DISEÑO Y DESARROLLO DE UNA HERRAMIENTA DE AYUDA AL PROFESIONAL MÉDICO PARA LA EVALUACIÓN OBJETIVA DE LA FONACIÓN

SARA GARCÍA NUÑO

PROYECTO DE SISTEMAS INFORMÁTICOS, FACULTAD DE INFORMÁTICA
UNIVERSIDAD COMPLUTENSE DE MADRID

Curso 2011/2012

Director: Rafael Caballero Roldán

SIC – FDI (UCM)

Codirector: Víctor José Osma Ruiz

ICS – EUITT (UPM)

Agradecimientos

En primer lugar quiero agradecer a Víctor Osma, codirector de este trabajo, por su esfuerzo y dedicación sin obtener nada a cambio. Gracias por su ayuda siempre que lo he necesitado y porque, en todo momento, ha hecho todo lo que estaba en su mano para que este trabajo saliera adelante. También a Juan Ignacio Godino que me apoyó en todos los trámites burocráticos necesarios para que este proyecto pudiese ser codirigido desde fuera de la Facultad de Informática. Tampoco puedo olvidar la colaboración de Nico y Juana que me han ayudado en todo lo que he necesitado durante el proceso de desarrollo.

Agradecer también a Don Rafael Caballero por aceptar dirigir el presente trabajo en condiciones excepcionales, con una codirección exterior a la UCM. También agradecerle su ayuda en todo momento.

Gracias a todos los compañeros y amigos de la universidad que me han acompañado y ayudado durante estos años así como a todos con los que he coincidido en el laboratorio de BYO del departamento de ICS de la Universidad Politécnica de Madrid y de los que he recibido ayuda cuando lo he necesitado.

Como no, mi agradecimiento a todos mis amigos de Villar de la Encina que siempre han estado ahí y que me han ayudado cuando lo he necesitado. Por supuesto, gracias a Rubén por aguantarme, animarme y estar a mi lado en los peores momentos.

Y, por último y más importante muchas gracias a toda mi familia, sobre todo a mis padres y hermana a los que debo tanto y que tanto me han ayudado. Por preocuparse por mí y por darme ánimos y cariño para poder llegar al final.

A mis padres

"No hay enigmas. Si un problema puede plantearse, también puede resolverse"

Ludwig Wittgenstein

"La verdadera sabiduría está en reconocer la propia ignorancia"

Sócrates

Resumen

Diseño y desarrollo de una herramienta de ayuda al profesional médico para la evaluación objetiva de la fonación.

En la actualidad hay una gran probabilidad de padecer una alteración vocal en algún momento de nuestras vidas. Por esta razón cada vez es más importante un buen diagnóstico de las mismas. El método más comúnmente empleado para detectar las patologías laríngeas consiste en la observación directa de las cuerdas vocales durante la fonación. Existen muchas herramientas y técnicas de visualización que permiten facilitar el trabajo del profesional médico. Algunas de las más extendidas son: la estroboscopia, la quimografíam (basada en vídeos de alta velocidad o en grabaciones estroboscópicas), los fonovibrogramas y los diagramas de área glotal. La mayoría de ellos requieren para su desarrollo un tratamiento de imagen orientado a la segmentación del área glotal.

En el presente trabajo se diseña y desarrolla una herramienta que permite al profesional médico visualizar grabaciones de las cuerdas vocales durante la fonación, segmentar los distintos fotogramas mediante un método que permite aislar el espacio glotal, realizar medidas de área de la glotis obtenida y guardar los resultados en formato DICOM.

El método que permite aislar la glotis en imágenes de la laringe consta de una transformada “Watershed” seguida de varios procesos de “Merging” que permiten obtener la glotis junto a otros tejidos de la laringe. La decisión final se toma mediante un predictor lineal que distingue la glotis del resto de elementos en función de su forma.

El resultado de este proceso de segmentación se guarda en imágenes con formato DICOM. Este formato es el estándar reconocido para el intercambio de imágenes médicas y contiene, además de la propia imagen, información de la misma, del paciente o del método empleado para su captura.

Palabras clave

Fonación

Glotis

Cuerdas vocales

Patologías del aparato fonador

Segmentación

Transformada Watershed

Merging

DICOM

Abstract

Design and development of a tool to support the medical professional for the objective evaluation of the phonation.

Currently there is a high probability of having a voice disorder at some time in our life. For this reason, it is more and more important an accurate diagnosis of them. The most useful method of detection larynx pathologies is direct observation of vocal folds vibration during the phonation process. There are many tools and visualization techniques to make the surgeon task easier and more accurate. Some ot the most used are: stroboscopy, kymography(based on high-speed videos or stroboscopic recordings), phonovibrography and glottal area diagrams. Most of them require imagen treatment for their development oriented segmentation glottal area.

In the present work is designed and developed a tool that allows the doctor to see professional recordings of the vocal cords during phonation, segmenting the frames using a method that isolates the glottal space, take measurements of glottis area and save results in DICOM format.

The method is isolate the glottis in images of the larynx consists of a transformation "Watershed" followed by several processes "merging" which lead to the glottis with other tissues of the larynx. The final decision is taken with a linear predictor that distinguishes the glottis from other items based on their shape.

The result of this segmentation process is stored in DICOM format images. This format is the recognized standard for sharing medical images and contains, besides the image itself, the same information, the patient or the method used to catch them.

Keywords

Phonation

Glottis

Vocal cords

Pathologies of the vocal apparatus

Segmentation

Watersheds transformed

Merging

DICOM

Tabla de acrónimos

ACR:	American College of Radiology - Colegio Americano de Radiología.
ASM:	Active Shape Models - Modelos de formas activas.
DICOM:	Digital Imaging and Communication in Medicine.
EGG:	Electroglotograma.
JND:	Just Noticeable Difference - Diferencias apreciables.
LOPD:	Ley Orgánica 15/1999 de 13 de diciembre de Protección de Datos de Carácter Personal.
NEMA:	National Electrical Manufacturers Association.
NNG:	Nearest Neighbor Graph - Grafo de vecinos más cercanos.
ORL:	Otorrinolaringología.
RAG:	Region Adjacency Graph - Grafo de regiones adyacentes.
RGB:	Red, Green, Blue - Rojo, verde, azul.
UML:	Unified Modeling Language - Lenguaje Unificado de Modelado.
VR:	Value Representation. Representación del valor.
YIQ:	Luminance, In-fase, Quadrature - Luminancia, fase, cuadratura.

Índice

Capítulo 1 ..... **11**
Introducción **11**
1.1 Motivación 11
1.2 Objetivos e hipótesis de trabajo 12
1.3 Solución planteada 13
1.4 Estado de la cuestión 13
1.5 Protección de datos 15
1.6 Organización del documento 15
Capítulo 2 ..... **16**
Fisiología y funcionamiento del sistema fonador. Patologías relacionadas **16**
2.1 Introducción 16
2.2 Fisiología de la laringe y del sistema fonador 17
2.2.1 Subsistema efector 17
2.2.2 Subsistema vibrador 18
2.2.3 Subsistema resonador 19
2.3 El fenómeno fonatorio 20
2.4 Patologías Vocales 20
2.4.1 Métodos de valoración de la alteración vocal 21
2.4.2 Clasificación de patologías 22
2.5 Métodos para diagnosticar y caracterizar patologías en las cuerdas vocales 26
2.5.1 Estroboscopia 27
2.5.2 Cinematografía de alta velocidad 28
2.5.3 Laringoscopia 28
2.5.4 Técnicas que permiten el análisis y la extracción de medidas a partir de la exploración de las cuerdas vocales en fonación. 30
Capítulo 3 ..... **33**
Técnicas orientadas a la segmentación de estructuras laríngeas. **33**
3.1 Introducción 33
3.2 Transformada “watershed” 34
3.2.1 Transformada “Watershed” por simulación de lluvia. 35
3.3 Operaciones de “merging” 36
3.3.1 Merging JND 38
3.4 Momentos 38
3.5 Predictor Lineal 39
3.6 Método de detección de la glotis utilizado 40

Capítulo 4 ..... **44**
Imágenes Dicom **44**
4.1 Introducción 44
4.2 Estructura del fichero DICOM 45
4.2.1 Preámbulo y prefijo 46
4.2.2 Meta-cabecera 46
4.2.3 Cabecera 47
4.2.4 Imagen 50
4.3 Elementos de la cabecera del fichero DICOM significativos en la aplicación 50
Capítulo 5 ..... **55**
Diseño y desarrollo de la aplicación **55**
5.1 Casos de uso 55
5.1.1 Cargar video 56
5.1.2 Visualizar el video 56
5.1.3 Seleccionar fotogramas 56
5.1.4 Guardar fotogramas originales como DICOM 57
5.1.5 Procesar (segmentar) fotogramas seleccionados 58
5.1.6 Guardar proceso completo (imágenes originales y segmentadas) 59
5.1.7 Cargar resultados 59
5.1.8 Visualizar los DICOM del grupo seleccionado 60
5.2 Diagramas uml 60
5.3 Verificación de clases 64
5.4 Herramientas de desarrollo 65
5.5 Problemas encontrados y soluciones 66
5.6 Manual de usuario 68
Capítulo 6 ..... **77**
Consideraciones finales **77**
6.1 Aportaciones originales 77
6.2 Conclusiones 78
6.3 Líneas Futuras de investigación 79
Capítulo 7 ..... **81**
Referencias bibliográficas **81**
7.1 Referencias 81

Índice de figuras

Figura 2.1. Localización dentro del cuerpo humano de los distintos elementos constitutivos del sistema respiratorio. Imagen extraída de [Anón.2008a]..... 17

Figura 2.2. Sección longitudinal de cabeza y cuello. Circundada en azul se localiza la laringe y dentro de ella el aparato fonador. Imagen adaptada de [Anón.2008a]..... 18

Figura 2.3. Localización de los distintos cartílagos que forman la estructura de la laringe. Figura adaptada de [Hixon2008]..... 18

Figura 2.4. Posiciones de las cuerdas vocales: abducción (a) - durante la respiración; y aducción (b) - durante la fonación. Vista desde la parte superior de la laringe. Imagen adaptada de [Hixon2008]. 19

Figura 2.5. (a) Nódulos en ambas cuerdas vocales. (b) Quiste en el pliegue vocal izquierdo. (c) Pólipo en cuerda vocal izquierda. [Osma-Ruiz2010] 23

Figura 2.6. (a) Edema de *Reinke*. (b) Laringitis crónica, con un pólipo en la cuerda vocal izquierda. (c) Carcinoma en comisura posterior de las cuerdas vocales. [OsmaRuiz2010]..... 24

Figura 2.7. Comparativa de visualización de una onda periódica: real (superior) y con luz estroboscópica (inferior). 27

Figura 2.8. (a) Procedimiento para realización de laringoscopia indirecta. (b) Imagen obtenida de la realización de laringoscopia indirecta..... 29

Figura 2.9. (a) Laringoscopio. (b) Esquema de posicionamiento del laringoscopio para observación de las cuerdas vocales. 29

Figura 2.10. (a) Fotograma a resolución completa para selección de la línea de interés. (b) Ejemplos de videoquimogramas, correspondientes a: cuerdas vocales normales (izquierda), cuerdas vocales de un paciente con una leve ronquera (derecha). Figuras tomadas de [Švec1996]. 31

Figura 2.11**.** Izquierda: Delimitación de los bordes de la glotis y preparación de la síntesis. Derecha: (A) Fonovibrograma completo. (B) Fases de apertura y cierre de las cuerdas vocales representadas sobre el..... 32

Figura 2.12. GAW de una secuencia de fotogramas correspondiente a una exploración de unas cuerdas vocales normales. Figura adaptada de [Yan2005]. 32

Figura 3.1. Resumen de las operaciones de tratamiento de imagen necesarias para sintetizar un quimograma..... 34

Figura 3.2. Aplicación de la transformada “Watershed” al gradiente de una imagen convertida a escala de grises: (a) imagen original; (b) transformada “Watershed”. 36

Figura 3.3. Cálculo del NNG de una imagen a partir de su grafo RAG..... 37

Figura 3.4. Umbral de visibilidad del sistema visual humano en función de los niveles de gris de los píxeles de una imagen..... 38

Figura 3.5. Ejemplo de histograma de los valores entregados por la función discriminante de *Fisher* para dos clases de elementos conocidos: (a) elementos de la clase 0; (b) elementos de la clase 1..... 40

Figura 3.6. Esquema del proceso seguido para la detección de la glotis..... 40

Figura 3.7. Ejemplo de segmentación de la glotis tras el primer proceso de “Merging”: (a) imagen original en escala de grises con divisiones “Watersheds” sobre-impresionadas; (b) divisiones “Watersheds” con la glotis resaltada en gris para una mejor distinción. 41

Figura 3.8. Ejemplo de segmentación de la glotis tras el segundo proceso de “Merging” (tercer paso del sistema): (a) imagen original en escala de grises con divisiones “Watersheds” sobre-impresionadas; (b) divisiones “Watersheds” con la glotis resaltada en gris para una mejor distinción. 42

Figura 3.9. Ejemplos de segmentación y detección de la glotis en imágenes laríngeas con distintas condiciones de iluminación, calidad, tamaño y posicionamiento del espacio glotal 43

Figura 4.1. Ejemplo de imagen DICOM perteneciente a un TAC de la cabeza..... 45

Figura 4.2. Estructura DICOM de los elementos de datos que forman la cabecera..... 48

Figura 4.3. Imagen explicativa del significado de los valores *WindowCenter* y *WindowWidth*. 53

Figura 5.1. Diagrama de casos de uso de la aplicación..... 55

Figura 5.2. Esquema de la organización de los diferentes paquetes dentro del proyecto. Se presentan las relaciones que hay entre ellos. 61

Figura 5.3. Diagrama UML de las clases del paquete Auxiliares. 61

Figura 5.4. Diagrama UML de las clases del paquete Dicom. 62

Figura 5.5. Diagrama UML de las clases del paquete Segmentación. 63

Figura 5.6. Diagrama UML de las clases del paquete Principal. 64

Figura 5.7. Pantalla que se abre cuando iniciamos la aplicación. 68

Figura 5.8. Pantalla obtenida una vez seleccionada la opción de *Abrir video nuevo*..... 69

Figura 5.9. Captura de pantalla de video cargado y fotogramas del mismo obtenidos. 70

Figura 5.10. Selección de fotogramas indicando los extremos. 71

Figura 5.11. Captura de Guardar las imágenes originales seleccionadas como DICOM..... 71

Figura 5.12. Resultado de segmentación de imágenes seleccionadas pulsando el botón *Procesar*. 72

Figura 5.13.Captura obtenida al guardar todo el proceso (imágenes originales y segmentadas) como DICOM. 73

Figura 5.14. Seleccionar un directorio para observar los DICOM que se encuentran en él..... 74

Figura 5.15. Pantalla de resultados para imágenes que se han segmentado. 74

Figura 5.16. Pantalla de resultados para imágenes que se han segmentado después de realizar medida del área de la glotis. 75

Figura 5.17. Pantalla de resultados para imágenes que no han sido segmentadas. 75

Figura 5.18. Imagen original DICOM guardada con nuestra aplicación y abierta con el visor *MicroDicom*. 76

Figura 5.19. Imagen segmentada DICOM guardada con nuestra aplicación y abierta con el visor *MicroDicom*. 76

Índice de tablas

Tabla 4.1. Algunas de las Sintaxis de transferencia existentes más utilizadas..... 46
Tabla 4.2. Codificación de los elementos cuya VR es OB, OW, OF, SQ, UT o UN..... 49
Tabla 4.3. Codificación de los elementos cuya VR sea diferente a la de los elementos de la tabla 4.1..... 49
Tabla 4.4. Codificación de los elementos con VR implícita..... 49
Tabla 4.5. Elementos de la cabecera del formato DICOM usados en la aplicación..... 51

Tabla 5.1. Tabla de verificación de clases para cada Caso de Uso..... 65

Capítulo 1

Introducción

1.1 MOTIVACIÓN

Hoy en día la voz juega un papel esencial en todos los ámbitos de nuestra vida, hasta tal punto que no sólo tiene valor como instrumento de comunicación y expresión de sentimientos, sino que existe, incluso, una preocupación creciente por tener un tono y un registro de voz agradables, lo que lleva a pensar que pequeñas afecciones que antes pasaban inadvertidas son causa ahora de visita al especialista.

La probabilidad de padecer una alteración vocal en la actualidad es muy alta. Esta probabilidad aumenta debido a que un gran número de profesionales utilizan su voz como instrumento de trabajo.

Los métodos más utilizados en la actualidad para el diagnóstico o evaluación de lesiones vocales son aquellos que permiten al especialista observar directamente la laringe. Dentro de estos sistemas destacan: la fibroscopia, que consiste en la introducción por la cavidad nasal de un cable flexible, de unos 5 milímetros de diámetro, conectado a una cámara; y la laringoscopia, que se basa en insertar a través de la boca un teleobjetivo para observar directamente los pliegues vocales.

El problema de los sistemas anteriores es que por sí solos no permiten visualizar el comportamiento dinámico de las cuerdas vocales, ya que la velocidad de vibración de estas es excesivamente elevada para ser captada por el ojo humano. Para resolver este inconveniente en el año 1878 aparece una técnica denominada estroboscopia, con la que se consigue una secuencia de las cuerdas vocales vibrando a una velocidad aparentemente menor que la velocidad real de fonación. Este hecho permite que el movimiento pueda ser observado a simple vista o grabado con una cámara de vídeo estándar.

Otra técnica muy importante que permite observar la fonación de las cuerdas vocales es la cinematografía de alta velocidad que, a diferencia de la estroboscopia, no se basa en una ilusión óptica sino que registra fotogramas a una velocidad mucho mayor que la que se podría conseguir con un video convencional. Esta técnica, en combinación de un laringoscopio y una fuente de luz continua ofrece al facultativo un sistema de diagnóstico completo. A pesar de esto, no está muy extendida en la rutina diaria de clínicas y hospitales debido a su elevado coste.

Un método económicamente más asequible, y que permite también visualizar los patrones de vibración real de las cuerdas vocales, es la videoquimografía. El sistema va representando en un monitor una única línea capturada de cada fotograma, una debajo de la otra, al mismo tiempo que se graban en una cinta de vídeo. La gran cantidad de parámetros que las imágenes quimográficas permiten calcular para caracterizar el comportamiento vibratorio de las cuerdas vocales hacen que este método empiece a aplicarse sobre secuencias previamente grabadas, ya sean de alta velocidad [Larsson2000] o estroboscópicas [Kim2003].

Existen otras técnicas que permiten calcular distintos parámetros sobre la vibración de las cuerdas vocales, tales como los fonovibrogramas [Lohscheller2008] y los diagramas de área glotal ("Glottal Area Waveform" - GAW) [Yan2006]; incluso hay técnicas que permiten la localización directa de ciertas patologías en las cuerdas vocales [Méndez-Zorrilla2008]. En todas ellas la detección de la glotis resulta una tarea básica.

El presente trabajo pretende diseñar y desarrollar una herramienta que permita al profesional médico visualizar grabaciones de las cuerdas vocales durante la fonación, así como el aislamiento de forma automática de la glotis en los fotogramas obtenidos a partir de dichas grabaciones. Además, una vez aislada la glotis, debe permitir realizar una medida del área de la misma, indicando al usuario el número de píxeles que ocupa. Después de haber obtenido tanto las imágenes originales como las segmentadas, el profesional de ORL podrá guardarlas como imágenes DICOM, que es el estándar reconocido mundialmente para el intercambio de imágenes médicas. La importancia de este formato radica en que permite guardar, además de la propia imagen, información de la misma relativa al paciente, la prueba realizada o el método empleado para su captura.

1.2 OBJETIVOS E HIPÓTESIS DE TRABAJO

Actualmente existen muchos trabajos de investigación que realizan segmentación de imágenes laríngeas. Sin embargo, se trata de un campo que se encuentra todavía en fase de desarrollo lo que hace que existan muy pocas herramientas para que puedan ser usadas por el profesional médico.

La falta de herramientas de este tipo se debe al hecho de que la segmentación del espacio glotal no es una tarea fácil ya que existen muchos factores que dificultan el proceso, tales como: ruido en la imagen, variabilidad en la iluminación, variabilidad de los niveles de gris presentes en la glotis, borrosidades, difuminado de bordes, movimientos de la cámara y/o del paciente, etc.

Teniendo en cuenta estas hipótesis de partida, se plantean los siguientes objetivos para el presente trabajo:

1. Presentar una revisión de los conceptos y definiciones asociados a los mecanismos de producción de la voz así como su mecanismo natural de generación. Recoger un resumen de las principales patologías que afectan al proceso y que son susceptibles de ser diagnosticadas mediante un análisis visual de las cuerdas vocales y/o de su comportamiento dinámico. Explicar las principales técnicas utilizadas en la actualidad para el diagnóstico de los problemas vocales.
2. Presentar una revisión de los conceptos y definiciones asociados a las principales técnicas de procesado digital de imagen empleadas en el método que se aplicará para realizar la segmentación del espacio glotal en imágenes laríngeas. Dentro de este objetivo se prestará una especial atención a métodos como la transformada "Watershed" y "Merging" que ofrecen una mayor facilidad para automatizar el proceso de segmentación.
3. Programar e integrar dentro de una aplicación de apoyo al profesional de ORL un sistema que, mediante combinación de las técnicas anteriormente citadas, permite

detectar la glotis en cualquier imagen de las cuerdas vocales de una forma totalmente automática.

4. Permitir la extracción de medidas a partir de la segmentación y detección de la glotis. Se presenta al usuario la medida del área glotal en píxeles para que sirva de ayuda al diagnóstico.
5. Estudiar el formato DICOM, que es el estándar reconocido mundialmente para el intercambio de imágenes médicas. Transformar tanto las imágenes originales obtenidas a partir de los videos de laringoscopia, como las imágenes obtenidas por la segmentación de la glotis de las anteriores en imágenes DICOM. Estas imágenes incluirán información relevante del paciente, como datos personales, y del método utilizado, para facilitar el trabajo del profesional médico.
6. Diseñar e implementar una aplicación más genérica que permita al profesional médico realizar la segmentación de la glotis de forma automática a partir de vídeos e imágenes estroboscópicas o de alta velocidad para realizar medidas de área que puedan servir de ayuda al diagnóstico y permitir evaluar lesiones vocales. El sistema deberá permitir seleccionar la fuente, dar la orden de iniciar el proceso, observar y almacenar los resultados en formato DICOM con los datos del paciente. Las imágenes obtenidas serán susceptibles de ser utilizadas posteriormente en otras aplicaciones, como visores DICOM de uso general.

1.3 SOLUCIÓN PLANTEADA

La solución planteada en el presente trabajo es una aplicación que permite cargar vídeos de alta o baja velocidad, obtenidos a partir de técnicas como es la laringoscopia, para obtener sus fotogramas. A partir de dichos fotogramas se permite realizar una segmentación de la glotis mediante un sistema que integra varias técnicas de segmentación de gran relevancia en el campo del tratamiento digital de imágenes.

Las técnicas utilizadas incluyen la transformada "Watershed", que junto con varios tipos de "Merging" y un proceso final de predicción lineal, hacen posible la detección automática de la glotis en las imágenes analizadas. La potencia del método se ve incrementada por la ausencia de cualquier tipo de inicialización y por no necesitar condiciones estrictas sobre las características de las imágenes a procesar.

Se permite procesar dichas imágenes y guardarlas en formato DICOM, uno de los más utilizados dentro del campo de la medicina debido a que permite almacenar conjuntamente la imagen y la información más relevante del paciente en un solo fichero.

Además, una vez segmentada la imagen y obtenida la glotis de la misma, se permite realizar medidas de área para obtener los píxeles que ocupa y poder ayudar al profesional médico a la realización de un diagnóstico o la evaluación de un tratamiento.

Aunque las imágenes así guardadas pueden recuperarse con cualquier visor DICOM de uso general, se ofrece también un visor dentro de la aplicación que permite obtener tanto la imagen como algunos datos del paciente. Todo ello facilitará el trabajo del profesional médico, permitiéndole procesar los fotogramas, guardarlos y abrir los resultados obtenidos de su almacenamiento en DICOM mediante una única aplicación.

1.4 ESTADO DE LA CUESTIÓN

En la actualidad, existen varios grupos de investigación que intentan facilitar el diagnóstico de los problemas y patologías de la voz y, para ello, realizan segmentación de imágenes laríngeas, más concretamente de la glotis. Esta técnica consiste en dividir una imagen en varias partes u objetos para, entre todos los resultantes, detectar aquel que representa al espacio glotal. El

objetivo de la segmentación es simplificar la imagen original y obtener las partes de ella que interesen al profesional médico para que sea más fácil de analizar y detectar posibles problemas o patologías en las cuerdas vocales de cara a su diagnóstico y tratamiento. Algunas de las técnicas más importantes usadas para la segmentación de la glotis son:

- Crecimiento de región.
- Contornos activos ("snakels").
- Modelos de formas activas.
- Watershed.

Hay otros grupos que además desarrollan técnicas más avanzadas, a partir de un proceso previo de segmentación del área glotal, y que permiten analizar y extraer medidas a partir de la exploración de las cuerdas vocales. A continuación se citan y explican brevemente las más importantes:

- Quimografía: técnica que permite inspeccionar movimientos aperiódicos en la vibración de las cuerdas vocales. Esta técnica ha sido, y es, aunque casi exclusivamente en el ámbito científico, una de las más utilizadas en el análisis del comportamiento dinámico de las cuerdas vocales, gracias a la multitud de parámetros que es posible calcular con ella y que permiten una buena caracterización del proceso de vibración [Herbst2009; Manfredi2006].
- Fonovibrogramas: tratan de representar la distancia ortogonal desde el borde interno de la cuerda vocal al eje central de la glotis. Con este método se consiguen representaciones muy claras de las fases de apertura y cierre de los pliegues, lo que permite caracterizar de forma muy precisa el comportamiento dinámico de los mismos [Lohscheller2008].
- Diagramas de área glotal: se trata simplemente de un esquema donde se representa el valor del área relativa ocupada por la glotis en cada uno de los fotogramas del vídeo, grabado durante una exploración laríngea [Woo1996].

Todas las técnicas que se han citado son más comunes en la investigación ya que en el campo clínico no existen prácticamente aplicaciones que permitan al profesional médico usar la segmentación para realizar medidas del área glotal con el fin de poder evaluar los problemas de la voz de determinados pacientes y ayudar al diagnóstico de los mismos. De hecho, el único grupo del que tenemos conocimiento es *Kay Elemetrics* [Anón.2011].

Esta compañía fue fundada en 1948 y, hasta la actualidad se ha dedicado al diseño y desarrollo de instrumentos electrónicos de medición. Muchos de ellos han servido durante años como herramientas de análisis de referencia de las señales acústicas en los diversos campos de la medicina, patología del habla y la lingüística, entre otros. Cabe citar que sacaron al mercado el primer espectrógrafo de sonido llamado *Sona-Graph* en 1951. En la actualidad, *KayPentax* (nombre con el que es conocido actualmente), distribuye sus productos a los principales líderes mundiales que se dedican a tratar problemas de la voz y el habla.

Aun así, ni KayElemetrics ni ningún otro grupo dedicado a la distribución de herramientas médicas ha introducido en las clínicas un sistema que permita detectar la glotis de forma totalmente automática. Tampoco se ha encontrado ninguna aplicación que permita, una vez aplicado el proceso de segmentación de la glotis, guardar el resultado en DICOM. Este aspecto puede ser muy útil para el profesional médico ya que, como hemos dicho anteriormente, es el formato estándar reconocido mundialmente para el transporte de imágenes médicas, lo que facilitaría los intercambios de datos para su análisis a través de un formato estándar.

Por estas razones aparece la necesidad de realizar una herramienta más genérica, que permita combinar funcionalidades existentes con otras que no se habían desarrollado hasta ahora y que, mediante esta fusión, pueda servir de ayuda y facilitar el trabajo a los profesionales en el tratamiento de la voz.

1.5 PROTECCIÓN DE DATOS

Todos los datos de pacientes usados en videos e imágenes para la realización de pruebas en la aplicación cumplen la Ley de Protección de Datos de Carácter Personal (LOPD) de manera que no sea posible identificar a ningún paciente.

Haciendo referencia a la salud, esta ley establece que si estos datos no son usados por los profesionales correspondientes en las instituciones y centros sanitarios, es preciso el consentimiento expreso de los titulares o la existencia de una Ley que permita el tratamiento de dichos datos.

1.6 ORGANIZACIÓN DEL DOCUMENTO

El presente trabajo se ha dividido en seis capítulos, que se enumeran y describen brevemente a continuación:

El actual **capítulo 1** trata de introducir el por qué del trabajo realizado, estableciendo además los objetivos que se persiguen con el mismo y las soluciones planteadas para poder llevarlo a cabo.

El **capítulo 2** se dedica a explicar la fisiología del aparato fonador, además de realizar una descripción de las principales patologías que pueden afectarlo, deteriorando el proceso de producción de la voz. También recoge una breve clasificación de las técnicas de ayuda al diagnóstico.

El **capítulo 3** presenta una descripción en detalle de las técnicas de procesado digital de imagen que se utilizan para segmentar la glotis en imágenes laríngeas en el presente trabajo.

Se explica, además, la combinación de las técnicas que se ha usado para llevar a cabo dicha segmentación.

El **capítulo 4** explica qué es el formato DICOM, de qué partes está compuesto y los elementos del mismo que se han utilizado para poder crearlos, modificarlos y guardar en ellos los fotogramas obtenidos de un video de la fonación de las cuerdas vocales.

El **capítulo 5** describe cómo está diseñada la aplicación. Para ello se presentan los casos de uso, diagramas UML, tablas de verificación de clases y las herramientas utilizadas para el desarrollo de la misma. Se incluyen además los problemas encontrados durante el diseño, las soluciones que se han adoptado y un manual de usuario de la aplicación para su fácil manejo.

Por último, en el **capítulo 6**, se realiza una discusión sobre las aportaciones originales del presente trabajo al estado del arte, se exponen las conclusiones y las futuras líneas de investigación.

Capítulo 2

Fisiología y funcionamiento del sistema fonador. Patologías relacionadas

2.1 INTRODUCCIÓN

En este capítulo se presenta una revisión de los conceptos y definiciones asociados a los mecanismos de producción de la voz [Baken2000; Garcia-Tapia1996; Hixon2008]. También se recoge un resumen de las principales patologías [Alonso- Hernández2008; Garcia-Tapia1996; Godino-Llorente2002; Jackson-Menaldi2002] que afectan al proceso y que son susceptibles de ser diagnosticadas mediante un análisis visual de las cuerdas vocales y/o de su comportamiento dinámico.

Se denomina **fonación** a la emisión sonora producida por el hombre, exhalada desde la boca en una espiración, por un reflejo llamado de sinergia neumofónica. Dentro de ella, la **voz** es el resultado de la transmisión en el medio aéreo de la vibración de las cuerdas vocales. Se puede definir la voz como el sonido que el aire expelido por los pulmones produce al atravesar las cuerdas vocales haciéndolas vibrar.

La voz es uno de los vehículos que el ser humano utiliza para poder comunicarse y transmitir ideas. Es, por tanto, uno de los muchos soportes físicos que permiten la comunicación. Es por esto que, en la actualidad, la voz juega un papel esencial en todos los ámbitos de nuestra vida, hasta tal punto que no sólo tiene valor como instrumento de comunicación y expresión de sentimientos, sino que existe, incluso, una preocupación creciente por tener un registro de voz agradable.

Resulta, pues, de especial trascendencia conocer a fondo la actividad del sistema fonador, de manera que puedan detectarse anomalías de funcionamiento o patologías del mismo causantes de posibles alteraciones en las características acústicas de la voz, lo que habitualmente se conoce como disfonía[1].

[1] El concepto, que proviene del griego, significa “voz difícil”, y aunque en un principio sólo se usó el término para referirse a alteraciones acústicas de la voz, hoy en día se acepta para hablar de la dificultad subjetiva de emitirla.

2.2 FISIOLOGÍA DE LA LARINGE Y DEL SISTEMA FONADOR

En la producción de la voz están involucrados tres subsistemas anatómicos que trabajan conjuntamente: el aparato respiratorio (subsistema efector), constituido básicamente por los pulmones y la tráquea; el sistema fonatorio (subsistema vibrador), situado en la laringe; y el sistema resonador, que modula la señal de voz al pasar por el tracto vocal y nasal [Garcia-Tapia1996; Rabiner1978].

2.2.1 Subsistema efector

Se presenta en la Figura 2.1. Está constituido, básicamente, por la tráquea y los pulmones, que descansan sobre un marco musculoesquelético fundamental dentro del proceso respiratorio. Los pulmones se localizan en el interior de la denominada caja torácica, formada por las costillas y delimitada en su base por el diafragma.

