

UNIVERSIDAD DE LAS PALMAS DE GRAN CANARIA
DEPARTAMENTO DE INFORMÁTICA Y SISTEMAS

PROGRAMA DE DOCTORADO
SISTEMAS INTELIGENTES Y APLICACIONES NUMÉRICAS EN INGENIERÍA



TESIS DOCTORAL

**Metodología del Análisis Acústico
para la Evaluación Clínica de la Voz**

D. Jesús Bernardino Alonso Hernández

Las Palmas de Gran Canaria

Julio 2006

Anexo I

**Don Luis Mazorra Manrique de Lara, Secretario del
Departamento de Informática y Sistemas de la Universidad de
Las Palmas de Gran Canaria,**

CERTIFICA,

Que el Consejo del Departamento, en su sesión extraordinaria de fecha 28 de abril de 2006, tomó el acuerdo de dar el consentimiento para su tramitación a la tesis doctoral titulada "**Metodología del Análisis Acústico para la Evaluación Clínica de la Voz**" presentada por el doctorando Jesús Bernardino Alonso Hernández (DNI 52847148) y dirigida por los Doctores Don Miguel Ángel Ferrer Ballester y Don José De León y De Juan

Y para que así conste, y a efectos de lo previsto en el Artº 73.2 del Reglamento de Estudios de Doctorado de esta Universidad, firmo la presente en Las Palmas de Gran Canaria, a veintiocho de abril de dos mil seis.



UNIVERSIDAD DE LAS PALMAS DE GRAN CANARIA
DEPARTAMENTO DE INFORMÁTICA Y SISTEMAS

PROGRAMA DE DOCTORADO
SISTEMAS INTELIGENTES Y APLICACIONES NUMÉRICAS EN INGENIERÍA



TESIS DOCTORAL

**Metodología del Análisis Acústico
para la Evaluación Clínica de la Voz**

AUTOR: D. Jesús Bernardino Alonso Hernández

DIRECTOR: Dr. D. Miguel Ángel Ferrer Ballester
Dr. D. José de León y de Juan

El Director

El Codirector

El Doctorando,

Las Palmas de Gran Canaria a 2 de mayo de 2006

*A mis Padres,
a quienes les debo todo lo que soy.*

AGRADECIMIENTOS

Quisiera agradecer a los directores de esta Tesis Doctoral, al **Dr. Miguel Ángel Ferrer Ballester** y a al **Dr. José de León y de Juan** su confianza, su ayuda y su paciencia, lo cual ha posibilitado la realización de este Trabajo.

Me gustaría destacar la ayuda, colaboración y apoyo recibido por parte del resto de los miembros del Grupo de Procesado Digital de Señales (GPDS) del Departamento de Señales y Comunicaciones de la Universidad de Las Palmas: el **Dr. Carlos Manuel Travieso González** y **D. José Cruz Gil**. Por otro lado, mi agradecimiento por la disponibilidad de los recursos puestos a mi disposición, en especial el uso del Centro de Cálculo “José Cruz”.

Quisiera además, tener una nota de agradecimiento al servicio de Otorrinolaringología (ORL) del Hospital General de G.C. Dr. Negrín por su ayuda en la elaboración de este Trabajo, en particular al Jefe de Servicio el **Dr. José María Cuyás de Torres** así como al **Dr. Juan Francisco Rivero Suárez** y al **Dr. Francisco Ayudarte Tapias**.

Agradezco a **mi familia** todo su apoyo y sus ánimos, que son sin duda, los verdaderos elementos motrices de este Trabajo. Debido a su implicación además en la edición de este texto, quiero agradecerle además su ayuda a mi hermana **Elena** y mi novia **Dalia**.

Y por último, quiero agradecer a todos **mis Amigos** su apoyo y aliento en el transcurso de la elaboración de esta Tesis Doctoral.

PRÓLOGO

“La ciencia avanza a pasos, no a saltos.”

Thomas B. Macaulay (1800-1859)

TABLA DE CONTENIDOS

1.	INTRODUCCIÓN	1
1.1	ANTECEDENTES	3
1.1.1	<i>Los métodos Visuales Indirectos</i>	<i>5</i>
1.1.2	<i>Los métodos Visuales Directos.....</i>	<i>6</i>
1.1.3	<i>Métodos Directos.....</i>	<i>10</i>
1.1.4	<i>Métodos Indirectos.....</i>	<i>10</i>
1.1.5	<i>Extracción de la Presión del Flujo Glotal.....</i>	<i>12</i>
1.1.6	<i>Métodos basados en la audición.....</i>	<i>14</i>
1.2	MOTIVACIÓN.....	16
1.3	OBJETIVOS	17
1.4	METODOLOGÍA.....	18
1.5	RESULTADOS.....	22
1.6	ESTRUCTURA DE LA MEMORIA.....	24
1.7	REFERENCIAS.....	26
2.	FISIOLOGÍA DE LA LARINGE: TRASTORNOS DE LA FONACIÓN	27
2.1	FUNCIÓN FONATORIA.....	27
2.2	PRODUCCIÓN DE LA VOZ.....	27
2.2.1	<i>Teorías sobre la Fonación.....</i>	<i>32</i>
2.2.2	<i>Acústica Vocal.....</i>	<i>34</i>
2.2.3	<i>Características de la Voz.....</i>	<i>36</i>
2.2.4	<i>Evolución de la Voz.....</i>	<i>37</i>
2.3	TRASTORNOS DE LA FONACIÓN: CONCEPTO DE DISFONÍA	38
2.3.1	<i>Clasificación de las Disfonías</i>	<i>39</i>
2.4	PATOLOGÍAS BENIGNAS	40
2.4.1	<i>Nódulos Vocales.....</i>	<i>40</i>
2.4.2	<i>Pólipos Vocales.....</i>	<i>42</i>
2.4.3	<i>Edema de Reinke</i>	<i>43</i>
2.4.4	<i>Úlcera de Contacto o Granuloma.</i>	<i>45</i>
2.4.5	<i>Laringitis Crónica</i>	<i>46</i>
2.4.6	<i>Papilomas.....</i>	<i>46</i>
2.4.7	<i>Cáncer de Laringe.....</i>	<i>47</i>
2.5	DISFONÍAS NEUROMUSCULARES (FUNCIONALES).....	48
2.5.1	<i>Disfonías Hipercinéticas</i>	<i>48</i>
2.5.2	<i>Disfonías Hipocinéticas.....</i>	<i>49</i>
2.5.3	<i>Parálisis Cordal</i>	<i>50</i>
2.5.4	<i>Fonoponosis</i>	<i>50</i>

2.6	REFERENCIA.....	51
3.	LA CALIDAD DE LA VOZ: FENÓMENOS IMPLICADOS.....	53
3.1	DOMINIO TEMPORAL.....	55
3.2	DOMINIO ESPECTRAL.....	57
3.3	DOMINIO CEPSTRAL.....	64
3.4	DOMINIO DEL MODELO INVERSO.....	68
3.5	DOMINIO NO LINEAL.....	73
3.5.1	<i>Estado del Arte: ¿por qué usar modelos no lineales?</i>	74
3.5.2	<i>Descripción de Fenómenos Físicos no lineales Implicados.....</i>	76
3.6	FENÓMENOS FÍSICOS IMPLICADOS EN LA CALIDAD DE LA VOZ.....	83
3.7	REFERENCIAS.....	86
4.	BASES DE DATOS.....	91
4.1	CARACTERÍSTICAS DE LAS BASES DE DATOS DE VOCES.....	93
4.1.1	<i>Influencia del Proceso de Grabación.....</i>	93
4.1.2	<i>Influencia de la Presencia de Ruido.....</i>	94
4.1.3	<i>El Contenido.....</i>	94
4.1.4	<i>Cantidades y Balanceadas.....</i>	95
4.2	BASES DE DATOS EMPLEADAS.....	96
4.2.1	<i>Base de datos Propia.....</i>	97
4.2.2	<i>Base de datos Comercial.....</i>	104
4.3	REFERENCIAS.....	105
5.	SELECCIÓN DE MEDIDAS.....	107
5.1	REVISIÓN DE LOS MÉTODOS.....	109
5.2	MÉTODO PROPUESTO.....	118
5.3	REFERENCIAS.....	123
6.	CUANTIFICACIÓN DE LA ESTABILIDAD DE LA VOZ.....	127
6.1	ESTADO DE LA TÉCNICA.....	128
6.1.1	<i>Dominio Temporal.....</i>	128
6.1.2	<i>Dominio espectral.....</i>	130
6.1.3	<i>Dominio del Modelo Inverso.....</i>	130
6.2	ANÁLISIS.....	132
6.3	APORTACIONES.....	132
6.4	NOMENCLATURA.....	133
6.5	VALORACIÓN DE LA RELEVANCIA DE LAS MEDIDAS DE CALIDAD DE LA VOZ.....	134
6.5.1	<i>Estudio: Sano – Patológico.....</i>	134
6.5.2	<i>Estudio: Sano – Disfonía Leve – Disfonía Moderada – Disfonía Severa.....</i>	135
6.6	EVALUACIÓN DE LAS MEDIDAS DE CALIDAD DE LA VOZ.....	136
6.6.1	<i>Estudio: Sano – Patológico.....</i>	137

6.6.2	<i>Estudio: Sano – Disfonía Leve – Disfonía Moderada – Disfonía Severa</i>	139
6.7	EVALUACIÓN	141
6.8	REFERENCIAS.....	143
7.	CUANTIFICACIÓN DE LA RIQUEZA ESPECTRAL	145
7.1	ESTADO DE LA TÉCNICA.....	146
7.1.1	<i>Dominio Espectral</i>	146
7.1.2	<i>Dominio Cepstral</i>	148
7.1.3	<i>Dominio del Modelo Inverso</i>	153
7.2	ANÁLISIS.....	153
7.3	APORTACIONES.....	154
7.3.1	<i>Dominio Espectral</i>	155
7.3.2	<i>Dominio Cepstral</i>	156
7.4	NOMENCLATURA.....	159
7.5	VALORACIÓN DE LA RELEVANCIA DE LAS MEDIDAS DE CALIDAD DE LA VOZ.....	160
7.5.1	<i>Estudio: Sano – Patológico</i>	160
7.5.2	<i>Estudio: Sano – Disfonía Leve – Disfonía Moderada – Disfonía Severa</i>	162
7.6	EVALUACIÓN DE LAS MEDIDAS DE CALIDAD DE LA VOZ	163
7.6.1	<i>Estudio: Sano – Patológico</i>	164
7.6.2	<i>Estudio: Sano – Disfonía Leve – Disfonía Moderada – Disfonía Severa</i>	166
7.7	EVALUACIÓN	168
7.8	REFERENCIAS.....	170
8.	CUANTIFICACIÓN DE LA PRESENCIA DE RUIDO.....	173
8.1	ESTADO DE LA TÉCNICA.....	174
8.1.1	<i>Dominio Temporal</i>	174
8.1.2	<i>Dominio Espectral</i>	174
8.1.3	<i>Dominio Cepstral</i>	180
8.1.4	<i>Dominio del modelo inverso</i>	180
8.2	ANÁLISIS.....	183
8.3	APORTACIONES.....	183
8.3.1	<i>Dominio Espectral</i>	184
8.4	NOMENCLATURA.....	185
8.5	VALORACIÓN DE LA RELEVANCIA DE LAS MEDIDAS DE CALIDAD DE LA VOZ.....	187
8.5.1	<i>Estudio: Sano – Patológico</i>	187
8.5.2	<i>Estudio: Sano – Disfonía Leve – Disfonía Moderada – Disfonía Severa</i>	188
8.6	EVALUACIÓN DE LAS MEDIDAS DE CALIDAD DE LA VOZ	190
8.6.1	<i>Estudio: Sano – Patológico</i>	190
8.6.2	<i>Estudio: Sano – Disfonía Leve – Disfonía Moderada – Disfonía Severa</i>	193
8.7	EVALUACIÓN	195
8.8	REFERENCIAS.....	196

9.	CUANTIFICACIÓN DEL COMPORTAMIENTO NO LINEAL	199
9.1	ESTADO DE LA TÉCNICA.....	200
9.1.1	<i>Modulaciones angulares.....</i>	<i>200</i>
9.1.2	<i>Estadísticos de Orden Superior (High order statistic (HOS)).....</i>	<i>206</i>
9.1.3	<i>Sistemas dinámicos no lineales deterministas: Sistemas Caóticos.....</i>	<i>206</i>
9.1.4	<i>Modelos No Lineales de Cuerdas Vocales</i>	<i>208</i>
9.2	ANÁLISIS.....	213
9.3	APORTACIONES	214
9.3.1	<i>Estadísticos de Orden Superior (HOS).....</i>	<i>214</i>
9.3.2	<i>Sistemas dinámicos no lineales deterministas: Sistemas Caóticos.....</i>	<i>221</i>
9.4	NOMENCLATURA.....	224
9.5	VALORACIÓN DE LA RELEVANCIA DE LAS MEDIDAS DE CALIDAD DE LA VOZ	225
9.5.1	<i>Estudio: Sano – Patológico</i>	<i>226</i>
9.5.2	<i>Estudio: Sano – Disfonía Leve – Disfonía Moderada – Disfonía Severa.....</i>	<i>227</i>
9.6	EVALUACIÓN DE LAS MEDIDAS DE CALIDAD DE LA VOZ	228
9.6.1	<i>Estudio: Sano – Patológico</i>	<i>228</i>
9.6.2	<i>Estudio: Sano – Disfonía Leve – Disfonía Moderada – Disfonía Severa.....</i>	<i>230</i>
9.7	EVALUACIÓN	232
9.8	REFERENCIAS.....	234
10.	APLICACIONES.....	239
10.1	SISTEMA DE DETECCIÓN AUTOMÁTICA DE TRASTORNOS DEL SISTEMA FONADOR	240
10.1.1	<i>Adquisición de Voz.....</i>	<i>242</i>
10.1.2	<i>Parametrización</i>	<i>243</i>
10.1.3	<i>Clasificación.....</i>	<i>245</i>
10.1.4	<i>Resultados.....</i>	<i>246</i>
10.1.5	<i>Test del sistema.....</i>	<i>248</i>
10.2	CUANTIFICADOR DEL GRADO DE DISFONÍA	251
10.2.1	<i>Sistema automático de medida del grado de disfonía propuesto.....</i>	<i>255</i>
10.3	SISTEMA DE DETECCIÓN AUTOMÁTICA DE PATOLOGÍAS DEL SISTEMA FONADOR EN HABLA CONTINUA	259
10.3.1	<i>Medidas de la calidad de la voz.....</i>	<i>260</i>
10.4	PROTOCOLO DE EVALUACIÓN	266
10.4.1	<i>Revisión de Experiencias Clínicas.....</i>	<i>267</i>
10.4.2	<i>Protocolo Propuesto.....</i>	<i>271</i>
10.5	REFERENCIAS	283
11.	CONCLUSIONES Y LÍNEAS FUTURAS	285
11.1	LÍNEAS FUTURAS	289
12.	RESULTADOS DE LA SELECCIÓN DE CARACTERÍSTICAS	291

12.1	MEDIDAS DE CALIDAD	292
12.2	VALORACIÓN DE LA RELEVANCIA DE LAS MEDIDAS DE CALIDAD DE LA VOZ	295
12.2.1	Estudio: Sano – Patológico	296
12.2.2	Estudio: Sano – Disfonía Leve – Disfonía Moderada – Disfonía Severa	302
12.3	EVALUACIÓN DE LAS MEDIDAS DE CALIDAD DE LA VOZ	308
12.3.1	Estudio: Sano – Patológico	309
12.3.2	Estudio: Sano – Disfonía Leve – Disfonía Moderada – Disfonía Severa	333
12.4	REFERENCIAS	357

ÍNDICE DE FIGURAS

Figura 3-1: Representación de la señal de voz en el dominio temporal correspondiente a la fonación de cinco vocales de forma sostenida de un locutor con voz de alta calidad (superior) y de un locutor con voz de baja calidad (inferior).....	56
Figura 3-2: Representación en el dominio temporal de un sonido sonoro sostenido....	57
Figura 3-3: Representación en el dominio temporal de un sonido sonoro sostenido durante un corto intervalo de tiempo para una voz de alta calidad (izquierda) y para una voz de baja calidad (derecha).....	57
Figura 3-4: Representación del espectro en escala logarítmica de una vocal sostenida producida por un locutor con voz de alta calidad (superior) y el espectro en escala logarítmica de la misma vocal producida por un locutor con voz de baja calidad (inferior).....	59
Figura 3-5: Envoltorio del espectro de una trama de voz donde aparece señalado los cuatro primeros formantes	60
Figura 3-6: Espectrograma estimado para una frecuencia de muestreo de 22050 Hz de una muestra de voz de alta calidad (superior) y una muestra de voz de baja calidad (inferior) durante la fonación de la cinco vocales....	62
Figura 3-7 : Valores de la frecuencia de pitch para las distintas tramas de una muestra de voz de alta calidad (color azul) y de una muestra de voz de baja calidad (color rojo). Ambas muestras corresponden a la vocal "a" sostenida de un locutor masculino.....	64
Figura 3-8: Representación en el dominio de los cepstrum de potencia de una muestra de voz de alta calidad (superior) y para una muestra de voz de baja calidad (inferior).....	65
Figura 3-9: Representación en el dominio de los cepstrum de potencia rectificadas de una muestra de voz de alta calidad (superior) y para una muestra de voz de baja calidad (inferior).....	66
Figura 3-10: Localización del primer y segundo armónico en el dominio de los cepstrum de potencia rectificadas.....	67
Figura 3-11: Modelo y parámetros LF	69
Figura 3-12: Modelo de fuentes y filtros (source/filter model) para la producción de voz.....	74
Figura 3-13: Interpretación clásica de la propagación del flujo de aire a través del tracto vocal para la producción de voz. [30]	75
Figura 3-14: Interpretación de la dinámica de fluidos no lineal de la propagación del flujo de aire a través del tracto vocal para la producción de voz [30].....	76
Figura 4-1: Número de muestras de voces sanas y voces patológicas	101
Figura 4-2: Distribución del porcentaje de muestras en función del sexo y su tipo.....	101
Figura 4-3: Distribución del porcentaje de las muestras patológicas según el grado de disfonía.....	102
Figura 4-4: Distribución del porcentaje de las muestras patológicas en función del sexo y su grado de disfonía.....	102
Figura 4-5: Distribución del porcentaje de muestras patológicas en función del grupo de patologías.....	103
Figura 4-6: Distribución del porcentaje de muestras patológicas del grupo "cuerdas sin lesión" entre las distintas patologías.....	103

Figura 4-7: Distribución del porcentaje de muestras patológicas del grupo “cuerdas con lesión” entre las distintas patologías.....	103
Figura 5-1: Esquema de un sistema automático de medida de calidad de voz.....	108
Figura 5-2: Esquema del procedimiento de selección de características de medidas de calidad de voz propuesto.....	121
Figura 7-1: Cepstrum complejos de una muestra de voz sana y de una muestra de voz patológica.....	149
Figura 7-2: Representación en el dominio de los cepstrum de potencia de una muestra de voz de voz de alta calidad (superior) y para una muestra de voz de baja calidad (inferior).....	150
Figura 7-3: Representación de las diferentes bandas en las que se divide el espectro de una trama de voz en el proceso de estimación del grado de sonoridad por trama.....	156
Figura 7-4: Localización del primer y segundo armónico en el dominio de los cepstrum reales.....	157
Figura 7-5: Localización del primer y segundo armónico en el dominio de los cepstrum derivados.....	158
Figura 8-1: Esquema de bloques del cálculo del parámetro GNE.....	181
Figura 9-1: Esquema de la estimación de las medidas de calidad de la voz basadas en modulaciones angulares.....	202
Figura 9-2: Espectro de una trama de voz (azul) y la respuesta del filtro pasobanda centrado en el primer formante (rojo). En el espectro se puede apreciar la envolvente espectral calculada mediante coeficientes LPC.....	203
Figura 9-3: Señal moduladora de AM de una señal de voz de alta calidad (voz sana) y de una señal de voz de baja calidad (voz patológica).....	204
Figura 9-4: Señal moduladora de FM de una señal de voz de alta calidad (voz sana) y de una señal de voz de baja calidad (voz patológica).....	204
Figura 9-5: Funciones de autocorrelación normalizada de la señal moduladora de AM para voz de alta calidad (azul) y para voz de baja calidad (roja).....	205
Figura 9-6: Envolventes de las funciones de autocorrelación normalizada de la señal moduladora de AM para voz de alta calidad (azul) y para voz de baja calidad (roja).....	205
Figura 9-7: Representación en el espacio de fases de una fonación sostenida de una muestra de voz para distintas calidades de voz.....	207
Figura 9-8 : Modelo de cuerdas vocales “Modelo de una masa”.....	209
Figura 9-9 : Modelo de cuerdas vocales “Modelo de dos masas”.....	209
Figura 9-10 : Modelo de cuerdas vocales “Modelo de múltiples masas”.....	210
Figura 9-11 : Modelo de cuerdas vocales “Modelo de una masa”.....	211
Figura 9-12 : Modelo de cuerdas vocales “Modelo de dos masas modificado”.....	212
Figura 9-13: Índice de bicoherencia de una trama de voz de alta calidad.....	215
Figura 9-14: Índice de bicoherencia de una trama de voz de baja calidad.....	215
Figura 9-15: Índice de bicoherencia integrado para una trama de voz de alta calidad y para una trama de voz de baja calidad.	217
Figura 9-16: Superposición de la representación de la estimación del espectro de potencia calculado por medio de la FFT (azul) y por medio del biespectro (rojo), para una trama de voz de alta calidad (voz sana) y para una trama de baja calidad (voz patológica).....	219
Figura 10-1 :Esquema general de sistema del detección automático de trastornos del sistema fonador.....	241

Figura 10-2: Señal de voz registrada correspondiente a la fonación de cinco vocales de forma sostenida.....	242
Figura 10-3: Esquema de la red de clasificadores del sistema de detección automático de trastornos del sistema fonador.....	245
Figura 10-4: Muestra de habla de un locutor de alta calidad de voz.....	261
Figura 10-5: Muestra de habla de un locutor de baja calidad de voz.....	262
Figura 10-6: Esquema general de un sistema para la caracterización de la voz de un locutor a partir de habla continua.....	263

ÍNDICE DE ILUSTRACIONES

Ilustración 1-1: Espejo Laríngeo	5
Ilustración 1-2 : Imagen de intervención quirúrgica usando Microlaringoscopio.....	6
Ilustración 1-3 : Imagen de un sistema de videoendoscopia	7
Ilustración 1-4 : Evaluación de un paciente por medio del fibroscopio	8
Ilustración 1-5 : Telelaringoscopia.....	9
Ilustración 1-6 : Instrumental utilizado en la transiluminación.....	10
Ilustración 1-7 : Imagen de la glotis obtenida por técnicas basadas en Tomografía.....	11
Ilustración 1-8: Imagen de la glotis obtenida por técnicas basadas en ultrasonidos	11
Ilustración 1-9: Electrodo utilizado en una electroglotografía y su sujeción y colocación en la garganta del paciente	13
Ilustración 1-10: Máscara de Rothenberg y su utilización	14
Ilustración 2-1 : Fisiología de la laringe.....	28
Ilustración 2-2 : Fisiología de las cuerdas vocales	29
Ilustración 2-3: Imagen de las cuerdas vocales	29
Ilustración 2-4 : Fase de abducción y fase de aducción de las cuerdas vocales.....	31
Ilustración 2-5 : Diferentes fases del movimiento de las cuerdas vocales	31
Ilustración 2-6 : Nódulos Vocales	40
Ilustración 2-7 : Pólipo vocal.....	43
Ilustración 2-8 : Edema de Reinke	44
Ilustración 2-9 : Úlcera de contacto o Granuloma.....	46
Ilustración 2-10 : Papiloma.....	47
Ilustración 2-11 : Cáncer de laringe	47
Ilustración 10-1: Esquema de un sistema de medida de calidad de la voz por medio de voz telefónica.....	260

ÍNDICE DE TABLAS

Tabla 1-1: Técnicas para la evaluación de la conducta laríngea	4
Tabla 2-1: Clasificación de las Difonías	40
Tabla 4-1 : Balance de la base de datos propia.....	100
Tabla 5-1: Clasificación de las técnicas de selección de características	111
Tabla 6-1: Medidas de calidad de la voz pertenecientes al Grupo I (Estabilidad de la voz).....	133
Tabla 6-2: Orden en función de la relevancia para el estudio Sano – Patológico	135
Tabla 6-3: Orden en función de la relevancia para el estudio Sano – Disfonía Leve – Disfonía Moderada – Disfonía Severa	136
Tabla 7-1: Medidas de calidad de la voz pertenecientes al Grupo II (Riqueza Espectral).....	159
Tabla 7-2: Orden en función de la relevancia para el estudio Sano – Patología.....	161
Tabla 7-3: Orden en función de la relevancia para el estudio Sano – Disfonía Leve – Disfonía Moderada – Disfonía Severa	162
Tabla 8-1: Medidas de calidad de la voz pertenecientes al Grupo III (Presencia de Ruido).....	186
Tabla 8-2: Orden en función de la relevancia para el estudio Sano – Patológico	187
Tabla 8-3: Orden en función de la relevancia para el estudio Sano – Disfonía Leve – Disfonía Moderada – Disfonía Severa	189
Tabla 9-1: Medidas de calidad de la voz pertenecientes al Grupo IV (Comportamiento no Lineal).....	225
Tabla 9-2: Orden en función de la relevancia para el estudio Sano – Patológico	226
Tabla 9-3: Orden en función de la relevancia para el estudio Sano – Disfonía Leve – Disfonía Moderada – Disfonía Severa	227
Tabla 10-1: Medidas de calidad de la voz utilizadas en la parametrización del sistema de detección automático de trastornos del sistema fonador	244
Tabla 10-2: Tasas de éxito (%) en la clasificación obtenidas con diferentes clasificadores en el sistema de detección automático de trastornos del sistema fonador usando la base de datos propia.....	246
Tabla 10-3: Tasas de éxito (%) en la clasificación obtenidas con diferentes clasificadores en el sistema de detección automático de trastornos del sistema fonador usando la base de datos <i>Voice Disorders Database Model 4337</i>	249
Tabla 10-4: Medidas de calidad de la voz utilizadas en la parametrización del sistema automático de medida del grado de disfonía	256
Tabla 10-5: Tasas de éxito (%) en la clasificación obtenidas con diferentes clasificadores en el sistema de detección automático de trastornos del sistema fonador.....	257
Tabla 10-6: Medidas de calidad estudiadas en el sistema para la caracterización de la voz de un locutor a partir de habla continua.....	264
Tabla 10-7: Ordenación de las medidas de calidad estudiadas en el sistema para la caracterización de la voz de un locutor a partir de habla continua en función de su relevancia	265
Tabla 10-8: Tasas de éxito (%) con Evaluación incremental comenzando con la característica más relevante en el problema de la cuantificación automática de la calidad de la voz	266

Tabla 10-9 : Medidas para cuantificar los distintos fenómenos físicos implicados en la medida de la calidad de la voz.	273
Tabla 10-10 : Rangos de normalidad de las diferentes medidas de la calidad de la voz.	282
Tabla 12-1: Medidas de calidad de la voz pertenecientes al Grupo I (Estabilidad de la voz)	292
Tabla 12-2: Medidas de calidad de la voz pertenecientes al Grupo II (Riqueza Espectral).....	293
Tabla 12-3: Medidas de calidad de la voz pertenecientes al Grupo III (Presencia de Ruido).....	294
Tabla 12-4: Medidas de calidad de la voz pertenecientes al Grupo IV (Comportamiento no Lineal).....	295
Tabla 12-5: Orden en función de la relevancia para el estudio Sano – Patología de la vocal “a”	296
Tabla 12-6: Orden en función de la relevancia para el estudio Sano – Patología de la vocal “E”.....	297
Tabla 12-7: Orden en función de la relevancia para el estudio Sano – Patología de la vocal “I”.....	298
Tabla 12-8: Orden en función de la relevancia para el estudio Sano – Patología de la vocal “O”	299
Tabla 12-9: Orden en función de la relevancia para el estudio Sano – Patología de la vocal “U”	301
Tabla 12-10: Orden en función de la relevancia para el estudio Sano – Disfonía Leve – Disfonía Moderada – Disfonía Severa de la vocal “a”	302
Tabla 12-11: Orden en función de la relevancia para el estudio Sano – Disfonía Leve – Disfonía Moderada – Disfonía Severa de la vocal “e”.....	303
Tabla 12-12: Orden en función de la relevancia para el estudio Sano – Disfonía Leve – Disfonía Moderada – Disfonía Severa de la vocal “i”.....	304
Tabla 12-13: Orden en función de la relevancia para el estudio Sano – Disfonía Leve – Disfonía Moderada – Disfonía Severa de la vocal “o”.....	306
Tabla 12-14: Orden en función de la relevancia para el estudio Sano – Disfonía Leve – Disfonía Moderada – Disfonía Severa de la vocal “u”.....	307

ÍNDICE DE GRÁFICOS

Gráfico 6-1: Evaluación secuencial de la relevancia para las medidas de calidad de la voz del grupo I usando el clasificador PNN en la clasificación entre voz sana y voz patológica.....	137
Gráfico 6-2: Evaluación secuencial de la relevancia para las medidas de calidad de la voz del grupo I usando el clasificador NN en la clasificación entre voz sana y voz patológica.....	137
Gráfico 6-3: Evaluación secuencial de la relevancia para las medidas de calidad de la voz del grupo I usando el clasificador RBF en la clasificación entre voz sana y voz patológica.....	138
Gráfico 6-4: Evaluación secuencial de la relevancia para las medidas de calidad de la voz del grupo I usando el clasificador SVM en la clasificación entre voz sana y voz patológica.....	138
Gráfico 6-5: Evaluación secuencial de la relevancia para las medidas de calidad de la voz del grupo I usando el clasificador KNN en la clasificación entre voz sana y voz patológica.....	139
Gráfico 6-6: Evaluación secuencial de la relevancia para las medidas de calidad de la voz del grupo I usando el clasificador PNN en la clasificación entre distintos grados de disfonía.....	139
Gráfico 6-7: Evaluación secuencial de la relevancia para las medidas de calidad de la voz del grupo I usando el clasificador NN en la clasificación entre distintos grados de disfonía.....	140
Gráfico 6-8: Evaluación secuencial de la relevancia para las medidas de calidad de la voz del grupo I usando el clasificador RBF en la clasificación entre distintos grados de disfonía.....	140
Gráfico 6-9: Evaluación secuencial de la relevancia para las medidas de calidad de la voz del grupo I usando el clasificador PNN en la clasificación entre distintos grados de disfonía.....	141
Gráfico 6-10: Evaluación secuencial de la relevancia para las medidas de calidad de la voz del grupo I usando el clasificador KNN en la clasificación entre distintos grados de disfonía.....	141
Gráfico 7-1: Evaluación secuencial de la relevancia para las medidas de calidad de la voz del grupo II usando el clasificador PNN en la clasificación entre voz sana y voz patológica.....	164
Gráfico 7-2: Evaluación secuencial de la relevancia para las medidas de calidad de la voz del grupo II usando el clasificador NN en la clasificación entre voz sana y voz patológica.....	164
Gráfico 7-3: Evaluación secuencial de la relevancia para las medidas de calidad de la voz del grupo II usando el clasificador RBF en la clasificación entre voz sana y voz patológica.....	165
Gráfico 7-4: Evaluación secuencial de la relevancia para las medidas de calidad de la voz del grupo II usando el clasificador SVM en la clasificación entre voz sana y voz patológica.....	165
Gráfico 7-5: Evaluación secuencial de la relevancia para las medidas de calidad de la voz del grupo II usando el clasificador KNN en la clasificación entre voz sana y voz patológica.....	166

