla 10. PT.5: Mejora de la herramienta		

## 6.4.- Diagrama Gantt

En la figura 8 se puede ver el diagrama de Gantt. Mediante este diagrama se representa la organización de las tareas de este TFM de una forma más visual.

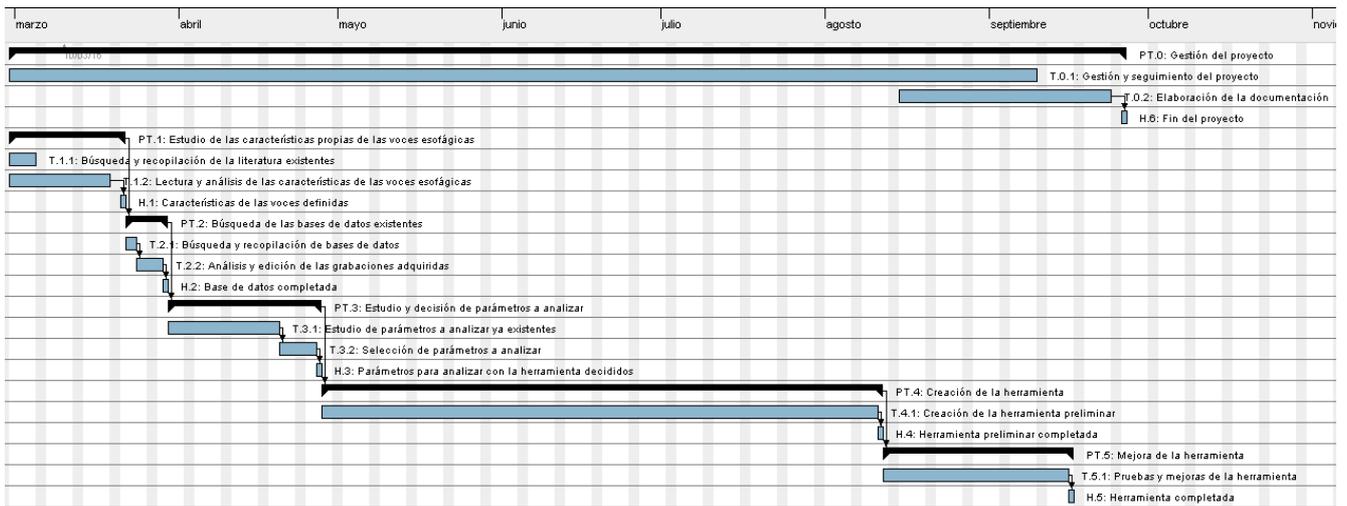


Figura 8. Diagrama de Gantt

## **7.- Descripción de la solución**

En este apartado se explica el proceso que ha sido necesario para llevar a cabo este TFM. Para la explicación de la realización de la herramienta de diagnóstico y evaluación para voces alaríngeas, este apartado se ha dividido en tres grandes bloques: creación de la base de datos, creación de la herramienta y representación de los resultados.

### **7.1.- Creación de la base de datos**

Una vez tomada la decisión de que no se grabaría la base de datos, sino que se utilizarían grabaciones ya existentes de personas sanas y personas con voz esofágica, se procede a la búsqueda y la edición de grabaciones para la creación de una base de datos para la posterior prueba del correcto funcionamiento de la herramienta creada.

#### ***7.1.1.- Búsqueda de bases de datos existentes***

Como se explica en el apartado de *Objetivos y alcance*, la idea inicial es la de buscar una base de datos donde las voces sanas y las voces esofágicas digan el mismo texto, que sean voces tanto masculinas como femeninas y de distintas edades. Finalmente se han descargado de diferentes páginas webs [8] [9], dos bases de datos que cumplen los requisitos mencionados. Estas grabaciones leen el siguiente texto en inglés:

*“When the sunlight strikes raindrops in the air, they act as a prism and form a rainbow. The rainbow is a division of white light into many beautiful colors. These take the shape of a long round arch, with its path high above, and its two ends apparently beyond the horizon. There is, according to legend, a boiling pot of gold at one end. People look, but no one ever finds it. When a man looks for something*

*beyond his reach, his friends say he is looking for the pot of gold at the end of the rainbow.”*

Lo que traducido al castellano dice:

*“Cuando la luz del sol golpea las gotas de agua en el aire, actúan como un prisma y forman un arco iris. El arco iris es una división de la luz blanca en muchos colores preciosos. Éstos toman la forma de un largo arco redondo, pasando por encima, y sus dos extremos aparentemente más allá del horizonte. Según la leyenda, hay una olla de oro en un extremo. La gente mira, pero nadie lo encuentra. Cuando un hombre busca algo fuera de su alcance, sus amigos dicen que está buscando la olla de oro al final del arco iris.”*

Antes de editar las grabaciones se analizan para determinar si son válidas para este proyecto. Aunque ambas bases de datos cumplen los requisitos, en el caso de las grabaciones de las voces sanas se puede apreciar un ruido de fondo que seguramente perjudicará los resultados. Por lo que, después de intentar eliminarlo, se decide buscar otras grabaciones.

Al no encontrar otra base de datos de voces sanas leyendo el mismo texto, se toma la decisión de utilizar una base de datos muy conocida de voces sanas adquirida de las bases de datos del laboratorio AHOLAB, la CMU Artic.

La base de datos CMU Artic fue diseñada con el propósito de investigar la síntesis de voz. Esta base de datos ha sido cuidadosamente registrada bajo condiciones de estudio y constan de cerca de 1.150 expresiones en inglés fonéticamente equilibradas. Se distribuyen como software libre, sin restricciones sobre el uso comercial o no comercial [10]. Esta base de datos no dice el mismo texto que la base de datos de voces esofágicas, pero cumple el resto de requisitos.

### 7.1.2.- Edición de las bases de datos

Utilizando el programa de edición de audio Audacity, se modifican las grabaciones para que cumplan las características deseadas mencionadas en el apartado *Análisis de alternativas*, como se puede ver en la figura 9. Audacity es un editor de audio gratuito con el que se puede grabar, reproducir, editar, importar y exportar archivos.

El primer paso es editar todas las grabaciones de estéreo a monofónico, con una frecuencia de muestreo de 16 kHz y comprobar que están con una resolución de 32 bits, características especificadas en *Análisis de alternativas*.

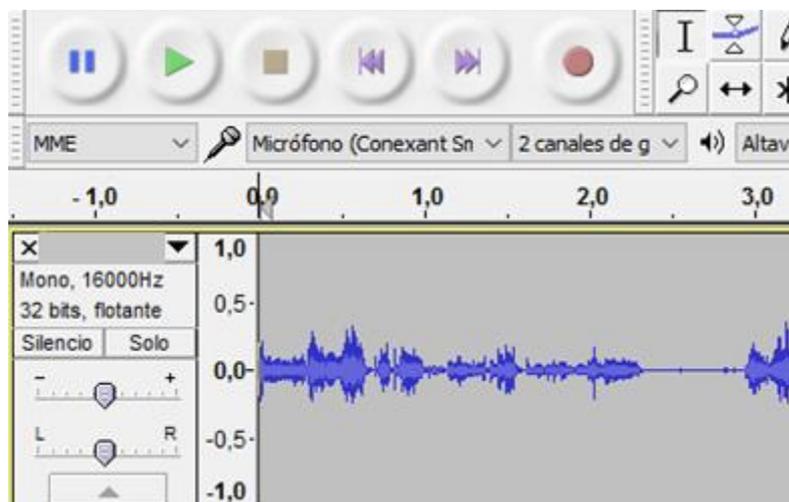


Figura 9. Edición de las grabaciones

La base de datos de las voces esofágicas tiene una resolución de 44,1 kHz y está formada por una única grabación por persona que lee el texto completo. Para que cumpla las características deseadas, se decide cortar las grabaciones creando un archivo por cada frase, obteniendo como resultado seis grabaciones por persona.

Al cortar las grabaciones por frases, se han respetado los silencios anteriores y posteriores de cada una de las frases. En el caso de que al crear la

herramienta se prefieran frases sin silencios iniciales o finales, se eliminarán de la misma forma que se han cortado las grabaciones con el programa de edición Audacity.

En el caso de la base de datos de la CMU Artic todas las grabaciones tienen una frecuencia de muestreo de 16 kHz y una resolución de 32 bits. Por lo que no ha sido necesario realizar ninguna modificación, simplemente comprobar que cumplían los requisitos.

**7.1.3.- Base de datos resultante**

En la tabla 11 se resume las características de la base de datos creada para este proyecto con las grabaciones de las dos bases de datos anteriores.

	Grabaciones esofágicas	Grabaciones sanas
<b>Nº total de grabaciones</b>	48	63
<b>Nº de hablantes</b>	8	7
<b>Nº de grabaciones por persona</b>	6	9
<b>Nº de mujeres</b>	7	2
<b>Nº de hombres</b>	1	5
<b>Duración de las grabaciones</b>	Entre 5 y 10 segundos	Máximo 6 segundos
<b>Tiempo de voz total</b>	161 segundos	330 segundos

Tabla 11. Resumen de las características de la base de datos creada

## 7.2.- Creación de la herramienta

Una vez analizada la literatura existente sobre las características y los parámetros que evalúan las voces esofágicas, se decide analizar los parámetros mencionados en el apartado *Análisis de alternativas*: la periodicidad, la regularidad, el ruido en frecuencia alta, la voz rota y la velocidad del habla.

A continuación se describen los cinco parámetros y se explica de qué forma se van a cuantificar los resultados.

### **7.2.1.- La periodicidad**

En la producción de la voz existen diferentes tipos de clasificaciones para los sonidos emitidos por una persona. Una de esas clasificaciones divide los sonidos en sonidos sonoros y sordos.

Los sonidos sonoros son producidos cuando la glotis se encuentra cerrada y el aire a su paso hace vibrar las cuerdas vocales que se abren y cierran de forma periódica. De esta forma se obtiene un flujo de aire periódico con una frecuencia fundamental ( $F_0$ ).

