



Universidad de Deusto

FACULTAD DE INGENIERÍA

TESIS DOCTORAL

Algoritmo de Parametrización de Imágenes Estroboscópicas de Cuerdas Vocales

Presentada por Dña. AMAIA MÉNDEZ ZORRILLA

dentro del Programa de Doctorado CIENCIA DE LA COMPUTACIÓN

Digirida por la Dra. Dña. BEGOÑA GARCÍA ZAPIRAIN

La directora

La Doctoranda

Septiembre 2012

A mi madre

Resumen

La realización de esta tesis ha sido motivada por el impulso de la ingeniería biomédica en los últimos años, el avance de la tecnología en las diferentes especialidades médicas y concretamente de las peculiaridades de la otorrinolaringología en este aspecto, la existencia de multitud de profesionales alrededor del estudio de las imágenes estroboscópicas de cuerdas vocales, y la falta de objetivación en la emisión de los diagnósticos de este área.

El principal objetivo es que esta tesis sea una aportación a la comunidad científica y médica en el ámbito de la otorrinolaringología, aplicando algoritmos de procesamiento digital de imagen como ayuda al diagnóstico. La aportación se centra en la parametrización y cuantificación de las características de las patologías de cuerdas vocales a partir de imágenes estroboscópicas de forma automatizada. Para ello, se han acotado las patologías a aquellas relacionadas con la morfología de la cuerda vocal, y patologías relacionadas con el movimiento o ausencia de movimiento de las cuerdas vocales. Para cumplir este objetivo, se han aplicado técnicas de análisis de texturas para la segmentación, concretamente los filtros de Gabor. Se ha utilizado el Análisis de Componentes Principales para realizar una clasificación de las imágenes en función de la patología, y se ha estudiado el movimiento mediante algoritmos de Block Matching. La extracción de los parámetros se ha efectuado usando todos los resultados de las técnicas previamente aplicadas.

Los resultados obtenidos experimentalmente sobre una base de datos de 1262 imágenes muestran ratios de éxito muy satisfactorios, siendo del 95,07% en la segmentación, alcanzando un 92,10% en la preclasificación previa a la medida de los parámetros. Por último, atendiendo a las secuencias y no a las imágenes individuales, el algoritmo ha propuesto el mismo diagnóstico que el médico en el 100% de las estudiadas.

Estos resultados demuestran la eficacia de los algoritmos propuestos y dejan la puerta abierta hacia el desarrollo de algoritmos para parametrizar el cáncer de cuerdas vocales, o hacia el estudio de la vascularización y el color que también pueden ser indicadores de patología.

Abstract

Biomedical engineering impact in recent years, technological breakthroughs in different medical specialties, otolaryngology peculiarities, the presence of different professionals around the world working on the study of voice disorders and stroboscopic images, together with the lack of diagnosis objectifying in this field have encouraged carrying out this thesis.

The main goal of this thesis is to be a contribution to scientific and medical communities by applying advanced image digital processing algorithms to support diagnosis in the field of otolaryngology. The contribution of this thesis is focused on the vocal cord pathology characteristics parameterization and quantification using stroboscopic images in an automatized way. To do it, this study has been limited to pathologies related either to vocal cord morphology or vocal cord movement. To achieve these objectives, texture analysis techniques for segmentation, namely Gabor filters, were applied. Principal Component Analysis was used to classify the different images according to the pathology, and movement has been studied using Block Matching algorithms. The extraction of the parameters was made using all the results of the techniques previously applied.

The results obtained experimentally on a 1262 images database show satisfactory success rates of 95.07% in the segmentation phase, reaching up to 92.10% on the preclassification previous to the parameters measurement. Finally, in response to sequences rather than to individual images, this algorithm has proposed the same diagnostic as the physician in 100% of the cases studied.

These results show the effectiveness of the proposed algorithms and leave the door open to the development of computer-aided diagnosis systems of vocal pathologies using vocal cords images.

Laburpena

Tesi hau burutzearen oinarria Ingeniaritza Biomedikoak azken urteetan izan duen bultzadan, medikuntza arloetan emandako aurrerapen teknologikoetan eta bereziki otorrinolaringologiak dituen berezitasunetan dago, non irudi estroboskopikoen ikerketan profesional ezberdin asko parte hartzen duten eta diagnostikoak egiterakoan objektibazio falta dagoen arlo honetan.

Tesi honen helburu nagusia zientzia-elkartean eta medikuntzan, otorrinolaringologia-alorrean, ekarri handia izatea da, horretarako irudi digitalen prozesamendu algoritmoak lagunduriko diagnostikoan erabiliz. Ekarriaren muina parametrizazioan eta ahots-kordetako patologien ezaugarrien kuantifikazioan da, irudi estroboskopikoen ikertze automatizatuaren bidez. Hori lortzeko ikerketa, ahots-korden morfologi patologietara eta ahots-korden mugimendu patologietara eda mugimendu falta dituzten patologietara mugatu da.

Helburu hau lortzeko segmentazioa lortzeko testura-analisirako teknikak erabili dira, Gabor-delako filtroak hain zuzen ere. Osagai Nagusien Analisia erabili da, patologiarekin arabera, irudien sailkapena egiteko eta ahots-korden mugimendua Block Matching-delako algoritmen bidez aztertu egin da. Parametroen ateratzea alde aurretik erabilitako tekniken emaitz guztien bidez lortu egin da.

Lortutako emaitza experimentalek, 1262 irudizko datu-base batetan oinarrituta, arrakasta ratio onak erakusten dute, 95,07% segmentazioan eta 92,10% parametroen aurretiko neurketen presailkapenean. Azkenik, irudien sekuentzietan onarrituta, eta ez banakako irudietan, algoritmoak 100% kasuetan medikuaren diagnostiko berdina proposatzen du.

Emaitza hauek proposatutako algoritmoen efikazia egiaztatzen dute eta ahots-korden-patologia arloan, algoritmoen garapenerako atea irekita usten dute ahots-korden mimbizia parametrizatzeko, baskularizazioa aztertze eta kolorea ikertzeko zeintzuk patologi adierazleak ere izan ditezke.

Agradecimientos

Llegado este punto, quiero hacer constar mi sincero agradecimiento a todas las personas que han contribuido en mayor o menor medida al desarrollo de esta tesis.

En primer lugar, a mi directora Begoña García Zapirain, por su paciencia, comprensión y apoyo a lo largo del camino recorrido, y sobre todo, por la confianza depositada en mí.

Gracias también a todos los integrantes de la unidad DeustoTech-Life, compañeros que considero amigos, que han estado presentes durante momentos críticos y siempre han tenido una sonrisa.

Especial mención he de hacer a Javier Vicente, por su tiempo, por su amistad incondicional desde hace ya casi 15 años, y por las continuas molestias que le he ocasionado en esta última etapa. No ha sido el único, tengo que mencionar la ayuda y el apoyo incondicional de Ainara Sudupe y Lourdes Pinto.

Recalcar también la colaboración del otorrinolaringólogo Agustín Pérez Izquierdo, que ha proporcionado parte de la base de datos necesaria para el desarrollo de esta tesis.

Por último, dar las gracias a mi familia, especialmente a Gorka Sáez Zorrilla, cuyo detalle en la revisión de los errores tipográficos, ha sido excepcional.

Muchas gracias a todos aquellos que han creído en este trabajo como yo o no, pero que en algún momento han influido o han estado presentes en mi vida durante los últimos años.

Índice

1.	INTRODUCCIÓN.....	1
1.1	JUSTIFICACIÓN DE LA INVESTIGACIÓN	3
1.2	HIPÓTESIS	4
1.3	OBJETIVOS	5
1.4	METODOLOGÍA.....	7
1.4.1	Elección de los participantes de la investigación	9
1.5	ESTRUCTURA DE LA TESIS	10
2.	ESTADO DEL ARTE	13
2.1	ANTECEDENTES MÉDICOS.....	15
2.1.1	Descripción del aparato de fonación.....	15
2.1.2	Descripción anatómica de las cuerdas vocales	17
2.1.3	Trastornos de las cuerdas vocales	19
2.1.3.1	Laringitis	22
2.1.3.2	Reflujo de ácido del estómago	23
2.1.3.3	Nódulos en las cuerdas vocales.....	23
2.1.3.4	Pólipos	24
2.1.3.5	Quistes.....	25
2.1.3.6	Parálisis de las cuerdas vocales	26
2.1.3.7	Cáncer de laringe	26
2.1.3.8	Disfonía espasmódica	27
2.1.3.9	Úlceras de contacto.....	27
2.1.3.10	Curvatura del pliegue vocal.....	28

2.1.3.11	Papilomatosis laríngea	28
2.2	ANTECEDENTES TÉCNICOS	29
2.2.1	Técnicas de captura de imágenes y software de análisis.....	29
2.2.1.1	Laringoscopia refleja.....	31
2.2.1.2	Microlaringoscopia laríngea	33
2.2.1.3	Exploración endoscópica	34
2.2.1.4	Estroboscopia.....	39
2.2.1.5	Videoquimografía	45
2.2.1.6	Software de Análisis	47
2.2.2	Alternativas de lenguajes para el desarrollo de algoritmos de procesado digital de imagen.....	48
2.2.2.1	ImageJ	48
2.2.2.2	SciPython.....	49
2.2.2.3	OpenCV	49
2.2.2.4	Matlab	49
2.2.3	Técnicas y métodos generales empleados en el procesado digital de imagen	50
2.2.3.1	Mejora de la imagen	51
2.2.3.2	Detección de bordes.....	64
2.2.4	Otras técnicas evaluadas	75
2.2.4.1	Contornos Activos	75
2.2.4.2	Transformada Wavelet.....	76
2.2.4.3	Filtro de Kalman.....	77
3.	BASE DE DATOS	79
3.1	DESCRIPCIÓN DE LA BASE DE DATOS.....	80
3.2	ETIQUETADO DE LA BASE DE DATOS.....	81

4.	DISEÑO	83
4.1	DISEÑO DE ALTO NIVEL	85
4.1.1	Etapa de Pre-Procesado	87
4.1.2	Etapa de Segmentación	87
4.1.3	Etapa de Detección del Movimiento	88
4.1.4	Etapa de Clasificación	89
4.1.5	Etapa de Análisis y Medida.....	89
4.2	DISEÑO DE BAJO NIVEL	91
4.2.1	Etapa de Pre-Procesado	91
4.2.2	Etapa de Segmentación	99
4.2.3	Etapa de Detección de Movimiento	106
4.2.4	Etapa de Clasificación	112
4.2.5	Etapa de Análisis y Medida.....	120
4.2.5.1	Análisis y Medida de parámetros relacionados con la morfología.....	123
4.2.5.2	Análisis y Medida de parámetros relacionados con el movimiento.....	129
4.2.6	Decisión	132
5.	RESULTADOS.....	135
5.1	CONSIDERACIONES PREVIAS.....	136
5.1.1	Entorno de desarrollo.....	136
5.1.2	Hardware utilizado para el desarrollo y pruebas	136
5.2	EVALUACIÓN DE RESULTADOS PARCIALES	137
5.2.1	Pruebas de la Etapa de Segmentación	137
5.2.2	Pruebas de la Etapa de Clasificación.....	147
5.2.2.1	Experimento de Identificación.....	148

5.2.2.2	Experimento de Clasificación.....	149
5.2.3	Resultados de la Etapa de Detección de Movimiento.....	152
5.2.4	Pruebas de la Etapa de Análisis y Medida	157
5.2.4.1	Análisis y Medida de las imágenes de cuerdas vocales clasificadas como morfológicamente patológicas.....	158
5.2.4.2	Análisis y Medida de las imágenes de cuerdas vocales clasificadas como morfológicamente sanas	163
5.3	REPRESENTACIÓN GRÁFICA DE LOS RESULTADOS	166
5.3.1	Representación gráfica de los resultados por secuencia.....	166
5.3.2	Representación gráfica reducida de los resultados por secuencia	170
5.3.3	Representación gráfica final de los resultados.....	173
5.4	INCIDENCIAS ENCONTRADAS	175
5.4.1	Incidencias referentes a la obtención de la base de datos.....	176
5.4.2	Incidencias en el diseño y desarrollo de los algoritmos	176
5.4.3	Incidencias durante la evaluación del sistema.....	177
6.	CONCLUSIONES.....	179
6.1	CUMPLIMIENTO DE OBJETIVOS.....	179
6.2	IMPACTO CIENTÍFICO	182
6.2.1	Publicaciones científicas	182
6.2.2	Propiedad intelectual.....	185
6.2.3	Proyectos de investigación y colaboraciones externas.....	185
6.3	FUTURAS LÍNEAS DE INVESTIGACIÓN	186
7.	BIBLIOGRAFÍA	189
	ANEXO A. GLOSARIO DE TÉRMINOS	201

Índice de Figuras

Figura 1. Etapas de la Investigación.....	8
Figura 2. Ligamentos de la laringe. Visión Posterior [Gray 18].	17
Figura 3. a) Diagrama de cuerdas vocales abiertas o en abducción. b) Cuerdas vocales cerradas o en aducción. c) En abducción durante una inspiración forzada. d) En aducción durante la fonación.....	18
Figura 4. Funcionamiento normal de las cuerdas vocales	21
Figura 5. Funcionamiento anormal de las cuerdas vocales	22
Figura 6. Cuerdas vocales con nódulos	23
Figura 7. Cuerdas vocales con pólipos	25
Figura 8. Quiste en cuerda vocal derecha	25
Figura 9. Úlcera de contacto. Parálisis unilateral y bilateral	26
Figura 10. Ejemplos de parálisis de cuerda vocal izquierda.....	26
Figura 11. Papilomatosis recurrente de cuerda vocal izquierda.....	28
Figura 12. Ejemplo de fibroscopio 1	36
Figura 13. Ejemplo de fibroscopio 2	36
Figura 14. Telelaringoscopia.....	37
Figura 15. Imagen de telelaringoscopia	39
Figura 16. Equipo de uso en Consultorio	40
Figura 17. Secuencia de Frames capturados con estroboscopia laríngea	41
Figura 18. Imagen capturada mediante estroboscopia digital	42
Figura 19. Equipo de Estroboscopia digital	43
Figura 20. Microlaringoscopia	44
Figura 21. Imagen de cuerdas vocales obtenida mediante VKG	46
Figura 22. Imagen de cuerdas vocales. a) original. b) Negativo de la a)	52

Figura 23. Imagen de cuerdas vocales en escala de grises	53
Figura 24. Histograma de la Figura 31	53
Figura 25. Histograma de una imagen en la cual no todos los niveles de grises son utilizados	55
Figura 26. Imagen de las cuerdas vocales a la que se le ha añadido ruido	60
Figura 27. Imagen de las cuerdas vocales tras aplicarle un filtro de promediado de máscara 3x3, todavía se pueden observar restos de espurios	60
Figura 28. Cuerdas vocales tras aplicarle el filtro de la mediana. Se recupera la imagen original, sin restos de ruido	61
Figura 29. Ampliación de imágenes por repetición	63
Figura 30. Modelo unidimensional y continuo de un borde ideal	64
Figura 31. Plantillas para detección de transiciones N, NO, O, SO	66
Figura 32. Plantillas para detección de transiciones S, SE, E, NE	66
Figura 33. Máscaras orientadas de Kirsch para la detección de bordes, en las direcciones E, NE, N y NO	66
Figura 34. Máscaras orientadas de Kirsch para la detección de bordes, en las direcciones O, SO, S y SE	67
Figura 35. a) Imagen Original de Yoda. b) Imagen tras aplicar el operador de Roberts	70
Figura 36. Ejemplo de segmentación realizada aplicando Contornos Activos	76
Figura 37. Aplicación del Filtro de Kalman al seguimiento del movimiento de las cuerdas vocales	77
Figura 38. Diagrama de Capas del Sistema propuesto	84
Figura 39. Estudio secuencial realizado en el diseño de los algoritmos de la capa de Middleware	85
Figura 40. Diagrama de Alto Nivel	86
Figura 41. Ejemplos de imágenes de poca calidad	91
Figura 42. a) Ejemplo de imagen con luz blanquecina. b) Ejemplo de Imagen con buena calidad	92

Figura 43. Diagrama de Bloques de la etapa de Pre-Procesado	93
Figura 44. Bloque A1 de la Etapa de Pre-Procesado	94
Figura 45. Bloque A2 de la etapa de Pre-Procesado	95
Figura 46. a) Imagen Original en formato RGB. b) Imagen en escala de grises. c) Imagen Original con pólipo, formato RGB. d) Imagen con pólipo en escala de grises.	96
Figura 47. Bloque A3 de la Etapa de Pre-Procesado	97
Figura 48. Bloque A4 de la etapa de Pre-Procesado	98
Figura 49. Diagrama de Bloques de la etapa de Segmentación.....	100
Figura 50. Bloque B1 de la Etapa de Segmentación	101
Figura 51. Bloque B2 de la etapa de Segmentación.....	102
Figura 52. Bloques B3 de la etapa de Segmentación	103
Figura 53. Detalle del Proceso iterativo del bloque B3	105
Figura 54. Diagrama de Bloques de la Etapa de Detección de Movimiento	107
Figura 55. Bloque C1 de la etapa de la etapa de Detección de movimiento.....	109
Figura 56. Bloque C2 de la etapa de la etapa de Detección de movimiento.....	109
Figura 57. Bloque C3 de la etapa de la etapa de Detección de movimiento.....	111
Figura 58. Bloque C4 de la etapa de la etapa de Detección de movimiento.....	111
Figura 59. Imagen Ejemplo. Ventana de búsqueda en violeta y Macrobloque en verde.	112
Figura 60. Diagrama etapa D	115
Figura 61. Diagrama de bloques del Clasificador de Patologías Vocales.....	116
Figura 62. Detalle del bloque D1 de la etapa de Clasificación	117
Figura 63. Resultados de los 4 primeros pasos descritos en el bloque D1 de la etapa de Clasificación.....	118
Figura 64. Resultado del módulo de Pre-Procesado D1.1.....	119

Figura 65. Ejemplo de Training Set.....	119
Figura 66. Diagrama de Bloques de la etapa: Clasificación, Análisis y Medida.....	120
Figura 67. Detalle de los módulos de la Etapa E.....	121
Figura 68. Detalle del Bloque E1 de la Etapa de Análisis y Medida	122
Figura 69. Detalle del Bloque E 1.2	124
Figura 70. a) Contorno del espacio glótico. b) Contornos interiores de las cuerdas vocales. Derecha (azul) e izquierda (rojo).....	125
Figura 71. Detalle del bloque E1.4.....	126
Figura 72. Cuerdas vocales sanas. Original, filtrada y resultado de derivada..	127
Figura 73. Cuerdas vocales patológicas. Original, filtrada y resultado de la derivada.	127
Figura 74. a) Ejemplo de cálculos de la segunda derivada. b) Intersección con la recta de regresión	128
Figura 75. Patologías localizadas y resaltadas en la imagen original	129
Figura 76. Detalle del Bloque E2.	130
Figura 77. Cuerdas vocales en las que se muestran vectores de movimiento...131	
Figura 78. Ejemplos de medida de ángulos de cuerda vocal a)derecha y b) izquierda.....	132
Figura 79. Diagrama de la Decisión/diagnóstico	133
Figura 80. Frame 13 de secuencia de cuerdas vocales sanas. a) Original. b) Resultado tras los filtros de Gabor. c) Resultado tras denoising. d) Máscara resultado del espacio glótico.....	138
Figura 81. Frame 56 de secuencia de cuerdas vocales sanas. a) Original. b) Resultado tras los filtros de Gabor. c) Resultado tras denoising. d) Máscara errónea del espacio glótico. e) Máscara resultado del espacio glótico	139
Figura 82. a) Imagen Original de Cuerdas Vocales con Nódulos. b) ROI de a). c) Imagen Original de cuerdas vocales con pólipo. d) ROI de c)	140
Figura 83. Ejemplo de Secuencia de cuerdas vocales sanas, Frame 60 al Frame 69. Imagen Original y ROI.	143

Figura 84. Representación gráfica de los resultados de segmentación	146
Figura 85. Ejemplo de Training Set con 25 imágenes de cuerdas vocales derecha e izquierda independientemente	151
Figura 86. Vectores de movimiento en a) una secuencia de cuerdas vocales con parálisis y b) una secuencia de cuerdas vocales sanas	153
Figura 87. Glottal Area Waveform de una secuencia de cuerdas vocales sanas	158
Figura 88. Imagen segmentada de cuerdas vocales con pólipo con contorno resaltado	159
Figura 89. Resultado gráfico del cálculo del área respecto a estimación del contorno de cuerda sana	159
Figura 90. Representación gráfica de los resultados de la secuencia Sana 1	167
Figura 91. Representación gráfica de los resultados de la secuencia Sana 2	168
Figura 92. Representación gráfica de resultados de secuencia Parálisis 39	168
Figura 93. Representación gráfica de resultados de secuencia Parálisis 40	169
Figura 94. Representación gráfica de resultados de secuencia Nódulos 11	170
Figura 95. Representación gráfica reducida de los resultados de la secuencia Sana 1	170
Figura 96. Representación gráfica reducida de los resultados de la secuencia Sana 2	171
Figura 97. Representación gráfica reducida de los resultados de la secuencia Parálisis 39	172
Figura 98. Representación gráfica reducida de los resultados de la secuencia Parálisis 40	172
Figura 99. Representación gráfica reducida de los resultados de la secuencia Quiste 35	173
Figura 100. Resultados del diagnóstico en función de la morfología	174
Figura 101. Resultados del diagnóstico en función del movimiento	175

Índice de Tablas

Tabla 1. Dimensiones medias de la laringe en el adulto europeo.....	16
Tabla 2. Comparativa entre Videoestroboscopia y Videoquimografía.....	31
Tabla 3. Etiquetado de las secuencias de la Base de Datos	81
Tabla 4. Especificaciones Técnicas del Hardware	137
Tabla 5. Resultados de segmentación por secuencia	145
Tabla 6. Experimento de Identificación. Resultados en %	148
Tabla 7. Experimento de Clasificación. Resultado de las Eigenfolds en %	150
Tabla 8. Resultados de Eigenfolds usando un Training Set de cuerdas vocales tratando de forma independiente la cuerda vocal derecha e izquierda	152
Tabla 9. Resultados de Eigenfolds tratando de forma independiente la cuerda vocal derecha e izquierda	152
Tabla 10. Cálculos estadísticos realizados sobre secuencias de cuerdas vocales sanas.....	155
Tabla 11. Cálculos estadísticos realizados sobre secuencias de cuerdas vocales con parálisis	157
Tabla 12. Medidas de área de la patología de cada cuerda vocal (cuerdas morfológicamente patológicas).....	161
Tabla 13. Medidas de área de la patología de cada cuerda vocal (cuerdas morfológicamente sanas)	163
Tabla 14. Resultados de las medidas de ángulos de secuencias con parálisis .	164
Tabla 15. Resultados de las medidas de ángulos de las secuencias sanas.....	165
Tabla 16. Resultados del análisis del movimiento según (61)	165
Tabla 17. Comunicaciones en Congresos	184
Tabla 18. Publicaciones en Revistas	184
Tabla 19. Publicación en libro	184

“La imagen es uno de los principales medios de interpretación y por ello su importancia es cada día mayor”.

Berenice Abbott

1.INTRODUCCIÓN

La presente tesis describe el trabajo realizado en el campo del procesado digital de imagen aplicado a la ingeniería biomédica. En concreto, la aportación se centra en la especialidad de otorrinolaringología.

Son de sobra conocidos la multitud de avances para mejorar la instrumentación y el diagnóstico médico en cualquier área de la medicina, tanto a nivel hardware como software [Zaidi+09] [Lonsdale+10], y más concretamente, el diagnóstico asistido por ordenador (CAD). Sin embargo, el otorrinolaringólogo es uno de los especialistas en cuya actividad diaria interviene ampliamente la *subjetividad* y su experiencia previa, en la emisión de los diagnósticos. A pesar de ser las patologías vocales comunes causa de amplia baja laboral [Ibermutuatur], actualmente no se dispone de mecanismos para *caracterizar* el funcionamiento de las *cuerdas vocales* de forma objetiva.

La laringe y las cuerdas vocales con su vibración son los instrumentos básicos en la emisión de la voz. Cualquier alteración en ellas, además de los problemas de salud, puede tener como consecuencia un problema de comunicación.

Por todas estas razones, en esta tesis se ha trabajado en el diseño y la aplicación de nuevos algoritmos de procesamiento digital de imagen, que implementados en un software, proporcionan al médico un diagnóstico automatizado previo (es importante reseñar que en ningún momento se pretende prescindir del especialista) y todas las herramientas y datos necesarios para que el diagnóstico final sea lo más riguroso y exhaustivo posible.

El primer problema a abordar es la falta de unificación y normalización en las imágenes capturadas, ya que se han estudiado videos e imágenes de cuerdas vocales procedentes de diferente instrumentación, hospital y especialista, por lo tanto ha sido necesario aplicar un pre-procesamiento previo sobre características tales como: brillo, resolución, color, etc.

Se han diseñado varias etapas de procesamiento tomando como base de algunas de ellas, algoritmos de análisis de texturas como los filtros de Gabor, técnicas estadísticas como Principal Component Analysis (PCA), o técnicas de Block Matching para el análisis del movimiento, cuya combinación y aplicación con otros desarrollos propios, ha dado como resultado el diagnóstico de las patologías vocales benignas más comunes:

- Patologías relacionadas con el **movimiento**, principalmente en este campo se puede encontrar la parálisis de cuerda vocal unilateral o bilateral.
- Patologías relacionadas con la **morfología**. Entre las más conocidas se pueden destacar: nódulos, pólipos, quistes, edema, etc.

Además, se han estudiado todos los test aprobados por la Sociedad Europea de Laringología [ELS] para el análisis de la videoestroboscopia y se han tomado

como referencia los parámetros aceptados por la comunidad médica. Por otra parte, se proponen nuevos parámetros para caracterizar el funcionamiento de las cuerdas vocales.

Todos los resultados obtenidos (sean imágenes o resultados numéricos) han sido contrastados con los diagnósticos y las opiniones de varios especialistas en otorrinolaringología.

1.1 JUSTIFICACIÓN DE LA INVESTIGACIÓN

En los últimos años, el estudio de las patologías de la voz ha despertado el interés de multitud de profesionales de este campo como logopedas, foniatras, o médicos especialistas en otorrinolaringología. Y son éstos últimos los que capturan y analizan las imágenes y vídeos de cuerdas vocales, objeto de este estudio.

Se plantea la pregunta ¿por qué estudiar las cuerdas vocales? La respuesta es clara, **la voz** se produce gracias a la **vibración de las cuerdas** al acercarse entre sí como consecuencia del paso del aire a través de la laringe [Titze08].

La voz es nuestro medio de comunicación, y los números son claros en este campo, los especialistas cada vez tienen en sus consultas más pacientes que sufren algún trastorno de la voz. El 30-40% de la población infantil en edad escolar sufren disfonías [SEORL], de las que entre el 60-80% está causada por nódulos.

Según la Sociedad Española de Otorrinolaringología (SEORL), alrededor del 5% de la población española [SEORL] sufre algún trastorno de voz que requiere atención por parte de un otorrinolaringólogo. La edad de mayor prevalencia es entre los 25 y 45 años. Además, el 22% de los docentes presentan alteraciones de la voz, siendo la segunda causa más frecuente de baja laboral en este grupo profesional. [Ibermutuatur].

La mayoría de los casos son disfunciones leves provocadas por el ritmo y los cambios producidos en nuestra sociedad. Pero hay otros, en los que el diagnóstico resulta complicado, y se ha conseguido llegar a él, gracias a los avances tecnológicos que han permitido crear instrumentación de gran precisión, dando como resultado la digitalización de las imágenes adquiridas (objeto de estudio).

En algunos casos, esta instrumentación lleva asociada casi siempre un software de análisis, que en realidad son aplicativos que contienen algoritmos matemáticos avanzados capaces de precisar características de la voz o las cuerdas vocales impensables en la época de Manuel García en 1854 [Fernandez+06] cuando la exploración se realizaba mediante la laringoscopia indirecta con el espejillo. Por todo ello, son varios los hospitales que hoy en día cuentan con una unidad de voz con software y hardware especializado [Casado+10].

Llegados a este punto, dentro del estado del arte se pueden distinguir dos apartados sobre los que profundizar. Por un lado, es importante enmarcar esta tesis dentro de su ámbito de aplicación: las cuerdas vocales (descripción anatómica, funcionamiento y patologías más comunes), las técnicas de captura y las posibilidades software existentes en el mercado. Y por otro, ahondar en las técnicas de procesamiento digital de imagen.

1.2 HIPÓTESIS

Con el fin de delimitar y concretar el ámbito de acción de esta tesis, en este apartado se presenta la hipótesis de partida.

A continuación, se enuncia la hipótesis que promueve toda la investigación:

No existe ningún software de evaluación objetiva de la voz en el mercado, que incluyendo el análisis de imágenes estroboscópicas, proporcione de forma automatizada, medidas objetivas sobre éstas, que avalen el diagnóstico asistido por ordenador de patologías vocales a través de algoritmos de procesado digital de imagen.

1.3 OBJETIVOS

En este punto, se enuncia el objetivo principal de esta tesis y los objetivos específicos que subyacen, todos ellos marcados al comienzo de este trabajo para que pudiese ser culminado con éxito

*El objetivo principal de esta tesis es el diseño y **desarrollo de los algoritmos** de procesado digital de imagen necesarios, para la extracción de parámetros objetivos a partir de videos e imágenes estroboscópicas de cuerdas vocales sanas y patológicas, sin la interacción del usuario, en este caso, del médico especialista.*

Evidentemente, con esta aplicación no se pretende sustituir al médico en su práctica diaria, simplemente proporcionarle las herramientas necesarias para avalar sus diagnósticos mediante datos objetivos. Por todo ello, se han definido los siguientes objetivos secundarios:

- O1. Crear una completa **base de datos** de videos estroboscópicos e imágenes estáticas de cuerdas vocales sanas y patológicas. Para ello, se ha contado con una base de datos comercial [LeBorgne06] con 15 registros, de ellos 7 son videos de cuerdas vocales sanas, y 8 videos de cuerdas vocales con diferentes patologías tales como: nódulos, pólipos, quiste, parálisis... Además, el especialista en otorrinolaringología

Agustín Pérez Izquierdo del Hospital de Basurto, ha aportado 30 grabaciones realizadas por él mismo para completar la base de datos.

- O2. Especificar las patologías de las cuerdas vocales objeto de estudio. Los desarrollos se realizarán sobre un número acotado de patologías, centrándose en las más comunes, objeto de numerosas consultas médicas y bajas laborales de los profesionales de la voz.
- O3. Realizar un completo estado del arte de los software de voz comerciales, así como las técnicas de procesamiento digital de imagen implicadas en esta tesis.
- O4. Diseñar un **algoritmo** matemático que contenga técnicas avanzadas de procesamiento digital de imagen, para poder satisfacer la medida de los parámetros anteriormente definidos y emitir una valoración avalada por ellos. En este punto se tendrán que satisfacer las siguientes necesidades:
 - ❖ Tener la capacidad de procesar vídeos y secuencias completas de frames (no sólo imágenes estáticas) con el objetivo de obtener información relativa al **movimiento** (en las patologías en las que aporte información crítica para el diagnóstico, como la parálisis unilateral o bilateral de cuerdas vocales).
 - ❖ Estudiar los distintos algoritmos de **segmentación**. Centrar el esfuerzo en la obtención de una óptima segmentación del espacio glótico. De ahora en adelante, en el documento el espacio glótico será nuestra región de interés (ROI).
 - ❖ Optimizar el tiempo procesado. Para ello se utilizarán técnicas estadísticas que definirán un pre-diagnóstico, y en función de él, los parámetros objetivos mensurables en cada caso.
 - ❖ Realizar las transformaciones geométricas necesarias para la

realización de medidas objetivas.

- O5. Desarrollar en software, todos los algoritmos diseñados y que nos permiten el diagnóstico/detección/evaluación de patologías vocales benignas.
- O6. Elaborar un exhaustivo plan de evaluación de los resultados de todo lo descrito anteriormente. Para ello, todos los diagnósticos automatizados se contrastarán con los emitidos por los especialistas, se realizarán análisis estadísticos de las medidas.
- O7. Definir líneas abiertas de futuro

1.4 METODOLOGÍA

En este apartado se expone el proceso de investigación llevado a cabo y el método de trabajo para cumplir los objetivos previamente descritos. La investigación se compone de cinco etapas, típicas en este tipo de trabajos (ver Figura 1).

Etapa 1. *Identificación del problema.* Durante esta etapa se estudia el estado del arte con el objetivo de detectar las necesidades y limitaciones actuales entorno al área de conocimiento de esta tesis, el procesado digital de imagen. Toda la documentación recopilada durante esta fase ha sido organizada y estructurada en dos grandes apartados de forma que permita su cómoda consulta. Por un lado la base teórica sobre la que se sustenta esta tesis y por otra, los trabajos relacionados analizados.

Etapa 2. *Enunciado de la Hipótesis.* Una vez detectada la oportunidad de la tesis, se plantea la hipótesis, el problema de partida y la contextualización, que encamina la investigación hacia la solución. En el caso de esta tesis, se puede enunciar el problema como:

La búsqueda de parámetros que ayuden a objetivar el diagnóstico de patologías benignas de cuerdas vocales a través de algoritmos de procesamiento digital de imagen implementados en el ordenador.

Etapas 3. *Planteamiento de la solución.* Llegados a este punto se propone una solución para dar respuesta al problema.



Figura 1. Etapas de la Investigación

Etapas 4. *Experimentación.* Durante esta etapa se recogen y analizan los datos que pueden ser indicadores del éxito de este trabajo, y que además, puedan plantear nuevos objetivos y retos futuros.

Etapas 5. Es la *Validación de la Hipótesis*, el resultado final de esta investigación.

El método de este trabajo se encuentra basado principalmente en el método *investigación-acción* [French+96] [Wadsworth98], aplicándose en la definición, refinamiento y validación de la solución planteada, aunque también se ha

utilizado el método basado en *experimentos de laboratorio* [Straub+04] [Boudreau+01].

1.4.1 Elección de los participantes de la investigación

Los tipos de participantes definidos según [Wadsworth98] se reflejan en esta tesis según los siguientes roles:

- **Investigador.** La persona que asume este rol es Amaia Méndez, apoyada por todos los componentes de la unidad DeustoTech-LIFE del Instituto Tecnológico de la Universidad de Deusto y el PhD. Agustín Pérez Izquierdo del Hospital de Basurto.
- **Objeto investigado.** Para cumplir los objetivos marcados, el objeto de esta investigación son los videos y las imágenes estroboscópicas de cuerdas vocales para concluir definiendo un algoritmo que nos permita distinguir entre patologías, y ayude a los especialistas en otorrinolaringología en su práctica diaria y a la emisión de diagnósticos más concretos y detallados, nutridos por parámetros objetivos.
- **Grupo Crítico de referencia.** Para realizar las pruebas finales y la validación del algoritmo se ha contado con un grupo de otorrinolaringólogos, entre los que cabe destacar el Doctor Agustín Pérez Izquierdo del Hospital de Basurto, que ha colaborado activamente aportando grabaciones de sesiones realizadas con sus pacientes y evaluando médicamente los resultados durante toda la investigación.
- **Beneficiarios.** Serán todos aquellos interesados o beneficiados por los resultados de esta tesis. En primer lugar tendremos a la Comunidad Médica especializada que podrá aportar más datos a sus informes y añadir a sus diagnósticos objetividad. Por otro lado, todas las empresas desarrolladoras de software médico, que podrían incluir estos algoritmos a sus próximas implementaciones. Y por último, la Comunidad

Científica que ha apoyado esta investigación a lo largo de todo su desarrollo, permitiendo su difusión.

1.5 ESTRUCTURA DE LA TESIS

En esta sección se describe brevemente la estructura del presente documento, resumiéndose el contenido de cada capítulo.

Esta disertación está estructurada en seis capítulos y dos anexos, de acuerdo a lo que se describe a continuación. El presente capítulo enmarca el contexto de la tesis. En él se justifica el estudio en relación a la hipótesis establecida y se exponen los objetivos generales y específicos de la disertación.

El capítulo 2 se muestra un recorrido por el estado del arte actual en varios aspectos: por un lado, la descripción anatómica de las cuerdas vocales, su funcionamiento y descripción de sus patologías más comunes, y de las técnicas de captura de videos e imágenes digitales de cuerdas vocales, así como de los software y herramientas disponibles en el mercado para la ayuda al diagnóstico de patologías vocales. Por otro lado, se establece un marco teórico de las técnicas más comunes de procesamiento digital de imagen.

El capítulo 3 describe la base de datos utilizada para el diseño y validación de los algoritmos propuestos en esta tesis.

En el capítulo 4 se propone el diseño de una metodología para el diagnóstico de patologías vocales con técnicas de procesamiento de señal a partir de videos estroboscópicos. Se explica a alto y bajo nivel el diseño de la solución propuesta, detallando los algoritmos y la aportación realizada.

El capítulo 5 evalúa la calidad de los algoritmos y funciones desarrolladas en esta tesis. En esta sección se detallan los resultados obtenidos durante toda la investigación, indicando los resultados parciales de cada uno de los bloques del sistema propuesto, y los resultados globales. Además, para validar los

resultados se ha contado en todo momento con el apoyo del especialista en otorrinolaringología.

Finalmente, el capítulo 6 está dedicado a las conclusiones y aportaciones generales derivadas de la investigación desarrollada en esta tesis doctoral. Además, para finalizar en este mismo capítulo se enumeran algunas de las líneas de investigación que quedan abiertas y que se pretenden cubrir en trabajos futuros.

El glosario de términos y aspectos que son complementarios para ilustrar los resultados y el diseño realizado en esta tesis, se presentan en los anexos.

“...para mi gran alegría, la glotis abierta ante mí, y tan bien, que incluso pude percibir una porción de la tráquea. Después, ya más tranquilo, comencé a observar lo que pasaba ante mis ojos: la manera como la glotis se abría y cerraba y como se movía en el acto de la fonación”.

Manuel García

2. ESTADO DEL ARTE

En este capítulo se establecen los principios y características de las cuerdas vocales, se describen las peculiaridades médicas de sus patologías, y se exponen las principales estrategias diagnósticas actualmente implantadas. Todo ello, con el objeto de posicionar en este contexto el sistema propuesto. Finalmente, se ofrece una perspectiva del estado del arte y una descripción de los métodos y técnicas de procesamiento empleados con carácter general.

Las patologías relacionadas con problemas en las cuerdas vocales, están adquiriendo cada vez más importancia en nuestra sociedad debido a las demandas crecientes de las profesiones que utilizan la voz como herramienta de trabajo.

La Otorrinolaringología es la especialidad responsable del abordaje de estas patologías en la que además intervienen otros profesionales médicos y no médicos como foniatras, logopedas, profesores de canto, etc.

Los avances en la tecnología audiovisual, informática, y telecomunicaciones han supuesto una mejora en los medios diagnósticos y terapéuticos de la patología vocal. Este rápido progreso, hace que las técnicas utilizadas en la patología laríngea, se queden obsoletas en breves periodos de tiempo.

Los avances sin duda más significativos han sido aquellas pruebas diagnósticas basadas en la imagen de la patología laríngea. La obtención de imágenes de mayor calidad, junto con los sistemas de digitalización, ampliación y edición de los fotogramas facilita enormemente la valoración de las distintas patologías.

El movimiento laríngeo ondulatorio imperceptible al ojo humano en condiciones normales, hoy en día no es problema, puesto que el perfeccionamiento en la captura y procesado de la imagen dinámica de la Estroboscopia, la fotografía ultrarrápida y videoquimografía (VKG), ha supuesto el diagnóstico de patologías que antes pasaban desapercibidas [Johns03], y la aplicación de técnicas de procesado digital de imagen a este campo de la medicina.

Los hallazgos laringoestroboscópicos son sin duda los más importantes y clarificadores en la patología vocal, puesto que genera imágenes ralentizadas del movimiento ondulatorio de la mucosa vocal, mediante una luz intermitente en desfase con la onda vocal.

De estas imágenes se pueden extraer datos relacionados con el cierre glótico, regularidad, simetría y amplitud de la onda mucosa. Sin embargo, la valoración de estos parámetros está expuesta a una interpretación subjetiva, en la cual entran en juego factores como la experiencia del foniatra u otorrinolaringólogo, tanto en la interpretación como en la captura de las imágenes.

El objetivo de esta tesis está centrado en el procesado y análisis digital de las imágenes que se obtienen en la laringoestroboscopia (el método más utilizado en España), que permita crear una parametrización de las cuerdas vocales para

apoyar el diagnóstico, y por lo tanto apoyar y objetivar la interpretación visual del médico especialista.

2.1 ANTECEDENTES MÉDICOS

La voz humana está producida en la laringe, cuya parte esencial, la glotis, constituye el verdadero órgano de fonación humano. El aire procedente de los pulmones es forzado durante la espiración a través de la glotis, haciendo vibrar los dos pares de cuerdas vocales, que se asemejan a dos lengüetas dobles membranáceas. Las cavidades de la cabeza, relacionadas con el sistema respiratorio y nasofaríngeo, actúan como resonadores [Felsa+10].

2.1.1 Descripción del aparato de fonación

El aparato de fonación está compuesto por:

- Laringe
- Cuerdas vocales

La laringe es un órgano de la fonación, un conducto aéreo y un mecanismo esfinteriano que se extiende desde el suelo de la lengua hasta la tráquea [Gray18]. Se proyecta ventralmente entre los grandes vasos del cuello y está cubierta anteriormente por la piel, las aponeurosis y los músculos depresores del hioides. Cranealmente se abre en la laringofaringe, de la cuál forma su pared anterior y caudalmente se continúa con la tráquea. En el hombre adulto se encuentra frente a las III, IV, V y VI vértebras cervicales, siendo esta porción algo más craneal en el niño y en la mujer adulta. En niños entre seis y doce meses el extremo superior de la epiglotis (la parte más craneal de la laringe) se encuentra a un nivel ligeramente superior al disco cartilaginoso existente entre la apófisis odontoides y el cuerpo del axis. Las dimensiones medias en el adulto europeo se pueden ver en la Tabla 1.

	Hombre	Mujer
Longitud	44 mm	36 mm
Diámetro Transverso	43 mm	41 mm
Diámetro anteroposterior	36 mm	26 mm

Tabla 1. Dimensiones medias de la laringe en el adulto europeo

La laringe cumple funciones en:

- La respiración
- Deglución
- Protección de la vía aérea
- Vocalización
- Defecación
- El levantamiento de objetos pesados.

Está compuesta básicamente por cartílagos y músculos, ambos cubiertos por mucosa (ver Figura 2). Sus estructuras conforman un sistema de válvula a diferentes niveles, y funciona creando un esfínter localizado en la unión entre el tracto respiratorio superior y la tráquea.

Las cuerdas vocales (en ocasiones también llamadas, pliegues vocales) se componen de los denominados pliegues verdaderos y pliegues falsos, ambos pares. Estos, junto con la **epiglotis** conforman el esfínter, el cual se abre para permitir el paso del aire durante la respiración (ver Figura 3a y 3c), y se cierra para permitir la fonación (ver Figura 3b y 3d), el paso de alimentos al esófago durante la deglución, y el atrapamiento de aire en los pulmones en ciertas actividades.

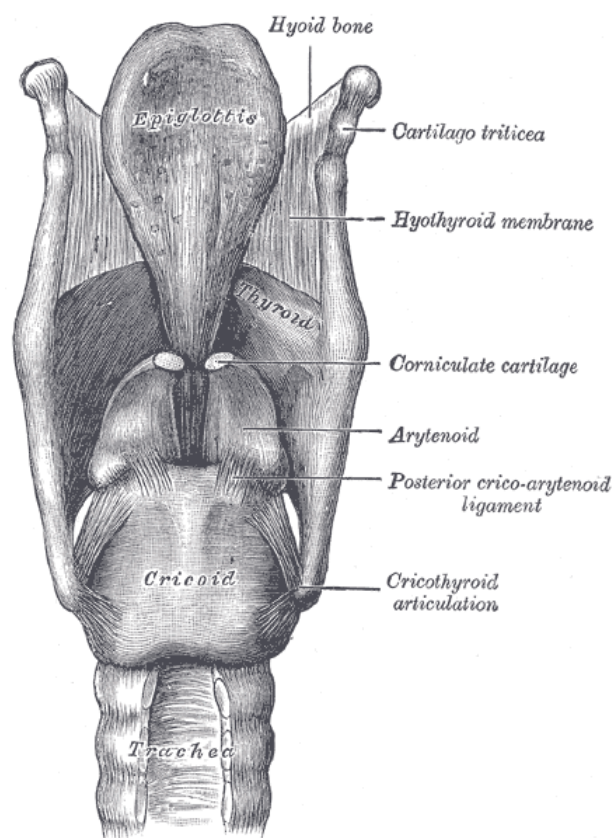


Figura 2. Ligamentos de la laringe. Visión Posterior [Gray 18].

Dado que son las cuerdas vocales el centro de la presente investigación, se detallan sus características y funcionamiento en el siguiente apartado.

2.1.2 Descripción anatómica de las cuerdas vocales

La cavidad laríngea se extiende desde la entrada de la laringe, por medio de la cual se comunica con la faringe, hasta el borde inferior del cartílago cricoides, a cuyo nivel se continúa con la cavidad traqueal. Está dividida en tres porciones por dos pares de pliegues mucosos, superior e inferior, que se proyectan desde las partes laterales de la cavidad hacia su interior. Los pliegues superiores se llaman **cuerdas vocales superiores**, quedando entre ellos fisura denominada hendidura del vestíbulo. Los inferiores están involucrados en la producción de la voz, por lo que se les denomina **cuerdas vocales inferiores**, y a la fisura existente entre ellos hendidura glótica, frecuentemente denominada **glotis** (ver Figura 3).

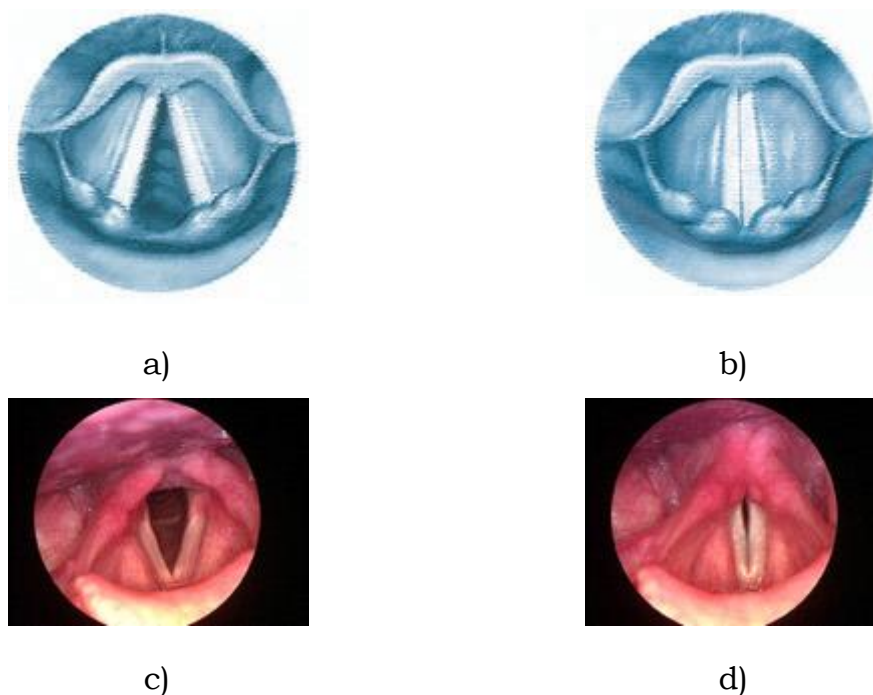


Figura 3. a) Diagrama de cuerdas vocales abiertas o en abducción. b) Cuerdas vocales cerradas o en aducción. c) En abducción durante una inspiración forzada. d) En aducción durante la fonación

Las **cuerdas vocales superiores** (falsas) son dos gruesos pliegues mucosos sonrosados que envuelven cada uno una estrecha banda de tejido fibroso, llamada ligamento vestibular, el cual se inserta por delante en el ángulo del cartílago tiroides inmediatamente por debajo del punto de fijación del cartílago epiglótico y por detrás en la cara anterolateral del cartílago aritenoides, un poco por encima de la apófisis vocal.

Las **cuerdas vocales inferiores** (verdaderas) son dos marcados pliegues mucosos blanquecinos que se extienden desde el centro del ángulo del cartílago tiroides hacia la apófisis vocal de los cartílagos aritenoides. Constituyen los límites laterales de la hendidura glótica en su parte anterior e intervienen en la producción de la voz. La cuerda vocal inferior está recubierta por un epitelio plano estratificado que se adhiere al ligamento vocal subyacente. Como consecuencia de la falta de la capa submucosa y vasos sanguíneos, la cuerda vocal inferior tiene una **coloración blanca nacarada**.

La hendidura glótica o glotis es una fisura situada entre las cuerdas vocales inferiores en su parte anterior y las bases y las apófisis vocales de los cartílagos aritenoides en su parte posterior. Su límite dorsal está constituido por la mucosa existente entre ambos cartílagos aritenoides a la altura de las cuerdas vocales inferiores. La porción delimitada por las cuerdas vocales inferiores se denomina glotis interligamentosa y representa las tres quintas partes de la longitud de toda la glotis. La longitud media de la glotis en el hombre adulto es de 23 mm y en la mujer de 17 mm. La glotis es la parte más estrecha de la laringe, y su anchura y forma varían con los movimientos de las cuerdas vocales inferiores y los cartílagos aritenoides durante la respiración y la fonación.