Gráfico 7-6: Evaluación secuencial de la relevancia para las medidas de calidad de la voz del grupo II usando el clasificador PNN en la clasificación entre distintos grados de disfonía	166
Gráfico 7-7: Evaluación secuencial de la relevancia para las medidas de calidad de la voz del grupo II usando el clasificador NN en la clasificación entre distintos grados de disfonía	167
Gráfico 7-8: Evaluación secuencial de la relevancia para las medidas de calidad de la voz del grupo II usando el clasificador RBF en la clasificación entre distintos grados de disfonía	167
Gráfico 7-9: Evaluación secuencial de la relevancia para las medidas de calidad de la voz del grupo II usando el clasificador PNN en la clasificación entre distintos grados de disfonía	168
Gráfico 7-10: Evaluación secuencial de la relevancia para las medidas de calidad de la voz del grupo II usando el clasificador KNN en la clasificación entre distintos grados de disfonía	168
Gráfico 8-1: Evaluación secuencial de la relevancia para las medidas de calidad de la voz del grupo III usando el clasificador PNN en la clasificación entre voz sana y voz patológica.....	190
Gráfico 8-2: Evaluación secuencial de la relevancia para las medidas de calidad de la voz del grupo III usando el clasificador NN en la clasificación entre voz sana y voz patológica.....	191
Gráfico 8-3 Evaluación secuencial de la relevancia para las medidas de calidad de la voz del grupo III usando el clasificador RBF en la clasificación entre voz sana y voz patológica.....	191
Gráfico 8-4 Evaluación secuencial de la relevancia para las medidas de calidad de la voz del grupo III usando el clasificador SVM en la clasificación entre voz sana y voz patológica.....	192
Gráfico 8-5: Evaluación secuencial de la relevancia para las medidas de calidad de la voz del grupo III usando el clasificador KNN en la clasificación entre voz sana y voz patológica.....	192
Gráfico 8-6: Evaluación secuencial de la relevancia para las medidas de calidad de la voz del grupo III usando el clasificador PNN en la clasificación entre distintos grados de disfonía	193
Gráfico 8-7: Evaluación secuencial de la relevancia para las medidas de calidad de la voz del grupo III usando el clasificador NN en la clasificación entre distintos grados de disfonía	193
Gráfico 8-8: Evaluación secuencial de la relevancia para las medidas de calidad de la voz del grupo III usando el clasificador RBF en la clasificación entre distintos grados de disfonía	194
Gráfico 8-9: Evaluación secuencial de la relevancia para las medidas de calidad de la voz del grupo III usando el clasificador PNN en la clasificación entre distintos grados de disfonía	194
Gráfico 8-10: Evaluación secuencial de la relevancia para las medidas de calidad de la voz del grupo III usando el clasificador KNN en la clasificación entre distintos grados de disfonía.....	194
Gráfico 9-1: Evaluación secuencial de la relevancia para las medidas de calidad de la voz del grupo IV usando el clasificador PNN en la clasificación entre voz sana y voz patológica.....	228

Gráfico 9-2: Evaluación secuencial de la relevancia para las medidas de calidad de la voz del grupo IV usando el clasificador NN en la clasificación entre voz sana y voz patológica.....	229
Gráfico 9-3 Evaluación secuencial de la relevancia para las medidas de calidad de la voz del grupo IV usando el clasificador RBF en la clasificación entre voz sana y voz patológica.....	229
Gráfico 9-4 Evaluación secuencial de la relevancia para las medidas de calidad de la voz del grupo IV usando el clasificador SVM en la clasificación entre voz sana y voz patológica.....	229
Gráfico 9-5: Evaluación secuencial de la relevancia para las medidas de calidad de la voz del grupo IV usando el clasificador KNN en la clasificación entre voz sana y voz patológica.....	230
Gráfico 9-6: Evaluación secuencial de la relevancia para las medidas de calidad de la voz del grupo IV usando el clasificador PNN en la clasificación entre diferentes grados de disfonía.....	230
Gráfico 9-7: Evaluación secuencial de la relevancia para las medidas de calidad de la voz del grupo IV usando el clasificador NN en la clasificación entre diferentes grados de disfonía.....	231
Gráfico 9-8: Evaluación secuencial de la relevancia para las medidas de calidad de la voz del grupo IV usando el clasificador PNN en la clasificación entre diferentes grados de disfonía.....	231
Gráfico 9-9: Evaluación secuencial de la relevancia para las medidas de calidad de la voz del grupo IV usando el clasificador PNN en la clasificación entre diferentes grados de disfonía.....	232
Gráfico 9-10: Evaluación secuencial de la relevancia para las medidas de calidad de la voz del grupo IV usando el clasificador KNN en la clasificación entre diferentes grados de disfonía.....	232
Gráfico 12-1: Evaluación secuencial de la relevancia para las medidas de calidad de la voz del grupo I usando el clasificador PNN en la clasificación entre voz sana y voz patológica.....	309
Gráfico 12-2: Evaluación secuencial de la relevancia para las medidas de calidad de la voz del grupo II usando el clasificador PNN en la clasificación entre voz sana y voz patológica.....	311
Gráfico 12-3: Evaluación secuencial de la relevancia para las medidas de calidad de la voz del grupo III usando el clasificador PNN en la clasificación entre voz sana y voz patológica.....	312
Gráfico 12-4: Evaluación secuencial de la relevancia para las medidas de calidad de la voz del grupo IV usando el clasificador PNN en la clasificación entre voz sana y voz patológica.....	313
Gráfico 12-5: Evaluación secuencial de la relevancia para las medidas de calidad de la voz del grupo I usando el clasificador NN en la clasificación entre voz sana y voz patológica.....	314
Gráfico 12-6: Evaluación secuencial de la relevancia para las medidas de calidad de la voz del grupo II usando el clasificador NN en la clasificación entre voz sana y voz patológica.....	315
Gráfico 12-7: Evaluación secuencial de la relevancia para las medidas de calidad de la voz del grupo III usando el clasificador NN en la clasificación entre voz sana y voz patológica.....	317

Gráfico 12-8: Evaluación secuencial de la relevancia para las medidas de calidad de la voz del grupo IV usando el clasificador NN en la clasificación entre voz sana y voz patológica.....	318
Gráfico 12-9: Evaluación secuencial de la relevancia para las medidas de calidad de la voz del grupo I usando el clasificador RBF en la clasificación entre voz sana y voz patológica.....	319
Gráfico 12-10 Evaluación secuencial de la relevancia para las medidas de calidad de la voz del grupo II usando el clasificador RBF en la clasificación entre voz sana y voz patológica.....	320
Gráfico 12-11 Evaluación secuencial de la relevancia para las medidas de calidad de la voz del grupo III usando el clasificador RBF en la clasificación entre voz sana y voz patológica.....	321
Gráfico 12-12 Evaluación secuencial de la relevancia para las medidas de calidad de la voz del grupo IV usando el clasificador RBF en la clasificación entre voz sana y voz patológica.....	322
Gráfico 12-13: Evaluación secuencial de la relevancia para las medidas de calidad de la voz del grupo I usando el clasificador SVM en la clasificación entre voz sana y voz patológica.....	323
Gráfico 12-14 : Evaluación secuencial de la relevancia para las medidas de calidad de la voz del grupo II usando el clasificador SVM en la clasificación entre voz sana y voz patológica.....	325
Gráfico 12-15 : Evaluación secuencial de la relevancia para las medidas de calidad de la voz del grupo III usando el clasificador SVM en la clasificación entre voz sana y voz patológica.....	326
Gráfico 12-16 : Evaluación secuencial de la relevancia para las medidas de calidad de la voz del grupo IV usando el clasificador SVM en la clasificación entre voz sana y voz patológica.....	327
Gráfico 12-17: Evaluación secuencial de la relevancia para las medidas de calidad de la voz del grupo I usando el clasificador KNN en la clasificación entre voz sana y voz patológica.....	328
Gráfico 12-18: Evaluación secuencial de la relevancia para las medidas de calidad de la voz del grupo II usando el clasificador KNN en la clasificación entre voz sana y voz patológica.....	330
Gráfico 12-19: Evaluación secuencial de la relevancia para las medidas de calidad de la voz del grupo III usando el clasificador KNN en la clasificación entre voz sana y voz patológica.....	331
Gráfico 12-20: Evaluación secuencial de la relevancia para las medidas de calidad de la voz del grupo IV usando el clasificador KNN en la clasificación entre voz sana y voz patológica.....	332
Gráfico 12-21: Evaluación secuencial de la relevancia para las medidas de calidad de la voz del grupo I usando el clasificador PNN en la clasificación entre diferentes grados de disfonía.....	333
Gráfico 12-22: Evaluación secuencial de la relevancia para las medidas de calidad de la voz del grupo II usando el clasificador PNN en la clasificación entre diferentes grados de disfonía.....	335
Gráfico 12-23: Evaluación secuencial de la relevancia para las medidas de calidad de la voz del grupo III usando el clasificador PNN en la clasificación entre diferentes grados de disfonía.....	336

Gráfico 12-24: Evaluación secuencial de la relevancia para las medidas de calidad de la voz del grupo IV usando el clasificador PNN en la clasificación entre diferentes grados de disfonía.....	337
Gráfico 12-25: Evaluación secuencial de la relevancia para las medidas de calidad de la voz del grupo I usando el clasificador NN en la clasificación entre diferentes grados de disfonía.....	338
Gráfico 12-26: Evaluación secuencial de la relevancia para las medidas de calidad de la voz del grupo II usando el clasificador NN en la clasificación entre diferentes grados de disfonía.....	339
Gráfico 12-27: Evaluación secuencial de la relevancia para las medidas de calidad de la voz del grupo III usando el clasificador NN en la clasificación entre diferentes grados de disfonía.....	341
Gráfico 12-28: Evaluación secuencial de la relevancia para las medidas de calidad de la voz del grupo IV usando el clasificador NN en la clasificación entre diferentes grados de disfonía.....	342
Gráfico 12-29: Evaluación secuencial de la relevancia para las medidas de calidad de la voz del grupo I usando el clasificador RBF en la clasificación entre diferentes grados de disfonía.....	343
Gráfico 12-30: Evaluación secuencial de la relevancia para las medidas de calidad de la voz del grupo II usando el clasificador RBF en la clasificación entre diferentes grados de disfonía.....	344
Gráfico 12-31: Evaluación secuencial de la relevancia para las medidas de calidad de la voz del grupo III usando el clasificador RBF en la clasificación entre diferentes grados de disfonía.....	346
Gráfico 12-32: Evaluación secuencial de la relevancia para las medidas de calidad de la voz del grupo IV usando el clasificador PNN en la clasificación entre diferentes grados de disfonía.....	347
Gráfico 12-33: Evaluación secuencial de la relevancia para las medidas de calidad de la voz del grupo I usando el clasificador PNN en la clasificación entre diferentes grados de disfonía.....	348
Gráfico 12-34: Evaluación secuencial de la relevancia para las medidas de calidad de la voz del grupo II usando el clasificador PNN en la clasificación entre diferentes grados de disfonía.....	349
Gráfico 12-35: Evaluación secuencial de la relevancia para las medidas de calidad de la voz del grupo III usando el clasificador PNN en la clasificación entre diferentes grados de disfonía.....	351
Gráfico 12-36: Evaluación secuencial de la relevancia para las medidas de calidad de la voz del grupo IV usando el clasificador PNN en la clasificación entre diferentes grados de disfonía.....	352
Gráfico 12-37: Evaluación secuencial de la relevancia para las medidas de calidad de la voz del grupo I usando el clasificador KNN en la clasificación entre diferentes grados de disfonía.....	353
Gráfico 12-38: Evaluación secuencial de la relevancia para las medidas de calidad de la voz del grupo II usando el clasificador KNN en la clasificación entre diferentes grados de disfonía.....	354
Gráfico 12-39: Evaluación secuencial de la relevancia para las medidas de calidad de la voz del grupo III usando el clasificador KNN en la clasificación entre diferentes grados de disfonía.....	355

Gráfico 12-40: Evaluación secuencial de la relevancia para las medidas de calidad de la voz del grupo IV usando el clasificador KNN en la clasificación entre diferentes grados de disfonía.....	356

CAPÍTULO 1

Introducción

El diccionario de la Real Academia Española de la Lengua define **voz** como “*sonido que el aire expelido de los pulmones produce al salir de la laringe, haciendo que vibren las cuerdas vocales*”. Sin embargo, la voz es algo más, da la capacidad de comunicación al ser humano, define la identidad de un sujeto, permite conocer su estado de ánimo e incluso su estado de salud. Por lo tanto, la voz es una de las características biológicas que mejor definen a cada individuo.

Cuando existe un problema en el sistema fonador se pueden producir vibraciones anómalas de las cuerdas vocales y la aparición de ruido glótico, que provocan que la voz presente alteraciones (voz ronca, rasposa, apagada, entrecortada, etc.). Esta degradación de la calidad de la voz se conoce como **disfonía**. La disfonía, es por tanto, un término general que describe un cambio anormal de la voz producida por diferentes motivos.

Las disfonías son causadas principalmente por la presencia de patología en las cuerdas vocales, tanto debido a diferentes lesiones orgánicas (como malformaciones, lesiones benignas adquiridas, inflamaciones, infecciones, lesiones precancerosas y cancerosas, traumatismos, o afecciones neurológicas, endocrinas y audiológicas) o

diferentes disfonías funcionales (tanto en voz hablada como en voz cantada), e incluso de origen psiquiátrico. Pueden existir otros motivos diferentes a los de carácter médico que produzcan cambios en la calidad de la voz, como por ejemplo factores relacionados con el estado anímico.

La disfonía afecta a la calidad de vida de quien la padece: problemas en las relaciones sociales, alteración del estado anímico y la modificación de hábitos de vida (se evitan los ambientes con altos niveles de ruido, etc).

Actualmente vivimos en una sociedad en la que prima las medidas preventivas en lo que respecta al padecimiento de cualquier alteración en la salud, y sin embargo el diagnóstico precoz es difícil y generalmente es inevitable recurrir al diagnóstico por medio de técnicas invasivas.

Los sistemas informáticos proporcionan herramientas para la autodetección de trastornos en la calidad de la voz. Por lo tanto, resulta de interés desarrollar estudios sobre medidas de la calidad de la voz de un locutor con el fin de poder ser utilizados en la identificación de la presencia de irregularidades en el sistema de fonación. Adicionalmente estas medidas podrían usarse para la detección del estado anímico o simplemente evaluar la calidad de un sistema de telecomunicación, aunque esto cae fuera del ámbito de este estudio.

1.1 Antecedentes

Durante los últimos años se han desarrollado numerosos métodos de análisis y observación de la función fonatoria, la cual comprende fenómenos mecánicos y aerodinámicos. La observación de los movimientos del pliegue de las cuerdas vocales sólo es posible visualmente, mientras que en la práctica la excitación aerodinámica sólo puede ser examinada indirectamente. La inspección visual permite observar los parámetros dinámicos (y estáticos) de la fonación mientras que el flujo glotal debe medirse indirectamente o por medio de sensores adicionales.

Las diferentes técnicas para la evaluación de la conducta laríngea pueden ser clasificadas en diferentes grupos teniendo en cuenta la estrategia seguida en el proceso de evaluación. Los diferentes grupos identificados son:

- Métodos visuales directos
- Métodos visuales indirectos
- Métodos directos
- Métodos indirectos
- Extracción de la Presión del Flujo Glotal
- Métodos basados en la audición

En la Tabla 1-1 se muestra una clasificación de diferentes técnicas utilizadas para evaluar la conducta laríngea atendiendo a la estrategia de observación.

TABLA 1-1: TÉCNICAS PARA LA EVALUACIÓN DE LA CONDUCTA LARÍNGEA

Método de observación	Técnica
Métodos visuales Directos	Exploración endoscópica¹: • <i>Endoscopia flexible o fibroscopia</i> • <i>Endoscopia rígida o Telelaringoscopia</i> ¹ Elementos Adicionales: - Estudio Estroboscópico - Video - Técnicas fotográficas de gran velocidad
Métodos visuales Indirectos	Laringoscopia Refleja²: • <i>Laringoscopia refleja convencional</i> • <i>Micro-laringoscopia refleja</i> ² Elementos Adicionales: - Estudio Estroboscópico - Video - Técnicas fotográficas de gran velocidad
Métodos directos	Fotoglوتografía Transiluminación
Métodos Indirectos	Tomografía Laminografía Ultrasonidos
Extracción de la Presión del Flujo Glotal	Señales directas: • <i>Microfófono de contacto</i> • <i>Electroglotografía</i> Mediante filtrado inverso: • <i>Señal registrada por medio de un micrófono</i> • <i>Señal de la presión obtenida por medio de la máscara de Rothenberg</i>
Técnicas basadas en la audición	Valoración Subjetiva de la Calidad de la Voz

1.1.1 Los métodos Visuales Indirectos

Laringoscopia refleja convencional

La laringoscopia refleja convencional ha sido desde el año 1855, en que el Profesor de Canto español Manuel García ideó el espejillo laríngeo, el método habitualmente utilizado para observar la laringe hasta nuestros días.

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.

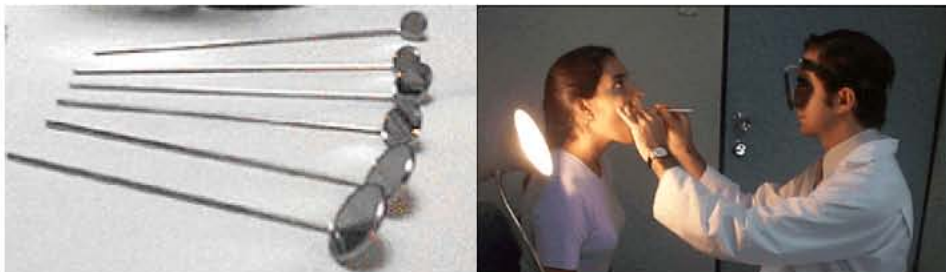


Ilustración 1-1: Espejo Laríngeo

Microlaringoscopia refleja

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 visores laterales, un monitor de video, una cámara de video, un magnetoscopio o cámara fotográfica, destacando la existencia de cámaras fotográficas de alta velocidad, y la posibilidad así de sustituir la iluminación convencional por luz estroboscópica, con lo que se obtienen imágenes visibles de las distintas fases del movimiento de las cuerdas vocales por medio de la iluminación de las distintas fases del movimiento captadas en distintos ciclos.



Ilustración 1-2 : Imagen de intervención quirúrgica usando Microlaringoscopio

1.1.2 Los métodos Visuales Directos

Exploración endoscópica

La utilización de la tecnología moderna ha modificado de forma extraordinaria la metodología de valoración visual de la laringe, haciendo posible al clínico la observación tanto del estado como del comportamiento laríngeo en condiciones altamente informativas. 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, e incluso se pueden aplicar técnicas estroboscópicas para estudiar el movimiento de las cuerdas vocales. El examen

repetido de los datos obtenidos para su análisis a partir la videoendoscopia, contribuye notablemente a la mejora del diagnóstico y de la docencia.



Ilustración 1-3 : Imagen de un sistema de videoendoscopia

Dependiendo del tipo de endoscopio que se utilice distinguimos dos métodos diferentes: la endoscopia flexible o fibroscopia y la endoscopia rígida o telelaringoscopia.

Endoscopia flexible o fibroscopia

Los fibroscopios son endoscopios constituidos por dos haces de fibras ópticas flexibles, que mediante una envoltura común constituyen una especie de cable en cuyo extremo se adapta una lente. Uno de los haces está destinado a la transmisión de la imagen mientras que el otro es el que propaga la luz.

Los fibroscopios están dotados de un sistema que mediante un leve control manual hace posible la torsión 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.

La vía de inserción habitual del fibroscopio es la nasal. El paso se realiza por el suelo de la fosa nasal. 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 áditus laríngeo o proseguir el descenso hasta observar las cuerdas vocales desde una distancia muy próxima que nos 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 podía 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.

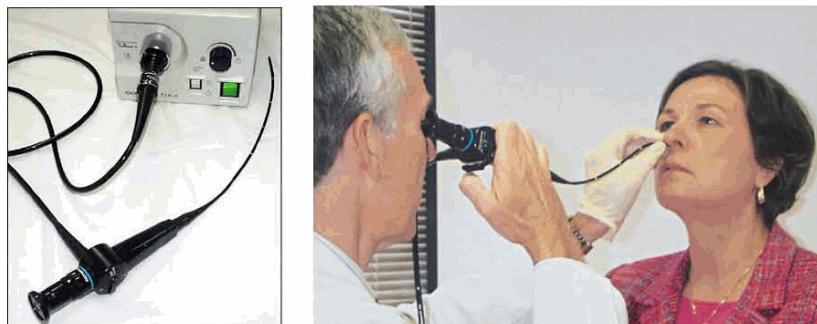


Ilustración 1-4 : Evaluación de un paciente por medio del fibroscopio

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 fibroendoscopio, lo que significa mejores prestaciones ópticas 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 telarlaringoscopios para laringología oscilan entre 4 y 10 mm.

Las prestaciones ópticas de los telarlaringoscopios superan en gran medida las de los fibroscopios, sobre todo en lo que se refiere a magnificación, iluminación en el 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, que son de máxima utilidad en relación al estudio estroboscópico.

La exploración de la laringe mediante telarlaringoscopia rígida se realiza utilizando como vía de introducción la cavidad oral. La posición del paciente depende del tipo de endoscopio. 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.

Tanto la fibroendoscopia como la telefaringoscopia 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.



Ilustración 1-5 : Telarlaringoscopia

1.1.3 Métodos Directos

Fotoglوتografía y Transiluminación

Una técnica de observación directa es la fotoglوتografía (PGG). Consiste en dirigir un foco de luz en el cuello del paciente, pero iluminando únicamente debajo de las cuerdas vocales. Posteriormente, por medio de un sensor fotoeléctrico situado en la faringe, es posible detectar el movimiento de las cuerdas vocales por medio de la cantidad de luz que capta el sensor. Para la utilización de esta técnica es necesario el uso de anestesia. Por medio de la modulación de la luz es posible captar características de la glotis. Esta técnica presenta diversas variaciones como la inserción del fotosensor por la cavidad nasal o la inversión del sistema insertando el foco de luz en la laringe y utilizando el fotosensor en el cuello. Esta última modificación permite el estudio de la conducta laríngea durante la articulación. Esta técnica se denomina transiluminación. En la Ilustración 1-6 se muestra el instrumental utilizado en la transiluminación.

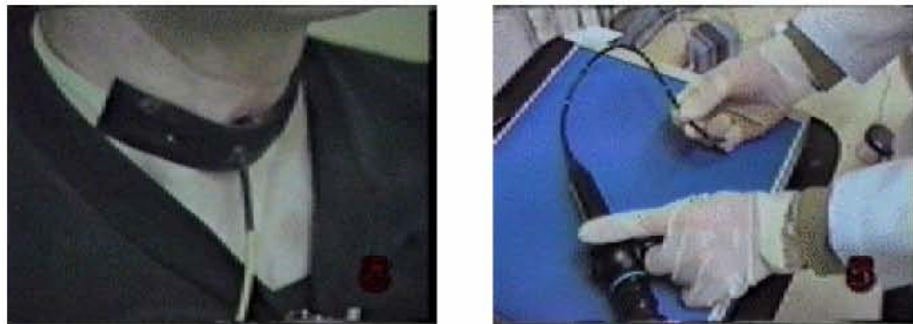


Ilustración 1-6 : Instrumental utilizado en la transiluminación

1.1.4 Métodos Indirectos

Tomografía y Laminografía

Otros métodos de inspección visual de la conducta laríngea son la radiografía y la tomografía por medio de ordenador. Estas técnicas permiten la obtención de imágenes estáticas de la glotis. Una modificación de esta última técnica denominada laminografía, permite mediante un movimiento alrededor de la zona glótica de la fuente de rayos X y de su sensor, obtener imágenes del movimiento glótico de toda una sección. Una vista

de la laringe es posible incluso durante la fonación. Adicionalmente, la combinación del estroboscopia y de la laminografía (estrobolaminografía - STROL) permite tomar imágenes del movimiento de los pliegues de las cuerdas vocales por medio de la sincronización de las dos técnicas.



Ilustración 1-7 : Imagen de la glotis obtenida por técnicas basadas en Tomografía

Ultrasonidos

Los ultrasonidos no se usan con asiduidad para observar las funciones de la laringe debido a la pobre calidad de las imágenes, sin embargo si es usada para observar los movimientos de las articulaciones del sistema fonador. En la Ilustración 1-8 se muestra una imagen de la glotis obtenida por técnicas basadas en ultrasonidos.

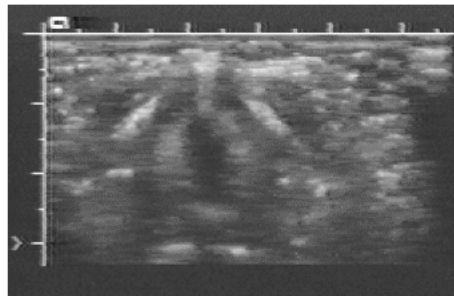


Ilustración 1-8: Imagen de la glotis obtenida por técnicas basadas en ultrasonidos

1.1.5 Extracción de la Presión del Flujo Glotal

Las medidas aeroacústicas del conducto laríngeo incluyen las medidas directas e indirectas de presión del flujo glótico.

Medidas Directas

Micrófono de contacto

Las medidas directas de presión del flujo glótico se realizan por medio de transductores de presión miniaturizados ajustados a un catéter los cuales se colocan sobre los pliegues de las cuerdas vocales. Este método se emplea raramente. Es necesario el uso de anestesia local para poder insertar los catéteres (ésto puede afectar al proceso de fonación). Esta técnica se utiliza con la finalidad de modelar el funcionamiento de los pliegues de las cuerdas vocales y de la excitación glótica, en tareas de investigación.

Electroglotografía

Existe otra técnica de estimar la forma de onda del flujo glótico, esta técnica se denomina electroglotografía (EGG). La EGG es una técnica que registra el comportamiento de la laringe mediante el cambio en impedancia eléctrica en la garganta durante el habla.

El principio básico de la electroglotografía consiste en aplicar una señal de alta frecuencia y con bajo voltaje y amperaje (fisiológicamente seguro), por medio de dos electrodos situados en la superficie de la garganta sobre el cartílago tiroideo. Los electrodos se hacen de cobre, plata u oro. Los electrodos tienen la forma de anillos o de rectángulos cubriendo un área comprendida entre 3 y 9 cm². Un tercer electrodo se usa a menudo como una referencia para las medidas de impedancia. Los electrodos normalmente son montados sobre una venda flexible, la cual permite ajustar la longitud y sostener los electrodos en una posición firme, además de permitir al paciente un habla cómodo y realizar una respiración no forzada. En la Ilustración 1-9 se muestran los electrodos utilizados en una electroglotografía, además también se pueden apreciar su sujeción y colocación en la garganta del paciente.



Ilustración 1-9: Electrodo utilizado en una electroglotografía y su sujeción y colocación en la garganta del paciente

Un generador de señal proporciona a los electrodos una corriente sinusoidal de una frecuencia que normalmente va de 300 kHz a 5 MHz. La corriente suministrada no es mayor de varios miliamperios. El voltaje entre los electrodos depende de la impedancia de los tejidos, siendo el valor típico de aproximadamente 0.5 V.

La variación de la impedancia medida mediante este sistema permite averiguar la calidad del contacto de las cuerdas vocales durante su vibración. Esto es así, ya que la impedancia es inversamente proporcional a la cantidad de superficie de contacto lateral entre los pliegues de las cuerdas vocales en la fase de cierre glótico.

Medidas Indirectas

Para la obtención de medidas de las características de funcionamiento de las cuerdas vocales y de la excitación glótica es necesario el uso de filtros inversos, por medio de los cuales, a partir de la eliminación del efecto del tracto vocal y del efecto de radiación de los labios es posible obtener la excitación glótica para su estudio. En el estudio de la voz por medio del filtrado inverso sólo se estudia la voz sonora. La principal ventaja de este método es la de tratarse de un método no invasivo.

Micrófono y Máscara de Rothengerg

Existen dos técnicas para realizar el filtrado inverso. La primera técnica se basa en grabar la voz mediante un micrófono de altas prestaciones y posteriormente, por medio de un modelado matemático de un filtro que plasma los efectos del tracto vocal y de la radiación de los labios, se obtiene aplicando un filtrado inverso la excitación glótica. El

inconveniente de este método es que no registra el nivel de continuo que presenta la señal de voz, mientras que por el contrario tiene la ventaja de recoger gran cantidad de información espectral (frecuencial) de la señal de voz.

La segunda técnica se basa en recoger el flujo de aire de la glotis por medio de una máscara en la cara (máscara de Rothenberg). La ventaja es que en este caso si se recoge el nivel de continuo de la señal, aunque las componentes espectrales registradas están muy limitadas. En esta técnica, y durante el proceso de grabación, es necesario que la máscara esté totalmente sellada a la cara, ya que la más mínima pérdida del flujo de aire afectaría en gran medida a los resultados del estudio. Para evitar movimientos musculares de la cara que provocaran este efecto generalmente en un estudio de este tipo se limitan las sílabas que se pronuncian. En la Ilustración 1-10 se muestra la máscara de Rothenberg y su utilización.