Los sonidos sordos son producidos con la glotis abierta y las cuerdas vocales relajadas. Por ellos el aire fluye sin obstáculos convirtiéndose en un aire turbulento. No hay vibración de las cuerdas vocales por lo que no existe ninguna periodicidad en este caso.

En la figura 10 se puede apreciar la periodicidad de la vocal a, sonido sonoro (parte superior de la figura) y la aperiodicidad de la consonante s, sonido sordo (parte inferior de la figura).

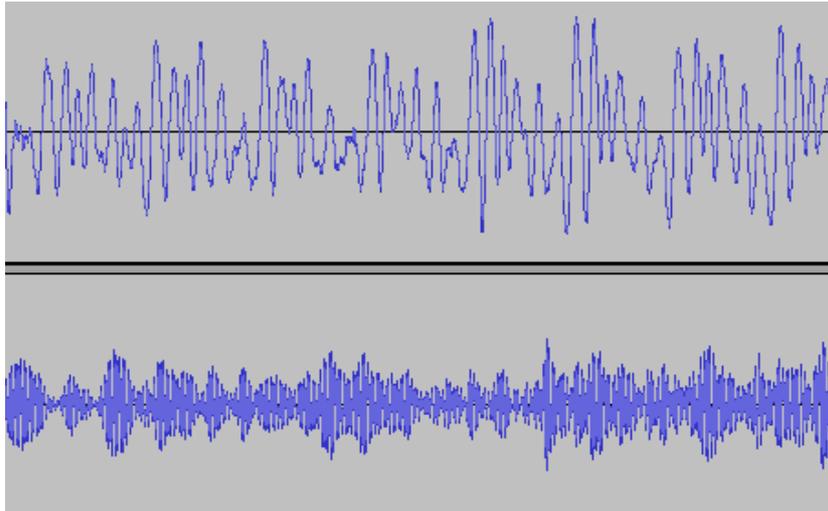


Figura 10. Periodicidad del fonema a y fonema s

A la hora de analizar las diferencias entre las voces sanas y las voces esofágicas es necesario estudiar la periodicidad, ya que las voces esofágicas son menos periódicas que las voces sanas [7]. Esto es resultado de la extracción de las cuerdas vocales en la cirugía y por lo tanto no pueden modular el aire cuando pasa por la glotis.

En este trabajo, el análisis de la periodicidad se basa en la suma de los armónicos residuales (*Summation of the Residual Harmonics*, SRH) [11]. SRH primero estima la envolvente espectral mediante el análisis de predicción lineal convencional y luego filtra inversamente la señal de voz. SRH no sólo estima la F0, sino también el nivel de periodicidad de la señal de voz.

El SRH es calculado mediante la fórmula (1) siendo  $E(f)$  la amplitud de la envolvente y  $N_{harm}$  el número de armónicos:

$$SRH(f) = E(f) + \sum_{k=2}^{N_{harm}} \left[ E(k \cdot f) - E\left(\left(k - \frac{1}{2}\right) \cdot f\right) \right] \quad (1)$$

Para poder evaluar de forma objetiva una característica de la voz esofágica, es necesario pensar cómo cuantificarla. En este TFM la periodicidad se cuantifica con la frecuencia máxima de la voz (*Maximum Voiced Frequency, MVF*). La frecuencia máxima es la línea que separa en dos el espectro de la voz en función de la periodicidad. Al analizar el espectro de la voz se obtienen dos grupos: componentes periódicos que se aprecian en las frecuencias bajas y componentes aperiódicos en las altas [12].

Como se puede ver en la figura 11, el espectro de un sonido sonoro va perdiendo la periodicidad a medida que aumenta la frecuencia, la MVF es la barrera entre la zona periódica y aperiódica, en este caso cercano a 2 kHz (marcado con un cuadrado rojo).

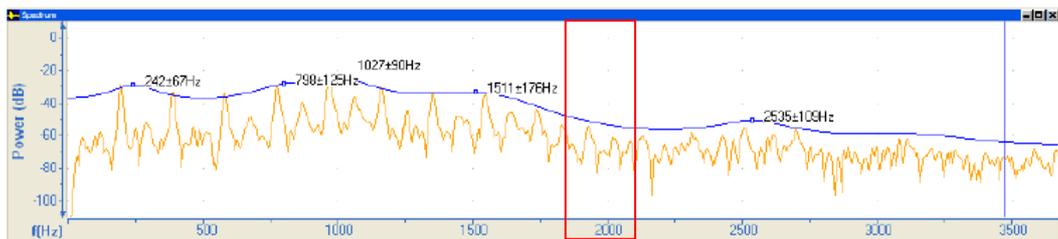


Figura 11. Espectro y formantes del fonema a

Para la medición de la MVF se utiliza un Toolbox de Matlab ya existente, “the GLOAT Toolbox”, creado por Thomas Drugman. Una de las funciones del Toolbox realiza la estimación de la MVF y de la F0, devolviendo dos vectores con los valores de éstas cada 10 ms. Después, se realiza una media de las partes sonoras de este vector, ya que los sonidos sordos y los silencios no tienen F0. Para detectar cuáles son las partes sonoras, la herramienta utiliza el vector de la estimación de la F0 que indica las partes sonoras. De esta forma se obtiene un único valor por audio para evaluar su periodicidad.

En las figuras 12 y 13 se puede ver la evolución y la media de la F0 y la MVF en una de las grabaciones de una mujer sana.

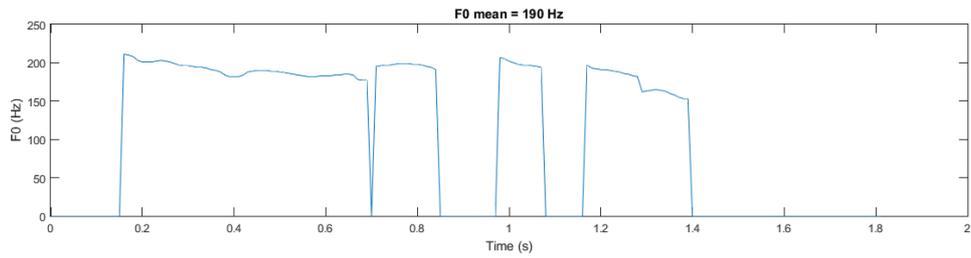


Figura 12. Evolución de la F0 de una mujer sana

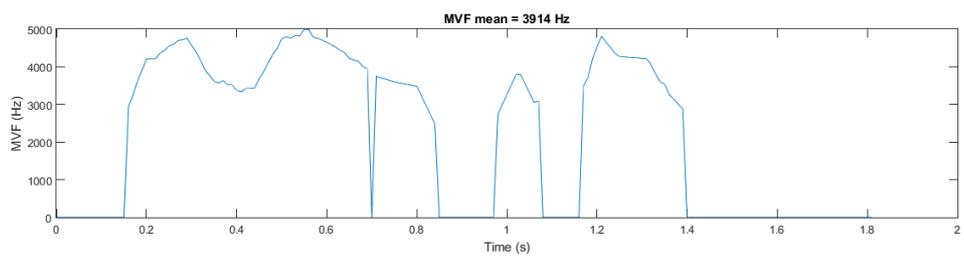


Figura 13. Evolución de la MVF de una mujer sana

Como se ha mencionado antes, las voces esofágicas son menos periódicas que las sanas, lo que supone una menor MVF ya que tienen menos componentes periódicas en su espectro. Esto se demuestra en la figura 14 donde se representa la evolución de la MVF de una mujer con voz esofágica.

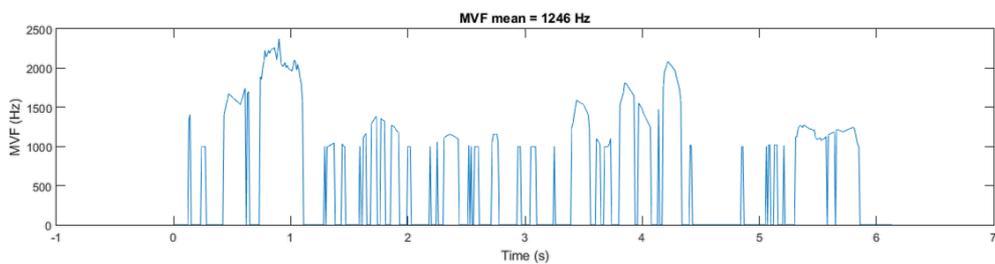


Figura 14. Evolución de la MVF en una mujer con voz esofágica

### 7.2.2.- La regularidad de la fonación

En las voces esofágicas se percibe una menor regularidad en la fonación en comparación con las voces sanas. Para estudiar este trastorno de la voz, se analiza la componente de la fase, ya que transmite una gran cantidad de información [13]. Una de las formas para analizarlo es el estudio del retardo de grupo, es decir, el tiempo que emplea un cierto punto de la envolvente de una onda en propagarse entre dos puntos determinados.

En este trabajo se propone estudiar el retardo de grupo del chirp (*Chirp Group Delay*, CGD), porque esta representación se basa en la fase de la señal de la voz y ha demostrado ser más adecuado para el análisis de sus irregularidades que el análisis de la amplitud de la transformada de Fourier [7].

Se define el CGD como la derivada negativa del espectro de la fase calculada a partir de la transformada  $z$  en un círculo que no sea el círculo de radio unidad. Dicho de otra forma, el espectrograma del CGD es una representación que se basa en un análisis de la transformada de Fourier evaluada en el plano  $z$  en un contorno de diferente radio que la unidad.

La transformada de Fourier del chirp se calcula según la fórmula 2, siendo  $\rho$  el radio del círculo del análisis [14]:

$$\tilde{X}(\omega) = X(z)|_{z=\rho e^{j\omega}} = \sum_{n=0}^{N-1} x(n)(\rho e^{j\omega})^{-n} = |\tilde{X}(\omega)|e^{j\tilde{\theta}(\omega)} \quad (2)$$

El CGD se calcula según la fórmula 3 [14]:

$$CGD(\omega) = - \frac{d(\tilde{\theta}(\omega))}{d\omega} \quad (3)$$

En las figuras 15 y 16 se ve un ejemplo del resultado del análisis del espectrograma del CGD (parte superior de la imagen) y la forma de onda (parte inferior de la imagen) de una voz sana y una voz esofágica. Ambas diciendo la palabra “*strikes*” (golpea en inglés) del texto del arcoíris.