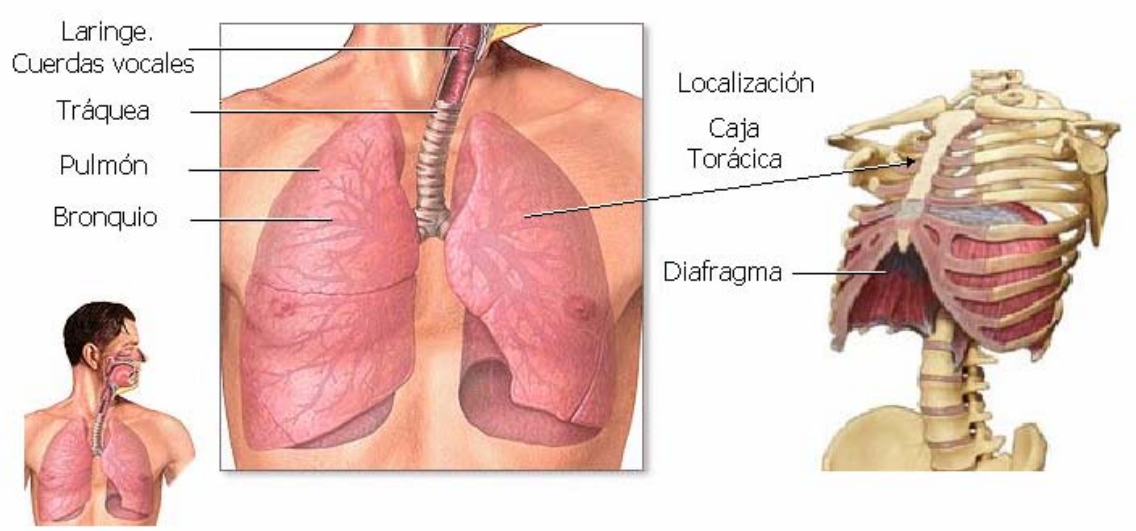


Figura 2.1. Localización dentro del cuerpo humano de los distintos elementos constitutivos del sistema respiratorio. Imagen extraída de [Anón.2008a]

Los pulmones constituyen la fuente de energía del sistema, y su misión es impulsar el aire a través de la tráquea para que llegue hasta los pliegues vocales, a los que hace vibrar.

El proceso respiratorio consta de dos fases bien diferenciadas: una activa, la inspiración; y otra pasiva (aunque a veces puede realizarse de forma voluntaria), la espiración.

- Fase de inspiración: en el interior del tórax, los pulmones se mantienen próximos a las paredes de la caja torácica debido a la presión que existe en su interior. Cuando el tórax se expande, gracias a un descenso del diafragma, los pulmones comienzan a llenarse de aire. En este momento, las costillas se levantan y se separan entre sí.
- Fase de espiración: La relajación de los músculos del tórax permite que éstos vuelvan a su estado natural contraído. El diafragma sube, presionando los pulmones y haciéndoles expulsar el aire por la tráquea. Aquí, las costillas descienden y quedan menos separadas entre sí, con lo que el volumen del tórax disminuye de nuevo. Se trata de la fase de espiración.

2.2.2 Subsistema vibrador

La laringe se localiza en el cuello, justo sobre el extremo superior de la tráquea, conectando a esta con la faringe (Figura 2.2). Está formada por las cuerdas vocales y una serie de cartílagos que sirven de soporte y protección. Todos ellos se encuentran unidos entre sí y/o con las estructuras periféricas mediante distintas membranas y ligamentos.

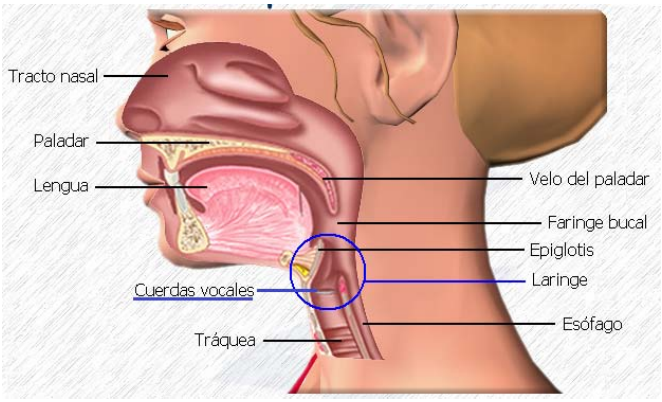


Figura 2.2. Sección longitudinal de cabeza y cuello. Circundada en azul se localiza la laringe y dentro de ella el aparato fonador. Imagen adaptada de [Anón.2008a].

La laringe está constituida por un marco o estructura cartilaginosa rodeada de tejidos. La pieza más prominente tiene forma de escudo, es el denominado cartílago tiroides. La parte inferior de la laringe está formada por una pieza circular, también cartilaginosa, denominada cartílago cricoides. Los cartílagos cricoides y tiroides constituyen la estructura básica que da forma de cavidad a la laringe. Justo encima del tiroides, se encuentra la epiglotis (Figura 2.3). Se trata de otra formación de tipo cartilaginoso que no participa en la fonación.

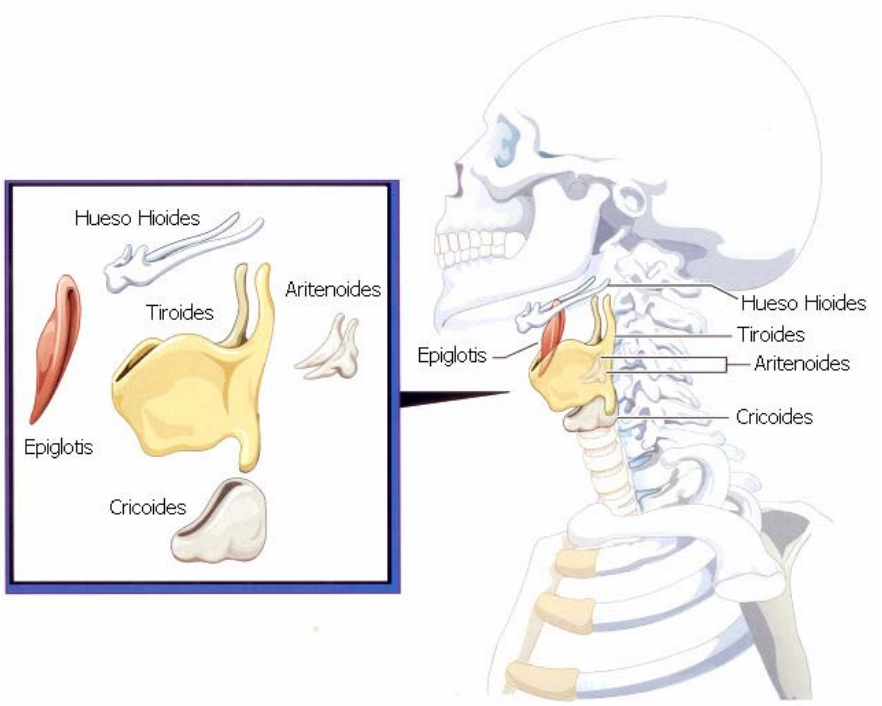


Figura 2.3. Localización de los distintos cartílagos que forman la estructura de la laringe. Figura adaptada de [Hixon2008].

En la laringe, como puede observarse en la Figura 2.2, el tracto aerodigestivo se divide en dos caminos diferentes: el aire inspirado atraviesa la tráquea hacia los pulmones, y los alimentos de la ingesta pasan por el esófago hacia el estómago. Este órgano desempeña, pues, tres funciones vitales: control del flujo de aire durante la respiración; protección de las vías aéreas durante la deglución; y producción de los sonidos para la voz.

En el centro de la estructura aparecen las cuerdas o pliegues vocales, que constituyen sin duda el elemento clave en la producción de la voz. Las cuerdas vocales forman parte de un oscilador capaz de generar una amplia gama de frecuencias ante el paso del aire a través de la glotis, variando su elasticidad, rigidez y viscosidad. Durante la fonación, los pliegues se cierran, a la vez que se acortan, siendo la mucosa que recubre las cuerdas y la capa más externa de éstas las que sufren el proceso de vibración. Por el contrario, durante la respiración, los pliegues vocales se prolongarán y abrirán, para dejar circular el aire de (y hacia) los pulmones, en un movimiento que se denomina de abducción. En la Figura 2.4 se presenta una vista de la parte superior de la laringe donde pueden distinguirse perfectamente ambas posiciones. Desde este foco, las cuerdas vocales forman una "V" invertida. El espacio comprendido entre los dos pliegues se denomina espacio glotal o glotis (Figura 2.4).

Figura 2.4. Posiciones de las cuerdas vocales: abducción (a) - durante la respiración; y aducción (b) - durante la fonación. Vista desde la parte superior de la laringe. Imagen adaptada de [Hixon2008].

2.2.3 Subsistema resonador

El subsistema resonador permite modular el sonido producido en parte gracias a la vibración de las cuerdas vocales, para convertirlo en los distintos fonemas.

Se compone de la faringe, la cavidad oral (boca), el tracto nasal, y el paladar blando o velo del paladar (Figura 2.3). La faringe tiene forma de cono invertido y está compuesta de músculos y varias capas membranosas. La cavidad oral continúa la faringe y no puede separarse de ésta. En ella se encuentran importantes elementos articuladores: mandíbula, lengua y labios. El tracto nasal se sitúa por encima de la boca, separado de ella por el paladar, y abarca desde las fosas nasales hasta el velo del paladar, el cual se encarga de regular el acoplamiento acústico entre las dos cavidades, a modo de válvula.

Asimismo, los ajustes en la posición de la lengua, las mejillas y los labios son responsables del cambio de las dimensiones y forma de la cavidad oral, afectando directamente a las características de resonancia de la voz.

2.3 EL FENÓMENO FONATORIO

La teoría más aceptada para explicar la vibración de los pliegues vocales es la mioelástica-aerodinámica, presentada por *J. Van den Berg* en 1958 y recogida en [Garcia-Tapia1996; Jackson-Menaldi1992]. Consta de 3 fases:

1. Durante la inspiración los pliegues vocales son abducidos hacia la posición intermedia o lateral, permitiendo el paso de aire del exterior a los pulmones. Cuando comienza la espiración, los músculos intrínsecos aductores hacen que las cuerdas se aproximen entre sí, contactando en la línea media, lo que combinado con el inicio de la espiración genera un aumento rapidísimo de la presión subglótica.
2. La presión del aire que viene de los pulmones produce una resistencia en las cuerdas vocales cerradas. Eventualmente esta presión se hace mayor que la fuerza que mantiene los pliegues unidos, por lo que se abren de forma momentánea para dejar salir el aire, liberándose así parte de la presión.
3. Al fluir el aire rápidamente por la laringe decrece la presión subglótica, lo que provoca un movimiento de aspiración de las cuerdas vocales hacia la línea media debido al efecto *Bernoulli*[2]. De esta forma, los pliegues vuelven a juntarse hasta la oclusión completa.
4. Este proceso se repite continuamente: apertura, disminución de la presión subglótica, cierre, aumento de la presión subglótica, apertura, etc.; es decir, el borde libre de las cuerdas vocales se mantiene en una vibración periódica.

Para que todo este proceso se lleve a cabo, se requieren una serie de condiciones, sin las cuales, el mecanismo de producción del ciclo vibratorio no sería factible: es necesario que la presión de aire sea suficientemente fuerte como para separar las cuerdas vocales; una glotis estrecha y un cuerpo muscular elástico; así como una mucosa suficientemente laxa, húmeda y libre de fijación al plano medio, con el fin de que sea capaz de ondular y desplazarse por una mínima presión negativa. La alteración de cualquiera de estas circunstancias puede afectar a la dinámica de la cuerda vocal, generando alteraciones de la voz. De todas ellas, la que más trascendencia tiene es la variación de las características físicas de la mucosa, pues la pérdida de ésta o su fijación al plano medio son incorregibles, causando una disfonía permanente.

2.4 PATOLOGÍAS VOCALES

Por patologías vocales entendemos aquellas que alteran la voz, es decir, toda aquella patología que pueda encontrarse a nivel laríngeo, que es donde se produce la fonación, o bien en las cavidades de resonancia, que es donde va a producirse la proyección vocal. Se entiende, pues, la voz patológica como voz anormal.

El resultado de una mala emisión vocal es la disfonía. La mayoría de las veces, el paciente tiene una adecuada percepción de su propia voz, y cuando acude a las consultas médicas es porque no la reconoce como normal. Cuando una persona acude a un profesional de ORL (Otorrinolaringología), lo hace movida por la percepción de alguna alteración en su voz que podría ser un síntoma de alguna enfermedad o de un mal funcionamiento de su órgano fonador.

Los síntomas que caracterizan y ayudan a reconocer la disfonía son los siguientes [Godino - Llorente2002]:

1. Dureza de la voz: consiste en la existencia de una tensión excesiva en la laringe, lo que produce una contracción demasiado grande de las cuerdas vocales. Como consecuencia se pueden producir lesiones en el borde libre de los pliegues e incluso nódulos laríngeos. La

[2] Efecto aerodinámico por el cual cuando la presión de flujo de un fluido permanece constante, al encontrar en su paso una estrechez, se produce un incremento en su velocidad que genera una disminución de presión en ese punto, lo que garantiza la conservación de la energía.

voz dura, se caracteriza por un comienzo explosivo de las frases, correspondiente al golpe de cierre de la glotis.

2. Aire en la voz: se produce cuando no hay un cierre completo de la glotis al final de cada ciclo de vibración, por lo que parte del aire espirado se pierde de forma turbulenta entre las cuerdas. Sus causas pueden ser muy variadas, por ejemplo: una parálisis laríngea o la existencia de alteraciones en la masa que impiden el cierre correcto de los pliegues vocales.
3. Ronquera: se refiere a la irregularidad o ausencia de vibración de las cuerdas vocales. La presencia de una masa en el borde de los pliegues puede entorpecer la vibración correcta de estas, o también puede ser motivo de ronquera la fijación de la mucosa al músculo, impidiendo que se forme la onda característica.
4. Alteraciones de la resonancia: el sonido que generan las cuerdas vocales es bastante plano y monótono, parecido al vuelo de un moscardón. Es en el momento de atravesar el tracto vocal cuando se modula esta señal. Un incorrecto funcionamiento de la cavidad resonante puede producir anomalías en los sonidos.
5. Rupturas o sonidos cuasi-periódicos: debidos a una producción pulsada de la voz que resulta en una baja e irregular frecuencia fundamental (en el rango 30-50 *Hz*). Se da cuando las cuerdas vocales son cortas, gruesas y están relajadas. Puede ir asociada a ronquera, pero su síntoma principal es la falta de periodicidad en la vibración.
6. Fatiga vocal: es la incapacidad de fonar durante periodos largos de tiempo (más de unos ciento veinte minutos) sin modificar el timbre[3], el tono o el volumen vocal (intensidad). Ciertas enfermedades nerviosas como el Parkinson o la esclerosis múltiple tienen sus primeros síntomas precisamente en un debilitamiento de la voz, que se vuelve más apagada debido a una insuficiente presión subglótica.
7. Aclaramiento vocal: se manifiesta por la necesidad continuada de carraspeo o tos laríngea. Su causa es la secreción excesiva de mucosidad, originada por la existencia de un nódulo o de cualquier lesión del borde libre de las cuerdas vocales.

Como se habrá observado, aunque son muchas las causas que pueden producir alteraciones de la voz, existen tres principales: defecto de cierre de la glotis; falta o irregularidad de vibración de las cuerdas vocales; y tensión excesiva de la laringe.

2.4.1 Métodos de valoración de la alteración vocal

Existen tres métodos distintos de valoración de la calidad vocal. Estos son: el análisis acústico, el análisis de la onda glotal y el análisis perceptual. Los dos primeros son métodos objetivos basados en la medida de parámetros de la señal de voz y señal glotal respectivamente, mientras que el análisis perceptual es una técnica de valoración subjetiva basada en el conocimiento experto, pudiéndose realizar de forma aural y/o visual, a través de distintas herramientas laringoscópicas.

1. Análisis acústico: a partir de la señal capturada con un micrófono y almacenada digitalmente en la memoria de un ordenador, se calculan un conjunto de medidas en el dominio temporal o frecuencial. Existe un importante número de parámetros acústicos utilizados para caracterizar las voces patológicas [Baken2000] como la frecuencia fundamental, la relación armónica a ruido, el índice de turbulencia de la voz o el índice de fonación suave.

[3] Característica acústica que permite diferenciar voces presentadas con la misma frecuencia fundamental y la misma intensidad.

2. Análisis de la onda glotal: se han desarrollado muchos parámetros a partir de la señal de EGG útiles para la clasificación de la calidad de la voz [Baken2000; McGillion2000]. A partir de la onda glotal son comunes las medidas: cociente de apertura, cociente de cierre, cociente de velocidad e índice de velocidad, descritas en [Godino-Llorente2002]. Asimismo, las medidas de frecuencia fundamental y las alteraciones en periodo y amplitud de la voz pueden extraerse también a partir de la señal electroglotográfica.
3. Análisis perceptual: el análisis perceptual es un juicio aural y/o visual de la fonación llevado a cabo por expertos. La evaluación aural suele hacerse aplicando escalas de valoración numérica, en términos tales como grado de ronquera, dureza, voz aérea y otros tipos de medidas cualitativas.

2.4.2 Clasificación de patologías

Para facilitar el diagnóstico de patologías vocales es importante contar con una buena clasificación de las mismas. Existen muchas variaciones dependiendo del autor que realice dicha clasificación. Dado que una patología vocal se debe a diferentes componentes orgánicos y funcionales, parece bastante adecuada la clasificación ideada por *Z.Milutinovic* [Milutinovic1996] en 1996, quien, atendiendo a la causa que las provoca, diferenció precisamente entre patologías funcionales y orgánicas.

2.4.2.1 *Patologías orgánicas*

Las disfonías orgánicas son las responsables de los problemas de voz más directamente relacionados con la vibración de los pliegues vocales, y las que producen alteraciones del habla más pronunciadas en el tiempo.

Entre las más habituales, se encuentran: nódulos, pólipos, laringitis aguda, quistes, edema de *Reinke*, cáncer de laringe, etc. A continuación se tratan de forma individual cada una de ellas explicando a qué son debidas, dónde se forman y cómo se tratan [Jackson-Menaldi2002].

Nódulos vocales

Es un tipo de tumor benigno que suele aparecer en las cuerdas vocales de personas con un mal uso vocal, que hablan muy alto, durante demasiado tiempo, o con una mala técnica. Son muy frecuentes en profesores o profesionales para los que la voz es su herramienta de trabajo y sobre todo en niños que gritan mucho durante sus juegos. También afectan bastante a las mujeres.

Se trata de pequeñas inflamaciones en forma de gránulos que suelen instalarse en la parte anterior de las cuerdas vocales (Figura 2.5-a).

Los nódulos son normalmente bilaterales, es decir, se forman en ambas cuerdas vocales, uno en frente del otro, aunque sus tamaños pueden ser asimétricos.

Al situarse en el borde de las cuerdas, justo en el punto de mayor fricción, se produce un defecto de cierre glótico y, por tanto, un escape de aire que tiene efectos en la intensidad, pudiendo llegar en ocasiones a producir una afonía bastante considerable e incluso la ausencia total de voz.

El diagnóstico se realiza por laringoscopia, aunque en muchas ocasiones se usa la estroboscopia.

Para tratar la enfermedad se emplean distintos tipos de ejercicios de fonación y logopedia, aunque a veces no llegan a desaparecer en su totalidad y es necesario recurrir a la cirugía, siempre que los síntomas sean especialmente molestos.

Quistes

Se trata de una lesión unilateral (en una sola cuerda vocal) de superficie lisa, que suele tener forma de tienda de campaña (Figura 2.5-b). Los quistes son redondeados, limitados por su propia pared epitelial y situados en el espesor de la cuerda, bajo la mucosa.

Pueden ser congénitos o adquiridos. En este último caso la afección surge por una retención de excreciones ante la obstrucción de algún conducto.

Producen ronquera crónica y alteraciones en la frecuencia fundamental de vibración de los pliegues vocales.

El diagnóstico se realiza mediante laringoscopia. Los quistes se pueden tratar con una terapia vocal, que busca reducir o eliminar el abuso de la voz, y/o con microcirugía.

Pólipos

Son neoformaciones en forma de péndulo que suelen aparecer en la parte anterior de las cuerdas vocales, como resultado de un uso excesivo de las mismas, aunque en ocasiones pueden tener un origen inflamatorio, infeccioso o incluso alérgico.

En realidad, los pólipos son un caso especial de nódulos con una componente inflamatoria mucho más elevada y que afecta en mayor grado a los hombres. Suelen aparecer de forma unilateral.

Tanto su sintomatología como su tratamiento son muy similares al de los nódulos, si bien, en este caso, la cirugía suele ser la solución más habitual.

Su diagnóstico se realiza mediante laringoscopia y, en la mayoría de los pólipos, suele ser necesaria la cirugía, ya que no mejoran lo suficiente con el tratamiento logopédico. La Figura 2.5-c presenta una lesión de este tipo.

(a) (b) (c)

Figura 2.5. (a) Nódulos en ambas cuerdas vocales. (b) Quiste en el pliegue vocal izquierdo. (c) Pólipo en cuerda vocal izquierda. [Osma-Ruiz2010]

Edema de Reinke

Esta enfermedad se manifiesta como una acumulación de líquido en las capas más superficiales de las cuerdas vocales (en una zona conocida como región de *Reinke*) que hace que adquieran una forma irregular, lo que dificulta su vibración.

Esta afección se produce principalmente en personas fumadoras o que se encuentran expuestas al humo del tabaco durante bastante tiempo, aunque también pueden contribuir otros factores como: reflujos gástricos, cambios hormanales (por ejemplo, hipotiroidismo) y un constante abuso de la voz. Son más frecuentes en las mujeres.

Como sintomatología se puede citar la ronquera, clásica de cualquier enfermedad relacionada con las cuerdas, y problemas en la función respiratoria.

Su diagnóstico se realiza bien con laringoscopia. Un método común de tratamiento consiste en practicar una incisión longitudinal en la parte afectada de las cuerdas, con el fin de extraer la acumulación de líquido característica de la afección.

La Figura 2.6-a muestra un edema de *Reinke* que afecta a ambas cuerdas vocales.

Laringitis crónica

Este término se utiliza para referirse a cualquier alteración patológica caracterizada por la inflamación duradera de la mucosa, hinchazón o edema de alguna parte de la laringe.

Generalmente se produce por: un uso excesivo de la voz, infecciones virales o bacterianas que afectan al tracto respiratorio, irritación debida al humo, o incluso irritación debida al reflujo ácido del estómago. Puede derivar hacia un tumor maligno.

Su diagnóstico se realiza con el estudio videoestroboscópico. La terapia vocal va a depender del tipo concreto de laringitis, aunque es muy común la necesidad de reposo, la corrección de malos hábitos en el uso de la voz y el tratamiento con antibióticos.

La Figura 2.6-b presenta un ejemplo de esta afección.

Carcinoma o Cáncer

Se produce por la aparición de células malignas en esa zona del cuello.

Muchas veces la enfermedad afecta también a las cuerdas vocales originando problemas en la voz como ronquera y cambio en la entonación. Suele ser causado por el humo del tabaco y el alcohol.

La sintomatología inicial es una ronquera persistente, seguida de una tos seca que puede llegar a producir la expulsión de un poco de sangre. Más adelante pueden originarse dificultades en la respiración o en la deglución, y originarse un bulto visible en el cuello debido a la inflamación de las glándulas próximas a la laringe.

El diagnóstico se hace mediante laringoscopia y biopsia del tejido afectado. El único tratamiento de esta enfermedad consiste en someter al paciente a distintas sesiones de radioterapia o extirparle la laringe mediante cirugía, en cuyo caso se le deberá practicar una traqueotomía permanente.

La Figura 2.6-c recoge un ejemplo de esta enfermedad.

Figura 2.6. (a) Edema de *Reinke*. (b) Laringitis crónica, con un pólipo en la cuerda vocal izquierda. (c) Carcinoma en comisura posterior de las cuerdas vocales. [OsmaRuiz2010]

2.4.2.2 *Patologías funcionales*

Se denominan así porque el paciente presenta los síntomas típicos de una disfonía pero, desde un punto de vista histológico, no se pueden observar pruebas evidentes de ésta en el interior de la laringe, y más concretamente en las cuerdas vocales. Este tipo de afecciones se suelen producir por un abuso de la voz o por diversas enfermedades, algunas de origen vírico o bacteriano.

Para este tipo de patologías se hacen especialmente importantes los métodos que permiten visualizar el movimiento de las cuerdas vocales como la videoquimografía o la estroboscopia.

Los siguientes puntos recogen algunas de las más significativas [Garcia-Tapia1996].

Parálisis de cuerda vocal

Como su propio nombre indica, esta enfermedad implica la parálisis total o parcial de una o de ambas cuerdas vocales, provocando un cierre incompleto.

Si la parálisis es unilateral el síntoma más común es una voz aérea y débil. Obliga al paciente a respirar más frecuentemente durante el habla. Si el problema es bilateral la situación es más seria, ya que ambos pliegues permanecen en la posición media, dificultando considerablemente la respiración.

La etiología es variada, pero suele deberse a afecciones de los nervios que estimulan los músculos de las cuerdas vocales. Sin embargo, puede también surgir por infecciones virales, por la atrofia del músculo tiroaritenoideo, o debido a una mala cirugía.

El diagnóstico se realiza principalmente mediante laringoscopia.

En cuanto al tratamiento, es necesario señalar que algunas parálisis unilaterales llegan a compensarse de forma natural, aunque con la ayuda de una terapia adecuada, en un tiempo nunca inferior a seis meses. En otros casos más complejos, o cuando la afección no remite, se hace imprescindible la corrección quirúrgica.

Trastornos parkinsonianos

El Parkinson es un síndrome neurológico manifestado por la combinación de temblor, reposo, rigidez y pérdida de reflejos posturales. Su etiología es desconocida. Según [Garcia-Tapia1996] un 40% de los pacientes afectados tienen alteraciones en la voz. La disfonía se caracteriza por una disminución de la sonoridad con monotonía, monointensidad e insuficiencia prosódica[4].

Es una enfermedad progresiva del sistema nervioso central que afecta al sistema autónomo y al sistema motor.

Suele provocar parálisis de las cuerdas vocales junto con dificultades respiratorias. La voz se vuelve monótona, con tensión y esfuerzo al hablar.

Esclerosis lateral amiotrófica

Es una enfermedad degenerativa del sistema nervioso central. Los síntomas más frecuentes son: habla borrosa[5], ronquera y disfagia. También se observa debilidad en la lengua, cuello, cara, velo del paladar y faringe.

Se puede detectar mediante un análisis laringoscópico.

[4] Término que hace referencia a una pronunciación correcta de las palabras y frases siguiendo las normas gramaticales establecidas.
[5] Pronunciación imprecisa y descuidada de los fonemas.

Corea de Huntington

Se trata de un desorden hipercinético[6] caracterizado por movimientos bruscos y carentes de propósito de cabeza, cuello y miembros [Garcia-Tapia1996], que se intensifican con la tensión emocional, desapareciendo durante el sueño. La voz es áspera; también hay monotonía, tensión y esfuerzo al hablar con un acortamiento del rango dinámico y del tiempo que la persona es capaz de permanecer hablando.

Disfonía espasmódica

Se trata de un grupo de trastornos de la voz caracterizados por problemas graves y espasmódicos de la aproximación de las cuerdas vocales, lo que hace muy difícil su control. La etapa inicial cursa con una ronquera inespecífica, asociada a "golpes vocales" y roturas de tono.

Aparece dificultad al respirar y dolor muscular en la parte superior del pecho.

Temblor esencial

Se trata de movimientos oscilatorios, involuntarios, relativamente rítmicos y carentes de propósito de la musculatura laríngea. Los temblores en las cuerdas vocales vienen asociados con los semejantes en las extremidades. Produce voz áspera y lenguaje hablado tenso y esforzado.

2.5 MÉTODOS PARA DIAGNOSTICAR Y CARACTERIZAR PATOLOGÍAS EN LAS CUERDAS VOCALES

A lo largo de los tiempos se han desarrollado muchos métodos que han ido resultando muy útiles para diagnosticar y caracterizar patologías en las cuerdas vocales. Según una clasificación establecida por *D.G. Childers* [Childers2000] éstos pueden dividirse en: glotográficos, centrados en medir aspectos de la apertura glotal de forma eléctrica (electroglotografía) u óptica (glotografía óptica); paramétricos, basados en la extracción de rasgos mediante procesado de la señal acústica (espectrogramas, medidas de intensidad, frecuencia fundamental, perturbaciones de frecuencia y de amplitud, medidas de ruido, etc.); y visuales, que permiten visualizar las cuerdas vocales en su forma estática (e.g., la laringoscopia) o dinámica (e.g., la estroboscopia).

Sin embargo, y a pesar de esta variedad, al final el experto en otorrinolaringología habitualmente recurre a la observación de la laringe y/o del movimiento vibratorio de las cuerdas vocales para confirmar o precisar sus evaluaciones.

Las imágenes en movimiento de la vibración de las cuerdas vocales proporcionan un excelente medio para evaluar todos los detalles de la función glotal. Ya en el año 1962, *G.P. Moore*, demostró que el análisis funcional de las cuerdas vocales podía revelar la causa de problemas en la voz [Moore1962]. Sin embargo, el asunto dista mucho de ser sencillo, debido a la gran velocidad con que se produce el proceso de fonación (100Hz en hombres y 250Hz en mujeres, aproximadamente) ya que a este ritmo es imposible que un humano o una cámara de vídeo estándar sean capaces de captar el movimiento real de las cuerdas.

A día de hoy existen, principalmente, 3 técnicas que logran resolver este problema. A saber: la estroboscopia, la cinematografía de alta velocidad y la quimografía. A continuación se estudian más detenidamente cada una de ellas.

[6] Grupo de trastornos que generan en el paciente falta de atención, hiperactividad e impulsividad en varios ambientes.

2.5.1 Estroboscopia

La luz estroboscópica fue introducida por primera vez en la laringe humana por *Toepler* (1866) [Garcia-Tapia1996] y fue impulsada definitivamente para la investigación por *M.J. Oertel* en 1878 [Oertel1878]. Con ella se consigue una visión de los pliegues vibrando a una velocidad aparentemente menor que la velocidad real a la que se está produciendo el proceso. Este hecho permite que el movimiento pueda ser observado a simple vista o grabado con una cámara de vídeo estándar. Sin embargo, la secuencia de fotogramas que se obtiene no refleja la vibración verdadera de las cuerdas vocales, ya que no se dispone de muestras de un único ciclo sino que se trata de una composición de diferentes periodos.

La técnica se desarrolla de la siguiente forma:

Se produce una secuencia de destellos luminosos de aproximadamente 0,1 milisegundos de duración (en ese tiempo las cuerdas vocales prácticamente no se mueven por lo que la imagen que se obtiene durante un único destello es bastante nítida). Los flashes de luz se encuentran retardados una pequeña fracción del periodo fundamental de fonación. Así, mediante submuestreo, se consigue un efecto de vibración lenta de las cuerdas que puede ser captado perfectamente por la cámara de vídeo.

La Figura 2.7 puede ayudar a entender el concepto de iluminación estroboscópica. En la gráfica superior se representa una onda que se repite periódicamente de forma similar a como lo harían las cuerdas vocales durante la vibración. La gráfica central representa los momentos en que se produce el destello luminoso, y por tanto los instantes de tiempo en que la imagen se hace visible. Como puede observarse, los flashes se encuentran retardados una pequeña cantidad (Δ) con respecto al periodo fundamental de la señal (T). Por último, en la gráfica inferior se representa la forma de onda que visualizaría el observador del proceso. Se trata de una curva con el mismo aspecto que la superior pero con una duración 7 veces mayor.

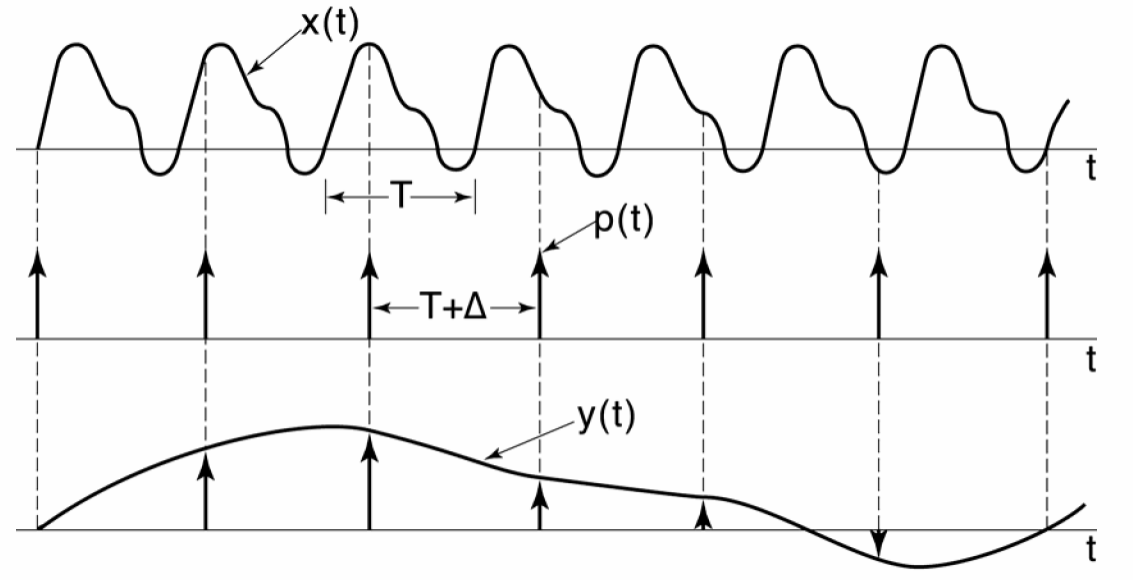


Figura 2.7. Comparativa de visualización de una onda periódica: real (superior) y con luz estroboscópica (inferior).