La porción inferior de la cavidad laríngea se extiende desde las cuerdas vocales inferiores hasta el borde inferior del cartílago cricoides. Su parte posterior es elíptica, mientras que la inferior, más ancha, se hace circular y se continúa con la cavidad traqueal. Sus paredes están formadas, en su parte superior, por el ligamento cricotiroides y, en la inferior, por la superficie interna del cartílago cricoides, recubiertas ambas por mucosa [Gray18].

2.1.3 Trastornos de las cuerdas vocales

La voz es un recurso extremadamente valioso para el ser humano: es el medio de comunicación más utilizado, una herramienta de trabajo para numerosos profesionales y una forma de expresión estética ante los demás.

Se estima que en España un 2% de la población sufre algún trastorno de la voz que requiere la atención de un especialista [Arbos+05]. En concreto, los profesionales de la voz (locutores, docentes, animadores, periodistas...) son los que consultan más rápidamente y los más afectados por los trastornos de la voz. De hecho suponen **la segunda causa más frecuente de baja laboral en este grupo profesional** [ISC10]. Asimismo, las mujeres sufren más trastornos de la voz que la población masculina.

Cuando hay mal uso de la voz se producen ciertas patologías vocales, las más frecuentes son las disfonías que provocan la pérdida parcial de la voz. Dentro de ellas se pueden encontrar las funcionales, cuando no hay daño en el órgano que produce la voz (laringe), y las orgánicas cuando hay daño demostrable.

Las **disfonías orgánicas** de base funcional son una de las más comunes, éstas pueden desarrollar nódulos, pólipos, quistes de retención, hemorragias, y son patologías que se producen por el mal uso vocal. Algunos de estos casos mejoran con la rehabilitación, con el reposo de la voz, con higiene vocal y ejercicios. Hay otros, sin embargo, que deben operarse porque los pacientes consultan demasiado tarde y un nódulo de seis u ocho meses necesita cirugía.

Las **disfonías emocionales**, en tanto, no presentan daño en el órgano estructural, pero la persona como una manera de protegerse del estrés genera una disfonía, situación que se ha podido apreciar más frecuentemente durante los últimos años.

Síntomas de una enfermedad de la voz pueden ser desde ronquera o sequedad crónica, pasando por la sensación de carraspera en la garganta, un tono de voz desagradable, limitaciones para hablar claramente, o simplemente periodos de pérdida de voz.

Entre las patologías más comúnmente encontradas están:

- Mal uso o abuso vocal
- Laringitis
- Reflujo de ácido del estómago
- Nódulos en las cuerdas vocales. [Serrano+10].
- Pólipos
- Quistes
- Parálisis de las cuerdas vocales
- Cáncer de laringe
- Disfonía espasmódica (intermitente)

- Úlceras de contacto
- Curvatura del pliegue vocal
- Papilomatosis laríngea
- Etc

A continuación, en la Figura 4, se observa cuál es el ciclo vibratorio normal de unas cuerdas vocales sin ninguna patología:

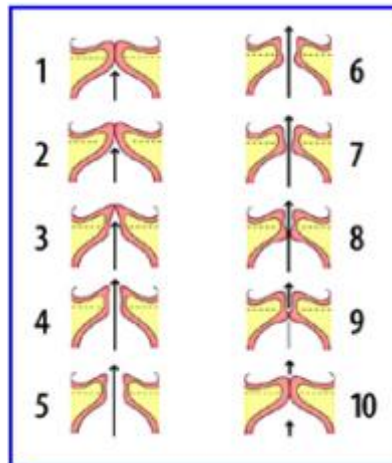


Figura 4. Funcionamiento normal de las cuerdas vocales

1. La presión del aire mueve hacia arriba las cuerdas vocales en posición cerrada.
- 2,3. La presión del aire abre las capas inferiores de vibración de las cuerdas vocales.
- 4,5. La presión del aire continúa hacia arriba, ahora las cuerdas vocales ya se han abierto.
- 6-10. La baja presión creada tras el rápido movimiento del aire, produce el efecto Bernoulli que causa el cierre de la parte inferior y después de la superior.
10. Cierre de las cuerdas vocales.

En la Figura 5 se puede ver cómo una disfunción (o imperfección) en las cuerdas vocales (no importa lo pequeña que sea) puede alterar su ciclo de vibración:

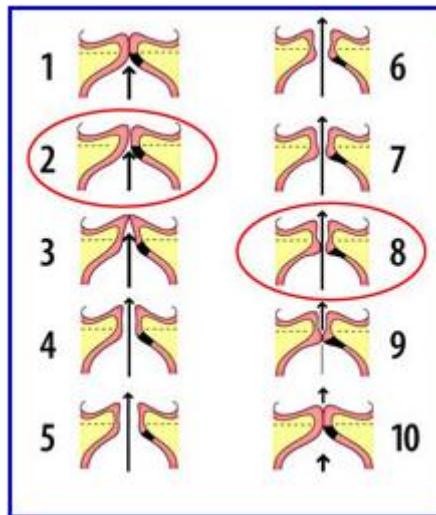


Figura 5. Funcionamiento anormal de las cuerdas vocales

1. Debido a que una de las cuerdas vocales tiene una imperfección, se incrementa la cantidad de aire que se necesita para abrir las cuerdas vocales
8. Por la misma razón, se decrementa la eficiencia del efecto Bernoulli para cerrar el pliegue vocal.

2.1.3.1 Laringitis

La laringitis es uno de los trastornos más frecuentes. Se trata de la inflamación de las cuerdas vocales, que generalmente tiene su origen en una infección viral de las vías aéreas altas. Cuando las cuerdas vocales se inflaman, vibran de manera diferente y producen ronquera. El mejor tratamiento en este caso es el reposo de la voz y una buena hidratación. Asimismo es importante extremar los cuidados ya que en las cuerdas vocales inflamadas aumenta el riesgo de padecer lesiones importantes o un daño permanente.

Los sistemas basados en el procesamiento digital de imagen son esenciales para avanzar en la prevención y detección precoz de estas lesiones.

2.1.3.2 Reflujo de ácido del estómago

El reflujo de ácido del estómago en la garganta puede dar lugar a diversos síntomas en la garganta, como ronquera, problemas al tragar y dolor de garganta. Es un trastorno de difícil diagnóstico ya que la mitad de los pacientes que padecen este reflujo laringofaríngeo no tienen ardor, un síntoma que tradicionalmente acompaña al reflujo gastroesofágico. Por su parte, el reflujo laringofaríngeo de presentación nocturna es el más difícil de diagnosticar ya que no cursa con los síntomas habituales, sino que presenta con irritación de garganta y ronquera al despertar.

2.1.3.3 Nódulos en las cuerdas vocales

Los nódulos vocales son **crecimientos benignos** (no cancerosos) sobre las cuerdas vocales producidos por el abuso de la voz. Los nódulos vocales son un problema frecuente de los cantantes profesionales. Los nódulos son pequeños y tienen forma de callos y **generalmente crecen en pares (uno en cada cuerda vocal)**, como se puede observar en la Figura 6.

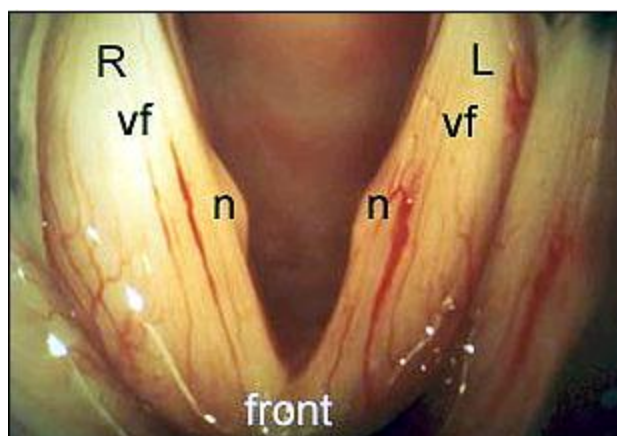


Figura 6. Cuerdas vocales con nódulos

Los nódulos normalmente se forman en las áreas de las cuerdas vocales que reciben la mayor parte de la presión cuando las cuerdas vocales se juntan y vibran (similar a la formación de un callo). El resultado es una voz ronca que se

fatiga fácilmente. Los nódulos se consideran benignos y con un diagnóstico temprano, responden bien al tratamiento.

En concreto, los nódulos desde 2003 han sido considerados como enfermedad profesional para docentes por la Comunidad Europea [UE03], y se han incluido en el cuadro de enfermedades profesionales en España en 2007 [Gañet+07].

Gracias a las características fisiológicas de los nódulos de las cuerdas vocales, se puede concluir que sería relativamente sencillo realizar un modelado de esta patología, a través de las imágenes obtenidas con instrumentos tales como el fibroscopio (del que se hablará posteriormente).

2.1.3.4 Pólipos

Un pólipo vocal es un **crecimiento blando, benigno** (no canceroso), parecido a una ampolla. Un pólipo normalmente **crece solo sobre una cuerda vocal** (principal diferencia con los nódulos, tal y como se puede apreciar en la Figura 7) y se produce a menudo por el consumo de tabaco durante un tiempo prolongado. Otras causas de los pólipos vocales incluyen hipotiroidismo (glándula tiroides hipofuncionante), reflujo gastroesofágico y mal uso continuo de la voz.

Dan lugar a voz ronca, a períodos de pérdida ocasional de voz y a sensación de fatiga excesiva de la voz durante períodos de habla prolongados.

De acuerdo con los resultados del examen estroboscópico, pueden ser tratados con terapia de la voz o con intervención médica. Una vez decidido el tratamiento **es imposible medir de manera objetiva la evolución del paciente**. El médico dispone de los instrumentos necesarios para obtener las imágenes, pero sólo dispone de su visión subjetiva y experiencia para concluir si el paciente está evolucionando correctamente o no.

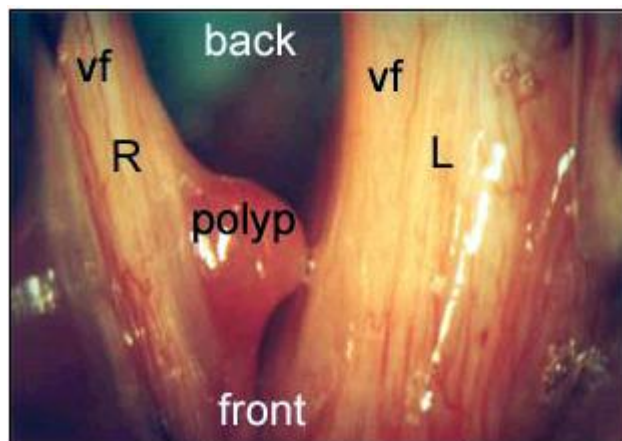


Figura 7. Cuerdas vocales con pólipos

2.1.3.5 Quistes

Inicialmente aparecen como pequeñas protuberancias en las cuerdas vocales. La voz comenzará a ser ronca y a menudo con un tono más bajo. A lo largo del tiempo, los quistes van aumentando de tamaño, causando un crecimiento de la rigidez de la cuerda vocal y aumentando problemas de la voz. **Un quiste se puede confundir con un nódulo** (ver Figura 8). Una de las maneras más confiables de **identificar un quiste** (y a menudo la única manera) es **el examen estroboscópico**.

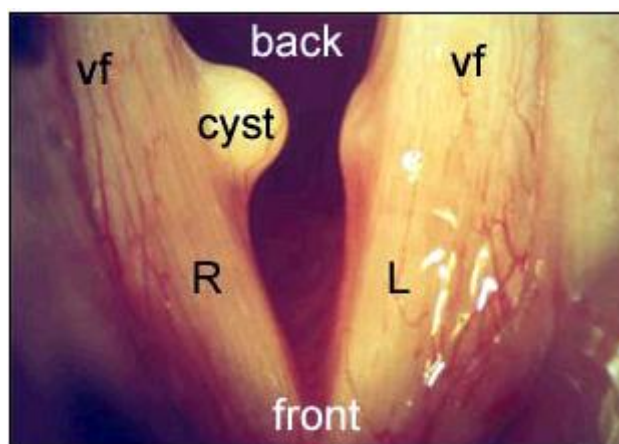


Figura 8. Quiste en cuerda vocal derecha

2.1.3.6 Parálisis de las cuerdas vocales

La parálisis (ver ejemplos en la Figura 9 y en la Figura 10) se produce cuando una de las cuerdas vocales (o ambas) deja de moverse. Suele estar acompañado por dificultad al tragar. Se puede observar después de golpes, lesiones o de una disfunción del nervio. El tratamiento comienza a menudo con terapia de la voz, para ver si la cuerda vocal puede ser fortalecida.

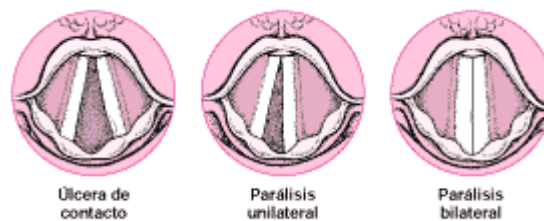
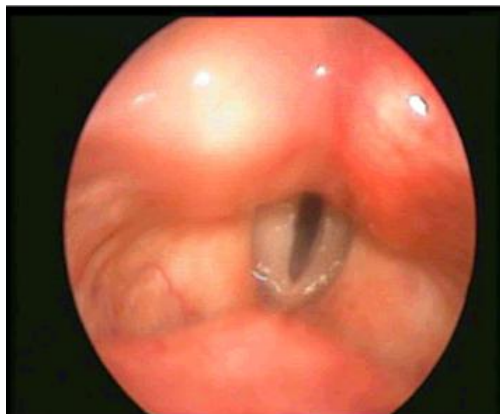


Figura 9. Úlcera de contacto. Parálisis unilateral y bilateral



a)



b)

Figura 10. Ejemplos de parálisis de cuerda vocal izquierda

2.1.3.7 Cáncer de laringe

El cáncer de laringe es una enfermedad muy seria que afecta directamente en la voz del paciente. Cuando el cáncer afecta a las cuerdas vocales la voz sufre un cambio en el tono, que se vuelve ronca y áspera. Otros signos son dificultad al tragar y respirar, dolor de oído o la presencia de un bulto en el cuello. Estos

síntomas suelen aparecer de manera bastante precoz, por lo que en cuanto se detecten es fundamental acudir al especialista. Su **diagnóstico temprano garantiza** que el tratamiento quirúrgico o radioterápico obtenga **un porcentaje de curación superior al 90%** [Armstrong+10].

Aunque entre las causas del cáncer de laringe están presentes varios factores, en su mayoría está relacionado con el consumo de tabaco y alcohol y con el virus del papiloma humano. Por eso hay que incidir en la importancia de que el cáncer de laringe se puede prevenir modificando los hábitos de vida; es decir, abandonando el tabaco y la ingesta de alcohol.

2.1.3.8 Disfonía espasmódica

Cuando hablar supone esfuerzo y fatiga, la voz puede sonar tensa, presionada y como con rápidas paradas y comienzos. Puede haber cambios incontrolados en tono e intensidad.

Después de la diagnosis correcta, muchos pacientes consideran la posibilidad de inyectarse botox. La inyección de cantidades muy pequeñas de una toxina en los músculos vocales produce una capacidad de producir una voz mejor. Esto es un tratamiento (no una curación) y da lugar a menudo, a la producción de voz mejorada durante semanas o meses, antes de que se requiera cualquier otra inyección.

2.1.3.9 Úlceras de contacto

Las úlceras de contacto (ver Figura 9) en las cuerdas vocales pueden ocurrir cuando se fuerza demasiado la voz para hablar. Cuando las cuerdas vocales son forzadas a juntarse excesivamente, pueden producirse llagas ulceradas. El tejido puede también erosionarse cercano a la zona de los cartílagos de la laringe. Las úlceras algunas veces son también causadas por reflujo gastroesofágico. Las úlceras de contacto pueden causar que la voz se canse con facilidad y puede producir dolor de garganta.

2.1.3.10 Curvatura del pliegue vocal

Un tipo anormal de movimiento del pliegue vocal cuando las cuerdas vocales solo se unen parcialmente para producir voz. La causa, a menudo se cree que es la atrofia del músculo (tono muscular pobre, pérdida de fuerza). El diagnóstico cuidadoso y exacto es importante para distinguir las causas de la inclinación.

2.1.3.11 Papilomatosis laríngea

La papilomatosis laríngea (ver Figura 11) es un trastorno raro causado por el papiloma virus humano (su acrónimo en inglés es HPV). Existen más de 60 tipos de HPV. La papilomatosis laríngea causa el crecimiento de tumores dentro de la caja de la voz, cuerdas vocales o el conducto del aire desde la nariz a los pulmones. La mayoría de los papilomas laríngeos (tumores) ocurren en niños antes de los tres años de edad. Los tumores normalmente crecen rápidamente y pueden variar en tamaño, provocando problemas de respiración y para tragar. Otros síntomas pueden incluir tos y ronquera.

El tratamiento de los papilomas laríngeos puede incluir la cirugía para extirpar los tumores. Otras opciones de tratamiento pueden incluir las siguientes:

- La quimioterapia
- Los antibióticos

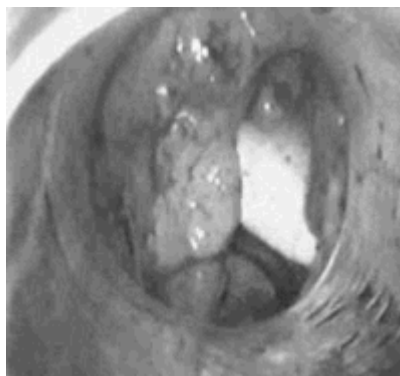


Figura 11. Papilomatosis recurrente de cuerda vocal izquierda

2.2 ANTECEDENTES TÉCNICOS

En esta sección se describen las principales técnicas de captura de imágenes de cuerdas vocales existentes en el mercado y que han sido utilizadas a lo largo de los últimos años, así como las técnicas de procesamiento digital más usadas para el análisis y extracción de características de la imagen. También se incluye una selección de los lenguajes de programación más usados para el desarrollo de algoritmos de procesamiento de imagen y se justifica la opción elegida.

2.2.1 Técnicas de captura de imágenes y software de análisis

Dado que esta investigación está centrada en el **estudio de las imágenes** (y vídeos) de cuerdas vocales, es objeto de este trabajo también el análisis de los distintos métodos de captura digital de éstas.

Hoy en día, la comunidad científica y médica acepta principalmente dos técnicas de captura, cuyos resultados pueden ayudar al diagnóstico de distintas patologías vocales. Por un lado está la **videoestroboscopia** y por otro la **videoquimografía** o captura de imágenes de alta velocidad. De éstas y otras técnicas se hablará en este apartado.

La forma de visualizar y estudiar las cuerdas vocales ha sido objeto de estudio desde varios siglos atrás. Se considera el primer descubridor de la laringoscopia indirecta con el espejillo a Manuel García en 1854 [Fernandez+06]. Al principio se le consideró un intruso ya que era compositor, tenor y profesor de canto, y no médico como cabía esperar, pero su idea de colocar un espejillo de dentista en la garganta e ir iluminando la laringe con la luz del sol reflejada en un espejo que sujetaba con su mano, examina las cuerdas vocales [Garcia47]. Posteriormente todas las sociedades laringológicas reconocieron su hallazgo.

Más adelante, fue Johann Nepomuk Czermak de Budapest [Nepomuk60] quien mejoró la técnica empleando luz artificial y espejillos de distintos tamaños y es

precisamente él quien consigue introducir la laringoscopia indirecta como método exploratorio principal.

En 1975 Stuckrad y Lakatos [Stuckrad+75] practicaron esta técnica con amplificación, es decir, con lupa. Pero no es hasta 1978 cuando Oertel [Oertel85] desarrolla la **laringoestroboscopia** que permite el examen de la vibración de las cuerdas vocales.

Este método exploratorio permite el diagnóstico por observación de patologías en estados iniciales, o que no necesariamente afectan a la morfología de la cuerda vocal, sino a su movimiento.

En los años 90, surge la **videokimografía** digital de alta velocidad de la laringe [Svec+96]. Esta técnica viene a solucionar los problemas de velocidad de captura de la videoestroboscopia. El ojo humano solo es capaz de 5 o 6 imágenes por segundo, mientras que la videoestroboscopia captura entre 25 y 30 frames por segundo [Poburka99], pero continúa siendo insuficiente para observar los movimientos dinámicos que tienen lugar en la laringe durante la fonación. Por eso, la videokimografía está siendo una herramienta muy valorada sobre todo en el campo de la investigación.

Desde el punto de vista de la exploración clínica, la videoestroboscopia digital es el método esencial y rutinario en el diagnóstico de los trastornos de la voz, ya que proporciona un grado de información extraordinario sobre el comportamiento del ciclo vibratorio y sus alteraciones.

Y es aquí donde se enmarca el **desarrollo de esta tesis**, precisamente en el diagnóstico de estas alteraciones, a partir de la **parametrización objetiva** de videos estroboscópicos.

Característica/técnica	Videoestroboscopia	Videoquimografía (VKG)
	(Tradicional)	(Captura de alta velocidad)
Frames/Seg	25-30	2000-4000
Posibilidad de diagnosticar tipología de patologías	Principalmente, patologías relacionadas con la morfología y el color	Específica para diagnosticar irregularidades (p.e. surcos o cicatrices), y cambios en la vibración de las cuerdas
Coste/Disponibilidad	Asequible Dependiente de la cámara. Disponibile en todos los hospitales.	Muy elevado. Orientada a fines de investigación. Disponibile en muy pocos centros hospitalarios.

Tabla 2. Comparativa entre Videoestroboscopia y Videoquimografía

Se utilizarán las imágenes capturadas con el estroboscopio, a sabiendas de que hay otros métodos de captura mejores, que proporcionan más frames por segundo y más información. La razón es que el uso de la videokimografía todavía es minoritario debido a su elevado coste, y que a priori el método implementado, será independiente del método utilizado en la captura de las imágenes, solo se observarán diferencias en la cantidad de medidas o precisión que se puedan realizar sobre las cuerdas.

2.2.1.1 Laringoscopia refleja

La laringoscopia refleja convencional ha sido desde el año 1855, en que el Profesor de Canto español Manuel García ideó el **espejillo laríngeo** [Fernandez+06], el método habitualmente utilizado para observar la laringe hasta nuestros días. La simplicidad de este método de exploración así como la calidad y precisión de las imágenes que permite obtener ha hecho que a pesar

de la tecnología de que disponemos hoy día no haya desaparecido, como era fácil suponer.

Aunque podría pensarse que está fuera de lugar como método de exploración laríngea y faríngea, dado que en gran medida ha sido sustituido con ventaja por otros métodos modernos, sigue siendo un **método de exploración laríngea válido**.

La laringoscopia refleja o indirecta consiste en la introducción en el fondo de la orofaringe de un espejillo laríngeo en el paciente, al que previamente se ha cogido y traccionado de la lengua hacia fuera. Sobre dicho espejillo se hace incidir un haz de luz. De esta manera se obtiene la visión de la laringe reflejada en el espejo. La realización de la laringoscopia refleja es satisfactoria sólo en el 80% de los casos en los que se intenta, cuando está bien realizada; sin embargo, esta cifra se reduce considerablemente si no se realiza adecuadamente. Distintos factores influyen en la realización correcta: la posición del paciente, una iluminación adecuada, el tamaño del espejillo, la habilidad por parte del observador y la tolerancia por parte del paciente. En lo que se refiere a la iluminación, ésta puede realizarse mediante un reflector frontal, el espejo frontal o el microscopio operatorio.

En cuanto a la posición del paciente éste debe estar sentado frente al observador con las rodillas juntas y el torso erguido. Habitualmente se coloca el foco luminoso a la derecha del paciente a unos 20 cm de distancia de la cabeza y en línea con su oreja. El observador se sienta frente a él con el espejo frontal colocado lo más próximo posible al ojo izquierdo de forma que con el ojo vea a través del pequeño orificio de su centro. De esta manera, ambos ojos quedan protegidos de la luz directa del foco luminoso. El espejo frontal se orienta manual para que el haz reflejado incida sobre el punto de unión de la úvula con el velo del paladar blando, donde se apoyará el espejito laríngeo. El espejillo

que se utiliza habitualmente es circular, de un diámetro aproximado de 21 a 25 mm respectivamente.

Una vez colocado el espejillo en su mango se sujeta con la mano derecha y después de calentarlo convenientemente para que no se empañe con el vapor de agua del aliento y comprobar que no está muy caliente, se prende la lengua del paciente con la interposición de una gasa entre los dedos pulgar y anular de la mano izquierda y se tira de ella delicadamente evitando dañar el área del frenillo lingual con los incisivos inferiores. El dedo índice de la mano izquierda se utiliza para elevar el labio superior con objeto de que no se interponga en el campo visual. El objetivo de la laringoscopia refleja es la visualización tanto de los márgenes como del inferior de la laringe en su totalidad.

2.2.1.2 Microlaringoscopia laríngea

Este tipo de laringoscopia indirecta aúna la utilización del espejillo con la visión magnificada y luz directa que supone el microscopio operatorio. Para esta exploración, el procedimiento básico de situación del paciente y posición de la cabeza no varía con respecto a la laringoscopia refleja convencional. En líneas generales, con la microlaringoscopia indirecta la visión estereoscópica es excelente y además se puede regular el grado de magnificación a voluntad del observador.

Estas características hacen posible no sólo una exploración sumamente descriptiva sino incluso realizar algunos procedimientos quirúrgicos de la laringe en condiciones realmente buenas.

Además, al microscopio se le pueden adaptar los visores laterales y la cámara de televisión o cámara fotográfica, así como sustituir la iluminación convencional por **luz estroboscópica**, con lo que se obtienen imágenes altamente demostrativas.

2.2.1.3 Exploración endoscópica

La utilización de las nuevas tecnologías ha modificado de forma extraordinaria la metodología de valoración visual de la laringe, haciendo posible al médico la observación tanto del estado como del comportamiento laríngeo en condiciones altamente informativas, es decir, se pueden tomar imágenes de las cuerdas vocales y observar su funcionamiento y movimiento vibratorio. En el momento actual, la utilización de endoscopios en combinación con una cámara de vídeo, un magnetoscopio y un monitor de televisión da lugar a lo que definimos como videoendoscopia.

Esta adaptación aporta la posibilidad de obtener el registro de imagen y sonido en tiempo real, lo que supone recoger una excelente documentación tanto visual como acústica.

El examen repetido de los datos obtenidos para su análisis detenido que posibilita la videoendoscopia, contribuye notablemente a la mejora del diagnóstico y de la docencia. Dependiendo del tipo de endoscopio que se utilice distinguimos dos métodos diferentes: **la endoscopia flexible o fibroscopia y la endoscopia rígida o telelaringoscopia.**

A. Endoscopia flexible o fibroscopia

Los fibroscopios son endoscopios constituidos por dos haces de fibras ópticas flexibles, que mediante una envolvente común constituyen una especie de cable en cuyo extremo se adapta un objetivo y un ocular. Uno de los haces está destinada a la transmisión de la imagen mientras que el otro es el que vehiculiza la luz.

Los fibroscopios están dotados de un sistema que mediante una leva de control manual hace posible la incurvación del extremo distal, lo que permite una gran amplitud y orientación del objetivo y por tanto de maniobra, lo que facilita su introducción y la observación de distintas regiones.

Técnica de la fibroendoscopia: antes de comenzar la exploración es necesario explicar al paciente en qué consiste y cuáles son las sensaciones que puede sentir, asegurándole que éstas no van a ser extremadamente desagradables ni dolorosas.

La vía de inserción habitual del fibroscopio es la nasal. El paso se realiza por el suelo de la fosa nasal o por el meato medio. El paciente, en posición cómoda, está sentado adecuadamente en el sillón de exploración, con la cabeza dirigida hacia delante y la barbilla ligeramente hacia abajo. Se introduce por la fosa nasal más amplia bajo control visual, hasta alcanzar la coana. En ese momento, se pide al paciente que respire por la nariz para relajar el velo del paladar y se procede a la incurvación de la punta del fibroscopio y se progresa hacia la mesofaringe. A partir de ese momento, la observación de las distintas zonas viene condicionada por el grado de introducción del fibroscopio. Se puede optar por una visión de conjunto, situándonos por encima del háditus laríngeo o proseguir el descenso hasta observar las cuerdas vocales desde una distancia muy próxima que ofrezca una visión magnificada de las mismas.

La exploración fibroendoscópica permite una exploración de la laringe fisiológica, sin que ésta se vea alterada por recurrir a maniobras extrañas como podían suponer la tracción de la lengua en el caso de la laringoscopia convencional. Durante la exploración fibroendoscópica se pide al paciente que emita las distintas vocales así como que inicie durante unos momentos una conversación normal, inspiración nasal brusca, voz cuchicheada, silbido, etc. con el fin de observar los mecanismos fisiológicos que tienen lugar durante los distintos tipos de fonación.

En la mayoría de los pacientes es posible la introducción del fibroscopio sin necesidad de vasoconstricción ni anestesia tópica.

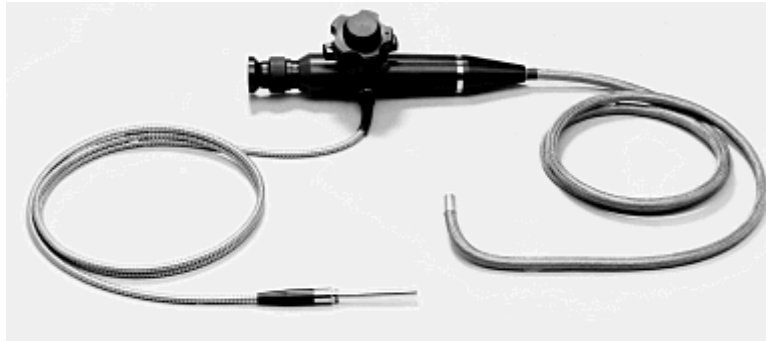


Figura 12. Ejemplo de fibroscopio 1



Figura 13. Ejemplo de fibroscopio 2

Gracias al fibroscopio (ver Figura 12 y Figura 13) se puede acceder a partes de la cavidad laríngea de forma poco invasiva, y de esta forma obtener imágenes y vídeos para posteriormente analizarlos y procesarlos. Y son los ingenieros de telecomunicaciones e informáticos los encargados de desarrollar el software necesario para apoyar a médico especialista en el diagnóstico.

B. Telelaringoscopia

El sistema óptico de los telelaringoscopios está formado por segmentos redondos de vidrio cuyos extremos están ópticamente tallados para generar en ellos una superficie cóncava o convexa que los configuran como lentes.

Su montaje en el interior de un tubo metálico es mucho más preciso y estable que el fibroscopio, lo que significa mejores prestaciones ópticas (obteniendo imágenes más precisas) y una mayor robustez que los sistemas clásicos. El sistema de conducción de luz utiliza, como el fibroscopio, un haz de fibras desordenadas cuyo extremo distal puede configurarse de distintas formas y

situarse en distintas posiciones con respecto a la lente, buscando la mejor distribución de la luz en el campo.

Los diámetros externos de los telelaringoscopios para laringología oscilan entre 7 y 10 mm y se pueden construir con diferentes ángulos de visión con campos también variables siendo los más utilizados los de 70° y 90°. La tolerancia durante la exploración es siempre mayor con el de 7 mm de diámetro.

Las prestaciones ópticas de los telelaringoscopios superan en gran medida las de los fibroscopios, sobre todo en lo que se refiere a magnificación, iluminación en ocular y definición.

La calidad de imagen es excelente y de gran magnificación, lo que permite obtener buenas fotografías e imágenes de vídeo de gran expresividad.

La exploración de la laringe mediante telelaringoscopia rígida se realiza utilizando como vía de introducción la cavidad oral (ver Figura 14 y Figura 15). La posición del paciente depende del tipo de endoscopio. Cuando se utiliza el de 90° debe procurarse que el ángulo que forma la parte inferior de la cavidad oral y la faringe sea aproximadamente de 90°. Si se emplea el de 70° el ángulo deseable es también de 70°, que se obtiene mediante una ligera extensión del cuello y la flexión de la cabeza o mediante la inclinación del paciente hacia delante hasta apoyar sus codos sobre los muslos y la elevación ligera de la barbilla.



Figura 14. Telelaringoscopia

Aunque como en la laringoscopia refleja no es siempre necesaria la exteriorización lingual, ésta es aconsejable en la mayoría de los casos. Tanto la presión que ejercen los dedos de la exploración sobre la lengua como la tracción hacia fuera deben ser las justas para obtener la posición deseada.

Un hecho frecuente que se da en la exploración telarlaringoscópica es el empañamiento de la lente. Para evitarlo se emplean distintos procedimientos que van desde la humectación de la lente en la cara interna de la mejilla del paciente a la utilización de un spray antiempañamiento o al calentamiento mediante inmersión de la punta del endoscopio en agua caliente.

Previamente a la inserción del telarlaringoscopio en la cavidad oral del paciente se explica en qué consiste y cuáles son las sensaciones que puede sentir. Asimismo, se le instruye en la manera de controlar la respiración y en la forma de fonar.

El uso de anestesia tópica se limita a aquellos casos en los que el reflejo nauseoso impide la exploración, hecho éste relativamente frecuente. La aplicación debe realizarse mediante spray dirigido a la cavidad orofaríngea ampliada por tracción lingual, evitando la aplicación oral para no producir saliva en exceso. El punto más reflexógeno es la base de la lengua, por lo que suele dar mejor resultado la anestesia por contacto mediante porta algodones curvo del surco glossofaríngeo lateral y base de la lengua.

Tanto la fibroendoscopia como la telarlaringoscopia pueden realizarse con **luz halógena** (lo habitual) o con **luz estroboscópica**. Esta última permite la **visualización de los fenómenos de ondulación de la mucosa a cámara lenta**, lo que permite en muchos casos un diagnóstico extremadamente preciso. Asimismo, las nuevas tecnologías que permiten la digitalización de las imágenes en tiempo real y posterior tratamiento de las mismas permiten extraer parámetros cuantitativos en lo que a la dinámica, morfología y color de las

estructuras laríngeas se refiere, permitiendo realizar diagnósticos mucho más precoces y precisos que lo que se podía realizar con las técnicas habituales.



Figura 15. Imagen de telelarínscopia

2.2.1.4 Estroboscopia

La endoscopia rígida y la fibroscopia tradicionales, sin luz fraccionada (luz estroboscópica), muestran la estructura y posibles lesiones de la laringe, pero no alteraciones en el movimiento de la mucosa que la recubre. Muchas patologías, en estadios iniciales, no muestran signos en las cuerdas vocales pero sí en el movimiento de la onda de la mucosa, que está alterado. Estudiarla ayuda a detectar lesiones muy pequeñas, imperceptibles con otros métodos. La laringoestroboscopia favorece un diagnóstico precoz. Ciertos cambios en la movilidad de la onda ya ofrecen indicios de la posible presencia de un tumor, lo que permite orientar rápidamente el tratamiento.

Los métodos dinámicos de visión de funcionamiento de las cuerdas vocales constituyen el principal avance en el área de diagnóstico. La luz estroboscópica es el instrumento que ha permitido a los especialistas detectar lesiones que antes no podían diagnosticar, o controlar con precisión su evolución [Quarteiro10].

El endoscopio con luz estroboscópica es la ayuda diagnóstica más precisa (ver ejemplo de equipo utilizado en la Figura 16), permite un análisis **mediante la captura de imágenes de las diferentes estructuras laríngeas tanto en fonación como en reposo**. Este método de diagnóstico clínico de consultorio, permite observar la función de las cuerdas vocales y analizar en detalle sus características vibratorias a saber: perceptuales (simetría, periodicidad, cierre glótico, amplitud, onda mucosa) y acústicas (frecuencia fundamental e intensidad).

Con ayuda de la luz halógena se vienen realizando los exámenes de laringe y cuerdas vocales. Pero esta **luz halógena no permite hacer una evaluación dinámica de las cuerdas vocales y de sus características, lo cual solo es posible con ayuda de la luz estroboscópica**.

Gracias a la fuente de luz estroboscópica, que genera pulsos de alta intensidad y frecuencia se puede llegar a sincronizar dichas frecuencias y así obtener como resultado una imagen de cuerda vocal estática o de movimiento en cámara lenta que nos permita analizar sus características.



Figura 16. Equipo de uso en Consultorio

A. Aplicaciones clínicas de la endoscopia laríngea:

- Visualización de estructuras laríngeas y análisis de su dinámica
- Observación de cuerdas vocales en cámara lenta o estática, como se puede observar en la Figura 17 y Figura 18.
- Diagnóstico y evaluación de disfonías

- Diagnóstico temprano de cáncer laríngeo y de su evolución
- Control de los efectos de la radioterapia
- Evaluación de las parálisis de cuerda vocal
- Ayuda en labor docente
- Evaluación del entrenamiento vocal en cantantes.



Figura 17. Secuencia de Frames capturados con estroboscopia laríngea

B. Resultados:

Las características evaluadas incluyen:

- **Frecuencia fundamental e intensidad**, que son registradas automáticamente por el equipo y se expresan en Hz y en dB respectivamente.
- **Onda mucosa**, que está disminuida por lesiones como cicatrices, edema, hiperplasia epitelial, lesiones orgánicas y deshidratación, entre otras.
- **Simetría**, se analiza evaluando cada pliegue vocal simultáneamente y clínicamente significa diferencias en posición, forma, grosor y elasticidad de cada pliegue vocal.
- **Periodicidad**, se refiere a la regularidad de las vibraciones sucesivas
- **Cierre glótico**, las alteraciones se pueden deber a paresia o parálisis de la cuerda vocal, rigidez del margen vibratorio, disfunción de la articulación cricoaritenoides, disfonía psicógena o lesiones orgánicas, entre otras. Este cierre glótico puede ser completo o incompleto o puede presentarse hiatos a diferentes niveles.

C. Beneficios de la integración con el ordenador:

- Permite la presentación de textos sobre la imagen, por ejemplo, identificación con el nombre, diagnóstico, fecha, control pre, postoperatorio, informaciones varias, frecuencia fundamental e intensidad.
- Establecer referencias sobre la propia imagen para su posterior **análisis, evaluación y procesamiento**.
- Adquisición de señales tanto de audio como de electroglotografía con visualización simultánea de las cuerdas vocales durante la vibración.
- Archivo y búsqueda fácil del paciente desde el control del computador.
- Almacenamiento de imágenes para posterior procesamiento, análisis y medición.



Figura 18. Imagen capturada mediante estroboscopia digital

D. Beneficios de la integración digital:

La estroboscopia digital es un sistema innovador de endoscopia/estroboscopia digital que adopta una nueva referencia para la calidad de imagen en exámenes estroboscópicos y endoscópicos grabados. Los exámenes de los pacientes son grabados directamente en el ordenador. La calidad de la imagen es excelente y equivale a estándares de transmisiones /emisiones profesionales. Es superior a grabaciones previas basadas en videocasetes SVHS. La digitalización del examen de vídeo también trae como beneficios: un examen puede ser ubicado y visto en segundos porque no hay necesidad de devolver o adelantar al

comienzo del mismo; **la comparación de exámenes** es fácil y directa ya que dos resultados pueden ser vistos a la vez.

Esta comparación es siempre dependiente de la percepción de cada especialista. **Los software comerciales no proporcionan medidas objetivas, ni un diagnóstico automatizado (sin interacción del usuario),** ni técnicas que certifiquen esta comparación.



Figura 19. Equipo de Estroboscopia digital

La estroboscopia digital está diseñada para capacitar al médico para realizar exámenes sin ayuda (ver Figura 19). Esto se hace posible utilizando el pedal, el cual controla las operaciones claves del sistema (i.e., fuente de luz, grabación/pausa del video, etc.). Permite realizar una sofisticada base de datos del paciente. Esta información es fácilmente actualizada antes y después del procedimiento. Toda la información del paciente y de la prueba están almacenadas en un formato estándar para una rápida generación de informes. Estos informes incluyen información crítica de la evaluación, base de datos, imágenes y videoclips, los cuales pueden ser enviados para una remisión al fisiatra o a otro colega por consulta.

E. Ventajas:

- **Calidad de imágenes superior** comparada a la grabación en video.
- Reproducciones inmediatas de los exámenes grabados
- Fácil y rápida comparación de los exámenes (pre y post - cirugía).
- Fuente de luz (xenon, halógena).

F. Limitaciones del estroboscopio:

- El estroboscopio solamente puede ser usado cuando la voz del paciente es bastante estable. Es decir, no es posible utilizarlo en pacientes con afonía severa o pérdida de voz.
- El estroboscopio proporciona una vista de la vibración de las cuerdas vocales (una imagen en dos dimensiones). Aunque esta visión proporciona mucha información, no se puede evaluar la zona situada bajo las cuerdas vocales (una imagen en tres dimensiones).
- Para investigar ciertos desórdenes vocales (cicatrices, surcos...) se recomienda el uso del microlaringoscopio (ver Figura 20) o la videoquimografía.



Figura 20. Microlaringoscopia

2.2.1.5 Videoquimografía

La videoquimografía se desarrolló en 1994 en Groningen, como una alternativa de bajo costo a los **sistemas de adquisición de imágenes de alta velocidad**, y fue especialmente diseñada para examinar la **vibración de las cuerdas vocales** [Popolo+08].

Desde 1994 el sistema ha sufrido cambios y mejoras significativas y en 1997 se convirtió en un producto comercial disponible para el público (actualmente se está empezando a comercializar en el mercado español con sustanciosas mejoras).

El sistema usa una cámara especial que puede trabajar en dos modos diferentes: estándar y de alta velocidad. En el modo estándar la cámara de la videoquimografía funciona como una cámara de vídeo comercial con una velocidad de imagen de 50 frames/s (de acuerdo con la televisión estándar). En el modo de alta velocidad, la cámara entrega imágenes desde una única línea a la velocidad de hasta 4000 frames/s. Las consecutivas imágenes se presentan una tras otra en el monitor (ver Figura 21), creando una nueva, la imagen de la videoquimografía que muestra el movimiento vibratorio de la parte seleccionada de las cuerdas vocales.

A. Ventajas de la videoquimografía:

Las ventajas de la VKG con respecto a otros sistemas de alta velocidad son las siguientes:

- La VKG es significativamente menos cara (el equipo, el coste de almacenamiento...)
- Menos cantidad de datos para ser procesados y guardados
- La duración de las muestras grabadas es virtualmente ilimitado

- El formato de la imagen es estándar, como el de cualquier vídeo comercial
- Ofrece una excelente resolución espacial
- Ofrece una excelente velocidad (4000 frames/s)

B. Desventajas de la videoquimografía:

- La falta de imagen completa en el modo de alta velocidad
- Las diferencias de fase anterior y posterior en la vibración de las cuerdas vocales no se registra
- La posición de medida tiene que ser seleccionada y apropiadamente ajustada antes de empezar a grabar y en el modo estándar.

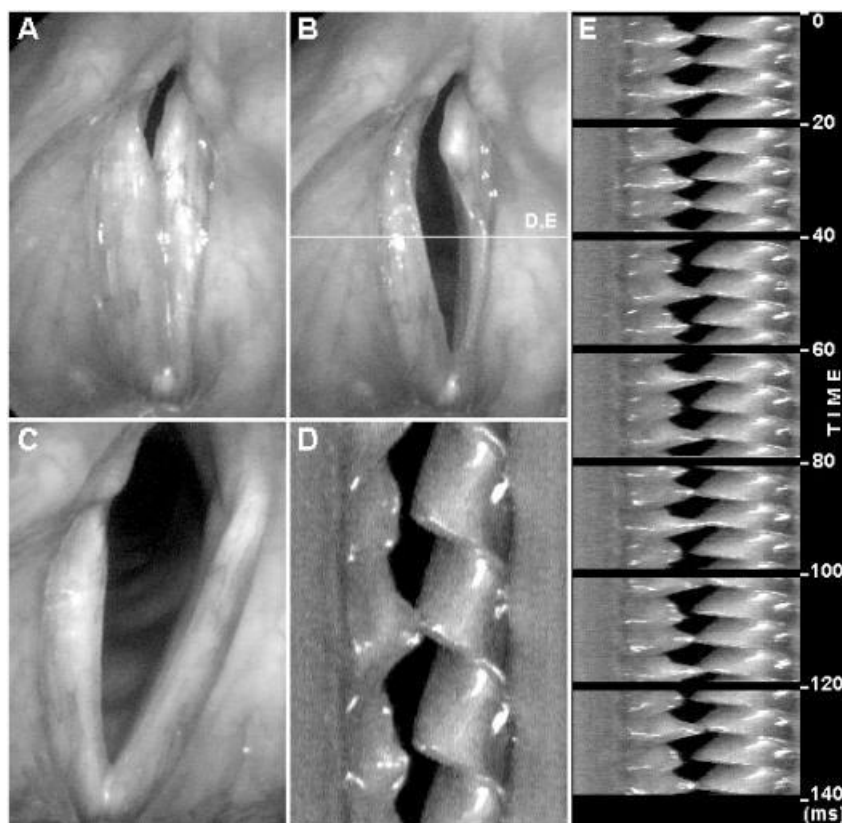


Figura 21. Imagen de cuerdas vocales obtenida mediante VKG

La videoquimografía ha sido empleada básicamente en la investigación de la voz y sus patologías., pero también en el ámbito clínico. Si nos centramos en la investigación, esta técnica es de gran ayuda en el estudio de las irregularidades

y los repentinos cambios en la vibración de las cuerdas vocales. La elevada resolución espacial de la VKG permite medir los pequeños desplazamientos de las cuerdas vocales, por ejemplo, para estudiar la laringe o la resonancia.

A nivel clínico este nuevo método supone una herramienta de diagnóstico complementaria a la estroboscopia y proporciona más datos y más información para el análisis posterior, y para obtener un buen diagnóstico. El inconveniente es su coste, y está bastante limitado al área de la investigación.

2.2.1.6 Software de Análisis

Recientemente, grupos de investigación internacionales en colaboración con médicos especialistas, están mostrando su interés por el desarrollo de software de análisis de características de las imágenes de cuerdas vocales, aunque habitualmente no llegan a comercializarse. Dado que las imágenes de VKG son de mejor calidad, el desarrollo de esta tipología de aplicaciones, suelen desarrollarse para su análisis, tal y como se puede observar en la literatura [Manfredi+07] [Manfredi+12].

Kay posee un software llamado KIPS (Kay's Image Processing Software) [KAY] orientado al procesado de imágenes de cuerdas vocales, provenientes de capturas estroboscópicas o de alta velocidad. Incluye multitud de herramientas para la edición y procesado de las imágenes, proporcionando al especialista información muy valiosa de apoyo a su diagnóstico, pero no emite un diagnóstico ni automatiza 100% el proceso.

Habitualmente, los software utilizados por otorrinolaringólogos, o cualquier especialista de la voz, sólo tienen en cuenta el análisis de voz, y forman parte de una unidad o laboratorio de la voz [Gonzalez+02].

Es importante recordar que **la medida de los parámetros acústicos no da el diagnóstico** de la lesión, pero puede indicar el grado de alteración de la disfonía. Los parámetros objetivos aceptados por la comunidad científica en el

análisis acústico son: pitch, jitter, shimmer y HNR (y sus variantes) y son indicativos de una buena o mala calidad vocal [Baken+00].

2.2.2 Alternativas de lenguajes para el desarrollo de algoritmos de procesamiento digital de imagen

Dado que la tesis Doctoral que se expone en esta memoria consiste en el diseño de un algoritmo de procesamiento digital de imagen se requiere implementar un programa software con el que poder validar los resultados del algoritmo diseñado. Por tanto, se ha tenido que elegir un lenguaje de programación en el que realizar dicha implementación. Para la elección del lenguaje de más adecuado se han tenido en cuenta diferentes parámetros:

- Conocimientos previos
- Facilidad de programación
- Librerías con funciones básicas de procesamiento de imagen
- Rapidez computo

Para poder tomar una decisión, estos parámetros se han evaluado en diferentes lenguajes: ImageJ, SciPython, OpenCV y Matlab.

2.2.2.1 ImageJ

ImageJ [ImageJ] es un programa de procesamiento de imagen desarrollado en Java por el Instituto Nacional de Salud de Estados Unidos [NIH]. Es multiplataforma, de código abierto y dispone de múltiples plugins desarrollados previamente y a disposición del público en general en su página web. Desarrollar algoritmos de procesamiento de imagen utilizando ImageJ para programadores experimentados en Java no es ningún inconveniente. No obstante, Java al no ser un lenguaje de programación orientado al cómputo matricial, el desarrollo de algoritmos no es trivial. Por otro lado, Java es un lenguaje que requiere compilación, y genera código interpretable por una máquina virtual. Esto hace que la velocidad de cómputo obtenida no sea tan

rápida como un lenguaje de programación que genere directamente código máquina.

2.2.2.2 SciPython

SciPython [SciPy] es un software orientado al desarrollo de aplicaciones en áreas como las matemáticas, la ciencia y la ingeniería. Entre sus puntos a favor la librería “Python Imaging Library” para el desarrollo de algoritmos de procesamiento de imagen. El lenguaje de programación Python cada vez se está imponiendo más en el ámbito del procesamiento de señal. Por ello, se han analizado las bondades de este lenguaje de programación para desarrollar la tesis doctoral, pese al hándicap inicial de no tener ninguna experiencia previa con él. Finalmente no se ha implementado la tesis doctoral en este lenguaje ya que las ventajas que ofrece son las mismas que ofrecen otros lenguajes de los cuales si se cuenta con un know-how previo.

2.2.2.3 OpenCV

OpenCV [Bradski08] [OpenCV] es una librería de programación de funciones para visión por computador en tiempo real. Esta librería es extremadamente potente, con una gran cantidad de funciones de procesamiento de imagen disponibles para su utilización. Permite un código muy rápido que se puede utilizar en aplicaciones de tiempo real. Estas cualidades son muy idóneas para un producto comercial que se podría derivar de esta tesis doctoral, pero para la prueba y validación del algoritmo diseñado no son tan relevantes. Como desventaja hay que decir que es necesario disponer de un alto conocimiento del lenguaje de programación C para poder desarrollar algoritmos utilizando esta librería.