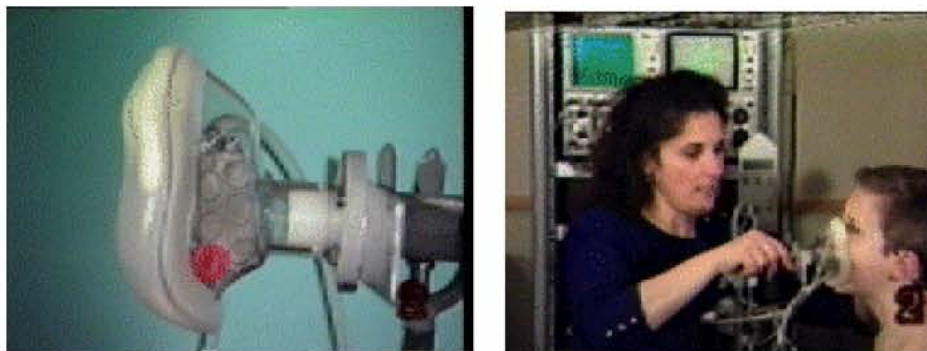


Ilustración 1-10: Máscara de Rothenberg y su utilización

1.1.6 Métodos basados en la audición

Actualmente una de las técnicas más usadas por los servicios de otorrinolaringología para realizar una inspección en primera instancia del estado de un paciente, es la inspección acústica de la voz. Para poder realizar una correcta evaluación de la voz del locutor, es preciso un correcto adiestramiento del profesional, así como un considerable bagaje profesional.

Una medida acústica de la calidad de la voz basada en una valoración auditiva es intrínsecamente difícil de establecer como referencia comparativa entre distintas voces y distintos evaluadores. En la bibliografía actual se han realizado varios intentos

de objetivar la medida de la calidad de voz por medio de medidas clínicas multidimensionales basadas en métodos auditivos. Existen ejemplos bien conocidos como son: la escala GRABS de Japón [1] y su formulario extendido desarrollado y aplicado en Europa [2], en Suecia se ha trabajado en otro juego de descriptores clínicos de la voz referentes a la percepción [3], o el conjunto de medidas [4] en el que se introduce un juego de características fonéticas el cual intenta agregar información de la excitación del tracto vocal en la medida de la calidad de la voz. Estos procedimientos de medida de la calidad de la voz pretenden alcanzar una medida objetiva a partir de una valoración subjetiva.

En la bibliografía actual existen diferentes trabajos en los que se proponen medidas objetivas de la calidad de la voz obtenidas a partir de una grabación de voz. Una característica común de la mayoría de estos trabajos es la de utilizar para el cálculo de las medidas de la voz la grabación de un sonido sonoro sostenido, principalmente una vocal. Esto es debido a que durante la fonación de este tipo de sonidos el sistema de producción de voz pone en funcionamiento gran parte de sus mecanismos (flujo glótico de aire constante, vibración de las cuerdas vocales de forma continuada, ...) permitiendo captar cualquier anomalía de estos mecanismos.

En estos trabajos, los distintos autores presentan diferentes conjuntos de medidas con la finalidad de cuantificar objetivamente la calidad de voz. Todos ellos tienen en común la necesidad de realizar diferentes medidas de la voz con la finalidad de recoger en el proceso de medida los distintos aspectos de las características acústicas de la voz.

Con el fin de suplir este adiestramiento y dar la posibilidad de realizar evaluaciones de este tipo a cualquier profesional no especializado, se han desarrollado aplicaciones que permiten cuantificar la calidad de la voz por medio de técnicas acústicas. Estas técnicas incluso permiten realizar una estimación del carácter de voz sana o el carácter de voz patológica a partir de una voz previamente registrada por medio de un micrófono.

1.2 Motivación

Las técnicas basadas en métodos de inspección visual directa son las técnicas que permiten obtener la mejor valoración del estado del sistema fonador de un locutor aunque sin embargo esta técnica presenta ciertos inconvenientes como el hecho de tratarse de técnicas invasivas, con la correspondiente incomodidad para el locutor que se preste a la evaluación, o el hecho de que este tipo de pruebas han de ser realizadas por profesionales del ámbito médico especializado.

Sin embargo en determinadas ocasiones no se dispone de los medios humanos y materiales adecuados, y resulta de interés el disponer de técnicas de evaluación de la calidad de la voz no invasivas y que resulten cómodas al locutor, por ejemplo para realizar el primer *screening* en centros de atención primaria, o bien para documentar de forma objetiva la evolución de un determinado tratamiento de un paciente, para documentar la evolución de una intervención quirúrgica o incluso en situaciones en las que el locutor no se encuentra presente como es el caso de la telemedicina.

Frente a esta situación resulta interesante evaluar la calidad de la voz por medio de técnicas basadas en la inspección acústica. Sin embargo los diferentes procedimientos de evaluación acústica basados en la percepción de la calidad de la voz presentan el inconveniente de su naturaleza subjetiva y por tanto, no garantizan la unicidad de las valoraciones obtenidas de diferentes evaluadores.

Surge por tanto, como medio de objetivar estas medidas de valoración acústicas, las distintas estrategias que se basan en la estimación de medidas objetivas que cuantifican numéricamente y de forma automática la calidad de la voz a partir de una grabación de audio. La situación actual de estas técnicas presentan el inconveniente de no existir un protocolo claro, ya que los sistemas comerciales actuales basan la cuantificación de la calidad de la voz en hasta 33 medidas diferentes, en las que es difícil extraer información útil en lo referente a una interpretación física de los valores numéricos.

1.3 Objetivos

El objetivo general de esta tesis es establecer una metodología para la evaluación acústica de la calidad clínica de la voz a partir de una grabación de audio.

Este objetivo general se desglosa en los siguientes objetivos concretos:

1. Identificar las diferentes características del comportamiento laríngeo susceptibles de ser medidas en la señal del habla.
2. Establecer relaciones entre las medidas objetivas de calidad de la voz más conocidas y estudiadas por la comunidad científica con los fenómenos físicos que las origina.
3. Justificar fenómenos físicos que caracterizan la calidad de la voz y que no son tenidos en cuenta por las anteriores medidas: comportamientos no lineales de la voz.
4. Validar la utilización de medidas de calidad de la voz que cuantifican su comportamiento no lineal.
5. Selección de las anteriores medidas de tal forma que la valoración de la calidad de la voz por medio de las medidas propuestas permita una clara cuantificación de los fenómenos físicos que intervienen en la producción de la calidad de la voz.

1.4 Metodología

Hasta ahora en la bibliografía actual nos encontramos diferentes escenarios en el estudio de las medidas de calidad de la voz. Por un lado nos encontramos los resultados obtenidos a partir de bases de datos de voces de diferentes calidades donde el número de muestras es reducido, en los que se presentan los resultados obtenidos para el estudio de una determinada medida o de un subconjunto de medidas. Por otro lado existen estudios donde, a partir de una base de datos de voces de diferentes calidades con un número significativo de muestras se realiza el estudio de conjuntos de medias e incluso se aplican técnicas de selección de características con el fin de cuantificar la relevancia de las diferentes medidas de calidad de la voz. Existen diferentes aspectos en la metodología de estos estudios que no permiten realizar una generalización de los resultados obtenidos.

Actualmente existe una única base de datos estándar de voces con diferentes calidades de voz, Kay Elemetrics Corporation [5], que permite realizar estudios que atribuyen a los resultados cierta consistencia. Sin embargo, dicha base de datos no resulta totalmente representativa debido a que el número de muestras de voces no es suficientemente alto. Han existido diferentes experiencias para crear una gran base de datos, con suficientes muestras de voz con distintas calidades de voz y diferentes locuciones, [7] es un buen ejemplo de ellas. Sin embargo, la creación de una gran base de datos con la suficiente diversidad de muestras es una tarea altamente difícil de llevar a cabo debido a la dificultad de localizar a los locutores con diferentes calidades de voz y a la necesidad de la certificación de la calidad de la voz por medio de un especialista clínico de la voz.

Debido principalmente a la inexistencia de una gran base de datos resulta necesario demostrar la independencia de los resultados de la cuantificación de la relevancia de las medidas de calidad de la voz con la base de datos.

Otro aspecto a tener en cuenta, es la gran cantidad de estudios en los que no se tiene en cuenta en el proceso de selección de medidas de la calidad de la voz todo el

conocimiento existente en el ámbito clínico respecto a la naturaleza de los trastornos del sistema fonador que producen una alteración de la calidad de la voz.

Por último, en gran cantidad de trabajos se utilizan sistemas de clasificación, bien con el fin de implementar sistemas de clasificación automático o con el fin de demostrar la validez de la medida o conjunto de medidas para diferenciar diferentes calidades de voz. Sin embargo en pocos estudios se busca dotar a los entornos clínicos de herramientas útiles para la cuantificación de la calidad de la voz a partir de los rangos de valores tomados por las medidas de calidad, máxime teniendo en cuenta que para la evaluación de la calidad de la voz se proponen conjuntos de medidas, en algunos casos de sistemas comerciales hasta 22 medidas [6], o medidas vectoriales difícilmente interpretables por profesionales clínicos de la voz. No existen trabajos en los que se asocie el valor numérico de una determinada medida de la calidad de voz con la cuantificación del comportamiento de un determinado mecanismo del sistema de producción de voz que permita un mejor uso clínico de las medidas de calidad de la voz obtenidas a partir de una grabación de la voz.

Para la consecución de los objetivos establecidos se precisó de un estudio detallado del funcionamiento del sistema fonador y los diferentes aspectos que intervienen en la degradación de la calidad de la voz, a la vez que una búsqueda y asimilación del estado de la técnica en lo referente a las medidas de calidad de la voz que se obtienen a partir de grabaciones de voz.

Una vez estudiadas las diferentes medidas de calidad de la voz, cada una de ellas fue implementada. Para este fin, fue necesario disponer de una gran variedad de muestras de voces de diferentes calidades. La estrategia que se siguió consistió en utilizar una colección de voces de locutores los cuales presentan diferentes patologías en el sistema fonador y voces de locutores sanos. Dentro de las muestras de voces de locutores patológicos se dispuso de muestras con distintos grados de afección consiguiendo disponer de un variado rango de calidades de voz. Para el estudio se dispuso de dos bases de datos.

Una de las bases de datos de voces fue grabada íntegramente para el desarrollo de este trabajo, gracias a la colaboración de los miembros del Servicio de

Otorrinolaringología del Hospital Doctor Negrín de Gran Canaria. La segunda base de datos de voces de diferentes calidades es de carácter comercial de aproximadamente 700 locutores.

Posteriormente fue estudiado el estado de la técnica en lo referente a métodos de selección de atributos, implementándolos e identificando el mejor método atendiendo a la efectividad en la selección y ordenación de los atributos en función de su relevancia y su coste computacional.

Posteriormente se estudió la posibilidad de utilizar la información sobre el comportamiento del sistema fonador combinado con técnicas heurísticas de selección de atributos, con el fin de identificar el subconjunto óptimo de medidas que mejor cuantifica la calidad de la voz. Además se buscó que las diferentes medidas de calidad fuesen interpretables, permitiendo asociar el valor de cualquiera de las medidas seleccionadas con un determinado fenómeno físico asociado al sistema de producción de voz.

A partir del procedimiento de selección de características desarrollado, se buscó el subconjunto óptimo de características siendo evaluado en diferentes sistemas de clasificación. Para este proceso se utilizó sólo una de las bases de datos. Con esta estrategia se consiguió encontrar el mejor subconjunto de características para la base de datos utilizada.

Con el fin de demostrar la independencia de la selección de características con la base de datos, se evaluó el subconjunto de características óptimo utilizando la segunda base de datos. Tras evaluar este nuevo sistema con diferentes clasificadores se estudió la similitud de los resultados obtenidos utilizando una y otra base de datos.

Se realizaron diferentes evaluaciones del subconjunto óptimo de características por medio de la implementación de varias aplicaciones. Las aplicaciones implementadas consistieron en un estimador automático de la presencia de patologías en el sistema fonador a partir de la fonación de las cinco vocales castellanas, un estimador automático del grado de disfonía igualmente a partir de la fonación de las cinco vocales castellanas, un sistema automático para estimar la presencia de patologías en el sistema

fonador a partir del habla continua adquirido a partir de un canal telefónico y un estudio estadístico sobre los rangos de valor para las diferentes medidas que cuantifican los distintos fenómenos físicos implicados en la caracterización de la calidad de la voz.

1.5 Resultados

La Tesis presenta las siguientes aportaciones:

1. Se presenta un estudio sobre el estado de la técnica en lo que respecta a medidas de la calidad de la voz a partir de medidas objetivas obtenidas de una grabación de voz.
2. En el estudio se identifican los diferentes fenómenos físicos que caracterizan la calidad de la voz desde la perspectiva de diferentes dominios de representación de la señal de voz, introduciendo de forma novedosa la caracterización de la calidad de la voz atendiendo a la presencia de alinealidades. El estudio continúa identificando y describiendo las distintas medidas de la calidad de la voz existentes en la bibliografía actual, realizándose un análisis crítico y aportando nuevas medidas de calidad de la voz.
3. Se describe el comportamiento de las cuerdas vocales por medio de cuatro características las cuales tienen asociadas sus respectivas manifestaciones físicas en la señal de voz: estabilidad de la voz, riqueza espectral, presencia de ruido y comportamiento no lineal. Las diferentes medidas de la calidad de la voz fueron clasificadas en los cuatro grupos, atendiendo a los fenómenos físicos que cuantifican, es decir atendido a cual de las cuatro características del funcionamiento de las cuerdas vocales caracteriza.
4. Se ordenaron las diferentes medidas de calidad de la voz correspondientes a cada una de las cuatro características del funcionamiento de las cuerdas vocales en función de su relevancia en la cuantificación de la calidad de la voz, siendo evaluada la relevancia en función de éxito obtenido en la clasificación automática.
5. Finalmente, se propone un protocolo de evaluación de la calidad de la voz a partir de medidas objetivas obtenidas a partir de una grabación de voz.

Como aplicación del trabajo se presentan diferentes aplicaciones: un estimador automático de la presencia de patologías en el sistema fonador a partir de la fonación de las cinco vocales castellanas, un estimador automático del grado de disfonía igualmente

a partir de la fonación de las cinco vocales castellanas, un sistema automático para estimar la presencia de patologías en el sistema fonador a partir del habla continua, adquirida a partir de un canal telefónico y un estudio estadístico sobre los rangos de valor para las diferentes medidas que cuantifican los distintos fenómenos físicos implicados en la caracterización de la calidad de la voz. Estas aplicaciones pretenden demostrar la valía de los resultados obtenidos en esta tesis. Debido a que en la etapa de clasificación de las diferentes aplicaciones se usa el protocolo de evaluación de la calidad de la voz, caracterizado por usar muy pocas medidas donde el valor de cada una de ellas puede asociarse a un determinado comportamiento del sistema fonador, y ya que los resultados son aceptables, teniendo en cuenta que se trata de un sistema de reconocimiento autónomo, estas aplicaciones demuestran la valía del protocolo de evaluación.

1.6 Estructura de la Memoria

La memoria de este trabajo está constituida por once capítulos y un anexo, los cuales se describen a continuación.

Tras el presente capítulo introductorio, en el capítulo dos, se describen las diferentes causas de la pérdida de calidad de la voz, centrándose en las patologías vocales. Este conocimiento sobre las causas de la pérdida de calidad vocal permitirá comprender los distintos fenómenos físicos que caracterizan la calidad de la voz.

En el capítulo tres, se describen las características de los diferentes fenómenos físicos que intervienen en la calidad y que se manifiestan desde diferentes dominios de representación de la señal de voz, los cuales son cuantificables por medio de diferentes medidas de la calidad de la voz. A modo de análisis, se realiza una síntesis de los diferentes fenómenos que intervienen en la alteración de la calidad de la voz, identificando cuatro grupos principales: estabilidad de la voz, riqueza espectral, presencia de ruido y comportamiento no lineal.

En el capítulo cuatro, se hace una revisión de las diferentes estrategias utilizadas a la hora de crear bases de datos para el estudio de medidas de calidad de la voz. Además, se realiza un análisis crítico de las diferentes estrategias. Este capítulo concluye con la presentación de las dos bases de datos de voces que se han utilizado en este trabajo para validar y justificar los resultados obtenidos.

En el capítulo cinco se presenta, desde la perspectiva de la minería de datos, el estado actual de los diferentes procedimientos de selección de características. Además, se presenta una estrategia para la selección de medidas de calidad de la voz en el que se combinan los métodos ciegos de selección de características y el conocimiento clínico sobre el fenómeno físico que cuantifica cada uno de las distintas medidas. En el procedimiento de selección se hace uso de la clasificación propuesta de los diferentes fenómenos físicos implicados en las medidas de la calidad de la voz.

En los capítulos seis, siete, ocho y nueve se plantea un estudio detallado sobre los distintos fenómenos físicos que cuantifican la calidad de la voz, correspondiendo cada capítulo al estudio de cada uno de los diferentes fenómenos físicos implicados: estabilidad de la voz, riqueza espectral, presencia de ruido y comportamiento no lineal, respectivamente. Para cada capítulo se presenta una descripción de las diferentes manifestaciones en los distintos dominios del fenómeno físico en estudio, la identificación de las distintas medidas de calidad de la voz asociadas al fenómeno y un análisis de los resultados de la ordenación de dichas medidas en función de su relevancia en la cuantificación de la calidad de la voz.

En el capítulo diez se presenta diferentes aplicaciones en las que se han evaluado las medidas de calidad de la voz propuestas. A modo de aplicaciones se presenta un estimador automático de la presencia de patologías en el sistema fonador a partir de la fonación de las cinco vocales castellanas, un estimador automático del grado de disfonía igualmente a partir de la fonación de las cinco vocales castellanas, un sistema automático para estimar la presencia de patologías en el sistema fonador a partir del habla continua adquirida a partir de un canal telefónico y un estudio estadístico sobre los rangos de valor para las diferentes medidas que cuantifican los distintos fenómenos físicos implicados en la caracterización de la calidad de la voz.

En el capítulo once se presentan las conclusiones obtenidas en este trabajo, realizándose un análisis de los resultados concluyendo con la propuesta de un protocolo de evaluación de la calidad de la voz para uso clínico a partir de medidas objetivas de calidad de la voz obtenidas a partir de una grabación.

Por último, en la memoria de este trabajo existe un anexo en el que se presentan las tablas de resultados obtenidos en los diferentes estudios y que han sido fruto de análisis en los capítulos anteriores.

1.7 Referencias

- [1] Hirano, M., *Clinical Examination of Voice*. New York, Springer-Verlag, 1981.
- [2] Dejonckere, P. H. Remacle, M. Fresnel-Elbaz, E. Woisard, V. Crevier-Buchman, L. Millet, B., "Differentiated perceptual evaluation of pathological voice quality: reliability and correlations with acoustic measurements". *Revue de Laryngologie Otologie Rhinologie*, vol. 117, no.2, pp. 219-224, (Bordeaux) 1996.
- [3] Hammarberg, B. y Gauffin, J., Perceptual and acoustic characteristics of quality differences in pathological voices as related to physiological aspects,. in O. Fujimura & M.Hirano (eds.), *Vocal Fold Physiology*, pp 283-303, 1995
- [4] Laver, J., *The Gift of Speech*, Edinburgh University Press, 1991.
- [5] Kay Elemetrics Corporation. *Disorder Voice Database Model 4337*. Massachusetts Eye and Ear Infirmary Voice and Speech Lab, Boston, MA. 1994.
- [6] Kay Elemetrics Corporation. *Multi-Dimensional Voice Program (MDVP)*, , [http://www.kayelemetrics.com], 2006
- [7] Pedro Gomez Vilda, Modelado y Clasificación Automática de Patrones de Voz patológica para su aplicación clínica sobre Internet, Proyecto CICYT: TIC2002-02273, 2002-2005.

CAPÍTULO 2

Fisiología de la Laringe: Trastornos de la Fonación

2.1 Función Fonatoria

Se interpreta la fonación como un proceso regulado por el sistema nervioso central y periférico que coordina la actividad motora y el acoplamiento funcional del resto de los elementos que participan en la generación de la voz y en su percepción, tanto a los niveles más bajo de comportamiento automático como al de aquellos en los que la participación de los centros superiores significan actividades fonatorias de mayor entidad, como son la articulación, el canto o la expresión de las emociones y la personalidad individual.

2.2 Producción de la Voz

Para producir la voz, el primer órgano implicado es el aparato respiratorio. Primero se inspira el aire, aumentando la presión dentro de los pulmones. Posteriormente se espira el aire, saliendo éste por la tráquea hacia la laringe.

Fisiología de la Laringe

La laringe está situada en el cuello, delante de la faringe, debajo de la base de la lengua y encima de la tráquea. La laringe está formada por cartílagos, que se unen por medio de ligamentos y que se mueven gracias a la acción de varios músculos. Uno de los cartílagos es la epiglotis, que actúa como una válvula deslizándose hacia abajo para cubrir la glotis, y cuya función consiste en impedir que los alimentos penetren por la tráquea o la laringe durante la deglución. La actividad de la epiglotis está controlada de manera automática por el sistema nervioso y, cuando falla, líquidos o partículas de alimentos pasan a la vía aérea; estas partículas o gotas se expulsarán hacia el exterior mediante un acceso de tos, siempre y cuando no sean demasiado grandes o bien queden impactadas en alguna de las vías aéreas por debajo de la laringe.

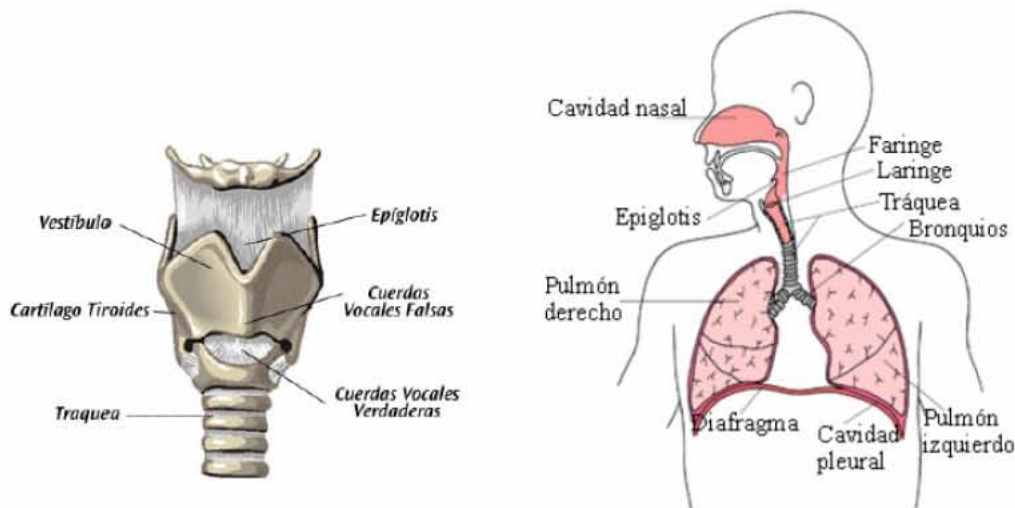


Ilustración 2-1 : Fisiología de la laringe

En el interior de la laringe se encuentran las cuerdas vocales que son unas estructuras músculo-membranosas alargadas constituidas por el propio tejido de la laringe. La vibración de las cuerdas vocales produce la voz. Cualquier enfermedad que dañe la laringe también afectará la capacidad de hablar.

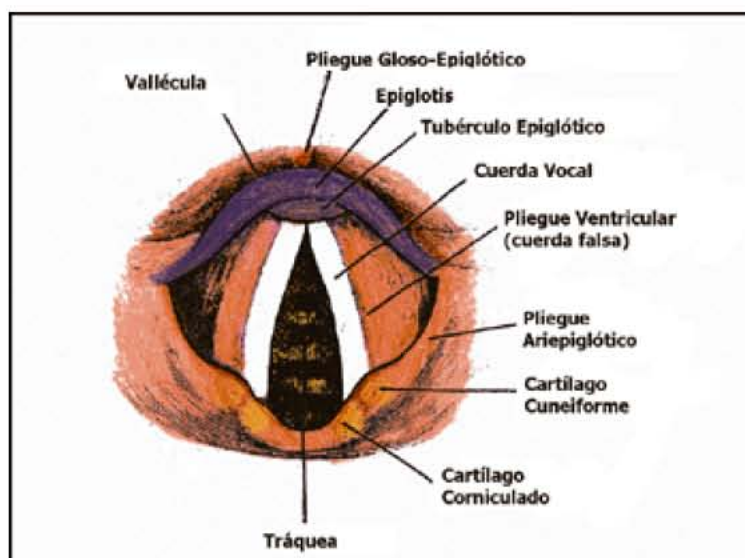


Ilustración 2-2 : Fisiología de las cuerdas vocales

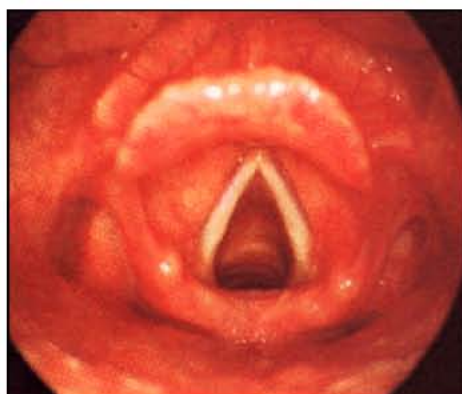


Ilustración 2-3: Imagen de las cuerdas vocales

El Fenómeno Fonatorio

Cuando el aire es expulsado de los pulmones, pasa entre las cuerdas vocales cerradas y las hace vibrar, produciendo sonidos que las cavidades de resonancia modulan en sonidos vocativos. A su vez, estos sonidos vocativos, mediante la articulación pueden dar lugar a diferentes unidades fonéticas como determinadas consonantes, fonemas, o palabras. Las cuerdas vocales están ancladas sobre piezas cartilaginosas de tal modo que, al moverse, pueden tensarlas o distenderlas, lo que producirá sonidos agudos o graves.

Actividad respiratoria. La fonación lleva aparejada la adopción de un ritmo respiratorio especial, fundamentalmente distinto del propio de la respiración tranquila. En la

fonación, el ritmo respiratorio es cadencioso y la duración de cada ciclo varía poco de un ciclo a otro. Las dos fases respiratorias tienen una duración similar, siendo tan sólo la espiración algo más prolongada que la inspiración. En la fonación, sin embargo, el ritmo respiratorio pierde esta regularidad porque la inspiración se acorta de manera considerable y adquiere el significado de impulso del gesto fonatorio.

La espiración, convertida en soplo fonatorio, se prolonga variablemente, entrecortada por pausas de bloqueo laríngeo y que corresponden a las naturales fluctuaciones que señalan la cadencia del habla espontánea.

En la respiración tranquila, sólo la fase inspiratoria precisa la intervención de una actividad muscular, o lo que es lo mismo, únicamente es activa la fase inspiratoria, ocurriendo de forma totalmente pasiva la espiración. Esta actividad inspiratoria corresponde a los músculos elevadores del tórax (escalenos), expansores del tórax (intercostales) y sobre todo al diafragma. En la fonación, sin embargo, la espiración pasa a ser activa, siendo realizada principalmente por los músculos intercostales internos y los músculos abdominales, si bien también interviene el diafragma.

Cavidades Resonantes. La calidad de los sonidos producidos por el sistema fonatorio, está influida por la posición de la lengua, de los labios y de la mandíbula que, en conjunto, actúan modulando la voz y articulando las palabras. Las cavidades de resonancia son indispensables para la emisión de la voz. Estas cavidades empiezan en la región supraglótica así como en la parte baja de la faringe (hipofaringe). La orofaringe, la cavidad bucal y la rinofaringe son las cavidades más importantes.

Las cuerdas vocales. Antes de producirse el sonido en las cuerdas vocales, se deben establecer diversas condiciones: primeramente las cuerdas vocales deben aproximarse al menos, hasta la posición fonatoria y posteriormente es necesario que las cuerdas se tensen y elonguen adecuadamente de forma previa a la producción de cualquier sonido. Finalmente, debe haber suficiente flujo aéreo desde los pulmones. Una vez establecidas estas condiciones puede comenzar la fonación sucediéndose una serie alternativa de cierres y aperturas de las cuerdas vocales. Así, partiendo de una abducción completa, la presión subglótica separa las dos cuerdas vocales hasta un punto en que su grado de elasticidad limita su excursión hacia fuera. A partir de este punto, la elasticidad actúa

invirtiendo el sentido del desplazamiento hacia la línea media, hasta recuperar la posición inicial, lo que lleva consigo la reducción del espacio entre ambas cuerdas vocales, el progresivo incremento de la presión subglótica, el aumento de la velocidad de flujo que reduce la presión intercordal, la oclusión completa y el inicio de un nuevo ciclo. De esta manera, cada ciclo vibratorio estaría formado por una fase abierta y una fase cerrada, y a su vez, la fase abierta estaría dividida en una fase de apertura, durante la cual las cuerdas vocales se separan alejándose de la línea media, y una fase de cierre, durante la cual se aproximan hasta su total contacto.