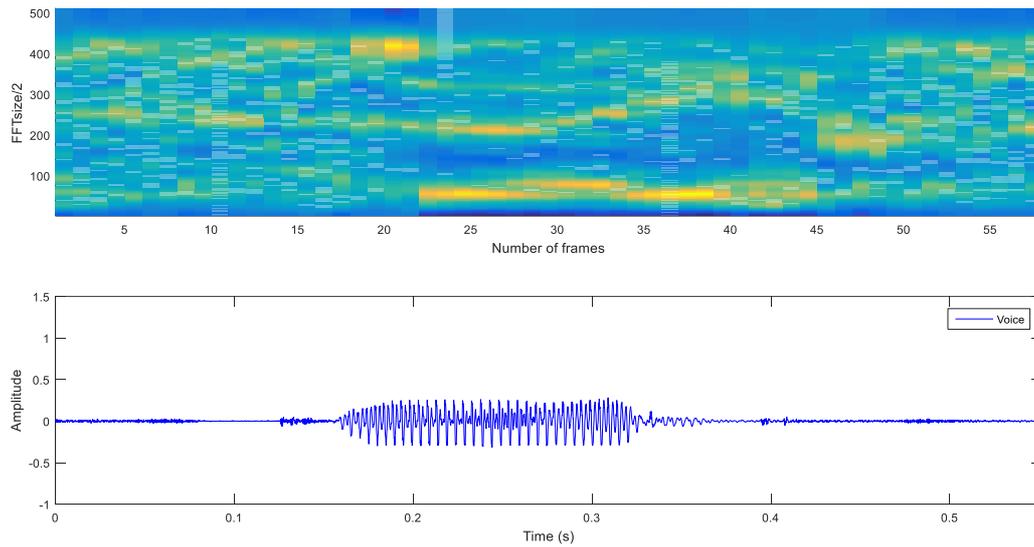


Figura 15. Espectrograma del CGD y forma de onda de una voz sana

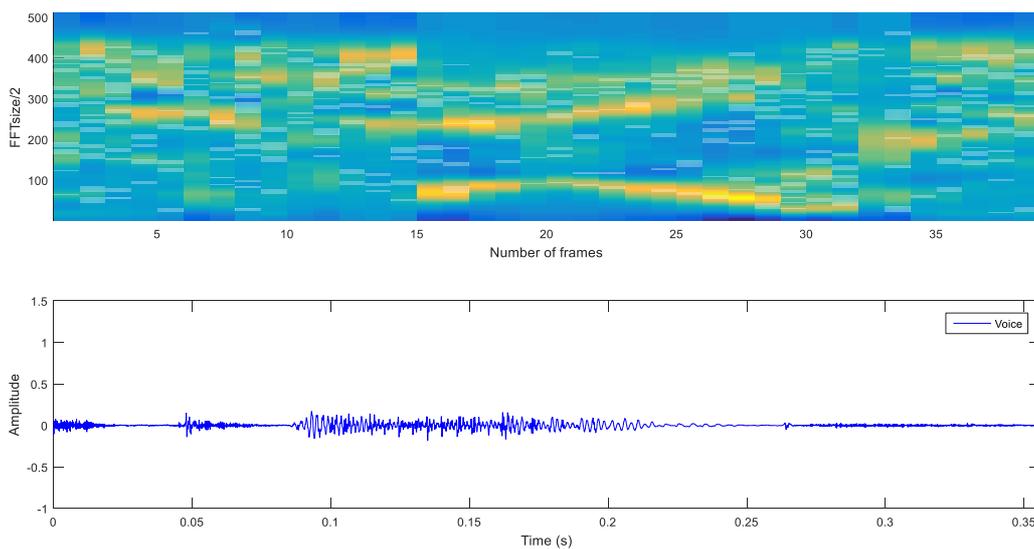


Figura 16 Espectrograma del CGD y forma de onda de una voz esofágica

Como se puede ver en las anteriores figuras, en ambos casos se distingue sin problemas los sonidos sonoros de los sonidos sordos de los fonemas de la palabra analizada. El poder distinguir los tipos de sonidos es porque este parámetro es dependiente de la periodicidad y por lo tanto, se deduce que también debería ser relevante a la hora de distinguir los dos tipos de voces, ya que como se ha explicado, las voces esofágicas son menos periódicas que las voces sanas.

Al comparar las figuras entre ellas, se puede ver que en el caso del espectrograma de la voz sana el espectrograma es más limpio que en el caso de la voz esofágica, es decir, no se aprecian tantas variaciones de colores en él. En cambio se ven muchas variaciones en el espectrograma de la voz esofágica.

Es necesario mencionar que la base de datos de voces esofágicas son voces muy entrenadas, por lo que al analizar otras voces, las diferencias serán aún más visuales.

Para el estudio del CGD se ha utilizado un código creado por Baris Bozkurt que calcula el espectro del CGD de un audio [14]. Y para poder cuantificar de alguna forma el espectro, se ha calculado la variación del CGD. Para ello se ha utilizado únicamente la parte de voz de la grabación, es decir, eliminando los silencios iniciales, finales e intermedios. La explicación de cómo la herramienta consigue diferenciar la voz del silencio, aparece en el apartado *La velocidad del habla*. Después se ha calculado la media de las diferencias entre muestras consecutivas.

Se explica a continuación otro ejemplo en el que se aprecia la variabilidad del CGD en los dos tipos de voces. En la figura 17 se puede ver el espectro del CGD (parte superior de la figura) de la voz de una mujer sana (parte inferior de la figura) diciendo la palabra "asa". Analizando el espectro se distinguen sin problemas las

zonas de sonidos sonoros (las vocales a) y las zonas de sonidos sordos (la consonante s).

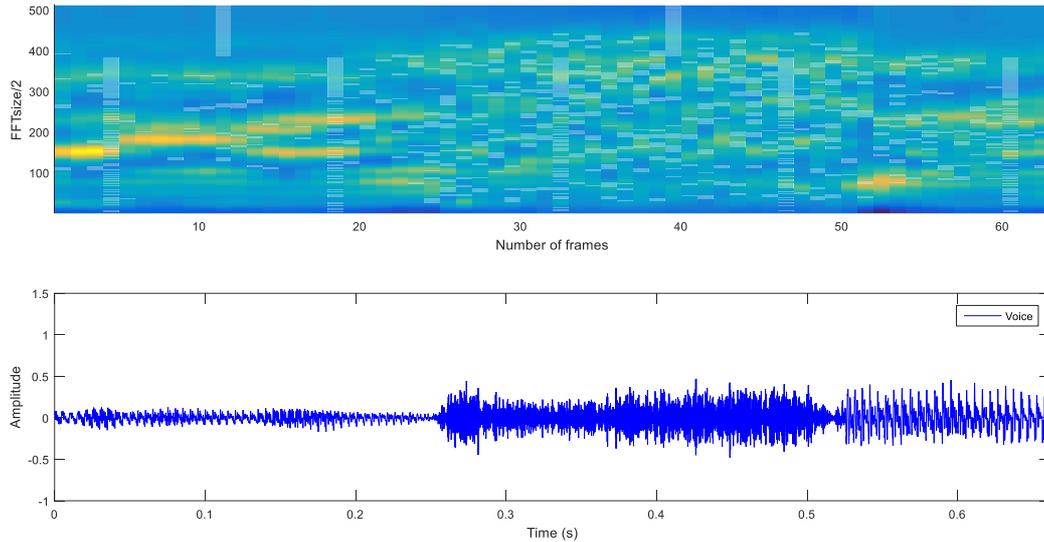


Figura 17. Espectro del CGD y la forma de onda de una voz sana diciendo “asa”

En la figura 18 aparece el espectrograma del CGD de una mujer con voz esofágica diciendo la palabra “sun” (sol en inglés). Además de que se distinguen fácilmente los tipos de sonidos, se puede apreciar una mayor variabilidad en el espectro de la voz esofágica.

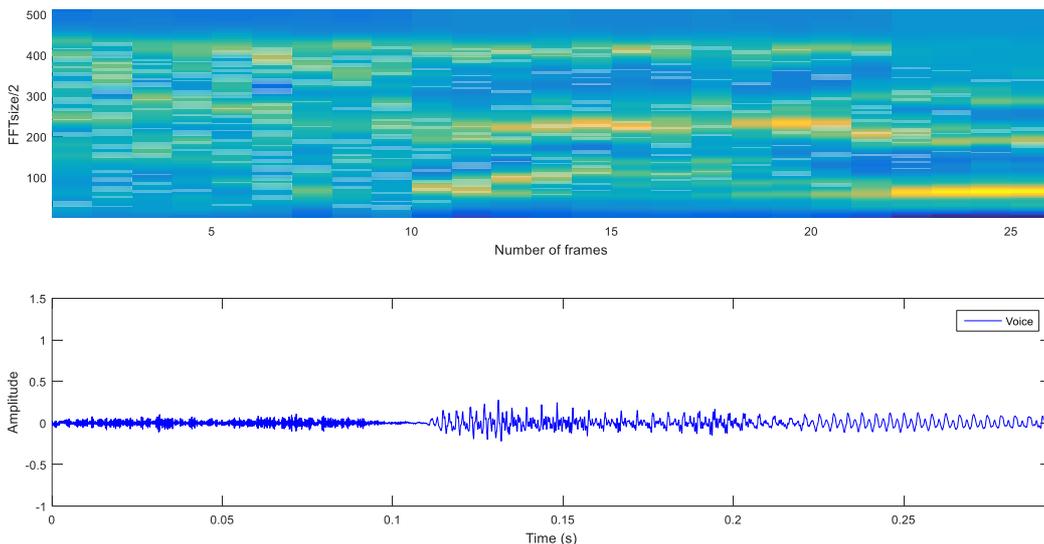


Figura 18. Espectro del CGD y la forma de onda de una voz esofágica diciendo “sun”

### 7.2.3.- El ruido en frecuencia alta

Algunos pacientes laringectomizados tienen dificultades para gestionar el flujo de aire que surge en la tráquea. Como consecuencia, cierto flujo de aire se escapa por el estoma, y esto, además de producir una voz más suave, también produce una voz más ruidosa, especialmente en las altas frecuencias (HF) [7]. Por este motivo el ruido en altas frecuencias es otro de los parámetros clave a la hora de evaluar las voces esofágicas, ya que es una forma de controlar que el paciente aprenda a gestionar el aire correctamente después de la cirugía.