2.2.2.4 Matlab

Matlab [Mathworks] es un lenguaje de alto nivel que permite realizar tareas de cálculo complejo de forma más *rápida* que los lenguajes tradicionales (C, C++,

Java, o Fortran). Además, Matlab dispone de una gran cantidad de toolboxes para el desarrollo de algoritmos de procesamiento digital de imagen, visión artificial y adquisición de imagen. Es un lenguaje de programación de *cálculo matricial de forma nativa*, cualidad que facilita el rápido desarrollo de algoritmos de procesamiento de imagen. Entre otras ventajas destaca su entorno de desarrollo, que permite un realizar test y pruebas de los algoritmos desarrollados de una forma muy rápida y sencilla. También es muy destacable que Matlab es un lenguaje de programación muy reconocido por toda la comunidad científica, lo que permite que los trabajos derivados de la tesis doctoral tengan una aceptación más sencilla entre los evaluadores de revistas científicas y conferencias internacionales.

Tras sopesar las diferentes alternativas se ha llegado a la conclusión que el **lenguaje de programación más adecuado es Matlab**. Se posee un gran conocimiento previo de este lenguaje de programación, permitiendo una implementación más sencilla y rápida del algoritmo. El hecho de que sea un lenguaje de programación interpretado no es un inconveniente, y aunque los algoritmos resultantes no sean muy eficientes computacionalmente, la implementación del algoritmo se ha realizado para verificar su validez, no para obtener un producto comercial. Para un producto derivado del trabajo aquí presentado, sí se debería optar por otro tipo de lenguaje de computación más eficiente en cuanto a rendimiento, y probablemente la elección más adecuada fuese OpenCV. Sin embargo, para los propósitos de esta tesis doctoral no cabe duda que la elección adecuada es Matlab.

2.2.3 Técnicas y métodos generales empleados en el procesamiento digital de imagen

Son técnicas de tratamiento digital de imagen todas aquellas manipulaciones realizadas sobre la imagen con el objetivo de:

- **Mejorar su calidad** eliminando degradaciones
- Conseguir una **representación eficiente** que permita almacenarla en un determinado soporte con el menor número posible de bits sin una pérdida apreciable de calidad o requerir el menor tiempo posible de transmisión.
- Extraer información relevante de cara a interpretar su contenido, tomar decisiones, etc.

Si bien existe una gran interrelación entre las técnicas utilizadas en cada uno de los campos anteriores, las aplicaciones prácticas son muy diferentes, y cubren un amplio abanico de posibilidades, desde la industria y la medicina a las aplicaciones militares y de ocio.

En este apartado se revisan las técnicas más utilizadas en el procesado digital de imagen [Cuevas+10] [Gonzalez+09] y que se consideran que podrían tener cabida dentro de esta investigación.

2.2.3.1 Mejora de la imagen

Son técnicas de mejora todos los algoritmos destinados a **acentuar, resaltar, contrastar**, etc, determinados aspectos de la imagen o **eliminar efectos no deseados** como, por ejemplo, desenfokes o ruidos.

Las imágenes médicas son el mejor ejemplo de aplicación en el que es necesario la mejora de las imágenes obtenidas. Las limitaciones del sistema de adquisición conducen a imágenes de escaso contraste, que dificultan la discriminación de los distintos elementos presentes en la imagen (pulmones, costillas, corazón...). Mediante técnicas de tratamiento de imagen es posible resaltar los distintos elementos de forma que sean más claramente identificables por el ojo humano.

Es importante destacar que mediante la mejora de imágenes:

- **No se añade información nueva** no presente en la imagen. Simplemente se resalta la existente, para que sea mejor apreciada por el ojo humano.
- **No existe un criterio para cuantificar el grado de mejora** que introducen las diferentes técnicas. La valoración de los resultados es subjetiva, y a menudo se trata de métodos empíricos que requieren procedimientos de prueba y error hasta obtener el resultado deseado.

En todas las técnicas de las que se va a hablar a continuación, se han utilizado imágenes de las cuerdas vocales para realizar las transformaciones. La razón, es el interés ya mostrado de realizar la tesis en el ámbito de la bioingeniería.

A. Negativo de una imagen

Esta técnica intercambia los niveles blanco y negro entre sí mediante la función:

$$f(u) = 2^B - u \quad (1)$$

Se obtiene el negativo digital de la imagen. La Figura 22 b el resultado de aplicar dicha función a la imagen original de la Figura 22 a.

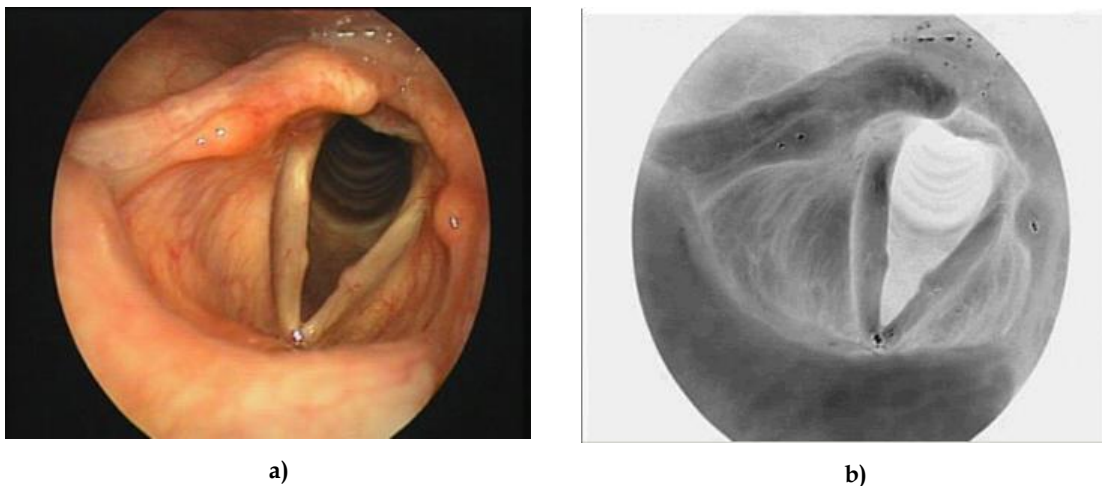


Figura 22. Imagen de cuerdas vocales. a) original. b) Negativo de la a)

B. Histogramas

El histograma de una imagen es una **representación gráfica de la frecuencia con la que los niveles de grises aparecen en dicha imagen** (ver Figura 23). El eje de abscisas indica los distintos niveles (discretos) de grises y en ordenadas se representa la frecuencia o, también a veces, el número de píxeles que poseen el nivel de gris (ver Figura 24).

El histograma se construye, simplemente rastreando toda la imagen y contabilizando el número de píxeles que poseen cada nivel de gris.



Figura 23. Imagen de cuerdas vocales en escala de grises

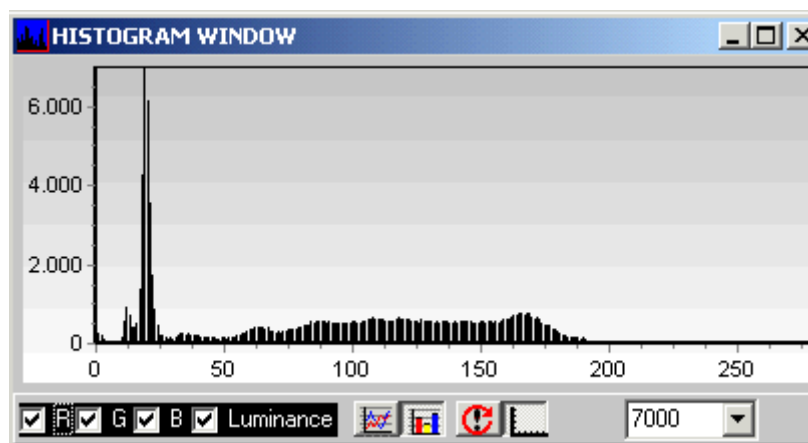


Figura 24. Histograma de la Figura 31

El histograma proporciona la información estadística sobre cómo están distribuidos los distintos niveles de grises de la imagen. Esta información puede

ser bastante útil para tareas tales como conocer si la digitalización se ha efectuado correctamente, decidir el valor de umbralización de una imagen, tener una estimación de brillo medio y contraste, etc.

Las operaciones de mejora basadas en el histograma son principalmente la ecualización y la especificación del histograma.

En la ecualización, el objetivo es obtener en la imagen de salida un histograma uniforme, mientras que en la especificación del histograma se fuerza a la imagen de salida a tener un determinado histograma.

C. Realce

El principal objetivo de las técnicas de realce es **aumentar el contraste mediante la redistribución de los niveles de grises de la imagen**. Para nuestros propósitos, la finalidad de este aumento de contraste no es mejorar la calidad de la imagen para una mejor visualización, sino obtener una imagen resultante que sea más adecuada que la original para una determinada aplicación.

D. Transformación de la escala de grises

La idea que subyace de este tipo de procesamiento no es más que la de aprovechar todos los niveles de la escala de grises.

Considérese, por ejemplo, la Figura 25 donde se muestra un histograma de intensidades. Se puede observar cómo existe una amplia banda de grises que no son utilizados en la imagen. Aprovechando estos niveles, la transformación de la escala de grises permite dilatar el rango de intensidades con el consiguiente aumento de contraste en la imagen.

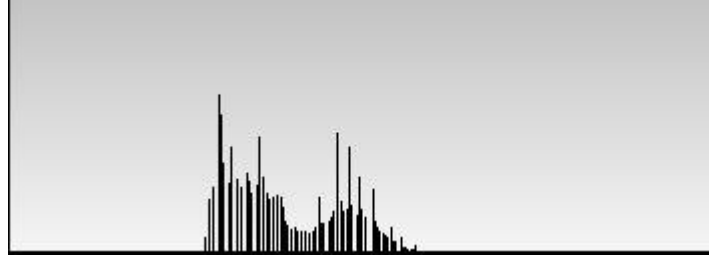


Figura 25. Histograma de una imagen en la cual no todos los niveles de grises son utilizados

La modificación de la escala de grises de una imagen f con niveles de intensidades discretos r_k (eje horizontal) viene dada por una transformación:

$$s_k = T(r_k) \quad k = 0, 1, \dots, L-1 \quad (2)$$

donde s_k son los niveles de grises (también discretos) de la imagen resultante, y L es el número de estos niveles.

Se supone que ambas escalas disponen del mismo rango de niveles, es decir si r_k toma valores entre 0 y 255, también los s_k se encuentran comprendidos en dicho rango. Esta suposición es usual en la mayoría de las aplicaciones en procesamiento de imágenes.

E. Realce local

El procedimiento general para la aplicación de un realce local consiste en definir una ventana $n \times m$ que es posicionada píxel a píxel por la imagen, operando solamente en dicho entorno en base a atributos de las intensidades de sus píxeles.

Dos de los atributos que más se utilizan son la media y la varianza. La media de intensidades da una medida del brillo medio, mientras que la varianza proporciona una estimación del contraste de la región considerada de la imagen. Así pues, cuanto mayor sea la media, mayor será el brillo medio, y mientras mayor sea la varianza, mayor será el contraste.

A continuación se muestra una transformación típica basada en estos conceptos, donde los valores de intensidad de la imagen de entrada $f(i,j)$ son modificados píxel a píxel obteniéndose una nueva imagen $g(i,j)$ de la siguiente forma:

$$g(i, j) = A(i, j)[f(i, j) - m(i, j)] + m(i, j) \quad (3)$$

con:

$$A(i, j) = K \frac{M}{\sigma(i, j)} \quad (4)$$

donde $m(i,j)$ y $\sigma(i,j)$ son la media y desviación típica calculadas sobre un entorno de vecindad centrado en el píxel (i,j) , M es la media global de $f(i,j)$, K es una constante que toma valores entre 0 y 1.

F. Suavizado

Las operaciones de suavizado tienen por objeto **reducir ruido y/o efectos espurios** que pueden presentarse en una imagen a consecuencia del proceso de captura, digitalización y transmisión. Su utilización es normalmente necesaria antes de la aplicación de un detector de bordes.

De acuerdo con la teoría de filtrado de señales, la reducción de ruido se puede realizar en el dominio de frecuencias mediante la aplicación de un filtro paso-bajo. Es sabido que la multiplicación de dos señales en el dominio de frecuencias equivale a la convolución de éstas en el dominio espacial, es decir:

$$G(W) = F(W)H(W) \leftrightarrow g(x) = f(x) \otimes h(x) \quad (5)$$

Donde $G(W)$, $F(W)$, $H(W)$ son transformadas de Fourier de las señales unidimensionales $g(x)$, $f(x)$ y $h(x)$, respectivamente. En el caso que se está tratando, imágenes digitales de las cuerdas vocales, estas funciones serán bidimensionales y discretas.

Habitualmente en el procesamiento de imágenes se trabaja en el dominio de la frecuencia, pero en algunos casos el coste computacional de la FFT es excesivo.

A continuación, se presentan las técnicas más utilizadas para suavizar una imagen.

G. Promediado del entorno

Bajo el nombre de promediado del entorno se agrupan una serie de técnicas que tienen en cuenta el nivel de gris de los píxeles vecinos sobre los que se pretende actuar. Sobre esta base se desarrollan una serie de variantes que difieren, principalmente en la forma de escoger los píxeles vecinos y en la ponderación que se le da a cada uno de ellos. En cualquier caso, esta operación se lleva a cabo mediante la convolución de la imagen con una ventana o máscara.

El principal inconveniente de estas técnicas es el enturbiamiento que se produce en la imagen, provocando el difuminado de los bordes, por lo que deben ser usados con cierta precaución.

Media simple

Dada una imagen $f(i,j)$, el procedimiento consiste en generar una nueva imagen $g(i,j)$ cuya intensidad para cada píxel se obtiene promediando los valores de intensidad de los píxeles de $f(i,j)$ incluidos en un entorno de vecindad predefinido, es decir, la imagen suavizada se obtiene como:

$$g(i,j) = \frac{1}{p} \sum_{(m,n) \in S} f(m,n) \quad (6)$$

Donde S es un subconjunto de “ p ” píxeles situados en el entorno de (i,j) , incluyendo (i,j) .

El tamaño de la máscara es importante. Cuanto mayor sea ésta se consigue una mayor reducción de ruido, pero a cambio se produce una mayor difuminación de los bordes.

Promediado selectivo y promediado ponderado

Si fuera posible realizar el promediado teniendo en cuenta sólo determinados píxeles vecinos, se podría evitar el difuminado de características de interés como bordes, esquinas, líneas y puntos de interés. Esta técnica, conocida como promediado selectivo, se puede abordar mediante un gran número de variantes.

Un posible enfoque consiste en añadir un detector de bordes y promediar sólo en los píxeles donde estos no hayan sido detectados. Otra idea más refinada es realizar un promediado direccional, es decir, si el borde es detectado se promedia sólo con los píxeles que se encuentran en la dirección del borde. El principal inconveniente de este tipo de promediado es el coste computacional que resulta de añadir el detector de bordes, lo que lo hace inviable en la mayoría de las aplicaciones.

Otra alternativa consiste en decidir qué píxeles vecinos entran en el promediado, en base al parecido con el píxel que se está promediando

Filtro de Gauss

Una distribución gaussiana con desviación típica σ y media μ viene dada por:

$$g_{\sigma} = \frac{1}{\sigma\sqrt{2\pi}} e^{-1/2\left(\frac{x-\mu}{\sigma}\right)^2} \quad (7)$$

La convolución de la función $g_{\sigma}(x)$ dada en la ecuación anterior con una señal $f(x)$ da lugar a una nueva señal suavizada $h(x)$, donde el valor en cada punto es el resultado de promediar con distintos pesos los valores vecinos a ambos lados de dicho punto. En este suavizado, la desviación típica σ juega un papel

importante a la hora de controlar el grado de suavizado de este operador. Cuanto mayor sea σ , más se tienen en cuenta los puntos lejanos, y, por tanto, se realizará un mayor suavizado. Conforme σ disminuye, los valores de los puntos vecinos van recibiendo menor ponderación, con lo que la reducción de ruido será menor.

Extendiendo la función anterior a dos dimensiones se tiene:

$$g_{\sigma}(x, y) = \frac{1}{2\pi\sigma^2} e^{-1/2\left(\frac{x^2+y^2}{\sigma^2}\right)} \quad (8)$$

Donde, además, se ha supuesto que la campana de Gauss se centra sobre la media.

Aunque este tipo de suavizado también tiene el problema del difuminado de los bordes, por las razones expuestas anteriormente (en el promediado del entorno), éste no es tan acusado como en el caso de la media simple.

Filtrado de la mediana

Una de las dificultades principales del promediado del entorno es que desdibuja los contornos y los detalles de la imagen. Una alternativa es el uso el filtro de la mediana, que consiste en reemplazar el nivel de gris de cada píxel por la mediana de los niveles de grises de los vecinos. A diferencia de los métodos vistos hasta ahora, esta técnica no se implementa en base a la convolución de la imagen con una determinada máscara (método lineal), sino que es un método no lineal.

Sea $\{a_1, a_2, \dots, a_N\}$ una secuencia discreta de un número impar de valores ordenados creciente o decrecientemente. Se define la mediana de esta secuencia como elemento $a_{(N-1)/2}$. Por ejemplo, si se consideran los 4-vecinos de "p" cuyas intensidades son 1, 2, 9, 4 y 5, el píxel "p" se reemplazará por el valor 4, ya que éste es el valor central en la secuencia ordenada 1, 2, 4, 5 y 9.

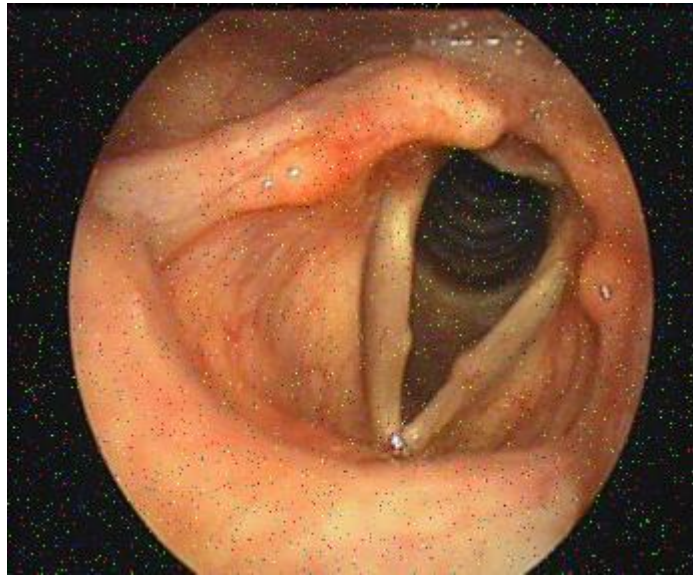


Figura 26. Imagen de las cuerdas vocales a la que se le ha añadido ruido



Figura 27. Imagen de las cuerdas vocales tras aplicarle un filtro de promediado de máscara 3x3, todavía se pueden observar restos de espurios

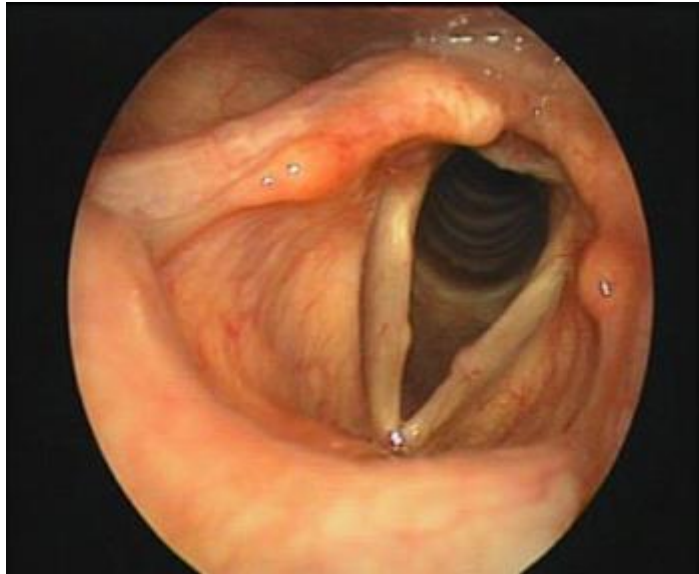


Figura 28. Cuerdas vocales tras aplicarle el filtro de la mediana. Se recupera la imagen original, sin restos de ruido

Puede afirmarse que, en general, **el comportamiento del filtro de la mediana es mejor que el del promediado del entorno** (ver Figura 27) **para eliminar efectos espurios y para preservar los bordes de la imagen** (ver Figura 28). Presenta, sin embargo, el inconveniente de que pierde los detalles finos como líneas delgadas, puntos aislados y redondea las esquinas.

Otro efecto que puede producir el filtro de la mediana es el desplazamiento de los bordes. Dicho problema debe ser tenido en cuenta en aplicaciones donde se requiere una gran precisión en la localización de éstos.

A la hora de implementar la mediana el procedimiento básico sería:

1. Ordenar los $2N+1$ valores (intensidades) del entorno de vecindad de cada píxel.
2. Varias posibilidades:
 - Tomar como intensidad del píxel la mediana de los valores anteriores.
 - Eliminar los k mayores y los k menores y promediar con el resto.
 - Etc.

Obsérvese que no es necesario volver a ordenar todos los valores de la ventana cada vez que ésta se desplaza a un nuevo píxel, sino que ésta se va reciclando, es decir, se eliminan de la secuencia ordenada los píxeles que se quedan fuera y se insertan nuevos.

Ampliación de imágenes

En ocasiones es necesario aumentar el tamaño de una imagen, o una parte de ella, con la finalidad de apreciar mejor sus detalles. Para ello existen diversos métodos.

Método 1: Repetición

Para pasar de una imagen I_1 de $M \times N$ píxeles a otra imagen I_2 de $kM \times kN$ ($k \geq 2$) el método más simple consiste en repetir cada píxel de todas las líneas k veces, y a continuación cada línea k veces, tal y como se puede ver en la Figura 29 con una imagen de cuerdas vocales. Este proceso es equivalente a:

- Intercalar k columnas y k filas entre sus píxeles, de valor 0, obteniendo una imagen de $kM \times kN$ píxeles.
- Convolucionar la imagen del punto 1 con la matriz unitaria H de $k \times k$ elementos, todos a 1.

$$H = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \text{ para un orden } K=2 \quad (9)$$

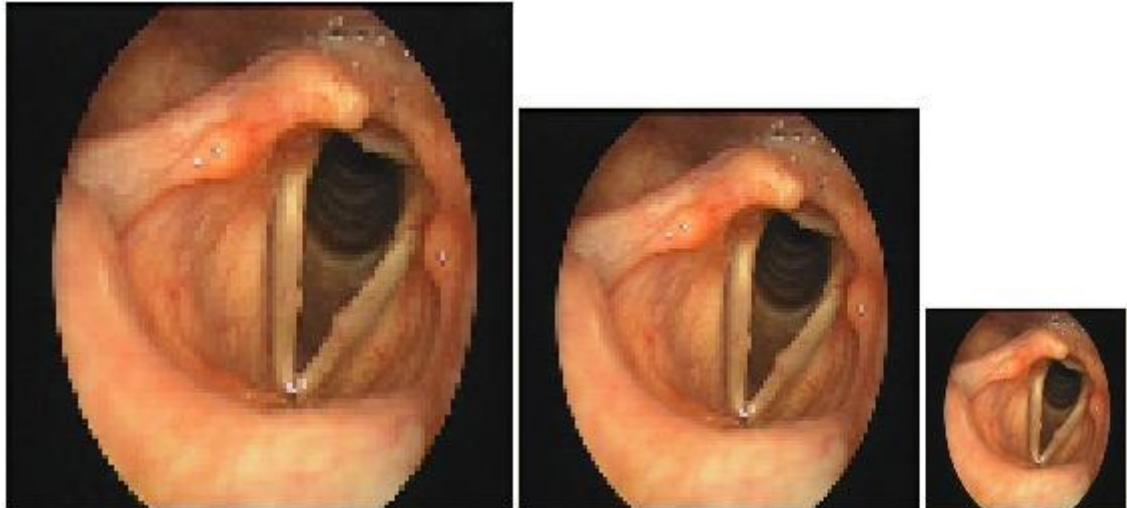


Figura 29. Ampliación de imágenes por repetición

Método 2: Interpolación lineal

Para evitar el efecto bloque del método anterior, el valor de los píxeles añadidos no se calcula por repetición sino por interpolación lineal, de forma que se produce una transición suave entre píxeles y, por tanto, no aparecen transiciones bruscas. Sin embargo, para órdenes de ampliación elevados, la imagen resultante queda borrosa.

El procedimiento consiste en ajustar una recta entre los píxeles de valor conocido de todas las líneas y después entre las columnas.

Mediante plantillas, las operaciones son semejantes al método de repetición, usando la plantilla:

$$H = \begin{bmatrix} 1/4 & 1/2 & 1/4 \\ 1/2 & 1 & 1/2 \\ 1/4 & 1/2 & 1/4 \end{bmatrix} \text{ para un orden } K=2 \quad (10)$$

Para una k genérica, hay que aplicar la plantilla H k veces, obteniendo una interpolación de orden k .

2.2.3.2 Detección de bordes

Los bordes de una imagen se pueden definir como transiciones entre dos regiones de niveles de gris significativamente distintos. Éstos suministran una valiosa información sobre las fronteras de los objetos que puede ser utilizada para la segmentación de imágenes médicas, reconocimiento de objetos, visión estéreo, etc.

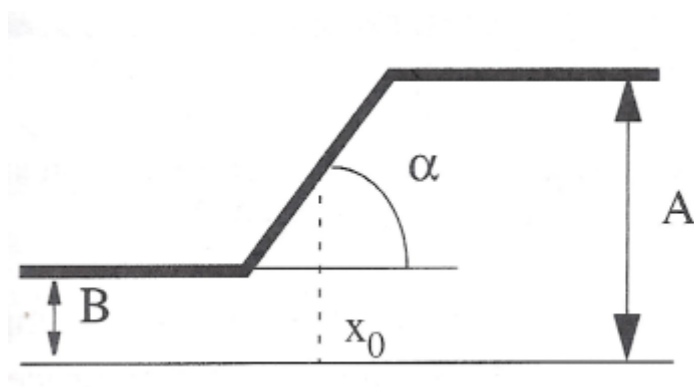


Figura 30. Modelo unidimensional y continuo de un borde ideal

En la Figura 30 se muestra un modelo unidimensional y continuo de un borde. Este modelo representa una rampa desde un nivel de gris bajo “B” a uno alto “A”, con los siguientes parámetros:

- Variación de intensidad $H = A - B$
- Ángulo de inclinación de la rampa “ α ”
- Coordenada horizontal “ x_0 ” donde se encuentra el punto medio de la rampa.

Un operador que proporcionara los valores “ x_0 ” y “ H ” daría unos datos valiosísimos sobre la imagen, ya que proporcionaría la amplitud del borde, y lo localizaría con exactitud dentro de la imagen.

En las imágenes reales los bordes nunca se ajustan totalmente al modelo anterior. Las causas para ellos son diversas, destacándose las siguientes:

- Las **imágenes** son **discretas**
- Están **afectadas por ruido diverso**

- El origen de los bordes puede ser muy diverso: bordes de oclusión, superficies de diferente orientación, distintas propiedades reflectantes, distinta textura, efectos de la iluminación (sombras y/o reflejos), etc.

Las circunstancias anteriores introducen una gran complejidad a la hora de detectar los bordes de una imagen. En particular, en este proceso se consideran tres tipos de errores:

- **Error en la detección.** Un operador es un buen detector si la probabilidad de detectar el borde es alta cuando éste realmente existe en la imagen, y baja cuando éste no existe.
- **Error en la localización.** Un operador localiza bien un borde cuando la posición que proporciona coincide con la posición real del borde de la imagen.

Ambos errores están estrechamente ligados a los problemas anteriormente señalados, y muy especialmente a la presencia de ruido de la imagen. En la práctica, la calidad en la detección y localización están en conflicto.

- **Respuesta múltiple.** Varios píxeles son detectados en un único borde.

La mayoría de las técnicas para detectar bordes emplean operadores locales basados en distintas aproximaciones discretas de la primera y segunda derivada de los niveles de grises de la imagen, si bien existen otras posibilidades para ello, como el empleo de patrones de bordes ideales.

A. Operadores basados en máscaras direccionales

Puesto que un borde ideal es esencialmente un patrón con forma de escalón, un método directo de detectar los bordes consiste en correlar con patrones en todas las orientaciones posibles. En la práctica, el número de orientaciones se limita a

8 (giradas 45 grados), tal y como se puede observar en la Figura 31 y en la Figura 32.

1	1	1	1	1	0	1	0	-1	0	-1	-1
0	0	0	1	0	-1	1	0	-1	1	0	-1
-1	-1	-1	0	-1	-1	1	0	-1	1	1	-1

Figura 31. Plantillas para detección de transiciones N, NO, O, SO

-1	-1	-1	-1	-1	0	-1	0	1	0	1	1
0	0	0	-1	0	1	-1	0	1	-1	0	1
1	1	1	0	1	1	-1	0	1	-1	-1	0

Figura 32. Plantillas para detección de transiciones S, SE, E, NE

Mediante las plantillas es posible detectar contornos independientes en ocho direcciones.

En la Figura 33 y en la Figura 34 se muestran las máscaras patrones propuestas por Kirsch (1971) [Maini+09].

El valor del gradiente en cada píxel de la imagen se obtiene como el máximo de las respuestas a cada una de las máscaras, mientras que la dirección será la indicada por la orientación de dicha máscara.

Aunque se han propuesto otras muchas máscaras direccionales, el alto coste computacional y la escasa precisión en la determinación del gradiente desaconsejan su utilización en la mayoría de las aplicaciones.

5	-3	-3	-3	-3	-3	-3	-3	-3	-3	-3	-3
5	0	-3	5	0	-3	-3	0	-3	-3	0	5
5	-3	-3	5	5	-3	5	5	5	-3	5	5

Figura 33. Máscaras orientadas de Kirsch para la detección de bordes, en las direcciones E, NE, N y NO

-3	-3	5	-3	5	5	5	5	5	-3
-3	0	5	-3	0	5	-3	0	-3	-3
-3	-3	5	-3	-3	-3	-3	-3	-3	-3

Figura 34. Máscaras orientadas de Kirsch para la detección de bordes, en las direcciones O, SO, S y SE

B. Operadores basados en la primera derivada (Gradiente)

La derivada de una señal continua proporciona las variaciones locales con respecto a la variable, de tal forma que el valor de la derivada es mayor cuanto más rápidas son estas variaciones.

Considérese una señal continua $f(x)$ que representa un borde en una dimensión. Parece lógico considerar que la localización exacta de este borde venga dada por el punto de inflexión de $f(x)$, es decir, por el máximo de la función $f'(x)$ (punto " x_0 "). De esta manera, la derivada de $f(x)$ permite determinar " x_0 " de una manera fácil, al tiempo que proporciona una estimación sobre la magnitud y el sentido de la variación (creciente si $f'(x)$ es positivo y, decreciente, en caso contrario).

En caso de funciones bidimensionales $f(x,y)$, la derivada es un vector que apunta en la dirección de máxima variación de $f(x,y)$ y cuyo módulo es proporcional a dicha variación. Este vector, denotado como $\nabla f(x,y)$, se denomina gradiente y se define como:

$$\nabla f(x,y) = \begin{bmatrix} \frac{\partial f(x,y)}{\partial x} \\ \frac{\partial f(x,y)}{\partial y} \end{bmatrix} = \begin{bmatrix} f_x(x,y) \\ f_y(x,y) \end{bmatrix} \quad (11)$$

Se puede observar que, de acuerdo con la expresión anterior, el gradiente en un punto (x,y) viene dado por las derivadas de $f(x,y)$ a lo largo de los ejes

coordenados ortogonales “x” e “y”. El módulo y la dirección del gradiente vienen dados por:

$$|\nabla f(x,y)| = \sqrt{(f_x(x,y))^2 + (f_y(x,y))^2} \quad (12)$$

Aproximaciones discretas del operador gradiente

En el caso bidimensional discreto, las distintas aproximaciones del operador gradiente se basan en diferencias entre los niveles de grises de la imagen. Por ejemplo, la derivada parcial $f_x(x,y)$ (gradiente de fila) puede aproximarse por la diferencia de píxeles adyacentes e la misma fila, es decir:

$$f_x(x,y) \approx G_F(i,j) = [F(i,j) - F(i,j-1)]/T \quad (13)$$

O también por la diferencia de píxeles separados:

$$f_x(x,y) \approx G_F(i,j) = [F(i,j+1) - F(i,j-1)]/2T \quad (14)$$

Donde se ha supuesto, sin pérdida de generalidad, que el índice de filas “i” crece de arriba abajo, y el de columnas “j” de izquierda a derecha. Puesto que las derivadas serán comparadas más tarde con un umbral ajustable, los factores de escala $1/T$ y $2/T$, pueden ser omitidos.

El gradiente de fila G_F y de columna G_C en cada punto se obtienen mediante la convolución de la imagen con sendas máscaras H_F y H_C , esto es:

$$G_F(i,j) = F(i,j) \otimes H_F(i,j) \quad (15)$$

$$G_C(i,j) = F(i,j) \otimes H_C(i,j) \quad (16)$$

donde $H_F(i,j)$ y $H_C(i,j)$ son las respuestas impulsionales del gradiente de fila y de columna, respectivamente.

Más utilizados que estos operadores, son los de **Roberts, Prewitt y Sobel**, de los que se hablará a continuación.

Operador de Roberts

La idea del operador de Roberts [Roberts65] es obtener una buena respuesta ante bordes diagonales a partir de la diferencia entre pares diagonales de píxeles. En este caso, las máscaras no representan las derivadas a lo largo de filas y columnas, sino en dos direcciones diagonales (45°) perpendiculares entre sí. Aunque esta circunstancia no afecta a la expresión para determinar el módulo del gradiente, sí afecta a la orientación, que en este caso viene dada por la expresión:

$$\alpha(i, j) = \frac{\pi}{4} + \arctan\left(\frac{G_2(i, j)}{G_1(i, j)}\right) \quad (17)$$

$$\text{donde } G_1(i, j) = F(i, j) - F(i + 1, j + 1) \quad (18)$$

$$G_2(i, j) = F(i + 1, j) - F(i, j + 1) \quad (19)$$

El operador de Roberts ofrece unas buenas prestaciones en cuanto a localización. El gran inconveniente es su extremada sensibilidad al ruido y, por tanto, con unas pobres cualidades de detección.

La Figura 35 a y b nos muestran la imagen antes y después de la aplicación del operador de Roberts.



Figura 35. a) Imagen Original de Yoda. b) Imagen tras aplicar el operador de Roberts

Operadores de Prewitt, Sobel y Frei- Chen

Los operadores de Prewitt, Sobel [Vicent+09] y Frei Chen intentan aliviar la deficiencia del operador de Roberts involucrando en la convolución una mayor número de píxeles del entorno de vecindad. Estos tres operadores pueden formularse de forma conjunta mediante las máscaras de convolución mostradas a continuación. Se puede observar que los gradientes de fila y columna están normalizados para proporcionar una ganancia unidad sobre la zona de pesos positivos y negativos.

$$\frac{1}{2+K} \begin{bmatrix} 1 & 0 & -1 \\ K & 0 & -K \\ 1 & 0 & -1 \end{bmatrix} \text{ Gradiente Fila} \quad (20)$$

$$\frac{1}{2+K} \begin{bmatrix} -1 & -K & -1 \\ 0 & 0 & 0 \\ 1 & K & 1 \end{bmatrix} \text{ Gradiente Columna} \quad (21)$$

El operador de **Prewitt** ($K=1$) es, en cierta forma, similar al de diferencia de píxeles separados introducido anteriormente, donde se involucran a los vecinos de filas/columnas adyacentes para proporcionar mayor inmunidad al ruido.

El operador de **Sobel** se distingue del operador de Prewitt en que se duplican los valores de los píxeles al norte, sur, este y oeste, es decir, $K=2$. De esta forma, cada píxel del entorno de vecindad es ponderado de acuerdo con la distancia de éste al central (píxel evaluado). Como consecuencia de ello, el Sobel es más sensible que el Prewitt a los bordes diagonales, mientras que éste lo es para los horizontales y verticales. En la práctica, sin embargo, no se aprecia gran diferencia entre ellos.

Por otra parte, **Frei y Chen** propusieron un valor de $K = \sqrt{2}$ con el fin de que el gradiente sea el mismo para bordes horizontales, verticales y diagonales.

Estos operadores 3×3 tienen unas prestaciones bastante similares. Particularmente, los dos primeros (Prewitt y Sobel) son muy populares, debido a que los valores de los coeficientes hacen que las operaciones aritméticas se puedan programar de una manera eficiente o implementar fácilmente en hardware.

Influencia del tamaño del operador

El tamaño del operador tiene una gran importancia en lo que concierne a sus características de detección y localización. Los operadores que se han presentado hasta ahora en este trabajo (2×2 y 3×3) y, por consiguiente, con un comportamiento deficiente en cuanto a la detección. Este problema puede mejorarse utilizando máscaras de mayor tamaño que actúan promediando sobre un mayor número de píxeles del entorno.

Las contrapartidas de utilizar máscaras más grandes son dos:

- El suavizado de la imagen conlleva la atenuación y enturbiamiento de los bordes y, por consiguiente, una peor localización de estos.
- El coste computacional es mucho mayor que en operadores pequeños ($O(N \times N)$, siendo N el tamaño de la máscara)

Como regla general, si el operador es pequeño la localización es más precisa, aunque es más sensible al ruido y por ello detectará bordes que no existen. Si el operador es grande entonces la localización es menos precisa, sin embargo, disminuye la sensibilidad al ruido y mejora la detección.

C. Operadores basados en la segunda derivada (Laplaciana)

Siempre que en la imagen se presenta un cambio de intensidades a lo largo de una determinada dirección, existirá un máximo en la primera derivada a lo largo de dicha dirección, y consecuentemente, un paso por cero en la segunda derivada. La dirección de interés será aquella ortogonal a la orientación local de los cruces por cero.

Puesto que, generalmente esta dirección no es conocida, parece claro que la utilización de un operador independiente de la orientación como la Laplaciana presenta evidentes ventajas computacionales. Asimismo, al ser una derivada segunda, tiene la ventaja de facilitar la localización precisa del borde.

En el dominio continuo la Laplaciana del borde de una función bidimensional $f(x,y)$ se define como:

$$\nabla^2 f(x,y) = \frac{\partial^2}{\partial x^2} f(x,y) + \frac{\partial^2}{\partial y^2} f(x,y) \quad (22)$$

La Laplaciana vale cero si $f(x,y)$ es constante o cambia linealmente su amplitud. El cambio de signo de la función resultante indica que en ese lugar existe un cruce por cero (cambio de positivo a negativo o viceversa) y, por tanto, indica la presencia del borde.

El método de Marr y Hildreth

Según los estudios realizados por Marr y Hildreth [Marr+80], los cambios de intensidad en imágenes reales tiene lugar sobre un amplio rango de escalas o resoluciones, de tal manera que un único filtro no puede actuar de forma

óptima en todas las escalas simultáneamente. Por ello se propone el suavizado previo de la imagen a distintas escalas para, luego, detectar los cambios de intensidad en cada uno de estos niveles. Esta propuesta está fundada en numerosas experiencias con seres vivos, donde parece ser que éste sería el procedimiento básico de detección de bordes y extracción de características en el primer nivel del proceso de visión.

En el suavizado de la imagen se deben tener en consideración dos circunstancias distintas:

- El suavizado de la imagen debe reducir el rango de escalas que pueden tener lugar los cambios de intensidad. Esto significa que en el dominio de frecuencias el filtro de suavizado debe tener un ancho de banda limitado, es decir, su actuación debe estar suficientemente localizada en frecuencias.
- Las discontinuidades de una imagen se originan por motivos bien diversos: cambios en la iluminación (sombras, fuentes visibles de luz, gradientes de iluminación, etc), distinta posición (distancia y/o orientación) del observador con respecto a las superficies visibles, y distintas propiedades reflectantes de las superficies. En el dominio espacial, todos estos cambios se producen, o pueden producirse, a distintas escalas espaciales, características de cada uno de estos fenómenos.

Por sus características de localización, la labor de detección de la discontinuidad recae en la segunda derivada (detección de pasos por cero), en particular, en la Laplaciana, por ser un operador diferencial de segundo orden independiente de la orientación del borde.

El método propuesto por Marr y Hildreth afirma que, analizando los cruces por cero a distintas escalas pueden inferirse características sobre la naturaleza de la discontinuidad, y si ésta es ocasionada por un único fenómeno o por varios a la

vez. Las características que se consideran son, además de los cruces por cero, “manchas”, terminaciones, segmentos, grupos, etc. Estas características dan lugar a una estructura, denominada “the raw primal sketch” que representa no solo cambios de niveles de intensidad, sino también su distribución y organización en la imagen. En la práctica, este operador es equivalente al propuesto por Canny.

Operador de Canny

El operador de Canny [Canny86] está basado en un desarrollo analítico de optimización partiendo de un modelo continuo unidimensional de un escalón.

Considérese una función escalón con amplitud “ h_E ” afectado por un ruido blanco gaussiano de desviación típica σ_n . Supóngase que la detección de bordes se va a llevar a cabo mediante la convolución de una función continua unidimensional $f(x)$ con una función respuesta impulsional antisimétrica $h(x)$, que tiene amplitud cero fuera del intervalo $[-W, W]$. El borde buscado de la función $f(x)$ se marcará donde aparezca el máximo local del gradiente, obtenido tras la convolución de $f(x)$ con $h(x)$. Para determinar la función $h(x)$ buscada se exige que ésta satisfaga los siguientes tres criterios:

- **Buena detección.** Se maximiza la amplitud de la relación señal-ruido (snr) del gradiente para obtener una alta probabilidad de marcar borde donde lo hay y baja probabilidad de marcarlo donde no lo hay. La relación señal-ruido para el modelo considerado es:

$$\text{snr} = \frac{h_E}{\sigma_n} S(h) \quad (23)$$

con

$$S(h) = \frac{\int_{-W}^0 h(x) dx}{\int_{-W}^W [h(x)]^2 dx} \quad (24)$$

- **Buena localización.** Los puntos del borde marcados por el operador deben estar tan cerca del centro del borde como sea posible. El factor de localización se define como:

$$LOC = \frac{h_E}{\sigma_n} L(h) \quad (25)$$

con

$$L(h) = \frac{h'(0)}{\int_{-W}^W [h(x)]^2 dx} \quad (26)$$

- **Respuesta única.** Debe haber una única respuesta para cada borde. La distancia entre picos del gradiente cuando sólo el ruido está presente, denotada por x_m , se establece como una fracción k del ancho del operador, es decir:

$$x_m = kW \quad (27)$$

2.2.4 Otras técnicas evaluadas

Durante el periodo de investigación para el desarrollo de esta tesis se han estudiado y aplicado a las imágenes objeto de estudio multitud de técnicas para intentar resolver los problemas que se planteaban. No obstante, después de ser evaluadas se determinó que no se ajustaban a las necesidades del algoritmo perseguido y fueron finalmente descartadas. A continuación se detallan las técnicas más importantes que pese a ser evaluadas no se han incluido en el algoritmo final de la tesis.

2.2.4.1 Contornos Activos

Estos contornos [Kass+88] modelan las fronteras entre un objeto, el fondo y el resto de objetos de la imagen. Permiten extraer los contornos de los objetos de interés basándonos en modelos o formas que utilizan información previa de la forma de los objetos. Los contornos activos o también llamados Snakes [Allin+04] [Chang+07] son mucho más robustos frente a la presencia de ruido y

otros elementos y permiten segmentar imágenes mucho más complejas, como las imágenes médicas objeto de esta tesis, por ejemplo.

Los resultados de segmentación proporcionados por esta técnica aplicados a las imágenes de cuerdas vocales son bastante positivos, tal y como se puede ver en la Figura 36, pero requieren de inicialización previa, cuestión que se ha querido evitar en el sistema propuesto en esta tesis. Por esta razón esta técnica fue descartada.

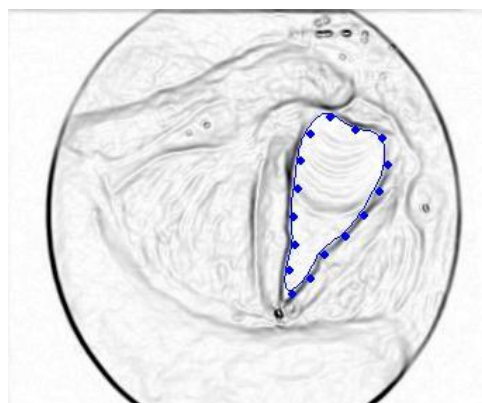


Figura 36. Ejemplo de segmentación realizada aplicando Contornos Activos

2.2.4.2 Transformada Wavelet

Las transformadas Wavelet [Cavalcanti+10] y otras funciones de análisis multiescalar son ampliamente usadas en el procesamiento digital de imagen para aplicaciones de eliminación/reducción de ruido, compresión y extracción de características.

Su aplicación en bioingeniería es muy extendido [Rees+08], y más concretamente en aplicaciones de imágenes médicas como: imágenes de ultrasonidos, imágenes de tomografía, o imágenes de resonancia magnética. Los métodos de captura de estas imágenes y por lo tanto, sus características son muy distintas, a las imágenes de cuerdas vocales objeto de estudio en esta tesis. Todo ello justifica que no hayan sido utilizadas en el diseño del sistema, pero sí evaluadas para el desarrollo de un futuro sistema de telemedicina que requiera de la transmisión de imágenes de alta resolución comprimidas.

2.2.4.3 Filtro de Kalman

En el ámbito del procesamiento digital de imagen el filtro de Kalman es un algoritmo recursivo que se utiliza para estimar la posición de un punto o característica en movimiento y la incertidumbre de la medida, en la siguiente imagen. Se trata de buscar la característica (punto, borde, esquina, región, etc.) en un área determinada de la siguiente imagen alrededor de la posición predicha, en la que estamos seguros de encontrar la característica dentro de un cierto grado de confianza.

El objetivo del filtro es la obtención de un estimador óptimo de las variables de estado de un sistema dinámico, basado en observaciones ruidosas y en un modelo de incertidumbre de la dinámica del sistema [Erturk02] [Di+08].

La Figura 37 muestra una aproximación del seguimiento del movimiento de las cuerdas vocales aplicando el filtro de Kalman, pero tomando como referencia siempre el rectángulo que contiene el espacio glótico, no el espacio glótico en sí, o las cuerdas vocales por separado, que hubiese sido lo deseable, y que se deja para futuras investigaciones.

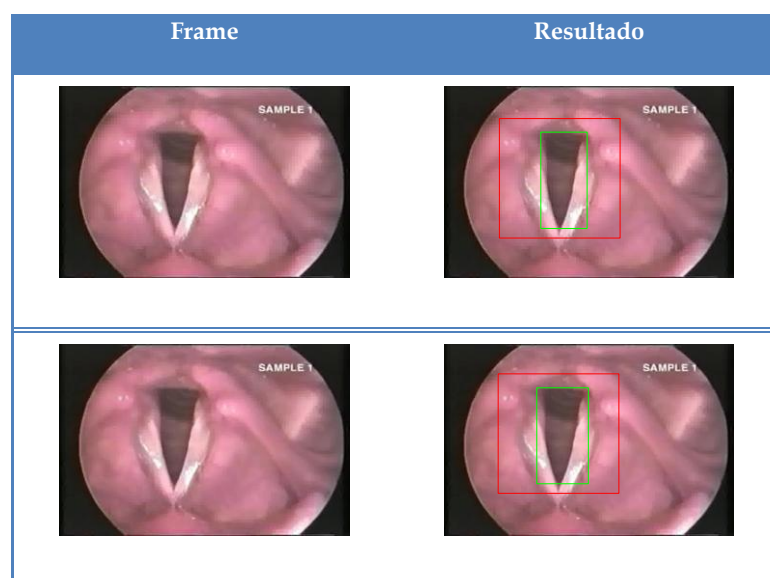


Figura 37. Aplicación del Filtro de Kalman al seguimiento del movimiento de las cuerdas vocales

The photographic image... is a message without a code.

Roland Barthes

3.BASE DE DATOS

Este es el capítulo dedicado a la descripción de la base de datos de imágenes de cuerdas vocales utilizada durante la realización de esta tesis, y principalmente, para realizar las pruebas de los algoritmos desarrollados.

Durante el desarrollo han sido varios los inconvenientes encontrados para dotar de contenido esta base de datos. Por un lado, la variabilidad de dispositivos de captura disponibles en el mercado es muy elevada, y por lo tanto su sensibilidad, funcionalidades aportadas, y en consecuencia, la calidad que se puede alcanzar en las capturas. Por otro lado, la experiencia del especialista en la técnica de grabación de imágenes estroboscópicas es muy importante y afecta directamente a la calidad de las imágenes capturadas.

Finalmente, se cuenta con una amplia base de datos que consta de las grabaciones realizadas por el Doctor Agustín Pérez Izquierdo en el Hospital de Basurto y Clínica Guimón, y de las grabaciones que contiene una base de datos

comercial “Laryngeal Videostroboscopic Images” (Dr. Wendy LeBorgne; Plural Publishing) [LeBorgne06].

3.1 DESCRIPCIÓN DE LA BASE DE DATOS

Las imágenes grabadas por el Doctor Pérez Izquierdo han sido adquiridas con ATMOS Endo-Stroboscope L, mientras que se desconoce la instrumentación utilizada en las grabaciones de la base de datos comercial.