Los sistemas de última generación no sólo resuelven muchos inconvenientes encontrados inicialmente, sino que además permiten el control mediante ordenador de la luz, de la cámara, del cable y de los sistemas de proyección, además de posibilitar la aplicación de técnicas de análisis de imagen para la evaluación y cuantificación objetiva de la vibración de las cuerdas vocales.

A pesar de todas las mejoras introducidas, y de tratarse del dispositivo de visualización más extendido en el entorno clínico, la luz estroboscópica sigue teniendo graves problemas [Baken2000] que han facilitado el desarrollo de otras técnicas [Olthoff2007]: el incumplimiento del teorema de Nyquist impide que puedan observarse movimientos aperiódicos de los pliegues vocales; y la necesidad de sincronización del flash con la frecuencia fundamental de fonación hace que el instrumento falle ante muchas patologías.

2.5.2 Cinematografía de alta velocidad

Este sistema, al contrario que la estroboscopia, no se basa en una ilusión óptica, sino que trata de registrar fotogramas a una velocidad mucho mayor que la de un vídeo convencional. Mientras que con la ayuda de la estroboscopia sólo es posible realizar un análisis de 25 imágenes por segundo, mediante la cinematografía de alta velocidad se puede obtener hasta un rango de 4000 ò 5000 imágenes a alta resolución.

La mayor frecuencia de imágenes de la cinematografía de alta velocidad permite la detección exacta de desviaciones periódicas en los movimientos de las cuerdas vocales individuales, así como la observación de las fases de vibración y de interrupción del movimiento de las cuerdas vocales, incluso sin producirse la fonación. Se consigue, por tanto, observar la vibración real de las cuerdas vocales con la posibilidad de analizar movimientos periódicos y aperiódicos de las mismas.

Los primeros desarrollos de **sistemas digitales de alta velocidad** se empiezan a producir a mediados de los 80, a cargo de *S. Kiritani, H. Hirose, H. Imagawa y K. Honda*. Conseguían velocidades de 1.000 y 2.000 fotogramas por segundo, a una resolución de 50x50 píxeles en escala de grises, que lograban aumentar hasta 4.000 pero reduciendo drásticamente la resolución [Hirose1988; Kiritani1986]. En el año 1993 consiguieron aumentar la resolución gracias a la introducción de un foco de luz más potente que permitía grabar 128x32 píxeles a 2.000 fotogramas por segundo [Kiritani1993].

En 1998, un grupo de investigación dirigido por *T. Wittenberg* desarrolla un dispositivo que alcanza una velocidad de 2.000 fotogramas por segundo a una resolución de 128x128 píxeles.

La cámara es capaz de grabar hasta 10.000 fotogramas por segundo reduciendo la resolución vertical hasta 16 píxeles. [Wittenberg1998]. En el año 2003, miembros del mismo grupo, consiguen mejorar la técnica alcanzando una resolución de 256x256 píxeles a 2.000 fotogramas por segundo [Eysholdt2003]. Algo más tarde, en el año 2006, el equipo de investigación ya lograba realizar grabaciones a 4.000 imágenes por segundo, manteniendo la resolución en el mismo nivel [Schwarz2006].

En la actualidad existen cámaras capaces de grabar hasta 120.000 fotogramas por segundo a una resolución de 128x16 píxeles, valor que puede ser aumentado disminuyendo la velocidad. Por ejemplo, la cámara 9710 de *KayPentax* permite registrar 2000 imágenes por segundo con una resolución de 512x512 píxeles y 4000 imágenes con 512x256 píxeles [Anón.2011]. Otro ejemplo lo vemos en la cámara *Photron ultima APX-RS* mantiene una resolución de 512x512 píxeles para grabaciones de 10.000 fotogramas por segundo, con posibilidad, incluso, de capturar imágenes en color [Anón.2009b].

En combinación con un laringoscopio especial y una fuente de luz continua, esta técnica ofrece al facultativo un sistema de diagnóstico completo, que con una inversión razonable permite una valoración más exacta y amplia de los parámetros patológicos y fisiológicos de la voz.

2.5.3 Laringoscopia

La laringoscopia es un examen visual del interior de la garganta, donde se encuentra la caja de la voz (laringe) con las cuerdas vocales.

La laringoscopia es un procedimiento eficaz para descubrir las causas de los problemas de voz y respiratorios, el dolor de garganta y oídos, las dificultades para tragar, los estrechamientos de la garganta (constricciones o estenosis) y las obstrucciones de las vías respiratorias. También puede ayudar a diagnosticar problemas en las cuerdas vocales.

Existen tres tipos de laringoscopia:

1. Laringoscopia indirecta
2. Laringoscopia por fibra óptica
3. Laringoscopia directa

En la figura 2.8 se puede visualizar cómo se lleva a cabo una laringoscopia indirecta y la imagen que se obtiene después de su realización.

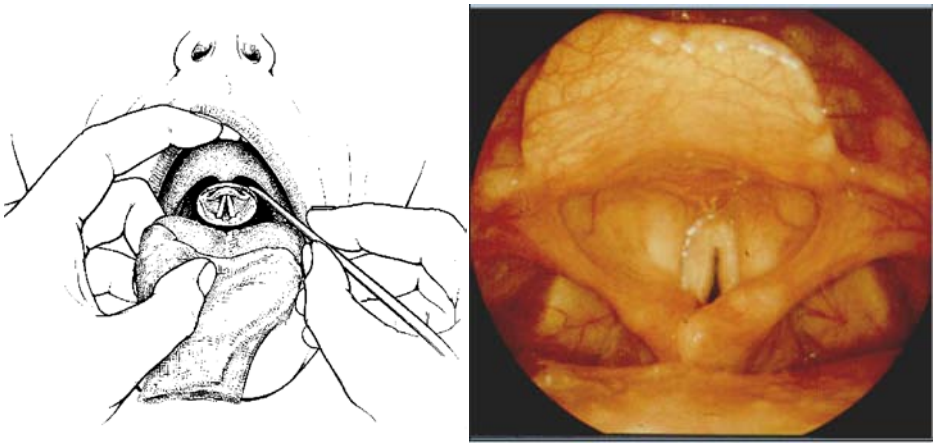
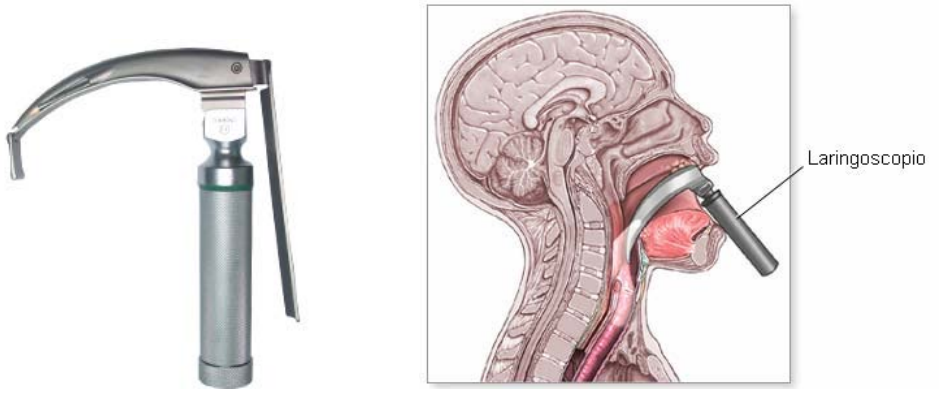


Figura 2.8. (a) Procedimiento para realización de laringoscopia indirecta. (b) Imagen obtenida de la realización de laringoscopia indirecta.

El procedimiento se lleva a cabo utilizando espejos y una fuente de luz dirigidos hacia el interior de la garganta (laringoscopia indirecta) o introduciendo un instrumento delgado (laringoscopio) a través de la nariz o la boca hasta la garganta en el caso de la laringoscopia por fibra óptica o la directa.
El laringoscopio es un instrumento médico simple que sirve principalmente para examinar la glotis y las cuerdas vocales. La Figura 2.9 muestra esquemáticamente el funcionamiento del laringoscopio.



(a) (b)

Figura 2.9. (a) Laringoscopio. (b) Esquema de posicionamiento del laringoscopio para observación de las cuerdas vocales.

El aparato se compone de una hoja que sirve para apartar la lengua y la epiglotis y un mango para manejar el instrumento. Al final de la hoja se encuentra usualmente una fuente luminosa e incluyen una pequeña cámara de video que permite ir observando en tiempo real y tamaño aumentado las cuerdas vocales y la glotis. Aún así, el laringoscopio por sí solo no permite visualizar el movimiento de las cuerdas vocales, se necesita combinar con estroboscopia o alta velocidad.

Los videos utilizados en la aplicación realizada en el trabajo actual han sido obtenidos a partir de un laringoscopio con luz estrobocópica.

2.5.4 Técnicas que permiten el análisis y la extracción de medidas a partir de la exploración de las cuerdas vocales en fonación.

Para finalizar, se hará referencia a otros métodos que, mediante diversas representaciones, permiten también visualizar y analizar el comportamiento dinámico de los pliegues vocales. Además nos van a permitir analizar y extraer medidas a partir de exploraciones realizadas de las cuerdas vocales.

2.5.4.1 *Quimografía*

La quimografía, al igual que la cinematografía de alta velocidad, es una técnica que permite inspeccionar movimientos aperiódicos en la vibración de las cuerdas vocales. El desarrollo de esta metodología se ha visto impulsado por las posibilidades que ofrece para caracterizar de forma objetiva el comportamiento dinámico de los pliegues mediante numerosos parámetros [Manfredi2006; Qiu2003; Sung1999; Švec2002] imposibles de calcular directamente sobre grabaciones de alta o baja velocidad.

El primer sistema de este tipo se denominó **fotoquimografía** y se basaba en una cámara fotográfica modificada de modo que se introducía una abertura estrecha en el frontal de la película. La rendija iba desplazándose continuamente en sentido vertical durante la exposición de la imagen, lo que permitía registrar el movimiento vibratorio de las cuerdas. El principal inconveniente de esta técnica era que, debido al movimiento de la abertura, se hacía imposible el análisis de la vibración en un punto fijo [Baken2000; Švec2002].

En 1994, *J.G. Švec* y su equipo, en colaboración con los laboratorios *Lambert Instruments BV* (Leutingewolde, Holanda), dieron vida a la **videoquimografía** [Švec1996]. El invento consistía en una cámara de vídeo capaz de trabajar en dos modos: uno normal, en el que se graba a 25 o 30 fotogramas por segundo; y otro rápido en el que la cámara funciona a una velocidad de 7.812,5 imágenes por segundo. En este último, el sistema toma una única línea de cada fotograma y va representándolas una debajo de otra en un monitor de televisión, al mismo tiempo que se graban en una cinta de vídeo VHS. El primer modo sirve para seleccionar la posición de la línea a extraer sobre una imagen completa de la laringe.

La idea era conseguir análisis similares a la alta velocidad digital, pero abaratando los costes de ésta usando la tecnología estándar de televisión. El sistema resuelve los dos problemas principales que había encontrado la quimografía hasta este momento: la grabación se obtiene de forma inmediata; y la línea que sintetiza cada fotograma representa siempre la misma posición vertical de las cuerdas vocales. En la Figura 2.10 se muestran dos videoquimogramas grabados con el equipo de *Švec*, junto a un fotograma capturado para seleccionar la línea de interés.

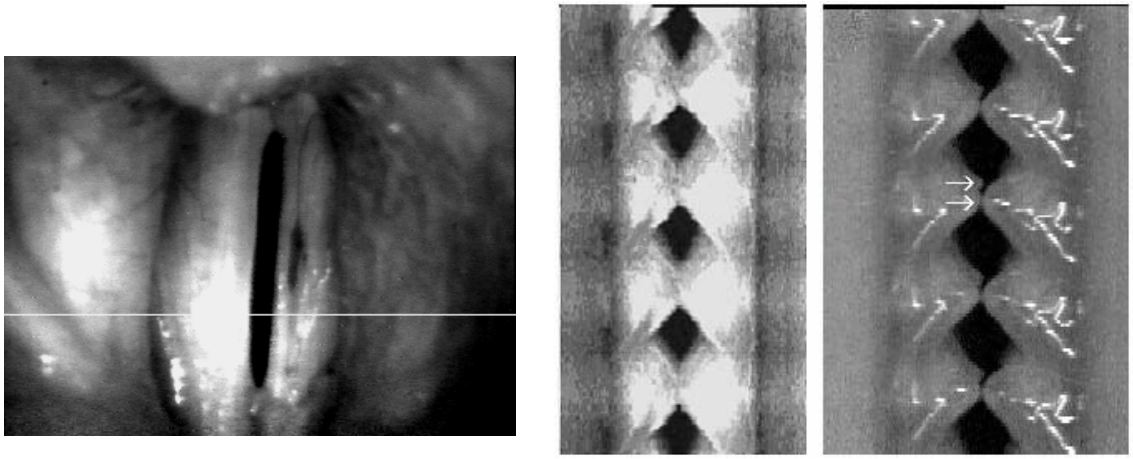


Figura 2.10. (a) Fotograma a resolución completa para selección de la línea de interés. (b) Ejemplos de videoquimogramas, correspondientes a: cuerdas vocales normales (izquierda), cuerdas vocales de un paciente con una leve ronquera (derecha). Figuras tomadas de [Švec1996].

Los principales inconvenientes que presenta la técnica son: en una exploración sólo se puede obtener un quimograma con la línea en una determinada posición; los movimientos de la cámara y/o del paciente durante la exploración introducen errores en la posición de la línea que se está capturando.

El último método que, hasta el momento, permite generar imágenes qimográficas, fue desarrollado en el año 1998 por *J.S. Lee*, *M.W. Sung* y otros investigadores [Lee2001; Sung1999], y lo denominaron **videoestroboquimografía**. En esta ocasión se trata de realizar el mismo proceso que en la videoquimografía pero con procesado de imagen, sobre videos ya grabados, lo cual es fundamental para una buena detección de la glotis.

2.5.4.2 *Fonovibrogramas*

Surgen de la mano de un grupo alemán de investigación [Lohscheller2008] Básicamente, lo que tratan de representar es la distancia ortogonal desde el borde interno de cada cuerda vocal al eje central de la glotis.

Con este método se consiguen representaciones muy claras de las fases de apertura y cierre de los pliegues, lo que permite caracterizar de forma muy precisa el comportamiento dinámico de los mismos (asimetrías de vibración entre cuerdas, periodicidad de vibración entre varios ciclos de un mismo pliegue, relaciones entre las fases de apertura y cierre, relaciones entre la apertura de la parte anterior y posterior, etc). El diagrama no sólo permite observar defectos de cierre e irregularidades de vibración a lo largo de toda la cuerda, sino que, por comparación de perfiles, refleja perfectamente las asimetrías que pudieran existir en el movimiento de un pliegue con respecto al otro

La Figura 2.11 muestra un ejemplo de fonovibrograma. El color rojo intensifica su luminosidad para representar las mayores aperturas, mientras que se vuelve oscuro donde la glotis está prácticamente cerrada.

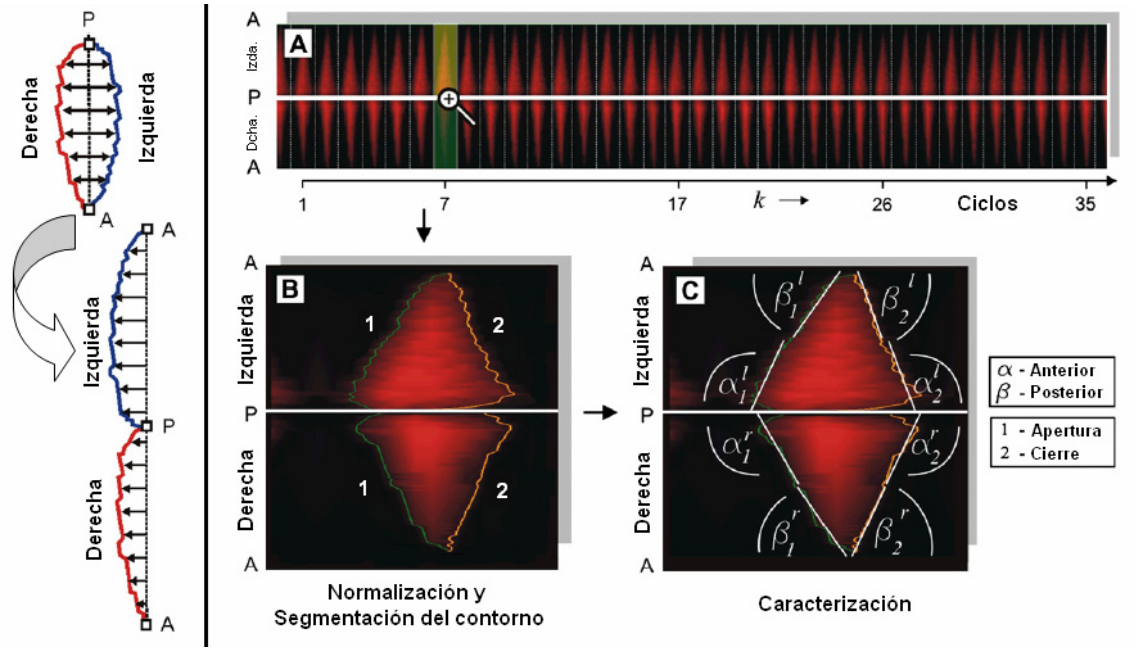


Figura 2.11. Izquierda: Delimitación de los bordes de la glotis y preparación de la síntesis. Derecha: (A) Fonovibrograma completo. (B) Fases de apertura y cierre de las cuerdas vocales representadas sobre el fonovibrograma. (C) Medición de distintos ángulos. Figura adaptada de [Dollinger2009].

2.5.4.3 *Diagramas de área glotal*

El **diagrama de área glotal** ("Glottal Area Waveform" - GAW) permite también caracterizar en cierta medida el comportamiento dinámico de las cuerdas vocales [Woo1996]. Se trata simplemente de un esquema donde se representa el valor del área relativa ocupada por la glotis en cada uno de los fotogramas del vídeo, grabado durante una exploración laríngea.

Los parámetros más comunes que se vienen calculando sobre la onda de área glotal se refieren a: duraciones de apertura y cierre de las cuerdas vocales, pico de área glotal, pendientes de ascenso y descenso de la onda, etc. [Woo1996]. Dos medidas más actuales son las perturbaciones de amplitud y frecuencia, con ellas se llega, incluso, a intentar la discriminación entre voces normales y patológicas [Yan2005; Yan2007].

En la Figura 2.12 puede observarse un ejemplo de una GAW sintetizada por el equipo de investigación de *Y. Yan* en el año 2005.

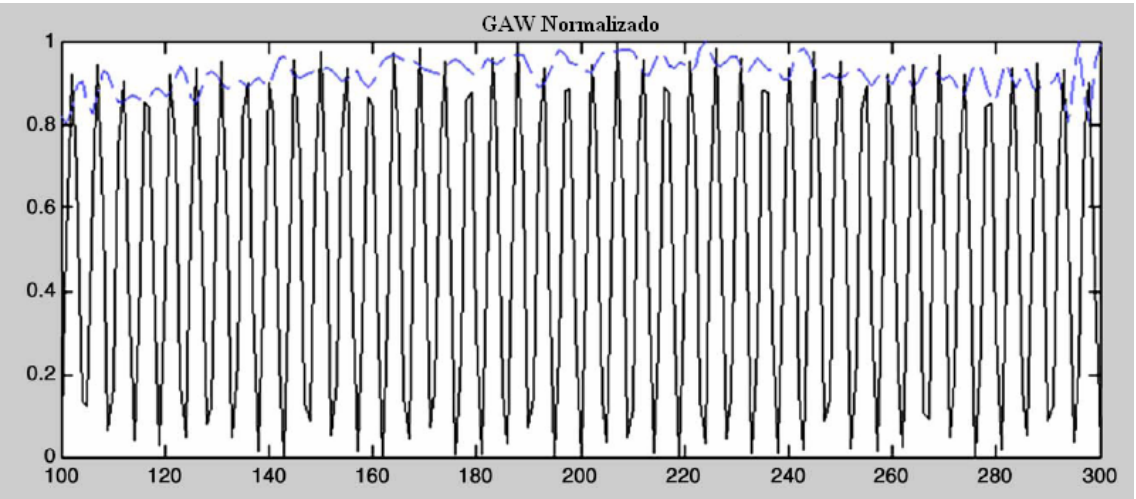


Figura 2.12. GAW de una secuencia de fotogramas correspondiente a una exploración de unas cuerdas vocales normales. Figura adaptada de [Yan2005].

Capítulo 3

Técnicas orientadas a la segmentación de estructuras laríngeas.

3.1 INTRODUCCIÓN

Para poner en práctica la mayor parte de las técnicas descritas en el capítulo anterior se necesita un proceso de tratamiento digital de imagen que permita alcanzar la representación inherente a la técnica.

La segmentación de la glotis es una operación fundamental del proceso y suele constituir uno de los primeros elementos de tratamiento a realizar sobre la imagen, imprescindible para cualquier otra operación posterior. La segmentación se define como el proceso de partición de la imagen en regiones que no estén solapadas, de forma que cada región sea homogénea y la unión de regiones adyacentes no lo sea.

La detección de la glotis en imágenes laríngeas no es una tarea sencilla, y sin embargo resulta una operación fundamental para el cálculo de numerosos parámetros de fonación ya sea directamente o a través de alguna representación (forma de onda glotal, fonovibrogramas, quimogramas…). Todas estas utilidades han hecho necesario el desarrollo de numerosas técnicas de tratamiento digital de imagen orientadas a la segmentación de la glotis de forma más o menos automática, desde las técnicas basadas en procesado clásico de imágenes hasta los modernos contornos activos [Yan2005]. El principal problema de todas estas técnicas es que resultan muy dependientes del punto de inicialización del proceso de segmentación, además de ser altamente sensibles al ruido.

Como ejemplo de segmentación, en la figura 3.1 se pueden observar algunas de las operaciones necesarias para sintetizar un quimograma digital: rectificación del ángulo de giro de la glotis; compensación de los desplazamientos que hubiese podido sufrir la glotis de un fotograma a otro; extracción de la línea deseada; y representación de todas las líneas obtenidas en forma de quimograma.

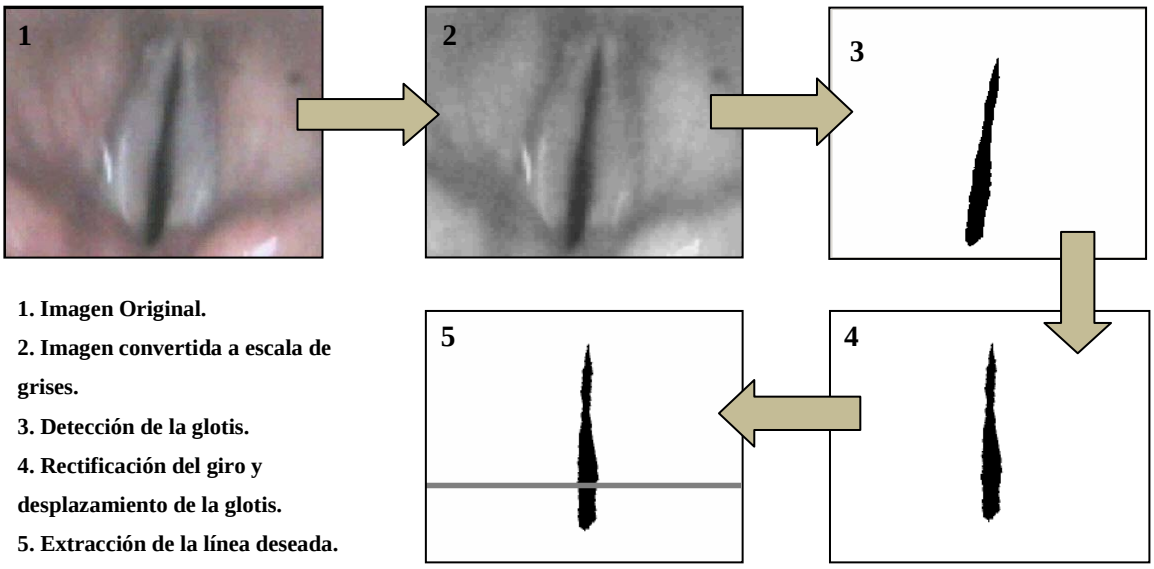


Figura 3.1. Resumen de las operaciones de tratamiento de imagen necesarias para sintetizar un quimograma.

No es objeto de este capítulo describir todos los métodos de segmentación existentes, ya que resultaría un tema demasiado extenso y complejo: existen cientos de técnicas en la literatura, pero no hay una que se pueda considerar válida para todos los tipos de imágenes; ni siquiera todos los métodos son equivalentes para la misma clase de imágenes. Por tanto, se dedicará esta sección a presentar la descripción y el estado del arte de las técnicas usadas para la segmentación de imágenes laríngeas en el trabajo actual, principalmente de técnicas avanzadas como la transformada "Watershed" y el "Merging".

3.2 TRANSFORMADA "WATERSHED"

La transformada "Watershed" es una de las herramientas más valoradas en el campo de la segmentación digital de imágenes ya que permite la segmentación de estructuras complejas que no pueden ser procesadas mediante otros métodos convencionales de procesamiento digital de imágenes.

El concepto de "Watersheds" proviene del campo de la topografía, concretamente hace referencia a la división de un terreno en función de sus cuencas de recepción de agua (o hidrográficas). La línea que separa las cuencas se denomina "línea Watershed". Desde este punto de vista, se puede considerar la imagen como una superficie topográfica donde cada píxel constituye un punto del espacio, situado a una altura u otra en función de su nivel de gris [Bleau1992a; Bleau2000; Gonzalez2004]. Tradicionalmente se considera el color blanco (nivel de gris 255 en una representación con 8 bits) como la máxima altura posible en la superficie asociada y el color negro (nivel de gris 0) como la mínima. El resto de valores suponen una altura, en la superficie asociada a la imagen, entre estos dos niveles límite.

A lo largo de los años se han desarrollado principalmente dos procesos para el cálculo de la transformada Watershed:

- Proceso basado en la inundación de la superficie topográfica en estudio, que ha sido previamente agujereada a la altura de sus mínimos [Beucher1979; Beucher1992]. El sistema consiste en sumergir la superficie poco a poco en un recipiente lleno de agua de modo que se van inundando las cuencas hidrográficas asociadas a cada mínimo. En un

momento determinado el agua procedente de la inundación de dos o más cuencas diferentes puede converger, lo que se impedirá mediante la construcción de un dique justo en ese lugar. Así, una vez que toda la superficie ha sido sumergida, sólo aflorarán los diques, que constituirán las líneas "Watershed" de la imagen. Las "watersheds" son, por tanto, cada una de las zonas delimitadas por las líneas anteriores.

- Simulación de un proceso de lluvia sobre la superficie asociada a la imagen [Osma-Ruiz2007]. Las gotas de agua que caigan sobre cada uno de los puntos de la superficie fluirán, siguiendo el camino descendente con mayor desnivel, hasta alcanzar un determinado mínimo. Ese punto es etiquetado entonces como perteneciente a la cuenca hidrográfica de dicho mínimo. Debido al proceso seguido, ningún punto pertenecerá explícitamente a una línea "Watershed", al ser todos etiquetados como pertenecientes a una determinada cuenca. Las líneas quedarían, pues, formadas por los bordes de los píxeles que separan las distintas cuencas [Bleau2000].

3.2.1 Transformada "Watershed" por simulación de lluvia.

A continuación se presentan y explican los conceptos más importantes para el proceso de cálculo de la transformada "Watershed" por simulación de lluvia [Bleau1992a; Bleau1992b; Bleau2000] [Osma-Ruiz2007]:

- Mínimo regional: punto, o grupo de puntos conectados con el mismo valor de gris, donde ningún píxel presenta un vecino con menor luminancia a la del conjunto. Es necesario definir una función de conectividad para determinar la vecindad de los puntos (suele ser 4 u 8).
- Camino descendente más pronunciado: es una sucesión de píxeles conectados, con origen en (m,n), donde cada punto presenta un nivel de gris estrictamente inferior al anterior. A partir de un punto se pueden encontrar varios caminos descendentes, ya que es posible que exista más de un vecino con un nivel de gris inferior al del píxel de partida. El más pronunciado es el que va recorriendo las caídas más abruptas.
- Cuenca de recepción de agua: formada por un mínimo regional y todos los puntos cuyo camino descendente más pronunciado acaba en ese mínimo [Bieniek2000].
- Transformada Watershed: matriz de puntos, del mismo tamaño que la imagen, donde cada elemento habrá sido etiquetado como perteneciente a una, y sólo una, cuenca de recepción de agua, a partir del cómputo de los caminos descendentes más pronunciados de cada uno de los píxeles de la imagen.

  La transformada Watershed donde más sentido tiene es sobre la imagen gradiente, ya que en ella se presentan los bordes con nivel de gris alto y las zonas homogéneas con nivel de gris bajo.

  El mayor problema de la transformada Watershed es la sobresegmentación: al aplicar la transformada, la imagen queda dividida en miles de regiones cuando sólo se esperaban unas pocas. Este fenómeno se debe principalmente al ruido de la imagen que genera bordes insignificantes.

  En la figura 3.2 se observa este problema. La imagen (3.2 - a) es un fotograma de unas cuerdas vocales humanas durante la fonación. En el dibujo (3.2 - b) se recoge el resultado de aplicar la transformada "Watershed" al gradiente de la imagen (3.2 - a) convertida a escala de grises.

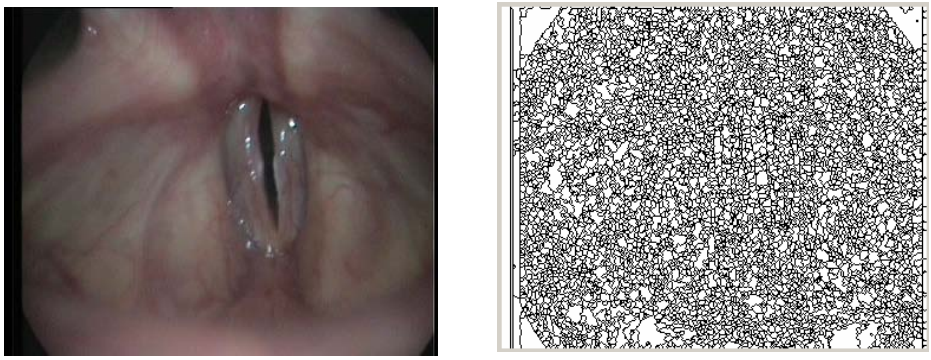


Figura 3.2. Aplicación de la transformada "Watershed" al gradiente de una imagen convertida a escala de grises: (a) imagen original; (b) transformada "Watershed".

La mejor solución a la sobresegmentación se basa en una unión posterior de las cuencas mediante algún proceso de "Merging". Además, existen técnicas de pre-procesado de imagen que pueden ayudar a paliar, en cierta medida, el problema. Lo más usual es aplicar algún tipo de filtrado u operación morfológica con lo que se consiguen eliminar muchos bordes insignificantes que pueden inducir a error o ralentizar el proceso de cálculo posterior.

3.3 OPERACIONES DE "MERGING"

Este método se basa en la unión por pares de las regiones en que previamente se hubiera dividido una imagen, atendiendo a algún determinado criterio de similitud. La transformada "Watershed" podría considerarse una operación de "splitting" inteligente [Gonzalez1992], desde el momento en que la división sigue un criterio lógico, como es la delimitación de bordes a partir del gradiente. La unidad base para el "Merging", en este caso, será la cuenca de recepción de agua.

Todos los métodos de "Merging" se basan en la realización de sucesivas iteraciones sobre la transformada "Watershed". En cada etapa, la herramienta calcula cuáles son las dos cuencas vecinas que pueden unirse con un menor coste y las fusiona. El proceso finalizará cuando sólo queden en la imagen el número de cuencas deseado o cuando el coste de unión supere un umbral establecido.