Con el fin de cuantificar la cantidad de ruido HF, se propone estimar el espectro promedio a largo plazo (*Long term average spectrum*, LTAS) para cada grabación. El LTAS refleja la contribución de la fuente glótica y el tracto vocal para la calidad de la voz. En el eje de abscisas muestra el nivel de presión sonora en decibelios, y en la ordenada, la frecuencia en hercios. Su cálculo se realiza con una función propia del Matlab.

Una vez calculada el LTAS, se calcula la amplitud del espectro y se normaliza en energía. Es necesario normalizar ya que la intensidad de una grabación no depende únicamente de la voz, sino también de la técnica de grabación en sí, el propio micrófono... De esta forma la frecuencia con más fuerza de la LTAS queda en 0 dB y el resto en unidades negativas [15]. Esto ayuda mucho a la hora de realizar una comparativa entre espectro como se hace más adelante.

Finalmente, el ruido HF se mide como la energía en el espectro a largo plazo en la banda superior a una frecuencia dada (1,5 kHz en este trabajo) con la fórmula 4, siendo las muestra entre  $n$  y  $N$  las correspondientes a las frecuencias de 1,5 kHz en adelante.

$$E = \sum_{f=n}^N |X(f)|^2 \quad (4)$$

En la figura 19 se puede ver la representación del espectro promedio a largo plazo de una voz sana (línea azul) y de una voz esofágica (línea verde). Como se ha dicho, en el caso de la esofágica se espera que al ser una voz más ruidosa, tenga un ruido HF más alto, y por lo tanto, la amplitud del espectro en alta frecuencia en el caso de la voz sana es menor que el de la voz esofágica.

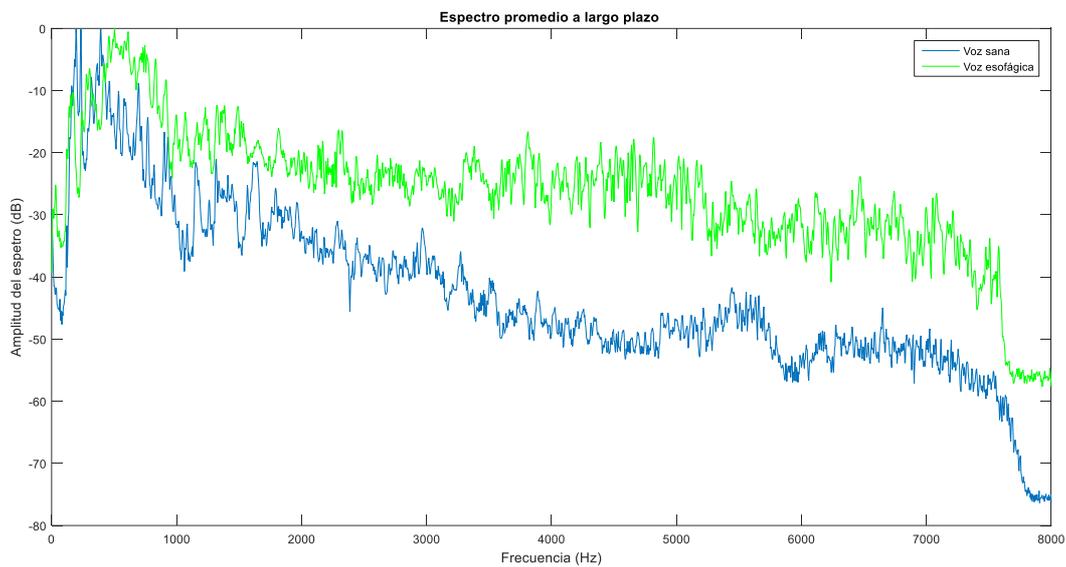


Figura 19. LTAS de una voz sana (azul) y una voz esofágica (verde)

#### 7.2.4.- La voz rota o quebrada

Para algunos pacientes, el habla esofágica se percibe como si estuvieran hablando con agua en la garganta. Esto es típicamente debido a problemas de deglución, que conducen a que la saliva y/o moco nasal de la garganta pueda dificultar el habla. El efecto de percepción resultante es, por lo general, similar a la de una voz rota o quebrada. La voz rota se define como una calidad áspera de la voz producida generalmente con un tono muy bajo y a menudo con periodicidad muy irregular [7], como se ha explicado anteriormente.

Para evaluar este parámetro se mide la proporción de voz rota, más conocido como *creakiness*. Para ello se utiliza el Toolbox de Matlab ya existente, “the GLOAT Toolbox”. Este algoritmo se centra en las características de la predicción lineal (*Linear Prediction*, LP) de la señal residual. En el caso de las voces esofágicas, se observa que aparecen picos secundarios inexistentes en el caso de la voz sana. El algoritmo se centra en detectar estos picos y en crear un parámetro resultante de la detección. Finalmente, utiliza un clasificador de decisión que, en función de un umbral, identifica las regiones de voz rota gracias al parámetro de la detección de los picos [16].

Una vez que el algoritmo devuelve las partes de voz rota, se calcula el *creakiness*. Primero se suprimen los silencios de la grabación, obteniendo así el número de tramas en las que el paciente dice la frase, eliminando silencios iniciales, finales e intermedios. El método para conseguir suprimir los silencios está detallado en el siguiente apartado, *La velocidad del habla*. Y después con el algoritmo, se detectan las partes con voz rota de la señal. Finalmente se dividen las tramas detectadas como voz rota entre las tramas de voz totales de la frase, de acuerdo a la fórmula 5.

$$Creakiness = \frac{\text{Tramas con voz rota}}{\text{Tramas con voz}} \quad (5)$$

En la figura 20 se representa la onda de la señal de una voz sana (color azul), la probabilidad de que en esa región exista voz rota (línea de color roja) y las regiones detectadas como zona de voz rota (línea de color verde). Como se ha comentado, para que la región sea clasificada como voz rota debe superar un umbral, en este caso se propone que la probabilidad supere el 0,3. Para esta voz el parámetro de *creakiness* es igual a cero.

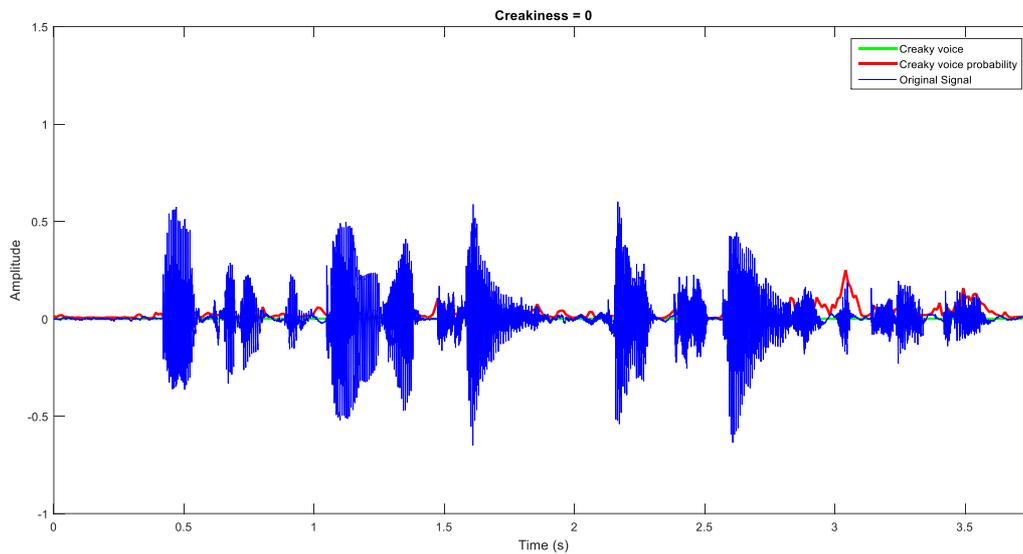


Figura 20. Detección de voz rota en voz sana

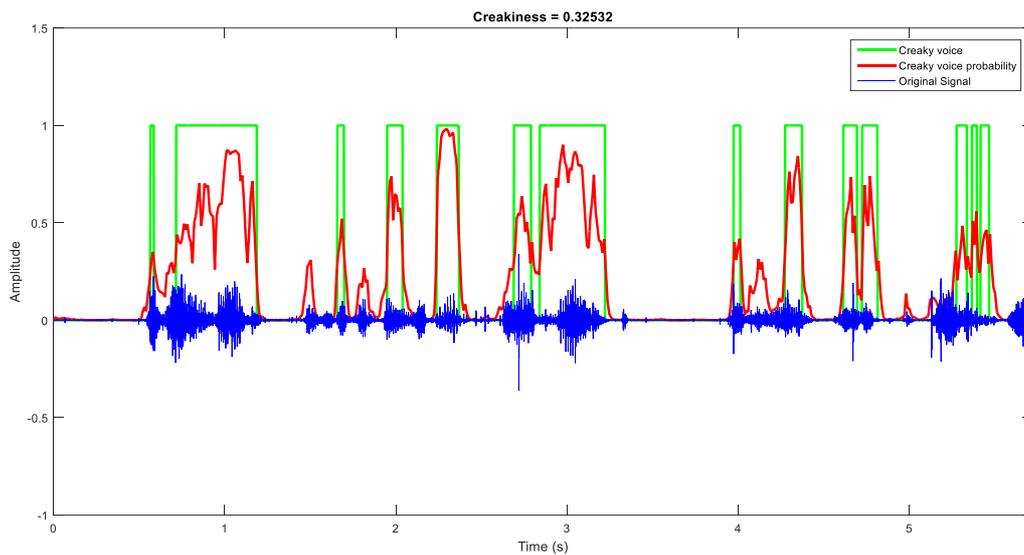


Figura 21. Detección de voz rota en voz esofágica

En la figura 21 se representa la misma figura pero para el caso de una voz esofágica. En este caso se ve que se detectan regiones de voz rota, ya que la probabilidad en estas zonas supera el umbral impuesto.