Por lo tanto, se dispone de una base de datos de 1262 imágenes provenientes de 45 grabaciones de diferentes pacientes con cuerdas vocales sanas y patológicas, principalmente patologías benignas.

Entre los 45 pacientes suponemos que habrá grabaciones de pacientes hombres y mujeres, aunque no podemos precisar con exactitud cuántos de cada sexo, porque no se dispone de esa información de las capturas de la base de datos comercial.

A continuación se describe el detalle de las grabaciones:

- 15 secuencias que pertenecen a la base de datos comercial “Laryngeal Videostroboscopic Images” (Dr. Wendy LeBorgne; Plural Publishing):
 - 7 secuencias con imágenes de cuerdas vocales sanas
 - 2 secuencias con imágenes de cuerdas vocales con nódulos
 - 2 secuencias con imágenes de cuerdas vocales con pólipos
 - 2 secuencias con imágenes de cuerdas vocales con parálisis
 - 1 secuencia con imágenes de cuerdas vocales con edema
 - 1 secuencia con imágenes de cuerdas vocales con quiste
- 30 secuencias proporcionadas por el Dr. Agustín Pérez Izquierdo del Hospital de Basurto:
 - 3 secuencias con imágenes de cuerdas vocales sanas
 - 8 secuencias con imágenes de cuerdas vocales con nódulos

- 7 secuencias con imágenes de cuerdas vocales con pólipos
- 5 secuencias con imágenes de cuerdas vocales con parálisis
- 3 secuencias con imágenes de cuerdas vocales con edema
- 4 secuencias con imágenes de cuerdas vocales con quiste

3.2 ETIQUETADO DE LA BASE DE DATOS

Todas las secuencias e imágenes de la base de datos han sido etiquetadas y nombradas según los siguientes criterios:

- Etiqueta identificando el tipo de patología, en caso de que la haya
- Número de secuencia
- Número de frame

La nomenclatura dada a las secuencias de cada patología se puede ver en la Tabla 3.

Diagnóstico	Etiqueta
Sana	NOR
Nódulo	N
Pólipo	P
Edema	PD
Quiste	Q
Parálisis	PA

Tabla 3. Etiquetado de las secuencias de la Base de Datos

Siguiendo el esquema planteado, el frame 1 de la secuencia 1, de las diagnosticadas sanas estaría etiquetado como: NOR1- 1.

Físicamente en el disco duro, se ha generado una estructura de archivos. Cada patología tiene su propia carpeta que contiene todas las secuencias así diagnosticadas. Y de cada una de ellas, se guarda el vídeo, los frames extraídos por el sistema planteado y la segmentación de cada uno.

La parte de la base de datos generada para esta tesis por el otorrinolaringólogo Agustín Pérez Izquierdo ha sido registrada como propiedad intelectual, dada la dificultad que se tiene en estos casos para disponer de imágenes.

“The mind loves the unknown. It loves images whose meaning is unknown, since the meaning of the mind itself is unknown”.

Rene Magritte

4.DISEÑO

En este capítulo se detalla el diseño de las etapas del sistema y de cada uno de los algoritmos de procesamiento digital de imagen que se han implementado para la solución presentada.

Como planteamiento general de diseño se ha considerado que todos los módulos/funciones de la solución funcionen en una única capa de middleware, que pueda ser accedida desde cualquier aplicación que desee proveerse de las funcionalidades desarrolladas, e interactúe con ellas.

Por otro lado, todos los módulos interactúan con cualquier capa inferior que le provea de imágenes, en este caso, la capa de Base de Datos, cuyo detalle se ha descrito en el capítulo 3. Este es el enfoque que se ha considerado óptimo y que se ha seguido a lo largo de todas las etapas (ver Figura 38).

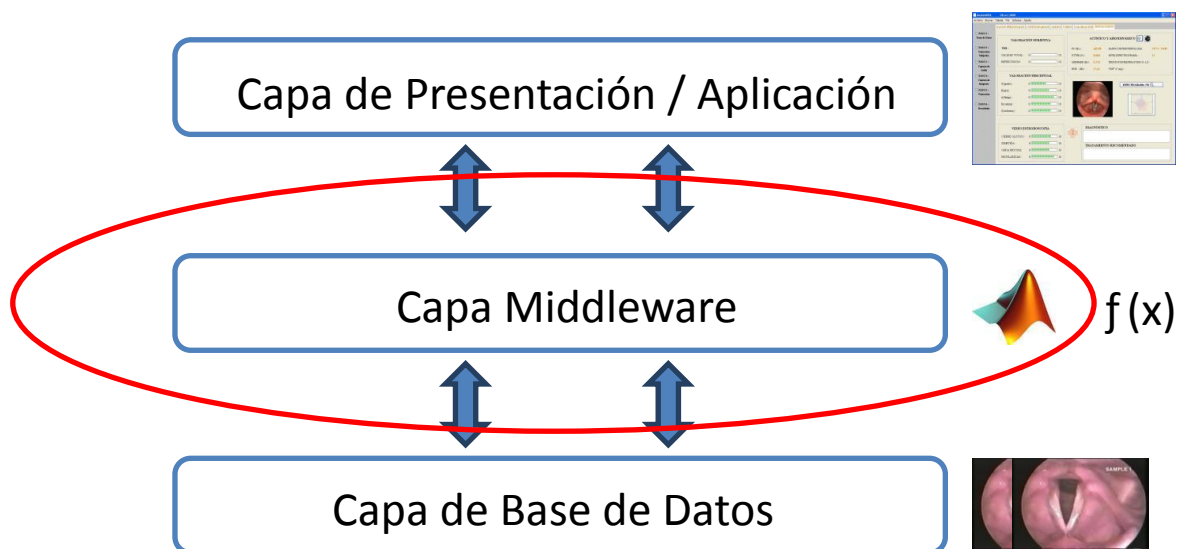


Figura 38. Diagrama de Capas del Sistema propuesto

Los algoritmos desarrollados, pertenecientes a la capa Middleware, normalizan, procesan, analizan y extraen características de distintas patologías de las cuerdas vocales. Principalmente patologías que afectan a la morfología de la cuerda vocal y que al menos, desembocan en disfonías, y en otros casos, patologías provocadas por un funcionamiento anormal en el movimiento y vibración de las cuerdas vocales, como puede ser la parálisis de cuerda vocal.

La capa de Aplicación se ocupa de la presentación de los datos al usuario en un interface gráfico amigable.

Por ello, se ha establecido un orden secuencial para la realización del estudio y la aplicación de los algoritmos (ver Figura 39). Se considera etapa 0, a la captura de las imágenes por parte del especialista, según el método descrito en la sección 2.2.1.4 de esta tesis. Una vez que se dispone de las imágenes, nos centramos en la extracción de todos los parámetros/características de las secuencias de imágenes que nos van a dar información sobre la patología o ausencia de ella que padece el paciente. Posteriormente, tratamos de clasificar/identificar la imagen comparando con la base de datos (con imágenes previamente clasificadas y diagnosticadas) de la que se dispone. Finalmente,

con toda la información se puede ofrecer un diagnóstico al especialista, apoyado en todos los parámetros objetivos extraídos durante el proceso.

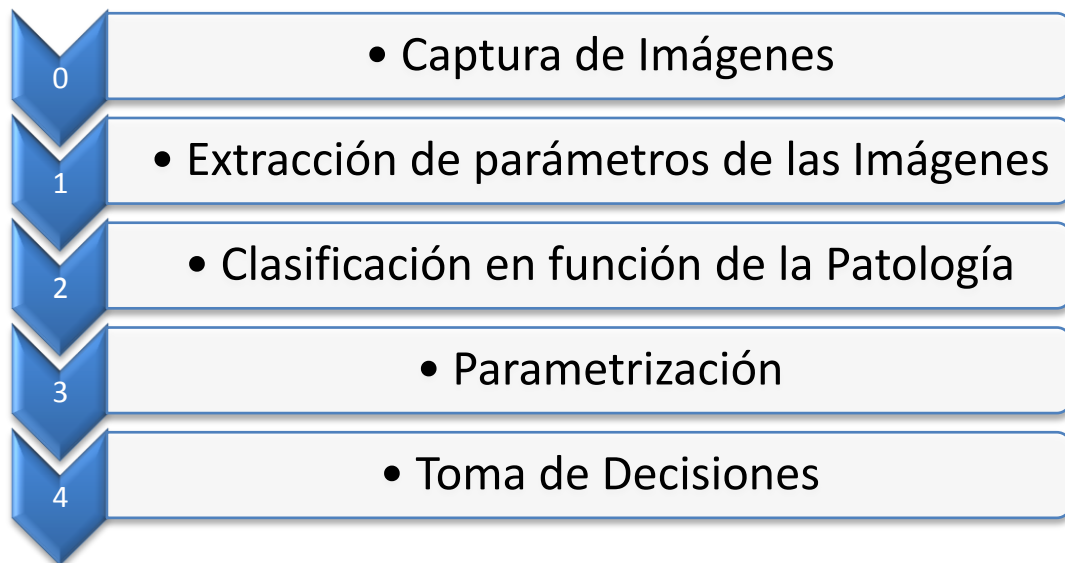


Figura 39. Estudio secuencial realizado en el diseño de los algoritmos de la capa de Middleware

La metodología seguida durante todo el proceso de diseño se ha mantenido constante, definiendo en cada caso los requerimientos y variables de entrada y salida de cada uno de los bloques de las etapas definidas, tal y como se describe en el punto siguiente. A partir de este momento, el diseño de los algoritmos se ha centrado principalmente en dar respuesta a los tres primeros pasos descritos en la Figura 39.

4.1 DISEÑO DE ALTO NIVEL

El Diseño de Alto Nivel establece fondo y forma del sistema considerándolo como un todo, como un conjunto de funciones que constituyen la estructura o arquitectura del sistema sin entrar en el detalle de cada una, ya que esto se hará en el Diseño detallado de bajo nivel.

En todas las aplicaciones, salvo en las más pequeñas, el primer paso para diseñar un sistema completo consiste en dividirlo en un pequeño número de componentes, bloques o etapas. Cada uno de los bloques principales de un

sistema abarca aspectos del sistema que comparten alguna propiedad común. Cada uno de los bloques o etapas comprende un paquete de funciones y sucesos interrelacionados que comparten un propósito común, que tienen un interface bien definido con los demás bloques o etapas (y que habitualmente pueden ser reutilizados en varios sistemas o aplicaciones).

El diagrama de bloques general propuesto para alcanzar los objetivos marcados se puede observar en la Figura 40.

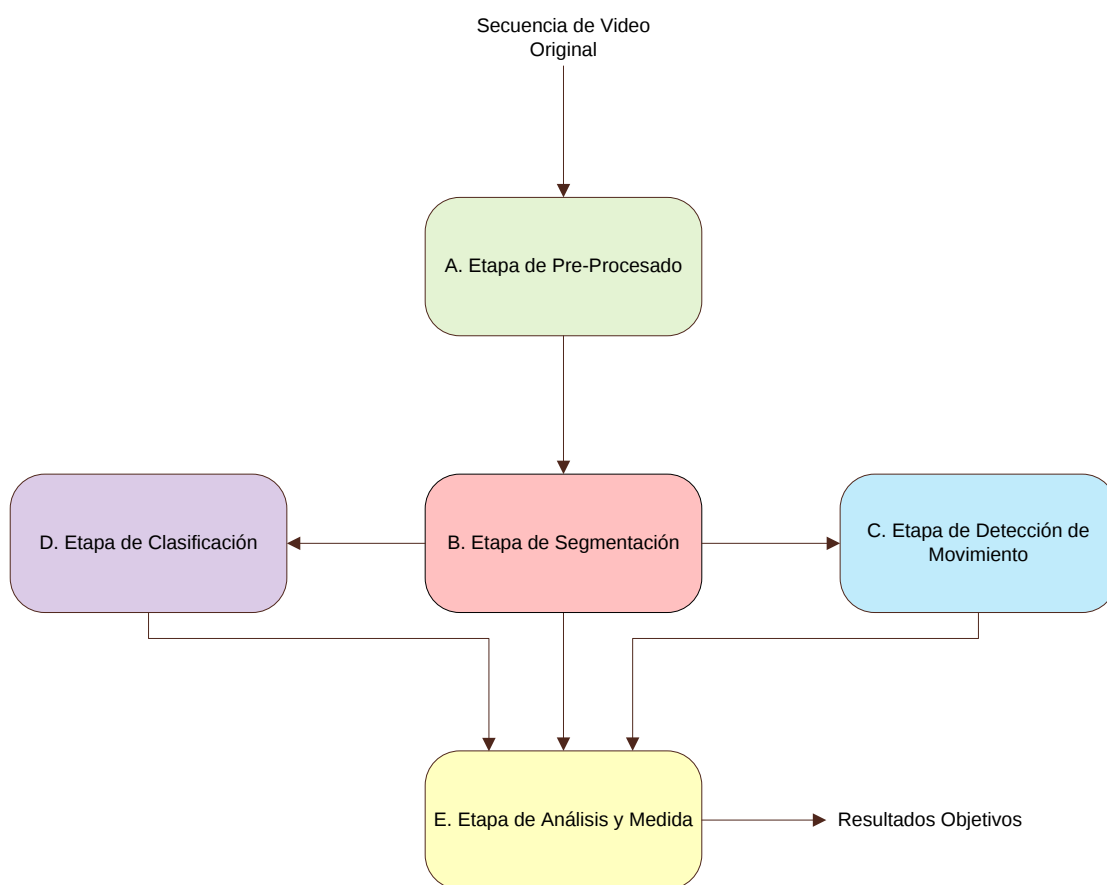


Figura 40. Diagrama de Alto Nivel

La entrada del sistema son los frames de cada una de las secuencias en formato original (en color, y con la calidad y resolución adquiridas por la cámara). Cada uno de los bloques aplica las transformaciones necesarias a las imágenes para obtener como resultado final un informe con datos objetivos que ayude y apoye

al otorrinolaringólogo en sus diagnósticos, y a evaluar procesos de rehabilitación, o la evolución tras una intervención quirúrgica.

A continuación, se describen los cinco grandes bloques que comprenden el diseño de alto nivel (cada uno identificado con un código de colores que se mantendrá a lo largo del capítulo, como se puede observar en la Figura 40).

4.1.1 Etapa de Pre-Procesado

Este primer bloque realiza las funciones necesarias unificación, normalización y estandarización de las características de los videos estroboscópicos de cuerdas vocales o imágenes de cuerdas vocales, para que puedan ser tratados con mayor sencillez y eficiencia en los bloques posteriores.

El *input* de este bloque es la grabación original proporcionada por el otorrinolaringólogo. También se puede analizar una imagen aislada, pero los resultados en algunas de las etapas tienen una fiabilidad menor. En algunas etapas es necesaria la comparación/el estudio de ciertos parámetros entre frames dentro de la misma secuencia. Hay que tener en cuenta la variabilidad de la grabación intraframe, por lo tanto, la comparación es una utilidad que proporciona mucha información.

El *output* de este bloque es una secuencia de frames en formato JPG normalizados en características y transformados a escala de grises, cuyo gradiente ha sido calculado (para facilitar la aplicación de los filtros de la etapa posterior).

4.1.2 Etapa de Segmentación

Por su criticidad, ésta es la etapa principal del sistema. De su éxito, depende en parte, la validez del sistema y el porcentaje de error obtenido en las etapas posteriores. Las tres etapas posteriores están basadas (e incluso utilizan como input) en el resultado de la etapa de segmentación.

Durante esta etapa se realizan las transformaciones necesarias para segmentar y aislar, en este caso, la región de interés. En este sistema la ROI con la que se va a trabajar a partir de este momento es el espacio glótico.

El *input* de este bloque son las imágenes gradiente resultado de la normalización realizada en el bloque A.

El *output* de este bloque es la secuencia de frames en los que el espacio glótico ha sido segmentado. Desde el bloque B, se presta atención principalmente a esta parte de la imagen y al estudio de las características del perímetro del espacio glótico.

4.1.3 Etapa de Detección del Movimiento

En esta etapa se aplican algoritmos de Block Matching, habitualmente utilizados para la compresión de imágenes en los estándares MPEG, y muy extendidos en multitud de aplicaciones [Goffredo+06]. El objetivo es detectar y estudiar el movimiento interframe de las cuerdas vocales.

Asumiendo siempre que las imágenes estroboscópicas **no son las más adecuadas para evaluar el movimiento** ni las vibraciones de las cuerdas, ya que el frame rate está entre 25 y 30 frames por segundo, muy inferior a la velocidad de vibración de las cuerdas vocales. Por esta razón, los resultados de esta etapa son relativos y tienen que ser completados con las medidas de la etapa E, e incluso otros estudios.

El *input* de este bloque es la secuencia de frames cuyo espacio glótico ha sido segmentado en el bloque B.

El *output* de este bloque son todos los vectores de movimiento calculados como fruto de la aplicación de los algoritmos de Block Matching.

4.1.4 Etapa de Clasificación

Con las etapas descritas hasta el momento, no se ha obtenido ningún resultado que nos pueda orientar en el diagnóstico, únicamente se ha extraído la región de interés y medidas de movimiento. Es en esta etapa, donde se realiza un pre-diagnóstico en función de los datos obtenidos en las etapas B y C. Se realiza una clasificación de las imágenes de entrada a través de varios algoritmos, para discernir entre patologías morfológicas y no morfológicas, comparando la entrada de esta etapa con una base de datos de imágenes previamente clasificadas y estudiadas. Los resultados de esta etapa, parcialmente, dependen de la elección y el tamaño de esa base de datos usada para comparar los input de esta etapa.

El bloque D tiene varios *input* a diferencia de los bloques anteriores. Para la ejecución de este bloque se necesitan tres inputs: la secuencia de frames originales (no el video, el resultado del primer bloque de la etapa A, como se puede ver en la Figura 43), la secuencia de frames con el espacio glótico segmentado (resultado del bloque B), y los vectores de movimiento calculados en el bloque C (output del bloque C).

El *output* de este bloque es un pre-diagnóstico/clasificación, en función de los algoritmos de identificación/clasificación aplicados.

4.1.5 Etapa de Análisis y Medida

Es en esta última etapa cuando finalmente se hacen los cálculos necesarios para llegar a unos resultados objetivos que nos permiten realizar una evaluación y un posible diagnóstico, e incluso discernir qué patología morfológica de las cuerdas vocales sufre el paciente. En el caso de las patologías relacionadas con el movimiento de las cuerdas vocales se aportarán parámetros que orienten al especialista hacia un estudio más profundo aportándole resultados objetivos,

pero sabiendo los resultados aportados puede que no sean 100% consistentes debido el frame rate de la captura disponible.

Realizadas las transformaciones necesarias, para llegar a esos resultados, se realizan medidas objetivas sobre las imágenes de cuerdas vocales, y es en su interpretación en lo que se basa la evaluación y/o el diagnóstico, y el valor de esta aportación.

El *input* de este bloque como el del bloque D es múltiple. Para calcular los parámetros objetivos que serán el resultado final de esta tesis, es necesario hacer cálculos y transformaciones sobre los resultados de los bloques B, C y D.

El *output* de este bloque será el conjunto de medidas objetivas que apoyen el diagnóstico final propuesto al especialista, que pueden formar parte del informe final entregado al paciente.

La novedad, complejidad y principal característica del sistema propuesto, es que no requiere de ningún tipo de inicialización, ni de interacción con el usuario durante su ejecución para la obtención del diagnóstico, evaluación y/o la medida de la efectividad del tratamiento (pre-diagnóstico asistido por ordenador).

4.2 DISEÑO DE BAJO NIVEL

En este apartado se describen con detalle cada uno de los módulos que componen las cinco etapas descritas anteriormente. En cada una de ellas, se detallan las transformaciones realizadas y el diseño de los algoritmos aplicados sobre los videos o secuencia de imágenes estroboscópicas, y la interacción entre cada uno de los módulos (detalle de entradas y salidas).

4.2.1 Etapa de Pre-Procesado

La falta de calidad y uniformidad de los videos estroboscópicos de cuerdas vocales (ver Figura 41) es una de las características de todas las grabaciones de la base de datos utilizada para las pruebas de este sistema.

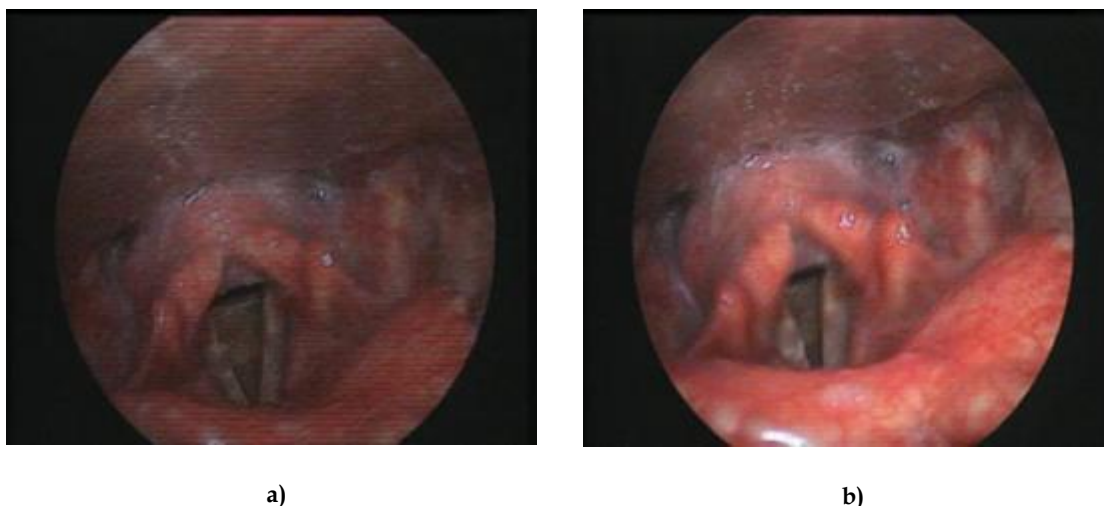


Figura 41. Ejemplos de imágenes de poca calidad

Los vídeos estroboscópicos son grabados por el especialista en otorrinolaringología en una sesión de consulta y exploración con el paciente. Por esta razón, la calidad de los vídeos depende de:

- El estroboscopio utilizado. La calidad de las grabaciones varía dependiendo de la gama del dispositivo utilizado.
- La experiencia del especialista en otorrinolaringología para hacer este tipo de exploraciones y grabaciones.

- El estado del paciente. Es decir, si el paciente permite con tranquilidad la exploración, se mueve mucho, está nervioso...

La calidad de las grabaciones varían mucho, incluso dentro de la propia grabación la calidad de los frames puede variar. Como ejemplo se pueden observar la Figura 41, que son dos frames que pertenecen a la misma secuencia y en los que se puede ver la diferente luminosidad que representan, o como algunos elementos de la imagen varían o cambian de ubicación (en este caso, el principal elemento es el espacio glótico).

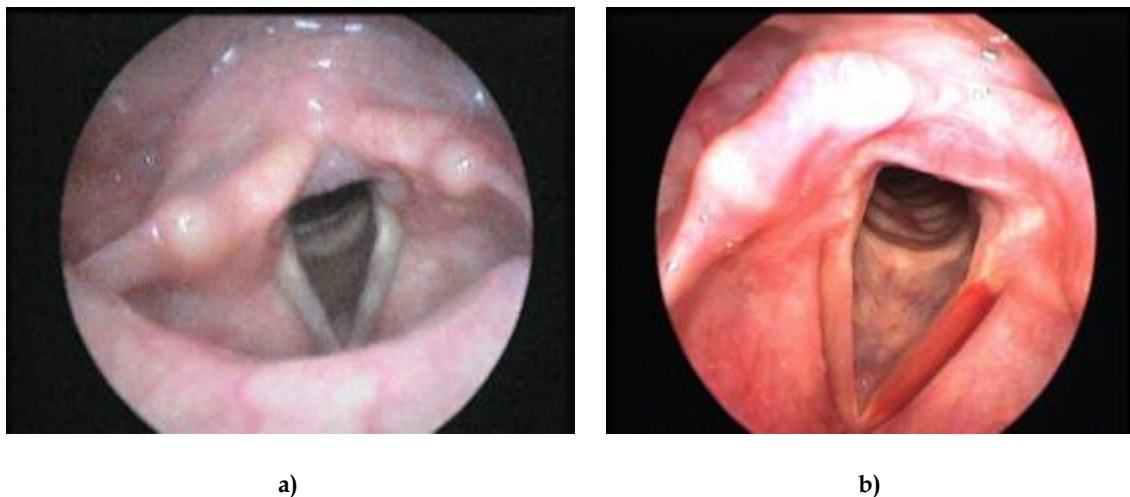


Figura 42. a) Ejemplo de imagen con luz blanquecina. b) Ejemplo de Imagen con buena calidad

La Figura 42 muestran dos frames de grabaciones distintas, realizadas con dispositivo, zoom e iluminación diferente. La Figura 42 b es un ejemplo de lo que se va a considerar a partir de ahora, imagen de buena calidad.

Por todo ello, es necesaria una etapa de pre-procesado para estudiar los parámetros previamente comentados y otros que se enumeran a continuación:

- Zoom utilizado (o no) por el otorrinolaringólogo.
- Iluminación utilizada durante la captura. Habitualmente las cámaras de los estroboscopios incluyen iluminación, dada la cavidad para la que se

le va a dar uso. El resultado se ve muy afectado si la luz usada es blanquecina.

- Brillos y artefactos que se generan en los frames y que dificultan la posterior segmentación de las regiones de interés.
- Intensidad de color de los frames capturados.

La etapa de Pre-Procesado (cuyo diagrama de bloques se puede ver en la Figura 43) consta de cuatro bloques principales para realizar las funciones de unificación y normalización de las características principales de las imágenes.

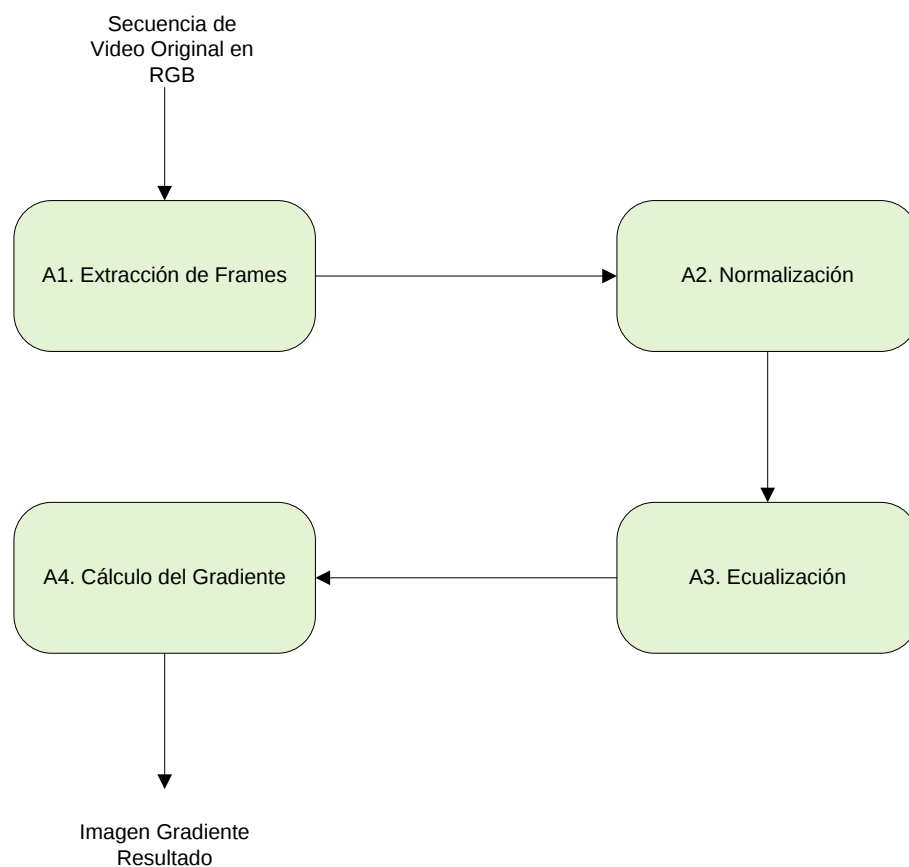


Figura 43. Diagrama de Bloques de la etapa de Pre-Procesado

A1. Extracción de Frames.

El primer bloque (A1 en la Figura 43) se centra en extraer todos los frames ($I(x,y)$) y guardar las imágenes de forma independiente. Este paso se realiza para trabajar con cada uno de ellas.

El *input* del bloque A1 (ver Figura 44) es un video digital con imágenes a color ($f(x,y,k)$ siendo k cada uno de los frames que componen el video). De él se elimina la parte de la grabación correspondiente al audio, que no es objeto de este trabajo, y que suele estar distorsionada debido a que los pacientes tienen introducida en su boca el estroboscopio.

El *output* del bloque A1 (ver Figura 44) son k imágenes en formato RGB, correspondientes a cada uno de los frames de la grabación. A partir de este bloque se hablará de frames independientes, no de K frames, ya que del bloque A1 saldrán los frames de la secuencia de forma secuencial, y en cada una de las etapas siguientes, se hará un procesamiento por frame, aunque el análisis final tenga en cuenta el resultado de toda la secuencia.

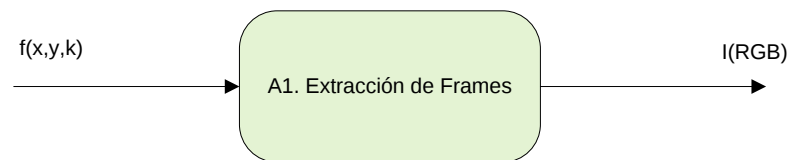


Figura 44. Bloque A1 de la Etapa de Pre-Procesado

$$\begin{pmatrix} R \\ G \\ B \end{pmatrix} = \begin{bmatrix} I_R(x,y) \\ I_G(x,y) \\ I_B(x,y) \end{bmatrix} \quad (28)$$

A2. Normalización. Paso a escala de grises.

Para aumentar la simplicidad y eficiencia de los algoritmos propuestos, y puesto que la información referente al color (RGB) no es determinante para el

estudio y diagnóstico de patologías morfológicas de cuerdas vocales, se va a trabajar a partir de esta etapa en escala de grises.

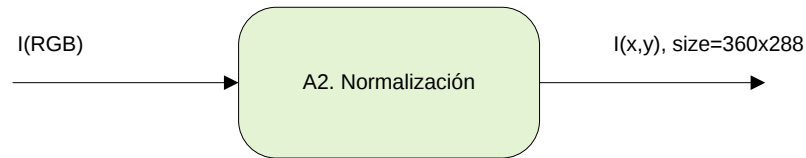


Figura 45. Bloque A2 de la etapa de Pre-Procesado

El *input* de este módulo son las imágenes que forman la secuencia introducida en el sistema (ver Figura 45).

El *output* de este módulo son las imágenes normalizadas, en escala de grises y con la resolución predeterminada.

Cada una de las imágenes está en formato de color RGB y las pasamos a escala de grises mediante la siguiente combinación lineal, según los criterios usados en televisión digital:

$$I(x, y) = w_R I_R(x, y) + w_G I_G(x, y) + w_B I_B(x, y) \quad (29)$$

Siendo $w_R=0,2989$, $w_G=0,5870$, y $w_B=0,1140$. Esta transformación la podemos observar en los ejemplos de la Figura 46.

Por último, y para que puedan ser comparadas las imágenes es indispensable que todas ellas tengan la misma resolución. El algoritmo diseñado es independiente de la resolución, y dentro de un mismo video todas tienen la misma, pero para poder comparar áreas, longitudes y otros parámetros (en etapas posteriores) dentro de un estudio y utilizando toda la base de datos, es necesario una referencia.

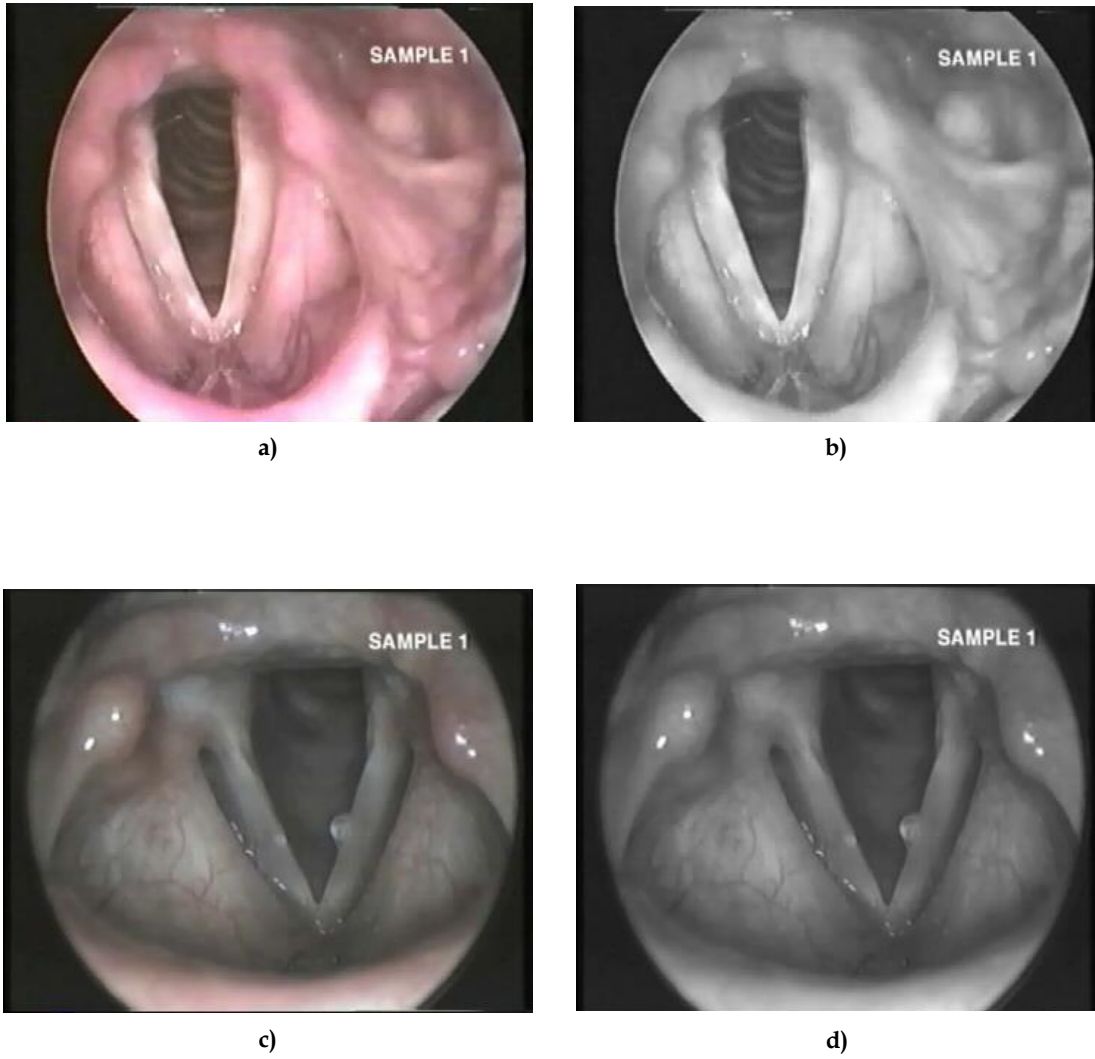


Figura 46. a) Imagen Original en formato RGB. b) Imagen en escala de grises. c) Imagen Original con pólipo, formato RGB. d) Imagen con pólipo en escala de grises.

La resolución escogida para trabajar es 360x288 y se calcula según un factor de corrección como se puede ver en la ecuación (30).

$$\begin{bmatrix} i' \\ j' \\ 1 \end{bmatrix} = \begin{bmatrix} S_i & 0 & 0 \\ 0 & S_j & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} i \\ j \\ 1 \end{bmatrix} \quad (30)$$

A3. Ecualización.

El hecho de pasar la imagen a escala de grises no soluciona los problemas de brillo e iluminación previamente comentados. Es más, dentro de un mismo

video se pueden encontrar diferencias significativas de estos parámetros, por eso es importante unificar las imágenes.

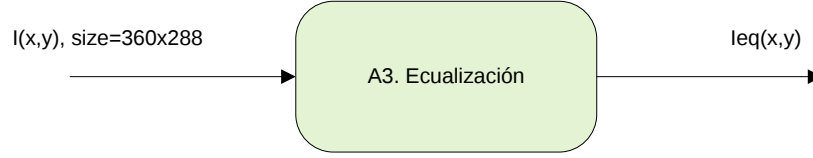


Figura 47. Bloque A3 de la Etapa de Pre-Procesado

El *input* de este módulo son las imágenes normalizadas, en escala de grises y con la resolución 360x288 (ver Figura 47).

El *output* de este módulo son las imágenes tras aplicarles la función de ecualización.

Para ello, se va a utilizar la función de ecualización f_{eq} (ecuación (31)) para obtener un histograma uniforme en todas ellas, es decir, una distribución equilibrada de todos los niveles de intensidad (a nivel de aproximación dado que se trata de distribuciones discretas).

El resultado de la ecualización maximiza el contraste de una imagen sin perder información de tipo estructural.

$$f_{eq}(p) = \left\lceil H(p) \cdot \frac{K-1}{MN} \right\rceil$$

M y N representan la resolución de la imagen en el intervalo $[0 \dots K-1]$ (31)

$H(p)$ es la función del histograma acumulativo

$$H(ip) = \sum_{j=0}^i h(j) \text{ para } 0 \leq i \leq K$$

Una vez que se dispone de la imagen en escala de grises y ecualizada se procede a calcular la imagen gradiente de cada frame, hecho que favorecerá el buen resultado de la posterior segmentación.

A4. Cálculo del Gradiente

Tras pasar las imágenes de cuerdas vocales a escala de grises, normalizarlas y ecualizarlas, se calcula el gradiente de cada una de ellas (ver Figura 48).

El *input* del módulo A4 son las imágenes ecualizadas resultado del bloque A3.

El *output* de este módulo son las imágenes con el gradiente calculado en cada una de ellas.

El método usado para el cálculo del gradiente, es el operador de Sobel [Jähne+99] en su aproximación tradicional (ver ecuación (32)). Tal y como se ha descrito en la sección 2.4.2.2, el operador de Sobel es más sensible a los bordes diagonales, que son los que más se asemejan a los bordes de la región de interés buscada.

Se utiliza un operador de Sobel de máscara 3x3, asumiendo que es más sensible al ruido que utilizar máscaras mayores, pero que realizará una localización más precisa de los bordes.

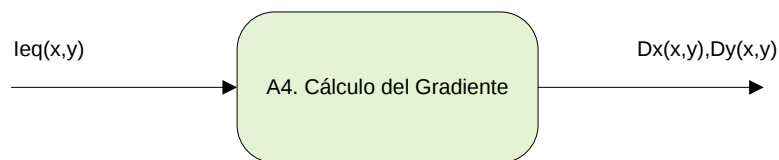


Figura 48. Bloque A4 de la etapa de Pre-Procesado

$$H_x^S = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix} \text{ y } H_y^S = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix} \quad (32)$$

Para obtener la magnitud gradiente, se calculan las aproximaciones de la derivada vertical y horizontal (ver ecuación (33)).

$$\nabla I(x, y) \approx \frac{1}{6} \begin{bmatrix} H_x^S & I \\ H_y^S & I \end{bmatrix} y \quad (33)$$

Se caracterizan los resultados de los filtros para cada uno de los diferentes sentidos como (34).

$$D_x(x, y) = H_x * I \text{ y } D_y(x, y) = H_y * I \quad (34)$$

La magnitud del borde es en ambos casos definida como la magnitud del gradiente:

$$E(x, y) = \sqrt{(D_x(x, y))^2 + (D_y(x, y))^2} \quad (35)$$

Y la dirección del gradiente en cada píxel es calculado a partir de:

$$\phi(x, y) = \tan^{-1} \left(\frac{D_y(x, y)}{D_x(x, y)} \right) \quad (36)$$

Para finalizar la etapa de pre-procesado, y previo a la etapa de segmentación, se aplica una transformación morfológica básica: Erosión, cuya definición es “El mínimo de la resta realizada entre los valores de la estructura de referencia y la región de la imagen con la cual se corresponde”.

$$(I \ominus H)(x, y) = \min_{(i, j) \in H} \{I(x + i, y + j) - H(i, j)\} \quad (37)$$

La estructura más comúnmente utilizada en este algoritmo es una estructura plana que equivale a un operador local mínimo.

4.2.2 Etapa de Segmentación

A continuación se detallan los módulos que componen la etapa de segmentación (ver Figura 49). Como se ha comentado, esta etapa queda lejos de ser algo trivial debido a las características de las imágenes estroboscópicas de

cuerdas vocales (descritas en la sección anterior) con las que se está trabajando, se puede afirmar que la etapa de Segmentación es la etapa principal del sistema, básica para que las etapas posteriores funcionen correctamente.

El objetivo buscado en esta etapa es la obtención del espacio glótico con la mayor exactitud posible y de forma automatizada, es decir, minimizando o eliminando cualquier posible interacción con el usuario durante el proceso (incluyendo la inicialización). El interés por segmentar correctamente el espacio glótico es múltiple: por un lado, para estudiar el borde interior de cada una de las cuerdas vocales, y por otro, para calcular el área del espacio glótico en cada una de las imágenes de la secuencia en caso de ser necesario (no muy interesante para medir el movimiento vibratorio de las cuerdas vocales dada la frecuencia de captura de las imágenes estroboscópicas).

Este procesado es crítico, ya que de sus resultados depende el cálculo de todos los parámetros objeto de estudio para la emisión del diagnóstico o evaluación de la patología, en caso de que exista.

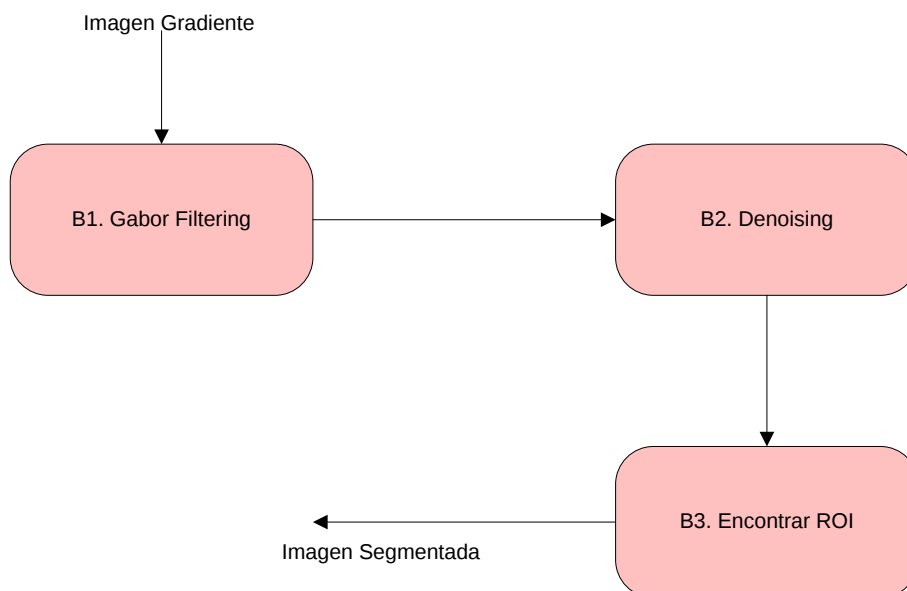


Figura 49. Diagrama de Bloques de la etapa de Segmentación

B1. Gabor Filtering

Los filtros de Gabor son filtros lineales cuya respuesta impulsional es una función sinusoidal multiplicada por una función gaussiana, se podría decir que son filtros casi paso banda. Por lo tanto, se han usado ampliamente en la segmentación de texturas.

En este módulo (ver Figura 50) se ha aplicado un algoritmo centrado en la aplicación de filtros de Gabor, que ha proporcionado la capacidad de resaltar las características más significativas en la imagen, gracias al análisis de texturas. En nuestro caso, la región de interés buscada es el espacio glótico.

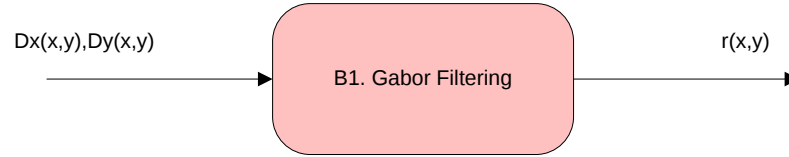


Figura 50. Bloque B1 de la Etapa de Segmentación

Siendo el *input* del módulo B1, cada frame de la secuencia de cuerdas vocales, $I(x,y)$, $x, y \in \Omega$ (Ω – el set de puntos de la imagen) es convolucionado con una función de Gabor bidimensional $g(x,y)$, $x, y \in \Omega$ para obtener la imagen de características de Gabor $r(x,y)$ según (38).

El *output* de este módulo son las imágenes en las que se aplican los filtros de Gabor y en las que se resaltan las texturas de la imagen.

$$r(x,y) = \iint I(\xi, n) = g(x - \xi, y - n) d\xi dn \quad (38)$$

Se usa la siguiente familia de funciones de Gabor:

$$g_{\lambda, \theta, \varphi}(x, y) = e^{-\frac{(x'^2 + y'^2)}{2\sigma^2}} \cos\left(2\pi \frac{x'}{\lambda} + \varphi\right) \quad (39)$$

Dónde:

$$x' = x \cos \theta + y \sin \theta$$

$$y' = -x \sin \theta + y \cos \theta$$

x e y son las coordenadas de los píxeles de la imagen en el rango $(-x/2, x/2)$ y $(-y/2, y/2)$. El parámetro σ determina la anchura del filtro. El parámetro λ es la longitud de onda y $1/\lambda$ la frecuencia espacial del factor armónico.

El parámetro ángulo θ ($\theta \in [0, \pi)$) especifica la orientación de la normal a la paralela de los lóbulos positivo y negativo de los Filtros de Gabor (esta normal es el eje x' en (39)). Finalmente, el parámetro ϕ es el desplazamiento de fase en el argumento del factor armónico.

Estos filtros tienen la capacidad necesaria para resaltar las características de una imagen con respecto a una cierta orientación y frecuencia. De esta forma, los Filtros de Gabor actúan como un filtro pasobanda gaussiano modulado por una senoide compleja.

La literatura [Clausi+00] recomienda usar ángulos cuya orientación esté separada 30°: 0°, 30°, 60°, 90°, 120°, y 150°.

Se han hecho pruebas con todos los ángulos recomendados, y dados los resultados empíricos, se ha decidido que la orientación a aplicar en este caso con los filtros explicados previamente es $\theta = \pi/3$.

B2. Denoising

Llegado este punto, aún es necesario eliminar mucha información para obtener la región de interés: el espacio glótico.

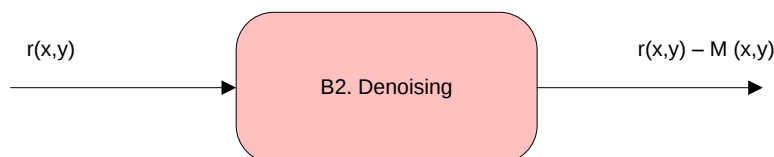


Figura 51. Bloque B2 de la etapa de Segmentación

El *input* de este módulo (ver Figura 51) son las imágenes filtradas a las que se les ha aplicado el filtro de Gabor.

El *output* del módulo B2 son las imágenes en las que ya se han eliminado artefactos para posteriormente ser capaces de encontrar automáticamente la región de interés. Se consideran artefactos toda la información adicional que aparece en la imagen como: líneas, puntos o pequeñas regiones cerradas.

Para intentar eliminar todo este “ruido” que todavía aparece en cada uno de los frames, se crea una máscara con la misma resolución que el frame estudiado ($M(x,y) = 1$) y se resta al resultado previo para eliminar el ruido de la imagen ($GaborResult - M(x,y)$).

B3 Encontrar ROI

Eliminado todo aquello que se ha considerado ruido en el bloque de denoising, todavía en la imagen quedan más regiones cerradas, y no son las que se están buscando. La búsqueda de la región de interés, es decir, el espacio glótico, es la función de este bloque B3.

En la Figura 52 se puede ver el input y output del bloque B3.

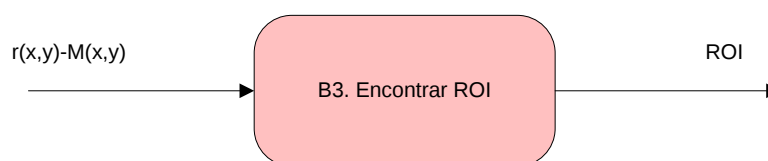


Figura 52. Bloques B3 de la etapa de Segmentación

El *input* del módulo B3 son las imágenes a las que previamente se les ha aplicado el denoising, pero en las que todavía no es sencillo encontrar la región de interés.

El *output* de este módulo es la imagen segmentada y en la que sólo parecen las regiones que cumplen los umbrales establecidos y que se describen a continuación.

En la Figura 53 se muestra que el detalle del bloque B3 y muestra que puede ser un proceso iterativo, si en la imagen se detecta más de una región cerrada, que suele ser lo habitual. En el hipotético caso de que sólo se detecte un área cerrada, se toma esa como ROI.