La técnica que se ha implementado en el presente trabajo se basa en el uso de dos tipos de grafos [Bueno2001; Haris1998; Shen2003], uno de ellos dirigido: el grafo de regiones adyacentes ("Region Adjacency Graph" - RAG); y el grafo de vecinos más cercanos ("Nearest Neighbor Graph" - NNG), que es el dirigido.

El RAG de una partición *K* de la imagen se define como un grafo *G* no dirigido formado por *K* nodos y un conjunto de bordes que los interconectan. Los nodos representan las cuencas en que está dividida la imagen, mientras que los bordes (*E*) conectan una región con todas y cada una de sus vecinas.

Además, a cada borde que une dos cuencas adyacentes se le asigna un peso que viene definido en función de la similitud de dichas cuencas.

Estudiar la unión de regiones directamente sobre el RAG supone una gran cantidad de operaciones, debido al elevado número de bordes que suelen aparecer en el gráfico. Para mejorar el rendimiento de la operación de "Merging" se define otro grafo, esta vez dirigido: el NNG. En este solamente se consideran aquellos bordes que suponen un menor coste de unión. De cada cuenca partirá, por tanto, uno de estos bordes, hacia su vecina más similar.

En la figura 3.3 se puede observar un ejemplo para cada uno de los dos grafos mencionados.

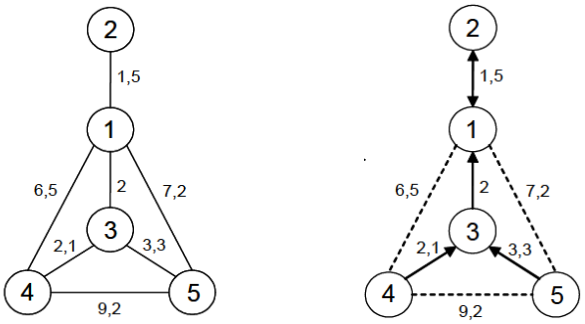


Figura 3.3. Cálculo del NNG de una imagen a partir de su grafo RAG.

El proceso de “Merging” quedaría como sigue [Osma-Ruiz2010]:

1. Se crea la estructura RAG apuntando los parámetros característicos de cada cuenca, así como una lista de sus vecinas y la vecina de menor coste (flechas dirigidas del grafo NNG).
2. Se crea la lista de CICLOS-NNG mediante escaneado de todas las cuencas, observando qué regiones se apuntan entre sí como vecinas de menor coste de unión (valor 1,5 en la figura 3.3). Esta lista se mantendrá siempre ordenada de menor a mayor coste.
3. Se extrae el primer ciclo de la lista y se realiza la unión de los nodos implicados. Esto afectará a la estructura RAG, primeramente porque será necesario asignar una etiqueta a la nueva cuenca fusionada. Además, habrá que recalcular los parámetros característicos de la región fusionada (nivel de gris, área, etc.) y actualizar la lista de vecinas con todas las que existían en las dos cuencas de origen.
4. La actualización del NNG es algo más compleja que la del RAG, realizada en el ítem anterior. Primero, se deben estudiar, en el nuevo RAG, todas las vecinas L_i de la cuenca fusionada para comprobar cuál es la nueva vecina de menor coste V_i de cada una de esas L_i. Si alguna V_i ha cambiado con respecto al momento anterior a la unión, habrá que actualizar la lista CICLOS-NNG para eliminar un posible ciclo en el que interviniese la cuenca L_i asociada (además de apuntar V_i como vecina de menor coste de L_i).
5. Como segundo paso de actualización del NNG, se comprobará cual de las vecinas de la nueva cuenca es la de menor coste, para apuntarla en el parámetro correspondiente de la estructura RAG reutilizada.
6. En el último paso de actualización del NNG, se deben recorrer de nuevo todas las vecinas L_i de la cuenca fusionada, comprobando si alguna de ellas se apunta mutuamente con su vecina de menor coste. Esto constituiría un nuevo ciclo que habría que incluir en la lista ordenada CICLOS-NNG.
7. Una vez actualizadas las estructuras, se volverá al paso 3 para realizar una nueva iteración de fusión, siempre que no se haya alcanzado alguna de las condiciones de parada: número de objetos deseado o superación de un umbral de coste establecido por parte del nuevo ciclo a considerar.

La mayor dificultad en el uso del “Merging”, en conjunción con la transformada “Watershed”, está en la elección de las características de la imagen a incluir en la función de coste, para lograr que el objeto deseado (en este caso la glotis) aparezca entre los finalmente segmentados. Posteriormente, se pueden discriminar unos objetos de otros en función de distintos parámetros (factores de forma, momentos, niveles de gris, etc.).

3.3.1 Merging JND

En este tipo de "Merging" la función de coste se calcula según el JND de los distintos niveles de gris de la imagen. El JND es una medida de la sensibilidad del sistema visual humano a los cambios de luminancia, basada en la incapacidad del ojo para diferenciar determinados cambios de nivel de gris. Por ejemplo, una persona que visiona una imagen en escala de grises no sería capaz de distinguir entre un nivel 80 y un nivel 85. Además, esta insensibilidad no sigue una pauta lineal, siendo el ojo menos sensible a los cambios de luminancia en valores oscuros que en claros.

En [Yang2005] se puede encontrar una expresión - ecuación (3.1) - que facilita el umbral de visibilidad T, en función de los distintos niveles de gris de una imagen I, por debajo del cual el ojo no es capaz de detectar los cambios de luminancia. Por supuesto, el nivel de gris 0 representa el negro y el 255 el blanco. La Figura 3.4 muestra una representación gráfica de la ecuación, en la que puede observarse claramente la gran insensibilidad que presenta el sistema de visión humano ante los cambios de niveles de gris en zonas oscuras, mientras que se comporta mejor para altas luminosidades.

$$T(n,m)=\begin{cases} 17\cdot\left(1-\sqrt{\frac{I(n,m)}{127}}\right)+3 & \text{Si } I(n,m)\le 127 \\ \frac{3}{128}\cdot\left(I(n,m)-127\right)+3 & \text{Resto} \end{cases} \qquad (3.1)$$

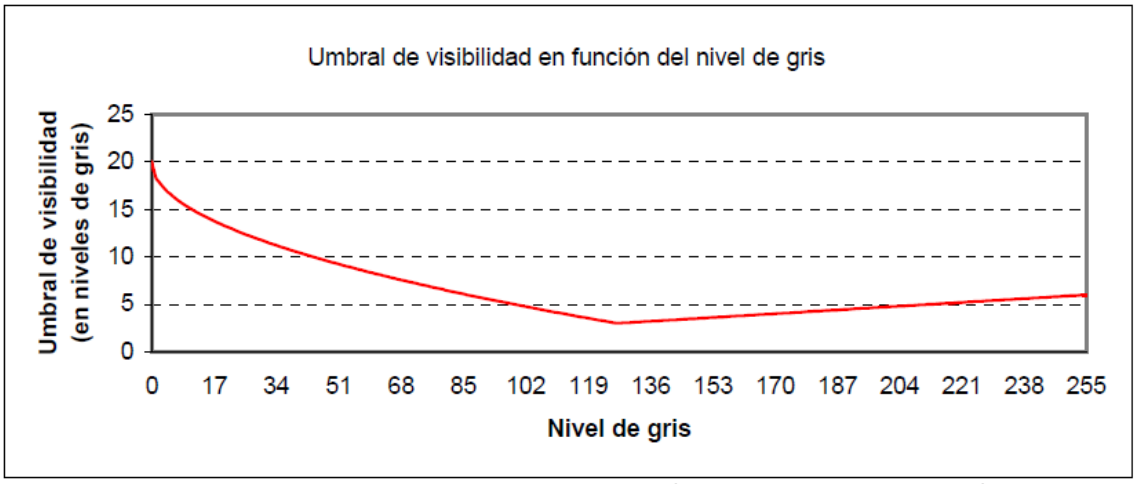


Figura 3.4. Umbral de visibilidad del sistema visual humano en función de los niveles de gris de los píxeles de una imagen.

El JND permite la unión de regiones con una flexibilidad mayor, como lo haría un observador humano, incapaz de distinguir entre niveles de gris suficientemente cercanos.

3.4 MOMENTOS

Los momentos son una familia de descriptores matemáticos de formas muy efectivos, que permiten extraer información de la masa total del objeto. Su significado es muy intuitivo y su cálculo no es demasiado lento.

Por contra, no son demasiado sensibles a los detalles finos; el cálculo de los momentos de orden alto, que son los que proporcionan mayor información, es complicado e inestable

numéricamente. Pero muchas aplicaciones, como es aquí el caso, tienen suficiente con los de orden bajo. Sólo depende de cuánto detalle se necesite y de cuánto ruido haya en el sistema.

Para una función continua bidimensional $F(x,y)$, el momento de orden $(p+q)$ está definido por:

$$m_{pq} = \iint_{-\infty}^{\infty} x^p y^q F(x,y) dxdy \qquad p,q = 0,1,2,\ldots$$

Hay varios tipos de momentos: **momentos binarios**, los cuales, al prescindir de la luminancia, pasan a ser meros descriptores de las características estructurales del objeto y **momentos centrales** que son aquellos expresados con el origen en el centro de masas del objeto.

Para hacer que los momentos calculados sean independientes del tamaño del objeto, hay que normalizarlos. El factor empleado es el momento M_{00}, el cual representa una medida de área, coincidiendo exactamente con esta cuando se trabaja con los momentos binarios, siempre suponiendo que la unidad de área es el píxel cuadrado.

Para que los momentos de una forma no dependan de su orientación, hay dos enfoques posibles: el primero consiste en girar arbitrariamente los objetos hasta colocarlos en una posición estándar. El inconveniente de este enfoque es que hay casos en que es difícil encontrar esa posición, debido a que algunas formas presentan simetrías que las hacen similares en diversas posiciones rotacionales. El segundo enfoque es utilizar una familia especial de momentos llamados **invariantes** [Gonzalez1992], que no se ven afectados por la rotación.

3.5 PREDICTOR LINEAL

Se trata de una sencilla técnica de clasificación que permite reconocer y asignar nuevos elementos dentro de clases previamente definidas, mediante un entrenamiento supervisado [Duda2001].

Cada uno de los elementos viene caracterizado por un conjunto de parámetros escogidos por su capacidad discriminante, de manera que permitan distinguir las distintas clases en estudio. A partir de una serie de unidades, así descritas, de las que se conoce a que clase pertenecen, se genera un modelo matemático discriminante contra el cual será contrastado un nuevo elemento de clase desconocida, para, en función de un resultado numérico, asignarlo a la clase más probable.

Un método de predicción lineal muy común es la función discriminante de *Fisher*, que permite clasificar elementos en dos clases mediante la función lineal recogida en la ecuación (3.2).

$$D = \sum_{i=0}^{N} u_i X_i \qquad (3.2)$$

Donde los X_i representan los distintos parámetros que definen a los elementos y las variables u_i son calculadas mediante entrenamiento supervisado para que la función lineal discriminante D separe lo mejor posible a los elementos conocidos de ambas clases.

La función lineal de *Fisher* produce una proyección del subespacio de parámetros inicial sobre un subespacio de dimensión uno buscando la mayor separabilidad posible entre las clases. La clasificación se realizará entonces estableciendo un umbral óptimo de decisión sobre este subespacio [Johnson1998].

La Figura 3.5 muestra un ejemplo de histograma de la función de *Fisher* para dos clases de elementos. En el eje de abscisas se representan los distintos valores que toma la función de $Fisher$ para todos y cada uno de los elementos, y en el eje de ordenadas el número de unidades que toman valor en pequeños rangos de la función. N es el número total de elementos considerados.

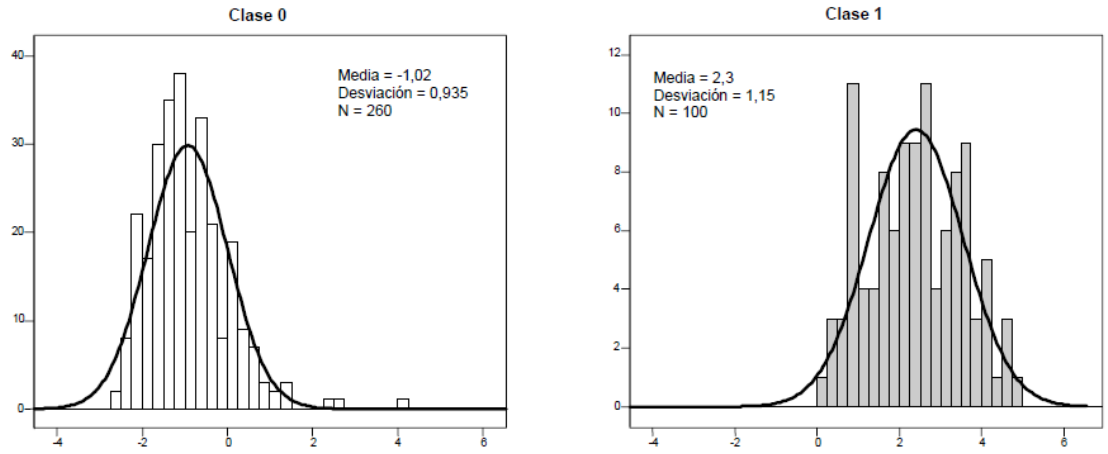


Figura 3.5. Ejemplo de histograma de los valores entregados por la función discriminante de *Fisher* para dos clases de elementos conocidos: (a) elementos de la clase 0; (b) elementos de la clase 1.

A la vista del histograma de la Figura 3.5 puede establecerse cuál es el mejor umbral que separa ambas clases de elementos, que no tiene por qué coincidir con la media de los centroides de ambas clases. Si, por ejemplo, se decidiese desplazar el umbral hacia la derecha de la media, se estará siendo más restrictivo para clasificar unidades como pertenecientes a la clase 1, por lo que muchos elementos que pertenecen realmente a esta serán clasificados como de clase 0.

Como ventaja se consigue que muy pocos elementos pertenecientes realmente a la clase 0 sean clasificados como pertenecientes a la clase 1. El posicionamiento de este umbral depende mucho del coste de decisión que se está dispuesto a asumir, ya que, desde el momento en que ambas clases se solapan, hay que contar con una determinada probabilidad de cometer errores, clasificando dentro de una clase unidades que realmente pertenecen a la otra.

3.6 MÉTODO DE DETECCIÓN DE LA GLOTIS UTILIZADO

El objetivo de este apartado es describir la técnica utilizada para la detección de la glotis en imágenes laríngeas en el trabajo actual. La técnica comienza con la aplicación de la transformada "Watershed" para realizar una división inicial de la imagen; a continuación se aplican dos procesos "Merging" que realizan una selección para quedarse con unos cuantos objetos, entre los que debe encontrarse la glotis. Finalmente, se realiza un proceso de predición lineal que realizará la discriminación final.

El método seguido para lograr la individualización de la glotis es el que se presenta de forma esquematizada en la Figura 3.4 [Osma-Ruiz2008a; Osma-Ruiz2008b]. Cada uno de los bloques se describe de forma detallada en las siguientes páginas.

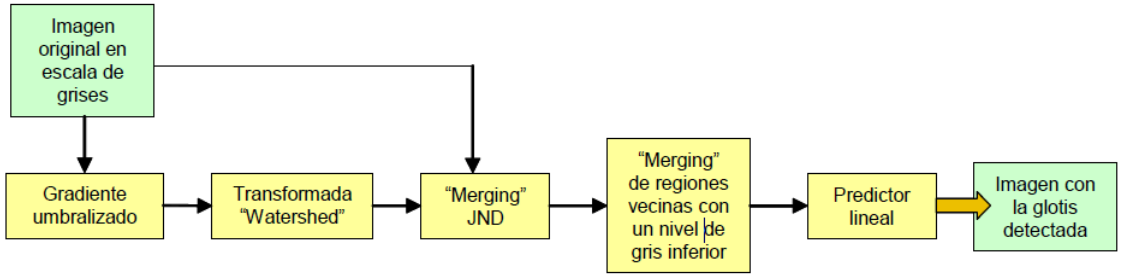


Figura 3.6. Esquema del proceso seguido para la detección de la glotis.

Transformada "Watershed" de la imagen gradiente

En el primer paso se convierte la imagen original en color (RGB) a escala de grises, mediante una transformación al modelo YIQ del que se toma la luminancia *Y* [Gonzalez1992].

Posteriormente se obtiene la transformada "Watershed" del gradiente de la imagen, previamente umbralizado con un valor de 2. Es decir, aquellos píxeles del gradiente que no superan el valor 2 son colocados a cero y, por tanto, son convertidos en mínimos que sólo podrán pertenecer al interior de una cuenca de recepción de agua. De esta forma se consigue reducir el número inicial de regiones en aproximadamente un 20%, eliminando regiones insignificantes debidas principalmente a la presencia de ruido en la imagen.

"Merging" basado en JND

El segundo paso consiste en una operación de "Merging" basada en JND. En este tipo de "Merging" la función de coste se calcula según el JND de los distintos niveles de gris de la imagen. El JND [Shen2003] es una medida de la sensibilidad del sistema visual humano a los cambios de luminancia, basada en la incapacidad del ojo para diferenciar determinados cambios de nivel de gris.

Tras el paso 2 la imagen quedará segmentada como muestra el ejemplo de la Figura 3.5. Puede observarse como la glotis aparece perfectamente segmentada, además de otras zonas de la imagen con nivel de gris homogéneo para el ojo. Las regiones claras de la imagen quedan unidas, como si de un único objeto se tratara.

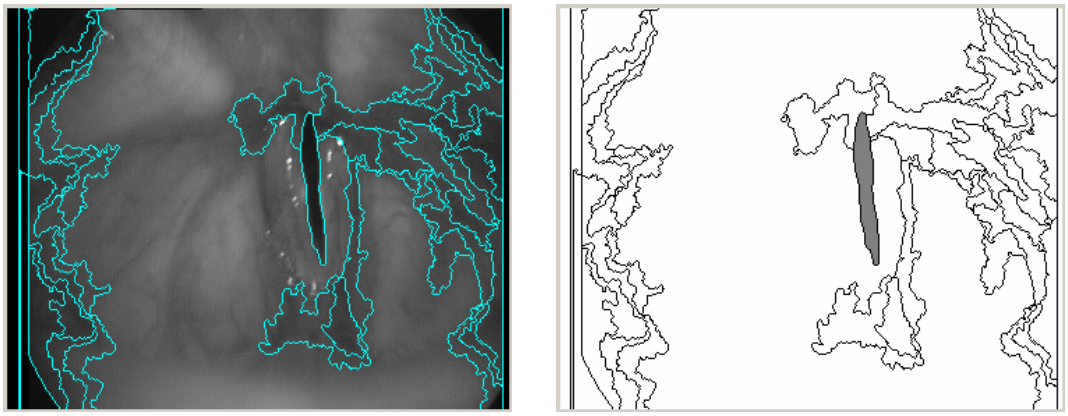


Figura 3.7. Ejemplo de segmentación de la glotis tras el primer proceso de "Merging": (a) imagen original en escala de grises con divisiones "Watersheds" sobre-impresionadas; (b) divisiones "Watersheds" con la glotis resaltada en gris para una mejor distinción.

Merging" de regiones vecinas con nivel de gris inferior

El tercer paso consiste en una segunda operación de "Merging", pero esta vez destinada a unir regiones que posean alguna vecina con nivel medio de gris inferior al suyo propio. El objetivo es reducir el número de objetos segmentados fusionando y, por tanto, descartando aquellas cuencas que no pueden ser la glotis. El sistema se basa en las decisiones que tomaría un observador humano al visualizar una imagen de la laringe, que esperaría encontrar la glotis como un objeto oscuro rodeado de tejidos más claros.

El proceso seguido para llevar a cabo esta operación de "Merging" es como sigue: para las dos cuencas adyacentes que se están tratando de unir, se comprueba si alguna de ellas no tiene

ningún vecino con un nivel de gris inferior; en este caso dicha cuenca es candidata a ser etiquetada como glotis y no se permitirá la unión; si ambas regiones presentan algún vecino con luminancia inferior, ninguna de las dos puede ser la glotis y se permitirá su fusión.

La división obtenida tras el tercer paso será como la mostrada en el ejemplo de la Figura 3.6, en la que se puede observar que el número de objetos presentes en la imagen es ya muy reducido. Entre ellos puede distinguirse, además de la glotis, el objeto de fondo (el más grande de la imagen), sombras (alguna con nivel medio de gris menor que el de la glotis, e.g. en la esquina superior derecha con nivel 22,1, frente a 37,4 de la glotis) y 2 ó 3 objetos muy oscuros (con nivel de gris inferior a 10) introducidos en los laterales de la imagen por defectos en el sistema de grabación de vídeo.

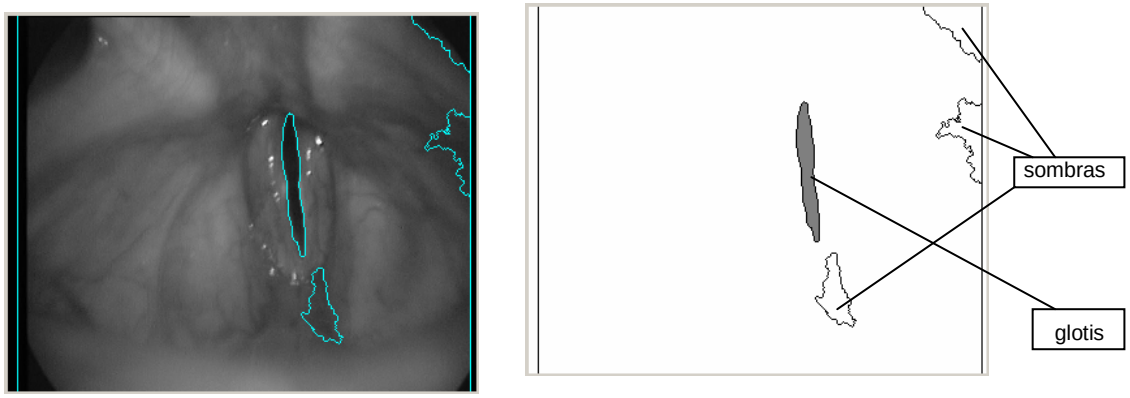


Figura 3.8. Ejemplo de segmentación de la glotis tras el segundo proceso de "Merging" (tercer paso del sistema): (a) imagen original en escala de grises con divisiones "Watersheds" sobre-impresionadas; (b) divisiones "Watersheds" con la glotis resaltada en gris para una mejor distinción.

Bloque de selección. Predictor lineal

El último paso está constituido por un proceso de clasificación mediante el que se pretende distinguir la glotis del resto de objetos presentes en la imagen en función de su forma. Los defectos laterales y el objeto de fondo son fáciles de eliminar ya que tienen características significativas muy diferentes a las del espacio glotal. Los artificios laterales se eliminan por tener un nivel medio de gris demasiado bajo, valor que nunca alcanza la glotis. El objeto de fondo se elimina porque es el que mayor área tiene de la imagen.

Para distinguir entre sombras y glotis se entrena un clasificador de *Fisher* discriminante con el 88% de las imágenes disponibles (98 fotogramas recogidos de 13 vídeos sobre un total de 15), lo que entrega 263 objetos sombra y 98 objetos glotis. El resto de las imágenes (13 fotogramas de los 2 vídeos restantes) se dejan fuera para posteriormente validar los resultados. La decisión de conservar los fotogramas de 2 vídeos exclusivamente para validación fue tomada con el fin de testear lo mejor posible el comportamiento del sistema ante variaciones de iluminación tanto intra-video como inter-vídeo. Además, los vídeos elegidos para prueba son representativos de dos tamaños de glotis distintos.

Las variables para la discriminación serán los 7 momentos invariantes y los 7 momentos invariantes binarios de los objetos que se tratan de distinguir [Gonzalez1992]. La idea es seleccionar la glotis dando la mayor importancia a su forma.

A continuación se muestran varios ejemplos de detección de la glotis después de la realización de todo el proceso completo (Figura 3.7). El resultado final con la glotis que se ha detectado aparece resaltado en color amarillo.

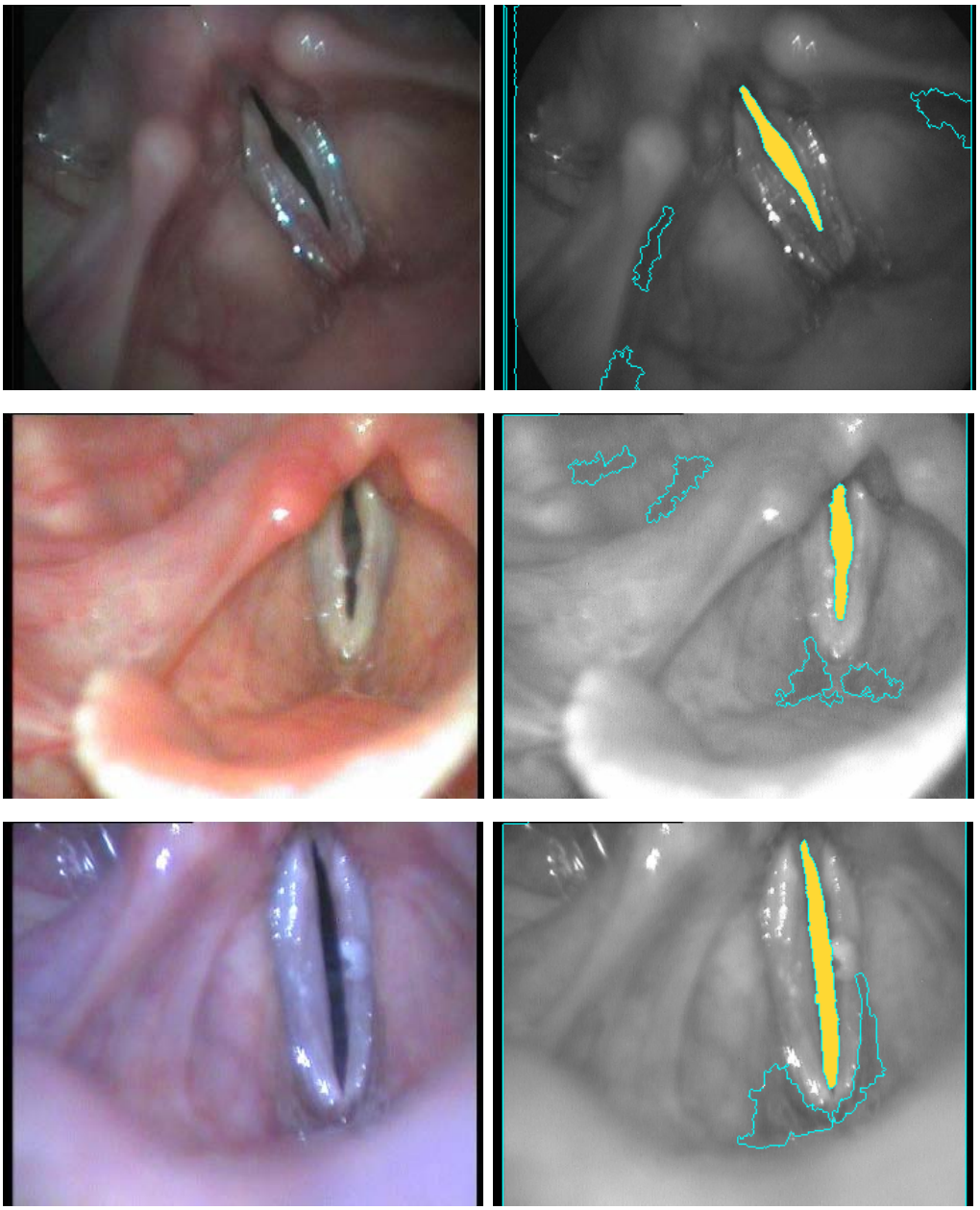


Figura 3.9. Ejemplos de segmentación y detección de la glotis en imágenes laríngeas con distintas condiciones de iluminación, calidad, tamaño y posicionamiento del espacio glotal

Con el proceso descrito, se detecta apropiadamente el espacio glotal, sin necesidad de inicialización, en el 98,98% de las casos de entrenamiento y en todos los casos reservados para validación.

El sistema funciona correctamente con distintas condiciones de iluminación y es capaz de detectar la glotis aunque ésta se encuentre parcialmente cortada en un lateral de la imagen debido a movimientos de la cámara y/o el paciente. Los resultados son también correctos cuando la glotis se encuentra dividida en dos por algún tipo de patología, aunque, en este último caso, suele ser necesario retocar el umbral de coste del primer proceso de “Merging” debido a las pequeñas dimensiones del espacio glotal.

Capítulo 4

Imágenes Dicom

4.1 INTRODUCCIÓN

DICOM (*Digital Imaging and Communications in Medicine)* es el estándar reconocido mundialmente para el intercambio de imágenes médicas [Anón.2009a]. Fue creado en los años 80 por un comité de la ACR (American College of Radiology) y NEMA (National Electrical Manufacturers Association). Este estándar es el mecanismo de codificación, almacenamiento y transmisión de imágenes aceptado universalmente por la comunidad médica. La cabecera de este formato permite almacenar información sobre el paciente, las condiciones en las que se tomó la imagen, y el formato interno de ésta.

La introducción de diferentes imágenes médicas en los sesenta y el uso de los ordenadores para el procesamiento de las mismas una vez adquiridas, llevó a ACR y NEMA a formar un comité conjunto para crear un método estándar para la transmisión de imágenes médicas y su información asociada. Este comité, formado en 1983, publicó en 1985 la versión 1.0 del estándar ACR-NEMA. Antes de esto, la mayoría de los dispositivos almacenaban las imágenes en un formato propietario y transferían ficheros de estos formatos propietarios a través de una red o en dispositivos de almacenamiento portátiles para llevar a cabo la comunicación de las imágenes.

En el año 1988 se publicó la versión 2.0 del estándar ARC-NEMA [Anón.2009a] que incluía a la versión 1.0 y sus revisiones así como comandos para mostrar dispositivos, para introducir nuevos esquemas de jerarquía de tipos de imágenes y para añadir información, consiguiendo de esta forma una descripción más específica de la imagen.

Con el lanzamiento de la versión 3.0 [Com08] se cambió el nombre a *Digital Imaging and Communications in Medicine* (DICOM), y se añadieron numerosas mejoras para las comunicaciones estandarizadas.

Lo que diferencia a las imágenes DICOM de otros ficheros de datos es que, además de la imagen en sí, incluyen información sobre la misma tal como el paciente al que corresponde o el médico que ha realizado la prueba. De esta manera, una radiografía contiene el ID del paciente evitando que la imagen pueda ser separada por error de su información. Esta característica hace que sea usado en el campo de la medicina ya que una imagen médica no tiene sentido por sí

sola, son necesarios los datos del paciente. Además, facilita la comunicación entre profesionales, ya que todos trabajarán con el mismo formato.

DICOM ha sido adoptado ampliamente por hospitales y está haciendo incursión poco a poco en las consultas particulares de los médicos.

En la figura 4.1 se muestra un ejemplo de imagen DICOM perteneciente a un TAC de la cabeza.

Figura 4.1. Ejemplo de imagen DICOM perteneciente a un TAC de la cabeza

4.2 ESTRUCTURA DEL FICHERO DICOM

Un archivo DICOM contiene por un lado la información de una o varias imágenes y por otro la información del contexto en el que se ha tomado la imagen. En el contexto de una imagen DICOM podemos encontrarnos con datos del paciente (nombre, apellidos, edad,…), del doctor que manda la prueba, del centro médico donde se realiza la prueba, de la prueba médica a la que corresponde la imagen, de la máquina que ha realizado la toma (parámetros de configuración de la máquina como por ejemplo la posición del paciente en cada toma), de las imágenes tomadas(número de tomas realizadas, separación entre cada imagen, dimensión de las imágenes,…). DICOM une la información visual con la información recogida en la prueba médica. Esta fusión es una de las mayores ventajas que posee el formato porque puede ser usada para mostrar y ubicar en pantalla información que de otra forma no sería posible.

El formato de fichero DICOM es muy complejo, debido a la gran cantidad de campos que se especifican en la cabecera, así como los varios tipos de cabecera que permite y la multitud de formatos en los que puede estar grabada la image.

Desde el punto de vista del implementador, un fichero DICOM se puede dividir en cuatro partes diferenciadas [Com08]:

- Preámbulo y prefijo identificativo del fichero.
- Meta-cabecera.
- Cabecera.

- Imagen (aunque desde el punto de vista del formato, la imagen es un elemento más de la cabecera).

4.2.1 Preámbulo y prefijo

El estándar DICOM especifica que un fichero en este formato tiene que comenzar con un preámbulo.

Este preámbulo tiene un tamaño fijo de 128 bytes y está pensado para tener un uso definido por la implementación. Por ejemplo, puede contener información sobre el nombre de la aplicación usada para crear el fichero, o puede tener información que permita a aplicaciones acceder directamente a los datos de la imagen almacenada en el fichero.