Como se ha comentado anteriormente, la base de datos de voces esofágicas son voces entrenadas y por lo tanto hay ciertos parámetros que se acercan bastante a los de las voces sanas. Aun así, en este caso, el *creakiness* detectado es de 0,33 aproximadamente.

### 7.2.5.- La velocidad del habla

Como es de esperar, la mayoría de los pacientes laringectomizados se enfrentan a algunos problemas cuando intentan mantener una cierta fluidez al hablar. En general tienen que hacer pausas para gestionar cómo respirar y hablar al mismo tiempo. Esta falta de fluidez se caracteriza por una velocidad de habla reducida, hablando más lento que una persona sana [7].

Para evaluar el parámetro de la velocidad se propone realizar una detección de la voz en la grabación. Para ello la herramienta sigue dos pasos: el primero es detectar cuándo comienza y cuándo finaliza el habla del paciente, detectando así los silencios iniciales y finales de las grabaciones; y el segundo consiste en diferenciar las tramas de sonido de las de silencios intermedios.

Para ello se utiliza un algoritmo de detección de la actividad de voz (*Voice Activity Detection*, VAD) del laboratorio de AHOLAB. Este VAD analiza el silencio inicial para calcular un umbral de ruido para hacer la clasificación de las tramas. Después, analizando la potencia del audio, ajusta el promedio de ruido y realiza el reconocimiento del nivel de sonoridad del ruido de fondo. De esta forma procede a calcular la relación señal a ruido (*Signal to Noise ratio*, SNR), y mediante umbrales se clasifican las regiones de voz y de silencio.

Para calcular la velocidad del habla (*Speaking Rate*, SR) primero se detectan las tramas de voz de la frase, suprimiendo todo los silencios. Y después se obtienen las tramas desde que el paciente comienza a decir la frase hasta que termina, con silencios intermedios incluidos. Finalmente se divide tal y como muestra la fórmula 6.

$$SR = \frac{\text{Tramas con voz}}{\text{Tramas de la frase completa}} \quad (6)$$

En la figura 22 se puede ver la forma de onda (en azul) y las partes que detecta como sonido (representado con la línea roja con amplitud 1) de una frase dicha por una mujer con voz esofágica.

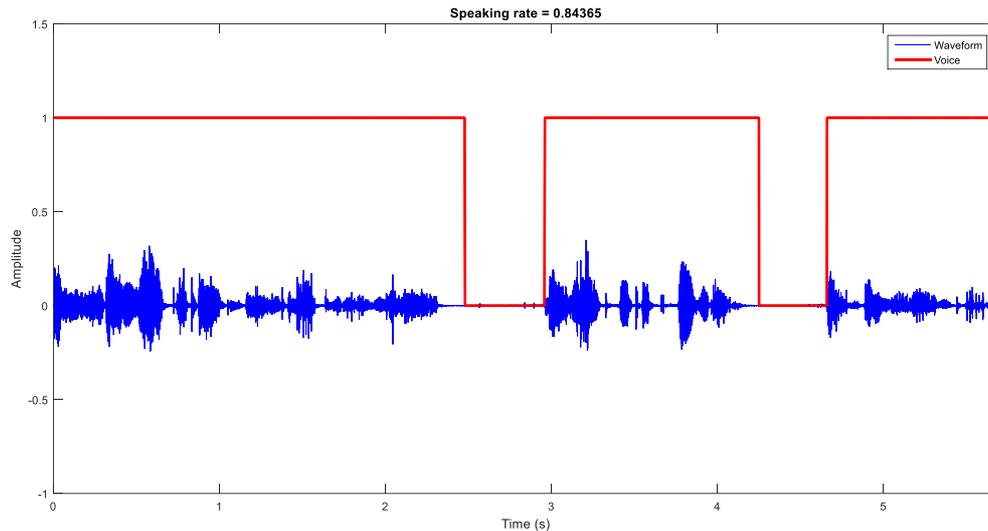


Figura 22. SR de una voz esofágica

Como se ha explicado, el SR se calcula dividiendo las tramas donde aparece la línea roja a 1 entre las tramas de la frase (aproximadamente 6 segundos en la figura 22), en este caso, SR es de 0,84 aproximadamente. Esto quiere decir que este paciente para decir esta frase habla el 84% del tiempo y el otro 16%, lo utiliza para hacer pausas, consecuencia de este tipo de voces.

Hay que tener en cuenta que esta base de datos está compuesta por frases cortas y que en general, para estas frases, una persona sana no necesita hacer pausas. Como se puede ver en la figura 23, una persona sana dice la misma frase anterior sin necesidad de hacer ninguna pausa, teniendo un SR del 100%.

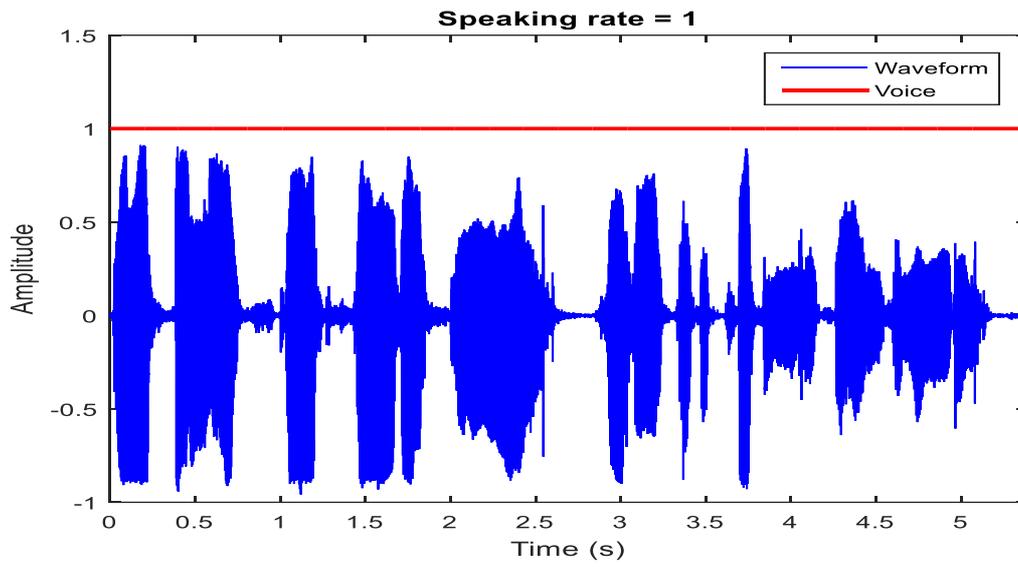


Figura 23. SR de una voz sana

### 7.3.- Representación de los resultados

Una vez creada la herramienta para analizar los parámetros por separado, es necesario prepararlos para finalmente visualizarlos en el pentágono propuesto por Drugman de la escala A4S.

Para representarlos en el pentágono es necesario que todos los parámetros estén evaluados entre el 0 y el 1, por lo que el primer paso es normalizarlos.

Se ha decidido normalizar los resultados de la base de datos creada en función del resto. Es decir, una vez calculados todos los resultados, se ha asignado al mejor de los casos la puntuación del 1 y al peor de los casos, el del 0. Por ejemplo, en el caso de la periodicidad, como una voz sana tiene un mayor MVF en comparación con una esofágica, al mayor MVF se la ha asignado el valor 1 y al peor el 0. El resto de resultados son puntuados en función de esto consiguiendo así que todos los valores estén entre 0 y 1.

Para concluir se procede a explicar lo que el logopeda ve a la hora de utilizar la herramienta. Cuando el logopeda seleccione una voz para que la herramienta la evalúe, ésta representa la evaluación de dicha voz con una imagen en la que aparece el pentágono y una tabla.

En la figura 24 se puede ver el pentágono de un ejemplo de una evaluación completa de una voz esofágica. En ella se puede ver la evaluación de los cinco parámetros normalizados, es decir, todo ellos evaluados entre el 0 y el 1.