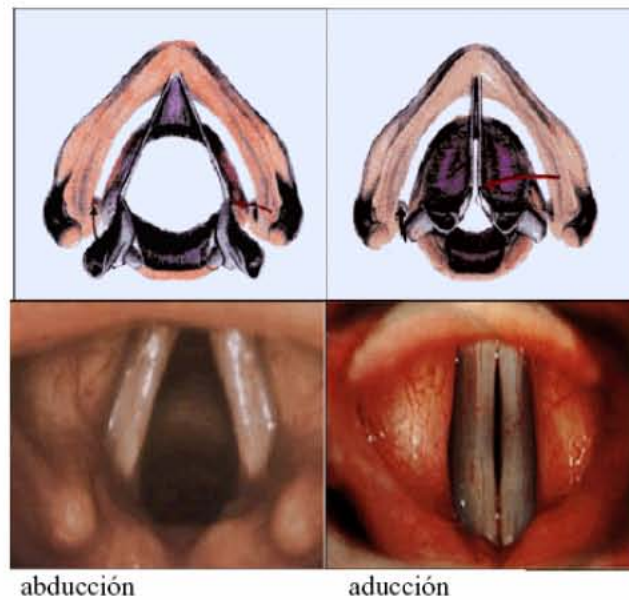


Ilustración 2-4 : Fase de abducción y fase de aducción de las cuerdas vocales

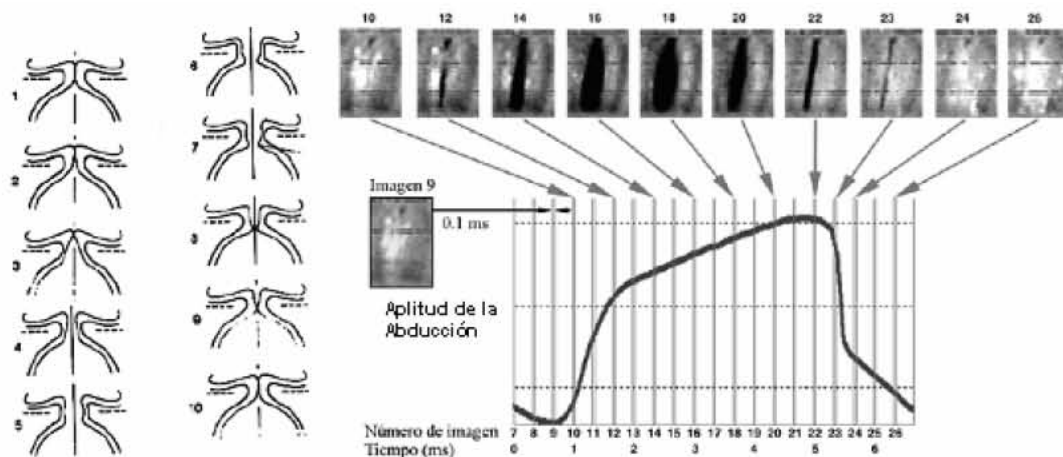


Ilustración 2-5 : Diferentes fases del movimiento de las cuerdas vocales

2.2.1 Teorías sobre la Fonación

La investigación de la función fonatoria de la laringe comenzó hace más de 300 años. Así ya Dodart en 1700 expone su teoría sobre la fonación, relacionando el paso de aire a través de la glotis como causa de la fonación. Posteriormente Ferrein en 1741, realiza los primeros experimentos sobre laringes escindidas de perros, presentando sus resultados en la Real Academia francesa de las Ciencias, donde expone la producción del sonido al insuflar aire a través de la tráquea y acuña el término de cuerdas vocales, vigente hasta la actualidad, por entender el comportamiento de los repliegues de la glotis, semejante al de un instrumento de cuerda. Un gran avance experimental se produce con los trabajos de Müller en 1843, realizados sobre cadáveres y laringes artificiales, asentando las bases para trabajos y teorías posteriores como la mioelástica.

Así llegamos a mitad del siglo pasado donde terminan de sentarse las bases para el conocimiento actual de la fisiología fonatoria, proponiendo Husson [3] la teoría Neurocronáxica, que es contestada posteriormente por Van der Berg [4] con la teoría Mioelástica-Aerodinámica, la cual es a su vez matizada por otros autores, como Perelló [5], quien formula la teoría Muco-Ondulatoria, vigente hasta la actualidad.

Husson, en su teoría Neurocronáxica, expone que la laringe es una “sirena de activación periódica” en la que los pliegues vocales poseen una función activa, de manera que la frecuencia de los impulsos motores procedentes del nervio recurrente condiciona la frecuencia de su vibración. De esta manera, el mecanismo regulador de la altura de los sonidos sería independiente del mecanismo que regula la intensidad los mismos. Los pliegues vocales, en esta teoría, se separan por influjo de un potencial de acción recurrencial que condiciona el escape de un poco de aire a través de la glotis y esto se traduciría en un mecanismo “impulso a impulso” recurrencial. Dado que un nervio no puede transportar más de 300 potenciales de acción por segundo, debido a la existencia de una fase refractaria, resulta difícil explicar la producción de sonidos por encima de esta frecuencia. Para ello Husson recurre a un fenómeno de bifascicidad ó trifascicidad, según el cual las fibras nerviosas estarían divididas en equivalentes grupos y actuarían alternativamente. Esta teoría no contempla ninguna función en la mucosa de los pliegues vocales.

Para Van den Berg, existe una unidad funcional compuesta por los pulmones y tráquea, la laringe como generador de sonidos y las cavidades vocales que forman el sistema resonador. Así en la teoría Mioelástica-Aerodinámica se postula que las cuerdas vocales actúan por el flujo de aire liberado por los pulmones, a través de la tráquea y la frecuencia fundamental de la vibración depende de la masa efectiva y de la flexibilidad de la cuerda vocal, concretamente de la porción vibrátil. Otro factor fundamental sería el valor de la presión subglótica y el área glótica efectiva durante el ciclo, lo que determina la resistencia efectiva de la glotis y a su vez condiciona la aparición del efecto Bernoulli a nivel glótico. Este efecto, produciría un fenómeno de succión a nivel glótico, como consecuencia del flujo aéreo fonatorio que provoca un descenso de la presión subglótica, lo que a su vez conlleva un nuevo cierre glótico y el aumento subsiguiente de la presión subglótica, hasta el siguiente escape aéreo.

Esta teoría cobra gran aceptación en detrimento de la Neurocronámica, pues da explicación a grandes lagunas creadas por esta última, como son la diferente longitud del nervio recurrente derecho e izquierdo para una misma velocidad de conducción, con el consiguiente desfase en la llegada del impulso a cada músculo vocal; también resulta difícilmente explicable la necesidad de un mínimo flujo aéreo para producir la vibración cordal, o la influencia de la presión subglótica en el tono.

Sin embargo, no todos los problemas quedan resueltos con esta teoría Mioelástica-Aerodinámica, pues la realidad es que la vibración del músculo vocal no puede ser objetivada. Sobre esta base, y a través de su experiencia en el uso de la estroboscopia, Perelló formula una teoría complementaria, pero extremadamente clarificadora, basada en el papel de la mucosa que recubre las cuerdas vocales. Este papel ya había sido mencionado por otros autores como Smith, pero sin darle el carácter pasivo que propone Perelló, sino como parte vibrante de una estructura segmentaria.

En la teoría muco-ondulatoria se considera la vibración de las cuerdas vocales, como una ondulación de la mucosa que las recubre. Las ondas durante la fonación se dirigen de abajo hacia arriba y de adelante a atrás, de tal manera que se propagan desde la subglotis, a través del borde de la cuerda continuando sobre su superficie y perdiéndose lentamente a la entrada del ventrículo; según esta onda desaparece, va apareciendo otra en la subglotis. Así, la succión causada por el efecto Bernoulli

anteriormente descrito, se efectuaría sobre una mucosa que descansa sobre una submucosa muy laxa.

De esta forma se produce una conjunción entre las teorías propuestas por Van Der Berg y Perelló, las cuales vienen a quedar definitivamente avaladas por los hallazgos histológicos sobre la estructura de la cuerda vocal, realizados por M. Hirano. Hirano [6] demuestra la existencia de dos estructuras funcionalmente diferentes; de un lado, estaría el músculo vocal, íntimamente adherido al ligamento vocal, y de otro, la membrana mucosa que se encuentra unida débilmente a este último y por tanto, con capacidad para moverse de diferente manera durante la vibración cordal.

2.2.2 *Acústica Vocal*

Se puede definir el sonido como la vibración mecánica de una masa o sistema de masas, capaz de producir una sensación auditiva. Considerando la vibración como un movimiento recurrente, al número de vibraciones completas por segundo se le denomina frecuencia y se mide en ciclos por segundos ó Herzios. Así el espectro de frecuencias audibles, se sitúa entre los 16 a 20.000 Hz. A la desviación máxima de este movimiento se le denomina amplitud y a la duración de una vibración completa se le denomina periodo.

Los movimientos recurrentes o cíclicos pueden dividirse en simples o complejos. Los simples producen el llamado movimiento armónico simple, que constituye la base o generador de otros sonidos. Son muy pocos los sonidos simples que existen en la naturaleza y están asociados a curvas sinusoidales. Toda vibración que no es simple es compleja, como ocurre con todos los sonidos del habla.

Así según el Teorema de Fourier, se establece que toda vibración compleja, puede descomponerse en una serie de movimientos armónicos simples, que en tal caso, se denominan armónicas de la vibración compleja. Si la vibración compleja es periódica, la primera armónica denominada fundamental, posee una frecuencia f_0 igual a la frecuencia F_0 de la onda compleja. En caso de que el sonido complejo sea un

fonema sonoro, la frecuencia glotal f_g es semejante a las anteriores y las siguientes armónicas f_n , que poseen frecuencias múltiplos de la primera. De esta manera:

$$f_0 = F_o = f_g \quad \text{y} \quad f_n = n \cdot f_0$$

La producción acústica del habla, se basa esencialmente en el fenómeno de la resonancia, entendiéndose por tal, la amplificación de un sonido de determinada frecuencia, por la acción de un cuerpo pasivo denominado resonador o filtro. El resonador, debido a sus características físicas tales como la dimensión, forma y grado de rigidez, posee una frecuencia natural de vibración, de tal manera que si un sonido proveniente de la fuente de energía posee la misma frecuencia, entra en resonancia, amplificándose el sonido. Sonidos de frecuencias mayores o menores a la del resonador, también son amplificados en menor grado y eventualmente atenuados. Así, el tracto vocal actúa como resonador del sonido glotal, pero los espectros sonoros para ser útiles en la comunicación oral, precisan ser modificados mediante variaciones en las cavidades de resonancia, de tal manera que se amplifican o se amortiguan determinadas frecuencias de la fuente de energía sonora.

Definimos así los formantes, como una zona del espectro vocal donde la función de transferencia del tracto vocal es máxima, o lo que es lo mismo, existen grupos de armónicos que presentan máximos relativos de intensidad. Se trata, por tanto, de frecuencias simples en las cuales la función de transmisión del tracto vocal, es más eficiente que en frecuencias vecinas. Dado que el tracto vocal actúa como un resonador tubular, su forma y tamaño condiciona la aparición de los formantes y por tanto, no es de esperar que sean exactamente iguales de un interlocutor a otro, debido a las variaciones anatómicas.

Por lo general, los cuatro primeros formantes se encuentran cerca de los 500, 1500, 2500 y 3500 Hz, siendo modificados por las diferentes posiciones de los articuladores, lo que condiciona la diversidad de los sonidos vocálicos. Así, el primer formante se modificaría con la apertura de la mandíbula y el segundo y tercer formante, con la forma del cuerpo de la lengua y con la posición de la punta de la lengua respectivamente.

2.2.3 Características de la Voz

Las cualidades acústicas de la voz son: *intensidad*, *tono*, *extensión* y *timbre*.

Intensidad: La intensidad de la voz, traduce en amplitud de la variación cíclica, el nivel de presión sonora correspondiente a dicho sonido. Ya desde los trabajos de Magendie y Luchsinger, se relaciona este parámetro con el flujo aéreo a través de la glotis y con el grado de contracción de la misma o resistencia glótica. Issihiki estudia la relación entre la intensidad vocal, la presión subglótica, el flujo aéreo y la resistencia glotal, encontrando en sus trabajos que en las frecuencias bajas de la fonación, el flujo aéreo permanece prácticamente invariable, o incluso, desciende ligeramente con los incrementos de la intensidad vocal, mientras que la resistencia glotal muestra una tendencia a aumentar. Por otra parte, el flujo aéreo en las altas frecuencias aumenta considerablemente, mientras que la resistencia glotal permanece prácticamente invariable, con los incrementos de intensidad. Por tanto, según estos resultados, las variaciones de intensidad en las frecuencias graves dependerían fundamentalmente de un control laríngeo y en las frecuencias agudas, de la actividad de los músculos respiratorios.

Tono: Viene determinado por la frecuencia de vibración glotal, que se caracteriza por ser una vibración compleja periódica y por tanto, según el Teorema de Fourier, está formada por una frecuencia fundamental y un conjunto de frecuencias múltiplos de ésta, denominadas armónicos. La frecuencia fundamental viene determinada por el número de vibraciones glotales completas por segundo.

Extensión: Al conjunto de frecuencias capaz de ser emitidas por una laringe, se le denomina extensión de la voz. En condiciones normales, la extensión de la voz es algo superior a las tres octavas, concretamente, 38 semitonos para los hombres y 37 para las mujeres, con un rango de frecuencias que oscilan entre los 78-698 Hz para las voces masculinas y de 139-1108 Hz para las femeninas.

Timbre: Es una cualidad fundamental del sonido vocálico, que está relacionada con la distribución acústica de energía en el espectro sonoro, es decir, el número y la intensidad relativa de los armónicos. El timbre vocal resulta por tanto del

enriquecimiento sonoro del sonido laríngeo por las cavidades de resonancia y es lo que nos permite diferenciar entre dos fonaciones de igual frecuencia fundamental e intensidad.

La laringe es un órgano cuyas características varían con cada individuo. La estructura laríngea y las características de la voz son un fiel reflejo de la edad del individuo, del sexo y del estado de salud. Esta maduración se prolonga a lo largo de los distintos períodos vitales, determinando así modificaciones estructurales y fónicas notables.

2.2.4 Evolución de la Voz

La evolución de la voz se puede clasificar en seis etapas: neonatal, primera infancia, segunda infancia, pubertad, estabilización y senectud. A continuación se explican las características de cada una.

Neonatal. Esta etapa va del nacimiento hasta los treinta o cuarenta días de vida. Las modificaciones fónicas (llanto y grito) se caracterizan por altas frecuencias. El ataque del sonido es brusco; de fuerte intensidad y modulación muy reducida: 784 Hz; el grito puede llegar a 1.318 Hz.

Primera infancia. Va desde el primer mes de vida hasta los seis años. Alrededor del primer mes de vida comienza a modificarse la voz. El ataque se hace menos brusco, coincidiendo con la instalación paulatina de la capacidad funcional hormonal propia del niño. A los dieciocho meses aparece la modulación vocal 523-784 Hz.

Segunda infancia. Desde los seis años hasta la iniciación de la pubertad. En esta etapa predomina la acción de la somatotrofina y de la tiroxina, las variaciones vocales llegan hasta una octava y media de extensión.

Pubertad. En este período se establecen las características de diferenciación sexual secundaria, siendo más notable en el varón que en la niña. Esta etapa se caracteriza por la acción de la estimulación hormonal hipofisaria que determina una intensa activación de las gónadas. Éstas comienzan a elaborar los esteroides sexuales que condicionan las

características morfológicas secundarias propias de cada sexo y sus capacidades funcionales definitivas. La mutación vocal se produce en el varón entre los trece y catorce años, y en las mujeres entre los catorce y quince años en los climas cálidos. Al cumplirse el descenso laríngeo se hace notable la disminución de los sonidos producidos. La muda fisiológica de la voz se caracteriza por signos evidentes, donde las dualidades fónicas son comunes. Es una voz que pasa del grave al agudo, produciendo los llamados “gallos”; hay disminución de la intensidad de la voz y a veces se instalan disfonías francas. En la mujer la laringe se desarrolla en altura sin alargamiento transversal y sagital.

Estabilización. En esta etapa la voz es estable y tiene las características propias de cada sexo.

Senectud. La sensibilidad vocal es más precoz en la mujer que en el hombre y se presenta más marcado en la voz cantada que en la hablada. Se produce una pérdida de los agudos, disminución de la extensión, pérdida de potencia y disminución de los armónicos. Este proceso se traduce a nivel laríngeo como pérdida de la tonicidad cordal (fonastenias). En la mujer, durante la menopausia, se agrava el tono de voz.

2.3 Trastornos de la Fonación: Concepto de Disfonía

Etimológicamente el concepto de disfonía proviene del griego y quiere decir “voz difícil”, y aunque en la actualidad es un término con múltiples definiciones, se acepta aquella que considera la disfonía como la alteración de una o más de las características acústicas de la voz, generalmente el timbre, pero también la intensidad o el tono. Incluso hoy tiende a aceptarse como disfonía la dificultad subjetiva para emitir la voz, aunque no se acompañe de alteraciones acústicas.

Más difícil aún resulta la definición de disfonía funcional, o para algunos autores como Le Huche [7], disfonía disfuncional, por cuanto la pretensión de separar trastornos puramente funcionales de los orgánicos, resulta enormemente dificultosa, ya que un trastorno funcional puede seguirse a continuación de una lesión orgánica, o viceversa. En cualquier caso, dada su enorme ventaja práctica, se mantiene esta

nomenclatura ayudándose de términos como “con lesión”, o “complejas”, para los casos con lesiones orgánicas asociadas a trastornos funcionales subyacentes.

Así, ya desde 1887, Frankel definió el concepto *mogifonía*, que describió como una ronquera sin lesión visible, es decir una alteración en el timbre de la voz en personas que “curiosamente” no presentaban alteración alguna en la exploración Laringoscópica. Posteriormente, Flateau y Gutzman describieron en 1906 el mismo trastorno explicándolo a través de una fatigabilidad anormal del órgano vocal, de donde viene el término de *Fonastenia* que los autores aplicaron a este trastorno.

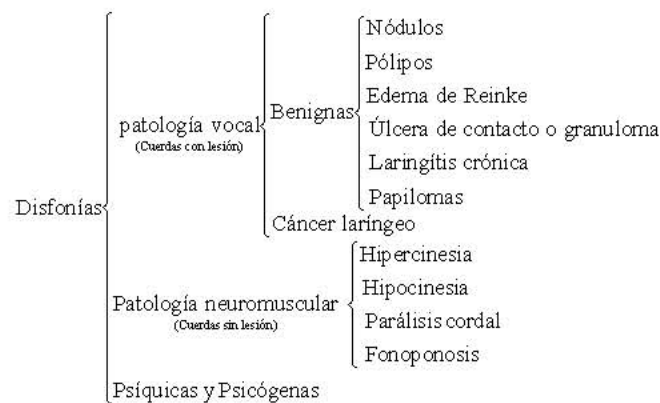
Ya en 1935 los trabajos de Tarneaud revisaron este concepto hipotético y accedieron a una mejor comprensión del problema, acuñando el término de disfunción. Tarneaud demostró que la desorganización de la función fonatoria es susceptible por sí sola de generar una alteración vocal. Para este autor la desorganización era resultado, sobre todo, de lo que denominó una “discordancia neumofónica”, es decir, una deficiencia de coordinación entre el fuelle o sistema respiratorio y el del vibrador o laringe. Además demostró que esta disfunción a través del mecanismo de sobreesfuerzo vocal, puede culminar en una lesión orgánica de la laringe.

2.3.1 Clasificación de las Disfonías

La clasificación de las enfermedades, cualquiera que sea su naturaleza, siempre ha entrañado una gran dificultad porque en ella se plasman puntos de vista que trascienden a la mera enunciación de una serie de patologías.

En este documento han sido clasificadas las disfonías en tres grupos [8] , las producidas por una patología vocal, las producidas por patologías neuromusculares, y las originadas por enfermedades psíquicas.

TABLA 2-1: CLASIFICACIÓN DE LAS DIFONÍAS



2.4 Patologías Benignas

Se incluyen dentro de esta clase a los pólipos, papilomas o nódulos formados, en su mayor parte, por forzar la voz de manera habitual. El síntoma es una ronquera gradual.

2.4.1 Nódulos Vocales

Son lesiones de pequeño tamaño, de coloración transparente, sonrosada o blanquecina descritas clásicamente como situadas en la unión del tercio anterior con el tercio medio de las cuerdas vocales. Esta situación se corresponde con el centro de la porción vibrátil de las cuerdas vocales, área de mayor impacto durante la fonación. Entre sus características se encuentran la de presentarse en la mayoría de los casos en ambas cuerdas vocales en situación simétrica, uno frente al otro, de ahí las denominaciones anglosajonas de *Kissing nodules* o francófona de *nodules bis-a-bis*.



Ilustración 2-6 : Nódulos Vocales

Podemos distinguir dos tipos de nódulos, los de presentación aguda y los crónicos. Como recoge Beckmann, los nódulos fueron descritos por primera vez por Hofer en 1904 como “nódulos de los cantantes”; Thost afirma que no serían nunca cantantes de primera línea. En el 95% de los pacientes se acompañan de un defecto de cierre glótico.

En muchas ocasiones una lesión como un pólipo o un quiste puede generar una lesión reactiva de aspecto nodular por el contacto en la cuerda vocal opuesta. Estas lesiones nodulares de contacto son histológicamente iguales a los nódulos vocales, pero no se engloban en este apartado.

Los nódulos de presentación aguda o iniciales son lesiones bilaterales puntiformes, generalmente translúcidas y de aspecto edematoso que se pueden encontrar, por ejemplo, en niños tras el llanto intenso o en cantantes después de un concierto. Suelen regresar a los pocos días con el reposo vocal. Los nódulos crónicos de aspecto blanquecino y fibrótico son de mayor tamaño que los agudos, pudiendo presentar un aspecto muy firme.

En la estroboscopia se observa una amplitud reducida en la ondulación de ambas cuerdas vocales que se mantiene simétrica; los nódulos agudos acompañan a la onda mucosa en su progresión, mientras que esto no sucede en los nódulos crónicos. La configuración del área glótica es irregular por un defecto de cierre anterior y posterior a los nódulos, dando lugar a la llamada “glotis en reloj de arena”.

Estas lesiones suponen, según las series de patología benigna, entre un 17 y un 24%. Los nódulos vocales son lesiones que predominan en mujeres y en la segunda década de la vida, aunque referidos a la edad infantil prevalecen en varones y en la edad adulta hay una clara prevalencia entre los profesionales de la voz.

Los nódulos son probablemente las lesiones cuya etiología traumática-fonatoria es más clara. En los casos agudos aparecen por un abuso vocal intenso en un corto periodo de tiempo, mientras que los nódulos crónicos aparecen como respuesta inflamatoria organizada ante la hiperfunción persistente. Aunque el síntoma principal es la disfonía muchos pacientes no presentan afectación mayor de la calidad vocal.

El tratamiento de los nódulos agudos es el reposo y los antiinflamatorios y en los casos crónicos el tratamiento inicial es logopédico debiendo dejarse la cirugía para aquellos casos en que la disfonía es limitante.

2.4.2 *Pólipos Vocales*

Son lesiones que aparecen, generalmente de modo unilateral, en el centro de la porción vibrátil de la cuerda vocal, aunque en ocasiones se localizan muy próximos a la comisura anterior. La mayor parte de los pólipos son de pequeño tamaño, aunque hay lesiones de dimensiones considerables. Su aspecto es muy variable, siendo su característica típica la apariencia angiomatosa. La implantación en la cuerda vocal puede ser mediante una base amplia, denominándose entonces *pólipos sésiles*, o mediante un pedículo, denominándose entonces *pólipos pediculados*. Ocasionalmente los pólipos pueden ser polilobulados.

Los pólipos se suelen asociar con frecuencia a lesiones por contacto en la cuerda contralateral. En algunas ocasiones el diagnóstico diferencial con los quistes es difícil, aún durante la laringoscopia directa.

La exploración con luz estroboscópica permite observar una clara asimetría en la vibración de las cuerdas; generalmente la amplitud de la vibración está disminuida en ambas cuerdas, aunque este hecho es más marcado en la cuerda que presenta la lesión. Este tipo de lesión predomina en varones, hasta casi un 80% de los pacientes, con una máxima incidencia entre los 30 y 50 años.

Son lesiones asociadas a la combinación de abuso vocal crónico y hábito de fumar, presente en casi el 80 % de los pacientes. En algunos pacientes con pólipos existen antecedentes de hematoma de las cuerdas vocales, pudiendo considerarse entonces como una organización de los hematomas.

La sintomatología más frecuente ocasionada por los pólipos es la disfonía crónica, presente en más del 80% de los pacientes; con mucha frecuencia producen necesidad de aclaramiento. Si el pólipo presenta un pedículo largo puede originar

disfonía intermitente, ya que el pólipo puede caer al espacio subglótico en la inspiración y las cuerdas contactar limpiamente durante la fonación. Si el pólipo es de gran tamaño puede ocasionar *disnea* o espasmos de glotis. Acústicamente se identifican laxitud del plano de cobertura y bitonalidad.

El tratamiento de esta entidad es quirúrgico en la mayor parte de los casos seguidos de rehabilitación foniátrica y abandono del hábito del tabaco.



Ilustración 2-7 : Pólipo vocal

2.4.3 Edema de Reinke

Según las distintas escuelas recibe nombres como laringitis crónica hipertrófica edematosa o pseudomixomatosa, degeneración o corditis polipoidea, fibromixoma o poliposis bilateral difusa. Dikkers y Schutte (1991) lo definen como una tumefacción sésil de las cuerdas, de asiento uni o bilateral y llena de un fluido gelatinoso, conservándose una buena movilidad de las cuerdas en la fonación, pero que origina una disfonía intensa y conlleva una mala calidad vocal. El edema es bilateral del 62.5 al 85% de los casos y más marcado en la porción posterior de las cuerdas. En ocasiones se pueden encontrar áreas esféricas pseudolobuladas sobre el edema global de la cuerda que recuerda la imagen de un pólipo.

Durante la exploración estroboscópica se observa aumento de la fase de cierre con un patrón asimétrico en el movimiento de ambas cuerdas vocales y una onda mucosa por lo general, aumentada de tamaño.

La distribución por sexos del edema de Reinke varía ampliamente según las series, aunque para algunos autores es una lesión que predomina en las mujeres, La edad de presentación oscila entre los catorce y setenta y ocho años, con una máxima incidencia entre los cuarenta y sesenta años.

Es una lesión asociada casi siempre al abuso vocal y a un gran consumo de tabaco. Myerson (1950) describe la lesión por primera vez en la literatura americana como “laringe de fumador”.

La disfonía aparece como principal síntoma en el 97% de los casos, seguida por la necesidad de aclaración y tos frecuentes. Los varones presentan un tono menor de 80 Hz y en las mujeres menor de 108 Hz, siendo esta frecuencia fundamental menor de lo normal y el rango dinámico está significativamente disminuido. En casos donde el edema sea muy extenso puede aparecer cierto grado de disfonía de origen laríngeo de instauración lenta, asociada a veces a estridor. Acústicamente la voz impresiona como de una gran laxitud del plano de cobertura, con gran irregularidad en la emisión, bitonalidad y frecuencia fundamental baja.

El tratamiento en primer lugar debe intentar eliminar los hábitos tóxicos e iniciar una terapia rehabilitadora, produciéndose la mejoría en muchos casos. En los casos de fracaso de esta primera postura ha de recurrirse a la microcirugía laríngea.

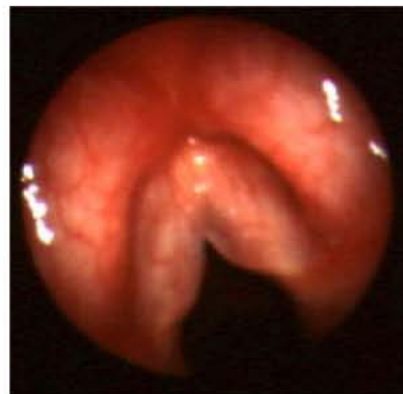


Ilustración 2-8 : Edema de Reinke

2.4.4 Úlcera de Contacto o Granuloma.

Se trata de una lesión sobreelevada recubierta generalmente por una mucosa muy delgada o de aspecto superficial ulcerado que afecta a una o ambas cuerdas vocales y que se asienta sobre las apófisis vocales de los aritenoides. Recibe también los nombres de úlcera de Jackson o de Holinger y Johnson.

El aspecto laringoscópico varía de unos casos a otros, siendo la presentación más común la de una masa esférica sobre las apófisis vocales de un aritenoides, en simetría con una lesión de menor volumen y a veces ulcerada sobre la apófisis vocal opuesta. Puede limitarse a una lesión ulcerada en un sólo lado o reconocerse como tal sobre la cara interna de ambas apófisis vocales. Durante la exploración se observa generalmente un defecto de cierre glótico posterior. La variación en su presentación se atribuye a diferentes estadios de la lesión. La forma ulcerada uni o bilateral se corresponde a la primera etapa de la lesión o a la última fase de curación.

Con luz estroboscópica no se aprecia alteración de la onda mucosa, ya que generalmente las cuerdas vocales no se encuentran afectadas.

Es una lesión propia del sexo masculino, con una incidencia máxima entre los 30 y 60 años. Suele producirse en los profesionales que utilizan mucho la voz como cantantes y profesores, así como en los pacientes afectados de reflujo gastro-esofágico.

Los síntomas son una disfonía caracterizada por una frecuencia fundamental baja que se acompaña de fatiga vocal y necesidad de aclarar la voz mediante tos o carraspeos frecuentes. Frecuentemente se acompaña de dolor y sensación de cuerpo extraño faríngeo.



Ilustración 2-9 : Úlcera de contacto o Granuloma

2.4.5 *Laringitis Crónica*

Se define como una serie de inflamaciones crónicas de la laringe con unas características comunes: afectan casi en exclusiva a las cuerdas vocales, sin alterar su movilidad; son claramente predominantes en el sexo masculino; el síntoma principal y prácticamente único es la disfonía; dependen de una amplia constelación etiológica de factores irritantes; y muestran en sus diferentes formas anatomoclínicas una distinta tendencia a la cancerización. Afecta en un 90% de los casos a varones en edades comprendidas entre los 30 y 60 años.

La disfonía de meses o años de evolución es el síntoma principal y casi exclusivo: intensa por la mañana, mejora al hablar para empeorar de nuevo vespertinamente con la fatiga vocal. No muestra tendencia espontánea a la curación. También presenta carraspeo para aclarar la voz, tos especialmente de forma irritativa aunque puede haber cierto componente productivo y parestesias, en forma de sequedad de garganta. La exploración laringoscópica permite clasificar diferentes formas de laringitis crónica, con una severidad y tipo de evolución distintos en cada una de ellas.

2.4.6 *Papilomas*

Son de etiología viral (*Papovavirus*). El 25% de los casos se presenta desde la infancia o adolescencia, y en el 75% se adquiere la enfermedad como adulto. Se observan lesiones exofíticas en pequeños racimos que comprometen fundamentalmente la glotis.

El tratamiento de elección es la extirpación vía laringoscopia directa con láser o con microcirugía endolaringea. No se ha demostrado utilidad del uso de antivirales ni de interferón. La radioterapia está contraindicada por su efecto carcinogénico.



Ilustración 2-10 : Papiloma

2.4.7 Cáncer de Laringe

La mayoría de los cánceres laríngeos son derivados de células escamosas que constituyen su epitelio de revestimiento. En la glotis en la mayoría de los casos se encuentran tumores escamosos queratinizantes bien diferenciados. El sitio más frecuentemente comprometido es el tercio anterior de las cuerdas vocales. Los tumores subglóticos tienden a ser carcinomas escamosos pobremente diferenciados. En general la frecuencia de metástasis es inversamente proporcional al grado de diferenciación. Otros tipos menos frecuentes son: el *Carcinoma Verrucoso* (1-2 %), el *Adenocarcinoma* y los *Sarcomas* (ambos menos del 1%).