Asumiendo que en la mayoría de las ocasiones se detectan varias regiones cerradas en la imagen tras los bloques B1 y B2, se van examinando una a una (bloque ROI_n de la Figura 53), hasta llegar a la que cumpla con las siguientes condiciones:

- Nuestra región de interés (ROI) nunca debería estar en el borde derecho o izquierdo de la imagen, dado que durante su adquisición se intenta que esté lo más centrada posible y que las imágenes no hayan sido cortadas por la parte correspondiente a las cuerdas vocales. Estas son las dos primeras condiciones que se plantean en la Figura 53.
- En el algoritmo se ha establecido que el área mínima de una región para que sea tomada en cuenta tiene que ser de 100 píxeles, sino no se considera. Esta es la tercera condición de la Figura 53. En las patologías estudiadas en esta tesis se tienen en cuenta principalmente los frames en los que las cuerdas vocales están abiertas (en las secuencias disponibles en nuestra base de datos, la mayoría de las secuencias no registran el cierre de las cuerdas vocales), por esta razón se ha establecido un umbral mínimo.

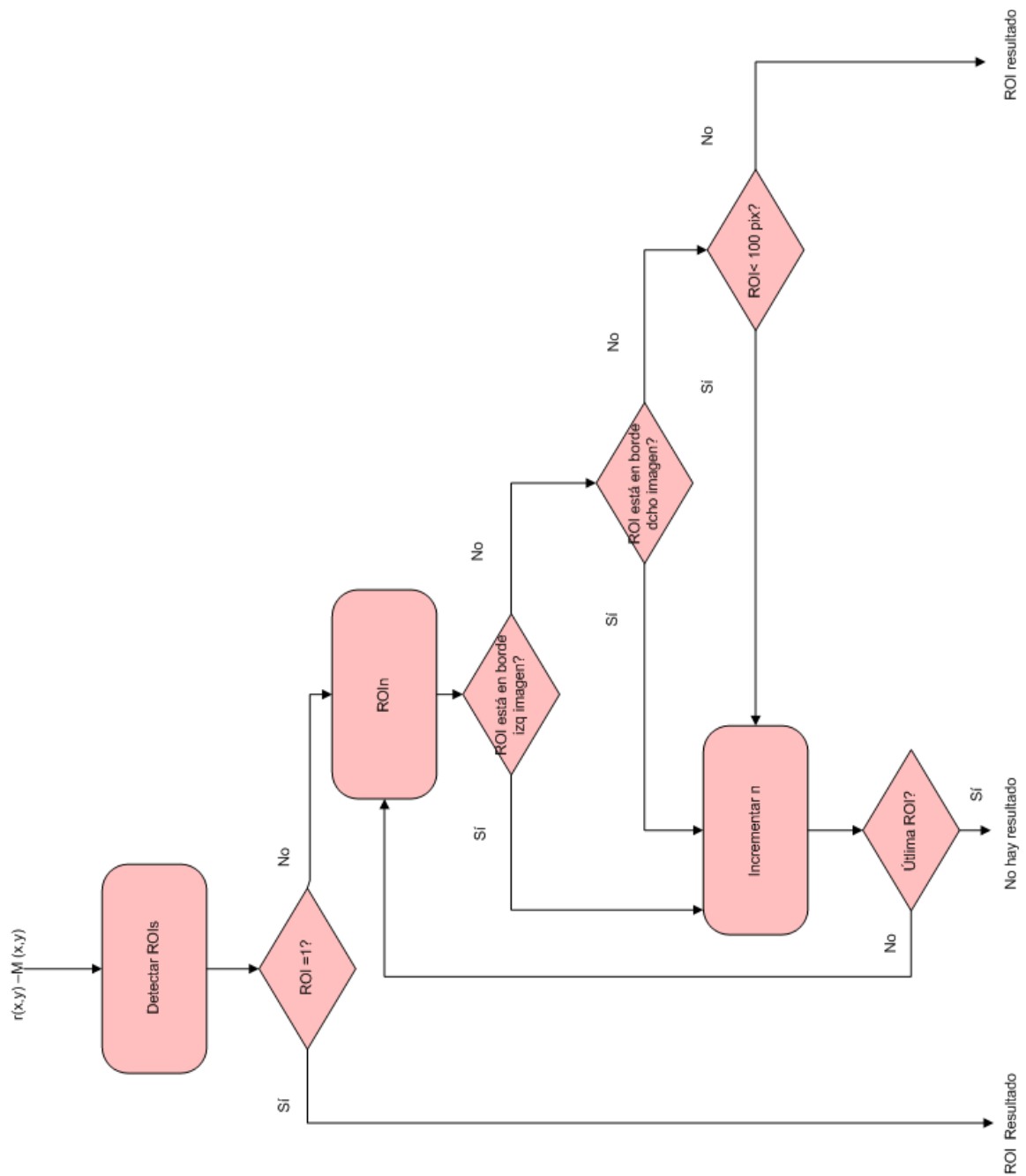


Figura 53. Detalle del Proceso iterativo del bloque B3

- Nos encontramos con la última ROI. El algoritmo se va ejecutando, analizando cada una de las ROI, hasta que encuentra una que cumple con las condiciones marcadas. La última condición de la Figura 53, es que hayamos analizado la última ROI, y no hayamos encontrado el resultado, es decir, no hayamos considerado ninguna de las regiones detectadas como la nuestra, el espacio glótico.

Basándonos en las premisas y umbrales propuestos, el sistema propone la región de interés, y a partir de este momento se pueden empezar a calcular parámetros objetivos como el área del espacio glótico e incluso obtener el Glottal Area Waveform (GAV) [Noordzij+00]. Como hemos indicado anteriormente, puede que no sea muy interesante en imágenes estroboscópicas, y más concretamente en las secuencias disponibles ya que no han registrado el ciclo completo de vibración de las cuerdas vocales.

Es evidente que la aplicación del banco de filtros de Gabor funciona perfectamente para extraer áreas de características especiales. No se trata de una segmentación minuciosa, para eso tenemos el operador de Canny, sino de establecer un método para que de una manera rápida se pueda extraer el área de interés.

4.2.3 Etapa de Detección de Movimiento

Aunque es sabido que las imágenes estroboscópicas no son las más adecuadas para el estudio del movimiento y la vibración de las cuerdas vocales, sí nos pueden aportar algunos datos sobre determinadas patologías como las parálisis (unilateral o bilateral), siempre teniendo en cuenta sus limitaciones, y que tras los resultados que se obtengan mediante este método, el estudio tiene que ser completado por el especialista.

El método ideal para los estudios de cuerdas vocales cuyo movimiento se ve afectado es la videoquimografía, pero actualmente en España, prácticamente

ningún hospital dispone de esta tecnología de captura debido a su elevado coste, y sólo se utiliza para ámbitos de investigación.

El diagrama de bloques de esta etapa se puede observar en la Figura 54.

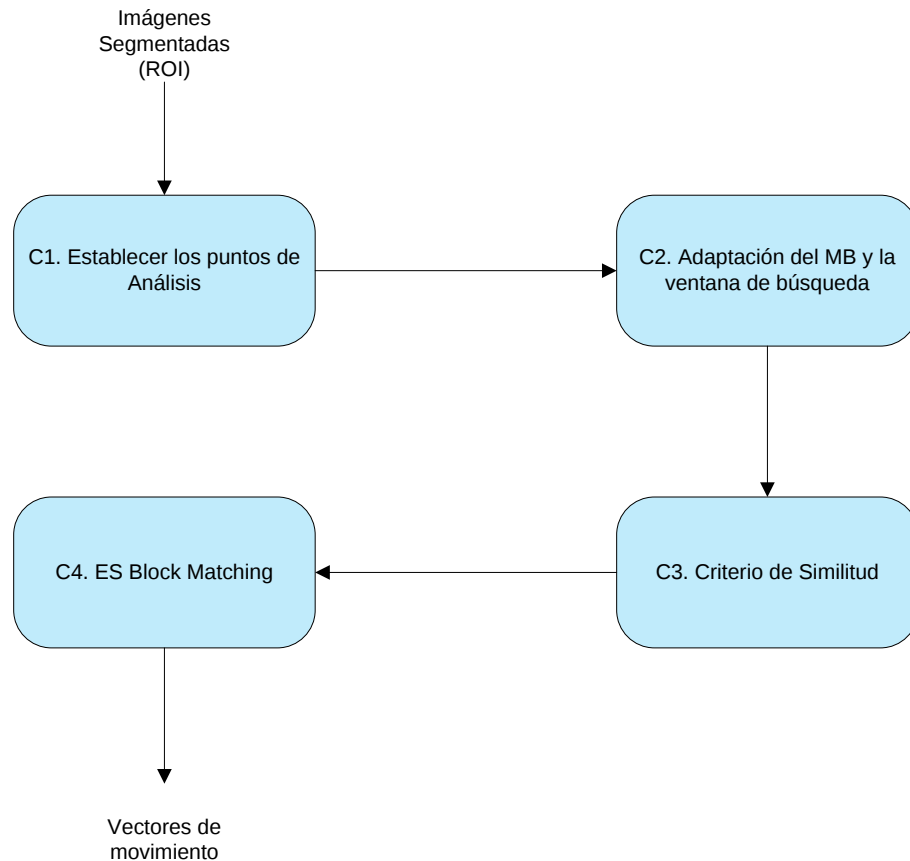


Figura 54. Diagrama de Bloques de la Etapa de Detección de Movimiento

El propósito general de los algoritmos de detección de movimiento, o en este caso, de Block Matching es encontrar un bloque del frame k que concuerde con un bloque en algún otro frame ($k-1$ para este estudio), aplicando el criterio de similitud y encontrando la máxima posibilidad de semejanza para ese Macro bloque (MB) en una ventana de búsqueda definida.

Para el estudio del movimiento de las cuerdas vocales (y específicamente para el estudio de las parálisis), se ha modificado el algoritmo general [Kim+98], tal y como se ve en la Figura 54, especialmente en los bloques C1 y C2.

C1. Establecer los puntos de análisis

Para establecer los puntos de los bordes de las cuerdas vocales cuyo movimiento va a ser analizado, es necesario decidir:

- Si vamos a estudiar todos los puntos que forman el borde de cada una de las cuerdas vocales, o
- Si vamos a acotar el estudio del movimiento a un número de puntos significativos, y con ellos extrapolar el movimiento general de cada una de las cuerdas vocales.

Se ha optado por la segunda opción, y estudiar solo un número acotado de puntos. Para establecer los umbrales se ha seguido el mismo criterio que se ha usado en toda la tesis y que han demostrado estudios previos [Armstrong+10], y es que mayoría de las patologías vocales se encuentran ubicadas en el tercio inferior de la cuerda vocal.

Para ello hemos realizado un paso previo de procesado para coger los puntos de las cuerdas que se encuentran entre el 10% inferior de la longitud de la cuerda y el 40%. El primer 10% de los puntos se ha descartado porque empíricamente se ve que no es representativo del movimiento de las cuerdas vocales, o incluso puede ser nulo, y que puede desvirtuar los resultados. Experimentalmente también se ha demostrado que suelen aparecer valores inconsistentes entre el 10% y el 30% de la longitud de las cuerdas. Por lo tanto, los puntos que han sido escogidos son del 30% hasta el 40% inferior de las cuerdas. Aunque es un rango muy pequeño de puntos, solo el 10%, son los suficientes para sacar conclusiones y para no sobrecargar mucho el procesado de este bloque C1.

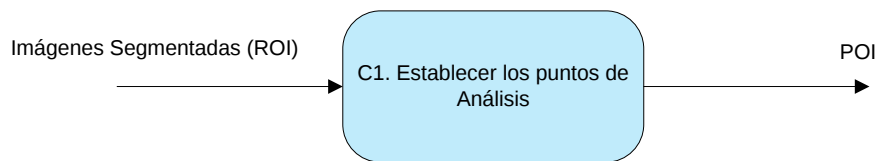


Figura 55. Bloque C1 de la etapa de la etapa de Detección de movimiento

El *input* del bloque C1 (ver Figura 55) son las imágenes con el espacio glótico segmentado.

El *output* nos muestra un 10% de los puntos del borde interior de cada una de cuerdas vocales y que atendiendo a los umbrales establecidos, son los puntos de interés.

C2. Tamaño adaptado del MB y de la Ventana de Búsqueda

El bloque C2 (ver Figura 56) muestra las decisiones tomadas para hallar el tamaño del macrobloque y la ventana de búsqueda, que a diferencia de la mayoría de las aplicaciones, no son de tamaño fijo.

El *input* de este bloque es la imagen con la segmentación del espacio glótico y los puntos de interés calculados en C1.

En nuestro caso, el tamaño del macrobloque (MBSize) **depende de la amplitud del espacio glótico en cada frame** (diferente en todos los frames debido al movimiento ondulatorio y a la vibración de las cuerdas vocales) y de **la posición** del dicho macrobloque a lo largo de los bordes de las cuerdas. Para nosotros, la posición se ha resuelto en el bloque C1, ya que en esta etapa sólo se van a estudiar los puntos establecidos (10%) en el bloque anterior.

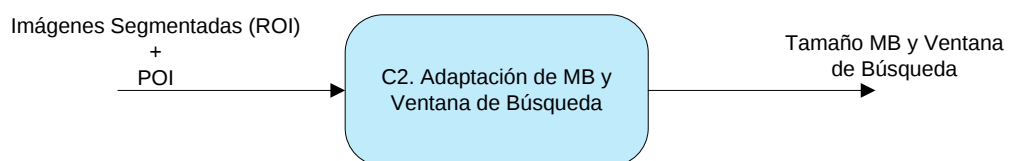


Figura 56. Bloque C2 de la etapa de la etapa de Detección de movimiento

Habitualmente, la relación entre el macrobloque y la ventana de búsqueda, es exactamente que esta última sea el doble. Esta relación funciona si los valores de ambos parámetros se mantienen fijos.

El caso de las cuerdas vocales es completamente diferente, dependemos de la apertura del espacio glótico y estos parámetros son recalculados en cada uno de los frames de la secuencia (ver Figura 59). Por lo tanto, y siendo fieles a la literatura [Chan+2002], el principal objetivo de este bloque es obtener un punto de interés de las cuerdas vocales y localizarlo en el centro del macrobloque. Dado que el movimiento de cada una de las cuerdas vocales de forma independiente lo que vamos a estudiar, se calcularán dos macrobloques de igual tamaño, pero estarán ubicados cada uno en una cuerda vocal.

El *output* de este bloque C2, son los tamaños para los macrobloques y la correspondiente ventana de búsqueda.

C3. Criterio de similitud

Para el cálculo del criterio de similitud se ha elegido como función de coste el Error Cuadrático Medio (Mean Squared Error (MSE)), cuya fórmula se muestra en (40).

Siendo N el tamaño del macrobloque, C_{ij} y R_{ij} son los pixeles que se comparan entre un macrobloque y el de referencia, respectivamente.

$$MSE = \frac{1}{N^2} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} (C_{ij} - R_{ij})^2 \quad (40)$$

El *input* de este bloque es toda la información proveniente de la etapa de segmentación y de los bloques C1 y C2.

El *output* del bloque C3 es el macrobloque de otro frame (normalmente el siguiente) que tras la comparación, se ha decidido que es el que más se parece al macrobloque de referencia.

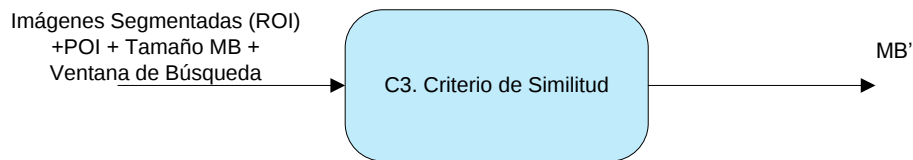


Figura 57. Bloque C3 de la etapa de la etapa de Detección de movimiento

C4. Exhaustive Block Matching

En este bloque es donde realmente se calculan los vectores de movimiento entre el macrobloque de referencia y el macrobloque calculado en C3 (ver Figura 58).

De todas las técnicas de Block Matching disponibles en la literatura [Barjatya04] se ha elegido el Block Matching exhaustivo. Esta técnica de block matching es la mejor para detectar y medir el movimiento porque se estudian todas las posibilidades a lo largo de la imagen, pero también es la que supone una mayor carga computacional y por lo tanto, tiempo. Dado que el algoritmo sólo ha sido evaluado en un número de puntos limitado, el tiempo de proceso alto que siempre se le achaca a este método queda reducido al mínimo.

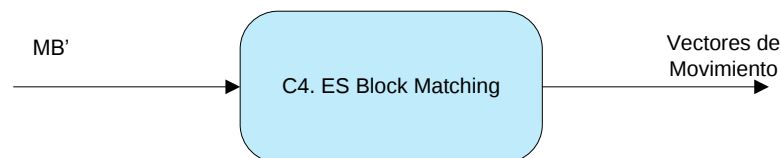


Figura 58. Bloque C4 de la etapa de la etapa de Detección de movimiento

El *input* del bloque C4 es el macrobloque salida del bloque C3, y el *output*, los vectores de movimiento que estábamos buscando.

Un ejemplo de vectores de movimiento se puede ver en la Figura 59 resaltados en rojo.

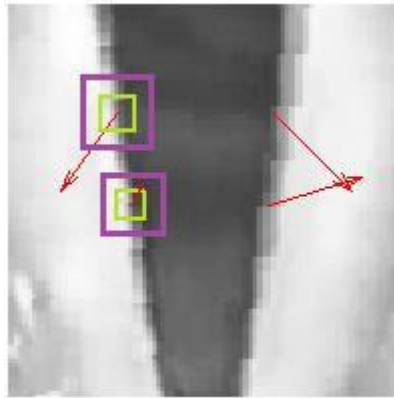


Figura 59. Imagen Ejemplo. Ventana de búsqueda en violeta y Macrobloque en verde.

4.2.4 Etapa de Clasificación

Para conseguir un pre-diagnóstico o clasificación de patologías se ha aplicado una clasificación estadística usando la técnica PCA (Principal Components Analysis).

Como alternativa a esta técnica existen principalmente técnicas de extracción de características que requieren realizar mediciones previas en las imágenes, lo cual suele ser una tarea bastante ardua. Además, aunque este método requiere de un consumo de memoria menor y puede llegar a ser más rápido, en el caso de estudio que nos ocupa, todo el sistema forma parte del post-procesado, nunca se realiza “online” por lo que todas las fuerzas se han concentrado en obtener una tasa de aciertos alta, y PCA ofrece mejores resultados.

D. Clasificación PCA

Para llevar a cabo esta primera clasificación de las patologías, se ha elegido el Análisis de Componentes Principales (PCA), también conocida como Eigenfaces. Esta técnica proviene del análisis de datos multivariable y su

desarrollo comenzó en 1901 en un contexto diferente [Noordzij+00]. Proyección de datos sobre el subespacio de componentes principales se llama también Transformada Hotelling o Transformada Karhunen Loeve-(KLT) de acuerdo con [Pearson01] [Hotelling33] [Karhunen47]. Este método ha sido muy útil en el reconocimiento de datos (especialmente en las caras) como puede verse en [Loève63] [Torre+01] [Pujol+01].

En términos matemáticos, se calculan los principales componentes de la distribución de las imágenes de cuerdas vocales (en este caso), o de los autovectores (eigenfolds para nosotros) de la matriz de covarianza del conjunto de imágenes, tratando la imagen como un punto (vector) en un espacio dimensional muy alto.

Una imagen de cuerdas vocales $I(x,y)$ es una array bidimensional de $N \times N$ de intensidad 8 bits. Una imagen también puede ser considerada como un vector de dimensión N^2 , de manera que una imagen típica de tamaño de 256×256 se convierte en un vector de dimensión 65.536, o un punto equivalentemente en un espacio de 65.536 dimensiones. Un conjunto de imágenes se asigna a un conjunto de puntos en este espacio enorme.

Las imágenes de cuerdas vocales, siendo similares en su configuración general, no se distribuyen al azar en este enorme espacio, y por lo tanto puede ser descrito por un subespacio dimensiones relativamente baja, de acuerdo con la literatura [Hotelling33].

Estos vectores definen el subespacio de cuerdas vocales. Cada vector de longitud N^2 , describe una imagen de $N \times N$, y es la combinación lineal de las imágenes originales de cuerdas vocales. Nos referiremos a estos vectores como “eigenfolds” porque son los autovectores de la matriz de covarianza correspondiente a las imágenes originales, y porque tienen una apariencia similar a las cuerdas vocales.

Siendo el Training Set de imagenes de cuerdas vocales $\Gamma_1, \Gamma_2, \dots, \Gamma_M$, entonces la media del grupo de imágenes se define por (41):

$$\Psi = \frac{1}{M} \sum_{j=0}^{M-1} \Gamma_j \quad (41)$$

Cada cuerda vocal se diferencia de la media por el vector:

$$\Phi_i = \Gamma_i - \Psi \quad (42)$$

Este grupo de vectores está sujeto al Análisis de Componentes Principales que busca un conjunto de vectores ortonormales M , que mejor describe la distribución de los datos. El vector K -ésimo se elige según (43):

$$\lambda_k = \frac{1}{M} \sum_{n=1}^M (u_k^T \Phi_n)^2 \quad (43)$$

Es el máximo sujeto a:

$$u_I^T u_K = \delta_{IK} = \begin{cases} 1, \text{if } I=K \\ 0, \text{en el resto de los casos} \end{cases} \quad (44)$$

Los vectores u_k y escalares λ_k son los autovectores y autovalores, respectivamente, de la matriz de covarianza.

$$C = \frac{1}{M} \sum_{n=1}^M \Phi_n \Phi_n^T = A A^T \quad (45)$$

Donde la matriz $A = [\Phi_1, \Phi_2, \dots, \Phi_M]$

Se diagonaliza la matriz de covarianza y se escogen los u autovectores asociados a los u autovalores más grandes.

La matriz C presenta una dimensión N^2 por N^2 elementos, y dado el tamaño de las imágenes, la diagonalización de la matriz requiere gran complejidad de cálculo y memoria.

Para reducir el coste computacional, se trabaja con la matriz $A^T A$, que presenta un tamaño mucho menor y depende del número de imágenes de la base de datos.

En este caso, a partir de los autovectores $\vec{v}_l$ se puede encontrar una expresión aproximada para los autovectores $\vec{u}_l$.

$$\vec{u}_l = \sum_{k=1}^M \vec{v}_{lk} \Phi_k \quad l=1, \dots, M \quad (46)$$

Para reconocer una imagen del test habrá que comparar sus coordenadas (Ω) en la base del subespacio vectorial obtenido a partir de la matriz de covarianza, con las coordenadas de cada una de las imágenes de cuerdas vocales modeladas (Ω_k). Para ello hay que proyectar cada una de las imágenes de cuerdas vocales de la secuencia que se desea clasificar en el espacio encontrado durante el proceso de entrenamiento. Usaremos únicamente los vectores más significativos. Por tanto, hay que proyectar cada una de las imágenes de la secuencia sobre las M' direcciones del espacio vectorial mediante:

$$\omega_k = \vec{u}_k^T (I - \varphi) \quad (47)$$

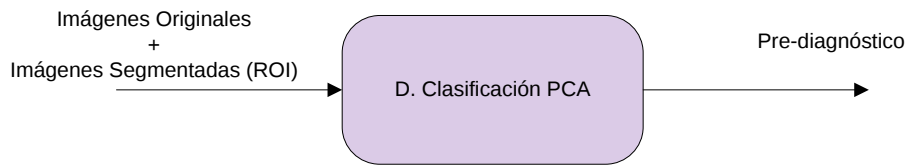


Figura 60. Diagrama etapa D

Descrita la técnica que se ha utilizado, y tal y como se puede observar en la Figura 60, el *input* de esta etapa son las imágenes originales y las segmentadas, con el objetivo de conseguir una pre-clasificación o un pre-diagnostico como *output*.

Módulos Principales de la clasificación PCA propuesta

En la Figura 61 se muestran los principales módulos de PCA. En este caso, la entrada al sistema son los frames originales y su correspondiente máscara, para obtener como resultado un pre-diagnóstico.

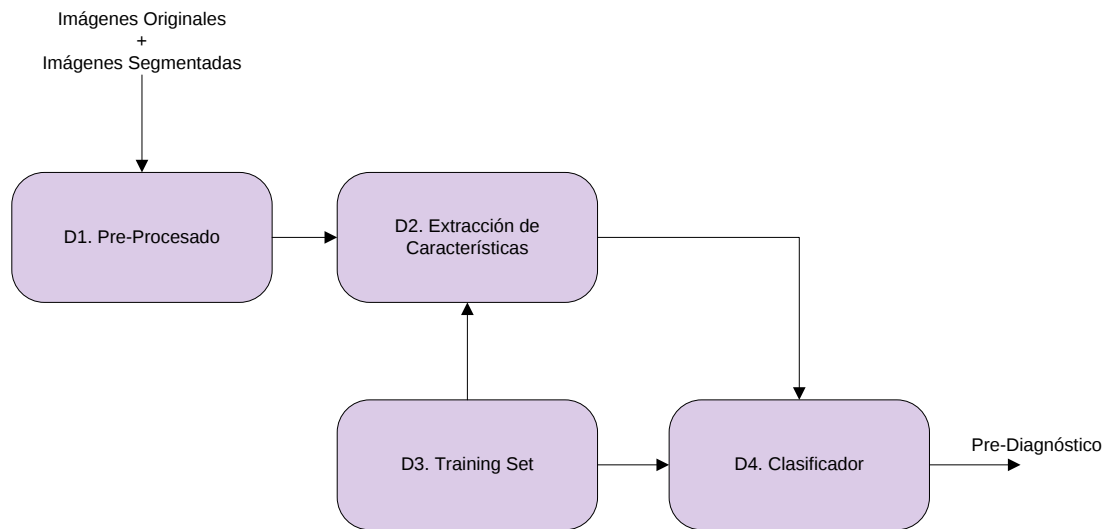


Figura 61. Diagrama de bloques del Clasificador de Patologías Vocales

- **D1. Módulo Pre-Procesado** (ver Figura 62). La entrada de este módulo son la imagen original y la máscara del espacio glótico. Para que PCA funcione correctamente, todas las imágenes han de tener un escalado comparable y la misma orientación. Por ello, es necesario, otra etapa de pre-procesado para homogeneizar todas las muestras. Se tienen en cuenta dos puntos principalmente: el vértice inferior y el centro de gravedad del espacio glótico.

D1.1. Calcular las coordenadas del vértice inferior.

D1.2. Calcular las coordenadas del centro de masas.

$$X_c = \frac{\sum(X * Pixel\ Values)}{\sum(Pixel\ Values)} \quad (48)$$

$$Y_c = \frac{\sum(Y * Pixel\ Values)}{\sum(Pixel\ Values)} \quad (49)$$

D1.3. Desplazar la imagen, ubicando el centro de masas como centro de la imagen. Con esto, todas las imágenes tendrán el centro de masas en las mismas coordenadas.

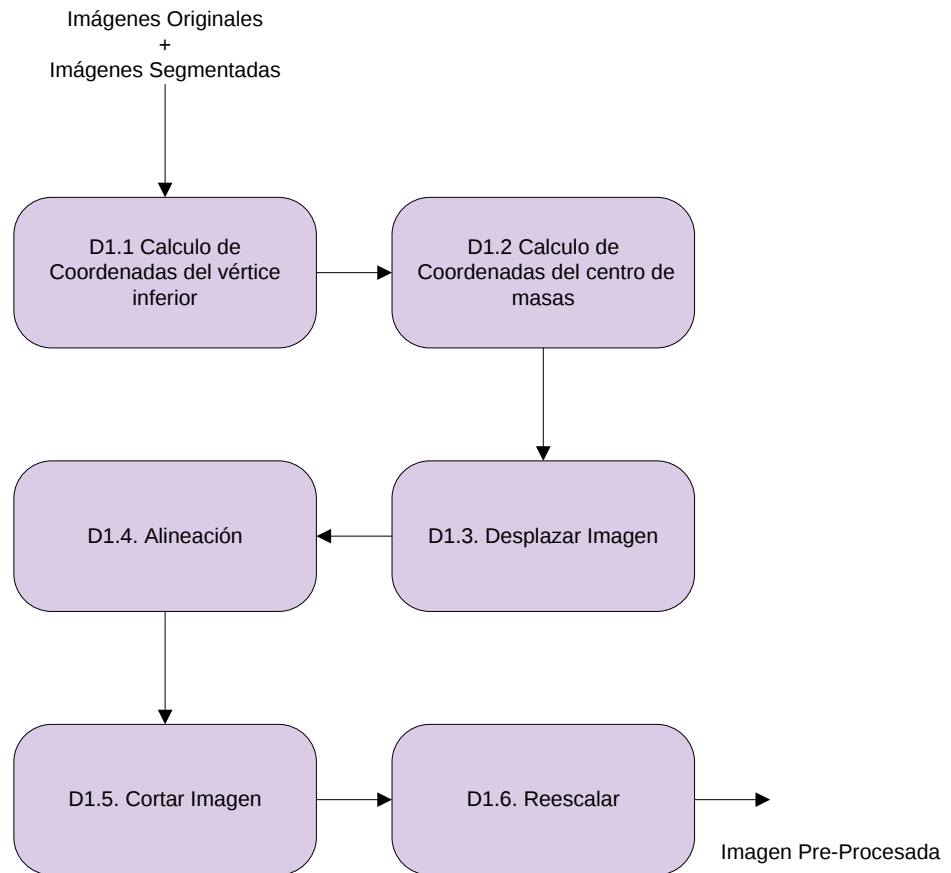


Figura 62. Detalle del bloque D1 de la etapa de Clasificación

$$Io(x, y) = I(x - x1, y - y1) \quad (50)$$

$$x1 = \frac{N}{2} - Xc \quad (51)$$

$$y1 = \frac{N}{2} - Yc \quad (52)$$

D1.4. El centro de masas y el vértice interior deben estar verticalmente alineados. Para ello se calcula la pendiente del vértice interior con

respecto al centro de masas, y se rota toda la imagen respecto al centro de la imagen, que en este caso, coincide con el centro de masas. Una vez se ha hecho esto, el centro de masas y el vértice inferior del espacio glótico están alineados.

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{pmatrix} \cos B & \sin B \\ -\sin B & \cos B \end{pmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \quad (53)$$

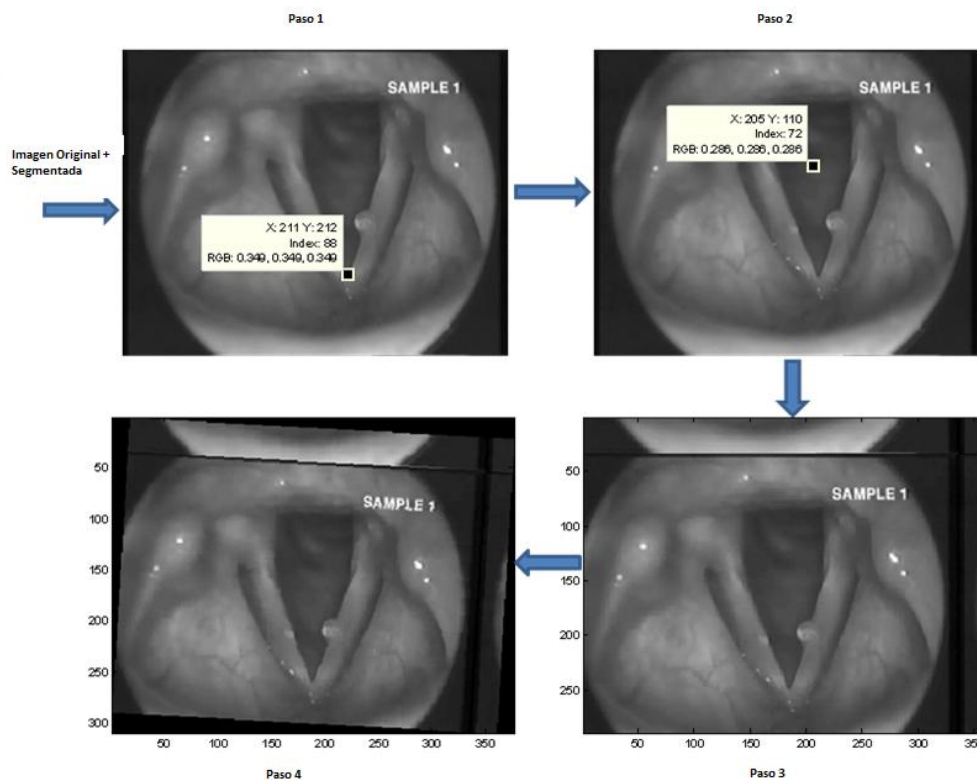


Figura 63. Resultados de los 4 primeros pasos descritos en el bloque D1 de la etapa de Clasificación

D1.5. El siguiente paso consiste en recortar la imagen cogiendo el margen entre el centro de masas y las coordenadas del vértice inferior del espacio glótico como puntos de referencia. Todas las imágenes ser cortan manteniendo las mismas proporciones.

D1.6. El último paso es reescalar la imagen, así todas las imágenes tendrán la misma resolución (200x200) y el algoritmo se podrá aplicar sobre las mismas componentes (ver Figura 64).

- **D2. Módulo Extraer Características.** Después de realizar un pre-procesamiento, la imagen de las cuerdas vocales se introduce en el módulo de extracción de características con el fin de encontrar las características más importantes, y que se van a utilizar para la clasificación. En otras palabras, este módulo es el responsable de componer un vector representativo de las imágenes de cuerdas vocales.



Figura 64. Resultado del módulo de Pre-Procesado D1.1

- **D3. Módulo Training Set.** Training sets se utilizan durante la fase de aprendizaje. Los módulos de extracción de características y de clasificación ajustan sus parámetros con el fin de lograr un rendimiento óptimo, haciendo uso de los Training Sets.

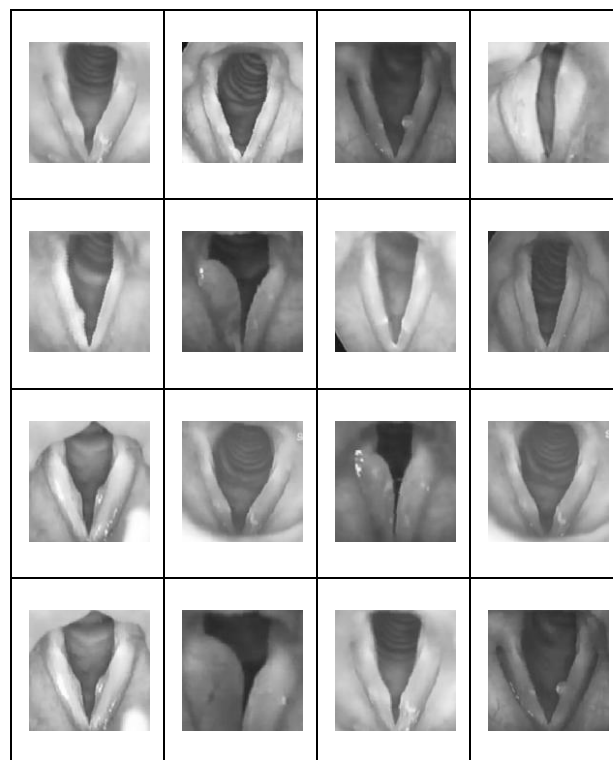


Figura 65. Ejemplo de Training Set

Durante este trabajo de investigación se han hecho pruebas con diferentes tamaños de Training Sets, cuyos resultados se muestran en la siguiente sección. En la Figura 65 podemos ver un ejemplo de Training Set que contiene 16 imágenes de cuerdas vocales.

- **D4 Módulo Clasificador.** En este módulo las características extraídas de la imagen de cuerdas vocales se comparan con las almacenadas. Después de eso, el resultado se calculará de acuerdo a la mínima distancia euclídeana entre ellos. Principalmente se debe distinguir entre patologías morfológicas e imágenes sanas o con parálisis.

4.2.5 Etapa de Análisis y Medida

En esta sección se describe el detalle de los módulos de los que está compuesta la etapa de Análisis y Medida.

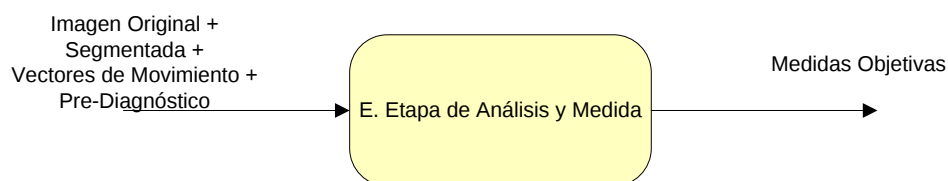


Figura 66. Diagrama de Bloques de la etapa de Clasificación, Análisis y Medida

El *input* de esta etapa son principalmente, todos los resultados de todas las etapas descritas hasta el momento (ver Figura 66).

- Imagen con la máscara del espacio glótico resultado de la segmentación realizada en la etapa B.
- Vectores de Movimiento tras aplicar el Exhaustive Block Matching bajo las condiciones descritas en la etapa C.
- Resultado de la clasificación o pre-diagnóstico de la etapa D.

El *output* de la etapa E son las medidas objetivas que nos llevarán a la emisión del diagnóstico.

Es importante recordar que la mayoría de las medidas calculadas son relativas, es decir, no se pueden realizar cálculos en cm, ya que como se ha comentado ampliamente, hay parámetros de la captura que no podemos controlar, e impiden mostrar medidas absolutas en el sistema de medida convencional.

El input que arranca todo el proceso de esta etapa es el resultado de la etapa D. Esta etapa proporciona un pre-diagnóstico, que para ser validado es necesario aplicar transformaciones matemáticas y obtener medidas objetivas que lo apoyen.

Para la etapa E, sólo nos fijamos en si la clasificación ha sido como cuerdas vocales sanas, o cuerdas vocales patológicas, a priori no se presta atención a la patología exacta resultado de la clasificación, aunque en el capítulo “Experimentación” mostraremos el grado de fiabilidad que tiene esta técnica con respecto a clasificar la patología exacta de las cuerdas vocales. A partir de este momento, el tratamiento de las cuerdas vocales clasificadas como “patológicas” y las cuerdas vocales clasificadas como “sanas” serán tratadas de forma diferente, tal y como se puede ver en la Figura 67.

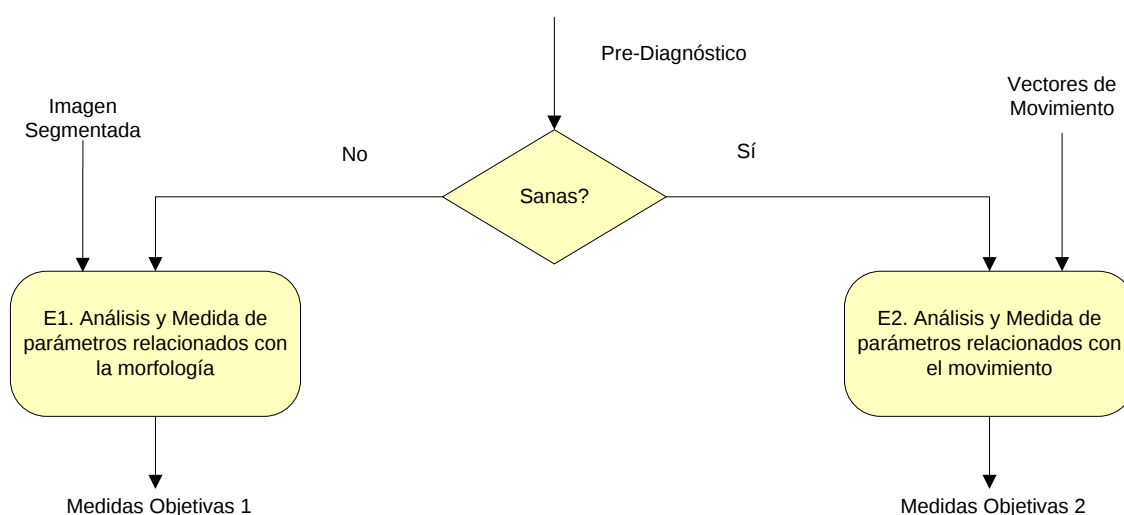


Figura 67. Detalle de los módulos de la Etapa E

Es importante aclarar que dentro de las patologías clasificadas como “sanas” (que se analizan en el bloque E2 de la Figura 67) se encontrarán también las

imágenes de cuerdas vocales cuya patología está relacionada con el movimiento y no con la morfología.

➤ **E1. Análisis y medida de parámetros relacionados con la morfología.**

Este gran bloque analiza todas las imágenes de las secuencias que han sido clasificadas como patológicas por el bloque D y calcula parámetros orientados a mediar las anomalías morfológicas de las cuerdas vocales. Este bloque se divide a su vez en otros tal y como muestra la Figura 68.

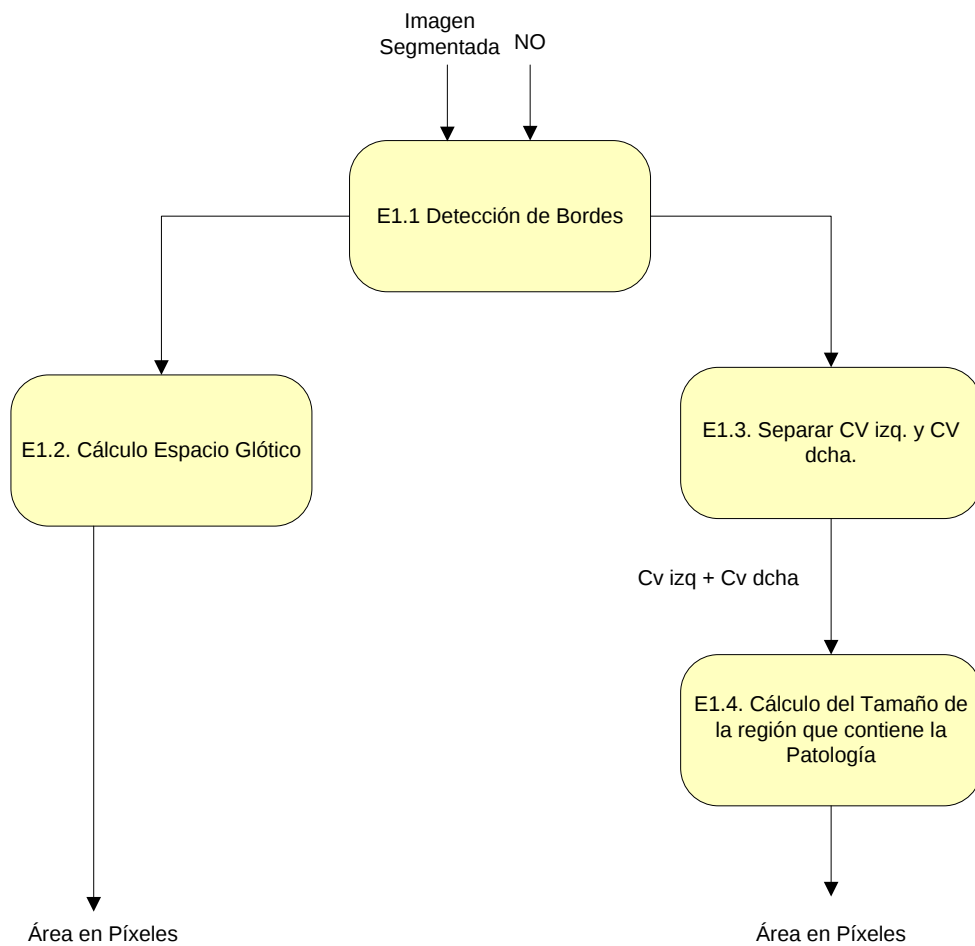


Figura 68. Detalle del Bloque E1 de la Etapa de Análisis y Medida

➤ **E2. Análisis y Medida de parámetros relacionados con el movimiento.**

Este segundo gran bloque analiza todas las imágenes de las secuencias que han sido clasificadas como no patológicas por el bloque D. De este podemos ver el detalle en la Figura 76.

4.2.5.1 Análisis y Medida de parámetros relacionados con la morfología

El bloque E1 tiene como *inputs* la imagen segmentada y el resultado de la clasificación que indica que la imagen de cuerdas vocales no es sana.

Los *outputs* van a ser medidas de áreas en píxeles, del espacio glótico y de la región que contiene y caracteriza a la patología correspondiente.

E1.1 Detección de bordes. La detección de bordes se realiza sobre la imagen que contiene la imagen con el espacio glótico segmentado. Para ello, se aplica el Filtro de Canny (ver ecuaciones 23-27) para extraer de forma rápida y sencilla el borde que forma la región cerrada del espacio glótico.

E1.2. Cálculo del Espacio Glótico. A partir de la segmentación de la etapa B es posible calcular el tamaño en píxeles del espacio glótico, para posteriormente mostrar el Glottal Area Waveform, que aportará información para el estudio del movimiento de las cuerdas vocales, aunque las imágenes estroboscópicas no son las más adecuadas para ello (por las características de la captura no podemos tener la certeza de disponer de todos los instantes del movimiento).

La medida del área del espacio glótico se hace de cada uno de los frames, y se estudia de forma especial el frame cuya área calculada es menor (asumiendo las limitaciones de las imágenes estroboscópicas), tal y como se puede observar en la Figura 69.

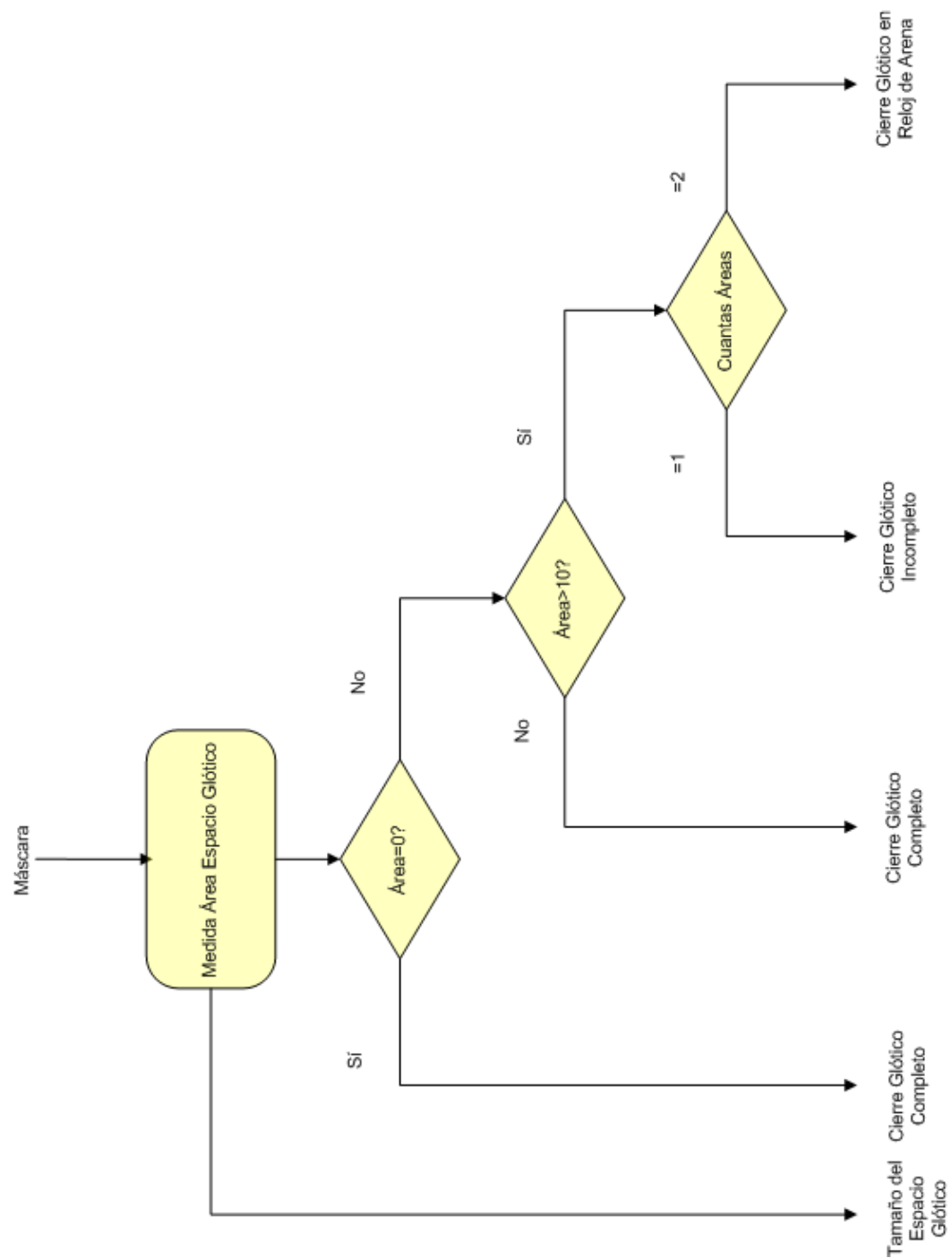


Figura 69. Detalle del Bloque E 1.2

Si existe algún frame en el que el área a estudiar es cero (o prácticamente cero), podemos concluir, que en lo que a cierre glótico se refiere, la cuerda es normal. Si por el contrario, existe un frame en el que el espacio glótico está dividido en varias áreas y/o el frame con menor área estudiada es mayor de 10 píxeles, se recomienda estudiar las cuerdas vocales con atención, ya que pueden ser objeto de un cierre glótico incompleto simple o en reloj de arena.

E1.3. Separar Cuerda Vocal Izquierda y Cuerda Vocal Derecha. Tras la detección de bordes, se extraen las líneas que corresponden a las cuerdas vocales derecha e izquierda, para a partir de este punto, trabajar con ellas de forma independiente (ver Figura 70).

Para mejorar la eficacia del algoritmo, y teniendo en cuenta que la mayoría de las patologías afectan al tercio inferior de la cuerda (continuando con el mismo criterio que en etapas anteriores), sólo se estudia el 80% de cada una de las cuerdas. Para llegar a establecer este umbral, también se han realizado pruebas empíricas con todas las imágenes de la base de datos para comprobar que estudiar por encima del 80% no se aportaba información adicional sobre las patologías de las cuerdas vocales.