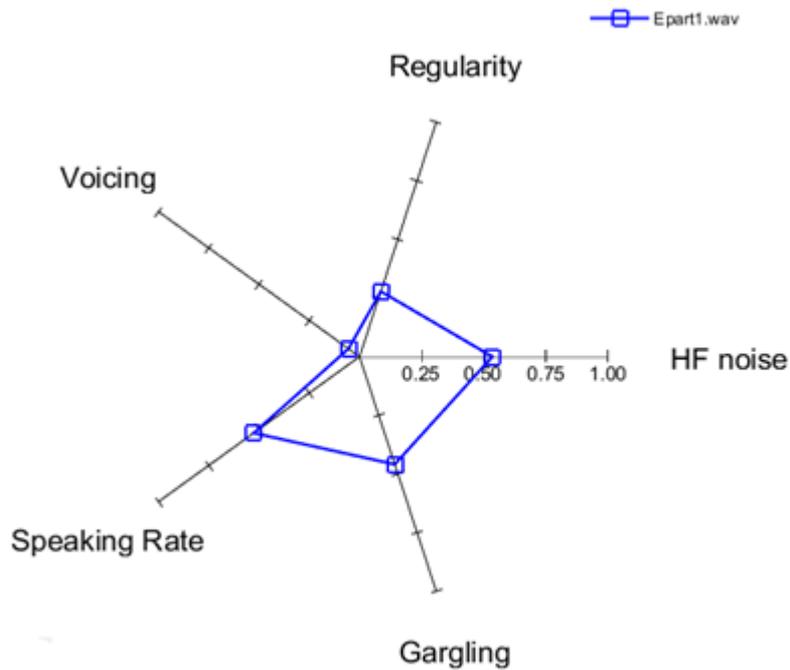


Figura 24. Evaluación de una voz esofágica

Parameters \ Axis Values	0	0.25	0.5	0.75	1
HF Noise	1.0879	0.9885	0.8892	0.7898	0.6905
Regularity	-10.0650	-10.6787	-11.2923	-11.9060	-12.5197
Voicing	1000	1.8835e+03	2.7670e+03	3.6504e+03	4.5339e+03
Speaking Rate	0.6246	0.7184	0.8123	0.9061	1
Gargling	0.3253	0.2440	0.1627	0.0813	0

Figura 25. Tabla proporcionada por la herramienta

Como se ha explicado, los resultados han sido normalizados en función de la base de datos, por ello se facilita al logopeda una tabla guía para conocer los valores reales de los ejes del pentágono para cada parámetro. En la figura 25 se ve una tabla en la que aparecen los valores de los parámetros equivalentes a los valores de los ejes del pentágono.

En el caso de que se quiera hacer un seguimiento de la evolución de un paciente, o se quiera comparar los resultados de varias señales al mismo tiempo, la herramienta permite representar más de una señal a la vez, como se puede ver en las figuras 26 y 27, dibujando el pentágono de cada señal de un color distinto y expresando en la leyenda de encima a cuál pertenece cada línea.

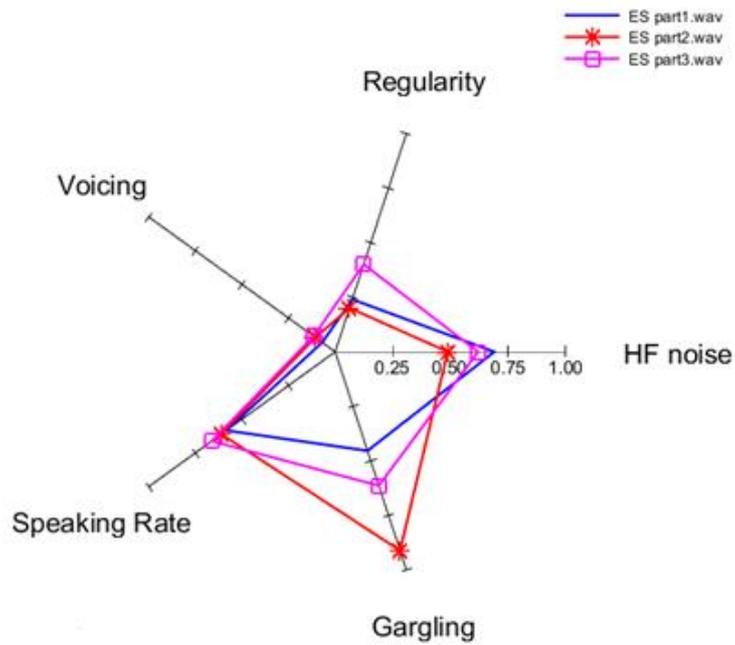


Figura 26. Evaluación de tres señales de la misma voz esofágica

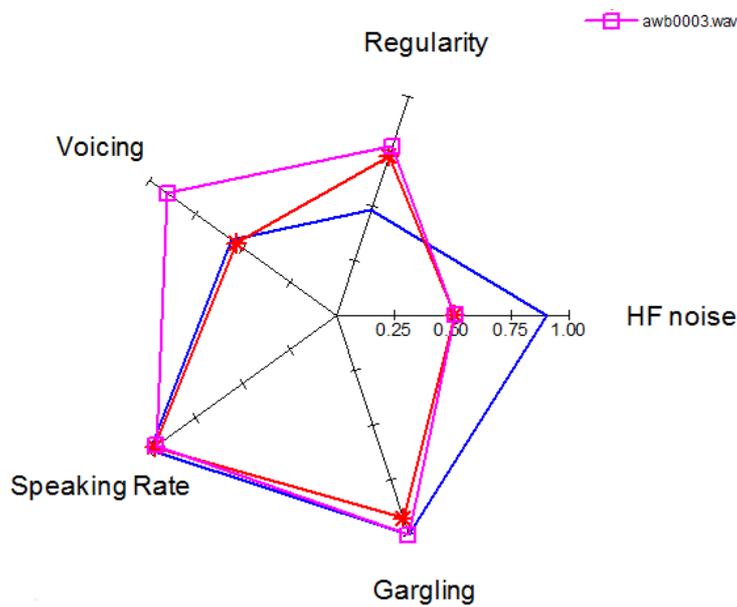


Figura 27. Evaluación de tres señales de la misma voz sana

## 8.- Aspectos económicos: costes

A continuación se presenta un presupuesto que detalla los costes de la realización y puesta en marcha del proyecto.

### 8.1.- Precios unitarios

#### 8.1.1.- Precios unitarios por mano de obra

Id.	Responsabilidad	Precio unitario (€/h)
P1	Director de Proyecto	60
P2	Ingeniero Junior	30

Tabla 12. Precios unitarios por mano de obra

#### 8.1.2.- Precios unitarios por recurso material

El material utilizado se divide en material fungible y material amortizable, tal y como se describe a continuación.

Id.	Concepto	Cantidad	Precio unitario (€)	Precio (€)
MATERIAL FUNGIBLE				
MF1	Material de Oficina / Otros	1	200	200

Tabla 13. Material fungible

Id.	Concepto	Precio (€)	Amortización (meses)	Precio unitario (€/mes)
MATERIAL HARDWARE				
MH1	Portátil lenovo G50	600,00	36	16,67
MH2	Impresora HP LaserJet CP1025 color	600,00	36	16,67
MH3	Memoria extraíble WD 150GB	70,00	36	2,33
MATERIAL SOFTWARE				
MS1	Sistema Operativo Windows 10	0,00	24	0,00
MS2	Matlab v2015	1.642,00	24	68,42
MS3	Microsoft Office 2013	150,00	24	6,25
MS4	Audacity	0,00	24	0,00

Tabla 14. Material amortizable

## 8.2.- Resumen económico de recursos humanos

En la siguiente tabla se muestran los costes derivados de los recursos humanos para cada paquete de trabajo del proyecto.

Paquete de trabajo	Horas		Importe (€)	
	P1	P2	P1	P2
PT.0: Gestión del proyecto	58	80	3.480	2.400
PT.1: Estudio de las características propias de las voces alaríneas	5	65	300	1.950
PT.2: Búsqueda de bases de datos existentes	2	20	120	600
PT.3: Estudio y decisión de parámetros a analizar	8	76	480	2.280
PT.4: Creación de la herramienta	15	300	900	9.000
PT.5: Mejora de la herramienta	10	60	600	1.800

Tabla 15. Coste de recursos humanos por paquete de trabajo

Id.	Nombre	Responsabilidad	Horas	Importe (€)
P1	Eva Navas Cordón	Director de Proyecto	98	5.880
P2	Ainhoa Larrea Murga	Ingeniero Junior	601	18.030
TOTAL RECURSOS HUMANOS (sin I.V.A)				23.910

Tabla 16. Resumen económico de recursos humanos

Los costes derivados de los recursos humanos, ascienden a **veintitrés mil novecientos diez euros (23.910 €)**, impuestos no incluidos y costes tanto directos como indirectos contemplados.

### 8.3.- Resumen económico de materiales

En la siguiente tabla se muestran los gastos incurridos en el proyecto debidos a los recursos materiales.

ID	Concepto	Precio unitario (€/día)	Días	Precio (€)
MATERIAL HARDWARE				
MH1	Portátil lenovo G50	0,55	140	77,00
MH2	Impresora HP LaserJet CP1025 color	0,55	15	8,25
MH3	Memoria extraíble WD 150GB	0,08	100	8,00
MATERIAL SOFTWARE				
MS1	Sistema Operativo Windows 10	0,00	140	0,00
MS2	Matlab v2015	2,28	90	205,2
MS3	Microsoft Office 2013	0,20	50	10,00
MS4	Audacity	0,00	5	0,00
MATERIAL FUNGIBLE				
MF1	Material de Oficina / Otros			200,00
<b>TOTAL RECURSOS MATERIALES (sin I.V.A)</b>				<b>508,45</b>

Tabla 17. Resumen económico de materiales

Los costes derivados de los recursos materiales, ascienden a **quinientos ocho euros con cuarenta y cinco céntimos (508,45 €)**.

### 8.4.- Resumen económico final

En la siguiente tabla se muestra el resumen económico final, teniendo en cuenta los costes derivados de los recursos materiales, los recursos humanos y los impuestos aplicables.

Concepto	Total (€)
Recursos humanos	23.910
Recursos materiales	508,45
<b>COSTE TOTAL (sin I.V.A)</b>	<b>24.409,45</b>

Tabla 18 Resumen del Presupuesto

El coste total de realización del proyecto “Herramienta de diagnóstico y evaluación para voces alaríngeas” asciende a **veinticuatro mil cuatrocientos nueve euros con cuarenta y cinco céntimos (24.409,45 €)**.

## 9.- Conclusiones

En este TFM se ha diseñado y desarrollado una herramienta para el diagnóstico y la evaluación de voces alaríngeas.

Una vez finalizado el proyecto se puede decir que se han cubierto los objetivos descritos anteriormente. Por un lado se ha realizado el estudio bibliográfico necesario para el correcto análisis de este tipo de voces, se ha creado una base de datos que cumple los requisitos necesarios para la comprobación del funcionamiento de la herramienta, se ha desarrollado la herramienta planteada, y posteriormente, se ha verificado el funcionamiento de la misma mediante distintas pruebas.