El estándar DICOM no especifica la manera en la que tienen que estar estructurados los datos en el preámbulo por lo que es el implementador el que se encarga de diseñarlo. Se puede optar por no usar el preámbulo y en este caso todos los bytes del mismo se deben poner a 0.

Lo que sigue al preámbulo es el prefijo que identifica a los ficheros DICOM. Este prefijo está compuesto con cuatro bytes que contienen la cadena de caracteres DICM. El propósito de este prefijo es poder identificar en la implementación si se trata de un fichero DICOM o no.

4.2.2 Meta-cabecera

Contiene datos generales sobre el fichero DICOM en cuestión. Uno de los más importantes es el referente a la Sintaxis de Transferencia que es un identificador único (UID) que describe la forma en que se va a codificar la cabecera.

En la tabla 4.1 se listan algunas de las sintaxis de transferencia existentes, junto con su identificador único y su descripción. El VR o Representación del Valor es un campo de los elementos de datos que está relacionado con la forma de codificación de los mismos, como podremos ver más adelante en el apartado dedicado a los elementos de datos de la cabecera.

Sintaxis de Transferencia	Identificador único	Descripción
Implicit VR *Little Endian*	1.2.840.10008.1.2	Sintaxis de transferencia por defecto. Para formato de imagen no encapsulado.
Explicit VR *Little Endian*	1.2.840.10008.1.2.1	Para formato de imagen no encapsulado (sin compresión). Se especifica el VR de cada elemento de la cabecera con codificación *Little endian* (el primer byte es el menos significativo).
Explicit VR Big *Endian*	1.2.840.10008.1.2.1	Para formato de imagen no encapsulado (sin compresión). Se especifica el VR de cada elemento de la cabecera con codificación *Big endian* (el primer byte es el más significativo).

Tabla 4.1. Algunas de las Sintaxis de transferencia existentes más utilizadas.

Todos los datos contenidos en la meta-cabecera están estructurados como elementos de datos (Data Elements) y deben ser codificados siempre usando la Sintaxis de Transferencia (Transfer Syntax) Explicit VR Little Endian.

4.2.3 Cabecera

La cabecera de este formato es extremadamente rica y permite almacenar información sobre el paciente, las condiciones en las que se tomó la imagen, y el formato interno de esta. La cabecera de un fichero DICOM es la parte que nos va a proporcionar la gran mayoría de la información que necesitamos para visualizar la imagen correctamente, incluyendo los propios datos de la imagen, ya que éstos son considerados por el estándar como un elemento más de la cabecera.

La cabecera de un fichero DICOM consiste en un conjunto de elementos de datos (Data Set) con toda la información necesaria, y está codificada según la sintaxis de transferencia indicada en la meta-cabecera. Existen cuatro tipos de cabecera según la sintaxis de transferencia: tres para imágenes en formato no encapsulado (como RGB) y una para imágenes en formato encapsulado (como JPEG o RLE).

4.2.3.1 *Elementos de datos*

La cabecera y meta-cabecera están formadas por una serie de campos con información importante referente a la imagen, incluyendo la propia imagen. En estos campos se encuentra información sobre el paciente, sobre el tipo de la imagen entre otras características.

A la información contenida en cada uno de los campos se le conoce como *Data Element* (Elemento de Datos). Un elemento de datos está constituido por los siguientes campos [Anón.2009a]:

- Etiqueta: sirve para identificar cada elemento de datos de manera única. Esta etiqueta está formada a su vez por un Número de Grupo (Group Number) y un Número de Elemento (Element Number). Estos números se presentan en hexadecimal.

  Al nombre del paciente, por ejemplo, corresponde el grupo 10 y el número de elemento 10, es decir, la etiqueta (0010 – 0010).

- Representación del Valor (VR): indica la forma en la que se codifica el valor del elemento. Este valor puede estar codificado como una cadena de caracteres o como enteros sin signo.
- Longitud del Valor (Value Length): como su nombre indica, es la longitud del campo Valor.
- Valor: es el valor del elemento de datos propiamente dicho. Está codificado según el campo VR y con la longitud que indica el campo Longitud del Valor. Siguiendo el mismo ejemplo de antes, en el elemento de datos Nombre del paciente (0010,0010) el valor podría ser “Juan Ruiz” con la longitud del valor igual a nueve y con la representación del valor “PN” (Person Name).

En la figura 4.2 podemos ver la estructura de la cabecera en la parte superior y los campos que forman cada uno de los elementos de datos en la parte inferior.

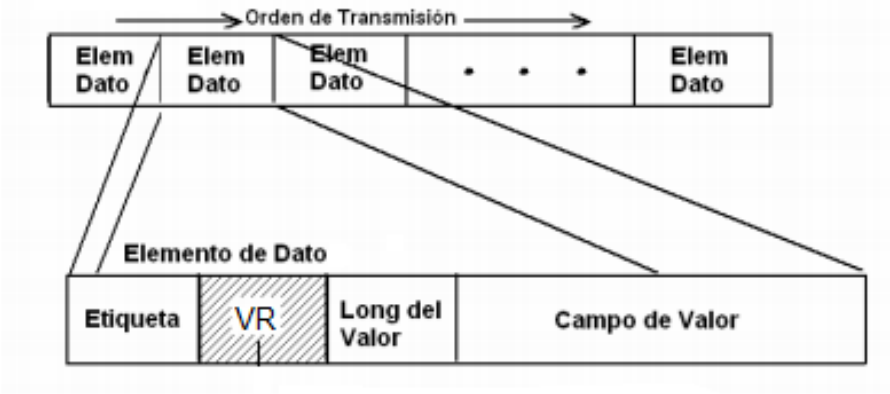


Figura 4.2. Estructura DICOM de los elementos de datos que forman la cabecera

Representación del valor (VR)

El estándar DICOM define una serie de VRs con diferentes características, con la intención de que el campo Valor de cada elemento de datos esté codificado correctamente según lo representa.

A continuación se resumen algunos de los más importantes junto con su descripción [Anón.2009a]:

- AT (Attribute Tag): par ordenado de enteros sin signo de 16 bits (igual que la codificación del campo Etiqueta de Elementos de Datos).
- CS (Code String): cadena de caracteres, siendo los espacios no significativos.
- DA (Date): cadena de caracteres con el formato yyyymmdd, siendo yyyy el año, mm el mes y dd el día.
- DS (Decimal String): cadena de caracteres representando un número en coma fija o flotante.
- FL (Floating Point Single): número en punto flotante de precisión simple.
- IS (Integer String): cadena de caracteres que representa un entero en base decimal.
- OB (Other Byte String): cadena de bytes. La codificación del contenido depende del campo Sintaxis de transferencia.
- OF (Other Float String): cadena de numerous en punto flotante de simple precisión.
- OW (Other Word String): cadena de palabras de 16 bits.
- PN (Person Name): cadena de caracteres de cinco componentes formado por apellidos, nombre de pila, segundo nombre, prefijo, sufijo.
- SQ (Sequence of Items): secuencia de cero o más ítems.
- UL (Unsigned Long): entero sin signo de 32 bits.
- UN (Unknown): cadena de bytes. La codificación del contenido es desconocida.
- UT (Unlimited text): cadena de caracteres que puede contener uno o más párrafos.

Existen dos tipos de codificación para los elementos de datos según la sintaxis de transferencia indicada en la meta-cabecera: VR explícita y VR implícita.

- Codificación VR explícita: dependiendo de su VR, los elementos se codifican de dos formas distintas.
  La tabla 4.1 recoge la codificación utilizada para los elementos en el caso de que su VR sea OB, OW, OF, SQ, UT o UN.

Etiqueta		**VR**		**Logitud del valor**	**Valor**
Número de grupo (entero sin signo de 16 bits)	Número de elemento (entero sin signo de 16 bits)	VR (OB, OW, OF, SQ, UT ó UN)	Reservado (2 bytes con valor 000h)	Entero sin signo de 32 bits	Valor del elemento codificado según la VR
2 bytes	2 bytes	2 bytes	2 bytes	4 bytes	Los bytes que indique el campo longitud del valor

Tabla 4.2. Codificación de los elementos cuya VR es OB, OW, OF, SQ, UT o UN.

En el caso de que la VR de los elementos no sea una de las mencionadas anteriormente la codificación usada en este caso la observamos en la tabla 4.2.

Etiqueta		**VR**	**Logitud del valor**	**Valor**
Número de grupo (entero sin signo de 16 bits)	Número de elemento (entero sin signo de 16 bits)	VR	Entero sin signo de 16 bits	Valor del elemento codificado según la VR
2 bytes	2 bytes	2 bytes	2 bytes	Los bytes que indique el campo longitud del valor

Tabla 4.3. Codificación de los elementos cuya VR sea diferente a la de los elementos de la tabla 4.1.

- Codificación VR implícita: es la que se usa en la sintaxis de transferencia por defecto y se diferencia de la anterior en que la VR de cada elemento de datos no se codifica en el fichero si no que el implementador tiene que consultar los documentos del estándar para saber qué VR corresponde con cada uno de los campos de la cabecera (en función de su etiqueta).
  En este caso, la codificación usada es la siguiente:

Etiqueta		**Logitud del valor**	**Valor**
Número de grupo (entero sin signo de 16 bits)	Número de elemento (entero sin signo de 16 bits)	Entero sin signo de 32 bits	Valor del elemento codificado según la VR
2 bytes	2 bytes	4 bytes	Los bytes que indique el campo longitud del valor

Tabla 4.4. Codificación de los elementos con VR implícita.

Por otra parte, en el documento del estándar [Anón.2009a] vienen todos los campos explicados detalladamente. Para cada uno de ellos, se nos especifica su función, la forma en que se debe codificar su valor y su obligatoriedad de uso.

DICOM divide la obligatoriedad de uso según la siguiente nomenclatura:

- Tipo 1: si un elemento de datos es de tipo 1, éste debe ser incluido obligatoriamente. La longitud del valor no puede ser cero, y debe tener un valor válido según su descripción.
- Tipo 2: si un elemento de datos es de tipo 2, éste debe ser incluido obligatoriamente. Sin embargo, se permite codificar el elemento con una longitud del valor igual a cero y sin campo Valor, solo en el caso de que el valor del elemento de datos de tipo 2 no sea conocido.
- Tipo 3: si un elemento de datos es de tipo 3, éste puede ser incluido opcionalmente, y la ausencia de un elemento de datos de este tipo no supone una violación del protocolo. Además, puede ser codificado con una longitud del valor igual a cero y sin campo Valor.

4.2.4 Imagen

A continuación de los elementos de datos que componen la cabecera se encuentran los pixeles de la imagen. Los datos de pixel de cualquier imagen son almacenados en el elemento de datos (7FE0,0010) utilizando distintas formas de codificación. El orden de lectura de los píxeles depende del algoritmo de compresión utilizado. Para una imagen plana sin compresión, es secuencial y corresponde a una disposición bidimensional de izquierda a derecha y de arriba abajo.

La información referente a las características de la imagen y de la estructura de los píxeles se encuentra en los elementos pertenecientes al grupo 0028. Los elementos que aparecen en este grupo indican el número de filas y columnas de la imagen y el número de bits con el que se representa, entre otros. La lectura e interpretación de estos datos es de suma importancia para mostrar la imagen correctamente.

4.3 ELEMENTOS DE LA CABECERA DEL FICHERO DICOM SIGNIFICATIVOS EN LA APLICACIÓN

Como se ha comentado anteriormente, la cabecera de un fichero DICOM está formada por un conjunto de elementos de datos que tienen información sobre el paciente, el médico e información de la propia imagen.

El número de elementos de datos de la cabecera es muy grande y, aunque en la aplicación se recorren todos y se almacenan, sólo utilizamos algunos de ellos. En la tabla 4.4 se muestran los elementos más usados y que mayor importancia han tenido durante el desarrollo de la aplicación junto con su descripción [Anón.2009a].

ELEMENTO DE DATOS	DESCRIPCIÓN
(0002-0010) Syntax transfer	Podemos tener tres casos: VR Implícita – Little Endian → 1.2.840.10008.1.2 VR Explícita – Little Endian → 1.2.840.10008.1.2.1 VR Explícita – Big Endian → 1.2.840.10008.1.2.2
(0008-0021) Series Date	Fecha de la realización de la prueba.
(0008-0080) Institution Name	Nombre de la institución que realiza esta prueba.
(0008-0090) Referring Physicians Name	Nombre del profesional médico que ha realizado la prueba.
(0008-103E) Series Descripcion	Descripción de la serie. Una serie es un conjunto de imágenes que corresponden a la misma exploración.
(0010-0010) Patients Name	Especifica el nombre completo del paciente.
(0010-0020) Patient Id	Número de identificación del paciente.
(0010-0030) PatientsBirthDate	Año de nacimiento del paciente. El formato de las fechas es: aaaammdd
(0020-0011) Series Number	Número de serie. Varias imágenes que pertenezcan a la misma serie o grupo compartirán este campo.
(0020-0013) Instance Number	Número de la imagen dentro de la serie. Para mantener un orden de imágenes dentro del grupo.
(0028,0002) Samples Per Pixel	Número de muestras en la imagen.
(0028-0010) Rows	Número de filas de la imagen.
(0028-0011) Columns	Número de columnas de la imagen.
(0028-0100) BitsAllocated	Número de bits asignados para cada muestra o píxel. Este valor será igual para todos los píxeles.
(0028-0101) BitsStored	Número de bits almacenados para cada muestra o píxel. Este valor será igual para todos los píxeles.
(0028,0102)High Bit	Bit más significativo para para los datos de muestra o píxeles. Este valor será uno menos que el valor de BitsStored.
(0028-1050) WindowCenter	Valor para ajustar el brillo y contraste de la imagen.
(0028-1051) WindowWidth	Valor para ajustar el brillo y contraste de la imagen.

Tabla 4.5. Elementos de la cabecera del formato DICOM usados en la aplicación.

Para comprender mejor cómo se almacenan los datos dentro de un DICOM y qué significan algunos de los campos descritos en la tabla 4.4 se va a explicar mediante un ejemplo:

El paciente Luis Rodríguez acude a la consulta del doctor López y este le realiza dos tacs, uno de ellos de la cabeza el día 5 de mayo de 2012 y el otro del tronco el día 11 de mayo del mismo año. El primero de ellos tiene 25 imágenes y el segundo 10. Las imágenes pertenecientes al primer tac se guardan en un grupo diferente a las realizadas por el segundo. A continuación se

describen los valores de las cabeceras para los DICOM obtenidos después de la realización de dichas pruebas.

En primer lugar vamos a mostrar los datos comunes a todas las imágenes (tanto en el grupo 1 como en el 2):

- El elemento (0010-0010) perteneciente al nombre del paciente tendrá el valor Luis Rodríguez.
- El elemento (0008-0090) referente al nombre del médico será doctor López.
- El número de identificación del paciente (0010-0020) también será igual en ambos grupos al igual que la información referente al año de nacimiento (0010-0030) o la edad del mismo (0010-1010).

Para el primer grupo (tac de la cabeza) algunos de los elementos de la cabecera de los DICOM que los forman serán los siguientes:

- El número de serie o de grupo, correspondiente a la etiqueta (0020-0011), tendrá el mismo valor para las 25 imágenes que formen parte del mismo, por ejemplo 1.
- El elemento (0008-103E), correspondiente a la descripción de la serie, también tendrá el mismo valor, concretamente “Tac de la cabeza”, por ejemplo.
- La fecha en la que se ha realizado la prueba (0008-0021) también coincidirá para estas 25 imágenes, en este caso será 5 de mayo de 2012 que en la cabecera DICOM aparecerá como 20120505.
- El valor del elemento Instance Number (0020-0013) correspondiente al número de la imagen dentro de la serie será diferente para cada imagen. Así, cada imagen tendrá un valor dependiendo del orden en el que se haya obtenido cada una de ellas (por ejemplo del 0 al 24).

En el caso del segundo tac correspondiente al tronco, las cabeceras de las 10 imágenes que lo forman tendrán en común los siguientes elementos de datos:

- El número de serie o de grupo correspondiente a la etiqueta (0020-0011) tendrá el mismo valor para las 10 imágenes que formen parte del mismo, por ejemplo 2.
- También tendrá el mismo valor el elemento (0008-103E), correspondiente a la descripción de la serie, que tendrá un valor “Tac del tronco”, por ejemplo.
- La fecha en la que se ha realizado la prueba (0008-0021) también coincidirá para estas 10 imágenes y tendrá el valor 20120511 que hace referencia al día 11 de mayo de 2012.
- El valor del elemento Instance Number (0020-0013) correspondiente al número de la imagen dentro de la serie será diferente para cada imagen. Así, cada imagen tendrá un valor dependiendo del orden en el que se haya obtenido cada una de ellas (del 0 al 9).

Con respecto a los campos que hacen referencia a características de la imagen, los dos últimos elementos que constituyen la tabla 4.4 (*WindowCenter* y *WindowWidth*) son muy importantes para poder representarla correctamente. A continuación se explica el por qué de la importancia de estos elementos:

En el momento de representar una imagen, su brillo y contraste son dos factores muy importantes. Al aumentar el valor de escala de grises de cada uno de los píxeles en una unidad, entonces se está aumentando el brillo de la imagen. De manera similar con la disminución del brillo. El contraste es una medida de la diferencia entre los valores altos y bajos en una imagen. Si dos pixeles adyacentes tienen una gran diferencia en los valores de brillo de la escala de grises, el contraste entre ellos se dice que es alto, y a la inversa, si dos pixeles adyacentes tienen una pequeña diferencia en la escala de grises, se dice que tienen un bajo contraste entre sí. Otra forma de representar el brillo y el contraste es a través de los parámetros *WindowCenter* y *WindowWidth*. Dicho en términos simples, *WindowWidth* es la diferencia entre el valor del

pixel más brillante y el menos que se muestra. *WindowCenter* es el valor medio entre el valor del pixel más brillante y el menos. Así, podemos distinguir cuatro valores importantes:

- El valor mínimo entre todos los valores de escala de gris de la imagen.
- El valor máximo entre todos los valores de escala de gris de la imagen.
- *Window Minimum*: El valor umbral inferior que se muestra como cero intensidad (negro) en la pantalla.
- *Window Maximum*: El valor umbral más alto que se muestra como la más alta intensidad (blanco) en la pantalla.

Los dos primeros valores dependerán de la imagen y el número de bits para su codificación, mientras que los dos últimos dependen de la configuración del usuario. Todos los píxeles de la imagen en escala de grises con menor intensidad que el valor de *Window Minimum* se muestran como negro (intensidad cero) en la pantalla, mientras que todos los píxeles de la imagen en escala de grises con mayor intensidad que el valor de *Window Maximum* se muestran como blanco (intensidad máxima, por lo general 255) en la pantalla.

WindowCenter hace referencia al valor medio entre valor umbral más bajo y el más alto, es decir, es el valor central. Cuanto mayor sea este número, más oscura apare la imagen, y viceversa. *WindowWidth* representa la diferencia entre el valor umbral más alto y el más bajo. Cuanto mayor sea la diferencia, mayor será el contraste. Ajustando estos dos valores, se pueden resaltar las características que se deseen de la imagen.

La figura 4.2 muestra los valores explicados anteriormente para una imagen de 16 bits. En este caso si nos centramos en los pixeles con intensidades entre 30000 y 50000 el valor de WindowCenter sería el valor medio (en este caso 40000) y el valor de WindowWidth sería la diferencia, es decir, 20000.

En el caso de centrarnos en valores entre 300 y 2000, el valor de WindowCenter sería 850 y el de WindowWidth 1700. Los valores dentro de la ventana se escalan para ser visualizados: por ejemplo, los valores por debajo de 300 pasarían a representarse como 0 y los mayores que 2000 tomarían el valor de 255. De la misma manera, valores intermedios como el 600 pasaría a representarse con 90 y, el 1500 se representaría con el valor 225 en la pantalla.

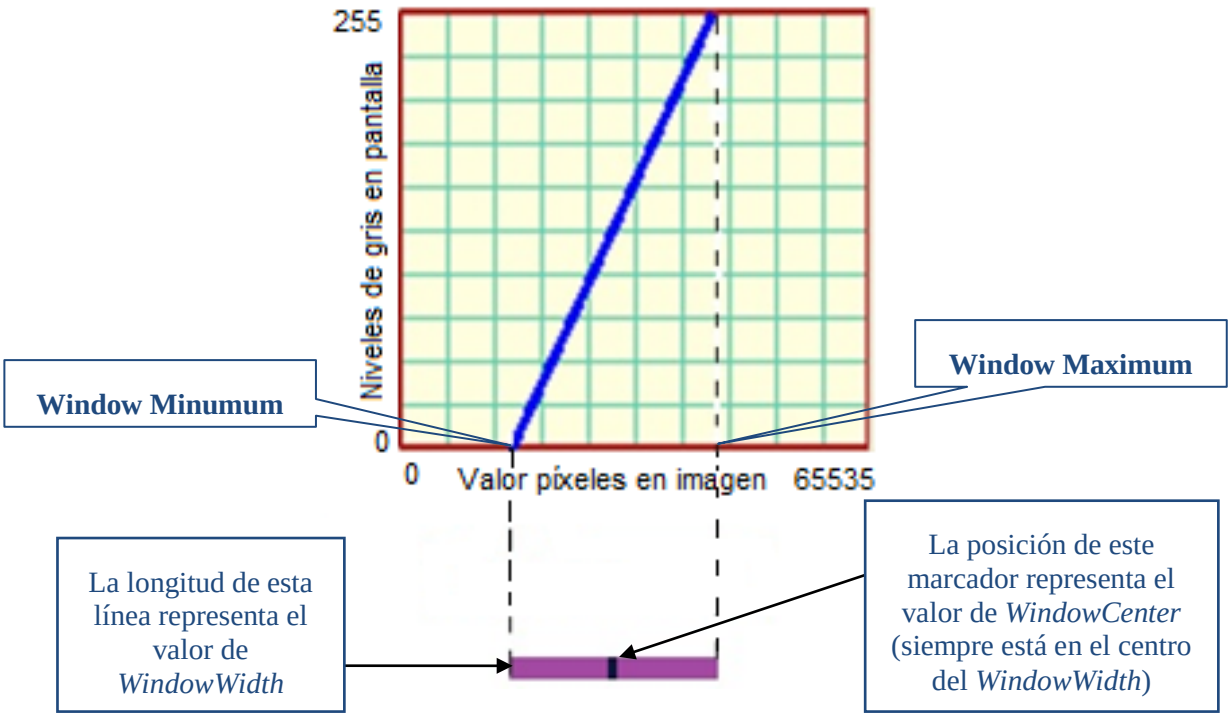


Figura 4.3. Imagen explicativa del significado de los valores *WindowCenter* y *WindowWidth*.

Una característica importante de estos valores de la cabecera (WindowCenter y WindowWidth) es que permiten obtener unas zonas u otras de la imagen ya que lo que hay fuera de la ventana se descarta poniéndolo a blanco o negro según corresponda. De esta manera nos quedamos sólo con la información de la imagen que se encuentra dentro de la ventana.

La importancia de esto radica en que moviendo la ventana, podemos centrar nuestra atención en las partes de la imagen que más nos interesen lo que facilitará el diagnóstico a los profesionales. Existen determinadas escalas que nos permiten saber el ancho y el centro de ventana que son necesarios para visualizar distintos tipos de tejidos. Así por ejemplo, una ventana adecuada para observar estructuras óseas en un tac de cabeza podría tener su centro en 400 con un ancho de 2000, mientras que si centramos la ventana en 50 y establecemos un ancho de 350 estaríamos capacitados para distinguir con mayor nitidez las zonas de tejido blando.

Capítulo 5

Diseño y desarrollo de la aplicación

5.1 CASOS DE USO

En este apartado se trata la especificación y la realización de los casos de uso de la aplicación realizada (AmeF). Un caso de uso es una descripción de los pasos o las actividades que deberán realizarse para llevar a cabo algún proceso. Los personajes o entidades que participarán en un caso de uso se denominan actores.

Para cada caso de uso se ha explicado en qué consiste, las precondiciones que se tienen que cumplir para que se pueda llevar a cabo y las postcondiciones si se produce fallo y si se produce éxito para cada caso. Además se ha añadido una secuencia que explica los pasos que tiene que ir realizando el actor (usuario de la aplicación) para la realización de cada caso de uso.

La figura 5.1 presenta un diagrama de los casos de uso, para poder verlo de manera más general antes de desarrollar cada uno de manera individual.

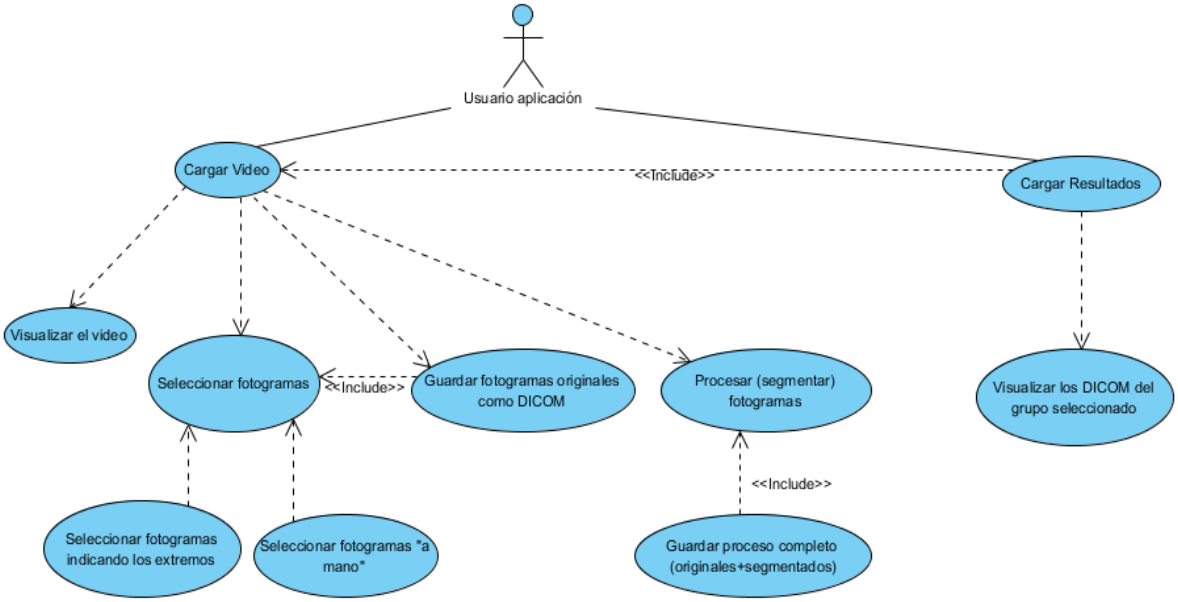


Figura 5.1. Diagrama de casos de uso de la aplicación

5.1.1 Cargar video

Es el caso de uso del que dependen los demás a excepción del caso de visualización de los resultados ya que, si el usuario no carga ningún video, no podrá realizar ningunas de las operaciones disponibles para ello.

En este caso, el usuario carga un video y de manera automática se va a dividir en los fotogramas que lo forman. De esta manera se podrá visualizar el video como tal o a partir de sus fotogramas.

Precondiciones: aplicación en funcionamiento.

Postcondiciones si éxito: se carga el video y se divide en fotogramas.

Postcondiciones si fallo: no se carga el video, se muestra un mensaje de error y se queda en la pantalla de inicio de la aplicación.

Actores: el usuario que va a utilizar la aplicación. Normalmente se tratará de un profesional médico que la utilizará para evaluar problemas en la voz del paciente que acude a su consulta.

Secuencia normal:

P1 → El usuario pulsa el botón de cargar video (ctrl+V) o selecciona *Cargar Video* del menú *Archivo.*

P2 → El usuario selecciona un .avi de los disponibles dentro de su equipo.

P3 → Se comprueba que se haya introducido un archivo con extensión correcta. Si error → S-1.

P4 → Se carga el video y se muestran sus fotogramas.

Secuencia alternativa:

S-1 → Si el sistema detecta que el archivo que se quiere abrir no tiene la extensión que admite la aplicación, se muestra un mensaje de error y no se carga el video. Se queda en la pantalla de inicio.

5.1.2 Visualizar el video

Una vez cargado el video en el sistema, el usuario podrá visualizar el mismo con los botones habilitados para ello. Este caso de uso es muy útil para el terapeuta, ya que a través del video puede identificar qué partes del mismo son más representativas para poder realizar la segmentación y poder obtener resultados.

En todo momento se puede saber en qué fotograma se encuentra el video para ver la relación con las imágenes que aparecen en la parte inferior.

Precondiciones: tiene que haber uno o más videos cargados en el sistema.

Postcondiciones: el usuario puede visualizar el video y moverse por él mediante los botones de reproducción.

Actores: el usuario que va a utilizar la aplicación.

5.1.3 Seleccionar fotogramas

Es un caso de uso del que heredan las dos formas posibles de seleccionar fotogramas que permite la aplicación: seleccionar los fotogramas indicando los extremos del video o haciendo clic sobre los mismos.

- **Seleccionar fotogramas indicando los extremos**

Una vez cargado el video, el usuario puede seleccionar los fotogramas del mismo con el que va a trabajar. En este caso concreto, existe una parte de la pantalla que está destinada a ello: *selección.*

De esta manera, por medio de los botones *extremo A y extremo B*, mientras el video se está reproduciendo, se puede indicar qué partes del video (fotogramas) se quieren seleccionar. Una vez concluida la selección, los fotogramas elegidos aparecerán seleccionados en la parte inferior de la pantalla.

Precondiciones: video cargado y reproduciéndose.

Postcondiciones: se seleccionan los fotogramas que se encuentran dentro del rango del video elegido.

Actores: el usuario que va a utilizar la aplicación.

Secuencia normal:

P1 → El usuario reproduce el video que previamente ha sido cargado.

P2 → Pulsa el botón *Extremo A* en el momento en el que quiera comenzar su selección.

P3 → Pulsa el botón *Extremo B* en el momento en el que quiera finalizar su selección.

P4 → Los fotogramas contenidos entre ambos extremos del video aparecen seleccionados.

P5 → Si en cualquier momento entre P2 y P3 el usuario pulsa el botón *Pausa* del reproductor, podrá mover aumentar el número de fotograma de los extremos por medio de las flechas disponibles en la selección.

- **Seleccionar fotogramas "a mano**"

Además de la posibilidad de seleccionar fotogramas indicando los extremos, el usuario podrá seleccionarlos o deseleccionarlos pulsando sobre cualquier fotograma en cualquier momento. Esta opción es útil si los fotogramas que se quieren seleccionar no son contiguos.

Precondiciones: video cargado.

Postcondiciones: los fotogramas seleccionados aparecen como tal en la secuencia inferior.

Actores: el usuario que va a utilizar la aplicación.

5.1.4 Guardar fotogramas originales como DICOM

Una vez que se han seleccionado los fotogramas, el usuario tiene disponible la opción de guardarlos como imágenes DICOM. De esta manera, se guardan las imágenes seleccionadas con la información que el usuario haya escrito del médico, paciente, nombre de grupo y la fecha actual.

Todas estas imágenes se guardarán dentro de un solo grupo, de manera que al visualizarlas aparecerán todas relacionadas y tendrán la misma información en su cabecera.

Precondiciones: debe haber algún fotograma seleccionado.

Postcondiciones si éxito: se guardan todos los fotogramas seleccionados con formato DICOM con la información introducida en los campos inferiores.

Postcondiciones si fallo: no se guardan las imágenes y se muestra un mensaje de error para indicarlo al usuario.

Actores: el usuario que va a utilizar la aplicación.

Secuencia normal:

P1 → El usuario presiona el botón de *Guardar Originales.* Si no hay ningún fotograma seleccionado S-1.

P2 → Aparece una pantalla para que se seleccione el directorio en el que se van a guardar las imágenes con formato DICOM.

P3 → Se guardan las imágenes en el directorio seleccionado. Si error →S-2.

P4 → Cuando termina el proceso se le indica al usuario con un mensaje.

Secuencia alternativa:

S-1 → Si al presionar el botón de *Guardar Originales* no hay ningún fotograma seleccionado, se indica con un mensaje de error y no se guarda ninguna imagen.

S-2 → Si una vez seleccionados los fotogramas que se quieren guardar y el directorio en el que se almacenarán se produce algún error en el proceso también se indicará y se le mostrará al usuario las imágenes que se han podido guardar y cuáles no.

5.1.5 Procesar (segmentar) fotogramas seleccionados

Después de seleccionar los fotogramas además de la opción de guardarlos como DICOM se puede aplicar una técnica de detección de la glotis por medio de segmentación de imágenes laríngeas.

El proceso de segmentación aplicado es el que se explicó en el anterior capítulo. Dependiendo del fotograma, la segmentación puede obtener resultados o no. El fotograma resultado aparecerá debajo del original y en éste se mostrará, o bien la glotis en caso de tener éxito el proceso, o bien se indicará que la imagen actual no ha obtenido resultados.