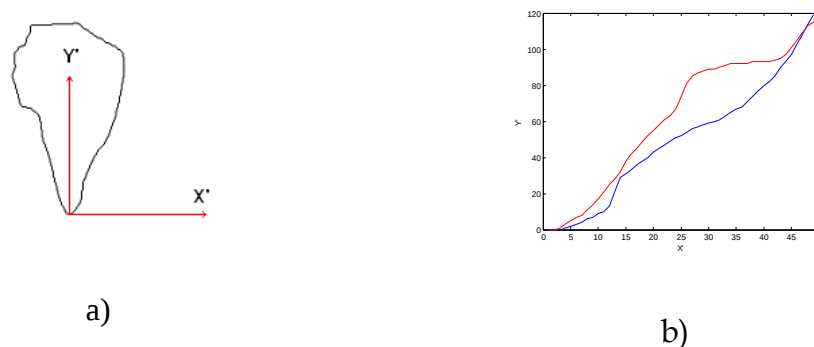


Figura 70. a) Contorno del espacio glótico. b) Contornos interiores de las cuerdas vocales. Derecha (azul) e izquierda (rojo).

E1.4. Cálculo del Tamaño de la región que contiene la patología. Para estudiar las características morfológicas cada una de las cuerdas vocales se realizan varios pasos que se pueden ver en la Figura 71.

- **E1.4.1 Cálculo de la derivada.** El primero de ellos es calcular la primera derivada, con el objetivo de localizar las irregularidades de la cuerda.

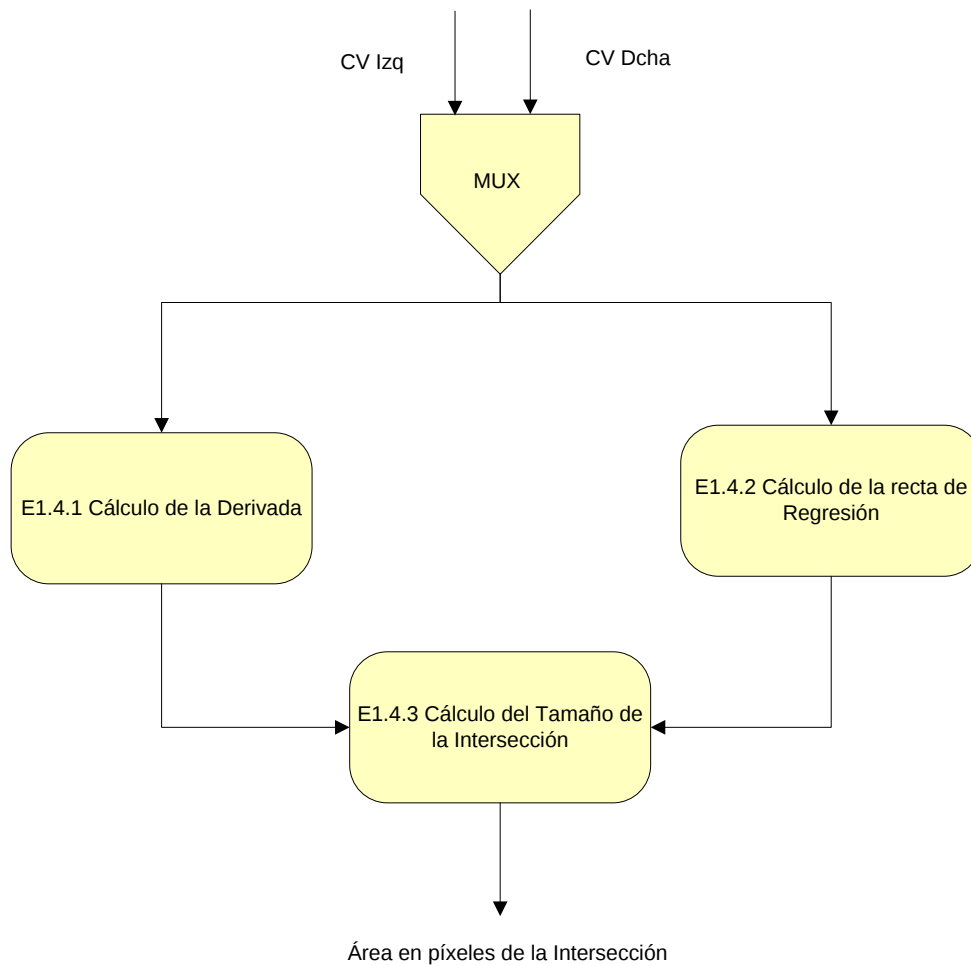


Figura 71. Detalle del bloque E1.4

Previo a la derivada, se aplica un filtro de suavizado para eliminar las irregularidades no significativas, tal y como se puede ver en la Figura 72 y Figura 73.

La Figura 72 muestra un segmento de cuerda vocal sana, posteriormente el mismo segmento tras aplicarle el filtro de

suavizado, y por último el resultado del cálculo de la derivada, en el que no se observan irregularidades significativas.

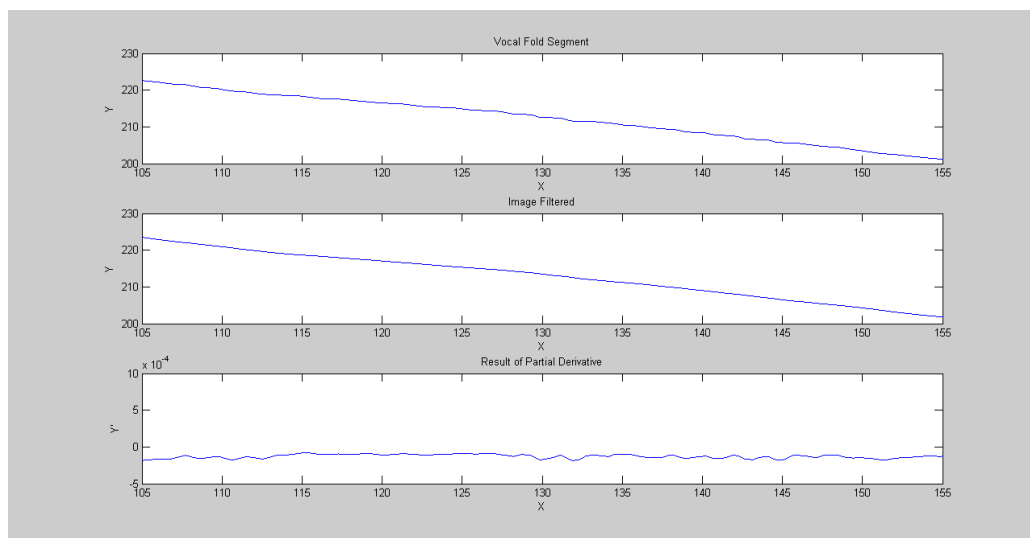


Figura 72. Cuerdas vocales sanas. Original, filtrada y resultado de la derivada.

Cuando se calcula la derivada a un segmento de cuerda vocal con alguna patología morfológica se observa el cambio en los resultados (ver Figura 73).

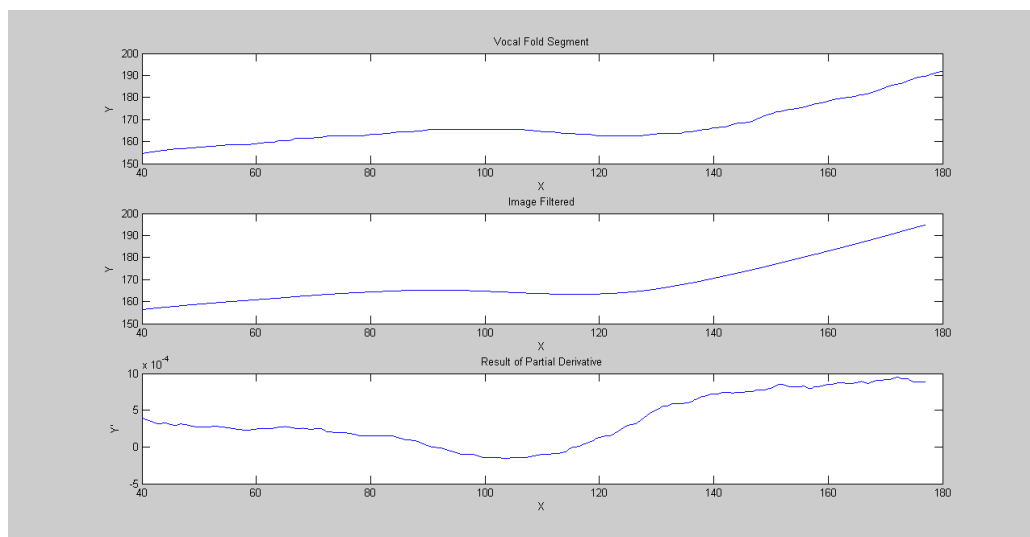


Figura 73. Cuerdas vocales patológicas. Original, filtrada y resultado de la derivada.

El siguiente paso es encontrar el punto de inicio de la patología en caso de que exista, usando la segunda derivada, tal y como se puede observar en la Figura 74a. Se calculan los máximos y los mínimos, y el punto medio entre ellos. Éste es el punto donde se considera el inicio de la

patología, siempre y cuando se encuentre la mitad inferior de la cuerda vocal. De lo contrario, se usa el punto con menor valor absoluto.

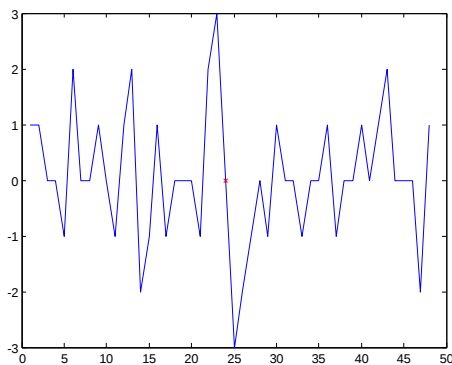
- **E1.4.2 Cálculo de la recta de regresión.** En paralelo es calculada la recta de regresión de cada una de las cuerdas vocales mediante métodos lineales como se puede ver en (54) y (55).

$$b = \frac{1/n \cdot \sum (x_i - x_m) \cdot (y_i - y_m)}{1/n \cdot (x_i - x_m)^2} \quad (54)$$

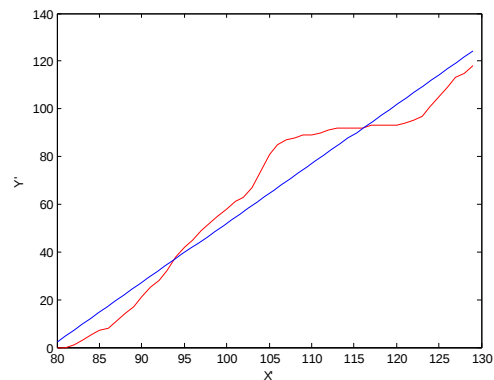
$$a = y_m - (b \cdot x_m) \quad (55)$$

Siendo b el gradiente de la recta de regresión y a el valor de la función en cero. En la ecuación (54), x_i y y_i representan el valor de la recta original en cada punto. En ambas ecuaciones x_m , y_m muestran el valor de todos los puntos en cada eje.

- **E1.4.3 Cálculo del tamaño de la intersección.** Por último, se obtiene el tamaño calculando el área del espacio existente de la intersección entre la recta de regresión y el segmento de la cuerda, tal y como se puede ver en la Figura 74b.



a)



b)

Figura 74. a) Ejemplo de cálculos de la segunda derivada. b) Intersección con la recta de regresión

El tamaño se calcula principalmente para hacer seguimiento ante la ejecución de un tratamiento o rehabilitación, y también es útil para diferenciar entre patologías morfológicas o evaluar su gravedad.

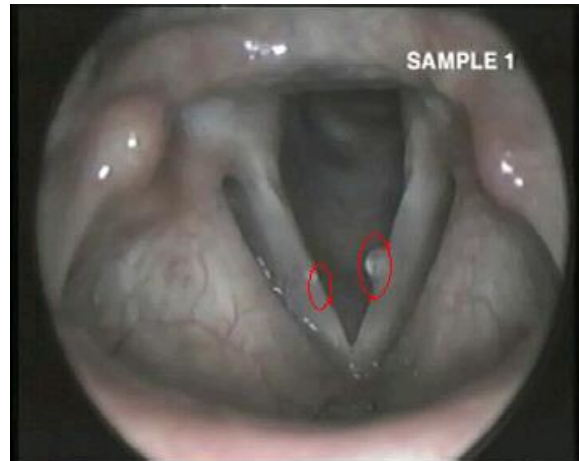


Figura 75. Patologías localizadas y resaltadas en la imagen original

Todas las transformaciones aplicadas hasta ahora nos permiten localizar automáticamente la patología en la imagen, tal y como se puede ver en Figura 75.

4.2.5.2 Análisis y Medida de parámetros relacionados con el movimiento

El bloque E2 (ver Figura 76) tiene como *input* los vectores de movimiento calculados en la etapa C y como *output* los ángulos que forman cada uno de los bordes de las cuerdas vocales con el índice inferior y el centro de masas.

E2.1. Cálculos estadísticos. Para analizar la información proporcionada por la etapa de Análisis de Movimiento, se han utilizado medidas estadísticas para medir la dispersión. El input de este bloque son los vectores de movimiento resultado del Block Matching.

Media: Dados n números $\{a_1, a_2, \dots, a_n\}$, la media aritmética se define como:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n a_i = \frac{a_1 + a_2 + \dots + a_n}{n} \quad (56)$$

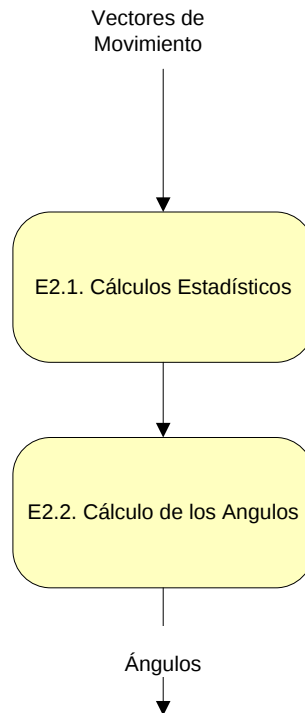


Figura 76. Detalle del Bloque E2.

Varianza: (σ^2) Dada una variable aleatoria X cuya media es $\mu = E(X)$, la varianza ($\text{Var}(X)$, también representada por σ^2_x o sólo σ^2) se define como:

$$\text{Var}(X) = E[(X - \mu)^2] = E(X^2) - \mu^2 \quad (57)$$

Desviación Standard: La desviación Stándard (σ) es la medida de la centralización o dispersión para variables ratio o intervalo, y para variables intervalo. Se define como la raíz cuadrada de la Varianza.

Coeficiente de Variación: Este coeficiente expresa la Desviación Standard como el porcentaje de la media aritmética, mostrando una mayor interpretación del grado de variabilidad de la Desviación Standard. Se calcula usando la siguiente fórmula:

$$C_v = \frac{\sigma}{\bar{x}} \quad (58)$$

El estudio de los resultados de todos ellos nos proporciona información para saber por cuerda vocal si ha habido movimiento, y por comparación, la relación entre el movimiento de la cuerda vocal derecha e izquierda.

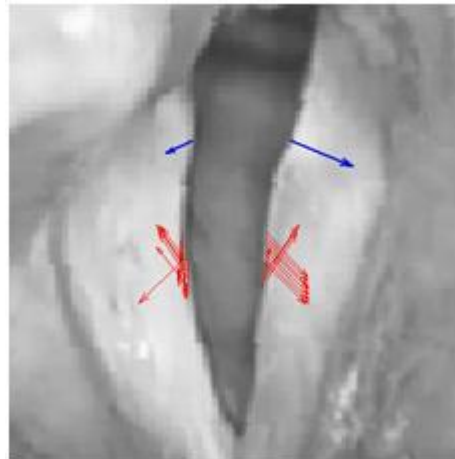


Figura 77. Cuerdas vocales en las que se muestran los vectores de movimiento.

E2.2. Cálculo de los ángulos. Este parámetro puede ser calculado de dos formas: a partir de la segmentación del espacio glótico o a partir de la información de movimiento.

1. A partir del espacio glótico segmentado. El ángulo es calculado estableciendo, por un lado la línea que forman el vértice inferior de las cuerdas vocales y el centro de masas del espacio glótico, y por otro la media de los ángulos calculados en cada punto en el tercio inferior de cada una de las cuerdas vocales.

$$Desviacion(\%) = \frac{|LAV - RAV|}{\max(LAV, RAV)} \times 100 \quad (59)$$

Siendo LAV: Left Angle Value y RAV: Right Angle Value

2. A partir de la información de movimiento. De cada una de las cuerdas vocales tenemos el movimiento medio horizontal y vertical, por ellos mediante (60) es posible calcular el ángulo que forman.

$$A = \text{atan}(v/u) \quad (60)$$

Siendo v el movimiento vertical y u el movimiento horizontal

En este segundo caso, es igualmente interesante realizar la comparación entre el ángulo de la cuerda vocal izquierda y derecha y medir su desviación como en (60).

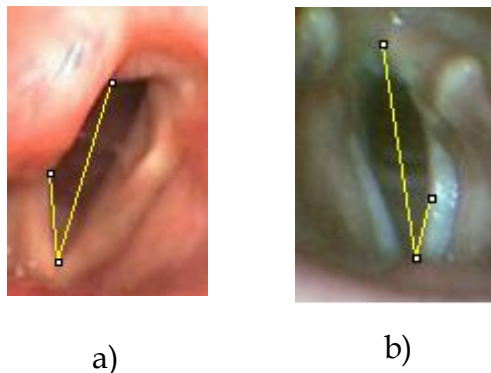


Figura 78. Ejemplos de medida de ángulos de cuerda vocal a) derecha y b) izquierda

Tanto el cálculo del área del espacio glótico segmentado como el cálculo de los ángulos en las cuerdas vocales, pueden ayudar también en la detección del cierre glótico total. La posibilidad de tener un cierre glótico parcial (o en reloj de arena) nos alarma ante algún tipo de patología, y puede ser causa de la disfonía del paciente.

4.2.6 Decisión

Una vez calculadas todas las medidas objetivas explicadas anteriormente, y basándonos en ellas, hay que tomar la decisión final para sugerir un diagnóstico al otorrinolaringólogo (ver Figura 79).

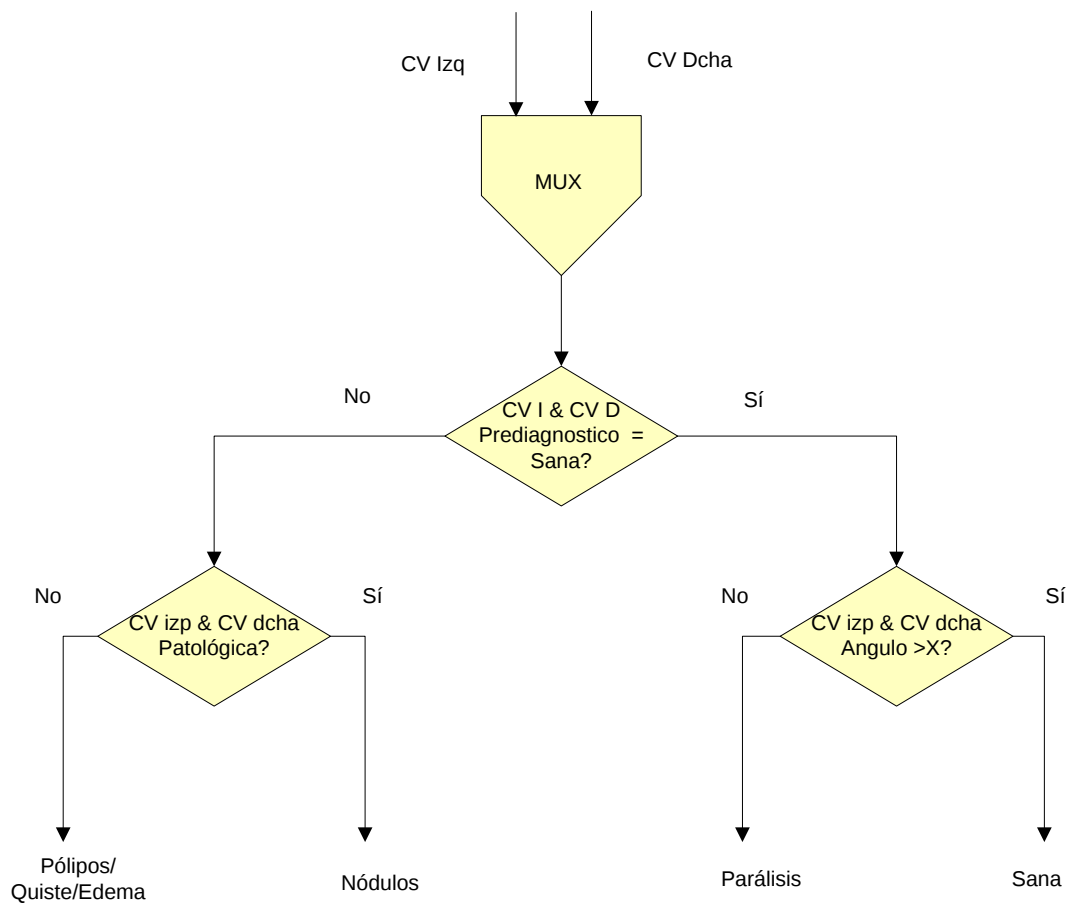


Figura 79. Diagrama de la Decisión/diagnóstico

En primer lugar se tiene en cuenta la primera clasificación a partir de la cual se han ejecutado unos pasos u otros. Si ha sido clasificada como “patología morfológica”, se tienen en cuenta las siguientes premisas:

- Si el problema detectado se encuentra en una cuerda o en ambas.
- En qué parte de la cuerda está la discontinuidad/patología.
- Tamaño

Si ha sido clasificada como “Sana”, tras el análisis se valida o no esta clasificación en función de:

- La media y la varianza calculadas por cuerda a partir de la información de movimiento.
- Los ángulos de apertura de cada cuerda calculados.

$$\sigma^2(ul) < 0,5 \cup \sigma^2(ur) < 0,5 \cap \sigma(\text{mean}(ul, ur)) < 1 \quad (61)$$

Siendo ul : el movimiento horizontal de la cuerda vocal izquierda y ur :
el movimiento horizontal de la cuerda vocal derecha

Si se detecta ausencia de movimiento, o un movimiento muy descompensado entre la cuerda vocal derecha y la izquierda, se puede concluir que el paciente puede sufrir parálisis de cuerda vocal, o al menos una anomalía en el funcionamiento normal de las cuerdas vocales. Empíricamente, esto se ha traducido en la fórmula (61), por cuyos resultados podemos distinguir entre sanas y patológicas (parálisis), y se confirma por los resultados de las medidas de desviación entre el ángulo derecho y el izquierdo.

“Dysphonia is not a singing problem. It's a voice box issue in the muscle on the voice, very different from having a nodule on the vocal cords, which I've never had. I'm lucky that I've never had that. It needs a long renewal time, and even today, I am still addressing it”.

Shania Twain

5.RESULTADOS

Una vez definida la solución propuesta a la problemática planteada, procedemos con la fase de experimentación y análisis de los resultados obtenidos: numéricos y gráficos. En esta tesis, cuyas aportaciones se enmarcan en el ámbito del procesado digital de imagen, los resultados gráficos son altamente significativos para ilustrar algunas de las etapas principales del desarrollo.

Tal y como se ha reflejado en el capítulo de “Diseño”, para ilustrar la experimentación realizada, se tendrá en cuenta que el sistema tiene cinco partes o etapas diferenciadas, y se mostrarán los resultados de cada una de ellas.

En primer lugar, en este capítulo, se abordan los aspectos más destacables sobre la implementación de los algoritmos y, en segundo lugar, la fase de evaluación, donde se describen las pruebas realizadas para la validación de la hipótesis y los resultados obtenidos en cada uno de los escenarios dentro del escenario global.

Se hará especial hincapié en los resultados de las patologías relacionadas con la morfología de las cuerdas vocales frente a las relacionadas con el movimiento,

por el tipo de imágenes estudiadas y por la muestra disponible en la base de datos, y por último se presentarán los resultados obtenidos en diferentes formatos.

5.1 CONSIDERACIONES PREVIAS

En este punto se describe el entorno software y el soporte hardware utilizado para el desarrollo de esta tesis, así como las decisiones que han tenido que ser tomadas durante la implementación, las incidencias encontradas y como han sido solucionadas.

5.1.1 Entorno de desarrollo

El lenguaje de desarrollo utilizado para implementar los algoritmos diseñados ha sido Matlab 7.6.0.324 (R2008a).

Las imágenes de la base de datos, descritas en el capítulo tres son las utilizadas para testear la validez de los algoritmos. Estas imágenes suponen una ingente cantidad de información, cuyo procesamiento puede ser lento y complicado, y se ha realizado 100% en Matlab.

5.1.2 Hardware utilizado para el desarrollo y pruebas

En esta investigación no ha sido necesario hardware específico, excepto el laringoscopio utilizado por los otorrinolaringólogos para realizar las capturas de las imágenes. Pero el desarrollo de esta tesis se ha centrado en el post-procesado de las imágenes proporcionadas por el especialista. Al no realizarse en tiempo real, no es necesario hardware de alto rendimiento de características especiales.

Los requisitos mínimos necesarios para el desarrollo y puesta en marcha de los algoritmos desarrollados son los siguientes:

Procesador	Genuine Intel ® CPU U4100 @ 1,30 GH
Sistema Operativo	Windows 7 Professional
Memoria RAM	4,00 GB

Tabla 4. Especificaciones Técnicas del Hardware

5.2 EVALUACIÓN DE RESULTADOS PARCIALES

En esta sección se describen los resultados obtenidos en las etapas de: segmentación, clasificación, detección del movimiento, y análisis y medida, evaluados durante el proceso de experimentación.

No se consideran resultados parciales, las salidas correspondientes a la etapa de pre-procesado, ya que en esa etapa se han realizado las transformaciones básicas necesarias para obtener buenos resultados en las siguientes etapas, y en sí mismos no son analizados como un resultado.

5.2.1 Pruebas de la Etapa de Segmentación

La etapa de segmentación es la primera y más importante de todo el proceso, ya que de sus resultados depende el éxito de las etapas posteriores. De esta etapa depende también la obtención de los resultados objetivos para la caracterización de las cuerdas vocales y la emisión de los diagnósticos.

En esta etapa podremos observar resultados gráficos y numéricos. Estos últimos para expresar el porcentaje de éxito en la segmentación dentro de la secuencia.

La Figura 80 muestra de forma gráfica los resultados de la etapa de segmentación, y de sus bloques intermedios, descritos en el apartado 4.2.2 del capítulo de “Diseño”. Además, la Figura 80d presenta el resultado de segmentación, la región de interés, del frame trece.

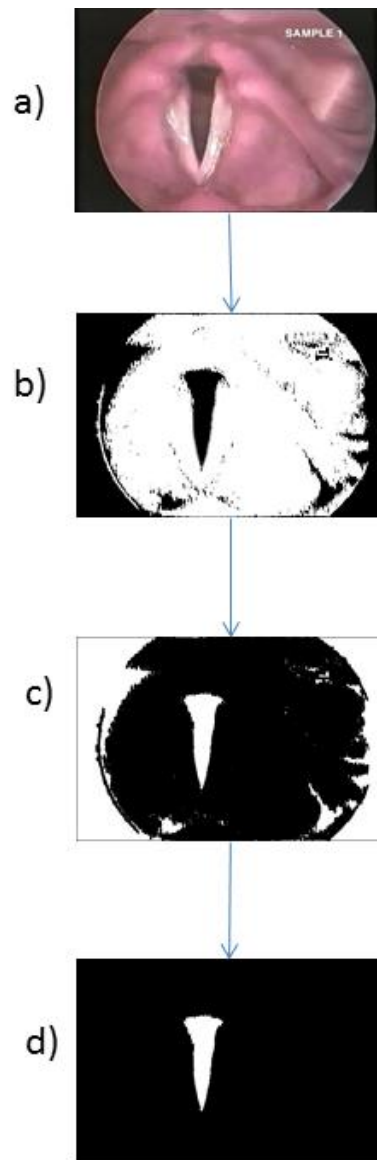


Figura 80. Frame 13 de secuencia de cuerdas vocales sanas. a) Original. b) Resultado tras los filtros de Gabor. c) Resultado tras denoising. d) Máscara resultado del espacio glótico

Pero, tal y como se ha mencionado en el capítulo de Diseño, la elección de la región de interés puede ser un proceso iterativo, si la región de interés no es encontrada en el primer proceso, tal y como se puede observar en la Figura 81.

La Figura 81 es un ejemplo en el que la ROI no ha sido encontrada en la primera iteración, por ello la Figura 81d tiene otra región que no es de nuestro interés dentro del considerado espacio glótico. Esta región tiene que ser eliminada ya

que aunque no afecta al estudio de los contornos de las cuerdas vocales, por el contrario, sí desvirtúa la medida del área del espacio glótico.

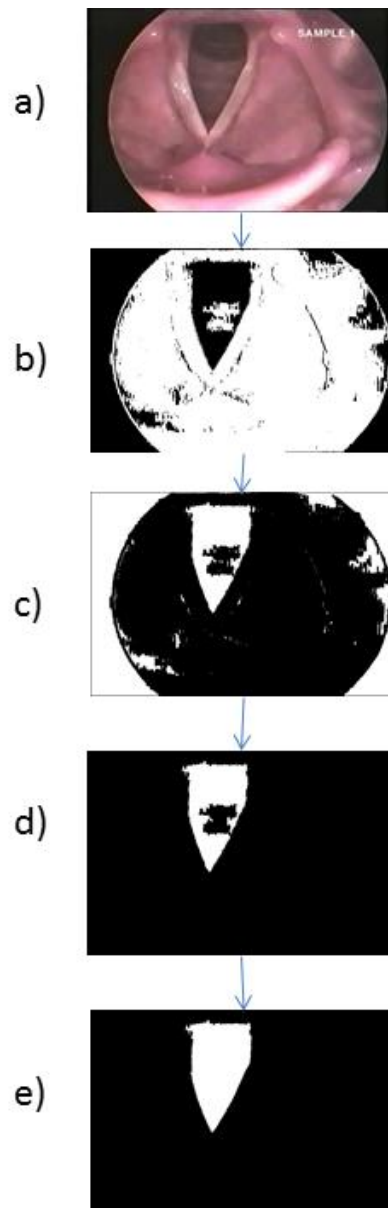


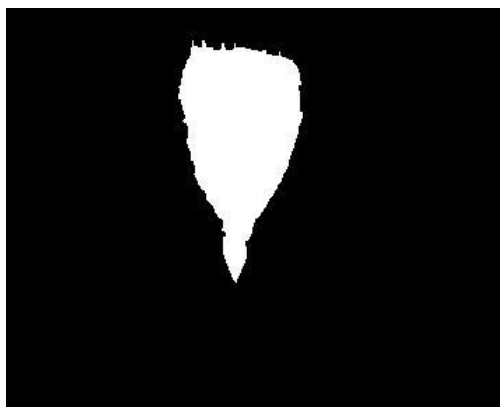
Figura 81. Frame 56 de secuencia de cuerdas vocales sanas. a) Original. b) Resultado tras los filtros de Gabor. c) Resultado tras denoising. d) Máscara errónea del espacio glótico. e) Máscara resultado del espacio glótico

En la Figura 82 se pueden ver dos ejemplos de imágenes de cuerdas vocales patológicas en las que tras aplicar los algoritmos de la etapa de segmentación, se ha obtenido como resultado la ROI (espacio glótico) buscada. La Figura 82a es un ejemplo claro de cuerdas vocales con nódulos, y la Figura 82c es una

imagen de cuerdas vocales con un pólipo. La Figura 82b y la Figura 82d son dos resultados exitosos de la segmentación.



a)



b)



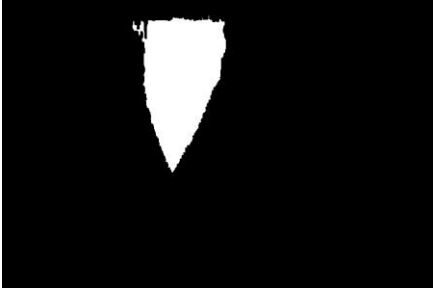
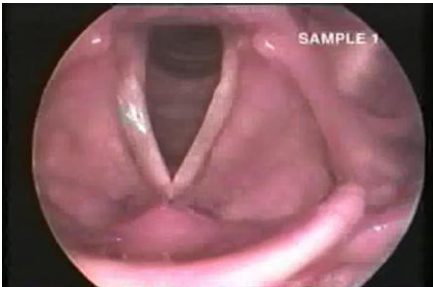
c)



d)

Figura 82. a) Imagen Original de Cuerdas Vocales con Nódulos. b) ROI de a). c) Imagen Original de cuerdas vocales con pólipo. d) ROI de c)

De esta forma se muestra la poca variación inter frame, aun cuando la estroboscopia no captura totalmente el movimiento vibratorio real de las cuerdas vocales. A continuación se pueden observar, en la Figura 83, diez frames de una grabación correspondiente a un paciente con cuerdas vocales sanas.

Frame Número	Imagen Original	Máscara de la ROI Obtenida
60		
61		
62		
63		

64



65



66



67



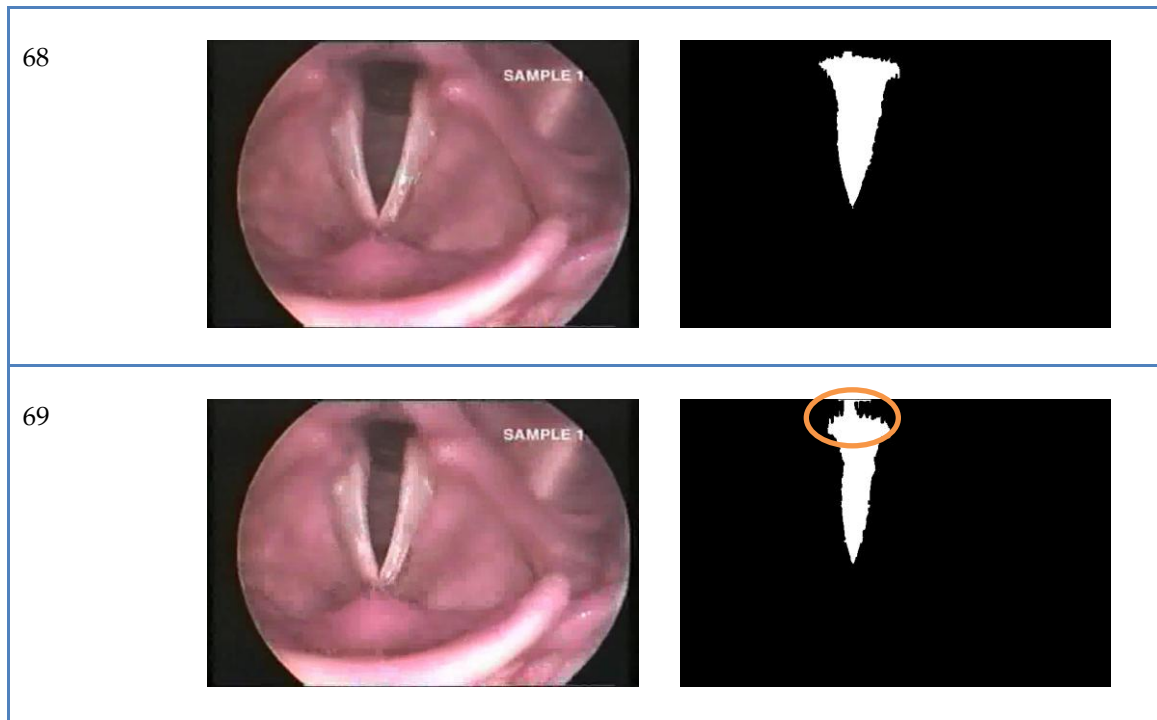


Figura 83. Ejemplo de Secuencia de cuerdas vocales sanas, Frame 60 al Frame 69. Imagen Original y ROI.

Además de múltiples ejemplos de segmentación exitosa, en la Figura 83 también podemos observar algunas ROI con pequeños errores, Frames: sesenta, sesenta y uno, sesenta y dos, y sesenta y nueve, mientras que en los frames sesenta y tres, y sesenta y cinco, no se produce.

Los anteriormente descritos errores no se consideran realmente como tal en la segmentación. La razón es que la mayoría de los cálculos de los parámetros que se obtienen en posteriores etapas utilizan el contorno de las cuerdas, pero solamente se tiene en cuenta el ochenta por ciento del contorno de cada una de las cuerdas, ya que las patologías habitualmente se localizan en el tercio inferior. El único parámetro que se ve afectado es el área del espacio glótico, pero como se ha comentado anteriormente la variación inter frame es mínima, y esta medida no tiene sentido si se hace solo sobre un único frame.

Además de los resultados gráficos presentados anteriormente, en la Tabla 5, se muestran las cuarenta y cinco secuencias estudiadas con su diagnóstico emitido

por el otorrinolaringólogo, el número de frames de cada una y el porcentaje de frames correctamente segmentados en cada una de ellas.

Secuencia	Diagnóstico Previo	Número de Frames	% éxito por secuencia
1	Sana	20	100%
2	Sana	20	90,00%
3	Sana	29	89,60%
4	Sana	20	100%
5	Sana	20	100%
6	Sana	15	100%
7	Sana	20	95,00%
8	Sana	31	96%
9	Sana	18	100%
10	Sana	28	88,90%
11	Nódulos	35	91,40%
12	Nódulos	29	93,10%
13	Nódulos	24	100%
14	Nódulos	20	93,20%
15	Nódulos	27	100%
16	Nódulos	19	100%
17	Nódulos	30	97,40%
18	Nódulos	17	98,10%
19	Nódulos	25	92,30%
20	Nódulos	34	87,90%
21	Pólipos	24	95,80%
22	Pólipos	50	98,00%
23	Pólipos	23	100%
24	Pólipos	28	100%
25	Pólipos	30	96,70%

26	Pólipos	27	96,20%
27	Pólipos	45	86,10%
28	Pólipos	41	84,50%
29	Pólipos	33	97,30%
30	Edema	19	94,70%
31	Edema	35	100%
32	Edema	29	83%
33	Edema	44	81,56%
34	Quiste	25	100%
35	Quiste	19	97,10%
36	Quiste	29	100%
37	Quiste	35	85%
38	Quiste	35	91,20%
39	Parálisis	19	94,70%
40	Parálisis	17	100%
41	Parálisis	25	100%
42	Parálisis	22	97,80%
43	Parálisis	41	93,90%
44	Parálisis	39	95%
45	Parálisis	47	96,73%

Tabla 5. Resultados de segmentación por secuencia

Si analizamos los resultados agrupando las secuencias por patologías podemos afinar un poco más:

- *Secuencias de Cuerdas Vocales Sanas.* La media de frames en estas secuencias es de 22,1 frames, obteniéndose una media de frames correctamente segmentados de 95,95%.
- *Secuencias de Cuerdas Vocales con Nódulos.* La media de frames en estas secuencias es de 26 frames, obteniéndose una media de frames

correctamente segmentados de 95, 34%. Aunque el número medio de frames de las secuencias es ligeramente superior, esto no afecta considerablemente al resultado porcentual de los frames correctamente segmentados.

- *Secuencias de Cuerdas Vocales con Pólipos.* La media de frames en estas secuencias es de 33,44 frames, obteniéndose una media de frames correctamente segmentados de 94, 96%.
- *Secuencias de Cuerdas Vocales con Edema.* La media de frames en estas 4 secuencias es de 31,75 frames, obteniéndose una media de frames correctamente segmentados de 89, 82%.
- *Secuencias de Cuerdas Vocales con Quiste.* La media de frames en estas 5 secuencias es de 28, 6 frames, obteniéndose una media de frames correctamente segmentados de 94, 66%.
- *Secuencias de Cuerdas Vocales con Parálisis.* Para la segmentación y la posterior clasificación, estas imágenes no se diferencian en cada de forma independiente, de los frames de cuerdas vocales sanas. La media de frames en estas 7 secuencias es de 30 frames, obteniéndose una media de frames correctamente segmentados de 96,88%.

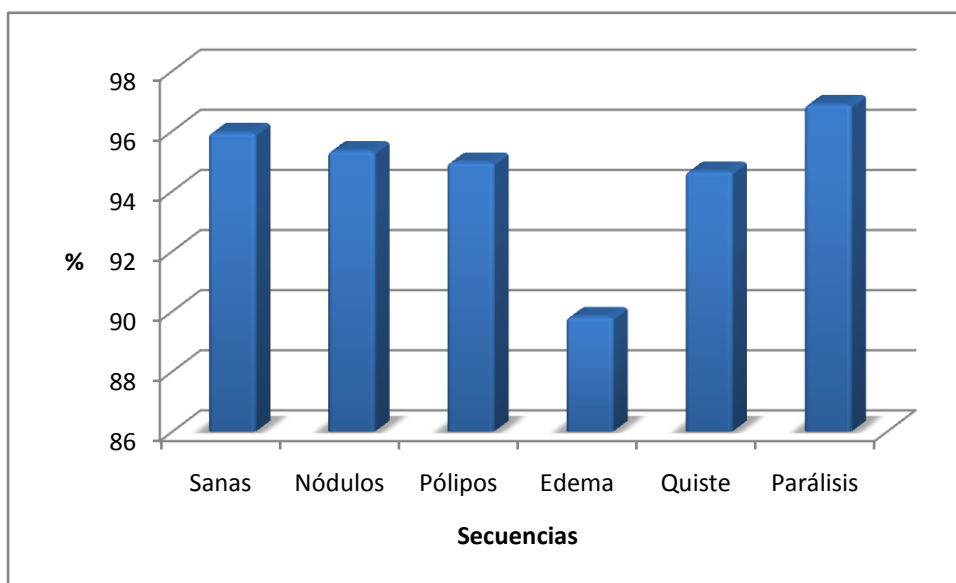


Figura 84. Representación gráfica de los resultados de segmentación

Observando los resultados en la Figura 84, se puede concluir que el porcentaje de frames correctamente segmentados por secuencia es mayoritariamente superior al 95% (concretamente 95,07%), llegando al 100% en muchos casos.

5.2.2 Pruebas de la Etapa de Clasificación

La etapa de clasificación es la tercera etapa del sistema, descrita en el apartado 4.2.3 del capítulo de “Diseño”, y se aplica para proporcionar un pre-diagnóstico de referencia al médico. También se usa como entrada en las etapas posteriores, y dependiendo del resultado de esta etapa se aplican unas transformaciones u otras en la etapa de “Análisis y Medida”.

El elemento crítico en la validación y éxito de esta etapa es la elección del Training Set. El número de imágenes que contiene el Training Set y cuáles tienen que ser estas imágenes son las decisiones más importantes que se han tenido que tomar en esta etapa, y evidentemente, de esta decisión también depende el porcentaje de éxito y el tiempo de proceso de esta etapa, que puede llegar a ser considerable. Todo este sistema es un post-procesado en sí mismo, y se asume que no se realiza en tiempo real.

Durante la experimentación se han realizado pruebas con diferentes Training Sets de 16, 18, 25 y 100 imágenes, siendo el elegido como definitivo uno compuesto por 100 imágenes.

Finalmente, con el método elegido, se han planteado dos experimentos dentro de esta etapa para probar la validez de la técnica PCA: uno para identificar imágenes y otro para clasificarlas.

5.2.2.1 Experimento de Identificación

La prueba que se plantea en este experimento es la clásica [Pujol+01], muy utilizada en procesos de identificación de personas contra una base de datos (o Training Set).

Para chequear la correcta identificación de frames de cuerdas vocales entre imágenes pertenecientes a la misma secuencia y que están contenidas en el Training Set, se han usado 362 imágenes, siendo 100 de ellas para el Training Set. Las 362 imágenes utilizadas para llevar a cabo este experimento pertenecen a frames de las 15 secuencias de la base de datos comercial.

Las pruebas de identificación se han llevado a cabo con otras 262 imágenes. Los resultados obtenidos e pueden ver en Tabla 6.

Número de imágenes en el Trainning Set	Número de Imágenes de Test	% de imágenes identificadas con la patología correcta	% de imágenes correctamente identificadas (sanas o patológicas)
100	262	81,9%	97,3%

Tabla 6. Experimento de Identificación. Resultados en %

En la Tabla 6 se puede observar que el porcentaje de imágenes identificadas exactamente con la patología diagnosticada por el otorrinolaringólogo es de 81,9%, mientras que si sólo tenemos en cuenta las identificadas correctamente con frames del Training Set que simplemente sean sanas o patológicas como las imágenes de test, el porcentaje sube a 97,3%.

Comparando con los resultados de la literatura [Supic08] [Du+07], usando eigenvectors puro los porcentajes varían de 75 a 96%, y en los que se combinan varios métodos para mejorar los resultados los porcentajes pasan a ser de entre 82 y 98%.

Como la identificación de imágenes no es algo que se vaya a dar en la consulta del especialista, ni aporta especialmente a la parametrización, en el apartado

siguiente se describen los resultados de la aplicación de la misma técnica pero para clasificación de imágenes según patología.

5.2.2.2 Experimento de Clasificación

Los resultados cambian si nos centramos en obtener resultados de clasificación, en lugar que de identificación. El concepto de clasificación es completamente distinto. En este caso, se utiliza el mismo Training Set de 100 imágenes usado en el experimento anterior, pero las imágenes con las que vamos a probar (imágenes de test) no pertenecen a las mismas secuencias que las imágenes del Training Set. De esta forma, se puede generalizar el proceso y aplicarlo a cualquier secuencia capturada.

Los resultados obtenidos se pueden ver en la Tabla 7. En este caso, el porcentaje de imágenes correctamente clasificadas con la patología exacta (nódulos, pólipo, edema, quiste...) desciende hasta el 71,8%, mientras que si simplemente atendemos a las clasificadas correctamente como sanas o patológicas el porcentaje es de 92,1%. Este último es el resultado que nos interesa porque de él depende en parte el éxito de la parametrización final.

La parte de clasificación es una parte crucial en esta tesis, por eso se han realizado varias pruebas con diferentes tamaños de Training Sets, e incluso se puede ver en la Tabla 7, experimentos realizados con el mismo tamaño de Training Set, pero que contienen diferentes imágenes.

Se demuestra que cambiar el Training Set hace que los resultados sean completamente distintos, véase en los Training Set de 25 imágenes testeados con 330 imágenes, el porcentaje de imágenes clasificados correctamente con la patología exacta varía del 67,23% al 71,20%. Un 4% de diferencia, que dependiendo de las imágenes del Training Set puede ser mayor.

En la Tabla 7 se demuestra que según vamos aumentando el número de imágenes del Training Sets los resultados mejoran, desde 85% de un Training Set de 16 imágenes a 92,09% del Training Set de 25 imágenes.

Número de imágenes en el Training Set	Número de Imágenes de Test	% de imágenes clasificadas con la patología correcta	% de imágenes correctamente clasificadas (sanas o patológicas)
25	330	67,23	92,09
25	330	71,20	87,65
18	330	69,98	86,34
16	330	68,51	85,00

Tabla 7. Experimento de Clasificación. Resultado de las Eigenfolds en %

Hasta ahora las pruebas realizadas han sido con frames que contienen las dos cuerdas vocales, pero atendiendo a la realidad, cada una de las cuerdas por separado puede ser clasificada como sana o patológica.

En las siguientes pruebas se han tratado cada una de las cuerdas vocales de forma independiente, aunque se han mantenido los mismos training Sets, pero ahora con el doble de imágenes, tal y como muestra la Tabla 8. Un ejemplo de este tipo de Training Set se puede ver en la Figura 85, reducido a 25 imágenes.