La herramienta de evaluación ha permitido un análisis objetivo de la voz, facilitando así al logopeda la evaluación de las voces de sus pacientes de una forma exacta. Además, gracias a la representación de los resultados mediante el pentágono de la escala A4S, el logopeda consigue una visualización intuitiva y fácil de interpretar para poder hacer un seguimiento del paciente o una comparativa entre voces de distintos tipos, sanas o esofágicas.

Finalmente para comprobar el correcto funcionamiento de la herramienta, se han realizado una serie de pruebas con las grabaciones de la base de datos creada. Se concluye que las pruebas realizadas proporcionan resultados satisfactorios en los parámetros analizados. Se recomienda que al comparar y analizar voces, éstas sean con el mismo contenido fonético, es decir, que lean el mismo texto, y que se graben bajo las mismas condiciones. Ya que como se puede ver en el anexo II, dos parámetros son dependientes del contenido fonético de la grabación: la regularidad de la fonación y el ruido en alta frecuencia. De esta forma se obtendrá un análisis más preciso.

## 10.- Bibliografía

[1] <http://seorl.net/>

REHABILITACIÓN DEL PACIENTE LARINGECTOMIZADO. Autores: Adriana María Penagos Vélez, Fernando López Álvarez, Faustino Núñez Batalla

[2] <http://www.otorrinoweb.com/es/temas-faringe-laringe/>

RECUPERACIÓN DE LA VOZ EN LARINGECTOMIZADOS.

[3] *Objective Assessment of Tracheoesophageal and Esophageal Speech Using Acoustic Analysis of Voice.* Autores: Ljiljana Siric, Dario Sos, Marinela Rosso and Sinisa Stevanovic.

[4] *A structured approach to voice range profile (phonetogram) analysis.* Autores: Arend M. Sulter, Hero P. Wit, Harm K. Schutte and Donald G. Miller

[5] [https://www.wevosys.com/products/lingwaves/lingwaves\\_phonetogram\\_plus.html](https://www.wevosys.com/products/lingwaves/lingwaves_phonetogram_plus.html)

lingWAVES Phonetogram Plus

[6] Estudio espectrográfico y comparativo de la voz normal, esofágica, traqueal y traqueoesofágica. Autores: Antonio Alarcos Llorach, Ángel Matesanz Sanz y Miguel González Pérez.

[7] *Tracheoesophageal speech: A dedicated objective acoustic assessment.* Autores: Thomas Drugman, Myriam Rijckaert, Claire Janssens, Marc Remacle.

[8] <http://dialectsarchive.com/>

Página de donde se obtiene la base de datos de las voces sanas leyendo el texto del arcoíris.

[9] <http://www.webwhispers.org/library/EsophagealSpeech.asp>

Página de donde se obtiene la base de datos de voces esofágicas leyendo el texto del arcoíris.

[10] *CMU ARCTIC databases for speech synthesis*. Autores: Kominek, Black. Tech. Rep. CMU-LTI-03-177, Carnegie Mellon University.

[11] *Joint Robust Voicing Detection and Pitch Estimation Based on Residual Harmonics*. Autores: Thomas Drugman, Abeer Alwan.

[12] *Maximum Voiced Frequency Estimation: Exploiting Amplitude and Phase Spectra*. Autores: Thomas Drugman, Yannis Stylianou.

[13] *Phase-based information for voice pathology detection*.

Autores: Thomas Drugman, Thomas Dubuisson, Thierry Dutoit.

[14] *Chirp group delay analysis of speech signals*.

Autores: Baris Bozkurt, Laurent Couvreur, Thierry Dutoit.

[15] *The Long-term average spectrum in research and in the clinical practice of speech therapists*. Autores: Suely Master; Noemi De Biase; Vanessa Pedrosa; Brasília Maria Chiari.

[16] *Improved automatic detection of creak*.

Autores: John Kane, Thomas Drugman, Christer Gobl.

[\*] <https://www.boe.es/buscar/doc.php?id=BOE-A-1999-23750>

Ley Orgánica 15/1999, de 13 de diciembre, de protección de datos de carácter personal.

---

## **11.- Anexo I: Normativa aplicable. Protección de datos**

Ya que para la realización de este proyecto ha sido necesario el uso de voces grabadas de personas reales, se debe de tener en cuenta la *Ley de Protección de datos* que defiende el derecho a que se requiera el previo consentimiento para la recogida y uso de los datos personales, el derecho a saber y ser informado sobre el destino y uso de esos datos y el derecho a acceder, rectificar y cancelar dichos datos. Su desarrollo se contiene en la *Ley Orgánica 15/1999, de 13 de diciembre, de protección de datos de carácter personal* que recoge el régimen general de la protección de los datos personales y que constituye una de las referencias legislativas a tener en cuenta [\*].

## 12.- Anexo II: Resultados

En este apartado se van a detallar los resultados obtenidos de la herramienta con la base de datos creada.

Para cada parámetro hay una figura que representa la distribución de los resultados una vez escalados. Esto sirve para poder comprobar si la herramienta ayuda a clasificar el tipo de voz, si es sana o esofágica, en función de los valores obtenidos en ese único parámetro. De esta forma se analiza cómo de representativo es este parámetro independientemente del resto.

Además de la figura, también se detallan en una tabla algunos datos de los resultados obtenidos para cada parámetro para la base de datos creada. Estos resultados son los valores antes de ser escalados para representarlos en el pentágono.

### *Resultados de la periodicidad*

En la figura 28 se puede ver el histograma con la distribución de los resultados del parámetro de la periodicidad, siendo las barras rojas las voces esofágicas y las azules las sanas.

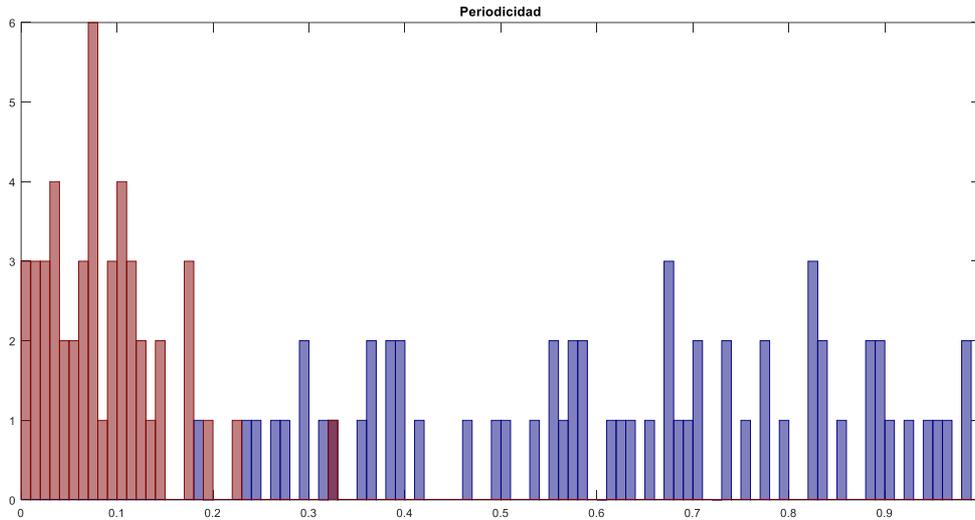


Figura 28. Histograma de la periodicidad

Como se puede ver en la figura 28, el parámetro de la periodicidad hace una buena clasificación de las voces, ya que separa los tipos de voces sin problema salvo en el caso de unas pocas señales. Para la base de datos utilizada, el rango de los valores de los resultados de las voces sana es mayor que el de las voces esofágicas.

En la tabla 19 se puede ver el valor máximo, el mínimo, la media y la mediana de los resultados para el parámetro de la periodicidad, tanto para las voces sanas como para las voces esofágicas.

	Máximo	Mínimo	Media	Mediana
Voces sanas	4533.9	1643.3	3223.8	3327.8
Voces esofágicas	2161.1	1000	1310.3	1262.5

Tabla 19. Valores orientativos de la periodicidad

### Resultados de la regularidad de la fonación

En la figura 29 se puede ver el histograma con la distribución de los resultados del parámetro de la regularidad, siendo las barras rojas las voces esofágicas y las azules las sanas.

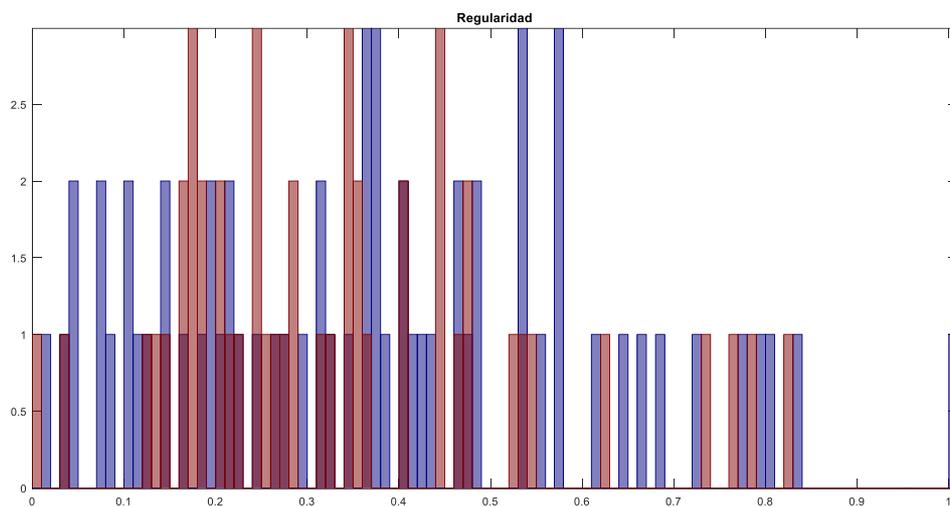


Figura 29. Histograma de la regularidad

Como se puede ver en la figura 29, el parámetro de la regularidad no hace una buena clasificación por sí solo en este caso.