Ilustración 2-11 : Cáncer de laringe

Los cánceres supraglóticos típicamente se presentan con dolor de garganta, *odinofagia*, *otalgia refleja*, cambios en la calidad de la voz, o nódulos cervicales. Los cánceres precoces de cuerdas vocales se detectan por una disfonía persistente.

El Sistema de clasificación es clínico, y se basa en la mejor estimación de la extensión de la enfermedad antes del tratamiento. El cáncer o tumor primario se diagnostica con laringoscopia ya sea indirecta con espejo o directa con fibroscopía. Es muy importante la inspección y palpación del cuello y laringe y determinar clínicamente el grado de compromiso de los ganglios. Se debe consignar una historia completa junto al examen ORL y general.

Hay una clara asociación entre el tabaco, el exceso de ingestión alcohólica y el desarrollo de cáncer de células escamosas del tracto aero-digestivo superior. Si un paciente con un cáncer continúa fumando y bebiendo alcohol, la probabilidad de curación inicial del cáncer en cualquier modalidad, disminuye, y el riesgo de un segundo tumor aumenta. El riesgo de un segundo tumor primario en pacientes con cánceres iniciales controlados es del 25%. Éste se presenta frecuentemente en el tracto aerodigestivo. Se ha demostrado en un estudio reciente que el tratamiento diario de estos pacientes con dosis moderada de *isotretinoína* por un año disminuye significativamente la incidencia de un segundo tumor.

2.5 Disfonías neuromusculares (funcionales)

2.5.1 Disfonías Hipercinéticas

Constituye una variedad de laringopatía funcional, producida por un exceso de tono de la musculatura intrínseca de la laringe. Se caracteriza porque los periodos de disfonía o afonía aparecen tras el esfuerzo vocal, mejorando con el reposo, y por presentar un comportamiento vocal de esfuerzo y una incoordinación fono-respiratoria.

Dentro de ella distinguimos un importante grupo que suele asociar lesión cordal, como consecuencia del sobreesfuerzo y un pequeño grupo en que la hipertonía se instaura sobre una lesión cordal.

En cuanto a las características acústicas de esta disfonía, existe una intensidad aumentada, el tono se halla descendido y el timbre con pocos armónicos y gran componente de ruido.

Exploratoriamente se observan unas cuerdas vocales con frecuencia congestivas, con una mucosidad característica en su tercio anterior, con hipertrofia de las bandas ventriculares y cierre antero-posterior del vestíbulo laríngeo. A la estroboscopia, la onda mucosa se encuentra disminuida en su amplitud, con fase de cierre aumentada pero de corta duración y con hiato posterior característico. La simetría y la periodicidad son variables según el grado de afectación.

El tratamiento se basa en el reposo vocal, supresión de hábitos tóxicos, tratamiento foniátrico y fonocirugía cuando existen lesiones asociadas susceptibles de tratamiento quirúrgico.

2.5.2 Disfonías Hipocinéticas

Se debe a un defecto de tono en la musculatura intrínseca de la laringe y generalmente se asocia un comportamiento de esfuerzo de la musculatura extrínseca. Se caracteriza porque el inicio de la fonación es dificultoso mejorando la voz a medida que se habla, aunque existe cansancio de forma precoz. Existe una importante incoordinación fono-respiratoria y acústicamente se caracteriza por una intensidad disminuida, el tono fundamental puede estar aumentado o disminuido, el timbre es pobre en armónicos y característicamente soplado.

A la exploración las cuerdas aparecen cortas y engrosadas, aunque en otras ocasiones aparecen de forma atrófica y con bordes libres arqueados. Los aritenoides basculan hacia delante con frecuencia, existiendo una hipertrofia de bandas y cierre antero-posterior del vestíbulo con mecanismo de hipertonía secundaria. A la estroboscopia la onda mucosa está aumentada de amplitud y suele ser asimétrica con una fase de cierre disminuida pero de larga duración y con una insuficiencia glótica posterior, longitudinal o en ojal.

El tratamiento se basa en la reeducación foniátrica y nunca en el reposo vocal.

2.5.3 Parálisis Cordal

Parálisis de cuerda (unilateral-bilateral) que consiste en la ausencia del movimiento de una o ambas cuerdas vocales.

2.5.4 Fonoponosis

Se produce conjuntamente un abuso y un mal uso vocal (pacientes que, aparte de elevar la intensidad de la voz, abusan de ésta mucho). El *abuso vocal* está relacionado con una higiene vocal pobre como: hablar excesivamente, chillar, gritar, excesivo aclaramiento de la garganta y demasiada tos, vocalizaciones forzadas, cantar con una técnica inadecuada, inhalación de polvo, humo y gases nocivos etc. Y el *mal uso vocal* está relacionado con hablar con una intensidad y/o tono más alto de lo que corresponde.

2.6 Referencia

- [1] Jackson Menaldi, *La voz normal*, Editorial medica panamericana, 1992.
- [2] Jackson Menaldi, *La voz patológica*, Editorial medica panamericana, 2002.
- [3] Husson R., “Sur la physiologie vocale”, *Ann. Oto-laryng*, vol.69, pp.124-137, 1953
- [4] Van Den Berg, “Myoelastic- aerodynamic Theory of Voice Production”, *Journal of Speech and Hearing Research*. vol.1, no.3, pp. 227-244, 1958
- [5] Perelló J., La théorie muco-ondulatoire de la phonation. *Ann. Oto-Laryng*, vol. 79, no. 9, pp. 722-725, 1962
- [6] Hirano M., “Morphological Structure of the Vocal Cord as a Vibrator and its Variations”, *Folia phoniatr.*, vol.6, pp.166-178, 1954.
- [7] Le Huche F. y Allali A, *La Voz*, Tomo II (“Patología Vocal”), Ed. Masson, Barcelona. 1993.
- [8] García Tapia R y Cobeta I., *Diagnóstico y Tratamiento de los trastornos de la Voz*, Ed. Garsi. Madrid, pp. 17-47.,1996.

CAPÍTULO 3

La Calidad de la Voz: Fenómenos Implicados

En este capítulo, se realiza un recorrido por los diferentes dominios de representación en los que se pueden estudiar la señal de voz con el fin de describir las diferentes manifestaciones, propias de cada dominio, que nos permite valorar la calidad de la voz. Cada una de las diferentes manifestaciones mencionadas, caracteriza un fenómeno físico que interviene en la generación de la voz, sin embargo el mismo fenómeno físico puede manifestarse en diferentes dominios, en forma de diferentes manifestaciones. Por tanto, en este capítulo se va a identificar el conjunto de fenómenos físicos necesarios describir, para realizar una correcta documentación de la calidad de voz de un locutor.

Medir la calidad de la voz consiste en cuantificar diferentes características acústicas de la voz las cuales permitan compararla respecto a ciertos patrones de referencia. En la bibliografía actual no existe una única medida que documente todas las características acústicas de la voz.

Una medida de la calidad acústica de la voz basada en una valoración auditiva es intrínsecamente difícil de establecer como referencia comparativa entre distintas voces y distintos evaluadores. En la bibliografía actual se han realizado varios intentos

de objetivar la medida de la calidad de voz por medio de medidas clínicas multidimensionales basadas en métodos auditivos. Existen ejemplos bien conocidos como son: la escala GRABS de Japón [1] y su formulario extendido, desarrollado y aplicado en Europa [2], o el utilizado en Suecia donde se ha trabajado en otro juego de descriptores clínicos de la voz referentes a la percepción [3], o como en el que se introduce un juego de características fonéticas las cuales intentan agregar información de la excitación del tracto vocal en la medida de la calidad de la voz [4]. Estos procedimientos de medida de la calidad de la voz pretenden alcanzar una medida objetiva a partir de una valoración subjetiva.

Existen diferentes trabajos en los se proponen **medidas objetivas** de la calidad de la voz obtenidas a partir de una grabación de voz. Una característica común de la mayoría de estos trabajos es la de utilizar para el cálculo de las medidas de la voz la grabación de un sonido sonoro sostenido, principalmente una vocal. Ésto es debido a que durante la fonación de este tipo de sonidos el sistema de producción de voz pone en funcionamiento gran parte de sus mecanismos (flujo glótico de aire constante, vibración de las cuerdas vocales de forma continuada, ...) permitiendo captar cualquier anomalía de estos mecanismos. El valor que toma cada una de las medidas de calidad de la voz depende de la vocal analizada [5].

En estos trabajos, los distintos autores presentan diferentes conjuntos de medidas con la finalidad de cuantificar objetivamente la calidad de voz. Todos ellos tienen en común la necesidad de realizar diferentes medidas de la voz con la finalidad de recoger en el proceso de medida los distintos aspectos de las características acústicas de la voz.

La calidad de la voz desde diferentes dominios

Una grabación de voz aporta diferentes características respecto a la calidad de la voz del locutor. Esta señal de voz registrada es posible representarla en diferentes dominios, lo cual resulta interesante ya que cada dominio en el que se puede representar la señal de voz, manifiesta de forma preferencial alguna de las características de la voz del locutor. Los principales dominios estudiados en el procesado de señales de voz [6] están centrados en los siguientes dominios:

- **Dominio Temporal**
- **Dominio Espectral**
- **Dominio Cepstral**

Sin duda alguna gran cantidad de los trabajos sobre procesado digital de señales de voz existentes en la bibliografía actual centran sus estudios en estos dominios. Sin embargo, nuevas líneas de trabajo presentan nuevos dominios derivados de los anteriores.

De igual forma resulta de interés estudiar nuevos dominios que cada vez toman mayor importancia en el procesado de señales: los **dominios que manifiestan el comportamiento no lineal de la señal de voz**.

A continuación se describen las características más relevantes de cada uno de los diferentes dominios. En este punto resulta necesario introducir el concepto de trama, ya que tanto para la visualización como para el análisis de estas características, resulta necesario segmentar la fonación con el fin de optimizar la obtención de información que aporta cada dominio. A cada una de estas unidades fruto de la segmentación se la denomina trama. La duración temporal de las tramas, dependiendo de cada característica, suele ser o bien constante (por ejemplo 20, 25, 100, 300 milisegundos) o bien en algunos casos depende de una relación directa con respecto a algún parámetro fisiológico (por ejemplo, tres veces la duración de un ciclo de apertura y cierre de la cuerdas vocales).

3.1 Dominio Temporal

En el dominio temporal se manifiestan ciertas características de la calidad de la voz las cuales permiten diferenciar entre una muestra de voz de alta calidad y una muestra de voz de baja calidad. En la Figura 3-1 se muestra una representación de los cinco golpes de voz correspondientes a las cinco vocales castellanas pronunciadas de forma sostenida (se han eliminado los silencios existentes entre cada vocal) para una voz de buena calidad y para una voz de mala calidad.

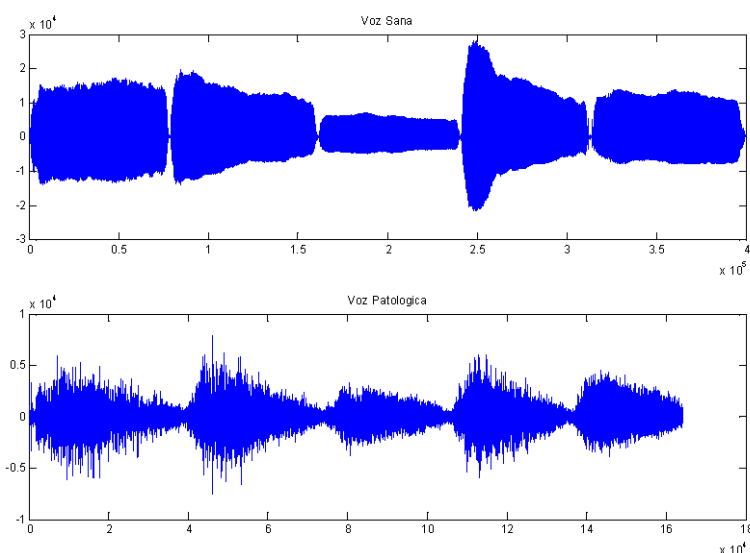


Figura 3-1: Representación de la señal de voz en el dominio temporal correspondiente a la fonación de cinco vocales de forma sostenida de un locutor con voz de alta calidad (superior) y de un locutor con voz de baja calidad (inferior)

En la anterior figura puede apreciarse que la señal de voz de un locutor con voz de alta calidad resulta más regular que la envolvente de señal de voz del locutor con voz de baja calidad. Este hecho resulta más significativo en intervalos de tiempo muy pequeños. Los fenómenos más característicos que permiten diferenciar entre voz de alta calidad y voz de baja calidad son:

- La energía de la señal contenida en un intervalo de tiempo pequeño, varía mucho de un intervalo al siguiente en las señales de voz de baja calidad respecto a las señales de voz de alta calidad. Esta característica de variabilidad también se manifiesta en variaciones en la forma de onda entre dos periodos consecutivos. Este fenómeno se puede observar en la Figura 3-2.

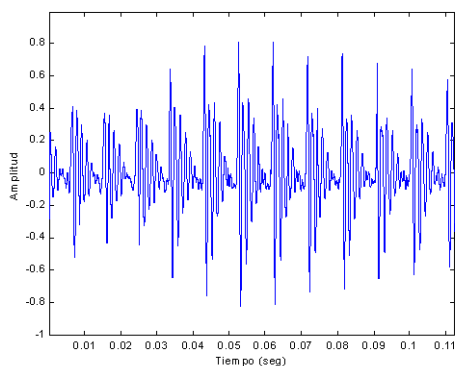


Figura 3-2: Representación en el dominio temporal de un sonido sonoro sostenido

- Otra característica que presenta las voces de baja calidad es la existencia de intervalos carentes de periodicidad lo que se identifica como la existencia de tramas sordas durante fonaciones sonoras sostenidas. Un ejemplo de este fenómeno se muestra en la Figura 3-3.

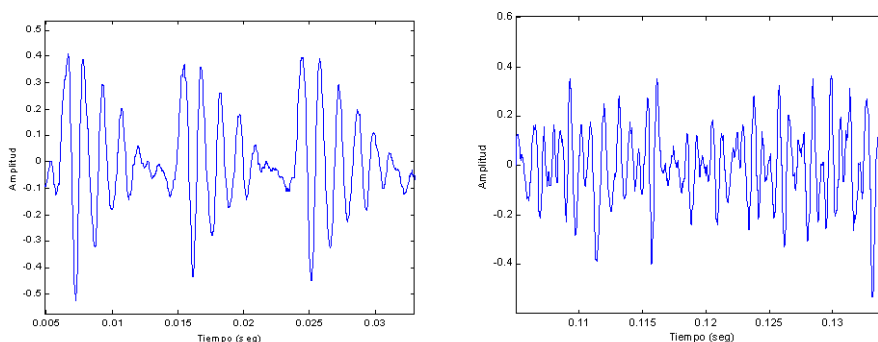


Figura 3-3: Representación en el dominio temporal de un sonido sonoro sostenido durante un corto intervalo de tiempo para una voz de alta calidad (izquierda) y para una voz de baja calidad (derecha)

3.2 Dominio Espectral

En este dominio es posible estudiar la contribución de potencia de cada una de las frecuencias (componentes espectrales) que forman la señal. Este estudio se realiza a partir de una nueva señal denominada espectro. El espectro se calcula a partir de una trama de voz.

El procesamiento espectral clásico considera la señal de voz estacionaria y ergódica cuando se estudia en intervalos cortos de tiempo, lo que nos permite definir el espectro de una trama de voz según la siguiente ecuación:

$$S_x(\omega) = \sum_{m=-\infty}^{+\infty} R_x(m) e^{-j\omega m} \quad [\text{Ec.3-1}]$$

siendo $R_x(m)$ la función de autocorrelación, la cual se define:

$$R_x(m) = E\{x(n)x(n+m)\} \quad [\text{Ec.3-2}]$$

Debido a que no es conocida la función densidad de probabilidad de la señal de voz para cada una de las tramas es necesario recurrir al uso de estimadores espectrales, como por ejemplo el periodograma. El cálculo de este estimador del espectro se realiza según:

$$\hat{S}_x(\omega) = \frac{1}{N} \cdot |X(\omega)|^2 \quad [\text{Ec.3-3}]$$

donde $X(\omega)$ es la transformada de Fourier de cada una de las tramas. Existen otros estimadores no paramétricos del espectro como el estimador de Welch y Blackman-Tukey, además de los diferentes métodos basados en los estimadores paramétricos del espectro.

De los parámetros que caracterizan una señal de voz desde el punto de vista espectral están:

- **Formantes:** son picos en el espectro de voz consecuencia de la cavidad vocal. Están caracterizados por su frecuencia central y su ancho de banda. Para un adulto las resonancias están centradas en torno a 500, 1500, 2500 y 3500 Hz. Existen pequeñas diferencias en las frecuencias de los formantes entre locutores de distinto sexo. Generalmente las frecuencias de los formantes de locutores femeninos son mayores que las frecuencias de locutores masculinos.
- **Frecuencia fundamental o “Pitch”:** es la frecuencia de vibración de las cuerdas vocales. En realidad no hay periodicidad total, este es el motivo de hablar de “casi periodicidad” en la frecuencia fundamental.

- **Timbre o matiz:** el movimiento vibratorio generador del sonido es, en general, un fenómeno complejo en el que intervienen simultáneamente, un movimiento vibratorio principal (vibración de las cuerdas vocales) y por otro lado, uno o más movimientos vibratorios secundarios. Los tonos secundarios (armónicos) resultan de las resonancias que el tono principal (armónico principal o *pitch*) produce en la cavidad o cavidades formadas en el tracto vocal. En este conjunto sonoro de tono fundamental (*pitch*) y tonos secundarios, el resonador predominante determina el timbre o matiz característico de cada sonido. Los sonidos son por su timbre, así como por su tono, agudos o graves, según la altura de la nota que corresponde a su resonador predominante.

El dominio espectral permite diferenciar las muestras de voz de alta calidad y las muestras de voz de baja calidad. En la Figura 3-4 se muestra una representación del espectro en escala logarítmica de una vocal sostenida producida por un locutor con voz de alta calidad y el espectro en escala logarítmica de la misma vocal producida por un locutor con voz de baja calidad.

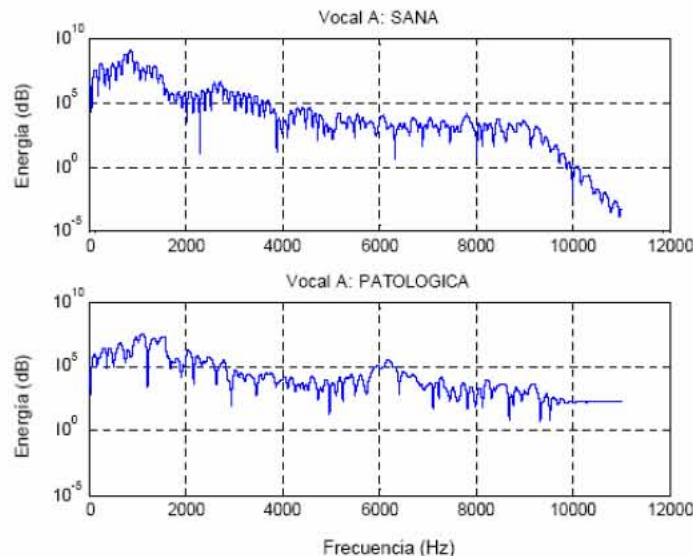


Figura 3-4: Representación del espectro en escala logarítmica de una vocal sostenida producida por un locutor con voz de alta calidad (superior) y el espectro en escala logarítmica de la misma vocal producida por un locutor con voz de baja calidad (inferior)

En las muestras de voz de un sonido sonoro sostenido producidas por un locutor con voz de baja calidad presentan las siguientes características diferenciales frente a las voces de alta calidad:

- Menor regularidad de la envolvente del espectro, principalmente en las bajas frecuencias.
- Mayor porcentaje de energía en las bajas frecuencias respecto a la energía total
- Presencia de bloques de energía en las altas frecuencias, los cuales son debido a la presencia de ruido glótico.
- Gran variación del espectro de una trama con respecto a las tramas contiguas resultando significativa la variación del espectro entre tramas contiguas con duración de tres periodos de *pitch*.

En la siguiente figura se muestra la envolvente del espectro para un sonido sonoro sostenido. Cada uno de los picos de la envolvente del espectro corresponde con los diferentes formantes. Los cuatro primeros formantes han sido utilizados ampliamente en trabajos de reconocimiento de voz [23] [24] . En la Figura 3-5 se muestra la posición de los cuatros primeros formantes para un sonido sonoro sostenido correspondiente a una vocal.

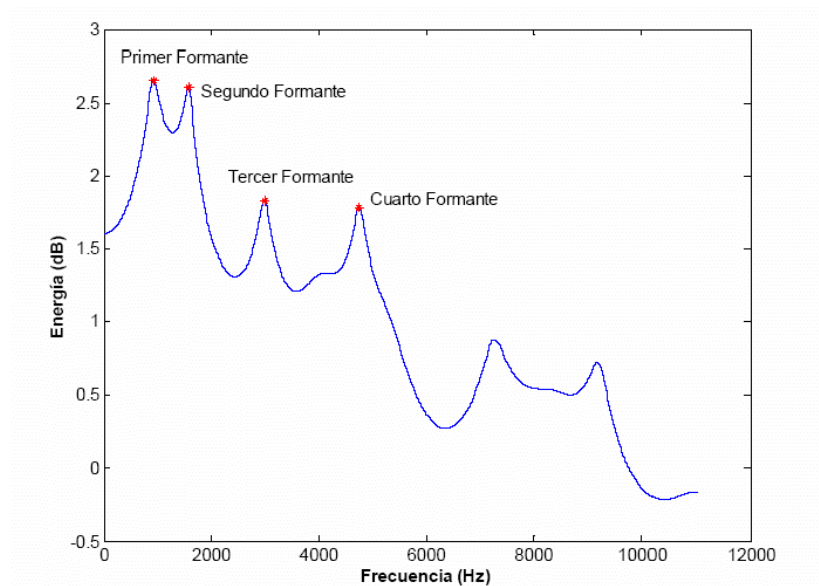


Figura 3-5: Envolvente del espectro de una trama de voz donde aparece señalado los cuatro primeros formantes

Una característica de las voces de alta calidad es la energía del espectro concentrada alrededor de determinados formantes, principalmente el primer y el tercer

formante, mientras que las voces de baja calidad se caracterizan por presentar gran componente de ruido alrededor de los formantes.

El problema de una representación espectral basada en el espectro es la pérdida de información en lo referente a su evolución temporal, es decir, el espectro no representa la variación espectral correspondiente a las diferentes tramas. Sin embargo, existe otra forma de representar la información espectral denominada espectrograma, la cual solventa esta carencia, realizando una representación del espectro a lo largo del tiempo. Para ello, representa la potencia de las diferentes componentes espectrales por medio de la intensidad de color utilizada en la representación, utilizando los ejes para representar las frecuencias y el tiempo. En la Figura 3-6 se muestra el espectrograma de una muestra de voz de alta calidad y una muestra de voz de baja calidad durante la fonación de las cinco vocales.

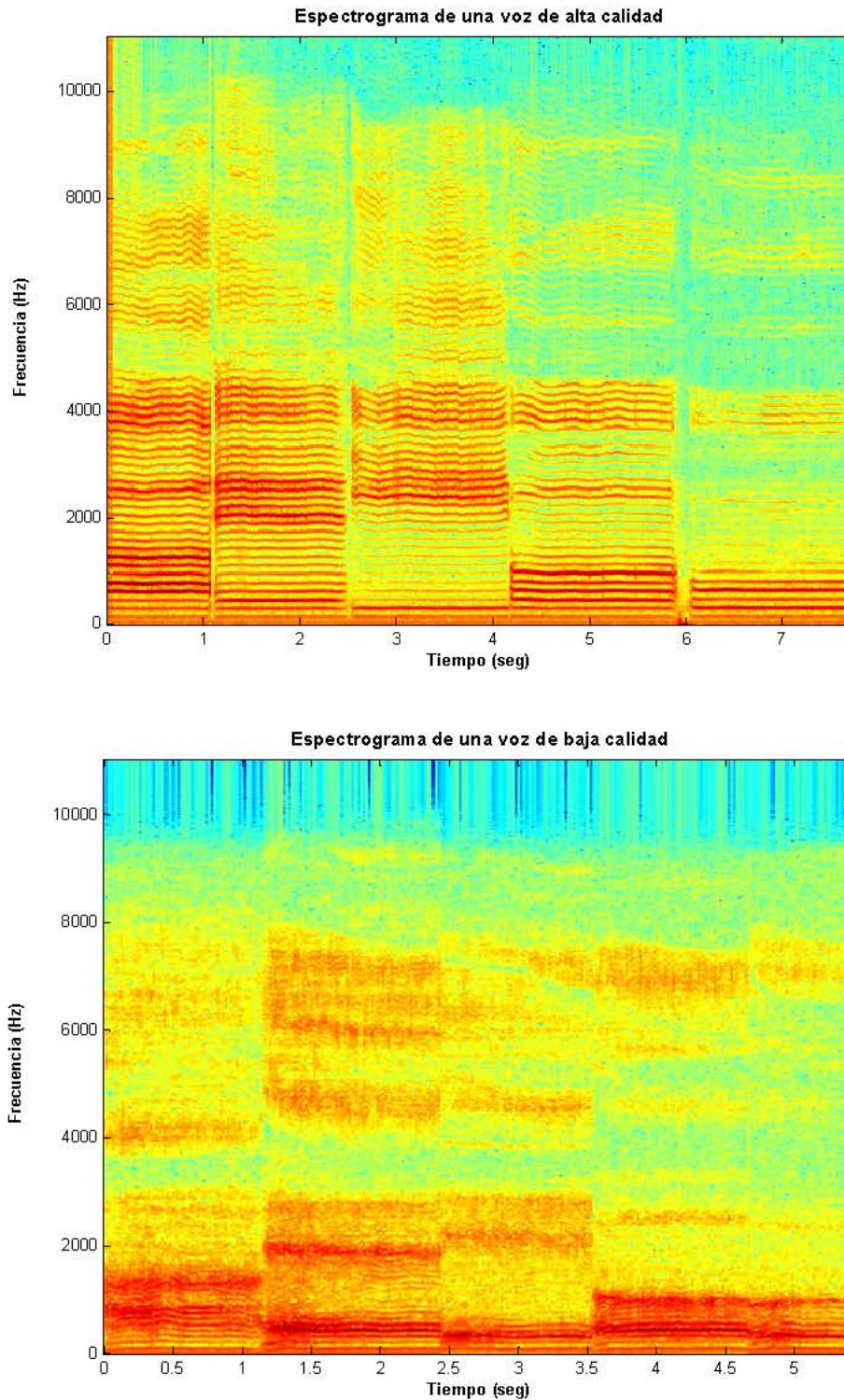


Figura 3-6: Espectrograma estimado para una frecuencia de muestreo de 22050 Hz de una muestra de voz de alta calidad (superior) y una muestra de voz de baja calidad (inferior) durante la fonación de la cinco vocales

En el espectrograma de voz se aprecian características las cuales nos permiten identificar particularidades que permiten valorar la calidad de voz. Durante la fonación de un sonido sonoro sostenido se aprecian las siguientes características:

- Las voces con alta calidad de la voz se caracterizan por presentar líneas paralelas, correspondientes a las componentes armónicas, principalmente en las bajas frecuencias. En las altas frecuencias también es común la existencia de las componentes armónicas, siendo cada vez menos paralelas. Esta característica nos da idea de la riqueza espectral de la voz. Por el contrario, las voces de baja calidad presentan muy poca componente armónica, estando ésta concentrada en las componentes de muy baja frecuencia.
- Las voces de baja calidad presentan zonas de energía irregulares, lo cual indica la presencia de cantidades anormales de ruido. Este fenómeno se manifiesta principalmente en altas frecuencias y alrededor de los formantes, lo cual es significativo de la presencia de ruido glótico.

Otra característica típica de las voces de baja calidad es la de presentar, durante la fonación de un sonido sonoro sostenido, variaciones en el ritmo de vibraciones de las cuerdas vocales, lo cual es equivalente a decir que existe variaciones de la frecuencia de *pitch*. Para el cálculo de la variación del valor de *pitch* es necesario el uso de un detector de *pitch* exacto, debido a que diferentes medidas de calidad de voz se fundamentan en los valores de *pitch*. Existe mucha bibliografía sobre detectores de *pitch*, destacando entre ellas [25] ,[26] y [27] .

En la Figura 3-7 se muestra los valores que toma la frecuencia de *pitch* para las distintas tramas de una muestra de voz sana y de una muestra de voz patológica.

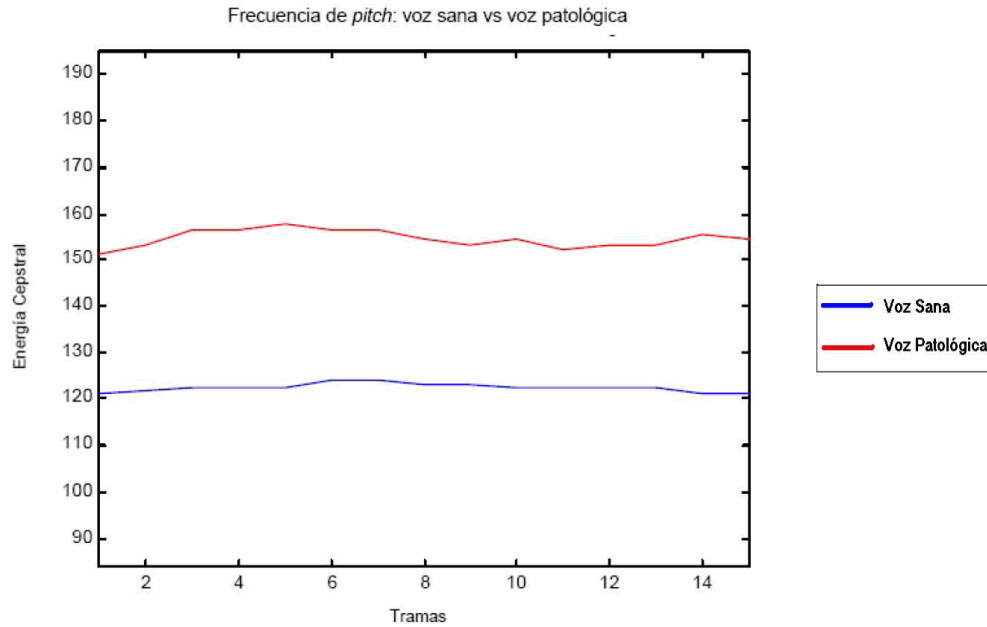


Figura 3-7 : Valores de la frecuencia de pitch para las distintas tramas de una muestra de voz de alta calidad (color azul) y de una muestra de voz de baja calidad (color rojo). Ambas muestras corresponden a la vocal "a" sostenida de un locutor masculino

Existen variantes al dominio espectral [6] , como por ejemplo el dominio mel-espectro, en el que la escala de frecuencias no es lineal permitiendo simular la percepción acústica, o el dominio de los LP (*Linear Predictor*), en el cual se estudia el efecto de la envolvente del espectro, o bien una representación utilizando MBE [9] en el que se pretende identificar qué partes de espectro corresponde a parte armónica y qué partes corresponde a las componentes de ruido.