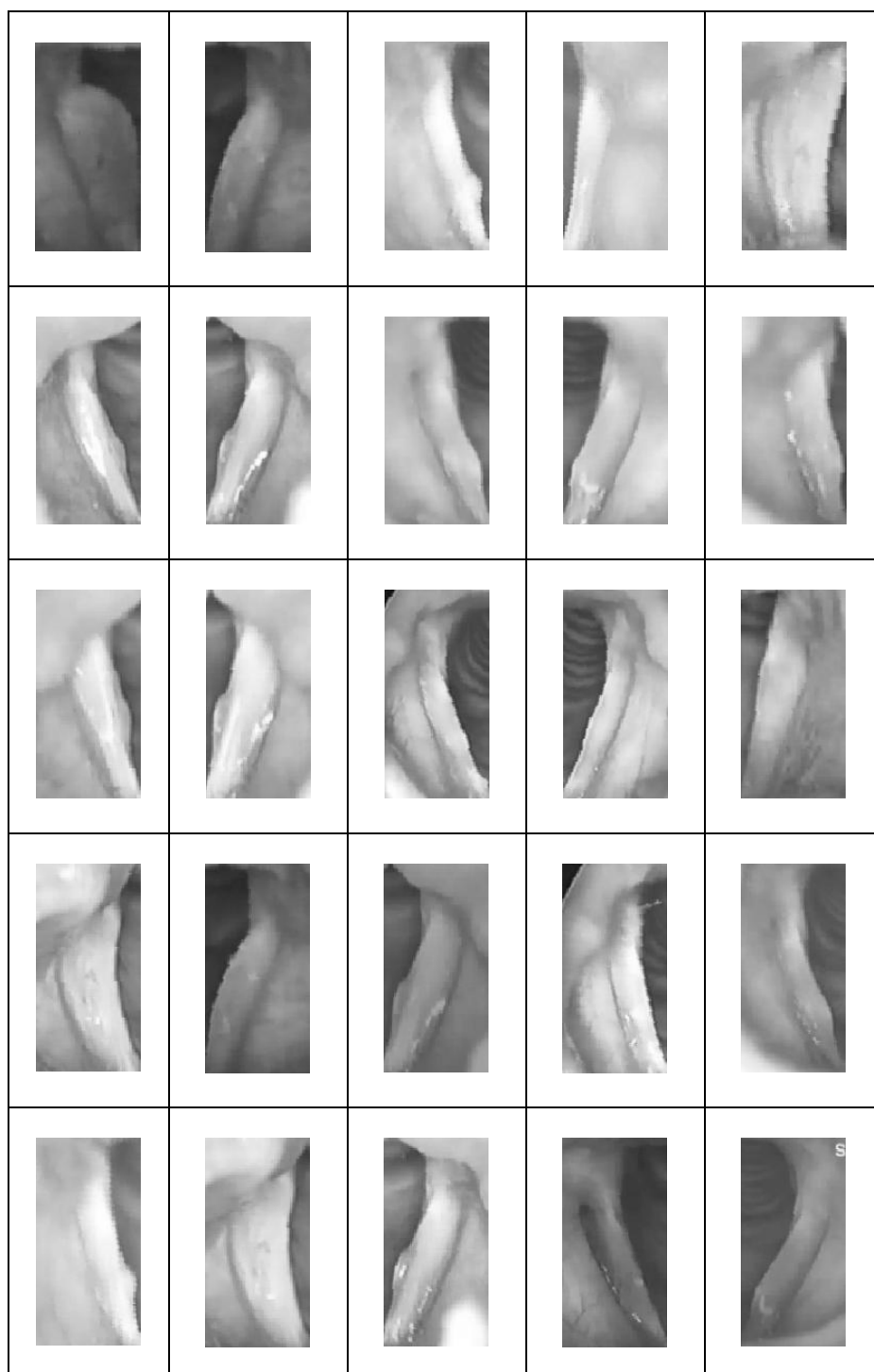


Figura 85. Ejemplo de Training Set con 25 imágenes de cuerdas vocales derecha e izquierda independientemente

Número de imágenes en el Training Set	Número de Imágenes de Test	% de imágenes clasificadas con la patología correcta	% de imágenes correctamente clasificadas (sanas o patológicas)
25 *2	330 *2	65,10%	95,04%
18*2	330*2	64,23%	90,56%
16*2	330*2	37,41%	80,21%

Tabla 8. Resultados de Eigenfolds usando un Training Set de cuerdas vocales tratando de forma independiente la cuerda vocal derecha e izquierda

Observando la Tabla 8, se ve que analizando cada cuerda por separado, los resultados mejoran hasta obtener un 95,04% de éxito diferenciando entre las cuerdas que tienen alguna patología relacionada con la morfología con las que no.

Número de imágenes en el Training Set	Número de Imágenes de Test	% de imágenes clasificadas con la patología correcta	% de imágenes correctamente clasificadas (sanas o patológicas)
100*2	900*2	71,80%	92,10%

Tabla 9. Resultados de Eigenfolds tratando de forma independiente la cuerda vocal derecha e izquierda

Por último, en la Tabla 9 se muestra el resultado considerado como final, utilizando un Training Set de 100 imágenes (200 porque se estudia cada cuerda vocal por separado) probado con 900 imágenes de test (1800 en realidad). El resultado final de esta etapa es que el 92,1% de las imágenes son correctamente clasificadas, y por tanto, la etapa de Análisis y Medida seguirá el proceso correctamente.

5.2.3 Resultados de la Etapa de Detección de Movimiento

La etapa de detección de movimiento se aplica únicamente a aquellas imágenes o secuencias de imágenes que han sido previamente clasificadas como “morfológicamente sanas”, y que realmente son sanas o tienen parálisis de cuerda vocal. En este punto del sistema se mide el movimiento en algunos

puntos del tercio inferior de cada una de las cuerdas vocales, tal y como se ha descrito en la sección 4.2.3.

Se realizan medidas estadísticas, principalmente, media, varianza y desviación típica, poniendo especial atención en la medida de la varianza.

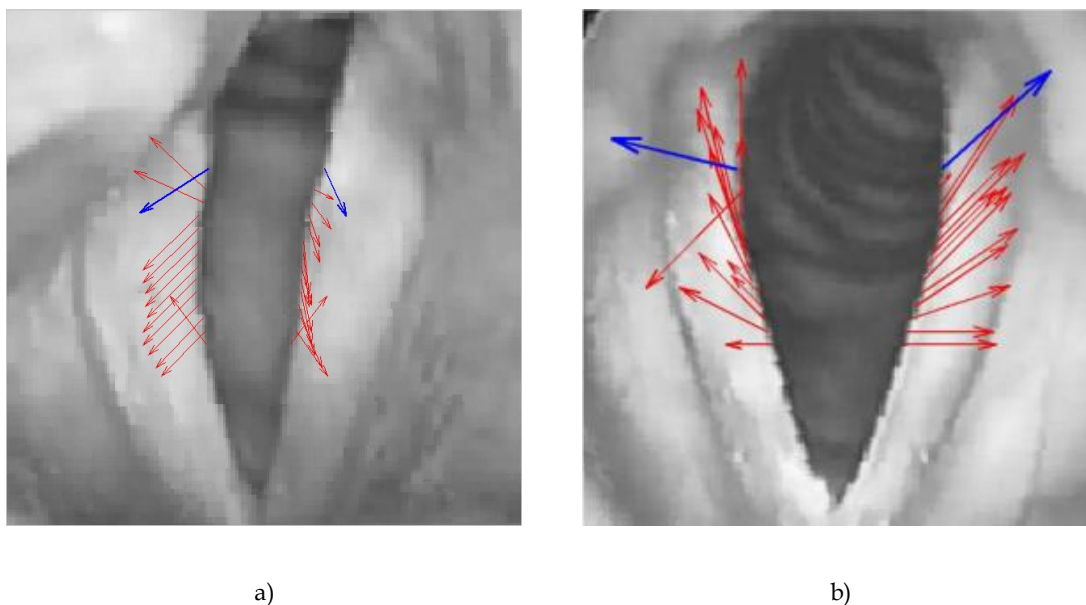


Figura 86. Vectores de movimiento en a) una secuencia de cuerdas vocales con parálisis y b) una secuencia de cuerdas vocales sanas

En la Figura 86 se muestran dos ejemplos de frames estudiados, con y sin parálisis. En rojo aparecen todos los vectores estudiados en el contorno de las cuerdas vocales y calculados mediante técnicas de block matching, y en azul el vector media de todos ellos.

Observando los vectores de la Figura 86a se puede ver cómo una de las cuerdas tiene menor movimiento, y además la cantidad de movimiento representada por el vector media es desigual. Sin embargo en la Figura 86b el movimiento representado es más sincrónico entre ambas cuerdas vocales.

En la Tabla 10 se pueden observar los cálculos estadísticos realizados sobre las secuencias de cuerdas vocales sanas.

Secuencia	Medida	Cuerda Derecha		Cuerda Izquierda	
		u	v	u	v
1	Media	-24,94	11,45	33,33	15,52
	Desviación Standard	16,77	37,47	24,81	41,34
	Varianza	28,13	140,45	61,58	170,91
2	Media	-1,97	2,82	4,12	2,45
	Desviación Standard	1,05	1,80	1,55	2,90
	Varianza	1,11	3,24	2,40	8,41
3	Media	-3,95	5,51	4,93	5,00
	Desviación Standard	0,71	1,092	1,54	2,08
	Varianza	0,50	1,19	2,27	4,34
4	Media	-2,77	2,26	3,64	3,16
	Desviación Standard	2,60	3,54	3,92	4,10
	Varianza	6,76	12,54	15,43	16,87
5	Media	-18,10	0,67	29,89	0,84
	Desviación	0,89	17,41	18,08	2,55

6	Standard				
	Varianza	0,80	30,31	32,71	65,29
	Media	-25,88	20,25	13,96	0,55
7	Desviación Standard	14,36	12,76	0,71	15,50
	Varianza	20,64	16,28	0,50	24,03
	Media	-23,09	0,07	21,57	-24,42
7	Desviación Standard	0,72	16,63	0,53	41,90
	Varianza	0,52	27,66	0,28	175,61
	Media				

Tabla 10. Cálculos estadísticos realizados sobre secuencias de cuerdas vocales sanas

En la Tabla 11 se muestran los resultados de las medidas realizadas sobre las secuencias de cuerdas vocales con parálisis.

Secuencia	Medida	Cuerda Derecha		Cuerda Izquierda	
		u	v	u	v
39	Media	-1,94	1,25	2,60	1,38
	Desviación Standard	0,67	2,31	0,65	2,47
	Varianza	0,45	5,32	0,42	6,13
40	Media	-3,51	-2,20	2,24	-3,17
	Desviación	0,64	4,76	0,31	5,23

	Standard				
	Varianza	0,41	22,67	0,09	27,42
41	Media	-0,34	2,78	1,01	3,06
	Desviación Standard	0,88	1,80	0,65	1,89
	Varianza	0,77	3,24	0,43	3,58
42	Media	-1,00	0,72	0,04	1,61
	Desviación Standard	3,70	3,29	0,59	1,83
	Varianza	13,72	10,83	0,35	3,36
43	Media	0,07	0,11	0,04	1,22
	Desviación Standard	0,67	1,41	0,59	1,50
	Varianza	0,45	1,98	0,35	2,25
44	Media	-0,88	1,61	0,52	3,11
	Desviación Standard	0,65	1,64	0,30	1,75
	Varianza	0,43	2,70	0,09	3,07
45	Media	-0,75	5,38	-0,09	4,38
	Desviación Standard	2,98	8,28	0,53	3,30

	Varianza	8,87	68,52	0,28	10,92
--	----------	------	-------	------	-------

Tabla 11. Cálculos estadísticos realizados sobre secuencias de cuerdas vocales con parálisis

Las secuencias 41, 42 y 45, atendiendo al valor de su varianza y a (61), sufren de parálisis unilateral de cuerda vocal izquierda. Sin embargo las secuencias 39, 40, 43, y 44 según la Tabla 11 padecen cuerdas vocales con parálisis bilateral.

5.2.4 Pruebas de la Etapa de Análisis y Medida

Esta es la etapa en la que se realizan las últimas pruebas del sistema y se puede llegar a sugerir un diagnóstico apoyado en los resultados objetivos obtenidos sobre la secuencia de imágenes capturada por el médico.

Se realizan las siguientes medidas en función de la clasificación obtenida en la etapa anterior de clasificación:

- Clasificadas como Patologías relacionadas con la morfología
 - Localización de la patología en cada una de las cuerdas de la imagen (en caso de que exista)
 - Medida del tamaño de la patología en cada una de las cuerdas (en caso de que exista)
- Clasificadas como morfológicamente sanas
 - Medidas relacionadas con el movimiento, mostradas en el apartado 5.2.3.
 - Medida de ángulos de apertura

Previo a mostrar los resultados de los parámetros anteriormente comentados, se muestran los resultados de la medida en píxeles del espacio glótico de una secuencia en forma gráfica, asumiendo que este parámetro puede no ser muy significativo utilizando imágenes estroboscópicas.

En la Figura 87 se muestra el Glottal Area Waveform de una secuencia de la base de datos de cuerdas vocales sanas. Como se puede apreciar, no se dispone

de información sobre el cierre de las cuerdas vocales, pero sí se puede observar gráficamente cómo se van abriendo progresivamente.

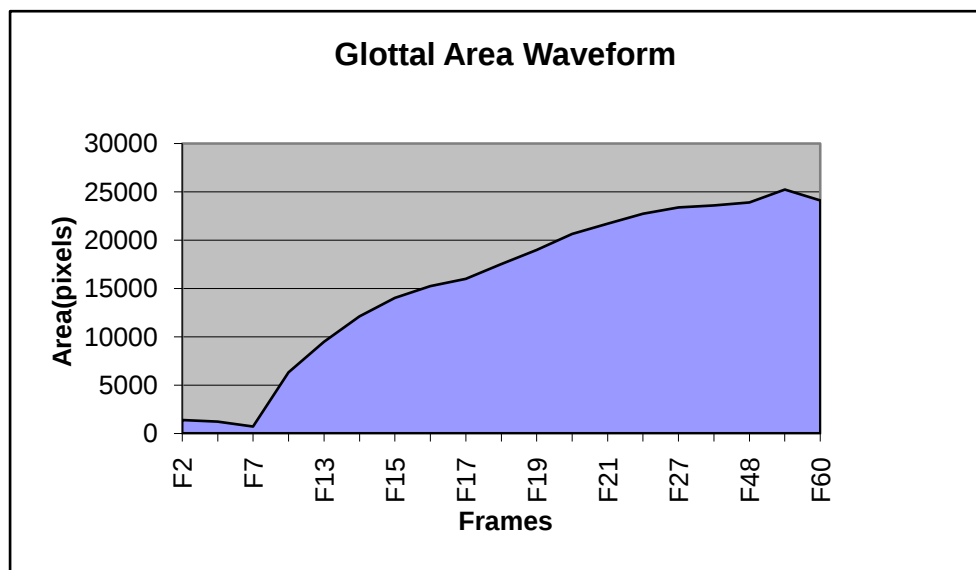


Figura 87. Glottal Area Waveform de una secuencia de cuerdas vocales sanas

5.2.4.1 Análisis y Medida de las imágenes de cuerdas vocales clasificadas como morfológicamente patológicas

Aunque a priori el sistema propuesto no analizaría la morfología de las cuerdas vocales sanas o con parálisis, ya que en la etapa anterior hubieran sido clasificadas como “sanas”, se han aplicado los mismos algoritmos que a las que sí tienen alguna patología relacionada con la morfología para observar las diferencias, poder establecer umbrales y poder comprobar la eficacia del método.

Entre las patologías que, en principio, concentra este grupo se encuentran:

- Nódulos
- Pólipos
- Quiste
- Edema

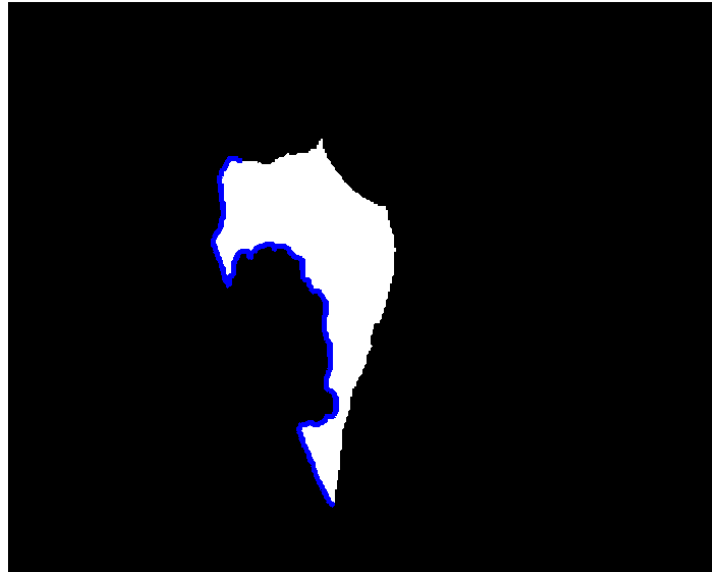


Figura 88. Imagen segmentada de cuerdas vocales con pólipo con contorno resaltado

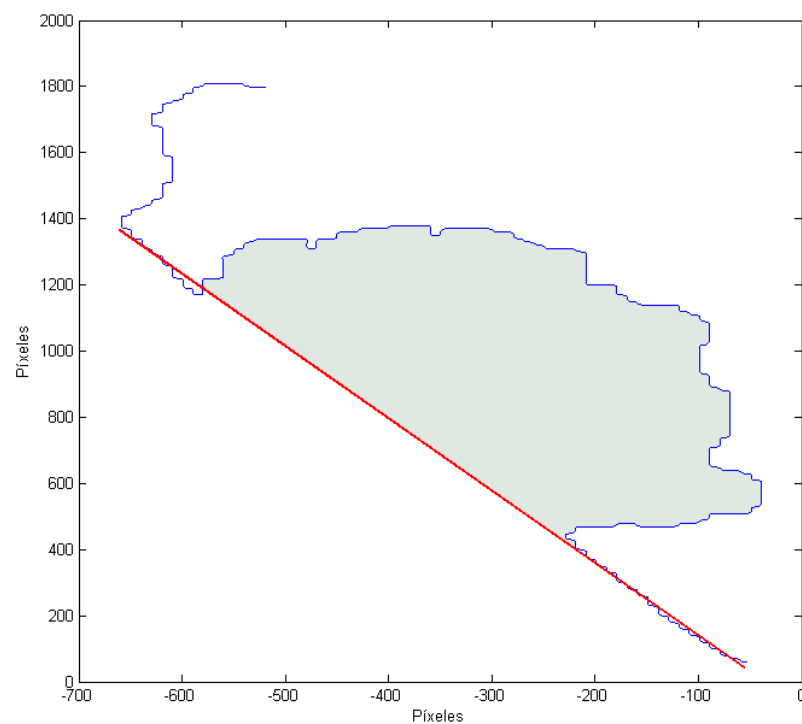


Figura 89. Resultado gráfico del cálculo del área respecto a estimación del contorno de cuerda sana

En la Figura 89 se puede observar el resultado gráfico de la medida del área de la patología, en este caso el pólipo de la Figura 88. En azul se muestra el contorno de la cuerda vocal afectada, en rojo, la estimación ideal del contorno de la cuerda vocal sana, y en verde el área a medir.

Los resultados de las medidas relacionadas con la morfología se pueden ver en la Tabla 12. En ella se muestran las secuencias de la 11 a la 38, que son las que corresponden al grupo de patologías previamente mencionadas. De cada secuencia se calcula el tamaño medio del área de la parte patológica de cada una de las cuerdas vocales y la desviación estándar calculada entre todos los frames.

Secuencia	Diagnóstico Previo	Tamaño de la patología de la cuerda vocal dcha. (en píxeles)	Desviación Standard	Tamaño de la patología de la cuerda vocal izq. (en píxeles)	Desviación Standard	Decisión Final
11	Nódulos	95,5	0,8	120	0,4	Nódulos
12	Nódulos	32,5	1,8	36,3	0,5	Nódulos
13	Nódulos	157	1,1	156	0,7	Nódulos
14	Nódulos	67	2,6	78	1,1	Nódulos
15	Nódulos	57,3	1,7	92	1,7	Nódulos
16	Nódulos	83,8	2,1	67	2,2	Nódulos
17	Nódulos	47,3	3,1	93	0,7	Nódulos
18	Nódulos	77,3	0,9	101,5	1,4	Nódulos
19	Nódulos	92,1	2,2	46,8	0,6	Nódulos
20	Nódulos	58,1	1,3	69,1	1	Nódulos
21	Pólipos	97,9	0,8	0	0	Cuerda Dcha. Patológica
22	Pólipos	2,3	0,1	101,8	2,1	Cuerda Izq. Patológica
23	Pólipos	1,1	0,1	543,1	1,1	Cuerda Izq. Patológica
24	Pólipos	0	0	56,3	2,1	Cuerda Izq. Patológica
25	Pólipos	0,6	0,2	79,7	1,9	Cuerda Izq. Patológica
26	Pólipos	87,2	1,1	2,1	0,7	Cuerda Dcha.

						Patológica
27	Pólipos	69,9	1,8	0	0	Cuerda Dcha. Patológica
28	Pólipos	49,9	0,9	0,8	0,3	Cuerda Dcha. Patológica
29	Pólipos	102	1,4	1,1	0,3	Cuerda Dcha. Patológica
30	Edema	0	0	925	0,6	Cuerda Izq. Patológica
31	Edema	867	2,5	0,5	0,1	Cuerda Dcha. Patológica
32	Edema	452	3,2	0	0	Cuerda Dcha. Patológica
33	Edema	0	0	729	2,9	Cuerda Izq. Patológica
34	Quiste	0	0	90,9	1,1	Cuerda Izq. Patológica
35	Quiste	0	0	59,2	1,1	Cuerda Izq. Patológica
36	Quiste	0	0	95,1	0,7	Cuerda Izq. Patológica
37	Quiste	110,5	0,5	1	0,2	Cuerda Dcha. Patológica
38	Quiste	92,1	2,1	0	0	Cuerda Dcha. Patológica

Tabla 12. Medidas de área de la patología de cada cuerda vocal (cuerdas morfológicamente patológicas)

Analizando con atención la Tabla 12, se observan diferencias significativas entre las patologías, aunque sólo se ha indicado qué cuerda es patológica, o en el caso de ser ambas, se considera que la patología que padece el paciente son nódulos.

Si nos fijamos en las patologías que aparecen en una única cuerda vocal, destaca por tamaño el edema, que según su definición es una inflamación de toda la

longitud de la cuerda vocal, por esa razón se observa que su tamaño en píxeles es mayor que en los quistes o pólipos.

Secuencia	Diagnóstico Previo	Tamaño de la patología de la cuerda vocal dcha. (en pixels)	Desviación Standard	Tamaño de la patología de la cuerda vocal izq. (en pixels)	Desviación Standard	Decisión Final
1	Sana	0,2	0	2,4	1,1	Morfológicamente Sana
2	Sana	0,1	0	0	0	Morfológicamente Sana
3	Sana	0,1	0	0	0	Morfológicamente Sana
4	Sana	5,2	2,9	0	0	Morfológicamente Sana
5	Sana	0	0	0	0	Morfológicamente Sana
6	Sana	0	0	0	0	Morfológicamente Sana
7	Sana	3,1	0,5	2,3	0,7	Morfológicamente Sana
8	Sana	0,3	0,1	0,8	0,3	Morfológicamente Sana
9	Sana	0	0	0	0	Morfológicamente Sana
10	Sana	0,1	0,2	0	0	Morfológicamente Sana
39	Parálisis	0	0	0	0	Morfológicamente Sana
40	Parálisis	0	0	0	0	Morfológicamente Sana
41	Parálisis	1,1	0	2,9	0,1	Morfológicamente Sana
42	Parálisis	0,9	0	0	0	Morfológicamente Sana
43	Parálisis	1,4	0,2	0	0	Morfológicamente Sana
44	Parálisis	5,6	1,1	6	2,2	Morfológicamente Sana

45	Parálisis	3,5	1,7	5,1	1,9	Morfológicamente Sana
----	-----------	-----	-----	-----	-----	-----------------------

Tabla 13. Medidas de área de la patología de cada cuerda vocal (cuerdas morfológicamente sanas)

En la Tabla 13 se muestran las mismas medidas que en la Tabla 12, pero en este caso las medidas se han realizado sobre las secuencias de cuerdas vocales previamente clasificadas como “morfológicamente sanas”. Aunque el sistema por defecto no hace estos cálculos sobre estas secuencias, se han querido mostrar para observar la clara diferencia que existe entre las medidas realizadas en ellas y en las anteriores. En estos casos, las áreas en píxeles son muy pequeñas o prácticamente cero, por lo tanto, se ha considerado que áreas por debajo del umbral de 20 píxeles no son significativas y no representan patología morfológica objeto de estudio/alarma.

5.2.4.2 Análisis y Medida de las imágenes de cuerdas vocales clasificadas como morfológicamente sanas

En este punto se analizan únicamente las secuencias clasificadas por el sistema y por el médico como sanas o con parálisis. Las características morfológicas de ambas son las mismas, pero tanto con las medidas de movimiento como con los ángulos de apertura se observan muchas diferencias, tal y como se muestra en la Tabla 14.

En la segunda columna de la Tabla 14, que corresponde al rango de medida del ángulo de las imágenes de la secuencia vemos que no en todas ellas disponemos de la información de cierre total (0 grados concierne al cierre glótico total), eso sólo ocurre en las secuencias 39, 40 y 42. En las otras cuatro secuencias no se dispone del momento de cierre de las cuerdas vocales.

En el capítulo 4 se ha explicado la forma de calcular el ángulo de apertura de cada una de las cuerdas vocales y en la Tabla 14 (ídem en la Tabla 15) se muestra la variación en grados del ángulo derecho e izquierdo. Pero el parámetro más importante y el que nos va a indicar si realmente estamos antes

una parálisis de cuerda vocal o no es la desviación entre en ángulo derecho y el izquierdo, calculada en %.

Secuencia	Rango Ángulo total (grados)	Δángulo Izq. (grados)	Δángulo dcho. (grados)	D(%)	Signo	Diagnóstico Final del Sistema
39	0 – 37,86	1,23	3,53	65,15%	-	Parálisis
40	0 – 39,63	4,11	0,6	85,40%	+	Parálisis
41	2,72 – 40,03	4,88	0,78	84,01%	+	Parálisis
42	0 – 10,68	0,64	0,45	29,68%	+	Parálisis
43	9,71 – 37,07	2,67	0,3	88,76%	+	Parálisis
44	44,82 – 67,82	0,69	3,03	25,48%	-	Parálisis
45	1,76 – 30,14	0,23	1,68	86,30%	-	Parálisis

Tabla 14. Resultados de las medidas de ángulos de secuencias con parálisis

Si la desviación está por debajo del 20%, como en la Tabla 15, estamos ante secuencias de cuerdas vocales sanas, si por el contrario es superior, como en la Tabla 14, nos encontramos con cuerdas vocales afectadas por parálisis.

Otro dato importante que aparece en estas tablas es la columna signo, que indica a cuál de las dos cuerdas afecta la parálisis.

Secuencia	Ángulo total (grados)	Δángulo Izq. (grados)	Δángulo dcho. (grados)	D(%)	Signo	Diagnóstico Final del Sistema
1	0 – 53,75	2,57	2,37	7,78%	+	Sana
2	0- 48,49	5,22	5,98	12,70%	-	Sana
3	0 – 46,28	3,46	2,88	16,76%	+	Sana
4	35-51,30	4,23	3,45	18,44%	-	Sana
5	18,34-39,89	2,39	2,13	10,88%	+	Sana
6	7,80 – 41,80	3,23	3,85	16,10%	+	Sana
7	24,1 – 49,65	1,23	1,47	16,33%	+	Sana

8	9,98 – 24,67	3,15	3,81	17,32%	-	Sana
9	21,45 – 51,78	1,09	1,24	12,10%	-	Sana
10	18,34 – 53,01	2,76	3,01	8,31%	+	Sana

Tabla 15. Resultados de las medidas de ángulos de las secuencias sanas

Otra forma de evaluar los resultados del movimiento es mediante la fórmula (61), a partir de los resultados de la sección 5.2.3. Esta es otra forma de distinguir entre sana y parálisis, en combinación o al margen de la medida de los ángulos mostrados en la Tabla 14 y Tabla 15.

Secuencia	Resultado de (61)
1	1
2	1
3	1
4	1
5	1
6	1
7	1
8	1
9	1
10	1
39	2
40	2
41	2
42	2
43	2
44	2
45	2

Tabla 16. Resultados del análisis del movimiento según (61)

Para interpretar Tabla 16, hay que saber que se ha etiquetado con “1” todos los resultados que cumplían (61) y por lo tanto indicaban que eran cuerdas vocales

sanas y con “2” a las que sufrían parálisis, y efectivamente coincide con el diagnóstico hecho por el otorrinolaringólogo de estas secuencias de la base de datos.

5.3 REPRESENTACIÓN GRÁFICA DE LOS RESULTADOS

En este apartado se pone de manifiesto la necesidad de exponer los resultados de una forma clara y fácilmente interpretable, porque tan importante es la correcta medida de los parámetros de cuerdas vocales como la representación de los resultados.

5.3.1 Representación gráfica de los resultados por secuencia

Los parámetros más representativos para caracterizar las patologías de las cuerdas vocales de los analizados en esta tesis son:

- Tamaño derecha y tamaño izquierda. Estos parámetros se refieren, al tamaño en píxeles que tiene la patología en caso de haberla.
- Desviación. Este parámetro mide la desviación entre el ángulo formado con la cuerda vocal derecha e izquierda. Es el resultado de aplicar la fórmula (59).
- Varianza de la cuerda vocal derecha e izquierda.
- Desviación I-d. Este parámetro es el resultado del cálculo de la desviación típica entre la media de los resultados obtenidos de la medida de movimiento en la cuerda vocal derecha e izquierda (en cada frame).

Para conseguir una representación gráfica normalizada de los seis parámetros y fácilmente reconocible, se han tenido que hacer algunas transformaciones en algunos de ellos. En el caso de los tamaños y la desviación, la normalización se ha conseguido simplemente dividiendo entre el valor límite entre cuerdas vocales sanas y patológicas (20 en el caso

de la desviación y 10 en el caso de los tamaños de las patologías de cada una de las cuerdas. Sin embargo, para poder seguir manteniendo la unidad como el límite de normalidad (no malignidad) de los parámetros de las cuerdas vocales, para la varianza del movimiento de las cuerdas vocales izquierda y derecha, y para la desviación I-d, se ha representado el inverso de cada uno de ellos.

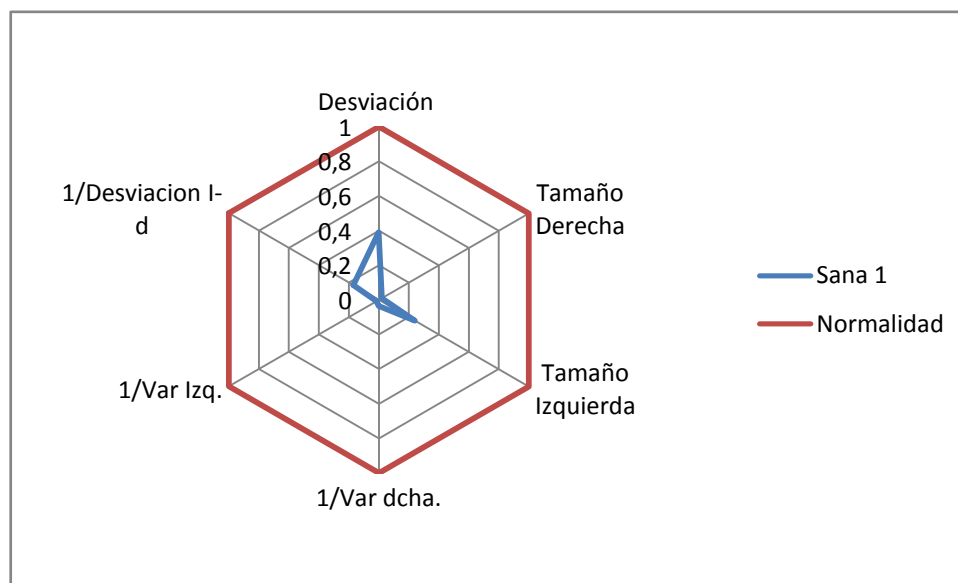


Figura 90. Representación gráfica de los resultados de la secuencia Sana 1

La Figura 90 y la Figura 91 muestran la representación gráfica de los 6 parámetros relativos a la secuencia “Sana 1” y “Sana 2”. En este caso la normalidad está representada mediante el hexágono rojo. Los resultados de la secuencia “Sana 1” y “Sana 2” se encuentran todos en el interior de este hexágono, por lo tanto, gráficamente se ve que las secuencias pertenecen a pacientes con cuerdas vocales sanas.

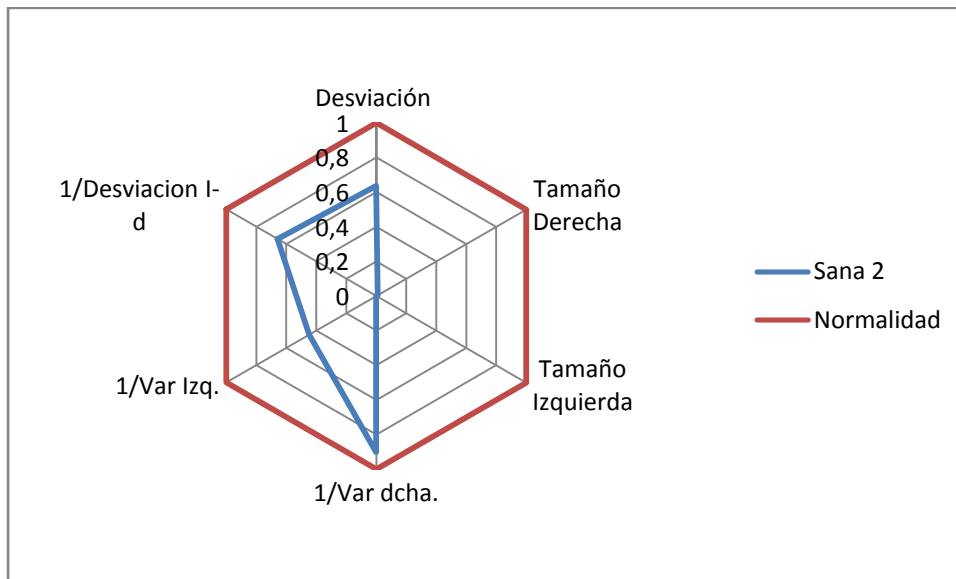


Figura 91. Representación gráfica de los resultados de la secuencia Sana 2

La Figura 92 y la Figura 93 son dos ejemplos de secuencias de cuerdas vocales patológicas. Se puede observar en ellas que cuatro de los seis parámetros se encuentran fuera del rombo de normalidad. Dado que en ambos casos son secuencias de cuerdas vocales con parálisis, los cuatro parámetros mencionados pertenecen a las medidas de movimiento.

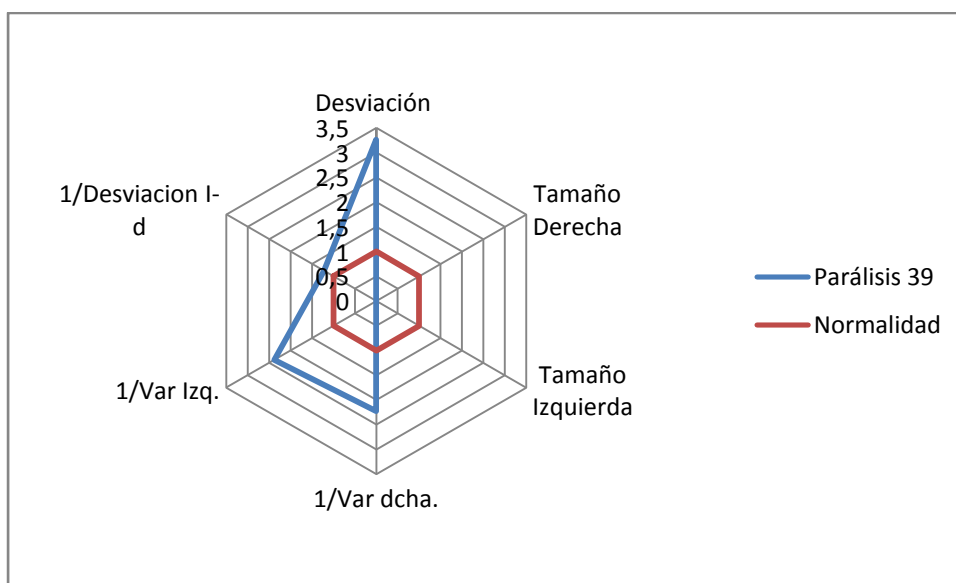


Figura 92. Representación gráfica de los resultados de la secuencia Parálisis 39

En la Figura 92 no se acentúa prácticamente ninguno de los tres parámetros relacionados con el movimiento. Simplemente observando el gráfico podríamos concluir que es una parálisis bilateral.

La Figura 93 muestra una descompensación entre los parámetros 1/Var Izq y 1/Var dcha. Aunque habría que hacer un estudio más a fondo, podríamos estar ante una parálisis unilateral de cuerda vocal.

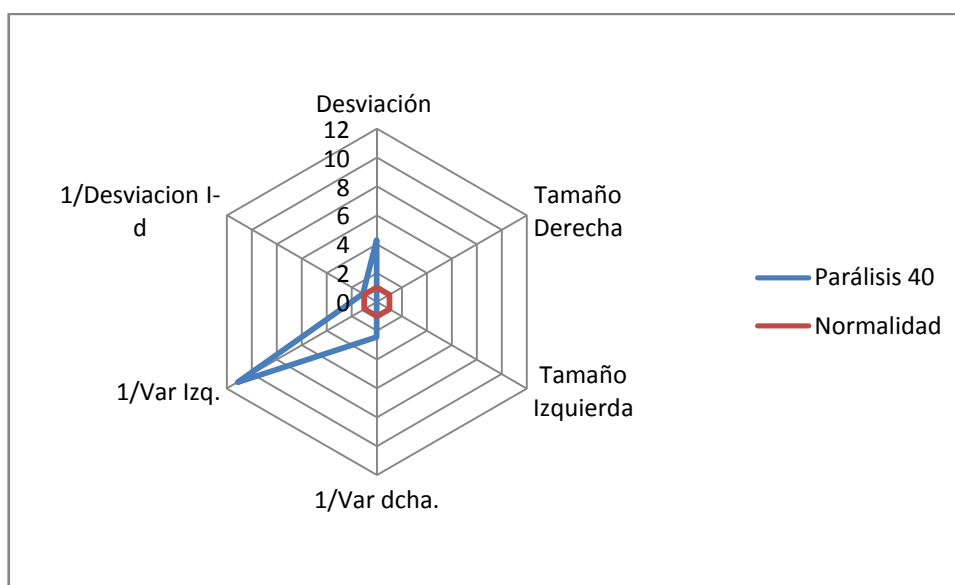


Figura 93. Representación gráfica de los resultados de la secuencia Parálisis 40

La Figura 94 es otro ejemplo de representación gráfica de resultados de una secuencia patológica, pero en este caso es una patología relacionada con el movimiento: nódulos. Es fácil de identificar en la figura porque los parámetros tamaño izquierda y derecha son los que resaltan y están completamente fuera del hexágono que representa la normalidad.

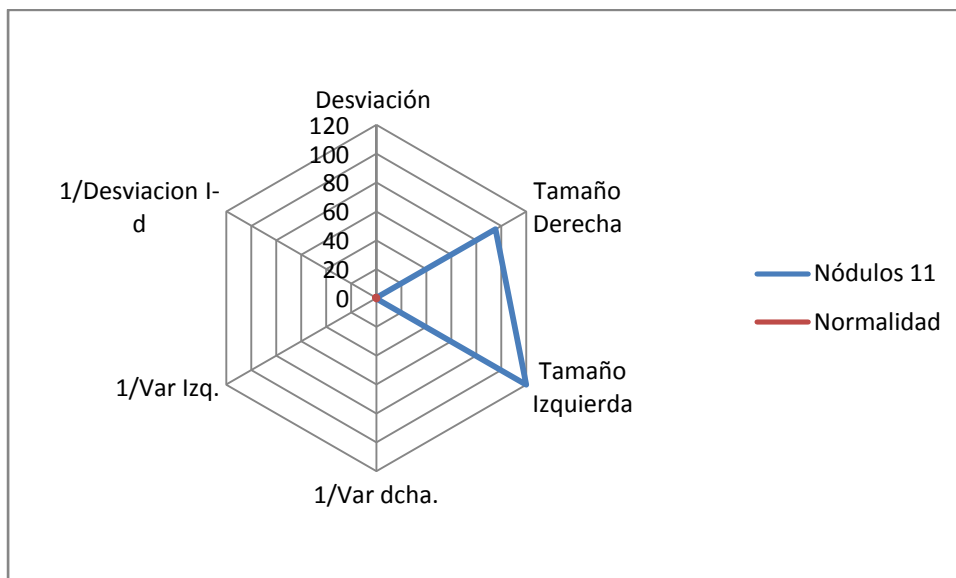


Figura 94. Representación gráfica de los resultados de la secuencia Nódulos 11

5.3.2 Representación gráfica reducida de los resultados por secuencia

En la versión reducida de los resultados, se presentan gráficamente menos parámetros, porque uno de ellos, el llamado “Movimiento” aúna las medidas realizadas para parametrizar el movimiento de las cuerdas vocales (según (61)), indicando si es normal o no.

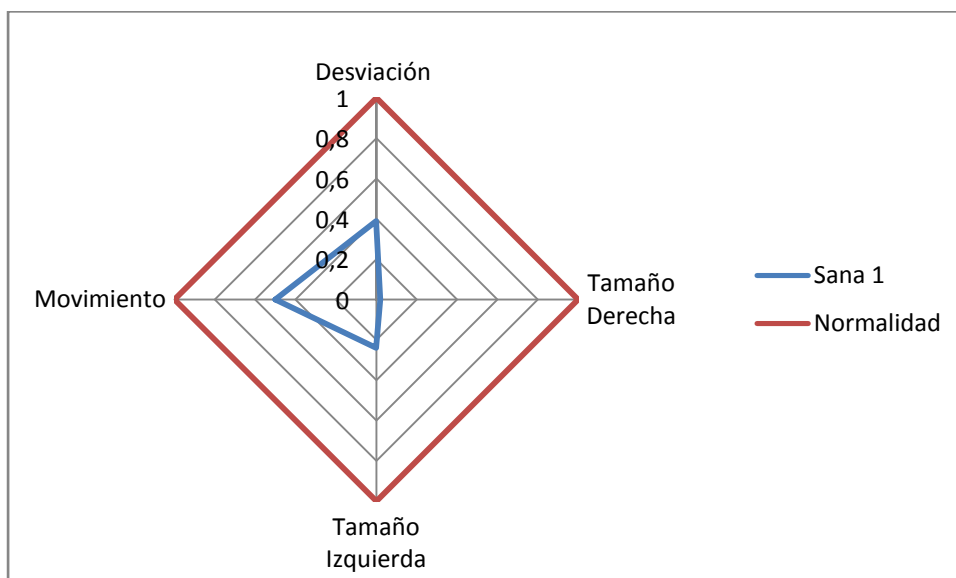


Figura 95. Representación gráfica reducida de los resultados de la secuencia Sana 1

La Figura 95, la Figura 96, la Figura 97 y la Figura 98 muestran la representación gráfica mencionada anteriormente. En todas ellas aparece el rango de normalidad representado en rojo, mediante el “rombo unidad”. Si en una secuencia todos los parámetros que se encuentren dentro de este rombo, podemos concluir que es una secuencia de cuerdas vocales sanas, como se puede ver en la Figura 95 y la Figura 96. Los parámetros propios de cada secuencia están representados en azul.

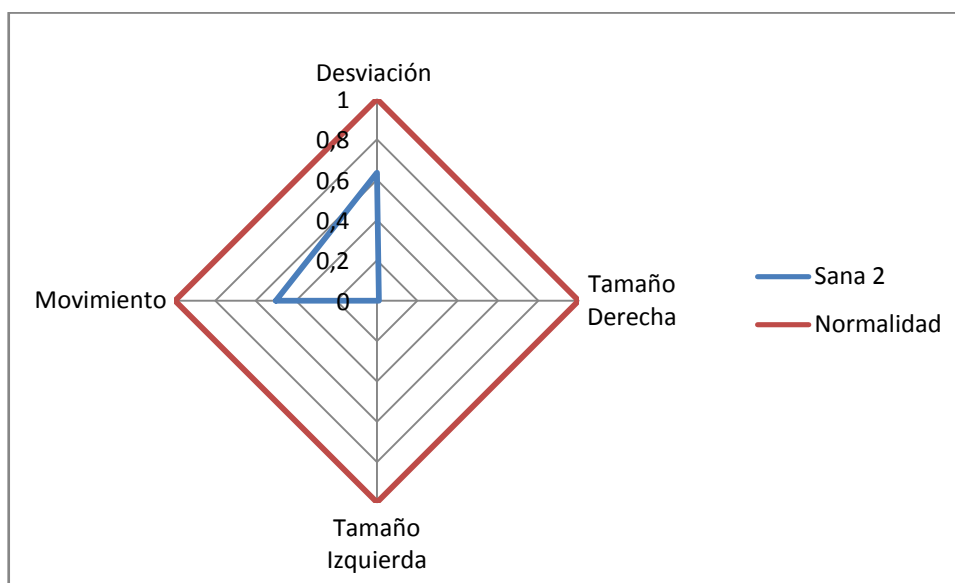


Figura 96. Representación gráfica reducida de los resultados de la secuencia Sana 2

La Figura 97 y la Figura 98 pertenecen a la representación gráfica de dos secuencias de cuerdas vocales que padecen parálisis. En este caso los parámetros medidos están fuera del rombo unidad, y por lo tanto, se puede ver que las cuerdas vocales están afectadas por alguna patología.

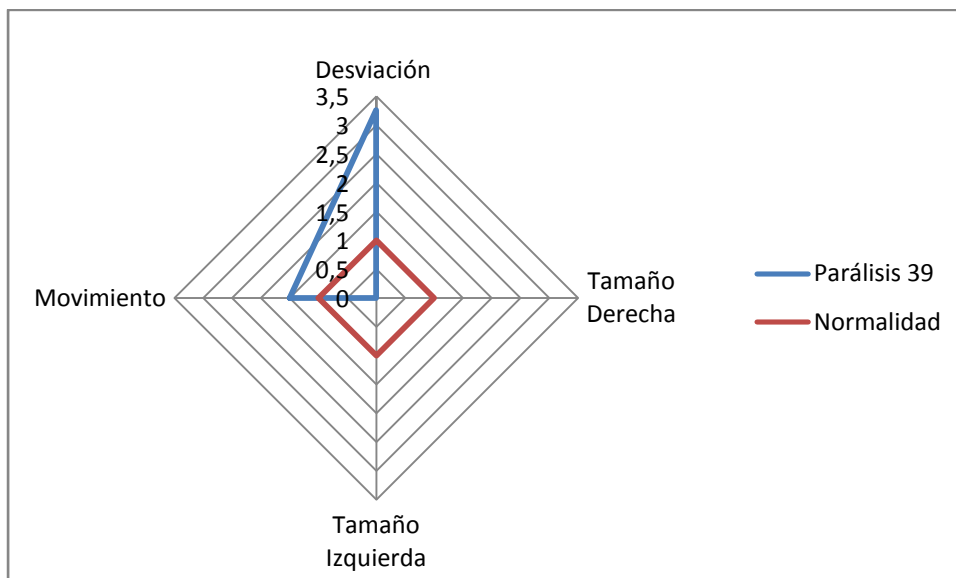


Figura 97. Representación gráfica reducida de los resultados de la secuencia Parálisis 39

Con estas representaciones gráficas se pretende facilitar la interpretación de la información por parte del médico otorrinolaringólogo, que la podrá adjuntar a sus informes, o incluso puede ser útil en la comparación de exploraciones en caso de evaluación de la evolución.

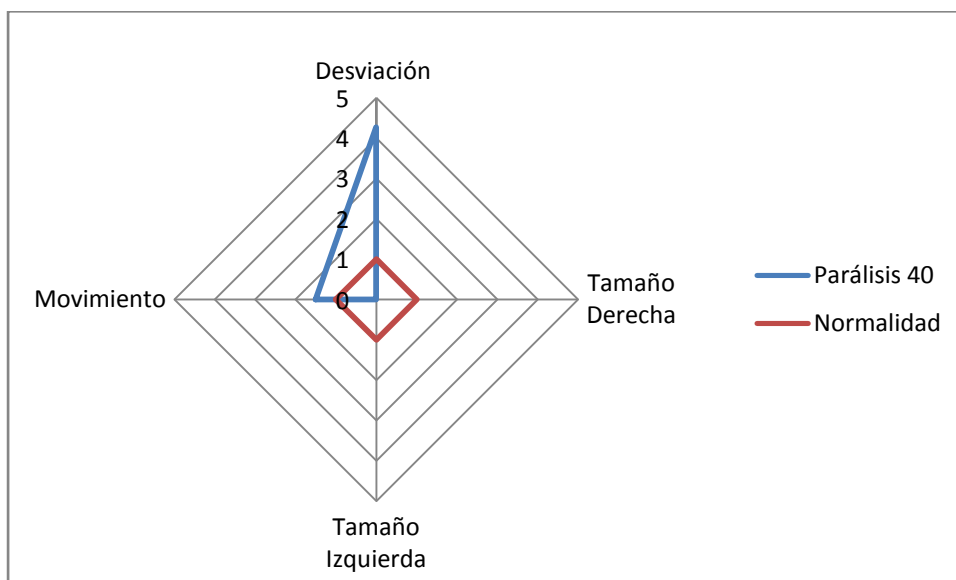


Figura 98. Representación gráfica reducida de los resultados de la secuencia Parálisis 40

La Figura 99 es completamente distinta a las anteriores. En ella se ven reflejados los resultados de una secuencia patológica, concretamente con quiste en la cuerda vocal izquierda. El rombo unidad aparece representado casi como un

punto en origen, mientras el parámetro tamaño izquierda aparece notablemente acentuado.

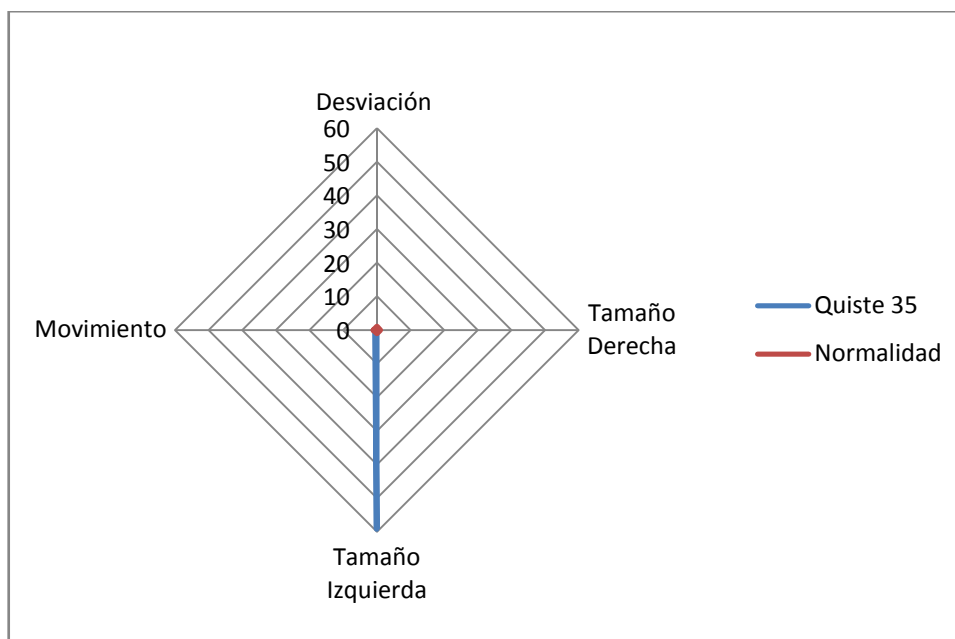


Figura 99. Representación gráfica reducida de los resultados de la secuencia Quiste 35

5.3.3 Representación gráfica final de los resultados

En este apartado se presentan los gráficos resumen de la clasificación automática realizada mediante el algoritmo propuesto, en base a las medidas realizadas.