Precondiciones: debe haber algún fotograma seleccionado.

Postcondiciones si éxito: se aplica un proceso de segmentación sobre los mismos para detectar la glotis.

Postcondiciones si fallo: no se produce ningún cambio si no se consigue aplicar el proceso.

Actores: el usuario que va a utilizar la aplicación.

Secuencia normal:

P1 → El usuario presiona el botón *Procesar.* Si no hay ningún fotograma seleccionado →S-1.

P2 → Se aplica el proceso de segmentación desarrollado para la detección de la glotis en las imágenes seleccionadas. Si error → S-2

P3 → Los resultados aparecen debajo de cada fotograma. Si la segmentación ha tenido éxito aparece la glotis detectada y en caso contrario se indica que no se han obtenido resultados.

Secuencia alternativa:

S-1 → Si al presionar el botón de *Procesar* no hay ningún fotograma seleccionado, se indica con un mensaje de error y no aplica el proceso a ninguna imagen.

S-2 → Si una vez seleccionados los fotogramas que se quieren procesar se produce algún error durante la segmentación también se indicará mediante un mensaje.

5.1.6 Guardar proceso completo (imágenes originales y segmentadas)

Una vez aplicada la segmentación para obtener la glotis en los fotogramas que se han seleccionado se puede guardar todo el proceso completo. En este caso se guardarán tanto las imágenes originales como las que se han creado resultado de la segmentación (que mostrarán la glotis de las originales).

Al igual que en el caso de uso de guardar las imágenes originales, el usuario puede establecer cierta información que aparecerá en las cabeceras de las imágenes guardadas como el nombre del médico, del paciente, el nombre del grupo en que se guardarán y la fecha. Las imágenes resultado de la segmentación se guardarán en un grupo diferente al grupo en el que se han guardado las originales.

Precondiciones: las imágenes seleccionadas deben haber sido segmentadas y, por tanto además de los fotogramas cargados al inicio también aparecerán las imágenes obtenidas al aplicar el proceso de detección de la glotis en la parte inferior.

Postcondiciones si éxito: se guardan todos los fotogramas seleccionados (los originales y los procesados) en formato DICOM con la información introducida en los campos inferiores.

Postcondiciones si fallo: no se guardan las imágenes porque no existen resultados o porque ha ocurrido algún error durante el proceso.

Actores: el usuario que va a utilizar la aplicación.

Secuencia normal:

P1 → El usuario presiona el botón de *Guardar Proceso Completo.* Si no se ha realizado anteriormente la segmentación de las imágenes seleccionadas S-1.

P2 → Aparece una pantalla para que se seleccione el directorio en el que se van a guardar las imágenes con formato DICOM.

P3 → Se guardan las imágenes en el directorio seleccionado. Si error S-2.

P4 → Cuando termina el proceso se le indica al usuario con un mensaje. Si algunas de las imágenes no obtuvieron resultados durante la segmentación no se guardarán y se informará al usuario.

Secuencia alternativa:

S-1 → Si al presionar el botón de *Guardar Proceso Completo* no hay ningún fotograma seleccionado o los que se han seleccionado no han sido segmentados, se indica con un mensaje de error y no se guarda ninguna imagen.

S-2 → Si una vez seleccionados los fotogramas que se quieren guardar y el directorio en el que se almacenarán se produce algún error en el proceso también se indicará y se le mostrará al usuario las imágenes que se han podido guardar y cuáles no.

5.1.7 Cargar resultados

Este caso de uso depende de que los anteriores se hayan llevado a cabo ya que, si no se han guardado los fotogramas de un video (ya sean originales o segmentados) como DICOM, no tendremos resultados que visualizar.

En este caso, se cargan las imágenes DICOM del directorio que se indique y la información que en los casos de uso anteriores se ha introducido. Las imágenes se cargan por grupos.

Precondiciones: aplicación en funcionamiento e imágenes DICOM en el directorio elegido.

Postcondiciones: se cargan las imágenes DICOM existentes en el directorio por grupos.

Actores: el usuario que va a utilizar la aplicación. Normalmente se tratará de un profesional médico para evaluar problemas en la voz del paciente que acude a su consulta.

Secuencia normal:

P1 → El usuario pulsa el botón de cargar resultados (ctrl+R) o selecciona *Cargar Resultados* del menú *Archivo.*

P2 → El usuario selecciona un directorio en el que se encuentren las imágenes DICOM que quiere visualizar.

P3 → Se cargan las imágenes de ese directorio que tengan extensión .dcm (correspondiente a las imágenes médicas DICOM).

5.1.8 Visualizar los DICOM del grupo seleccionado

Una vez que se han cargado los grupos con las imágenes DICOM del directorio elegido podemos visualizarlos, ya sean las imágenes originales únicamente o con sus respectivas imágenes segmentadas.

Como se ha explicado en casos de uso anteriores, cuando se guarda el proceso completo las imágenes originales y las segmentadas se guardan con números de grupo diferente. Sin embargo, en nuestro visor se han unido ambos grupos para así poder ver qué imagen segmentada corresponde con cada imagen original.

Precondiciones: se han tenido que cargar anteriormente los resultados contenidos en un directorio (imágenes DICOM).

Postcondiciones: se visualizan las imágenes DICOM guardadas y la información de la cabecera que se modificó.

Actores: el usuario que va a utilizar la aplicación.

Secuencia normal:

P1 → El usuario selecciona uno de los grupos que aparecen en la parte izquierda de la pantalla.

P2 → En la parte inferior aparecen todas las imágenes que corresponden a ese grupo y nos indica si existen imágenes segmentadas y en caso afirmativo se muestran.

5.2 DIAGRAMAS UML

En esta sección se realiza una descripción del diseño de la arquitectura de la aplicación AmeF.

El objetivo es explicar de qué clases está compuesto el proyecto. Usando diferentes visiones arquitecturales como UML se pretende presentar el proyecto a un nivel más detallado, también haciendo referencia a otros sistemas de representación como son los casos de uso.

El proyecto se encuentra dividido en cuatro paquetes. En la figura 5.2 podemos ver un esquema general de los mismos y las relaciones existentes entre ellos.

Existe un paquete *Auxiliares* cuyas clases son utilizadas por los otros dos paquetes que se encuentran en un nivel superior. A su vez, los paquetes *Dicom* y *Segmentación* van a ser utilizados por las clases que componen el paquete *Principal*, que se encuentra en el nivel superior y que contiene las clases principales del proyecto así como los formularios (pantallas) que componen el mismo.

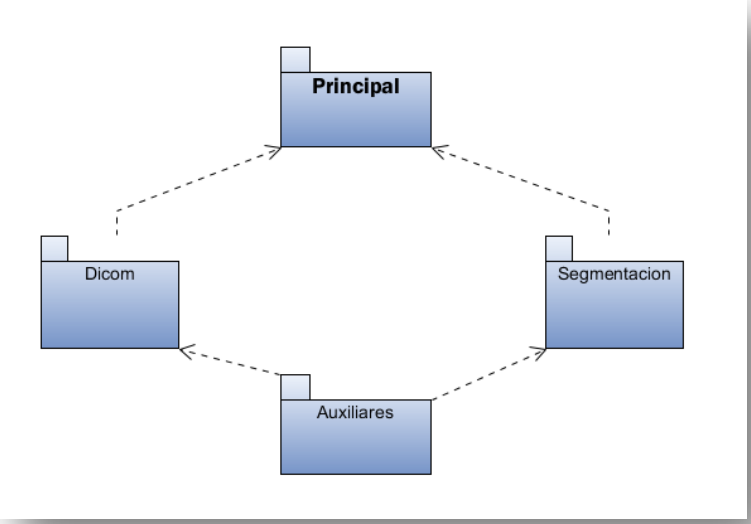


Figura 5.2. Esquema de la organización de los diferentes paquetes dentro del proyecto. Se presentan las relaciones que hay entre ellos.

A continuación explicamos cada paquete por separado, mostrando un diagrama UML con las clases que los componen y las relaciones entre las clases del mismo paquete. Con la intención de dar una visión esquemática más clara de la estructura se ha decidido prescindir de los métodos y atributos de las clases.

- **Auxiliares**

En este paquete están las clases auxiliares que se utilizan en los paquetes superiores (figura 5.3). La mayoría de las clases son estructuras como es el ejemplo de *Lista*, *Reserva1D, Reserva2D* y *Reserva3D* que son arrays de una, dos o tres dimensiones respectivamente.

Se dispone también de una clase TImagen que permite cargar una imagen desde un archivo, obtener información de la misma, insertarla en un BitMap para poder observarla en la interfaz y guardarla en un fichero.

Además el paquete contiene dos clases: *Funciones* y *FuncionesOrdenación;* con funciones básicas (como calcular máximos, mínimos o inicializar estructuras) que se van a utilizar en la mayoría de las clases que componen el resto del proyecto.

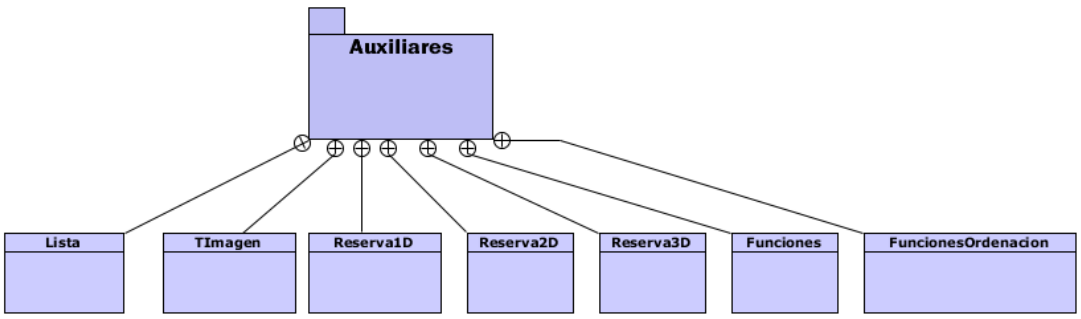


Figura 5.3. Diagrama UML de las clases del paquete Auxiliares.

- **Dicom**

El paquete Dicom contiene las clases encargadas de leer una imagen o un directorio con imágenes Dicom. También permite, dada una imagen y una cabecera de elementos de datos, escribir estas imágenes con formato .dcm.

Como se puede observar en la figura 5.3 está compuesto por dos clases. La primera de ellas *CargarDicom* contiene los métodos que se encargan de leer la cabecera de un Dicom, con funciones para obtener el número de grupo, número de elemento, VR y longitud del valor y en función de estos valores obtiene el valor para cada elemento de datos.

Todas estas funciones son usadas dentro de la clase *CodeDICOM* para poder leer la cabecera de estas imágenes y permitir así obtener toda la información contenida en ellas. Además permite escanear directorios que contienen más de una imagen y dada una cabecera guardar varias imágenes como Dicom dentro de un mismo grupo.

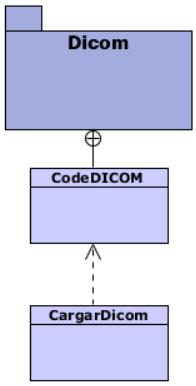


Figura 5.4. Diagrama UML de las clases del paquete Dicom.

- **Segmentación.**

En este caso disponemos de una clase para cada una de las operaciones que van a componer el proceso de segmentación que se va a aplicar a las imágenes. La unión de estas operaciones en un orden determinado va a permitir detectar la glotis de los fotogramas seleccionados.

Parte de los algoritmos que se implementan en estas clases se encontraban ya realizados antes de emprender el presente proyecto y durante la realización del mismo se ha procedido a estructurarlos y completarlos.

La clase *TDeteccionGlotis* permite realizar una detección de la glotis en una imagen segmentada anteriormente. El método de detección usa los momentos para discriminar la glotis en una imagen dividida en objetos según un predictor lineal entrenado con 100 imágenes de 15 vídeos de las cuerdas vocales, en las que previamente se ha segmentado la glotis mediante el proceso detallado en los apartados anteriores. La clase *TDescripcionObjetos* es utilizada por la anterior y su función es realizar operaciones de cálculo de parámetros sobre los objetos de una imagen.

En *TWatersheds* se crea un objeto de este tipo que calculará las cuencas Watersheds de una imagen. Este proceso entrega las cuencas con un número distinto y consecutivo en los píxeles que forman parte de cada cuenca diferente. El procedimiento en sí no calcula líneas watersheds, aunque existen métodos dentro de esta clase que permiten hacer el cálculo de las mismas. En esta clase están contenidas todas las funciones que se encargan de aplicar el proceso de Merging que se lleva a cabo después de la división inicial obtenida con Watershed.

En la clase *TConversionColor* se realiza una conversión de color con el método YIQ. El formato de color YIQ representa una división entre la luminosidad (cantidad de luz percibida) y

la información sobre el color. El ojo humano es mucho más sensible a la luminosidad que al color, cosa que se aprovecha para realizar la conversión.

La clase *TOperacionesDePunto* realizan operaciones de punto sobre imágenes, es decir, aplica diferentes operaciones a cada píxel que compone la imagen. Los parámetros que afectan a cada imagen vienen establecidos en las propiedades de la clase. Un ejemplo de este tipo de operaciones es la umbralización de una imagen, donde si un pixel es mayor que un umbral, se cambia su valor a 255 (blanco) y si es menor se lleva a 0 (negro).

Vectorial es una clase que implementa una lista enlazada de objetos genéricos con las operaciones típicas de insertar, eliminar u obtener el número de nodos que la componen. La clase *Proceso* guarda todos los procesos. Cada proceso es una lista de operaciones, cada una de ellas con sus parámetros correspondientes.

Operación es una clase genérica que representa una operación gráfica para tratamiento de imágenes. Una operación tiene un nombre y un tipo que puede ser conversión de color, gradiente, watershed, merging, detección de glotis o una operación no definida.

Las demás clases como *TGradiente* o *TFiltrado*, como el propio nombre de la clase indica, realizan gradientes o filtrados según una plantilla que reciben como entrada.

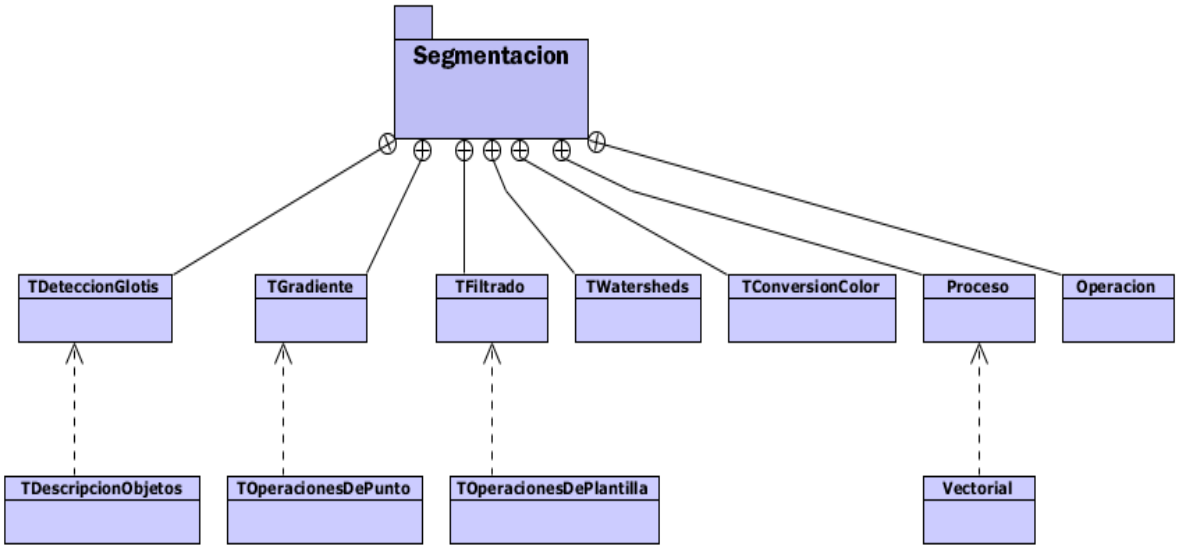


Figura 5.5. Diagrama UML de las clases del paquete Segmentación.

- **Principal**

En este paquete se encuentran los formularios de las pantallas que componen la aplicación así como las clases más importantes que utilizan métodos de las anteriores. El diagrama del mismo se puede observar en la figura 5.6.

Además de las pantallas y las clases principales también forman parte de esta pantalla la clase *TReproductorVideo* y *TProcesadorVideo* que permiten (mediante DirectShow) reproducir el video que se ha cargado durante la ejecución de la aplicación y descomponer ese video en fotogramas.

La clase *Esperar* es una pantalla que se muestra cuando se abre un video y, como consecuencia del tamaño del mismo, tarda más de lo esperado.

En el segundo nivel del diagrama tenemos la pantalla que nos permite elegir un directorio de nuestro sistema (*AbrirDirectorio)*, la que nos muestra la información sobre el proyecto, autor, fecha y lugar de realización (*AcercaDe)*, la pantalla que nos permite visualizar el video y sus fotogramas, procesarlo y realizar diferentes operaciones sobre él (*Edicion*) y por último la clase *Resultados* que se encarga de todo lo referente a la visualización de las imágenes Dicom que se han guardado anteriormente.

Todas estas clases dependen de la clase principal (*PantallaPadre)* en la que, dependiendo de la opción que se elija, se cargara sobre ella las pantallas del nivel inferior. Esta clase se encarga de llamar a todas las demás y del correcto funcionamiento de toda la aplicación en general. Son pantallas MDI que consisten en un conjunto de interfaces o ventanas que residen debajo de una ventana madre.

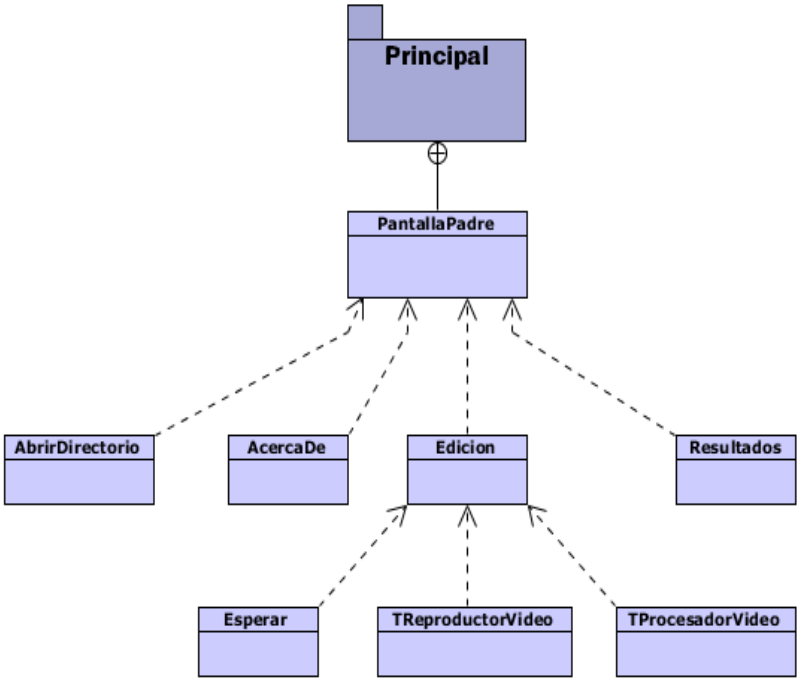


Figura 5.6. Diagrama UML de las clases del paquete Principal.

5.3 VERIFICACIÓN DE CLASES

El objetivo de este apartado es verificar que todos los casos de uso que se han presentado en apartados anteriores se encuentran implementados por una o varias clases del diagrama UML.

Para ello se ha realizado una tabla en la que las filas contienen las diferentes clases del proyecto y en las columnas aparecen los distintos casos de uso explicados anteriormente. La tabla está dividida en tres partes correspondientes a los tres paquetes que forman el proyecto. El paquete *auxiliares* no se ha incluido porque, como se ha visto en el apartado anterior por medio de un diagrama, las clases que lo componen se utilizan dentro de los demás paquetes por lo que, a su vez, también estarían incluidos dentro de la tabla.

Para cada una de las clases representadas en la tabla 5.1, se ha señalado en qué caso de uso se utilizan con el fin de asegurar que todos ellos quedan cubiertos, así como que todas las clases tienen utilidad para algún o algunos casos de uso.

	Cargar Video	Ver video	Seleccionar fotogramas	Guardar originales en DICOM	Procesar fotogramas	Guardar proceso completo	Cargar resultados	Ver resultados
PRINCIPAL								
Pantalla Padre	X	X	X	X	X	X	X	X
Abrir Directorio	X						X	
Edicion	X	X	X	X	X	X		
Esperar	X						X	
TReproductorVideo		X						
TProcesadorVideo		X	X					
Resultados							X	X
DICOM								
CodeDICOM				X		X	X	X
CargarDicom				X		X	X	X
SEGMENTACION								
TDeteccionGlotis					X			
TGradiente					X			
TFiltrado					X			
TWatersheds					X			
TConversionColor					X			
Proceso					X			
Operacion					X			

Tabla 5.1. Tabla de verificación de clases para cada Caso de Uso.

5.4 HERRAMIENTAS DE DESARROLLO

En esta sección se detallan las herramientas usadas durante el desarrollo del proyecto, así como algunas de las ventajas e inconvenientes que hemos encontrado en ellas.

Herramienta de desarrollo de la aplicación

El desarrollo de la aplicación se ha realizado en el lenguaje de programación C++ en el entorno de desarrollo C++ Builder XE.

Las principales ventajas encontradas es que Builder XE es un entorno que incluye herramientas que permiten un desarrollo visual de arrastrar y soltar componentes sobre la aplicación, lo que

hace que la parte visual sea más fácil de implementar. La gran difusión que tiene el lenguaje de programación utilizado hace que sea posible encontrar soluciones a los problemas que aparecen acerca de la sintaxis del propio lenguaje, así como los cambios en el mismo debidos a la nueva versión de Builder.

Los mayores problemas encontrados han sido sobre todo debidos a que se trata de un entorno que se encuentra en constante evolución y, aunque ya se había usado anteriormente, la versión empleada es muy actual e introduce numerosos cambios respecto a versiones anteriores. Estos cambios no están documentados actualmente lo que provoca más dificultades al intentar utilizar nuevos componentes o nuevas utilidades del mismo.

Herramientas para pruebas

Además de la herramienta de desarrollo principal, se han utilizado otros programas para la realización de pruebas. Entre ellos cabe destacar el visor de imágenes DICOM "MicroDicom".

Este programa ha permitido conocer mejor la estructura de este tipo de imágenes médicas para poder desarrollar funciones que permitan usarlas en la aplicación. Las principal ventaja que se puede destacar es que dada una imagen o conjunto de imágenes DICOM permite visualizarlas y nos muestra todos los campos que forman la cabecera, además de permitir la modificación de algunos de los parámetros de la imagen.

Herramientas de control de versiones

En cuanto al control de versiones se ha utilizado Subversion, más concretamente se ha usado una de las interfaces más prácticas y más sencillas para interactuar con el repositorio: *Tortoise SVN*.

Aunque en este caso no ha habido varias personas accediendo al proyecto, es también muy útil para tener organizadas todas las versiones que se han realizado hasta el momento, recuperar versiones antiguas de todos los archivos y examinar la historia de cómo y cuándo cambiaron sus datos. El manejo de esta herramienta es muy sencillo debido a que se instala como una extensión del escritorio, lo que significa que se puede seguir trabajando con las herramientas conocidas y que no hay que cambiar a una aplicación diferente cada vez que se necesiten las funciones del control de versiones.

Otras herramientas

Además de las herramientas citadas anteriormente cabe citar la utilización de DirectX, que es una colección de APIs creadas en un principio para facilitar tareas relacionadas con la programación y ejecución de juegos bajo Windows.

En nuestro caso concreto, se ha utilizado una API de las muchas que componen DirectX: DirectShow que se utiliza para reproducir audio y vídeo con transparencia de red. De hecho, la reproducción de video en nuestra aplicación se ha realizado únicamente con componentes de esta API.

Para facilitar el manejo de DirectShow, se ha usado una herramienta visual que permite crear y probar gráficamente los filtros de DirectShow: Graph Edit.

5.5 PROBLEMAS ENCONTRADOS Y SOLUCIONES

Durante el desarrollo del presente trabajo han surgido muchos problemas, tanto en la investigación inicial como durante el proceso de programación. En esta sección se explican los más significativos y las soluciones aplicadas en cada caso.

Inicialmente el trabajo se centro en la investigación del formato DICOM. Los problemas encontrados se debieron a que se trataba de un tema totalmente desconocido y del cual existía muy poca información. Es por esto por lo que la mayoría de la información descrita en el

capítulo dedicado a ello se obtuvo mediante la investigación de ejemplos de imágenes por medio de visores comerciales DICOM que muestran los contenidos de las cabeceras.

Una de las dificultades encontradas es que el formato de las cabeceras no es igual para todos los elementos de datos por lo que hubo que distinguir para cada uno de ellos (dependiendo de la VR), la manera de obtener el valor. Para cada elemento de datos se tuvo que guardar el grupo, elemento y demás información con el fin de realizar después distintas funciones y obtener el valor concreto de cada uno dependiendo de esa información.

Una vez que ya se conocía perfectamente el formato de las cabeceras, el mayor problema encontrado fue dibujar la imagen contenida en el fichero DICOM. Se probaron varias técnicas de procesado de imagen pero con ninguna de ellas se obtuvo el resultado esperado. La imagen obtenida no aparecía en escala de grises como se observaba en los visores. Esto se debía a que podía estar configurada según varios formatos: RAW, JPEG-LOSSLESS, etc. y se tuvieron que estudiar para poder realizar correctamente la codificación y decodificación de la misma.

Además, como ya se ha explicado en el capítulo 4, los elementos de la cabecera "Window Center" y "Window Width" desempeñan un papel muy importante a la hora de dibujar la imagen y poder guardarla con un formato conocido. Por tanto, hubo que realizar pruebas para obtener el valor deseado de estos elementos que hiciera que la imagen se observara correctamente.

Con respecto a la parte de segmentación de imágenes para la detección de la glotis los mayores problemas se derivaron del estudio de las funciones para su estructurado e integración en la aplicación mediante clases. Había que buscar una manera de relacionar todas las operaciones existentes para poder aplicarlas en conjunto, por lo que se creó una clase proceso que tuviera una lista de operaciones, en un orden indicado y cada una con unos parámetros determinados. Para que esto sea posible, también se creó una clase de operación genérica que permitiese acoger los distintos tipos de operaciones disponibles para después poder incluirlos en el proceso.

La opción de modificar cierta información de la cabecera de los DICOM antes de guardarlos también fue difícil debido a que inicialmente las funciones que se realizaron en el paquete *Dicom* sólo guardaban imágenes nuevas a partir de un DICOM ya existente, tomando sus cabeceras para guardar todas las imágenes nuevas. Tener que modificar la información de las cabeceras llevó mucho tiempo debido a que había que modificar toda la cabecera de la imagen si la longitud del nuevo valor que se quería introducir era mayor o menor que el existente. Todos los elementos que se encontraban a continuación del que se quería modificar había que moverlos, reservar o liberar espacio y colocarlos de nuevo en su lugar. Además, como los elementos de datos que componen los DICOM también guardan información de la longitud del mismo, esta información también tuvo que ser modificada.

En cuanto al número de grupo en el que se guardan los fotogramas como DICOM se tomó la decisión de guardar las imágenes originales dentro de un grupo y las segmentadas en otro. Sin embargo, una vez creada la parte de resultados se querían mostrar ambos grupos como uno sólo para que cada imagen tuviera al lado su segmentada correspondiente. Con este fin se decidió que el número de grupo con el que se guardarían las imágenes segmentadas sería el número de grupo asignado a las imágenes originales más 1000. En el momento de representarlas se va comprobando y, para cada grupo, si existe otro que cumpla esta condición se toma como un sólo grupo para que la visualización sea más clara.

Otro problema encontrado se debió al uso de librerías de DirectShow para la realización de clases que permitieran la reproducción y procesado de videos. Hubo que entender mediante manuales y libros el funcionamiento de estas librerías e investigar cómo se podían reproducir videos por medio de la unión de los filtros que lo forman. Uno de los problemas encontrados al final de la implementación se debió a que una de las clases que se habían hecho inicialmente no funcionaba como se deseaba. En concreto la clase *TReproductorVideo* que se encarga de reproducir un fichero de video no lo hacía correctamente: la calidad de la reproducción no era

buena y se veían líneas. El problema se producía porque uno de los filtros que formaban el grafo que se construyo para reproducir no funcionaba bien con las nuevas versiones de Windows por lo que hubo que reemplazarlo por otro más actual con el que se solucionó el problema. El programa utilizado *GraphEdit* ayudó en gran medida a encontrar la solución al problema planteado, ya que permite la unión de filtros gráficamente y la prueba del resultado del gráfico obtenido. Una vez que se comprueba que el grafo responde como se esperaba ya se puede pasar al código del programa con la seguridad de que funcionará correctamente.

La mayor dificultad a la hora de trabajar con los filtros dentro del DirectShow es la existencia de múltiples versiones para una misma operación y la escasa documentación existente sobre los requisitos y la funcionalidad de cada uno, por lo que, es necesario realizar pruebas para poder saber si cumple con los requisitos que se necesitan.

5.6 MANUAL DE USUARIO

La aplicación *AmeF* se puede definir como un sistema de síntesis de información para ayuda al diagnóstico médico de patologías vocales.

Este manual de introducción a *AmeF* ha sido elaborado con la intención de ofrecer la información necesaria para el uso de este sistema. Se explican los pasos que hay que seguir para el correcto funcionamiento de la aplicación, así como ejemplos de capturas de pantallas para su mejor comprensión.

En la figura 5.7 se muestra la pantalla que nos aparecerá al abrir la aplicación. En este momento tenemos dos opciones: abrir un video nuevo para obtener sus fotogramas, segmentarlos y guardarlos; o abrir un directorio de resultados para poder observar las imágenes DICOM guardadas anteriormente.

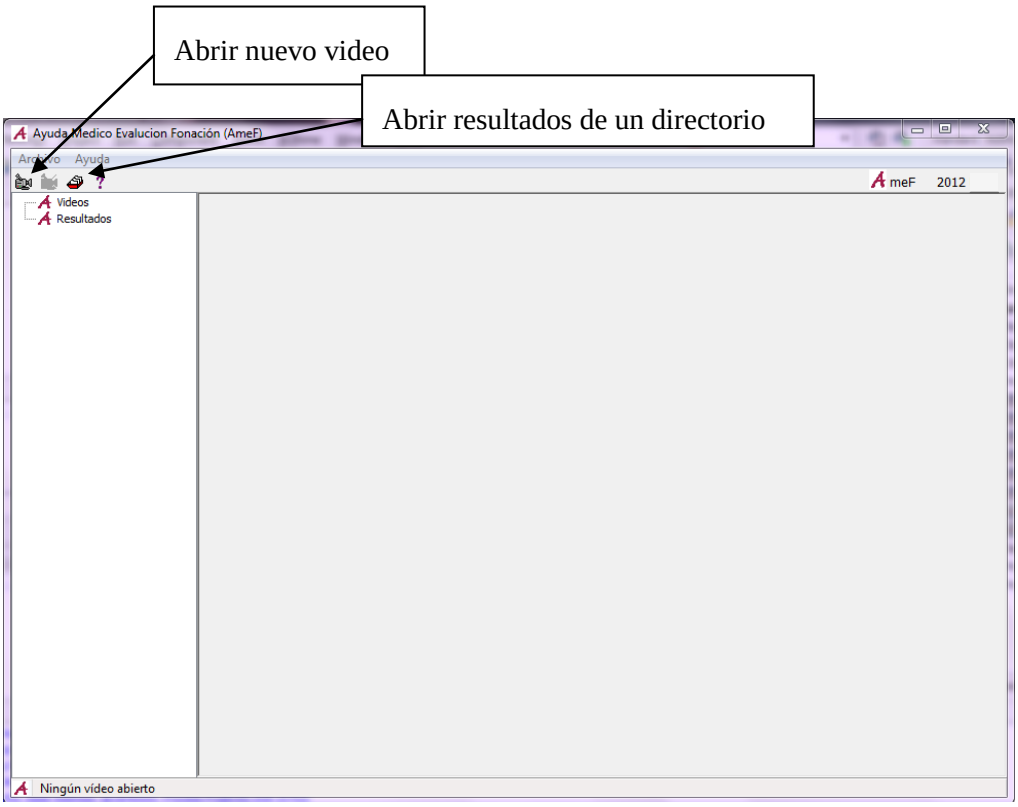


Figura 5.7. Pantalla que se abre cuando iniciamos la aplicación.