3.3 Dominio Cepstral

Los *cepstrum* complejos y los *cepstrum* de potencia han sido ampliamente utilizados en el procesamiento digital de señales. Los *cepstrum* complejos [7] [8] de una secuencia discreta $\{x(n)\}$ de longitud N , cuya transformada de Fourier es $X(\omega)$, se definen como:

$$c_x(n) \equiv F^{-1}\{\log(X(\omega))\} \quad [\text{Ec.3-4}]$$

donde $F^{-1}\{\}$ denota la transformada inversa de Fourier.

Los *cepstrum* de potencia, también llamados *cepstrum*, se definen como:

$$p_x(n) \equiv F^{-1} \left\{ \log(|X(\omega)|^2) \right\} \quad [\text{Ec.3-5}]$$

La relación que existe entre los *cepstrum* complejos y los *cepstrum* de potencia es la siguiente:

$$p_x(n) = c_x(n) + c_x(-n) \quad [\text{Ec.3-6}]$$

El dominio de los *cepstrum* de potencia es el comúnmente conocido como dominio *Cepstral*. Por medio de una representación *cepstral* es posible obtener información sobre la envolvente del espectro, la riqueza espectral e identificar las componentes armónicas y las componentes de ruido.

En la Figura 3-8 se muestra una representación de los *cepstrum* para una muestra de voz de alta calidad y para una muestra de voz de baja calidad.

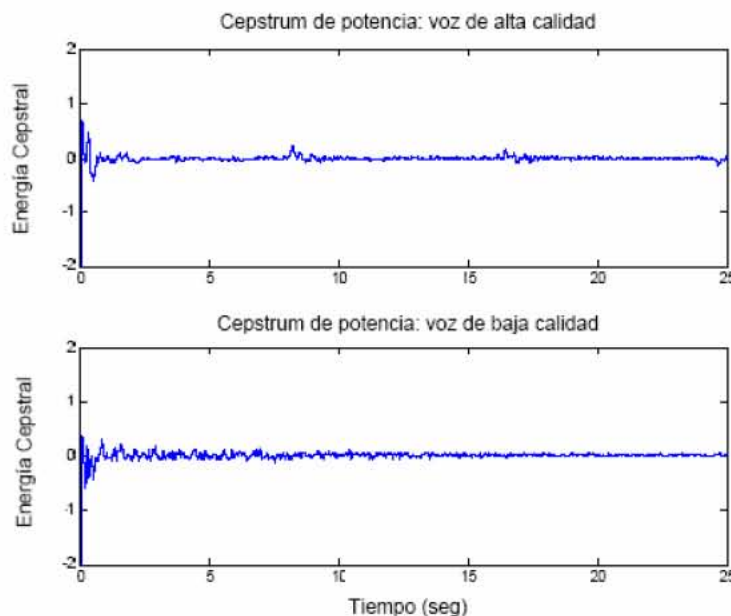


Figura 3-8: Representación en el dominio de los *cepstrum* de potencia de una muestra de voz de alta calidad (superior) y para una muestra de voz de baja calidad (inferior)

Es posible apreciar de forma más descriptiva la información que aportan los *cepstrum* de potencia realizando un rectificado, es decir, igualando los valores negativos a cero. En la Figura 3-9 se muestra una representación de los *cepstrum* de potencia rectificados aplicando un zoom en la zona de mayor interés.

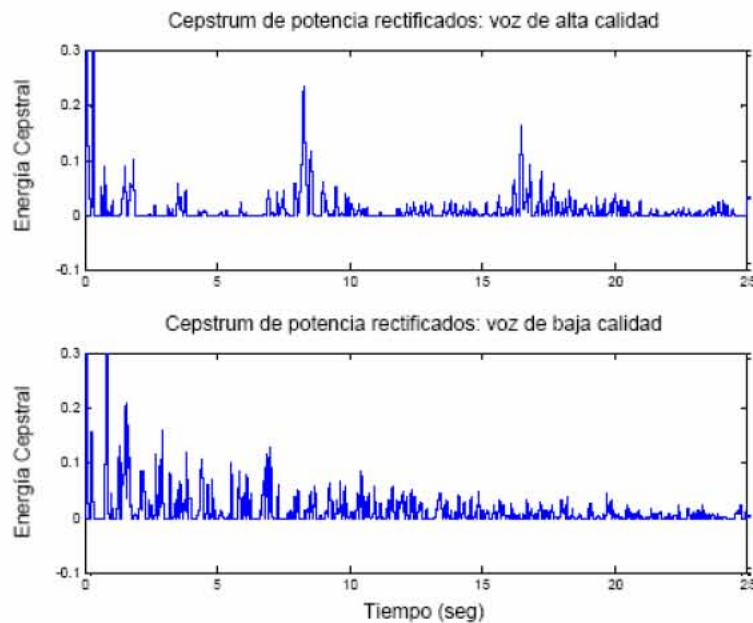


Figura 3-9: Representación en el dominio de los cepstrum de potencia rectificados de una muestra de voz de alta calidad (superior) y para una muestra de voz de baja calidad (inferior)

Por medio de una representación cepstral es posible identificar características que permiten valorar la calidad de la voz. En el caso de una fonación de un sonido sonoro sostenido, realizando una representación cepstral de diferentes tramas con duración equivalente a tres periodos de *pitch*, pueden identificarse las siguientes características:

- La riqueza espectral de una muestra de voz puede cuantificarse por medio de la amplitud y anchura de la componente cepstral correspondiente al *pitch*. La existencia de un pico de una amplitud considerable indica la presencia notable de energía en dicha componente armónica, siendo ésta una característica de las voces de alta calidad. También, una anchura reducida del pico cepstral, correspondiente al *pitch*, para una trama de tres periodos de *pitch*, indica la alta estabilidad de la frecuencia de *pitch* para tres periodos consecutivos, también siendo ésta una característica de las voces de alta calidad.
- La presencia de ruido en diferentes regiones de los *cepstrum* puede cuantificarse por medio de la relación entre: la componente armónica (las componentes cepstrales correspondientes al *pitch* y sus armónicos), y la componente de ruido (el resto de las componentes cepstrales).

En el dominio *cepstral* se manifiesta con claridad la presencia de componentes armónicas. En la Figura 3-10 se aprecia la localización del primer armónico (el *pitch*) y el segundo armónico.

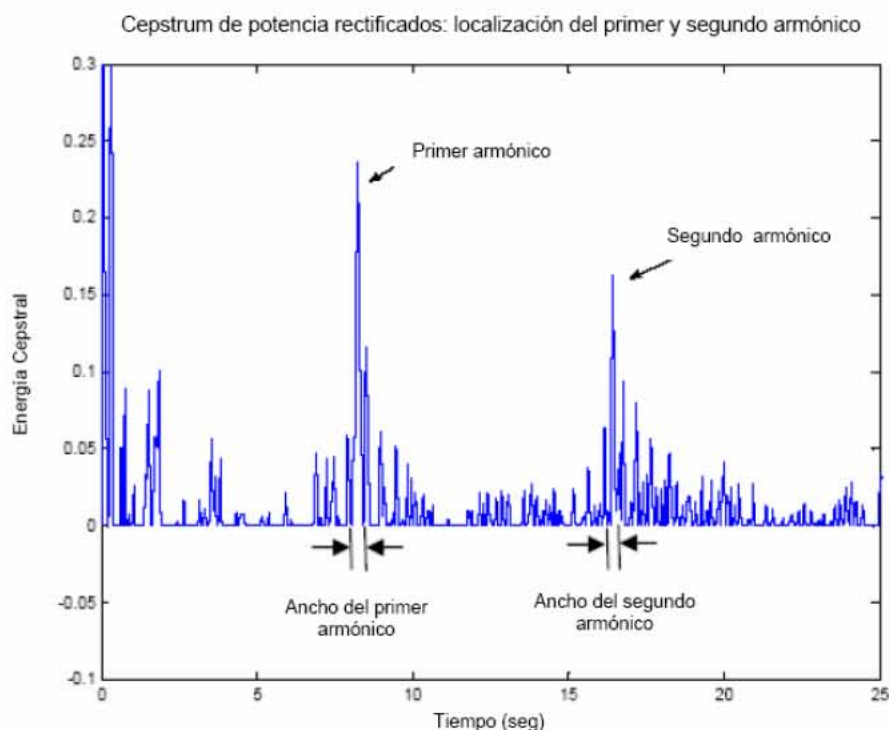


Figura 3-10: Localización del primer y segundo armónico en el dominio de los cepstrum de potencia rectificadas

Características como la amplitud y la estrechez del pico *cepstral* correspondiente al segundo armónico son, al igual que ocurre con el pico del *pitch*, aspectos que diferencian a las muestras de voz de alta y baja calidad.

La presencia de un pico *cepstral* correspondiente al primer armónico el cual resulte más estrecho que el pico *cepstral* correspondiente al segundo armónico resulta una característica propia de las voces de alta calidad. Esta característica matemática en el dominio de los *cepstrum* se identifica con el fenómeno físico consistente en la capacidad de tener una frecuencia de vibración de las cuerdas vocales (frecuencia del *pitch*) muy estable, sumándose la capacidad del locutor de generar armónicos de frecuencia del *pitch* con variaciones a corto intervalo de tiempo. Un ejemplo de este fenómeno se aprecia en la voz proyectada, donde el locutor es capaz de mantener el *pitch* constante y para cada uno de sus armónicos presentar ligeras variaciones respecto a los múltiplos exactos de la frecuencia de *pitch*, manifestándose más significativamente

este fenómeno en los armónicos de altas frecuencias. Esta característica se denomina *vibratto* [29] .

Existen otros dominios similares [6] como los *delta-cepstrum*, los *mel-cepstrum*, y los *mel-delta-cepstrum*, etc, los cuales permiten destacar ciertas particularidades que se manifiestan en el dominio cepstral.

3.4 Dominio del Modelo Inverso

En este dominio se estima, a partir de la señal de voz, la forma de onda del pulso de aire que se produce debido a las cuerdas vocales durante la fonación de un sonido sonoro, es decir, la forma de onda de la señal residual (forma de onda del flujo glótico). Esta estimación de la forma de onda se obtiene mediante un filtro inverso sobre la señal de voz, donde se elimina el efecto del tracto vocal y el efecto de la radiación en los labios.

Las técnicas de filtrado inverso están apoyadas en el modelo de producción de voz basado en filtros lineales [10] . La idea fundamental de las técnicas de filtrado inverso es a partir de la señal de voz registrada por medio de un micrófono obtener la señal de flujo glótico U_g o la señal derivada del flujo glótico dU_g . La señal del flujo glótico es la señal que excita al tracto vocal y que se obtiene a la entrada de la glotis.

Se ha experimentado con la posibilidad de obtener el flujo glótico a partir del flujo de aire emanado de los labios durante la producción de voz. Para obtener el flujo de aire se ha utilizado la máscara de Rothengerg. Esta técnica tiene el inconveniente de no captar frecuencias superiores a 1500 Hz, además de la pérdida de aire a través del pliegue entre la máscara y la cara.

Actualmente se puede conseguir la señal del flujo glótico por medio de la electroglotografía (EGG). Esta técnica consiste, por medio de la implantación de unos electrodos externos, medir la variación de resistividad que existe en el exterior de las cuerdas vocales cuando se induce una corriente eléctrica. Esta prueba se realiza durante la fonación de un sonido sonoro. Esta técnica tiene el inconveniente de la dependencia con la correcta ubicación de los electrodos para obtener unos resultados correctos.

Con la finalidad de cuantificar la calidad de la voz, usualmente no es suficiente el cálculo de U_g y dU_g , resultando necesario parametrizar las señales U_g y dU_g , para posteriormente evaluar estos parámetros. Existen técnicas de parametrización en el dominio del tiempo, en el dominio frecuencial y en el dominio tiempo-frecuencia. La elección del dominio óptimo generalmente depende de la aplicación. El modelo de parametrización más utilizado es el modelo Liljencrants-Fant (*LF model*)[28]. Para realizar la parametrización de la voz basta con un número pequeño de periodos del *pitch* de un sonido sonoro sostenido. En la Figura 3-11 se muestran los parámetros del modelo LF.

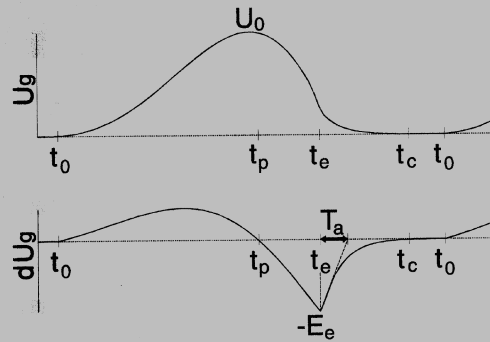


Figura 3-11: Modelo y parámetros LF

Los parámetros del modelo LF pueden ser divididos en tres grupos:

Amplitudes:

E_e : Fuerza de la excitación.

$$E_e = \min(dU_g)$$

U_o : Pico del flujo glotal.

$$U_o = \max(U_g)$$

Momentos:

t_o : Momento de apertura.

t_p : Momento del pico en U_g .

$$t_p = \arg \max(U_g)$$

t_e : Momento de excitación.

$$t_e = \arg \min(dU_g)$$

t_c : Momento de cierre.

Intervalos de tiempos de duración:

T_o : Duración de un periodo de *pitch*.

$$T_o = 1/F_o$$

T_a : Duración del intervalo entre t_e y la proyección de la tangente de dU_g en t_e sobre el eje de tiempo.

Además de estos parámetros, existen otros parámetros que se calculan a partir de los parámetros LF. Ejemplos de ellos son:

- “*Open quotient*” (OQ): Es la relación entre el tiempo en el que las cuerdas vocales están abiertas y el tiempo equivalente al periodo fundamental.

$$OQ = \frac{(t_c - t_o)}{T} \quad [\text{Ec.3-7}]$$

- “*Speed quotient*” (SQ): Es la relación entre el tiempo durante el cual están abriéndose las cuerdas vocales y el tiempo durante el cual están cerrándose. A este parámetro también se le suele de nominar “*Skewness*” o RK.

$$OQ = \frac{(t_p - t_o)}{(t_c - t_p)} \quad [\text{Ec.3-8}]$$

- “*Closing quotient*” (CQ): Es la relación entre el tiempo durante el cual se cierran las cuerdas vocales respecto al periodo fundamental.

$$OQ = \frac{(t_c - t_p)}{T} \quad [\text{Ec.3-9}]$$

Existen dos métodos para estimar los parámetros LF [11] : el método directo de estimación (*Direct Estimation Method: DE method*) y el método de estimación por ajuste (*Fit Estimation Method: FE method*).

El método de estimación directa calcula los parámetros directamente de U_g y dU_g , por medio de simples operaciones aritméticas como: mínimos, máximos, argumento mínimo y argumento máximo. Además, también se usan los cruces por ceros. Este método tiene varios inconvenientes. Uno de los inconvenientes que presenta

es el debido a que se procesan señales discretas, y por lo tanto en la práctica, los puntos claves de U_g y dU_g pueden no coincidir con ninguna de sus muestras. Otro de los inconvenientes que presenta este método es lo susceptible que resulta a las perturbaciones tales como el ruido o el rizado de los formantes. Por el contrario, este método presenta la ventaja de resultar una técnica relativamente sencilla.

El método de estimación por ajuste calcula los parámetros LF mediante una técnica iterativa. Esta técnica consiste en partir de una forma de onda glótica sintética y por cálculo de error (basados en la forma de onda) ir ajustando los parámetros LF de la onda sintética hasta que se obtenga un error aceptable.

En tareas de síntesis de voz, la calidad de la voz sintetizada depende de dos factores [12] : la inteligibilidad y la naturalidad. La inteligibilidad depende de la capacidad del sintetizador de reproducir los formantes (frecuencias, anchos de banda y transiciones), mientras que la naturalidad depende de la forma de onda de la excitación.

La forma de onda derivada del flujo glótico (dU_g) se puede reconstruir a partir de los parámetros LF. Para la reconstrucción se necesitan cuatro parámetros. La reconstrucción se divide en dos fases. La primera fase comprende desde el momento en el que comienza a abrirse la glotis (las cuerdas vocales) hasta el momento en el que comienza a cerrarse la glotis. Esta fase se modela por medio de la siguiente expresión:

$$\frac{dU_g}{dt} = E_o \cdot e^{\alpha t} \cdot \sin \omega_g t \quad t_o \leq t \leq t_e \quad [\text{Ec.3-10}]$$

donde:

E_o es el valor máximo de dU_g

$\alpha = B\pi$,donde B es el ancho de banda de la exponencial creciente

$\omega_g = 2\pi F_g$, donde $F_g = 1/(2t_p)$

La segunda fase del modelo es un segmento de una exponencial en la cual se cierra la glotis (las cuerdas vocales). Esta fase se representa como:

$$\frac{dU_g}{dt} = -\frac{E_e}{t_a \mathcal{E}} \left[e^{\mathcal{E}(t-t_e)} - e^{\mathcal{E}(t_c-t_e)} \right] \quad t_e \leq t \leq t_c \quad [\text{Ec.3-11}]$$

donde:

E_e es la máxima amplitud negativa de dU_g

t_c es el instante de cierre glótico completo

t_a duración del intervalo entre t_e y la proyección de la tangente de dU_g en t_e sobre el eje del tiempo.

Siendo el valor de ε calculado por la siguiente relación:

$$t_a \varepsilon = 1 - e^{\varepsilon(t_c - t_e)} \quad [\text{Ec.3-12}]$$

donde para valores pequeños de t_a , ε es aproximadamente igual a $1/t_a$

Existe un requerimiento para el modelo LF: el área de dU_g debe ser cero.

$$\int_0^{t_o} dU_g(t) dt = 0 \quad [\text{Ec.3-13}]$$

Existen técnicas en las que se realiza el filtrado inverso para lo cual es necesario ajustes manuales o resulta necesario en alguna fase del proceso una señal electroglotográfica (EGG) [13] .

Existen diferentes técnicas propuestas en las bibliografía actual que permiten obtener la señal de flujo glótico U_g a partir de la señal de voz, como la técnica basada en el algoritmo *IAIF* (*Iterative Adaptive Inverse Filtering*) [14] o la basada en una estimación por medio del método de los mínimos cuadrados [15] . Es posible obtener la señal residual por medio de la combinación de filtros de Wiener y filtros de Kalman [22] , con el fin de atender a la no estacionalidad de las características de un sonido sonoro sostenido, máxime en el caso de las voces de baja calidad.

Existen otros modelos para generar la forma de onda glótica además del LF que se ha utilizado con éxito en síntesis de voz. Las señales generadas por estos modelos también son susceptibles de ser modeladas con el fin de cuantificar la calidad de la voz. Un buen ejemplo de ellos es el modelo paramétrico de Rosenberg [60] . Este modelo ha sido utilizado para realizar un estudio sobre *jitter* y *shimmer* [61] .

Veldhuis propuso el modelo Rosenberg++ [57] [59] para caracterizar la derivada glotal, como una extensión del modelo paramétrico propuesto por Rosenberg. Estas modificaciones fueron introducidas para aumentar la flexibilidad del modelo original, incorporando control sobre la fase de retorno y asimetría del pulso glotal. El modelo R++ se diferencia del modelo LF en que el modelo R++ utiliza un polinomio de tercer orden para representar la forma de onda durante la fase de apertura.

Otro buen ejemplo es el modelo propuesto por Fujisaki y Ljungqvist [58] también conocido como modelo FL. La derivada del pulso glotal está formada por dos segmentos polinómicos, lo cual permite modificar de forma sencilla el número de parámetros (hasta 6), y por lo tanto el grado de detalle del modelo.

3.5 Dominio No Lineal

En el contexto de procesamiento digital de señales, tradicionalmente se ha utilizado estimadores y modelos con el fin de caracterizar un determinado sistema, un ejemplo es el sistema de producción de voz. Estos modelos generalmente suelen ser sistemas lineales basados en estadísticos de primer y de segundo orden, donde el uso de dichos modelos presenta ciertas limitaciones como enmascarar cierta información como la de fase, la presencia de alinealidades o comportamientos dinámicos no lineales. Por lo tanto muchos de los modelos existentes para diversos sistemas no contemplan algunas características provocando caracterizaciones menos realistas de los sistemas.

Por los motivos expuestos, en los diferentes ámbitos de actuación del procesamiento digital de señales se están revisando los diferentes procedimientos aplicando técnicas nuevas, principalmente técnicas no lineales en el cálculo de estimadores y modelos, aumentándose las posibilidades del procesamiento digital de señales.

La señal de voz no es una excepción, mostrándose en la literatura actual nuevas líneas de investigación que trabajan en la búsqueda de modelos que permitan caracterizaciones más realistas de la calidad de la voz.

3.5.1 Estado del Arte: ¿por qué usar modelos no lineales?

En la ciencia, a menudo se usan aproximaciones lineales a fenómenos no lineales. Un ejemplo es la característica voltaje-corriente de un transistor el cual tiene un rango de operaciones cuasilineal. Si el transistor trabaja en el rango lineal, el modelo lineal funciona perfectamente. La ventaja de usar este modelo, es que resulta fácil de estimar y de adaptar a sistemas tiempo-variante.

El modelo más popular de caracterización del sistema de producción de voz es un sistema tiempo-variante, basado en teorías acústicas lineales, consistente en un modelo de fuentes y filtros (*source/filter model*).

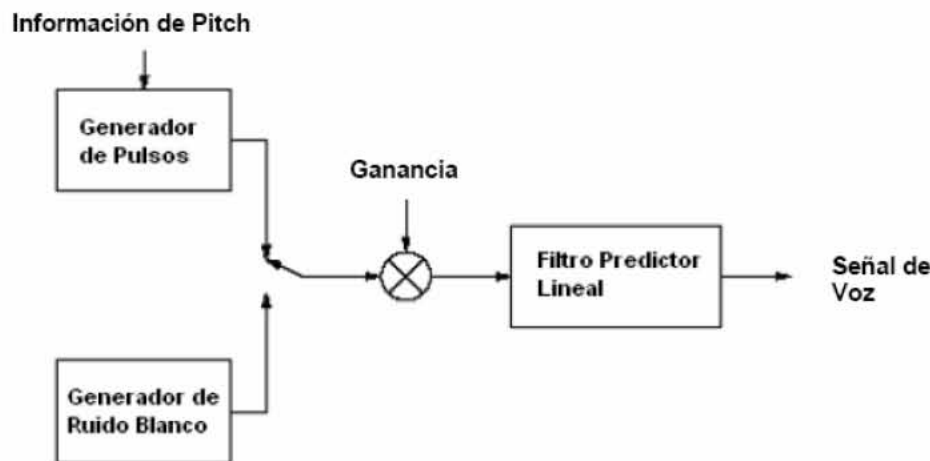


Figura 3-12: Modelo de fuentes y filtros (source/filter model) para la producción de voz

En el modelo lineal de fuentes y filtros para la producción de voz, se usa la teoría lineal de tubos acústicos para obtener un modelado de los filtros digitales, suponiendo una propagación plana de los sonidos.

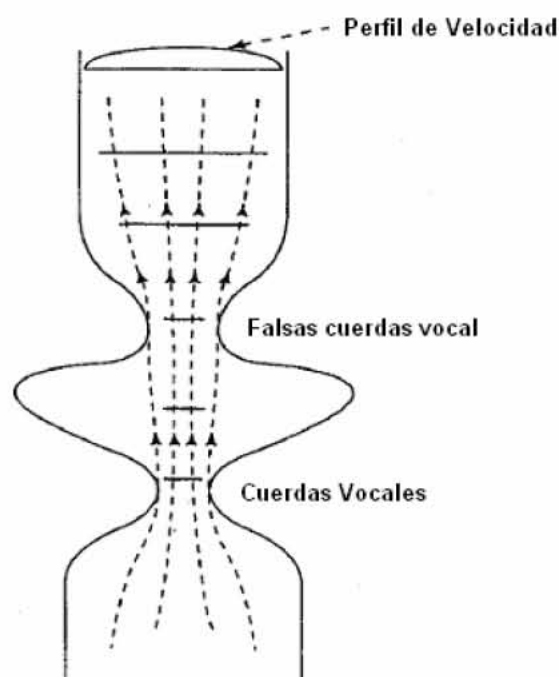


Figura 3-13: Interpretación clásica de la propagación del flujo de aire a través del tracto vocal para la producción de voz. [30]

Aunque este modelo, en principio, pueda resultar suficientemente satisfactorio en ciertas áreas de la Ingeniería de Telecomunicación como codificación y reconocimiento de voz, hay evidencias de que la producción de voz incorpora comportamientos no lineales. En efecto, la fisiología de la laringe sugiere que la idealización de la propagación planar no puede ocurrir [30]. Las medidas de presión sonora y las medidas de variación de volumen que se obtienen de la voz humana se adaptan de forma más realista a un modelo no lineal de fluido de dinámicos tal y como se representa en la Figura 3-14.

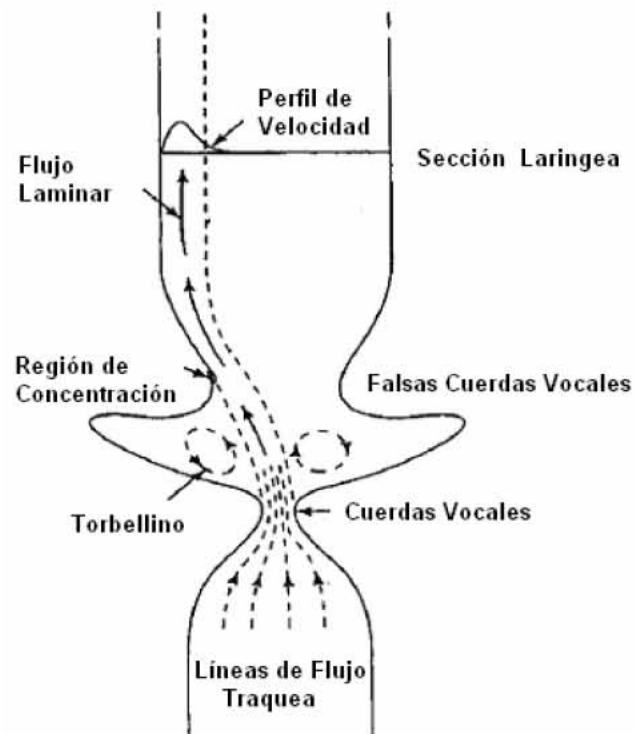


Figura 3-14: Interpretación de la dinámica de fluidos no lineal de la propagación del flujo de aire a través del tracto vocal para la producción de voz [30]

Las cavidades existentes entre las cuerdas vocales y las falsas cuerdas vocales pueden provocar un torbellino el cual excita al tracto vocal durante la fase de cierre de las cuerdas vocales [30], donde la propagación de la onda sonora permite líneas de flujo de aire, con alta velocidad de flujo, la cual en determinadas zonas golpea las paredes de la laringe.

3.5.2 Descripción de Fenómenos Físicos no lineales Implicados

Existen diferentes evidencias para considerar que la voz tiene un comportamiento no lineal. Se han identificado diferentes aspectos de la señal de voz en los que se ha clasificado los resultados de distintos trabajos en los que se describen diferentes fenómenos relacionados con el comportamiento no lineal de la voz, identificándose los siguientes grupos de fenómenos:

- Las Cuerdas Vocales
- Las Turbulencias del flujo de aire
- Propagación no plana de la onda sonora

- Estadísticos de orden superior
- Comportamiento Caótico

En las siguientes secciones se describen cada uno de los diferentes conjuntos de fenómenos.

3.5.2.1 Las Cuerdas Vocales

Hay varios rasgos sobre la oscilación de los pliegues vocales que muestran un comportamiento no lineal. En un modelo lineal la salida es proporcional a la entrada, sin embargo se observa que para una excitación estable se produce una forma de onda variable de la señal de voz debido a la oscilación de los pliegues de las cuerdas vocales, mostrándose no sólo variaciones en la amplitud espectral de la señal de voz, sino además en la frecuencia fundamental.

Se han identifican cuatro tipos de secuencias de *pitch* [42] :

- Tipo 1: Señal de voz prácticamente periódica
- Tipo 2: Señal de voz que tiende a resultar aperiódica, presentando ciertas características de periodicidad
- Tipo 3: La señal de voz resulta periódica pero con una modulación de amplitud
- Tipo 4: La señal de voz resulta con amplitud constante pero con una modulación de frecuencia, donde la señal moduladora presenta dos posibles valores con una relación entre ellas de 0.5.

En el caso de la modulación de frecuencia, los autores de [42] indican que la señal moduladora presenta dos posibles valores con una relación entre ellas de 0.5. A la menor de las frecuencias las denomina subarmónico. Además, en dicho trabajo se apunta a que el efecto de los subarmónicos se manifiesta o bien en la modulación de amplitud o bien una modulación de frecuencia, existiendo la posibilidad de que se produzcan los dos fenómenos.

La presencia de subarmónicos en voces patológicas (pólipos, edemas de Reinke, y disfonías ventriculares) ha sido estudiada [46] , ubicándose la presencia de subarmónicos en las bajas frecuencias. A partir del espectrograma, se detectó que en un

31% de las muestras de voz usadas en el estudio existía presencia de subarmónicos, asociándose dichos subarmónicos a cambios cualitativos en la conducta del sistema vibratorio.

Existen dos posibles teorías para justificar la presencia de subarmónicos; según la teoría de Svec [44] la frecuencia subarmónica se debe a la combinación de dos modos vibracionales (bifonación) cuyas frecuencias tiene la relación 3:2, y por otro lado, según la teoría de Titze [43] la producción de los subarmónicos se debe a asimetrías mecánicas o geométricas entre las cuerdas vocales.

Sin embargo, en [45] se indica que la presencia de bifonación, la presencia de dos frecuencias principales, puede ser inducida por asimetría entre cuerdas derecha-izquierda o por desincronización en la vibración anterior-posterior: en dicho trabajo se demuestran los resultados aplicando técnicas quimográficas. Por lo tanto, las teorías de Svec y Titze, convergen en el origen de los subarmónicos.