Las figuras que se presentan a continuación (Figura 100 y Figura 101) muestran los resultados obtenidos de cada una de las secuencias estudiadas (1-45), y su ubicación en función del umbral marcado en rojo. En ambas gráficas, por encima de este umbral se considera que las secuencias son de cuerdas vocales patológicas y por debajo, a cuerdas vocales sanas (todo ello, en función de los parámetros estudiados).

La Figura 100 muestra los resultados de las 45 secuencias en función de los resultados obtenidos de las medidas morfológicas. Las secuencias están ordenadas en el eje X, y el área en el eje Y. Cuando ambas cuerdas vocales son

patológicas (nódulos) se ha representado únicamente el área mayor de las dos (que se ha considerado más significativa).

Siguiendo con la Figura 100, se puede observar claramente que las secuencias de cuerdas vocales sanas (1-10) están por debajo del umbral, al igual que las parálisis (39-45), que atendiendo únicamente a la morfología presentan cuerdas vocales normales o sanas. Por lo tanto, tenemos concentradas encima del umbral a todas las secuencias patológicas. Aquellas secuencias en las que se han encontrado patológicas las dos cuerdas se representan en rojo (11-20), mientras que aquellas que solo afectan a una de las dos cuerdas se muestran en verde (21-38).

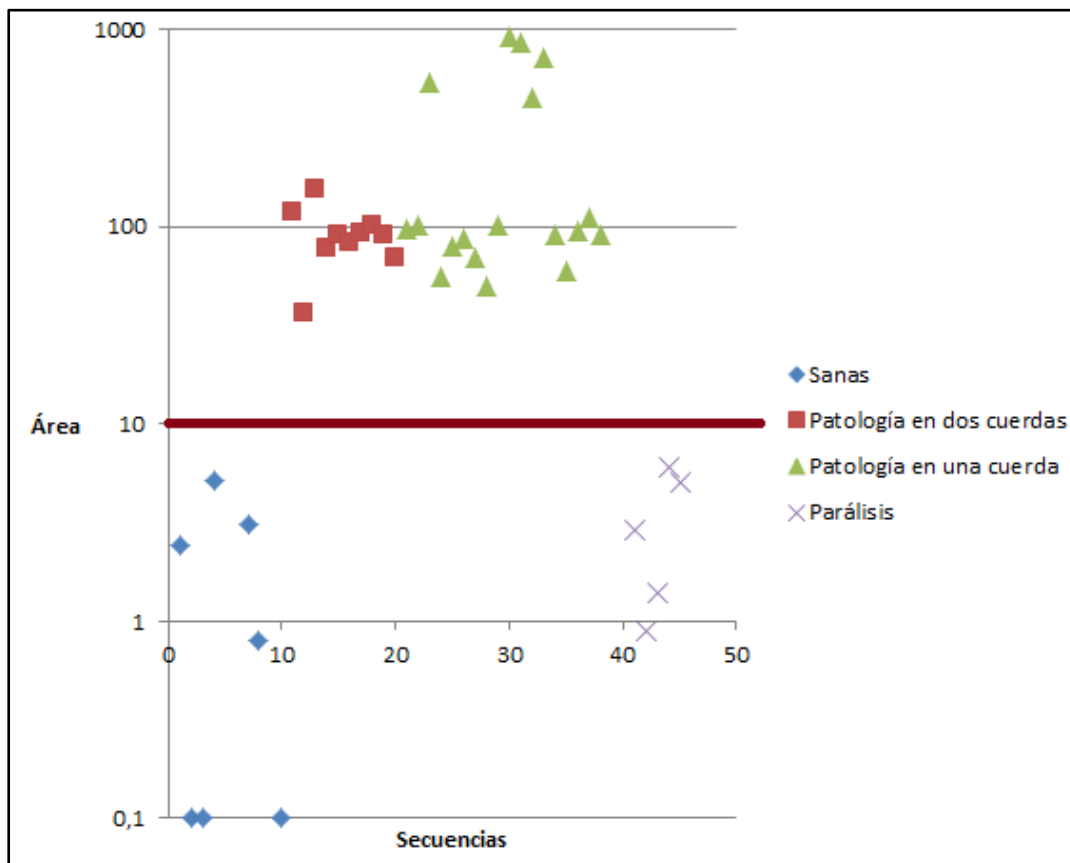


Figura 100. Resultados del diagnóstico en función de la morfología

Dado que no hemos podido ver la diferencia entre sanas y parálisis en la Figura 100, la Figura 101 nos permite discernir entre ellas. En este caso, también tenemos en el eje X los números de secuencia, y ahora el eje Y proporciona

información sobre la desviación de la medida de los ángulos izquierdo y derecho. Se puede observar que todas las secuencias pertenecientes a cuerdas vocales sanas (1-10) están por debajo del umbral, y las secuencias de parálisis por encima.

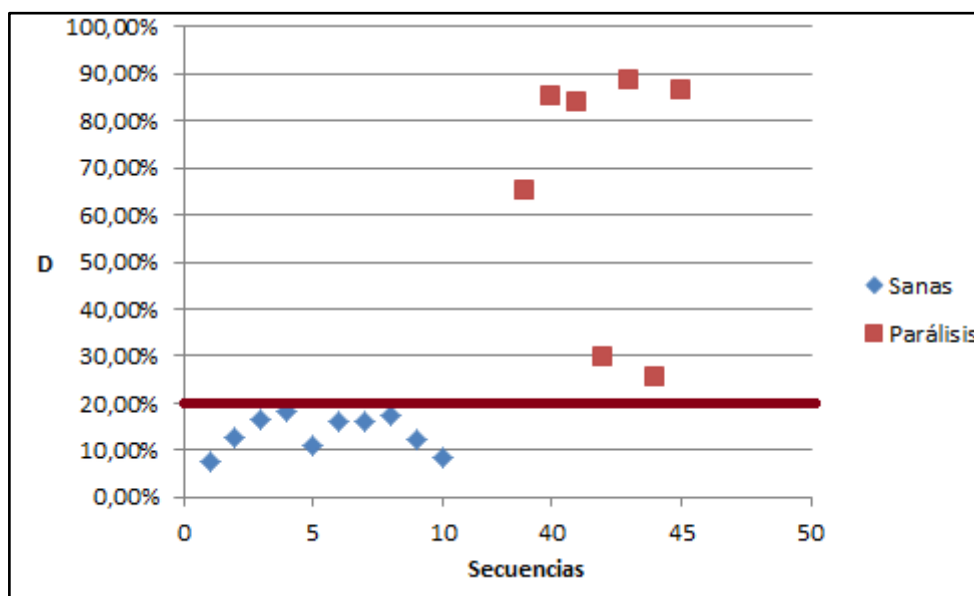


Figura 101. Resultados del diagnóstico en función del movimiento

Finalmente, atendiendo al estudio de las secuencias de imágenes y no a los frames aislados, se ha obtenido un 100% de éxito en el diagnóstico de las secuencias (coincidiendo con el emitido por el otorrinolaringólogo) utilizando la parametrización propuesta.

5.4 INCIDENCIAS ENCONTRADAS

Durante el desarrollo de la presente tesis no ha habido incidencias significativas que resaltar, pero sí muchos pequeños inconvenientes y contratiempos que se han tenido que superar.

A continuación, se describen las incidencias en los tres momentos críticos del desarrollo de esta tesis: la obtención de la base de datos, el diseño del algoritmo y la evaluación de los resultados.

5.4.1 Incidencias referentes a la obtención de la base de datos

Con respecto a la base de datos se dos fueron las metas planteadas: obtener imágenes de calidad y provenientes de diferentes especialistas, y tener un número representativo del mayor número de patologías. Bien, pues ha habido incidencias en ambas.

En el comienzo, la mala calidad de las imágenes fue uno de los caballos de Troya más duros a superar. Concretamente, hubo que pasar por un periodo en el que el especialista probó instrumentación de diferentes fabricantes, y por tanto, las capturas estaban desenfocadas, los frames no contenían la información requerida sobre el espacio glótico y las cuerdas vocales, etc.

El otro gran problema encontrado ha sido la cantidad de imágenes disponibles para analizar, por esta razón se recurrió a una base de datos comercial. Ha sido difícil poder llegar a un número de imágenes representativo para sacar conclusiones sobre los resultados obtenidos.

5.4.2 Incidencias en el diseño y desarrollo de los algoritmos

El diseño y desarrollo de los algoritmos es realmente el corazón de la tesis y donde se realizan las aportaciones. Se ha alargado en el tiempo debido a múltiples problemas que se han ido solucionando.

Solucionado el tema de la base de datos, y con la experiencia adquirida tanto para la captura de imágenes como en la aplicación de las técnicas de procesado digital de imagen, pudimos empezar a aplicar distintos algoritmos superando los primeros problemas.

Uno de los principales problemas a resolver ha sido la unificación de las características propias de este tipo de imágenes. Debido a esto, también es imposible realizar medidas en centímetros, y por esta razón se han efectuado en píxeles.

Con los problemas encontrados para componer una buena base de datos, han sido múltiples las técnicas testeadas para cumplir los objetivos marcados. Entre ellas, técnicas basadas en contornos activos o snakes, con las que se obtenían buenos resultados de segmentación, pero que requerían de inicialización por el usuario o por el sistema, o la aplicación de técnicas de análisis de movimiento para realizar la segmentación. Esto demuestra que desde el principio, la segmentación ha sido la principal preocupación, y objeto de muchas pruebas y desarrollos porque no se conseguían los resultados requeridos.

Las incidencias de la etapa de clasificación se centraron en las decisiones que se tuvieron que tomar en la elección del Training Set, con respecto al tamaño y al contenido en sí. Es decir, cuáles eran las imágenes más representativas de cada una de las patologías de la base de datos. Para ello, se contó con las inestimables aportaciones del Dr. Agustín Pérez Izquierdo, que resolvió nuestras dudas en múltiples ocasiones. Con respecto al tamaño se hizo una revisión de la literatura y se fueron sacando conclusiones empíricamente,

5.4.3 Incidencias durante la evaluación del sistema

Durante la experimentación, ha habido varias claves que han hecho que esta etapa fuese bastante dura. Principalmente han sido el tiempo de proceso y la interpretación de los resultados, para dar respuesta a la parametrización de las cuerdas vocales.

Para la obtención de los resultados, todos los frames de todas las secuencias han sido introducidos en el sistema, es decir, los algoritmos se han aplicado a todos ellos. Por ello, y debido al elevado número de imágenes analizadas, la incidencia principal de esta etapa ha sido el tiempo invertido en la ejecución, llegando a ser de varios minutos por secuencia.

Todo este proceso ha sido más costoso porque no se ha contado con un cluster para computación paralela, que hubiese aliviado este problema.

La necesidad de exhibir los resultados de una forma gráfica, unificando y normalizando todas las medidas en un rango común también ha generado indecisiones sobre si era la manera adecuada. Pero finalmente, la elección tomada permite al médico especialista interpretar los resultados de una forma fácil y sencilla.

“My vocal cords lived their own life, wild and free.”

Robert A. Heinlein

6.CONCLUSIONES

En este capítulo se va a comprobar el cumplimiento de los objetivos indicados en la introducción de la tesis, y se van a plantear las líneas que se abren como continuación de este trabajo de investigación cuyas aportaciones se han demostrado en el capítulo de “Experimentación”, y que se resumen ahora como conclusiones de la tesis. Se indica el impacto científico que han tenido los algoritmos diseñados en esta tesis en forma de: publicaciones científicas en diferentes foros internacionales que avalan los resultados de esta investigación y registros de propiedad intelectual.

6.1 CUMPLIMIENTO DE OBJETIVOS

Como resumen de todo lo expuesto en apartados anteriores, se pueden obtener una serie de conclusiones de las cuales se destaca una muy importante: es posible proporcionar una ayuda al diagnóstico de patologías vocales a través del procesado digital de imagen de imágenes de cuerdas vocales, y la extracción de medidas objetivas. Además, se ha demostrado que no es necesaria la interacción con el usuario (en este caso, el médico especialista en

otorrinolaringología) durante el proceso de análisis y procesado de las imágenes de cuerdas vocales.

Después del estudio previo, se determinaron qué tipo de imágenes íbamos a usar, los parámetros objetivos que nos interesaba medir y las carencias actuales existentes en el mercado. Sabiendo esto, enfocamos el trabajo en la determinación de las líneas de acción para la objetivación de parámetros que nos ayuden al diagnóstico, y demostramos el gran potencial de este trabajo.

En paralelo, y para el desempeño de la tesis, se fue creando la *base de datos* de secuencias de imágenes de cuerdas vocales, que ha sido alimentada con nuevos registros durante varios años, para cumplir el primer objetivo planteado (O1).

Esto nos permitió alcanzar los siguientes objetivos planteados (O2 y O3) y abordar el *diseño, desarrollo e implementación de un algoritmo* para el procesado de imágenes estroboscópicas de cuerdas vocales (O4 y O5).

Aunque el procesado digital de imagen no es una tarea fácil, y menos si se cuenta con imágenes de las características de las que centran el desarrollo de esta tesis, los *resultados* obtenidos y que se han mostrado en el capítulo 4 “Resultados”, son muy *satisfactorios* en cuanto a la *veracidad de los diagnósticos*. La verificación y el contraste han sido siempre realizados de la mano de un especialista en otorrinolaringología que o bien nos proporcionaba las imágenes previamente diagnosticadas, o validaba el resultado dado por el algoritmo.

Cumpliendo O5, se ha desarrollado un algoritmo dividido principalmente en dos partes: una para realizar la segmentación del espacio glótico de las cuerdas vocales, y otro para discernir entre cuerdas vocales sanas, con patologías morfológicas y otras, como puede ser las relacionadas con el movimiento.

En relación al tema de la segmentación, en el capítulo 4 se puede observar cómo el 95,07% de las imágenes que contiene la Base de Datos han sido segmentadas correctamente, después de aplicar el algoritmo propuesto. Este ratio ha sido calculado sin tener en cuenta los pequeños errores de segmentación como el

que se muestra en la Figura 83 en el frame 69, y que no afectan a las etapas posteriores del algoritmo, ni al diagnóstico (teniendo en cuenta que siempre la ubicación de estas pequeñas desviaciones en la segmentación es en la parte superior del espacio glótico). Se han presentado los resultados de la segmentación por secuencias agrupadas por patologías, y en todas se supera el ratio del 95%, excepto en las secuencias de cuerdas vocales con edema, en las que este ratio por debajo del 90%.

Para diferenciar y optimizar los procesados en función de la patología detectada, se ha desarrollado/aplicado el algoritmo estadístico PCA, para realizar una primera clasificación entre cuerdas vocales con patología morfológica y con ausencia de ella. Se ha demostrado que la técnica de Principal Component Analysis da buenos resultados utilizando un Training Set de imágenes significativas de cada patología. En este bloque se han obtenido resultados del 92,1% con un Training Set de 100 imágenes (200 en realidad, ya que se han separado cuerda derecha e izquierda). No se incluyen las pruebas con Training Sets mayores, ya que el tiempo de procesamiento aumentaba exponencialmente, llegando a ser de varios minutos.

Las medidas objetivas realizadas permiten al especialista cuantificar el tamaño de la patología que está describiendo/visualizando, pudiendo de esta forma, proporcionar más información al paciente, para afinar el tratamiento e incluso medir la evolución en procesos post-operatorios o de rehabilitación vocal. El sistema, sin prescindir del médico, puede sugerir un diagnóstico apoyado en los resultados obtenidos.

De estos resultados podemos concluir que el algoritmo diseñado funciona adecuadamente, y lo más importante, evitando al máximo la interacción con el usuario. Y es justo en la interacción con el usuario donde se diferencia de los pocos software comerciales que hay en el mercado sobre la temática que nos ocupa.

Actualmente, el software comercial de referencia en imágenes low speed de otorrinolaringología es KIPS de Kay Elemetrics [KAY]. La principal diferencia entre los algoritmos presentados y la aplicación comercial es la automatización del proceso, ya que en este caso los autores desean evitar 100% la interacción con el usuario durante el procesamiento y las medidas.

Una de las cosas inherentes de las medidas de las imágenes de cuerdas vocales es que son siempre en píxeles (y/o relativas), no teniendo la información necesaria para proporcionar medidas reales en cm.

6.2 IMPACTO CIENTÍFICO

El área de trabajo abordado en esta tesis ha dado lugar a publicaciones científicas y registros de propiedad intelectual y algunos proyectos de investigación.

6.2.1 Publicaciones científicas

A continuación, en la Tabla 17 se indican cronológicamente las ponencias presentadas en los siguientes congresos internacionales.

Congreso	Detalle
Biosignal 2006	Autores: García, B., Méndez, A., Vicente, J., & Ruiz, I. Título: "Diagnosis of vocal folds nodules and polyps using derivative techniques" [Garcia+06].
ISSPA 2007	Autores: Méndez, A., García, B., Vicente, J., Ruiz, I., & Sánchez, K. Título: "Objective Model of Vocal Folds, Based on Glottal Closure, Opening Angles and Mophologic Criteria" [Méndez+07].
ISPRA 2008	Autores: Méndez, A., García, B., Ruiz, I., & Iturricha, I. Título: "Vocal Folds Objective Evaluation Measuring Vascularization Level and Texture" [Méndez+08a]
Biomed 2008	Autores: Méndez, A., García, B., Ruiz, I., & Iturricha, I. Título: "Vocal folds Vascularization Level Analysis using Hough Transform"[Mendez+08b]
ISIVC 2008	Autores: Méndez, A., Ibn-Elhaj, E., García, B., Iturricha, I., & Ismaili

	Alaoui, E. Título: “Glottal Area Segmentation using Wiener Motion Estimation and Neighborhood Algorithm” [Méndez+08c]. Autores: Méndez, A., Osma, V., García, B., Sáenz-Lechón, N., Ruiz, I., & Fraile, R. Título: “Diagnosis of Vocal Folds Morphological Pathologies by means of Advanced Image Processing Methods”. [Méndez+08d]
ISSPIT 2008	Autores: Méndez, A., García, B., & Iturricha, I. Título: “Glottal Area Segmentation without Initialization using Gabor Filters” [Mendez+08e].
EMBC 2009	Autores: Méndez, A., Ismaili Alaoui, E., García, B., Ibn-Elhaj, E., & Ruiz, I. Título: “Glottal Space Segmentation from Motion Estimation and Gabor Filtering” [Mendez+09a].
ICIP 2009	Autores: Ismaili Alaoui, E., Méndez, A., Ibn-Elhaj, E., & García, B. Título: “Keyframes and Analysis in Vocal Folds Recordings using Hierarchical Motion Techniques and Texture Information” [Ismaili+09].
ISSPA 2010	Autores: Méndez, A., El-Zehiry, N., García, B., & Elmagraby, A. Título: “Pathological Vocal Folds Diagnosis using modified Active Contour Models” [Méndez+10a].
ISSPIT 2010	Autores: Méndez, A., & García, B. Título: “Vocal Folds Paralysis Study using a pre-processing stage of Gabor Filtering and Chan Vese Segmentation”. [Mendez+10b].
CE-ORL-HNS 2011	Autores: Pérez, A., García, B., Méndez, A., Santaolalla, F. Título: “Creación de una herramienta informática para el estudio clínico de la voz” [Perez+11].
ISSPIT 2011	Autores: Méndez, A., Lopetegui, E., García, B. Título: “Vocal Folds Paralysis Detection using an Adapted Block Matching Algorithm” [Mendez+11a].
Biomed 2012	Autores: Méndez, A., Lopetegui, E., García, B. Título: “FLDA and PCA classification supported by an adapted block matching algorithm to diagnose vocal folds paralysis” [Mendez+12a].
ISCCSP 2012	Autores: Méndez, A., Lopetegui, E., García, B. Título: “Vocal Folds Paralysis Clasiffication using FLDA and PCA algorithms suported by an Adapted Block Matching Algorithm” [Mendez+12b].

Otolaryngology 2012	Autores: Méndez, A., García, B., Pérez, A. Título: "Image Processing Techniques to Support Vocal Pathologies Diagnosis" [Mendez+12c]
------------------------	---

Tabla 17. Comunicaciones en Congresos

Además de publicaciones en congresos, en la última fase de la tesis se ha hecho especial hincapié en difundir los resultados en revistas:

Revista	Detalle
Majlesi Journal of Electrical Engineering	Autores: Méndez, A., El-Zehyri, N., García, B., & Elamghraby, A. Título: "Pathological Vocal Folds Features extraction using a modified Active Contour Segmentation" [Mendez+10c].
IET Computer Vision	Autores: Méndez, A., García B. Título: "Automated Diagnosis of Vocal Fold Paralysis using a combination of Active Contour and Gabor Filtering Segmentation and Angle Measurement". (Pendiente de aceptación).
SpringerOnline	Autores: Méndez, A., García B., Pérez, A. Título: "Computer Aided Tool for Diagnosis of ENT Pathologies using Digital Signal Processing of Speech and Stroboscopic Images". (Pendiente de aceptación).
Biomedical Engineering Online	Autores: Méndez, A., García B. Título: "Vocaligram: Morphological and Movement Parameterization of Vocal Folds using Image Processing Techniques". (Pendiente de aceptación).

Tabla 18. Publicaciones en Revistas

Por último, con la temática de esta tesis se ha participado en un capítulo de libro:

Libro	Detalle
Medical Imaging in Clinical Practice	Autores: Méndez, A., García, B. Título: "Vocal folds Stroboscopic image processing for Otolaryngology". (Pendiente de publicación). ISBN: 980-953-307-710-4

Tabla 19. Publicación en libro

6.2.2 Propiedad intelectual

Algunos resultados obtenidos durante esta tesis han sido objeto de registro de propiedad intelectual, y se describen a continuación:

➤ Registro de Software

Autores: María Begoña García Zapirain, Amaia Méndez Zorrilla, Ibon Ruiz Oleagordia, Agustín Pérez Izquierdo.

Nº de Asiento Registral: 00/2009/4936.

Título del registro de propiedad intelectual: AnalisisVOX. Software de ayuda al diagnóstico de patologías de cuerdas vocales según modelo objetivo.

Clase de obra: Programa de ordenador.

➤ Registro de Base de Datos

Autores: María Begoña García Zapirain, Amaia Méndez Zorrilla, Agustín Pérez Izquierdo.

Nº de Asiento Registral: Pendiente

Título del registro de propiedad intelectual: Grabaciones de imágenes estroboscópicas de cuerdas vocales

6.2.3 Proyectos de investigación y colaboraciones externas

En los comienzos, el primer proyecto de investigación básica con subvención externa, corresponde a la convocatoria SAIOTEK 2005-2006, de la Consejería de Educación del Gobierno Vasco.

El proyecto fue finalizado de forma satisfactoria, aunque sus resultados y soluciones no coinciden con el diseño presentado en esta tesis, pero dicho proyecto supuso el primer contacto que la autora de esta tesis tuvo con la temática y los primeros resultados presentados a la comunidad científica.

Posteriormente en 2008, y para incentivar las colaboraciones, se desarrolló el proyecto de análisis de procesamiento de imágenes de cuerdas vocales, financiado por el programa PCI del Ministerio de Asuntos Exteriores y Cooperación.

Desde 2007 a 2009, el proyecto Darevoz (Diagnóstico Asistido Remoto de la voz a partir de medidas biométricas y otras Parametrizaciones) financiado por el Ministerio de Educación y Ciencia, permitió a la doctoranda colaborar con otros equipos a nivel nacional (Universidad Politécnica de Madrid), y profundizar en las técnicas de procesamiento digital de imagen en la temática objeto de esta tesis.

El trabajo desarrollado a lo largo de esta tesis ha dado lugar a publicaciones y colaboraciones con otros grupos de trabajo de reconocido prestigio como el liderado por el profesor Elhassane Ibn-Elhaj (INPT, Rabat) o con el profesor Adel Elmagraby (University of Louisville, EEUU). Los resultados han estado presentes en una serie de foros científicos relacionados con el tema, donde ha recibido un reconocimiento por parte de la comunidad técnica que trabaja en áreas similares.

6.3 FUTURAS LÍNEAS DE INVESTIGACIÓN

La investigación abierta en esta tesis supone un avance en la aplicación de las nuevas tecnológicas al ámbito de la medicina y más concretamente a la otorrinolaringología y el diagnóstico asistido por ordenador (CAD). El trabajo presentado aborda el cálculo de algunos parámetros objetivos aplicados al diagnóstico de patologías benignas, pero basándonos en estos resultados, podemos extraer nuevas vías de investigación.

En relación a las patologías de la voz que pueden ser analizadas y diagnosticadas mediante la captura de imágenes digitales, queda fuera del alcance de esta tesis la *parametrización del cáncer*. Algunos de los parámetros definidos en esta tesis pueden ser aplicados, pero un diagnóstico eficiente del cáncer pasaría por un análisis más exhaustivo, no solo de la morfología, sino de características de color y textura.

Una patología muy interesante, que no ha sido objeto de estudio en esta tesis, es la *disfonía espasmódica*. Es un trastorno de baja prevalencia que suele quedar

relegado, a favor de patologías con mayor incidencia. La disfonía espasmódica es un trastorno neuromotor, que es el causante de los espasmos. Tiene que ser abordado desde la atención primaria, la otorrinolaringología, la neurología y la psicología, y por tener este carácter multidisciplinar resulta atractivo para futuros estudios. Los análisis en el ámbito de la ingeniería biomédica deberían abordar el procesamiento de señales neuronales (capturadas mediante un electroencefalograma), de imágenes y de voz, lo que podría suponer un reto para una futura tesis.

Si se desea continuar con patologías relacionadas con el movimiento de las cuerdas vocales (como las parálisis unilaterales o bilaterales), la aplicación de algoritmos de tracking (por ejemplo, *Filtros de Kalman*) y la caracterización matemática del espacio glótico puede aportar un gran avance en la predicción de movimientos y patologías de las cuerdas vocales.

Con todos los parámetros obtenidos en esta tesis, se podría cambiar la etapa de clasificación, por un sistema basado en redes neuronales o algoritmos de lógica difusa para la interpretación de la imagen en sí [Udelhoven+00] [Ding+01].

Por otro lado, la caracterización de la voz viene dada por el cálculo de los parámetros acústicos: pitch, jitter, shimmer y relación armónico ruido, principalmente, que si bien no es aceptado como método diagnóstico, sí ofrece resultados orientativos. La combinación de los resultados de estos parámetros junto con los resultados obtenidos mediante el análisis de la imagen puede dar lugar a interesantes conclusiones, sobre todo para la objetivación de las enfermedades de la voz como baja laboral.

Si nos orientamos la continuación de este trabajo hacia el desarrollo de proyectos de innovación e investigación aplicada, es clara la línea de desarrollo de nuevos software de ayuda al diagnóstico que contengan los algoritmos presentados en esta tesis y otros que puedan aportar información. En este tipo de aplicaciones es muy importante combinar la anamnesis que realiza el

médico, la información subjetiva que aporta el paciente, y todos los parámetros objetivos que puedan aportar información para el diagnóstico y la medida de la eficacia de los tratamientos aplicados por el médico, en este caso, por el otorrinolaringólogo. Si a estos softwares se les dota de la posibilidad de interacción remota entre profesionales, estaríamos ante los inconvenientes que puede ocasionar la transmisión de secuencias de imágenes (habitualmente muy pesadas). Para resolverlo podríamos pensar en usar el mismo esquema que *JPEG200* [Skodras+01], y usar las *Transformadas Wavelet* [Lamard+06].

Otro reto pendiente de alcanzar es la inclusión de información en la propia imagen. Este tema es muy interesante para la inclusión de metadatos y se podrían aplicar *técnicas de Watermarking* [Giakoumaki+03] [Coatrieux+00] que no afectasen a la calidad del post-procesado de la imagen.

En definitiva, se puede decir que son varias las líneas que quedan abiertas, tanto para la investigación sobre la aplicación de algoritmos de procesamiento digital de imagen para el diagnóstico de patologías vocales, como para el desarrollo de proyectos de aplicación de dicha investigación.

7.BIBLIOGRAFÍA

- [Allin+04] Allin, S., Galeotti, J., Stetten, G. Dailey, S.H. (2004). Enhanced snake based segmentation of vocal folds. *In proc. ISBI 2004*. 812- 815 Vol 1.
- [Arbos+05] Arbós, D., & Pérez, R. (2005). Dos millones de españoles sufren trastornos de la voz. *Ondasalud*. Recuperado el 1 de Septiembre de 2011, <http://www.ondasalud.com>
- [Armstrong+10] Armstrong, W., Vokes, D., & Maisel, R. (2010). Malignant tumors of the larynx. *Cummings Otolaryngology Head and Neck Surgery* (5th ed.). Elsevier.
- [Baken+00] Baken, P., & Orlikoff, R. (2000). *Clinical Measurement of Speech and Voice* (2nd ed.). San Diego: Singular Publishing Group.
- [Barjatya04] Barjatya A. (2004). Block Matching Algorithms for Motion Estimation. *In DIP 6620 Spring*. Final Project Paper.
- [Boersma+11] Boersma, P., & Weenink, D. (s.f.). *PRAAT*. Recuperado el 1 de Septiembre de 2011, de <http://www.fon.hum.uva.nl/praat/>

- [Boudreau+01] Boudreau, M., Gefen, D., & Straub, D. (2001). Validation in IS Research: A State of the Art Assessment. *MIS Quarterly*, 25 (1), 1-16.
- [Bradski08] Bradski, G., & Kaehler, A. (2008). *Learning OpenCV: Computer Vision with the OpenCV Library*. O'Reilly Media Inc.
- [Canny86] Canny, J. (1986). A Computational Approach to Edge Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8 (6).
- [Casado +10] Casado Morente, J. C., Casanova Castillo, R., Hernández Rubiño, A., Gómez Velasco, C., Fernández Ruiz, E., Piédrola Maroto, D., y otros. (2010). Procedimientos para la integración de una Unidad de Voz en el funcionamiento de un Servicio/Área de ORL y sus resultados. *Acta Otorrinolaringológica Española*.
- [Cavalcanti+10] Cavalcanti N., Silva, S., Bresolin, A., Bezerra, H., Guerreiro, A. (2010). Comparative analysis between wavelets for the identification of pathological voices. *Proceedings of the 15th Iberoamerican congress conference on Progress in pattern recognition, image analysis, computer vision, and applications*. Sao Paulo. Brasil.
- [Chan+02] Chan, M.H.; Yu, Y.B.; Constantinides, A.G. (2002). Variable size block matching motion compensation with applications to video coding. *IEE Proceedings in Communications, Speech and Vision*, 137 (4).
- [Chang+07] Chang, J.S., Kim, E.Y., Hyun Park S. (2007). Lip Contour Extraction Using Level Set Curve Evolution with Shape Constraint. *Human-Computer Interaction, Part III, LNCS* 4552, 583-588.
- [Clausi+00] Clausi D.A, Jernigan M. E. Designing Gabor Filters for

- optimal texture separability. *Pattern Recognition* 33. 1835-1849. 2000.
- [Coatrieux+00] Coatrieux, G. Maitre, H., Sankur, B., Rolland, Y., Collorec, R. (2000). Relevance of watermarking in medical imaging. *Proceedings of International Conference of IEEE Engineering in Medicine and Biology society. EMBC 2000*.
- [Cuevas+10] Cuevas, E., Zaldívar, D., & Pérez, M. (2010). *Procesado Digital de Imágenes con Matlab y Simulink*. Madrid: Ra-Ma.
- [Dhawan10] Dhawan, A. P. (2010). *Medical Image Analysis*. Piscataway, NJ 08854: IEEE Press Editorial Board.
- [Di+08] Di, M., Joo, E.M., Beng, L.H. (2008) A comprehensive study of Kalman filter and extended Kalman filter for target tracking in Wireless Sensor Networks. *Proceedings of International Conference on System, Man and Cibernetics*. 2792 – 2797.
- [Ding+01] Ding, C., and Dubchak, I. (2001). Multi-class protein fold recognition using support vector machines and neural networks. *Bioinformatics* 17 (4).
- [Du+07] Du ,L., Jia, Z., Xue L. (2007). Human Face Recognition Based on Principal Component Analysis and Particle Swarm Optimization -BP Neural Network. *Third International Conference on Natural Computation (ICNC 2007)*. Haikou, China.
- [ELS] European Laryngological Society. *ELS*. Recuperado el 1 de Octubre de 2011, de <http://www.elsoc.org/>
- [Erturk02] Ertürk, S.(2002). Real-Time Digital Image Stabilization Using Kalman Filters. *Real-Time Imaging*, 8 (4), 317–328.
- [Felsa+10] Felsa, S., & Bilston, L. (2010). Dynamic modelling of the oral, pharyngeal and laryngeal complex for biomedical applications. *Computer Methods in Biomechanics and Biomedical Engineering* , 13 (4), 441–442.

- [Fernández+06] Fernández González, S., Vázquez de la Iglesia, F., Marqués Girbau, M., & García-Tapia Urrutia, R. (2006). Manuel P. García. *Revista Medica Univ. Navarra* , 50 (3), 14-18.
- [French+96] French, W., & Bell, C. (1996). *Organisational Development: Behavioral Science*. London: Prentice-Hall.
- [Gañet+07] Gañet Benavente, R., Serrano Estrada, C., & Gallego Pulgarin, M. I. (2007). Patología vocal en trabajadores docentes: influencia de factores laborales y extralaborales. *Arch. Prev. Labor.* , 10 (1), 12-17.
- [García47] García, M. (1847). *Traité complet du chant*. Paris.
- [García+06] García, B., Méndez, A., Vicente, J., & Ruiz, I. (2006). Diagnosis of Vocal Folds Nodules and Polyps using Derivative Techniques. *Proceedings of 18th Biental international Eurasip Conference, Biosignal 2006*. Brno.
- [García+09] García, B., Ruiz, I., Méndez, A., & Mendezona, M. (2009). Objective characterization of oesophageal voice supporting medical diagnosis, rehabilitation and monitoring. *Computers in Biology and Medicine* , 39, 97-105.
- [Giakoumaki+03] Giakoumaki, A., Pavlopoulos, S., Koutouris, D. (2003). A medical image watermarking scheme based on wavelet transform. *Proceedings of International Conference of IEEE Engineering in Medicine and Biology society*. EMBC 2003.
- [Goffredo+06] Goffredo M, Schmid M, Conforto S, D'Alessio T. (2006). A markerless sub-pixel motion estimation technique to reconstruct kinematics and estimate the centre of mass in posturography". *Medical Engineering & Physics*, 28 (7), 719-726.

- [González+02] González, J., Cervera, T., & Miralles, J. L. (2002). Análisis Acústico de la voz: Fiabilidad de un conjunto de parámetros multidimensionales. *Acta Otorrinolaringol Esp*, 53, 256-268.
- [González+09] González, R., Woods, R., & Eddins, S. (2009). *Digital Image Processing Using MATLAB*. Gatesmark Publishing.
- [Gray18] Gray, Henry. (1918). *Anatomy of the Human Body*. Philadelphia: Lea & Febiger.
- [Henry1918] Henry, G. (1918). *Anatomy of the Human Body*. Philadelphia: Lea & Febiger.
- [Hotelling33] Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, 24, 417-441.
- [Ibermutuamur] Mutua de accidentes de trabajo y enfermedades profesionales de la Seguridad Social. (s.f.). *Ibermutuatur*. Recuperado el 25 de Agosto de 2011, de <http://ibermutuamur.es/salud>
- [ImageJ] ImageJ. (s.f.). Recuperado el 16 de Julio de 2012, de <http://rsbweb.nih.gov/ij/docs/index.html>
- [ISC10] ISC. (2010). *Guía de Valoración de incapacidad Laboral para médicos de Atención Primaria*. Madrid: Escuela Nacional de Medicina del Trabajo. Instituto de Salud Carlos III.
- [Ismali+09] Ismaili Alaoui, E., Méndez, A., Ibn-Elhaj, E., & García, B. (2009). Keyframes Detection and Analysis in vocal Folds Recordings using Hierarchical motion Techniques and Texture Information. *Proceedings of International conference of Image Processing. ICIP 2009*. El Cairo. Egipto.

- [Jähne+99] Jähne, B., Scharr, H., & Körkel, S. (1999). *Principles of filter design*. In *Handbook of Computer Vision and Applications*. Academic Press.
- [Johns03] Johns, Michael M. (2003). Update on the etiology, diagnosis, and treatment of vocal fold nodules, polyps, and cysts. *Current Opinion in Otolaryngology & Head & Neck Surgery*, 11 (6), 456-461.
- [Karhunen47] Karhunen K. (1947). Über lineare methoden in der Wahrscheinlichkeitsrechnung. *Annales Academiae Scientiarum Fennicae, Seried A1: Mathematica-Physica*, 37, 3-79.
- [Kass+88] Kass, M., Witkin, A., Terzopoulos, D. (1988) Snakes: Active contour models. *International Journal of Computer Vision*. Vol. 1, Num. 4.
- [KAY] Kay Elemetrics. *Kay Pentax*. Recuperado el 1 de Septiembre de 2011, de <http://www.kayelemetrics.com/>
- [Kim+98] Kim W, Choi J.-S, and Kim J.-T. "Adaptive motion estimation based on spatio-temporal correlation". *Signal Process.: Image Commun.*, vol. 13, pp. 161-170, 1998.
- [Lamard+06] Lamard, M., Daccache, W., Cazuguel, G., Roux, C., Cochener, B. (2006). Use of a JPEG-2000 Wavelet Compression Scheme for Content-Based Ophtalmologic Retinal Images Retrieval. *Proceedings of International Conference of IEEE Engineering in Medicine and Biology society. EMBC 2006*.
- [LeBorgne06] LeBorgne, W. (2006). *Laryngeal Videostroboscopic Images*. Plural Publishing.

- [Lim+11] Lim, H., Park, HW. (2011). A symmetric motion estimation method for motion-compensated frame interpolation. *IEEE Trans Image Process*, 20(12), 3653-8.
- [Loève63] Loève, M. (1963). *Probability Theory*. Van Nostrand.
- [Lonsdale+10] Lonsdale, M., & Beyer, T. (2010). Dual-modality PET/CT instrumentation—Today and tomorrow. *European Journal of Radiology*, 73 (3), 452-460.
- [Maini+09] Maini, R., & Aggarwal, H. (2009). Study and Comparison of Various Image Edge Detection Techniques. *International Journal of Image Processing (IJIP)*, 3 (1).
- [Mathworks] Mathworks. (s.f.). Recuperado el Septiembre de 2011, de <http://www.mathworks.es/>
- [Manfredi+07] Manfredi, C., Bocchi, L., Cantarella, G., Peretti, G., [Guidi, G.](#), [Mezzatesta, C.](#) Objective parameters from videokymographic images: a user-friendly interface. *Proceedings of INTERSPEECH 2007*, 1222-1225
- [Manfredi+12] Manfredi, C., Bocchi, L., Cantarella, G., Peretti, G. (2012) Videokymographic image processing: Objective parameters and user-friendly interface. *Biomedical Signal Processing and Control*, 7, 192– 201
- [Marr+80] Marr, D., & Hildreth, E. (1980). Theory of Edge Detection. *Proceedings of the Royal Society of London. Series B, Biological Sciences*, (págs. 187-217).
- [Mazinan+12] Mazinan, A.H., Amir-Latifi, A. (2012). Applying mean shift, motion information and Kalman filtering approaches to object tracking. *ISA Trans*, 51(3), 485-97.

- [Méndez+07] Méndez, A., García, B., Vicente, J., Ruiz, I., & Sánchez, K. (2007). Objective Model of Vocal Folds based on Glottal Closure, Opening Angles and Morphologic Criteria. *Proceedings of International Symposium on Signal Processing and its Applications. ISSPA2007*. Sharjah, EAU.
- [Méndez+08a] Méndez, A., García, B., Ruiz, I., & Iturricha, I. (2008). Vocal folds objective evaluation measuring vascularization level and texture. *Proceedings of Proceedings of the 7th WSEAS International Conference on Signal Processing, Robotics and Automation. ISRA 2008*. Cambridge.
- [Méndez+08b] Méndez, A., García, B., Ruiz, I., & Iturricha, I. (2008). Vocal folds Vascularization Level Analysis using Hough Transform. *Proceedings of BIOMED 2008*. Innsbruck.
- [Méndez+08c] Méndez, A., Ibn-Elhaj, E., García, B., Iturricha, I., & Ismaili Alaoui, E. (2008). Glottal Area Segmentation using Wiener Motion estimation and Neighborhood. *Proceedings of Proceedings of International Symposium on Image/Video Communications over fixed and mobile networks. ISIVC 2008*. Bilbao.
- [Méndez+08d] Méndez, A., Osma, V., García, B., Sáenz-Lechón, N., Ruiz, I., & Fraile, R. (2008). Diagnosis of Vocal Folds Morphological Pathologies by means of Advanced Image Processing Methods. *Proceedings of International Symposium on Image/Video Communications over fixed and mobile networks. ISIVC2008*. Bilbao.
- [Méndez+08e] Méndez, A., García, B., & Iturricha, I. (2008). Glottal Area Segmentation without Initialization using Gabor Filters. *Proceedings of International Symposium on Signal Processing and Information Technology. ISSPIT 2008*. Sarajevo.
- [Méndez+09] Méndez, A., Ismaili Alaoui, E., García, B., Ibn-Elhaj, E., & Ruiz, I. (2009). Glottal Space Segmentation from Motion

Estimation and Gabor Filtering. *Proceedings of International Conference of IEEE Engineering in Medicine and Biology society. EMBC 2009*. Minneapolis, EEUU.

- [Méndez+10a] Méndez, A., El-Zehiry, N., García, B., & Elmagraby, A. (2010). Pathological Vocal Folds Diagnosis using modified Active Contour Models. *Proceedings of International Symposium on Signal Processing and its Applications. ISSPA 2010*. Kuala Lumpur, Malasia.

- [Méndez+10b] Méndez, A., & García, B. (2010). Vocal Folds Paralysis Study using a pre-processing stage of Gabor Filtering and Chan Vese Segmentation. *Proceedings of ISSPIT 2010*. El Cairo. Egipto.

- [Méndez+10c] Méndez, A., El-Zehyri, N., García, B., & Elamghraby, A. (2010). Pathological Vocal Folds Features extraction using a modified Active Contour Segmentation. *Majlesi Jorunal of Electrical Engineering* , 4 (4).

- [Mendez+11a] Méndez, A., Lopetegui, E., García, B.(2011). Vocal Folds Paralysis Detection using an Adapted Block Matching Algorithm. *Proceedings of International Symposium on Signal Processing and Information Technology. ISSPIT 2011*. Bilbao.

- [Mendez+12a] Méndez, A., Lopetegui, E., García, B.(2012). FLDA and PCA classification supported by an adapted block matching algorithm to diagnose vocal folds paralysis-*Proceedings of BIOMED 2012*. Insbruck.

- [Mendez+12b] Méndez, A., Lopetegui, E., García, B.(2012). Vocal Folds Paralysis Clasiffication using FLDA and PCA algorithms suported by an Adapted Block Matching Algorithm. *Proceedings of ISCCSP 2012*. Roma.

- [Mendez+12c] Méndez, A., García, B., Pérez, A. (2012). Image Processing

- Techniques to Support Vocal Pathologies Diagnosis. In *Proceedings of Otolaryngology 2012*. (Aceptado).
- [NIH] National Institutes of Health. (s.f.). Recuperado el 16 de julio de 2012, <http://www.nih.gov/>
- [Nepomuk60] Nepomuk Czermak, J. (1860). *Du laryngoscope*. Paris.
- [Noordzij+00] Noordzij, P., Woo, P. (2000). Glottal area waveform analysis of benign vocal fold lesions before and after surgery. *Ann Otol Rhinol Laryngol*, 109(5),441-6.
- [Oertel85] Oertel, M. (1985). Das laryngoskopische untersuchung. *Arch Laryngol Rhinol* , 3, 1-16.
- [OpenCV] Open Source Computer Vision. (s.f.). Recuperado el 16 de julio de 2012, <http://opencv.willowgarage.com/>
- [Pearson01] Pearson, K. (1901). On lines and planes of closest fit to systems of points in space. *Philosophical Magazine*, 2, 559-572.
- [Perez+11] Pérez, A., García, B., Méndez, A., Santaolalla, F. (2011). Creación de una herramienta informática para el estudio clínico de la voz. *Proceedings of CE-ORL-HNS*. Barcelona.
- [Poburka99] Poburka, B. (1999). A New Stroboscopy Rating Form. *Journal of Voice* , 13 (3), 403-413.
- [Popolo+08] Popolo, P., & Titze, I. R. (2008). Qualification of a Quantitative Laryngeal Imaging System Using Videostroboscopy and Videokymography. *Ann Otol Rhinol Laryngol* , 117, 404-412.

- [Pujol+01] Pujol A., Vitrià J, Lumbreras F, Villanueva J. J. (2001). Topological Principal Component Analysis for face encoding and recognition. *Pattern Recognition Letters*, 22, 769-776.
- [Quarteiro+10] Quarteiro, A. L., Dedivitis, R. A., & Pfuetzenreiter, E. G. (2010). Análise videolaringoestroboscópica de pacientes submetidos à radioterapia para tratamento de câncer glótico. *Braz. j. otorhinolaryngol.* , 76 (1).
- [Rees+08] Rees J.M, Regunath G, Whiteside S.P, Wadnerkar M.B, Cowell P.E, et al..(2008). Adaptation of wavelet transform analysis to the investigation of biological variations in speech signals. *Medical Engineering & Physics*, 30 (7), 865-871.
- [Roberts65] Roberts, L. (1965). Machine perception of three-dimensional solids. In J. T. Tippet (Ed.), *Optical and electro-optical information processing*.
- [SciPy] SciPhyton. (s.f.). recuperado el 16 de Julio de 2012, <http://www.scipy.org/>
- [Skodras+01] Skodras, A. Christopoulos, C., Ebrahimi, T. (2001). The JPEG 2000 still image compression standard. *IEEE Signal Processing Magazine*, 18 (5).
- [SEORL] Sociedad Española de Otorrinolaringología y Patología Cérvico-Facial.(s.f.). SEORL. Recuperado el 1 de Octubre de 2010, de <http://www.seorl.net>
- [Serrano+10] Serrano, R. L., Wilder, F., & Mansilla, M. (2010). Correlato Clínico Acústico en nódulos de cuerdas vocales. *Proc of. Congreso Iberoamericano de voz cantada y hablada*, (págs. 77-88). Buenos Aires.
- [Straub+04] Straub, D., Gefen, D., & Boudreau, M. (2004). *The ISWorld Quantitative, Positivist Research Methods Website*. Recuperado el 30 de Agosto de 2011, de <http://dstraub.cis.gsu.edu:88/quant/>

- [Stuckrad+75] Stuckrad, H., & Lakatos, I. (1975). A new magnifying laryngoscope (epipharyngoscope). *Laryngol Rhinol Otol* , 54 (4), 336-40.
- [Supic08] Supic, H. (2008). Improving Face Recognition Rate by Combining Eigenface Approach and Case-based Reasoning. *Proceedings of the World Congress on Engineering*. Vol I WCE 2008, July 2 - 4, 2008, London, U.K.
- [Svec+96] Švec, J., & Schutte, H. (1996). Videokymography: High-speed line scanning of. *Journal of Voice* , 10 (2), 201-205.
- [Titze08] Titze, I. (2008). The human instrument. *Sci. Am.* , 94-101.
- [Torre+01] Torre, De la F., Black, M. J. (2001). Robust Principal Component Analysis for Computer Vision. *Proceedings of Conf. on Computer Vision (ICCV'01)*.
- [UE03] UE. (2003). Recomendación de la Comisión de 19/09/2003 relativa a la lista europea de enfermedades profesionales. Bruselas.
- [Udelhoven+00] Udelhoven, T., Naumann, D., Schmitt, J. (2000). Applied Spectroscopy, 54 (10), 1471-1479.
- [Vicent+09] Vicent, O., & Folorunso, O. (2009). A Descriptive Algorithm for Sobel Image Edge Detection. *Proceedings of Informing Science & IT Education Conference (InSITE)*.
- [Wadsworth98] Wadsworth, Y. (1998). What is participatory Action Research? *Action Research*.
- [Zaidi+09] Zaidi, H., & Prasad, R. (2009). Advances in multimodality molecular imaging. *J Med Phys* , 34 (3), 122-128.

ANEXO A. GLOSARIO DE TÉRMINOS

AOI: Area of Interest

CAD: Computer Aided Diagnosis

CV: Cuerdas Vocales

ELS: European Laryngological Society

ES: Exhaustive Search

FFT: Fast Fourier Transform

GAW: Glottal Area Waveform

JPG: Joint Photographic Experts Group

HPV: Human papillomavirus. papilomatosis laríngea

LAV: Left Angle Value

MB: Macrobloque

MSE: Mean Square Error

MPEG: Moving Picture Experts Group

NIH: National Institute of Health

PCA: Principal Component Analysis

RAV: Right Angle Value

RGB: Red-Green- Blue

ROI: Region of Interest

SEORL: Sociedad española de Otorrinolaringología y Patología Cervico-Facial

SW: Search Window

VKG: Videoquimografía