Hay que tener en cuenta que para analizar los resultados de este parámetro sería conveniente comparar los dos tipos de voces leyendo el mismo texto, ya que este parámetro es dependiente del contenido fonético. Por lo tanto no se puede sacar una conclusión clara sobre la distribución de los resultados de este parámetro con la base de datos utilizada.

En la tabla 20 se puede ver el valor máximo, el mínimo, la media y la mediana de los resultados para el parámetro de la regularidad, tanto para las voces sanas como para las voces esofágicas.

	Máximo	Mínimo	Media	Mediana
Voces sanas	-10.0977	-12.5197	-10.9841	-10.9675
Voces esofágicas	-10.0650	-12.0828	-10.8960	-10.7928

Tabla 20. Valores orientativos de la regularidad

### Resultados del ruido de frecuencia alta

En la figura 30 se puede ver el histograma con la distribución de los resultados del parámetro del ruido de frecuencia alta, siendo las barras rojas las voces esofágicas y las azules las sanas.

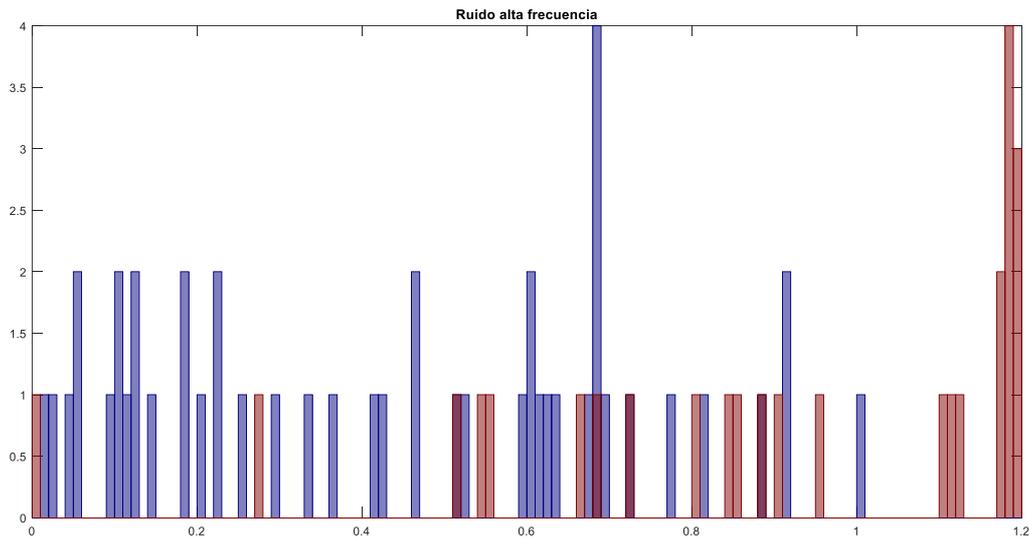


Figura 30. Histograma del ruido de frecuencia alta

Como se puede ver en la figura 30, el parámetro del ruido de alta frecuencia no hace una buena clasificación por sí solo en este caso.

Este parámetro es dependiente del contenido fonético y de las características con las que se hayan grabado las voces. En este caso las dos bases de datos utilizadas para la creación de la base de datos estudiada no están grabadas al mismo volumen y no leen el mismo texto. Por lo que para analizar los resultados de este parámetro sería conveniente comparar los dos tipos de voces leyendo el mismo texto y controlando las características de las grabaciones. Por lo tanto no se puede sacar una conclusión clara sobre la distribución de los resultados de este parámetro con la base de datos utilizada.

En este histograma se puede ver que los valores escalados no se encuentran entre 0 y 1. Esto ocurre porque, como se ha explicado anteriormente, el 1 se le asigna al mejor de los casos, para este parámetro eso sería la mínima energía entre las voces sanas. Como se puede ver en la tabla 21, existen señales con mejores resultados entre las esofágicas, algo no realista pero producido por el volumen de las grabaciones y por el contenido fonético.

En la tabla 21 se puede ver el valor máximo, el mínimo, la media y la mediana de los resultados para el parámetro del ruido de alta frecuencia, tanto para las voces sanas como para las voces esofágicas.

	Máximo	Mínimo	Media	Mediana
Voces sanas	0.3285	0.1269	0.2125	0.2159
Voces esofágicas	0.2390	0.0425	0.1121	0.1051

Tabla 21. Valores orientativos del ruido de alta frecuencia

### Resultados de la voz rota

En la figura 31 se puede ver el histograma con la distribución de los resultados del parámetro de la voz rota, siendo las barras rojas las voces esofágicas y las azules las sanas.

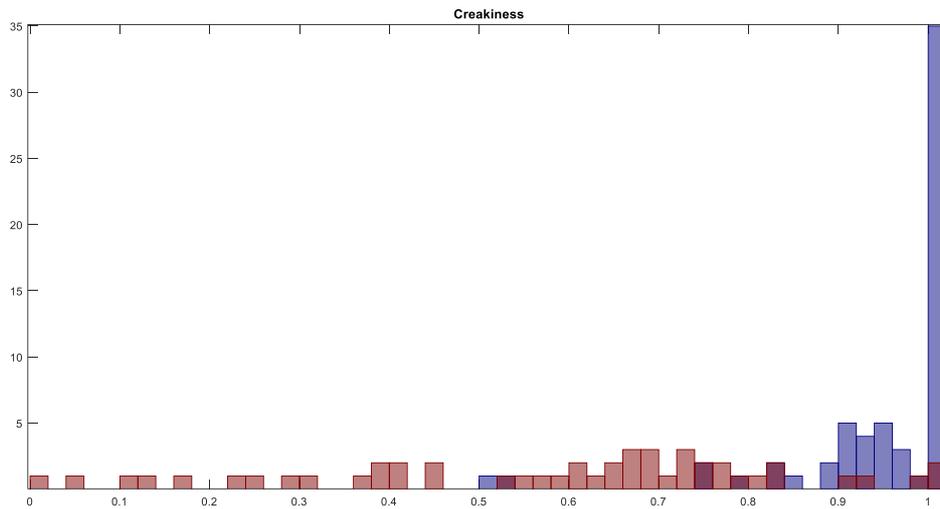


Figura 31. Histograma de la voz rota

Como se puede ver en la figura 31, el parámetro de la voz rota hace una buena clasificación de las voces, ya que separa los tipos de voces sin problema salvo en el caso de unas pocas señales.

Para la base de datos utilizada, los valores de los resultados de las voces sanas son la mayoría prácticamente uno, ya que el *creakiness* en las voces sanas debería ser prácticamente inexistente. Este parámetro es independiente del contenido fonético por lo que no se ve afectado por que las voces digan cosas distintas.

En la tabla 22 se puede ver el valor máximo, el mínimo, la media y la mediana de los resultados para el parámetro de la voz rota, tanto para las voces sanas como para las voces esofágicas.

	Máximo	Mínimo	Media	Mediana
Voces sanas	0.1623	0	0.0179	0
Voces esofágicas	0.3253	0	0.1362	0.1136

Tabla 22. Valores orientativos de la voz rota

Los resultados obtenidos para este parámetro son valores muy optimistas, para voces esofágicas no tan entrenadas como las que se han utilizado en la base de datos, los resultados de las voces esofágicas serían más cercanos a 1.

### Resultados de la velocidad del habla

En la figura 32 se puede ver un histograma con la distribución de los resultados del parámetro de la velocidad del habla, siendo las barras rojas las voces esofágicas y las azules las sanas.

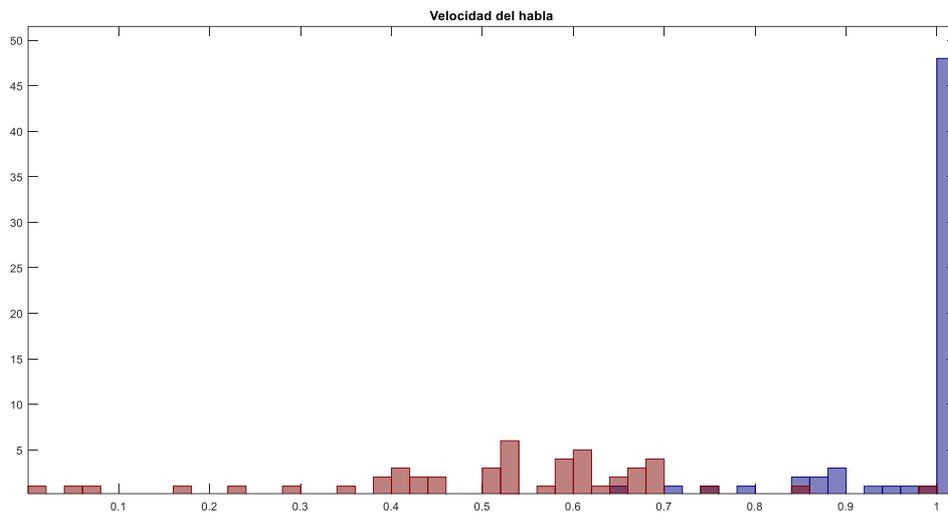


Figura 32. Histograma de la velocidad del habla

Como se puede ver en la figura 32, el parámetro de la velocidad del habla hace una buena clasificación de las voces, ya que separa los tipos de voces sin problema salvo en el caso de unas pocas señales. Para la base de datos utilizada, los valores de los resultados de las voces sanas son la mayoría prácticamente uno, ya que son frases cortas y las voces sanas no necesitan hacer pausas para decirlas.

En la tabla 23 se puede ver el valor máximo, el mínimo, la media y la mediana de los resultados para el parámetro de la velocidad del habla, tanto para las voces sanas como para las voces esofágicas.

	Máximo	Mínimo	Media	Mediana
Voces sanas	1	0.8684	0.9873	1
Voces esofágicas	0.9947	0.6246	0.8195	0.8264

Tabla 23. Valores orientativos de la velocidad del habla

Los resultados obtenidos para este parámetro son valores muy optimistas, para voces esofágicas no tan entrenadas como las que se han utilizado en la base de datos, los resultados de las voces esofágicas no serían tan cercanos a 1.