En esta línea, los autores de [47] indican la existencia de subarmónicos en las voces de alta calidad, afirmando que existen asimetrías en las fases de la onda mucosa de las cuerdas vocales de la población en general, estimándose en 10.5% de la población utilizada para el estudio, siendo este comportamiento susceptible de ser modelado por modelos de fonación no lineales. Estos comportamientos no lineales, los autores lo asocian con la presencia de subarmónicos. Indica que se manifiestan, estudiando las cuerdas vocales a partir de imágenes estroboscópicas, características periódicas de asimetrías en las fases laterales. En el estudio se añade que las asimetrías no necesariamente indican anormalidades. Indica el estudio además, que desde el punto de vista mecánico, asimetrías en la amplitud de la onda, o aperiodicidades de la onda, implican significativas diferencias en las masas o propiedades viscoelásticas entre las cuerdas vocales. También indican, que algunas lesiones afectan a las propiedades viscoelásticas de una de las cuerdas respecto a la otra, y esto producirá una asimetría o en la fase de cierre entre las cuerdas, o en la amplitud de la apertura o en ambas. Los autores apuntan a lo interesante de determinar el umbral de significancia de las asimetrías (fase, amplitud, o periodo) por alteración de las características físicas de las cuerdas ya que resultará de gran interés en el modelado de los sistemas de fonación y en particular en la detección de manifestaciones de asimetrías resultado de un proceso

patológico. Esto es, diferencias en las propiedades mecánicas de las cuerdas (elasticidad, viscosidad, tensión o forma) afectan al balance entre las fuerzas aerodinámicas en la glotis, caracterizándose la fonación por la excesiva presencia de ruido, subarmónicos, o la completa ruptura de la estructura armónica. En otras palabras, la presencia de subarmónicos presupone que las dos cuerdas vibran asimétricamente, o que un segundo modo de vibración coexiste (en función de la teoría utilizada para la explicación).

Otro aspecto interesante es el efecto de la viscosidad de la mucosa de las cuerdas vocales en aspectos como su vibración, estudiado por [48] . En dicho trabajo se apunta a que la presencia de mucosa en la superficie de las cuerdas vocales genera tensión superficial y causa adhesión, lo cual es una fuente de vibraciones no lineales en las cuerdas vocales, y por lo tanto propone la necesidad del uso de modelos no lineales para explicar el comportamiento de las cuerdas vocales.

Otras teorías, como la propuesta por [41] , apuntan a que las cuerdas vocales muestran comportamientos no lineales ya que en la transición de una fonación sorda a una fonación sonora se pasa por un estado de equilibrio. Además, los autores destacan la calidad de los modelos de masas múltiples, los cuales incluyen no linealidades que se acoplan entre los elementos de masas. Ésto se basa en el conocimiento del cartílago que constituye la laringe el cual tiene un comportamiento no lineal al estirarse.

3.5.2.2 Turbulencias del flujo de aire

Cuando se produce una fonación sorda se produce una constricción en el tracto vocal el cual ocasiona turbulencias. La turbulencia es un efecto no lineal el cual ocurre debido a una interacción entre el flujo aéreo y el acústico. Hay evidencias suficientes para considerar que el ruido en esa cavidad, es decir la turbulencia, es caótica, considerándose el mismo fenómeno que en los sonidos fricativos.

3.5.2.3 Propagación no plana de la onda sonora

Usualmente, el modelo del tracto vocal asume que la propagación de la onda sonora se realiza por medio de una propagación plana. Algunos autores, como por ejemplo [49] ,

sugieren que la propagación no es plana debido a la presencia de torbellinos en la producción de la excitación del tracto, pudiéndose producir torbellinos periódicos.

3.5.2.4 Estadísticos de Orden Superior (HOS)

Tradicionalmente se ha modelado la voz, en particular el tracto vocal, mediante sistemas lineales como por ejemplo por medio de modelos basados en los LPC, pero hay evidencias [31] de que el modelo de producción de voz es un sistema altamente no lineal.

El uso de técnicas basadas en estadísticos de alto orden permite identificar alinealidades en el proceso de producción de voz, lo que a su vez permite modelar el proceso de producción de voz de forma más realista. El uso de estadísticos de tercer orden (*cumulants* de tercer orden y el biespectro) permite obtener información sobre el *skewness* de la señal de voz y de la presencia de alinealidades de segundo orden.

Teniendo en cuenta las propiedades que presenta el biespectro, estas permiten utilizar a la señal de voz, en particular a la voz sonora, para analizar la relación de fase y las características de la estructura armónica. Cuestiones éstas, que no pueden ser estudiadas por medio de un análisis convencional, es decir, mediante el análisis espectral basado en transformadas de Fourier de tramas de corta duración.

La principal dificultad a la hora de estimar el biespectro estriba en que la señal de voz no es una señal estrictamente estacionaria, aunque puede considerarse estacionaria a intervalos cortos de duración, y por este motivo la señal de voz suele modelarse por medio de modelos obtenidos a partir de una segmentación temporal de la voz en “tramas de corta duración”. Las duraciones propuestas de estas tramas varían: se propone una duración comprendida entre 16 y 25 milisegundos [33], también de 37,5 milisegundos con un sobreinventanado del 50% [34], o la duración es de 15 milisegundos [35], mientras que otros autores proponen una duración de tres veces el *pitch* [36].

El tracto vocal se puede modelar mediante un modelo autorregresivo (AR) [37] [38]. El módulo del espectro del modelo AR basado en *cumulants* de tercer orden es

prácticamente inmune al ruido Gaussiano (blanco y coloreado), no siendo así para el módulo del espectro del modelo AR estimado a partir de la autocorrelación [37] . El modelo AR basado en *cumulants* supone que la voz sigue una función de distribución no Gausiana, siendo esto más cierto para sonidos de vocales que para sonidos fricativos.

El biespectro de la señal de voz es una potente herramienta para la identificación de locutores [36] [39] . La magnitud y fase del biespectro son distintas para cada locutor, por ejemplo, al pronunciar una misma vocal. También se podría reconocer vocales con la información del biespectro. Debido a que tanto el espectro como el biespectro es capaz de modelar determinados comportamientos de la excitación glótica como por ejemplo ruido en altas frecuencias o el ruido alrededor del primer formante, es posible utilizar estos modelos espectrales para cuantificar la calidad de la voz.

Resulta interesante el incluir en la fase de clasificación, información de medidas de distancias tanto de la magnitud como de la fase del biespectro. Técnicas de reconocimiento de locutores [39] basadas en *bicepstrum* presentan mejores resultados que los obtenidos con *cepstrum*, siendo mejores los resultados cuando existe ruido aditivo Gaussiano.

La manipulación del biespectro presenta ciertos inconvenientes, en el reconocimiento basado en tramas, debido a que es una función de dos variables frecuenciales, el coste computacional es elevado, la varianza del estimador del biespectro es mayor que la varianza del estimador del espectro y en el caso de utilizar bancos de filtros para el reconocimiento, el cálculo de los filtros tiene un elevado coste computacional, además de una gran dificultad para ser modelados. Se ha propuesto el uso de *cumulants* unidimensionales [40] , lo que se consigue mediante poliespectros integrados que dan lugar a biespectros y triespectros unidimensionales. Esta característica permite la utilización de bancos de filtro con escala “Mel” en sistemas de reconocimiento de señales ruidosas.

3.5.2.5 Comportamiento Caótico

Una vez se apuntó que la señal de voz no es lineal [49] , se planteó la cuestión de si podría ser de naturaleza caótica o fractal. En esta línea se ha estudiado la dimensión

fractal, por parte de diferentes autores, coincidiendo entre ellos en la baja dimensionalidad de la señal de voz de alta calidad. Además, las conclusiones obtenidas a partir del Espectro de Lyapunov indican que es posible diferenciar entre vocales y fricativas, atribuyéndosele un comportamiento no lineal a las vocales y caótica a las fricativas.

En [50] se afirma que la cantidad de alinealidades en el sistema vocal es una importante determinación de la presencia de fonación anormal, representando relaciones no lineales entre ciertas medidas de calidad de la voz y el correspondiente comportamiento físico de las cuerdas (tensión en las cuerdas). Por otro lado, se indica que un determinado valor de una medida puede corresponder a diferentes comportamientos físicos de las cuerdas. Además, propone que las patologías laríngeas sean clasificadas en función del grado de no linealidad, apuntando a lo interesante de utilizar la mínima dimensión del espacio de fases para caracterizar la voz. De hecho, propone una escala para clasificar las disfunciones vocales basada en la dimensionalidad del *attractor*:

- Disfunciones Hipodimensionales : pocas dimensiones
- Disfunciones Eudimensionales : un número medio
- Disfunciones Hiperdimensionales : muchas dimensiones

En este trabajo, en [50] , se apunta que la información que aporta la dimensionalidad puede estar relacionada con la cantidad de masa de las cuerdas puestas en juego.

Han sido usadas otras medidas del comportamiento no lineal basadas en técnicas de modelado de señales caóticas tales como los coeficientes de Lyapunov [51] . Además, se ha utilizado la dimensión de correlación [53] [54] como parámetro con el que se puede distinguir voces sanas de patológicas. También se ha utilizado la dimensión fractal [55] con el fin de discriminar entre voces de un locutor sano y voces de locutores con patologías en el sistema fonador.

Según [52] , se indica que la combinación de la vibración de las cuerdas vocales, con el efecto de la presión subglótica (que da lugar a los resonadores sub y

supraglóticos), las propiedades visco-elásticas de las cuerdas, y el efecto Bernoulli, da lugar a comportamientos no lineales, ocasionando señales subarmónicas, diferentes tipo de modulaciones y señales caóticas. Apuntando además que estas características del movimiento de las cuerdas vocales: desincronizaciones de la cuerda derecha con la izquierda o la desincronización horizontal y vertical del movimiento de cada una de las cuerdas, dan lugar a comportamientos no lineales.

3.6 Fenómenos físicos implicados en la calidad de la voz

En los capítulos anteriores se ha realizado un recorrido por los diferentes dominios de representación en los que se puede estudiar la señal de voz, describiendo las diferentes manifestaciones, propias de cada dominio, que permiten valorar la calidad de la voz. Cada una de las diferentes manifestaciones mencionadas, caracteriza un fenómeno físico que interviene en la generación de la voz y sin embargo el mismo fenómeno físico puede manifestarse en diferentes dominios, en forma de diferentes manifestaciones. En este apartado del capítulo se va a identificar el conjunto de fenómenos físicos necesarios describir para realizar una correcta documentación de la calidad de voz de un locutor.

En este trabajo se han identificado, siendo ésta una de las aportaciones en esta tesis doctoral, cuatro fenómenos físicos cuantificables que permiten caracterizar la calidad de la voz. Los cuatro fenómenos físicos identificados son los siguientes:

- **Estabilidad de la voz:** es la capacidad de un locutor de producir, durante la producción de un sonido sonoro sostenido, un flujo de aire con una intensidad constante para excitar las cuerdas vocales. Este fenómeno físico se cuantifica a partir de medidas de estabilidad de la señal de voz.
- **Riqueza espectral:** se define como la capacidad de generar, durante la fonación de un sonido sostenido, un movimiento periódico de las cuerdas vocales y de producir una excitación sonora del tracto vocal con gran cantidad de componentes espectrales. Este fenómeno físico se cuantifica por medio del cálculo de la estabilidad de la frecuencia de *pitch* y por la presencia de armónicos con energía elevada en las diferentes bandas de frecuencias.

- **Presencia de ruido:** se define como la presencia de ruido glótico en la voz, durante la fonación de un sonido sonoro sostenido, debido a carencias en la fase de cierre de las cuerdas vocales. Este fenómeno físico se cuantifica por medio de la presencia de ruido no estacionario en la voz.
- **Comportamiento no lineal:** este fenómeno físico es ocasionado por un funcionamiento anómalo de las cuerdas vocales, debido a irregularidades en las masas de cuerdas vocales que interviene en la fase de cierre de cuerdas, existencia de asimetrías en el movimiento de las cuerdas y factores relacionados con la mucosa de las cuerdas. Estos fenómenos son cuantificables por medio de la identificación de comportamientos no lineales en la señal de voz.

Una voz de calidad anormal presenta al menos uno de los valores correspondientes a la cuantificación de los cuatro fenómenos físicos fuera de los rangos de normalidad. Este procedimiento de cuantificación de la calidad de la voz permite identificar calidades de voz anómalas de diferente origen, es decir, la voz de un locutor con baja calidad debida a un estado incipiente de una patología orgánica presentará probablemente la alteración de la “Estabilidad de la voz” mientras que otro locutor con un estado incipiente de una patología funcional presentará valores anormales de “Presencia de ruido”. A medida que la calidad de la voz de un locutor empeora aumentará el número de fenómenos físicos y la desviación de los valores de la cuantificación de dichos fenómenos respecto a los valores de normalidad.

Estos cuatro diferentes fenómenos físicos son cuantificables en diferentes dominios de representación de la señal de voz, existiendo diferentes medidas objetivas de la calidad de la voz que cuantifican con mayor o menor precisión un mismo fenómeno físico.

En capítulos posteriores se realizará una revisión del estado de la técnica de las diferentes medidas objetivas de calidad de la voz que cuantifican cada uno de los diferentes fenómenos físicos. Para cada medida se presenta una descripción y una valoración de su calidad para cuantificar el fenómeno físico que mide. Sin embargo, resulta necesario previamente, describir las muestras de voz que serán utilizadas para evaluar cada una de las medidas (capítulo 4). También, resulta necesario realizar una

revisión de los diferentes métodos que permiten evaluar las medidas de calidad de la voz, así como la descripción del proceso de evaluación de las diferentes medidas (capítulo 5).

3.7 Referencias

- [1] Hirano, M., *Clinical Examination of Voice*. New York, Springer-Verlag, 1981
- [2] Dejonckere, P. H. Remacle, M. Fresnel-Elbaz, E. Woisard, V. Crevier-Buchman, L. Millet, B., "Differentiated perceptual evaluation of pathological voice quality: reliability and correlations with acoustic measurements". *Revue de Laryngologie Otologie Rhinologie*, vol. 117, no.2, pp. 219-224, (Bordeaux) 1996.
- [3] Hammarberg, B. y Gauffin, J., Perceptual and acoustic characteristics of quality differences in pathological voices as related to physiological aspects,. in O. Fujimura & M.Hirano (eds.), *Vocal Fold Physiology*, pp 283-303, 1995
- [4] Laver, J., *The Gift of Speech*, Edinburgh University Press, 1991
- [5] Jacques Koreman y Manfred Pützer, "Finding Correlates of Vocal Fold Adduction Deficiencies", *Phonus* 3, Institute of Phonetics, University of the Saarland, vol. 3, pp.155-178, 1997.
- [6] Picone, J. W., "Signal modeling techniques in speech recognition". *Proceedings of the IEEE*, vol.81, no.9, pp.1215–1247, Sept. 1993.
- [7] Athiana P. Petropulu y Chrysostomos L. Nikias, "The Complex Cepstrum and Bicepstrum: Analytic Performance Evaluation in the Presence of Gaussian Noise", *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol.38, no.7, pp. 1246–1256, Julio 1990.
- [8] Athiana P. Petropulu y Chrysostomos L. Nikias, "Analytic Performance Evaluation of the Bicepstrum", *IEEE Transactions on Acoustics, Speech and Signal Processing*, ICASSP-89, vol.4, pp.2337-2340, 1989.
- [9] Lam, K.F.; Chan, C.F., "MBE synthesis of speech coded in LPC format", *International Conference on Acoustics, Speech, and Signal Processing*. ICASSP-95, vol.1, no.1, pp. 668 – 671, 1995.
- [10] J.L. Flanagan, *Speech Analysis, Synthesis and Perception, Second*, Expanded Edition, Ed. Springer, 1983.
- [11] Helmer Strick, "Automatic parametrization of voice source signal: a novel evaluation procedure is used to compare methods and test the effect of low-pass filtering", Internal report, Dept. of Language and Speech, University of Nijmegen, [<http://lands.let.kun.nl/TSPublic/strik/>], 1997
- [12] Donald G. Childers y Chun-Fan Wong, "Measuring and Modeling Vocal Source-Tract Interaction", *IEEE Transactions on Biomedical Engineering*, vol.41, no.7, pp.663-670, July 1994.
- [13] Krishnamurthy, A, "Two channel (speech and egg) analysis for formant and glottal inverse filtering", *Acoustics, Speech, and Signal Processing*, ICASSP '84, vol.9, pp.60 - 63 , Mar 1984.
- [14] Paavo Alku, "An Automatic method to estimate the Time-Based Parameter of the Glottal Pulseform", *Acoustics, Speech and Signal Processing*, ICASSP-92, 1992:2:29-32.
- [15] Wong, D.; Markel, J.; Gray, A., Jr .,"Least squares glottal inverse filtering from the acoustic speech waveform", *IEEE Transactions on Acoustics, Speech, and Signal Processing* , vol.27, no.4, pp. 350– 55, 1979
- [16] Godino-Llorente, J. I, Aguilera-Navarro, Santiago , Hernández-Espinosa, Carlos, Fernández-Redondo, Mercedes ,Gómez-Vilda, Pedro, "On the selection of meaningful speech parameters used by a pathologic/non pathologic voice register classifier". *Sixth*

European Conference on Speech Communication and Technology (EUROSPEECH'99), pp. 563-566, Budapest, Hungary. September-1999.

[17] Hernández-Espinosa, C.; Fernández-Redondo, M.; Gómez-Vilda P. "Diagnosis of Vocal and Voice Disorders by the Speech Signal ". *IEEE-INNS-ENNS International Joint Conference on Neural Networks*, vol.4, pp. 253-258, 2000.

[18] Jo, C.; Kim, K.; Kim, D.; Wang, S. "Screening of Pathological Voice from ARS using Neural Network". *Pathological voice; ARS; Neural Networks*. 2001.

[19] Salvatore, A.P.; Thome, N.; Gorss, C.M.; Cannito, M.P., "Neural network approach to speech pathology", *42nd Midwest Symposium on Circuits and Systems*, 1999, vol.439 – 442, 999.

[20] Umapathy, K.; Krishnan, S.; Parsa, V.; Jamieson, D.G. "Discrimination of pathological voices using an adaptive time-frequency approach". *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol.4, pp.3852 -3855, 2002.

[21] Alonso, J. B.; de León, J.; Alonso, I.; Ferrer, M. A. "Automatic Detection of pathologies in the voice by Hos based parameters". *Eurasip journal on Applied signal procesing*, vol.2001, no. 4, pp.275-284, 2001

[22] Marcelo de Oliveira Rosa, Carlos Pereira, Marcos Grellet, "Adaptive estimation of residue signal for voice pathology diagnosis", *IEEE Transactions on biomedical Engineering*, vol.47, no.1, pp.96-104, 2000.

[23] J.L.Flanagan, *Speech Analysis Synthesis and Perception*, Ed. Springer-Verlag Heidelberg New York, 2ed, 1972.

[24] Eric Keller, *Fundamentals os Speech Synthesis and Speech Recognition*, Ed. Wiley, 1994.

[25] Maria Rangoussi, Anastasios Delopoulos, Michail Tsatsanis, "On the use of Higher-order statistics for robust endpoint detection of speech", *proc IEEE*, 1993

[26] Ali Alkulaibi, J.J. Soraghan, T.S. Durrani, "Fast HOS simultaneous Voiced/Unvoiced and Pitch Estimation Using 3-level Binary Speech Signals", *Proceedings of 8th IEEE Signal Processing Workshop on Statistical Signal and Array Processing*, 1996, pp. 194 - 197 , June 1996

[27] R. Sankar, "Pitch extraction algorithm for voice recognition applications", *Proceedings of the Twentieth Southeastern Symposium on System Theory*, pp.384–387, March 1988.

[28] Edward L. Riegelsberger y Ashok. Krishnamurthy, "Glottal source estimation,: methods of appying the LF-model to inverse filtering", *proc. IEEE*, 1993.

[29] Jackson Menaldi, *La voz normal*, Editorial medica panamericana, 1992.

[30] Hansen, J.H.L.; Gavidia-Ceballos, L. y Kaiser, J.F. "A nonlinear operator-based speech feature analysis method with application to vocal fold pathology assessment". , *IEEE Transactions on Biomedical Engineering*, vol.45, no.3, pp.300-313, March 1998.

[31] J. W. A. Fackrell y S. McLaughlin, "The Higher-Orden Statistics of Speech Signals", *Techniques for Speech Processing and their Application*, pp.711-716, 1994.

[32] Boyan Boyanov, Stefan Hadjitodorov y Ognyan Krumov, "Analisís of Voiced Speech by means of Bispectrum", *Communications and Signal Processing*, COMSIG-90, pp.2267-2268, 1990

[33] Ralph Fulchiero y Andreas S. Spanias, "Speech Enhancement Using the Bispectrum", *Acoustics, Speech and Signal Processing*, ICASSP-93, vol.2, pp.367-370, 1993.

[34] Asunción Moreno, Sergio Tortola, Josep Vidal y J.A.R. Fonollosa, "New Hos-Based Parameter estimation Methods for Speech Recognition in Noisy Environments",

International Conference on Acoustics, Speech, and Signal Processing, ICASSP-95., vol.1, pp. 429 – 432, 1995.

[35] Matthias Frohlich, Dirk Michaelis y Hans Werner Srube, “Acoustic ‘Breathiness Measures’ in the Description of Pathologic Voices”, *Acoustics, Speech and Signal Processing*, ICASSP-98, vol.2, pp.937-940,1998.

[36] Boyan Boyanov, Stefan Hadjitodorov y Ognyan Krumov, “Analisis of Voiced Speech by means of Bispectrum”, *Communications and Signal Processing*, COMSIG-90, vol.1, pp.2267-2268,1990.

[37] K. K. Paliwal y M. M. Sondhi, “Recognition of Noisy Speech Using Cumulant-Based Linear Prediction Analisis”, *Acoustics, Speech and Signal Processing*, ICASSP-91, vol.1, pp.429-432,1991.

[38] S. Seetharaman y M. E. Jernigan, “Speech Signal Reconstruction Based on Higher Orden Spectra”, *Acoustics, Speech and Signal Processing*, ICASSP-88, vol.1, pp.703-706, 1988.

[39] Stanley Wenndt y Sanyogita Shamsunder, “Bispectrum Features for Robust Speaker Identification”, *Acoustics, Speech and Signal Processing*, ICASSP-97, vol.2, pp.1095-1098,1997.

[40] Asunción Moreno y Miquel Rutllán, “Integrated Polispectrum on Recognition”, *Spoken Language*, ICSLP-96, vol.2, pp.1281-1284, 1996.

[41] Hansen, J.H.L.; Gavidia-Ceballos, L.; Kaiser, J.F. "A nonlinear operator-based speech feature analysis method with application to vocal fold pathology assessment". *IEEE Transactions on Biomedical Engineering*, vol.45, no. 3, pp.300-313, March 1998.

[42] Xuejing Sun and Yi Xu, “Perceived Pitch of Synthesized Voice with Alternate Cycles”, *Journal of Voice*, vol.16, no. 4, pp. 443-459, 1995

[43] Tiltze, IR., *Principles of Voice Production*, Englewood Cliffs, NJ: Prentice-Hall, Inc, 1994 .

[44] Svec JG, Schutte HK, Miller DG.“ A Subharmonic vibratory pattern in normal vocal folds”, *Journal of Speech and Hearing Research*, vol.39, no.1, pp.135-143,1996.

[45] Jürgen Neubauer, Patrick Eysholdt, Ulrich Eysholdt, Hanspeter Herzel, “Spatio-temporal analysis of irregular vocal fold oscillations: Biphonation due to desynchronization of spatial models”, *J. Acoustical Society of America*, vol.110, no. 6, pp.3179-3192, December 2001.

[46] Nuñez Batalla F, Suarez Nieto, C., Muñoz Pinto, C., Baragaño Rio, L., Álvarez Zapico, M.J., Martínez Ferreras, A., “Estudio Espectrográfico de la disfonía: Subarmónicos”, *Acta Otorrinolaring. Esp.*, vol.52,no. 1, pp.52-56, 2000.

[47] C. Michael Haben, Karen Kost, George Papagiannis, “Lateral Phase Mucosal Asymmetries in the Clinical Voice Laboratory”, *Journal of Voice*, vol.17,no. 1, pp.3-11,2003.

[48] Stephane Ayache, Maurice Ouaknine, Philippe Dejonkere, Pierre Prindere, Antoine Giovanni, “Experimental Study of the Effects od Surface Mucus Viscosity on the Glottic Cycle”, *Journal of Voice*, vol.18, no.1, 2004.

[49] H. M. Teager and S. M. Teager, “Evidence of nonlinear sound production mechanisms in the vocal tract,” in *Proceedings of the NATO Advanced Study Institute on Speech Production and Modelling*, (Bonas, France), pp.241–261,July 1989.

[50] Robert F. Orlikoff, R. J. Baken, “Curing Diagnosis: Improving the Taxonomy of Phonatory Dysfunction”, *Sixth Conference on Advances in Quantitative Laryngology*. Hamburg, Germany, April 2003

[51] Ping Yu, Maurice Ouaknine, Joana Revis, Antonine Giovanni, “Objective Voice Analysis for Dysphonic Patients: A Multiparametric Protocol Includinf Acoustic and Aerodynamic Measurements”, *Journal of Voice*, vol.15, no. 4, pp.529-542, 2001.

- [52] Wilden, H. Herzel, G. Peters, G. Tembrock, "Subharmonics, Biphonation and Deterministic Chaos in Mammal Vocalization", *The International Journal of Animal Sound and its Recording, Bioacoustics*, vol.9, pp.171-196, 1998.
- [53] Yu Zhang y Jiang J.J., "Nonlinear dynamic analysis in signal typing of pathological human voices", *Electronics Letters*, vol.39, no. 13, June 2003.
- [54] Jiang J.J. y Yu Zhang, "Nonlinear dynamic analysis of speech from pathological subjects", *Electronics Letters*, vol.38, no.6, March 2002.
- [55] Accardo A., Fabbro F. y Mumolo E., "Analysis of normal and pathological voices via short-time fractal dimension", *Proceedings of the Annual International Conference of the IEEE of Engineering in Medicine and Biology Society*, vol.14, no.3, 1270-1271, 1992.
- [56] Boyan Boyanov, Stefan Hadjitodorov y Ognian Krumov, "Analysis of Voiced Speech by means of Bispectrum", *Communications and Signal Processing*, COMSIG-90, pp. 2267-2268, 1990.
- [57] M. E. Lee, "Acoustic models for the analysis and synthesis of the singing voice", Doctoral Thesis, School of Electrical and Computer Engineering, Georgia Institute of Technology, U.S.A., March 2005.
- [58] H. Fujisaki, M. Ljungqvist, "Proposal and evaluation of models for the glottal source waveform", *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, Tokyo, vol. 4, pp. 1605-1608, 1998.
- [59] R. Veldhuis, "A computationally efficient alternative for the Liljencrants-Fant model and its perceptual evaluation", *Journal of the Acoustical Society of America (JASA)*, vol. 103, pp. 566-571, 1998.
- [60] Rosenberg A. E., "Effect of Glottal Pulse Shape on the Quality of Natural Vowels", *Journal Acoustical Society of America*, Vol. 49, No.2(part 2), 1971.
- [61] D. Michaelis, M. Fröhlich, H. W. Strube, E. Kruse, B. Story, I. R. Titze, "Some simulations concerning jitter and shimmer measurement", *Proceedings of the 3rd International Workshop on Advances in Quantitative Laryngoscopy*, 1998.

CAPÍTULO 4

Bases de datos

Para poder evaluar adecuadamente los resultados que se presentan en este trabajo, resulta necesario disponer de una base de datos de voces con diferentes calidades. En este capítulo de la memoria, se analizan diferentes aspectos que resulta necesario tener en cuenta a la hora de utilizar una base de datos de voces y se finaliza presentando las bases de datos utilizadas en este trabajo.

Ante la necesidad de una base de datos de voces con diferentes calidades, se ha utilizado las grabaciones de locutores con patologías en el sistema de producción de voz como muestras de baja calidad de voz y como muestras de voz con alta calidad grabaciones obtenidas de locutor sin dichas patologías.

Las bases de datos de las muestras de voces de diferentes calidades de voz utilizadas en este trabajo han permitido estudiar sus características, tanto las caracterizables por medio de modelos lineales como las caracterizables por modelos no lineales. Para poder generalizar las conclusiones obtenidas a partir de la base de datos, resulta necesario que la base de datos de voz disponga de suficiente número de voces de cada tipo.

Además, a partir de las bases de datos de las muestras de voces de diferentes calidades de voz ha sido posible implementar diferentes aplicaciones basadas en sistemas de clasificación automática de voz. Los sistemas de clasificación automática, de forma general necesitan para su correcto funcionamiento, una fase de entrenamiento, la cual se genera en el bloque de “clasificación”. La fase de entrenamiento consiste en “enseñar” al clasificador a diferenciar entre distintas unidades de clasificación. Para este fin es necesario disponer, en la fase de entrenamiento, de suficientes muestras de las distintas unidades de clasificación.

Por lo tanto, en el contexto de la problemática presentada en esta tesis doctoral, por un lado resulta importante para la fase de evaluación de las diferentes medidas de la calidad de la voz el disponer de las suficientes muestras de voces con el fin de poder extraer conclusiones lo más generales posibles, y por otro lado, para las diferentes aplicaciones de clasificación automática de voz que se presentan en este estudio, resulta fundamental que la base de datos sea suficientemente grande para conseguir una mayor “generalización” en el proceso de entrenamiento.

No sólo es importante el número de muestras de voces, además si la base de datos presenta diferentes tipos de voces resulta necesario que los diferentes tipos estén balanceados, es decir, que exista un número similar de unidades de cada tipo para que no predomine un determinado tipo tanto en la caracterización como en el entrenamiento.

Al tratarse de bases de datos de voces serán igualmente importantes otros aspectos tal como es el protocolo de grabación de las voces o las condiciones acústicas en el proceso de grabación.

En el siguiente apartado se estudia las diferentes consideraciones que han de tenerse en cuenta a la hora de crear o utilizar una base de datos de voces.

4.1 Características de las bases de datos de voces

En el contexto de este trabajo, el estudio de las medidas de la calidad de la voz a partir de grabaciones de voz, resulta fundamental la calidad de las grabaciones ya que la valoración de la calidad de la voz se estima únicamente a partir de la grabación. Si la grabación de voz no cumple unos mínimos de calidad, cualquier estimación corre el peligro de ser errónea.