Cuando el usuario quiere abrir un nuevo video, pulsa en el icono que aparece a la izquierda en la pantalla principal y le aparece una nueva ventana en la que elegirá el video que desea abrir. Una vez elegido, pulsará *Abrir* y se cargará el video seleccionado. En la figura 5.8 vemos la pantalla que nos aparece al solicitar la apertura de un video nuevo.

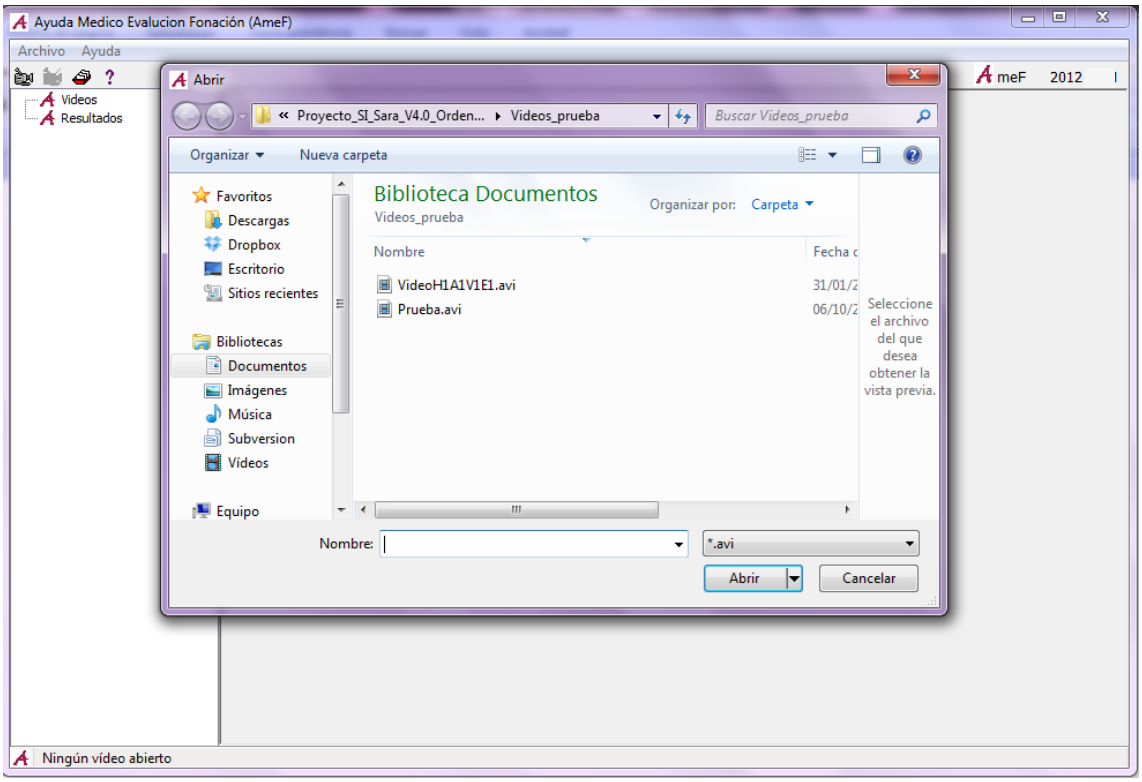


Figura 5.8. Pantalla obtenida una vez seleccionada la opción de *Abrir video nuevo*

Una vez seleccionado el video, su nombre aparece en la parte izquierda de la pantalla con el resto de videos que se encuentran cargados hasta ese momento (Figura 5.9). El video elegido se carga en la pantalla central en la que podremos visualizarlo por medio de los botones que aparecen en la parte inferior de la zona de visualización. Además podremos ver en qué fotograma se encuentra en cada momento el video por medio del número que aparece en la parte derecha de la barra de reproducción (en color azul).

Se puede observar que en este momento ya se ha dividido el video en fotogramas. Además de los fotogramas de la parte superior, aparecen imágenes en la parte inferior que inicialmente se muestran en color negro: ahí será donde se situarán los resultados de la segmentación de la glotis.

El escritorio se encuentra divido en 3 zonas en función de las operaciones que permiten realizar los distintos elementos que lo integran:

- Árbol de vídeos y resultados: permite abrir y cerrar vídeos, que serán introducidos o eliminados, respectivamente, en o de una lista construida al efecto. A través del árbol es posible navegar entre las distintas exploraciones que se quieran analizar. Además de los vídeos también permite abrir directorios que contengan imágenes DICOM para observar los grupos existentes. En la parte de resultados del árbol aparecerán todos los grupos de ese directorio.
- Visualización: se localizan todos los elementos que gestionan la monitorización del vídeo. Esta puede realizarse de forma convencional (zona superior) o mediante una barra de

desplazamiento que permite navegar por todos los fotogramas de la exploración (zona inferior).

- Procesado: por un lado, tenemos la opción de procesado de los fotogramas para detección de la glotis y, por otro, nos permite guardar las imágenes originales o todo el proceso (originales y segmentadas) en imágenes DICOM con la información que aparece en la parte inferior sobre el paciente.

En la figura 5.9 podemos observar la pantalla que nos aparecerá cuando carguemos un nuevo video en la aplicación.

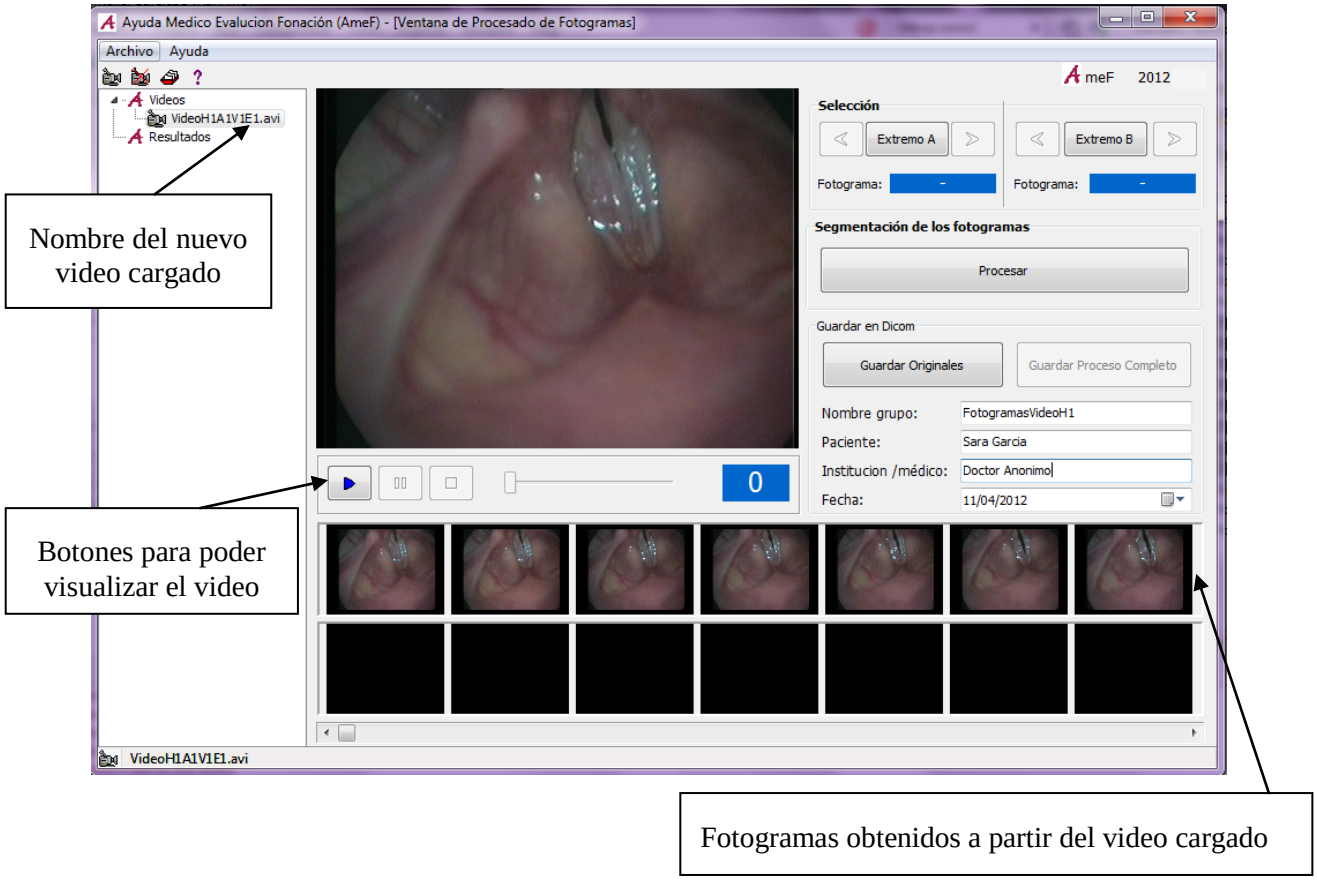


Figura 5.9. Captura de pantalla de video cargado y fotogramas del mismo obtenidos.

Una vez cargado el video podemos seleccionar los fotogramas para guardarlos o procesarlos. Por medio de los botones *ExtremoA* y *ExtremoB* podemos seleccionarlos mientras el video se está reproduciendo. Sólo tenemos que pulsar el botón *ExtremoA* cuando queremos que comience la selección y *ExtremoB* cuando queremos que termine. Una vez realizada esta selección nos aparecerán marcados con color rojo los fotogramas que se encuentren dentro esos dos extremos (Figura 5.10). Además, con los botones que aparecen a los lados podemos aumentar o disminuir el valor de ambos límites con un ajuste más fino hasta fijarlo en el deseado.

Además de este tipo de selección, siempre podremos seleccionar y deseleccionar fotogramas pulsando sobre la imagen correspondiente. Para saber cuál es el número de fotograma correspondiente a cada imagen sólo tenemos que colocar el ratón encima y nos aparecerá dicha información.

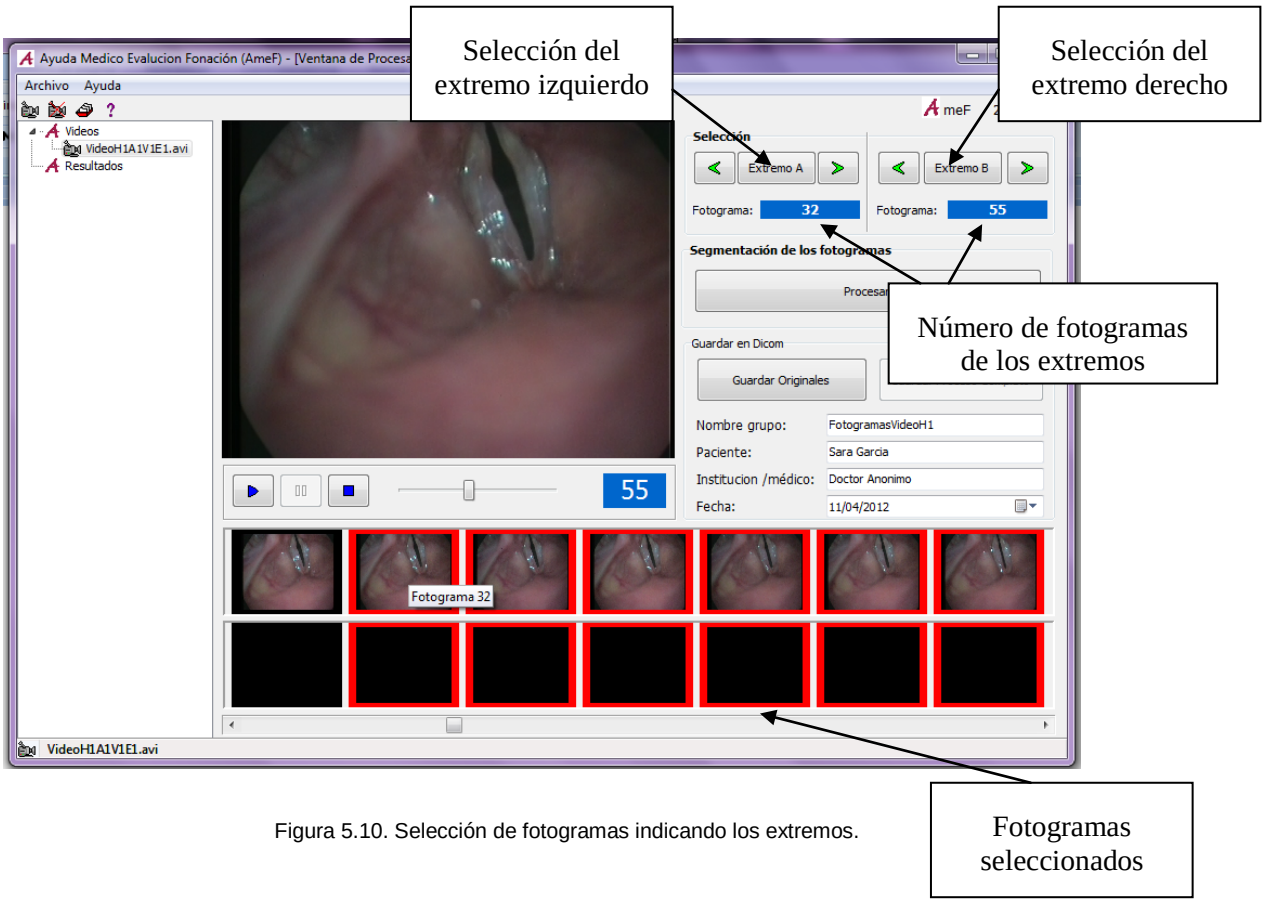


Figura 5.10. Selección de fotogramas indicando los extremos.

Una vez que hemos seleccionado las imágenes podemos llevar a cabo dos operaciones sobre ellas: guardar las originales como DICOM o procesarlas (obtener la glotis mediante segmentación). En la figura 5.11 se muestra una captura de la primera opción.

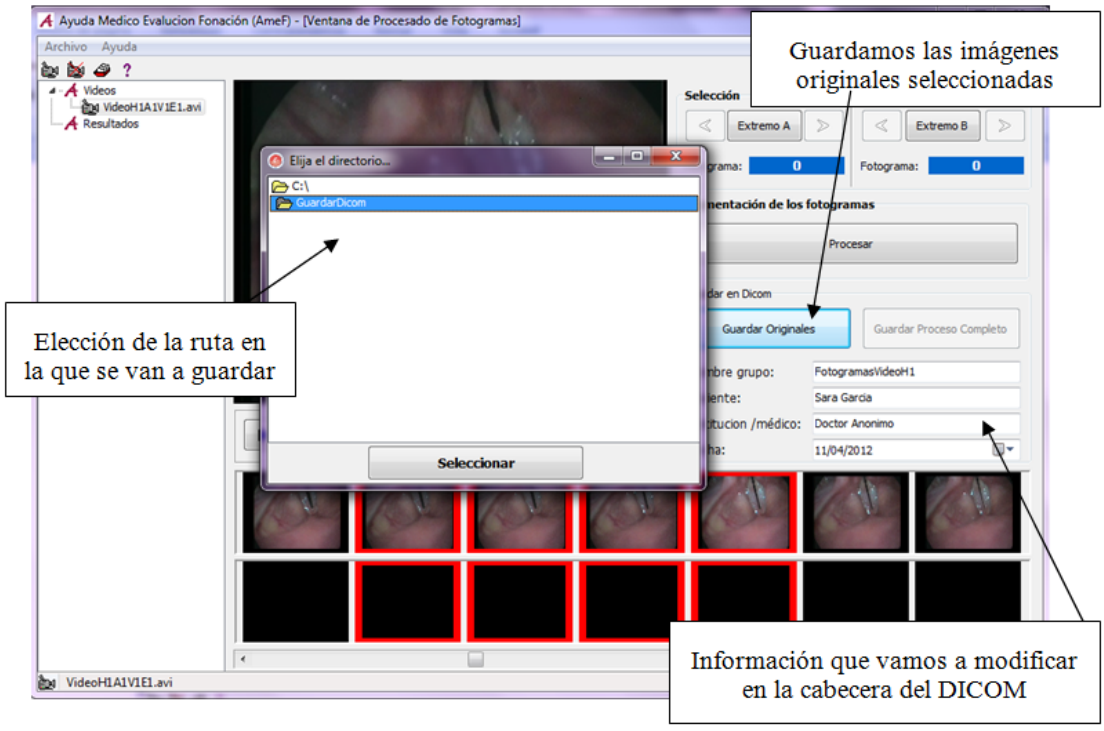


Figura 5.11. Captura de Guardar las imágenes originales seleccionadas como DICOM.

En este caso pulsamos el botón *Guardar Originales* y nos aparecerá una ventana para seleccionar la carpeta en la que queremos almacenarlos (Figura 5.11). Después de elegir la capeta y pulsar *Aceptar*, una ventana nos informará del éxito del proceso o de si se ha producido algún error.

Como podemos observar, debajo de los botones de guardar hay varios campos de texto para poder introducir información sobre el paciente, médico, la fecha o el nombre del grupo. Esta información se modificará para cada una de las imágenes guardadas. Todos estos fotogramas, una vez guardados como DICOM, pertenecerán al mismo grupo, es decir, les vamos a asignar a todos el mismo número para el campo *Series Number*.

En la figura 5.12 se muestra una captura del momento en el que, después de seleccionar las imágenes, pulsamos el botón de *Procesar*. Vamos a obtener debajo de cada imagen original otro fotograma resultado de aplicar el proceso de detección de la glotis referenciado anteriormente. En algunos casos esta operación no se podrá realizar; bien porque la glotis se encuentra demasiado cerrada o bien porque el proceso no se realiza correctamente. En este caso se mostrará una imagen en la parte inferior indicando que no se han obtenido resultados.

Para poder realizar la segmentación debe haber alguna imagen seleccionada, en caso contrario el programa devolverá un error y lo informará al usuario.

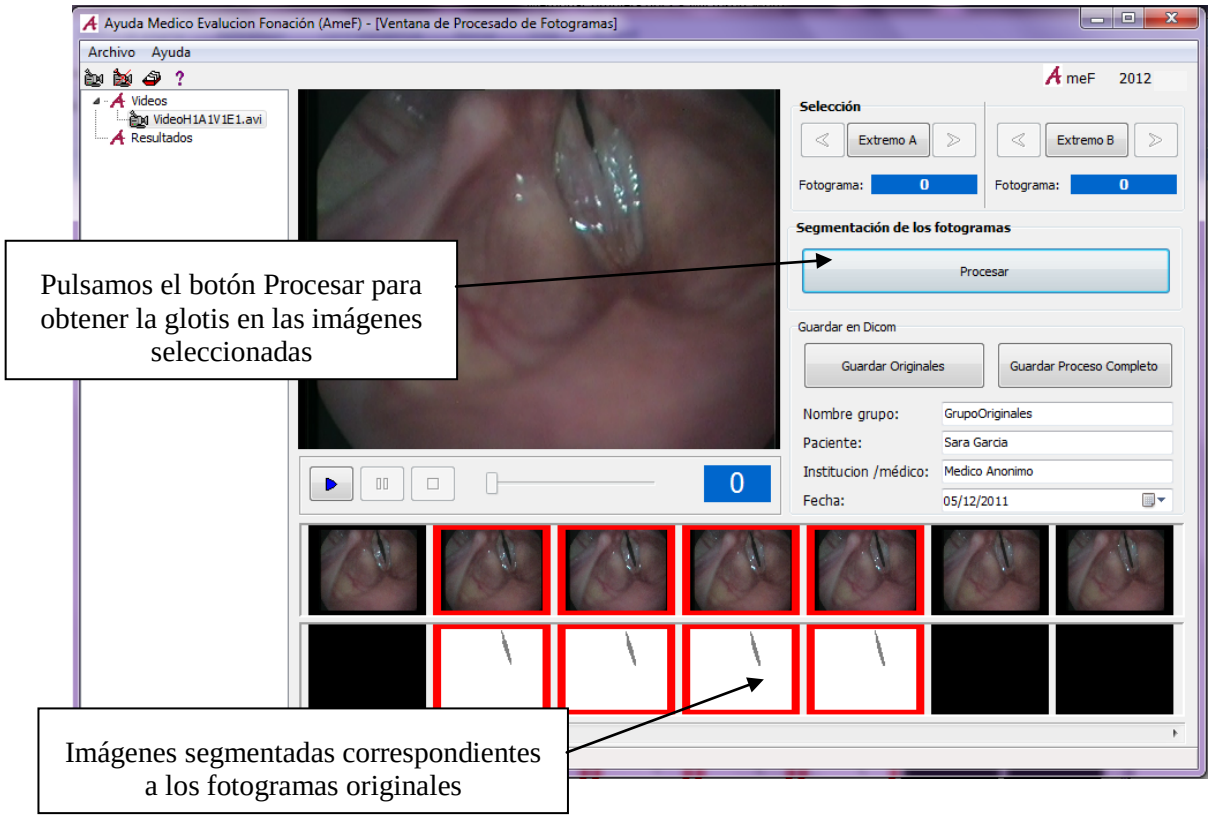


Figura 5.12. Resultado de segmentación de imágenes seleccionadas pulsando el botón *Procesar*.

Una vez seleccionadas y procesadas las imágenes la aplicación nos ofrece la opción de guardar el proceso completo (que hasta este momento se encontraba desactivada).

Al igual que en el caso de guardar las imágenes originales, tenemos la posibilidad de modificar cierta información de las cabeceras de los DICOM en los que se van a guardar estas imágenes. Las imágenes DICOM originales formarán parte de un grupo distinto al de los procesados.

Finalmente, elegimos la carpeta en la que queremos guardar todas las imágenes (tanto las originales como las obtenidas del proceso de segmentación).

El proceso completo realizado se muestra en la figura 5.13.

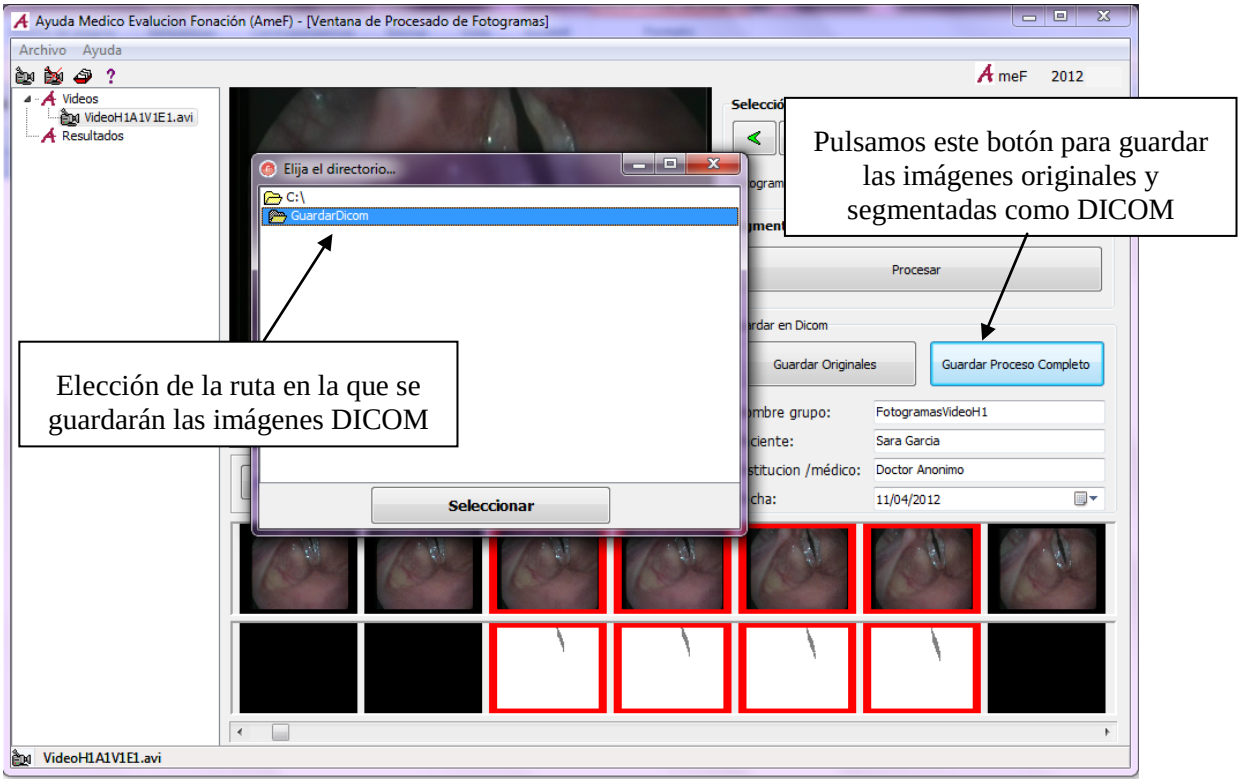


Figura 5.13.Captura obtenida al guardar todo el proceso (imágenes originales y segmentadas) como DICOM.

Como ya se ha explicado anteriormente en los Casos de Uso, además de operar con el vídeo, el sistema nos ofrece la opción de abrir y visualizar las imágenes DICOM que se han guardado mediante el proceso que se acaba de explicar.

En este caso, pulsamos el icono de los resultados y nos aparece una pantalla para seleccionar el directorio del cual queremos obtener las imágenes DICOM. Una vez seleccionado, en la parte izquierda de la pantalla (en la rama de Resultados), nos van a aparecer todas las imágenes con este formato que hay en ese directorio agrupadas según el grupo al que pertenecen.

En la figura 5.14 vemos una captura de lo que nos aparece al pulsar el botón de *Abrir Resultados*.

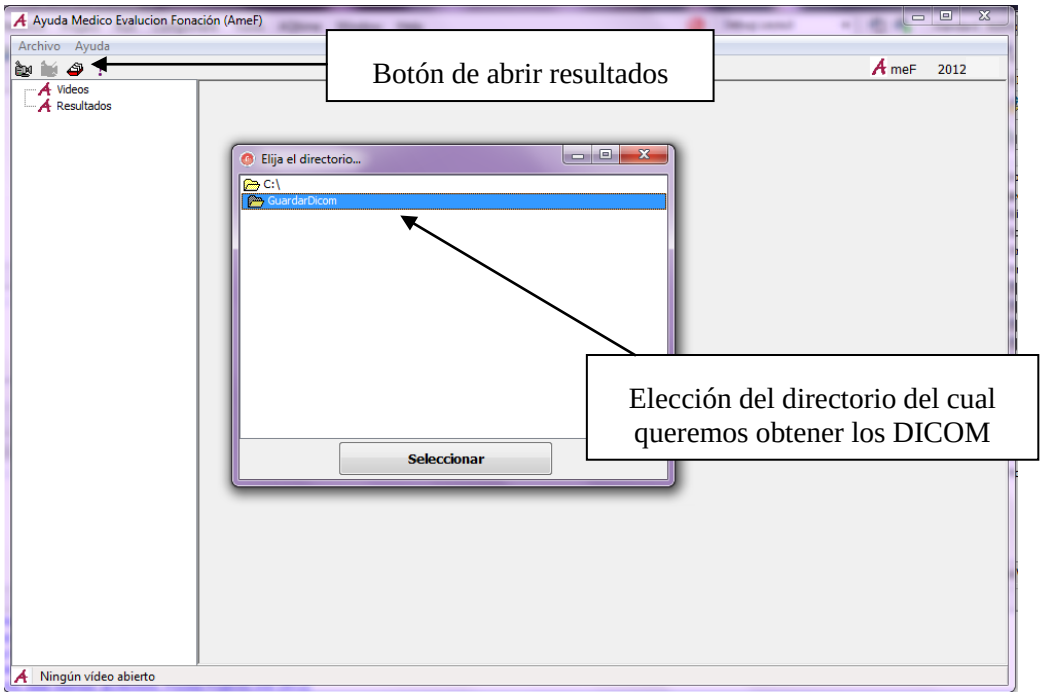


Figura 5.14. Seleccionar un directorio para observar los DICOM que se encuentran en él.

Una vez seleccionada la carpeta de los resultados, veremos en la parte izquierda de la pantalla todos los grupos de imágenes DICOM que hay en ella (Figura 5.15). A partir de este momento podemos seleccionar el nombre del grupo del cual queremos ver las imágenes y obtendremos en la parte inferior todas las imágenes que forman parte de él. Encima de éstas se muestra la imagen seleccionada (en color verde) aumentada e inmediatamente debajo de ella la información de la cabecera que hemos modificado antes de guardarlas.

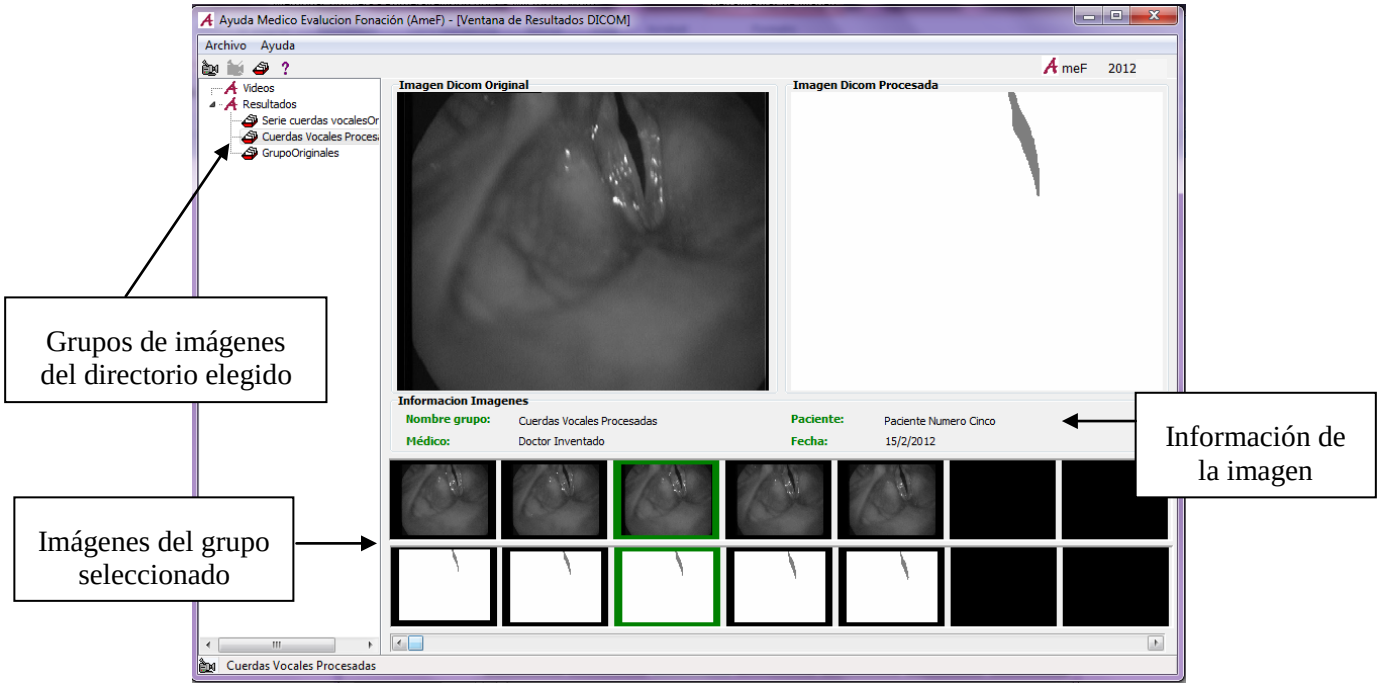


Figura 5.15. Pantalla de resultados para imágenes que se han segmentado.

Además de mostrar la imagen original y su segmentada correspondiente, si hacemos clic sobre uno de sus fotogramas podremos obtener la medida del área de la glotis en el mismo.

La figura 5.16 nos muestra una captura de pantalla de la aplicación después de hacer clic encima de la imagen segmentada, momento en el que nos muestra los píxeles que ocupa la glotis.

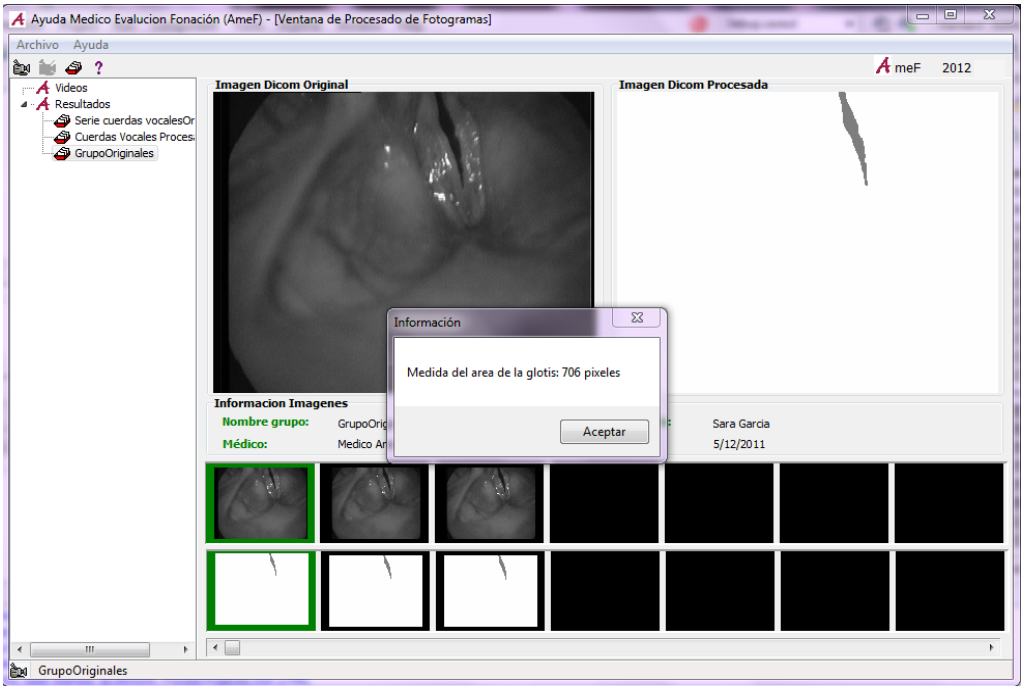


Figura 5.16. Pantalla de resultados para imágenes que se han segmentado después de realizar medida del área de la glotis.

En el caso de que sólo se hayan guardado las imágenes originales (no se ha realizado el proceso de segmentación), también podemos visualizarlas en la pantalla de resultados. La única diferencia con respecto a la pantalla mostrada anteriormente es que en este caso el programa nos indica que no hay imágenes procesadas y sólo muestra las originales, por tanto no permitirá obtener medidas de volumen. Podemos ver un ejemplo de captura en la figura 5.17.

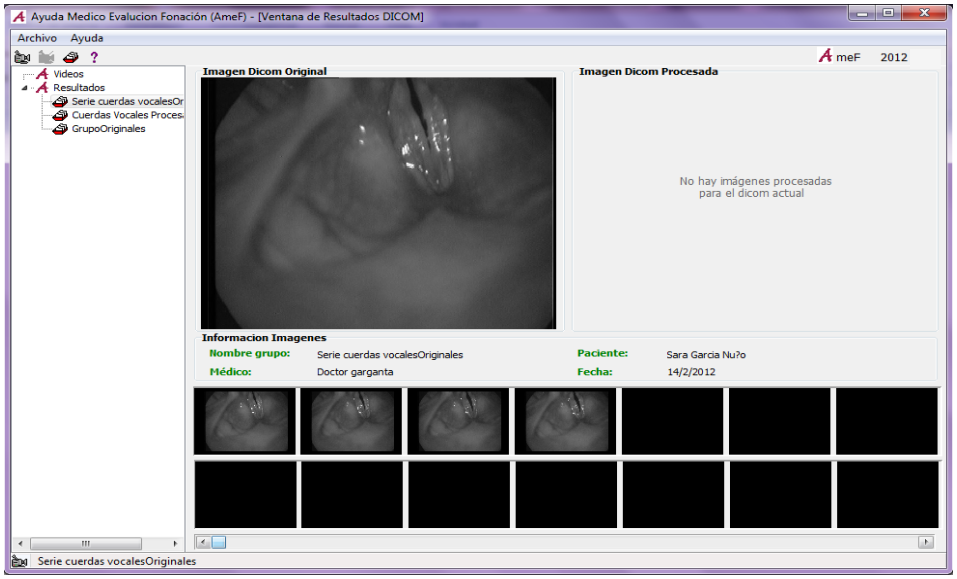


Figura 5.17. Pantalla de resultados para imágenes que no han sido segmentadas.

Todas las imágenes grabadas y convertidas a formato DICOM mediante esta aplicación se pueden abrir con un visor de DICOM estándar. En la figura 5.17 podemos observar una de las imágenes originales guardadas. En la parte izquierda de la pantalla se observan los grupos que hay en el directorio (el mismo que se ha abierto con nuestra aplicación) y las imágenes que contiene cada uno de ellos.

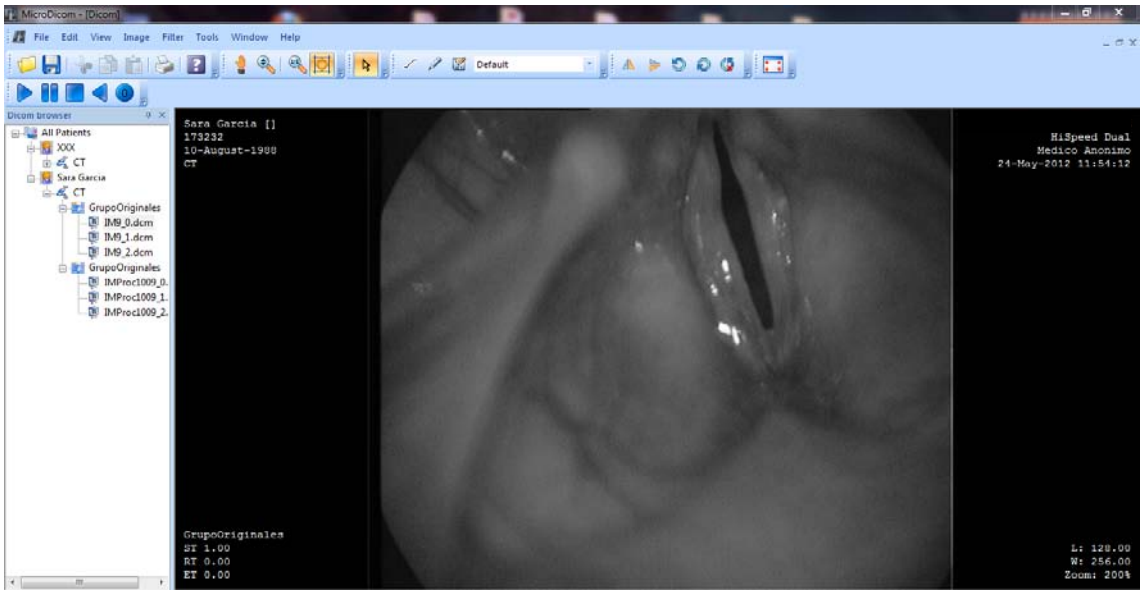


Figura 5.18. Imagen original DICOM guardada con nuestra aplicación y abierta con el visor *MicroDicom*.

De la misma manera, en la imagen 5.18 se muestra una de las imágenes guardadas en formato DICOM después de aplicar el proceso de segmentación.

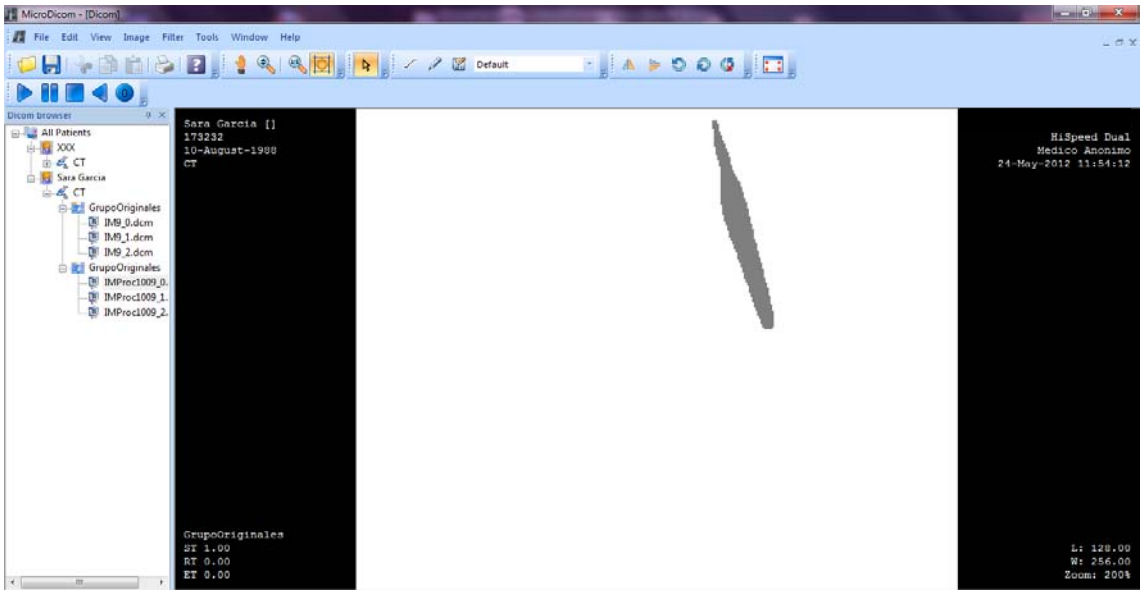


Figura 5.19. Imagen segmentada DICOM guardada con nuestra aplicación y abierta con el visor *MicroDicom*.

Capítulo 6

Consideraciones finales

6.1 APORTACIONES ORIGINALES

El presente trabajo propone una aplicación para la detección de la glotis en imágenes laríngeas y su posterior almacenamiento en el formato de imágenes médicas más usado actualmente: DICOM.

La operación de segmentación para detección de la glotis es fundamental para el desarrollo y correcto funcionamiento de muchos sistemas que permiten al profesional médico visualizar y caracterizar el movimiento vibratorio de las cuerdas vocales durante el proceso de fonación para la detección y diagnóstico de disfunciones del sistema fonador.

Además, su posterior almacenamiento en imágenes DICOM permite que el profesional médico pueda disponer de la imagen acompañada de toda la información del paciente y pueda visualizar todo ello en una misma aplicación, así como facilitar el intercambio de este tipo de imágenes gracias a que se encuentran guardadas en el formato estándar más utilizado para imágenes médicas.

A continuación se comentan las aportaciones originales del presente trabajo fin de carrera:

- Se han diseñado clases que permiten visualizar videos y procesarlos mediante la API de *DirecShow*. Estas clases, además de mostrar el video, permiten obtener todos los fotogramas que lo componen.
- Se han diseñado y programado un conjunto de clases que permiten explorar un directorio compuesto por imágenes DICOM obteniendo para cada una de ellas toda la información disponible en la cabecera. También se ha conseguido guardar un conjunto de imágenes como DICOM modificando la información del paciente que contiene.
- Se ha diseñado y desarrollado una aplicación que permite segmentar la glotis en vídeos e imágenes de la laringe para su posterior almacenamiento en imágenes con formato DICOM guardando, además de la propia información de la imagen, información importante para el profesional médico.
- La aplicación anterior incluye la posibilidad de realizar medidas de área en las imágenes obtenidas después de aplicar el proceso de segmentación y detección de la glotis.

- Se ha diseñado una aplicación que permite visualizar grupos de imágenes DICOM permitiendo comparar la imagen original con la obtenida a partir de aplicarle el proceso para detección de la glotis descrito en este trabajo. El visor también permite acceder a la información que guardan los DICOM en sus cabeceras.

6.2 CONCLUSIONES

Las patologías que pueden afectar a la producción de la voz son muchas y muy variadas. No obstante, su efecto común suele ser la generación de diferentes grados de dificultad para producir una vibración correcta de las cuerdas vocales durante la fonación, asociada a un defecto de cierre que agrava la situación.

El análisis de la vibración de los pliegues vocales resulta fundamental para el profesional de ORL a la hora de diagnosticar disfunciones del sistema fonador. Dicho análisis puede abordarse desde un punto de vista perceptual y/o desde un punto de vista objetivo. En el primer caso el análisis está condicionado por aspectos subjetivos a diferencia del punto de vista objetivo en el que el estudio se basa en la extracción de parámetros que permitan caracterizar el proceso de fonación. Independientemente del tipo de análisis empleado, la elevada velocidad con que se produce el movimiento de las cuerdas vocales supone un gran problema a la hora de visualizar el proceso con una mínima precisión. Para resolver este inconveniente se han ido desarrollando, a lo largo del último siglo, distintos métodos que de una u otra manera permiten captar el movimiento. Dentro de las técnicas subjetivas se pueden situar la estroboscopia y las grabaciones de alta velocidad; sin embargo, cada vez están cobrando una mayor importancia las técnicas objetivas que facultan al especialista para, además de visualizar, cuantificar el movimiento: técnicas quimográficas, diagramas de área glotal y fonovibrogramas.

El objetivo de estas técnicas es completar el análisis perceptual del profesional de ORL con medidas más objetivas de la vibración de los pliegues vocales, proporcionando datos exactos que permitan observar la evolución de un paciente.

Todas las técnicas objetivas citadas anteriormente necesitan de un procesado de imagen orientado a la segmentación del espacio glotal, bien como parte de su desarrollo, bien para solucionar diversos errores introducidos durante la exploración.

Para abordar la segmentación de la glotis existen muchas técnicas en el estado del arte, desde los más sencillos métodos clásicos, hasta los más complejos modelos de formas activas, pasando por algunos de gran popularidad como el crecimiento de región y los contornos activos.

En este trabajo se ha utilizado un sistema para la detección del espacio glotal basado en la transformada “Watershed” y distintos tipos de “Merging”, necesarios para solucionar el problema de sobresegmentación. El método trata de simular los pasos que seguiría un agente humano para la determinación de la glotis en una imagen de la laringe (búsqueda de una zona alargada, con nivel de gris cuasi-homogéneo y localmente rodeada por otra zona mucho más clara, las cuerdas vocales).

Todo el proceso de segmentación se realiza de una forma muy eficaz, con tiempos de ejecución bastante reducidos. Así, para la segmentación de 10 imágenes, la aplicación emplea un tiempo medio de 14 segundos lo que supone que para cada imagen tarda una media de 1,5 segundos aproximadamente. De la misma manera, para guardar en DICOM todo el proceso completo (tanto los fotogramas originales como los procesados) en 10 imágenes emplea 2,3 segundos, que es un tiempo muy reducido teniendo en cuenta que se tienen que guardar 20 imágenes (originales y segmentadas) con toda la información necesaria en la cabecera. En este último caso, emplearía 0,23 segundos para cada par de imagen original – segmentada y, por tanto, 0,115 segundos para cada imagen.

El formato DICOM es el mecanismo de codificación, almacenamiento y transmisión de imágenes aceptado universalmente por la comunidad médica. Lo que diferencia a las imágenes

DICOM de otros ficheros de datos es que, además de la imagen en sí, incluyen información sobre la misma tal como el paciente al que corresponde o el médico que ha realizado la prueba. Esta característica hace que sea ampliamente usado en el campo de la medicina, ya que una imagen médica no tiene sentido por sí sola, son necesarios los datos del paciente.

Un fichero DICOM contiene, por una parte, una cabecera y, por otra, todos los datos correspondientes a la imagen almacenada. La cabecera está formada por un conjunto de elementos de datos y almacena información sobre el paciente, el tipo de escáner, las dimensiones de la imagen, las condiciones en que se tomó y el formato interno de esta. El formato de fichero DICOM es muy complejo, debido a la gran cantidad de campos que se especifican en la cabecera, así como los varios tipos de cabecera que permite y la multitud de formatos en los que puede estar grabada la imagen.

Todas las herramientas anteriores se han integrado en una aplicación que permite la segmentación de la glotis en fotogramas resultado de la descomposición de un video. Una vez procesados o antes de ello, la aplicación permite el almacenamiento de los fotogramas como DICOM dentro de grupos modificando cierta información relevante de las cabeceras. El sistema actúa como un visor DICOM permitiendo ver todas las imágenes que se encuentran dentro de un grupo así como la información contenida en ellas. Además, ofrece la posibilidad de realizar medidas de área en las imágenes DICOM obtenidas después de aplicar el proceso de segmentación de la glotis.

6.3 LÍNEAS FUTURAS DE INVESTIGACIÓN

A pesar de los buenos resultados obtenidos a lo largo de la investigación, los trabajos desarrollados en este proyecto no cierran, ni mucho menos, ninguno de los campos tratados en el mismo. Por el contrario, profundizar en estos aspectos, abre multitud de vías posibles para la mejora del sistema. A continuación se proponen distintos avances susceptibles de llevarse a cabo:

- Existen muchas ampliaciones que podrían realizarse en el futuro y que serían de gran utilidad en el campo de la medicina y, en nuestro caso concreto en el campo de la fonación. Algunas de las ampliaciones posibles pasarían por la realización de quimogramas, fonovibrogramas o diagramas de área glotal dentro de nuestra aplicación, ya que estos constituyen técnicas muy útiles para extraer medidas a partir de la exploración de las cuerdas vocales en fonación. Además también se podría guardar los resultados de los mismos en imágenes DICOM junto a todo el resto de información del paciente.
- Se podría permitir al usuario variar las operaciones de segmentación y sus parámetros asociados, de forma que se puedan aplicar distintos procesos de segmentación sobre los fotogramas seleccionados. De esta manera, el usuario podría visualizar las diferencias entre los distintos procesos y escoger el que mejor se adapte a sus necesidades en cada momento.
- Permitir aplicar un proceso distinto o el mismo proceso pero con variación de sus parámetros para cada uno de los fotogramas. Así se podrían guardar dentro de un mismo grupo imágenes segmentadas a partir de procesos distintos lo que permitiría al profesional médico tratar de mejorar los resultados de la segmentación automática de la glotis en los fotogramas en los que estos no sean exactamente los esperados.
- Otra posible ampliación sería permitir al usuario que pueda variar más elementos de la cabecera DICOM de la imagen que va a guardar. En la aplicación que se presenta en el trabajo actual sólo se pueden modificar algunos datos importantes del paciente, del médico, el nombre del grupo y de la fecha de la exploración. Se podría extender para poder modificar también características de la imagen como el número de filas y columnas

o el valor del *WindowCenter* y *WindowWidth*. Esto último sería muy interesante ya que, mediante la variación de estos dos elementos de ofrecería al profesional de ORL la posibilidad de centrarse, y mostrar en pantalla, sólo la información de la imagen que le resulte de interés.

- Por último también se podría ampliar la aplicación para que realizara el proceso inverso: dado un conjunto de imágenes DICOM, permitir obtener el video resultado de la unión de todas ellas en un orden establecido. Además, ya que actualmente se pasa de las imágenes en formato bmp a formato DICOM, la aplicación podría transformar imágenes DICOM a bmp, poder visualizarlas y guardarlas con distintos formatos. De la misma manera, se podría obtener un video resultado de la unión de las imágenes de la glotis ya discriminada para ver únicamente este elemento.

Capítulo 7

Referencias bibliográficas

7.1 REFERENCIAS

[Alonso-Hernández2008] Alonso-Hernández, J. B., Travieso-González, C. M., Ferrer-Ballester, M. A., de León-y de Juan, J. y Godino-Llorente, J. I., *La evaluación acústica del sistema fonador*, Vicerrectorado de calidad e innovación educativa de la Universidad de Las Palmas de Gran Canaria, 2008.

[Anón.2008] Vocal parts. General structure. 2008. Blue Tree Publishing. www.bluetreepublishing.com.

[Anón.2009a] Dicom Standard. 2009. Medical Nema. http://medical.nema.org/

[Anón.2009b] "High-speed cameras", www.photron.com, 2009.

[Anón.2011] "KayPentax", www.kayelemetrics.com, 2011.

[Baken2000] Baken, R. J. y Orlikoff, R. F., *Clinical measurement of speech and voice*, 2 ed., Singular Thomson Learning™, 2000.

[Beucher1979] Beucher, S. y Lantuéjoul, C., "Use of watersheds in contour detection", en *Proceedings of the International Workshop on Image Processing: Real-Time Edge and Motion Detection/Estimation*, vol. 132, pp. 2.1-2.12, Sept.1979.

[Beucher1992] Beucher, S., "The watershed transformation applied to image segmentation", *Scanning Microscopy*, vol. 6, pp. 299-314, 1992.

[Bieniek2000] Bieniek, A. y Moga, A. N., "An efficient watershed algorithm based on connected components", *Pattern Recognition*, vol. 33, no. 6, pp. 907-916, 2000.

[Bleau1992a] Bleau, A., De Guise, J. y LeBlanc, R., "A new set of fast algorithms for mathematical morphology: I-Idempotent geodesic transforms", *CVGIP: Image Understanding*, vol. 56, no. 2, pp. 178-209, Sept.1992.

[Bleau1992b] Bleau, A., De Guise, J. y LeBlanc, R., "A new set of fast algorithms for mathematical morphology: II-Identification of topographic features on grayscale images", *CVGIP: Image Understanding*, vol. 56, no. 2, pp. 210-229, Sept.1992.

[Bleau2000] Bleau, A. y Leon, L. J., "Watershed-based segmentation and region merging", *Computer Vision and Image Understanding*, vol. 77, no. 3, pp. 317-370, Mar.2000.

[Bueno2001] Bueno, G., Musse, O., Heitz, F. y Armspach, J. P., "Three-dimensional segmentation of anatomical structures in MR images on large data bases", *Magnetic Resonance Imaging*, vol. 19, no. 1, pp. 73-88, 2001.

[Cootes2000] Cootes, T. F., *An introduction to active shape models*, Oxford: Oxford University Press, 2000.

[Childers2000] Childers, D. G., *Speech processing and synthesis toolboxes*, New York: John Wiley & Sons, 2000.

[Com08] DICOM committee, editor. PS 3.10: Media Storage and File Format for Data Interchange. DICOM Standard, 2008.

[Dollinger2009] Dollinger, M., Lohscheller, J., McWhorter, A. y Kunduk, M., "Variability of normal vocal fold dynamics for different vocal loading in one healthy subject investigated by phonovibrograms", *Journal of Voice*, vol. 23, no. 2, pp. 175-181, 2009.

[Duda2001] Duda, R. O., Hart, P. E. y Stork, D. G., *Pattern Classification*, 2 ed., Wiley-Interscience, 2001.

[Eysholdt2003] Eysholdt, U., Rosanowski, F. y Hoppe, U., "Vocal fold vibration irregularities caused by different types of laryngeal asymmetry", *European Archives of Oto rhinolaryngology*, vol. 260, no. 1, pp. 412-417, 2003.

[Garcia-Tapia1996] Garcia-Tapia, R. y Cobeta, I., *Diagnóstico y tratamiento de los trastornos de la voz*, 1 ed., Garsi, 1996.

[Godino-Llorente2002] Godino-Llorente, J. I., "Estrategias para la detección automática de patología laríngea a partir del registro de la voz", Tesis doctoral, Facultad de Informática, Universidad Politécnica de Madrid, 2002.

[Gonzalez1992] Gonzalez, R. C. y Woods, R. E., *Digital Image Processing*, Addison-Wesley, 1992.

[Gonzalez2004] Gonzalez, R. C., Woods, R. E. y Eddins, S. L., "Segmentation using the watershed transform," in Gonzalez, R. C., Woods, R. E., and Eddins, S. L. (eds.) *Digital Image Processing Using MATLAB* NJ, USA: Pearson Prentice Hall, 2004, pp. 417-425.

[Haris1998] Haris, K., Efstratiadis, S. N., Maglaveras, N. y Katsaggelos, A. K., "Hybrid image segmentation using watersheds and fast region merging", *IEEE Transactions on Image Processing*, vol. 7, no. 12, pp. 1684-1699, Dic.1998.

[Hixon2008] Hixon, T. J., Weismer, G. y Hoit, J. D., *Preclinical speech science*, Plural Publishing, 2008.

[Hirose1988] Hirose, H., "High-speed digital imaging of vocal fold vibration", *Acta Oto Laryngologica*, vol. 105, no. 458, pp. 151-153, 1988.

[Jackson-Menaldi1992] Jackson-Menaldi, M. C. A., *La voz normal*, Editorial médica panamericana, 1992.

[Jackson-Menaldi2002] Jackson-Menaldi, M. C. A., *La voz patológica*, Editorial Médica Panamericana, 2002.

[Johnson1998] Johnson, R. A. y Wichern, D., *Applied multivariate statistical analysis*, 4 ed., Prentice-Hall, 1998.

[Kass1988] Kass, M., Witkin, A. y Terzopoulus, D., "Snakes: active contour models", *International Journal of Computer Vision*, vol. 1, pp. 321-331, 1988.

[Kim2003] Kim, D. Y., Kim, L. S., Kim, K. H., Sung, M. W., Roh, J. L., Kwon, T. K., Lee, S. J., Choi, S. H., Wang, S. G. y Sung, M. Y., "Videostrobokymographic analysis of benign vocal fold lesions", *Acta Oto Laryngologica*, vol. 123, no. 9, pp. 1102-1109, 2003.

[Kiritani1986] Kiritani, S., Honda, K., Imagawa, H. y Hirose, H., "Simultaneous high-speed digital recording of vocal fold vibration and speech signal", en *Proceedings of IEEE ICASSP'86*, vol. 11, pp. 1633-1636, Abr.1986.

[Kiritani1993] Kiritani, S., Hirose, H. y Imagawa, H., "High-speed digital image analysis of vocal cord vibration in diplophonia", *Speech Communication*, vol. 13, pp. 23-32, 1993.

[Larsson2000] Larsson, H., Hertegard, S., Lindestad, P. A. y Hammarberg, B., "Vocal fold vibrations: high-speed imaging, kymography, and acoustic analysis: a preliminary report", *Laryngoscope*, vol. 110, no. 12, pp. 2117-2122, 2000.

[Lee2001] Lee, J. S., Kim, E., Sung, M. W., Kim, K. H. y Park, K. S., "A method for assessing the regional vibratory pattern of vocal folds by analysing the video recording of stroboscopy", *Medical & Biological engineering & Computing*, vol. 39, no. 3, pp. 273-278, 2001.

[Lohscheller2007] Lohscheller, J., Toy, H., Rosanowski, F., Eysholdt, U. y Dollinger, M., "Clinically evaluated procedure for the reconstruction of vocal fold vibrations from endoscopic digital high-speed videos", *Medical Image Analysis*, vol. 11, pp. 400-413, 2007.

[Lohscheller2008] Lohscheller, J., Eysholdt, U., Toy, H. y Dollinger, M., "Phonovibrography: mapping high-speed movies of vocal fold vibrations into 2D-diagrams for visualizing and analyzing the underlying laryngeal dynamics", *IEEE Transactions on Medical Imaging*, vol. 27, no. 3, pp. 300-309, 2008.

[Manfredi2006] Manfredi, C., Bocchi, L., Bianchi, S., Migali, N. y Cantarella, G., "Objetive vocal fold vibration assessment from videokymographic images", *Biomedical signal processing and control*, vol. 1, no. 2, pp. 129-136, 2006.

[McGillion2000] McGillion, M. A., "Automated analysis of voice quality", Tesis doctoral, University of Manchester Institute of Science and Technology, 2000.

[Méndez-Zorrilla2008] Méndez-Zorrilla, A., Osma-Ruiz, V. J., García-Zapirain, B., Sáenz-Lechón, N., Ruiz-Oleagordia, I. y Fraile, R., "Diagnosis of vocal folds morphological pathologies by means of avanced image processing methods", en *Proceedings of ISIVC 2008*, vol. 1, trabajo 240, Jul.2008.

[Milutinovic1996] Milutinovic, Z., "Classification of voice pathology", *Folia Phoniatrica et Logopaedica*, vol. 48, pp. 301-308, 1996.

[Moore1962] Moore, G. P., White, F. D. y Von Leden, H., "Ultra high speed photography in

laryngeal physiology", *Journal of Speech and Hearing Disorders*, vol. 27, no. 2, pp. 165-171, 1962.

[Oertel1878] Oertel, M. J., "Über eine neue 'laryngostroboskopische' untersuchungsmethode des kehlkopfes", *Zentralbl.f.d.Mediz.Wissenschaften Heft*, vol. 16, pp. 81-82, 1878.

[Olthoff2007] Olthoff, A., Woywod, C. y Kruse, E., "Stroboscopy versus high-speed glottography: a comparative study", *Laryngoscope*, vol. 117, no. 6, pp. 1123-1126, 2007.

[Osma-Ruiz2007] Osma-Ruiz, V. J., Godino-Llorente, J. I., Sáenz-Lechón, N. y Gómez-Vilda, P., "An improved watershed algorithm based on efficient computation of shortest paths", *Pattern Recognition*, vol. 40, no. 3, pp. 1078-1090, 2007.

[Osma-Ruiz2008a] Osma-Ruiz, V. J., Godino-Llorente, J. I., Sáenz-Lechón, N. y Fraile, R., "Segmentation of the glottal space from laryngeal images using the watershed transform", *Computerized Medical Imaging and Graphics*, vol. 32, no. 3, pp. 193-201, 2008.

[Osma-Ruiz2008b] Osma-Ruiz, V. J., Sáenz-Lechón, N., Godino-Llorente, J. I. y Fraile, R., "Detección del espacio glotal en imágenes laríngeas mediante transformada watershed y merging JND", en *Actas CASEIB 2008*, vol. 1, pp. 1-4, Oct.2008.

[Osma-Ruiz2010] Osma-Ruiz, V. J., Tesis: "Contribución al procesado digital de imágenes para la caracterización de patologías laríngeas", Escuela Universitaria de Ingeniería Técnica de Telecomunicaciones UPM, 2010.

[Palm2001] Palm, C., Lehmann, T. M., Bredno, J., Neuschaefer-Rube, C., Klajman, S. y Spitzer, K., "Automated analysis of stroboscopic image sequences by vibration profiles", en *Proceedings of the 5th International Workshop on Advances in Quantitative Laryngology, Voice and Speech Research*, vol, 1, Groningen, Netherlands, pp. 1-7, Abr.2001.

[Qiu2003] Qiu, Q., Schutte, H. K., Gu, L. y Yu, Q., "An automatic method to quantify the vibration properties of human vocal folds via videokymography", *Folia Phoniatrica et Logopaedica*, vol. 55, pp. 128-136, 2003.

[Rabiner1978] Rabiner, L. R. y Schafer, R. W., *Digital processing of speech signals*, Englewood Cliffs, NJ: Prentice Hall, 1978.

[Schwarz2006] Schwarz, R., Hoppe, U., Schuster, M., Wurzbacher, T., Eysholdt, U. y

Lohscheller, J., "Classification of unilateral vocal fold paralysis by endoscopic digital highspeed recordings and inversion of a biomechanical model", *IEEE Transactions on Biomedical Engineering*, vol. 53, no. 6, pp. 1099-1108, 2006.

[Shen2003] Shen, D. F. y Huang, M. T., "A watershed-based image segmentation using JND property", en *Proceedings of IEEE ICASSP 2003*, vol. 3, pp. 377-380, Abr.2003.

[Sung1999] Sung, M. W., Kim, K. H., Koh, T. Y., Kwon, T. Y., Mo, J. H., Choi, S. H., Lee, J. S., Park, K. S., Kim, E. J. y Sung, M. Y., "Videostrobokymography: a new method for the quantitative analysis of vocal fold vibration", *The Laringoscope*, vol. 109, no. 11, pp. 1859-1863, 1999.

[Švec1996] Švec, J. G. y Schutte, H. K., "Videokymography: high-speed line scanning of vocal fold vibration", *Journal of Voice*, vol. 10, no. 2, pp. 201-205, 1996.

[Švec2002] Švec, J. G. y Šram, F., "Kymographic imaging of the vocal fold oscillations", en *Proceedings of ICSLP 2002*, vol. 2, pp. 957-960, Sept.2002.

[Vincent1991] Vincent, L. y Soille, P., "Watersheds in digital spaces: an efficient algorithm based on immersion simulations", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 13, no. 6, pp. 583-598, Jun.1991.

[Wittenberg1998] Wittenberg, T., Tigges, M., Mergell, P. y Eysholdt, U., "Visualization of hoarseness by imaging techniques", en *Proceedings of VOICEDATA'98*, pp. 57-63, 1998.

[Woo1996] Woo, P., "Quantification of videostrobolaryngoscopic findings - Measurements of the normal glottal cycle", *Laryngoscope*, vol. 106, no. 3, pp. 1-27, 1996.

[Yan2005] Yan, Y., Ahmad, K., Kunduk, M. y Bless, D., "Analysis of vocal-fold vibrations from high-speed laryngeal images using a Hilbert transform-based methodology", *Journal of Voice*, vol. 19, no. 2, pp. 161-175, 2005.

[Yan2006] Yan, Y., Chen, X. y Bless, D., "Automatic tracing of vocal-fold motion from highspeed digital images", *IEEE Transactions on Biomedical Engineering*, vol. 53, no. 7, pp. 1394-1400, Jul.2006.

[Yan2007] Yan, Y., Damrose, E. y Bless, D., "Functional analysis of voice using simultaneous high-speed imaging and acoustic recordings", *Journal of Voice*, vol. 21, no. 5, pp. 604-616, 2007.

[Yang2005] Yang, X. K., Ling, W. S., Lu, Z. K., Ong, E. P. y Yao, S. S., "Just noticeable distorsion model and its applications in video coding", *Signal Processing: Image Communication*, vol. 20, pp. 662-680, 2005.

Sara García Nuño, autora del proyecto y abajo firmante autoriza a la Universidad Complutense de Madrid a difundir y utilizar con fines académicos, no comerciales y mencionando expresamente a sus autores, tanto la propia memoria, como el código, la documentación y/o el prototipo desarrollado.

Madrid, a 20 de Junio de 2012.