4.1.1 Influencia del Proceso de Grabación

Es importante la correcta elección de los sistemas *hardware* de adquisición (sistemas digitalizadores) de señal de audio pues éstos pueden afectar a las medidas de calidad de la voz.

Los sistemas de adquisición del tipo tarjetas de sonido de propósito general son potencialmente aceptables, siempre y cuando se utilice un preamplificador que adecúe el rango dinámico [1] . Por otro lado el error de discretización afecta a los valores de determinadas medidas, por lo que suele ser común utilizar dispositivos de calidad profesional [1] [2] , ya que los componentes de bajo costo son vulnerables a las condiciones ambientales en lo referente a interferencias electromagnéticas, vibraciones mecánicas, humedad y temperatura. Además, los sistemas de adquisición profesionales son más vulnerables al ruido, es esta virtud la que los capacita para obtener medidas más exactas de calidades de voz.

También resulta importante la sensibilidad de los sistemas de adquisición y principalmente la dependencia del tipo de micrófono utilizado, recomendándose [1] los micrófonos que se sostienen sobre la cabeza y que permiten una distancia constante hacia la boca.

En la influencia del micrófono, destaca su comportamiento en frecuencia y por lo tanto, destaca el efecto de la distorsión frecuencial. Los mejores resultados se obtienen con micrófonos capacitivos frente a los micrófonos dinámicos [1] [2] . Existe una variabilidad en las medidas de calidad de la voz en función de la distorsión frecuencial, estimando que el error que se comete en las medidas, cuando no se tiene

compensado el efecto de la distorsión espectral, está comprendido entre el 0.7% y el 5% [2] . Además, resulta importante tener en cuenta el ancho de banda del micrófono.

4.1.2 Influencia de la Presencia de Ruido

Determinas medidas de la calidad de la voz resultan más vulnerables al ruido, recomendando unos autores [3] un nivel de relación de calidad entre la voz y el ruido superior a 25dB, aunque otros autores [4] son más estrictos e indican que la relación señal ruido debe de ser de al menos 30 dB para producir unos resultados válidos, recomendándose un nivel de 42 dB y especial cuidado en evitar los ruidos no estacionarios.

4.1.3 El Contenido

En la práctica totalidad de los trabajos sobre sistemas automáticos de medidas objetivas de la calidad de la voz a partir de una grabación, se utiliza una grabación de un sonido sonoro sostenido como muestra de voz para el análisis. El motivo es que durante la fonación de un sonido sonoro sostenido se garantiza que el sistema fonador pone en movimiento a las cuerdas vocales. Este hecho es importante, ya que cualquier anomalía en este movimiento de las cuerdas vocales se plasmará en el resultado del audio producido.

De igual forma, casi en la totalidad de estos trabajos, los sonidos sonoros sostenidos consisten en vocales sostenida. Respecto a qué vocal utilizar, existen diferentes estrategias: usar una única vocal o usar múltiples vocales. En el caso de usar múltiples vocales resulta necesario tener en cuenta los resultados presentados en el trabajo [4] , donde el autor demuestra la dependencia de las medidas de la calidad de la voz con la vocal que se pronuncia.

También existen bases de datos en la que se incluye habla continua de las muestras de voz, en las que se recurre a pasajes típicos o frases concienzudamente estudiadas, a partir de las cuales es posible identificar características del locutor.

4.1.4 Cantidades y Balanceadas

La base de datos se utilizará con diferentes propósitos, donde para cada caso existe un valor óptimo de número de muestras de voces. A continuación se describen los diferentes ámbitos de utilización de dicha base de datos:

- Con el propósito de realizar un estudio cuantitativo de las características de la señal de voz, desde la perspectiva de diferentes dominios de representación de la señal de voz y con el fin de identificar los diferentes fenómenos físicos que intervienen en la caracterización de la calidad de la voz, resulta necesario disponer del mayor número de muestras posibles para poder extraer conclusiones con un carácter general.
- Con el propósito de analizar el comportamiento de las diferentes medidas de la calidad de la voz, y con el fin de garantizar la exactitud desde el punto de vista estadístico, resulta necesario un número significativamente alto de muestras que permita obtener resultados lo más consistentes posibles.
- Con el objetivo de implementar diferentes sistemas de clasificación automática en el ámbito de la voz, resulta importante el número de muestras empleadas en el proceso de entrenamiento y test, en especial en la fase de entrenamiento. Por lo tanto, resulta interesante disponer del mayor número de muestras posibles con el fin de obtener una correcta generalización en el entrenamiento, y evitar problemas como el sobreentrenamiento de un conjunto pequeño de muestras.

Por lo tanto, por los motivos expuestos, resulta necesario disponer de suficientes muestras de voz. Además del número de muestras, resulta necesario que la base de datos se encuentre balanceada, en el sentido que las diferentes unidades de clasificación presentes en la base de datos se encuentren en un porcentaje equitativo de tal forma que resulten representativos respecto a la totalidad de la base de datos.

Respecto a las diferentes unidades en las que se puede dividir la base de datos existen diferentes estrategias:

- Diferenciar entre voces sanas y voces representativas de cada una de las diferentes patologías del sistema fonador. Esta estrategia consiste en tener el mismo número de muestras de sanas e igualmente para cada una de las diferentes patologías. Esta estrategia presenta el inconveniente de que voces correspondientes a diferentes patologías presentan las mismas características acústicas, o el inconveniente de que existen diferentes manifestaciones de una misma patología en función de la ubicación en las cuerdas vocales. Además, es común presentar diferentes patologías vocales simultáneamente.
- Diferenciar entre voces sanas y diferentes grados de disfonía. Teniendo en cuenta que el grado de disfonía se suele etiquetar en leve, moderada y severa, esta estrategia consiste en usar igual número de muestras para voces sanas, voces con disfonías leves, voces con disfonía moderadas y voces con disfonías severas.
- Diferenciar entre voces sanas y voces patológicas. En este sentido, consiste en tener igual número de muestras correspondientes a cada uno de los diferentes grupos.

4.2 Bases de datos empleadas

En este trabajo se utilizarán dos bases de datos, una base de datos de voces propia, con unos criterios propios en lo que respecta a equipamiento a usar y el protocolo de grabación, y una base de datos de distribución comercial.

El motivo de utilizar dos bases de datos está originado en uno de los objetivos de esta tesis doctoral, el cual consiste en obtener un procedimiento de evaluación de la calidad de la voz en el que la elección de las medidas de calidad sea lo más independiente posible de la base de datos utilizada en ese estudio. Por ello, se ha planteado realizar la identificación de las medidas de calidad de la voz con la base de datos propia, y validar el procedimiento de evaluación de la calidad de la voz utilizando la base de datos comercial.

4.2.1 Base de datos Propia

Se ha creado una base de datos en este trabajo la cual es fruto de la coincidencia de dos líneas de investigación, por un lado del que fue inicialmente el Grupo de Procesado Digital de la Señal, Grupo de Investigación de la Universidad de Las Palmas de Gran Canaria, y un grupo de investigadores profesionales facultativos de la otorrinolaringología, miembros del Servicio de Otorrinolaringología de Hospital Doctor Negrín de Gran Canaria. Actualmente, ambos grupos de investigación se han fusionado en el Grupo de Procesado Digital de Señales Biológicas de la Universidad de Las Palmas de Gran Canaria. Fruto de esta colaboración, se ha creado una línea de investigación dirigida al estudio de técnicas para cuantificar la calidad de la voz de un locutor, dando lugar a la creación de una base de datos de grabaciones de voz con diferentes calidades. Las grabaciones de voz correspondientes a calidades anormales, han sido obtenidas a partir de locutores con trastornos laríngeos o del sistema fonador, contemplándose un espectro amplio de patologías con trastorno en la fonación y contemplándose diferentes grados de disfonía.

4.2.1.1 Análisis del material y del protocolo de grabación

Frente a las recomendaciones de [1] ,[2] ,[3] y [4] en las que, resumiendo a groso modo se recomienda el uso de dispositivos de grabación de calidad profesional, micrófonos capacitivos que se sostienen sobre la cabeza y que permiten una distancia constante hacia la boca, y recomendaciones de relaciones señal ruido de 42 dB, en esta tesis doctoral se ha cuestionado ciertos aspectos de estas recomendaciones estándar.

Ante el diseño de las pautas a seguir en la creación de una base de datos, se entiende en este trabajo que resulta fundamental tener en todo momento una visión del objetivo final del estudio. En este trabajo se tiene por objetivo diseñar un procedimiento para la evaluación de la calidad de la voz de uso clínico, el cual, sea implementable y pueda ser usado por el mayor número posible de profesionales clínicos de la voz. Esta pretensión de intentar ampliar lo máximo posible el número de usuarios potenciales presenta varios condicionantes al procedimiento para la evaluación de la calidad de la voz: que el equipamiento sea económico y de fácil adquisición, que las condiciones

acústicas necesarias para el uso no sean muy restrictivas y que el proceso de grabación tenga una complejidad y tiempo de duración lo más reducido posible.

Debido a que el equipamiento, las condiciones acústicas y el procedimiento de grabación de la base de datos condicionarán el uso de los resultados obtenidos de un estudio a partir de dicha base de datos, para la creación de la base de datos propia se han tenido ciertas consideraciones:

- La base de datos se ha grabado utilizando un PC con tarjetas de sonido convencionales. Ésto permite que el producto final se pueda ejecutar sobre cualquier PC de propósito general, presente prácticamente en cualquier entorno clínico.
- Atendiendo a la realidad de los recursos disponibles en los entornos clínicos, en este trabajo se ha considerado que en lo que respecta a las condiciones acústicas sobre las que se debe realizar el proceso de grabación, han de ser condiciones realistas en las que se asegure unos mínimos de calidad.
- Se considera en este trabajo que resulta muy difícil garantizar una correcta ubicación del micrófono y el locutor, por lo que protocolos de grabación muy rígidos darán lugar a grabación que incumple el protocolo y que por lo tanto los resultados estarían condicionados. Por este motivo, en este trabajo, se considera que para la ubicación del micrófono y el locutor debe existir ciertos grados de libertad, de tal forma que los resultados que se obtengan sean independientes de aspectos como la distancia entre el micrófono y el locutor, el ángulo, el modelo del micrófono, etc.
- En lo que respecta al contenido de la grabación de voz, este aspecto toma mayor importancia principalmente en los servicios médicos de masas, donde no es común disponer de un tiempo dilatado por paciente para su evaluación. Además, teniendo en cuenta la diversidad de la población, resulta difícil utilizar textos leídos, ya que circunstancias como baja calidad de lectura debido a causas como bajos niveles culturales o la presencia de otros trastornos que no son frutos de este estudio, hacen difícil la obtención de medidas objetivas coherentes de la calidad de la voz a partir del habla continua. Por este motivo se ha optado por el uso de vocales sostenidas.

Otro hecho del que resulta necesario ser conscientes, en la realidad de la actividad clínica, es la de que los profesionales clínicos disponen de un tiempo limitado para la evaluación de cada paciente, y en muchas ocasiones les resulta imposible la

realización de alguna actividad investigadora. También es necesario tener en cuenta el carácter periódico de las visitas de los pacientes lo que produce que el número de evaluaciones no se corresponda con la de nuevos pacientes. Además, resulta muy poco frecuente encontrar profesionales clínicos dispuestos a colaborar en estudios acentuadamente técnicos. Las consecuencias de esta realidad han producido que la base de datos que se ha creado para este estudio tenga un volumen considerablemente limitado.

4.2.1.2 Características

La grabación de voz ha sido realizada en un centro hospitalario, en una habitación en la que se ha tenido en cuenta los niveles de contaminación acústica, intentándose obtener un compromiso entre muy bajos niveles de ruido, lo cual facilita el estudio de la grabación, y niveles de ruido realistas, ya que si no fuese así resultaría imposible aplicar a las condiciones normales de trabajo cualquier resultado obtenido a partir de la base de datos.

Para el proceso de grabación se ha utilizado un PC multimedia, Pentium III a 550 MHz, con 128 MBytes de memoria RAM, con tarjeta de sonido SoundBaster de 16 bits, y un micrófono de sobremesa modelo VIVANCO MF 15/13166. Se ha utilizado un programa informático desarrollado explícitamente para tal fin (DISFONÍA), el cual facilita el proceso de grabación así como la toma de datos correspondiente: grado de disfonía, patología asociada, sexo y edad. Para la realización del proceso de grabación se recomienda como punto de partida que el locutor se encuentre sentado frente al micrófono, con una distancia entre la boca y el micrófono de 25 centímetros, aproximadamente.

La aplicación utilizada para el proceso de grabación fue desarrollada utilizando el Borland C++ versión 5.03, utilizándose en la fase de grabación la técnica de doble buffer. Esta técnica permite registrar audio durante tiempo indefinido sin perder muestras [8] .

Además, el sistema realiza un estudio cualitativo de la muestra de voz por medio de la representación de diversas gráficas (espectrogramas, espectros, evolución temporal del *pitch*,...) que resultan útiles como herramientas para monitorizar la calidad del audio registrado.

El contenido de las grabaciones corresponde a la fonación de las cinco vocales castellanas de forma sostenida y no susurrada.

Las grabaciones de voz se almacenan en un formato propio de un entorno matemático MATLAB® versión 4. Este formato consiste en audio digital, con una frecuencia de muestreo de 22050 KHz y con una resolución de 16 bits por muestra. Este formato es fácilmente transportable a otros formatos como WAV, AU, etc.

En lo que respecta al balance de la base de datos, en la Tabla 4-1 se presenta los datos pormenorizados por sexos y grados de patologías.

TABLA 4-1 : BALANCE DE LA BASE DE DATOS PROPIA

Grupo de Muestras de Voces		Hombres	Mujeres	Total	
Muestras de voces sanas		43	44	87	
Muestras de voces patológicas	Disfonía leve	7	19	26	93
	Disfonía moderada	22	14	36	
	Disfonía severa	18	13	31	

A continuación, en las siguientes figuras se presentan las características de la base de datos de forma gráfica mediante relaciones porcentuales.



Figura 4-1: Número de muestras de voces sanas y voces patológicas

En lo que respecta al sexo de los locutores de las distintas muestras de la base de datos se observa la distribución en la Figura 4-2.

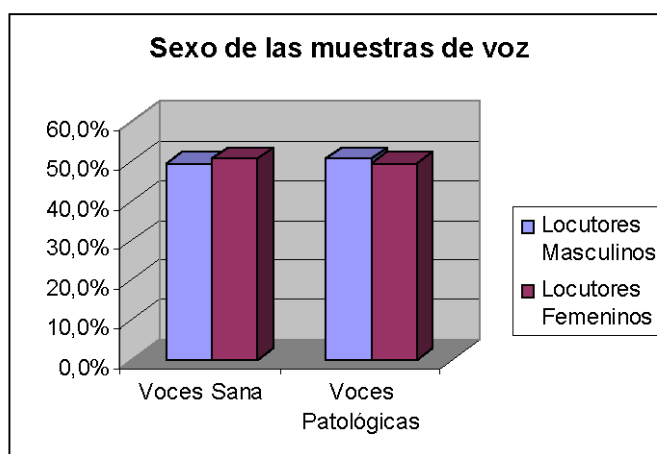


Figura 4-2: Distribución del porcentaje de muestras en función del sexo y su tipo

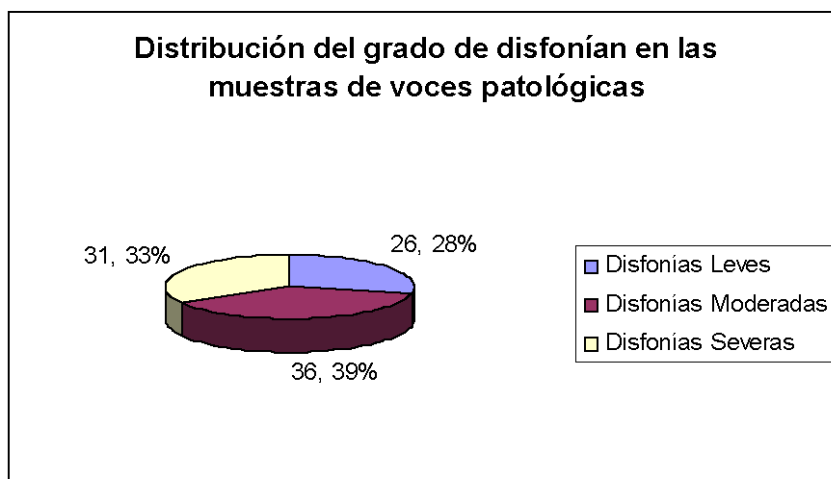


Figura 4-3: Distribución del porcentaje de las muestras patológicas según el grado de disfonía

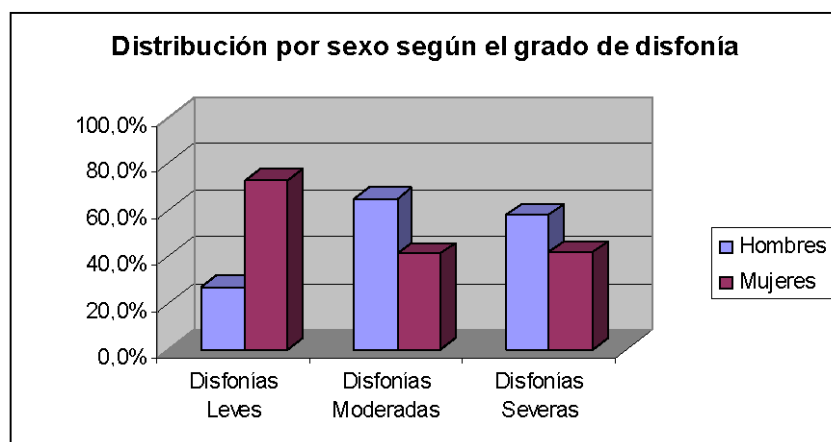


Figura 4-4: Distribución del porcentaje de las muestras patológicas en función del sexo y su grado de disfonía

Las muestras patológicas se han dividido en dos grupos principales: cuerdas sin lesión y cuerdas con lesión. En la Figura 4-5 se muestra una comparativa porcentual del número de muestras patológicas que componen cada grupo.

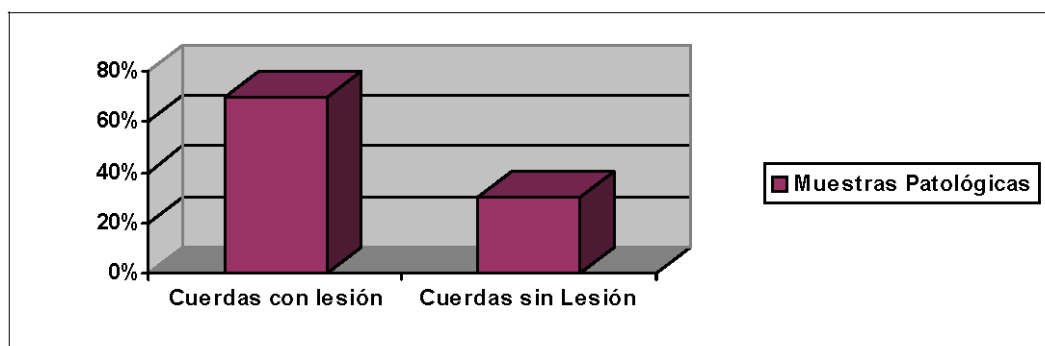


Figura 4-5: Distribución del porcentaje de muestras patológicas en función del grupo de patologías

Además se muestra en la Figura 4-6 y Figura 4-7 el número de muestras de las distintas enfermedades de cada grupo de patologías.

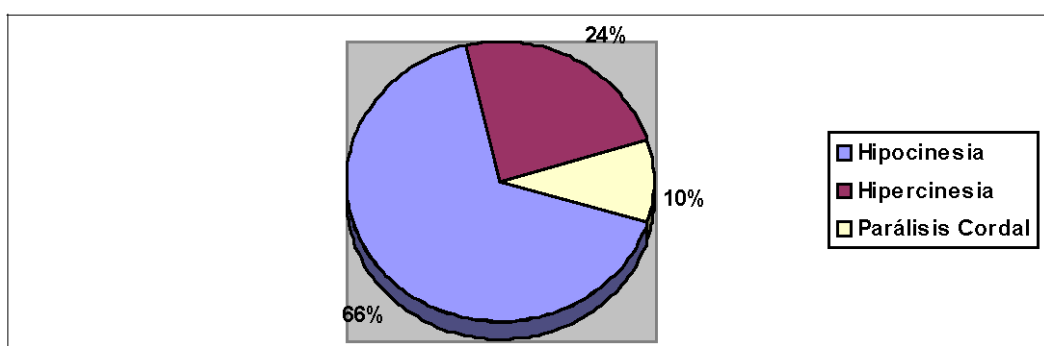


Figura 4-6: Distribución del porcentaje de muestras patológicas del grupo “cuerdas sin lesión” entre las distintas patologías

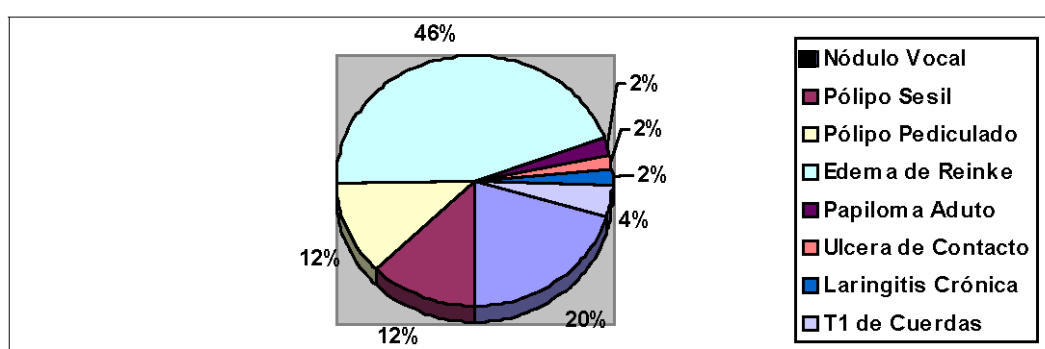


Figura 4-7: Distribución del porcentaje de muestras patológicas del grupo “cuerdas con lesión” entre las distintas patologías

En la base de datos se graban y se almacenan datos referentes al paciente: descripción de la patología, sexo, y el grado de disfonía. El campo “descripción de la patología” consiste en una anotación no normalizada describiendo el estado del sistema

fonador del locutor, mientras que el campo “grado de disfonía” evalúa subjetivamente la calidad de la voz permitiéndose etiquetarlas en: voz sana, disfonía leve, disfonía moderada y diafonía severa. Ya que no se almacenan datos personales del locutor que permitan identificarlo, y teniendo en cuenta que la base de datos tiene carácter de material docente dedicado a la investigación, la base de datos no necesita ser dada de alta en la Agencia Nacional de Protección de Datos.

4.2.2 Base de datos Comercial

Esta base de datos es la *Voice Disorders Database Model 4337* [6] siendo desarrollada por el *Massachussets Eye and Ear Infirmary Voice and Speech Lab* y distribuida comercialmente por la empresa *KAY Elemetrics*. Esta base de datos ha sido utilizada ampliamente en diferentes estudios de medidas de calidad de la voz, y por lo tanto, se ha convertido en la base de datos de referencia.

Contiene 1400 muestras de voz de aproximadamente 700 sujetos. Incluye la fonación sostenida de la vocal “a” y una frase del pasaje *Rainbow*, de pacientes con una amplia variedad de patologías del sistema fonador de carácter orgánico, neurológico, traumático y psicológico, al igual que 53 muestras de voces sanas.

Todos los pacientes han sido evaluados en un extenso examen de la función vocal incluyendo estudios videoestroboscópicos, aerodinámicos y fonatorios. En la información de cada locutor se incluye información sobre el diagnóstico, edad, sexo, datos sobre hábitos tales como si es fumador, etc.

La metodología empleada para grabar las muestras de voz, según describe el distribuidor, es la siguiente. Se ha usado un micrófono capacitivo en una cámara acústicamente aislada, usando una distancia de la boca al micrófono de 15 centímetros. La señal fue registrada con un grabador DAT con una frecuencia de muestreo de 44.1 Khz, utilizándose un preamplificador calibrado. A partir de la cinta DAT, fue digitalizada utilizando el equipo CSL modelo 4300, en unos casos frecuencia de muestreo 25 Khz y en otros casos 50 KHz, usándose en cada caso el filtro *antialiasing* adecuado. Los ficheros fueron grabados en el formato CSL-DOS, formato .nsp de *Kay Elemetrics*.

4.3 Referencias

- [1] Dimitar D. Deliyski, Maegan K. Evans and Heather S. Shaw, "Influence of Data Acquisition Environment on Accuracy of Acoustic Voice Quality Measurements", *Journal of voice*, vol.19 , no.2 ,pp.176-186 , 2005.
- [2] Vijay Parsa, Donald G. Jamieson, and Bradley R. Pretty, "Effects of Microphone on Acoustic Measures of Voice", *Journal of Voice*, vol.15,no.3, pp.331-343, 2001.
- [3] Cecyle Perry Carson, Dennis R.-S. Ingrisano, and K. Donald Eggleston, "The Effect of Noise on computer-Aided Measures of Voice: A comparison of CSpeechSp and Multi-Dimensional Voice Program Software Using the CSL 4300B Module and Multi-Speech for Windows", *Journal of Voice*, vol.17,no.1,2003.
- [4] Dimitar D. Deliyski, Heather S. Shaw, and Maegan K. Evans, "Adverse Effects of Environmental Noise on Acoustic Voice Quality Measurements", *Journal of voice*, vol.19 , no.1 ,pp.15-28 , 2005.
- [5] Jacques Koreman y Manfred Pützer, "Finding Correlates of Vocal Fold Adduction Deficiencies", *Phonus 3*, Institute of Phonetics, University of the Saarland, pp.155-178, 1997
- [6] Kay Elemetrics Corporation. Disorder Voice Database Model 4337. Massachusetts Eye and Ear Infirmary Voice and Speech Lab, Boston, MA. 1994.
- [7] <http://www.kayelemetrics.com>, [Consulta 10/10/05]
- [8] Jesús B. Alonso Hernández, Itziar Alonso González y Miguel A. Ferrer Ballester, "Procesado de Voz en Tiempo Real sobre PC para Aplicaciones Multimedia", *Actas IV Jornadas de Informática*, pp.343-352, 1998.

CAPÍTULO 5

Selección de Medidas

Un procedimiento útil para estudiar la viabilidad de las medidas de calidad de la voz es el uso de sistemas de clasificación, ya que si tras transformar el audio en un vector de medidas de calidad es posible por medio de un sistema de clasificación supervisado, obtener tasas de éxito aceptables, identificando diferentes unidades de clasificación relacionadas con distintos grados de calidad de voz, podemos asumir que en el vector de medidas de calidad existe suficiente información como para cuantificar la calidad de la voz.

Es posible particularizar el modelo general de reconocimiento de patrones para un sistema automático de medida de la calidad de la voz, identificando los siguientes bloques del modelo de funcionamiento:

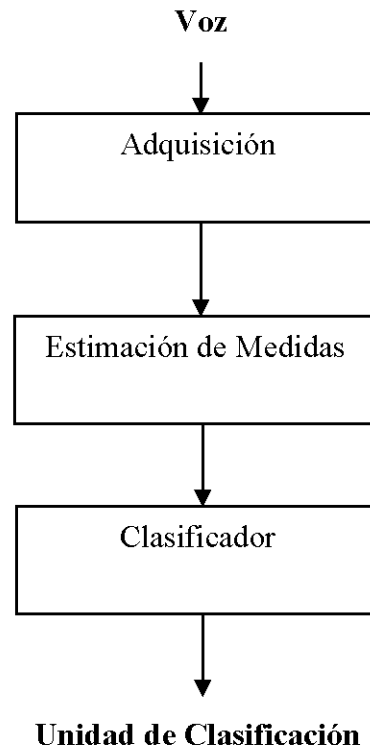


Figura 5-1: Esquema de un sistema automático de medida de calidad de voz

A partir del módulo de “adquisición” se digitaliza la señal de voz. Posteriormente el audio digitalizado es caracterizado por medio de un conjunto de medidas de calidad de voz, denominadas en el contexto de reconocimiento de patrones, entradas o características. Y por último, el clasificador, a partir del vector de características, etiqueta la voz en función de un conjunto de unidades de clasificación.

Existen diversos motivos por los que resulta importante realizar una selección de las entradas o características que se aplican a la entrada de un sistema de clasificación. Los cuatro principales motivos son los siguientes:

1. Al reducir el número de entradas se reduce el tamaño del sistema clasificador usado y de esta forma se simplifica el modelo utilizado para resolver el problema.
2. Se reduce la cantidad de datos necesarios para el entrenamiento.
3. Se reducirá el tiempo de entrenamiento ya que la arquitectura del sistema clasificador es menor.

4. Se podrá conseguir un incremento en la capacidad de generalización del sistema clasificador, es decir, un incremento en la eficiencia de la aplicación construida.

Este proceso de selección de medidas, en el contexto del reconocimiento de patrones, es denominado en la bibliografía común como “*feature selection*”, “*feature subset selection*” (FSS) o “*Input variable selection*” (IVS).

Debido a que en la bibliografía actual sobre medidas de calidad de la voz, existe gran cantidad de ellas, es conveniente, atendiendo a uno de los objetivos de esta tesis doctoral, obtener un subconjunto reducido de medidas de calidad de la voz que permitan realizar una correlación entre su valor numérico y el fenómeno físico que cuantifica, y por tanto resulta necesario aplicar técnicas de selección de características. A continuación se realiza una revisión del estado de la técnica en lo que respecta a selección de características.

5.1 Revisión de los métodos

Para elegir un método de selección de características es importante considerar el coste computacional que conlleva. En cualquier método existe una función objetivo la cual evalúa una determinada combinación de características. Entre los distintos grupos en los que se han ordenado los distintos métodos, podemos distinguir los que usan algún aspecto del clasificador como función objetivo, y los que usan diversas funciones objetivo que son independientes del clasificador.

Otra forma de agrupar a los métodos de selección de características, es en función de si realiza una ordenación de la importancia de las distintas características o si por el contrario simplemente propone cual es el subconjunto de las características evaluadas las cuales proporcionan una mayor eficiencia en la función objetivo.

La mayoría de los métodos permiten obtener una ordenación de los parámetros candidatos acorde con la importancia de los mismos, pero no suministran cuál debe ser el subconjunto mínimo y útil de entradas que debemos seleccionar. Una solución a este problema es evaluar las características, en el clasificador que se utilizará, buscando el

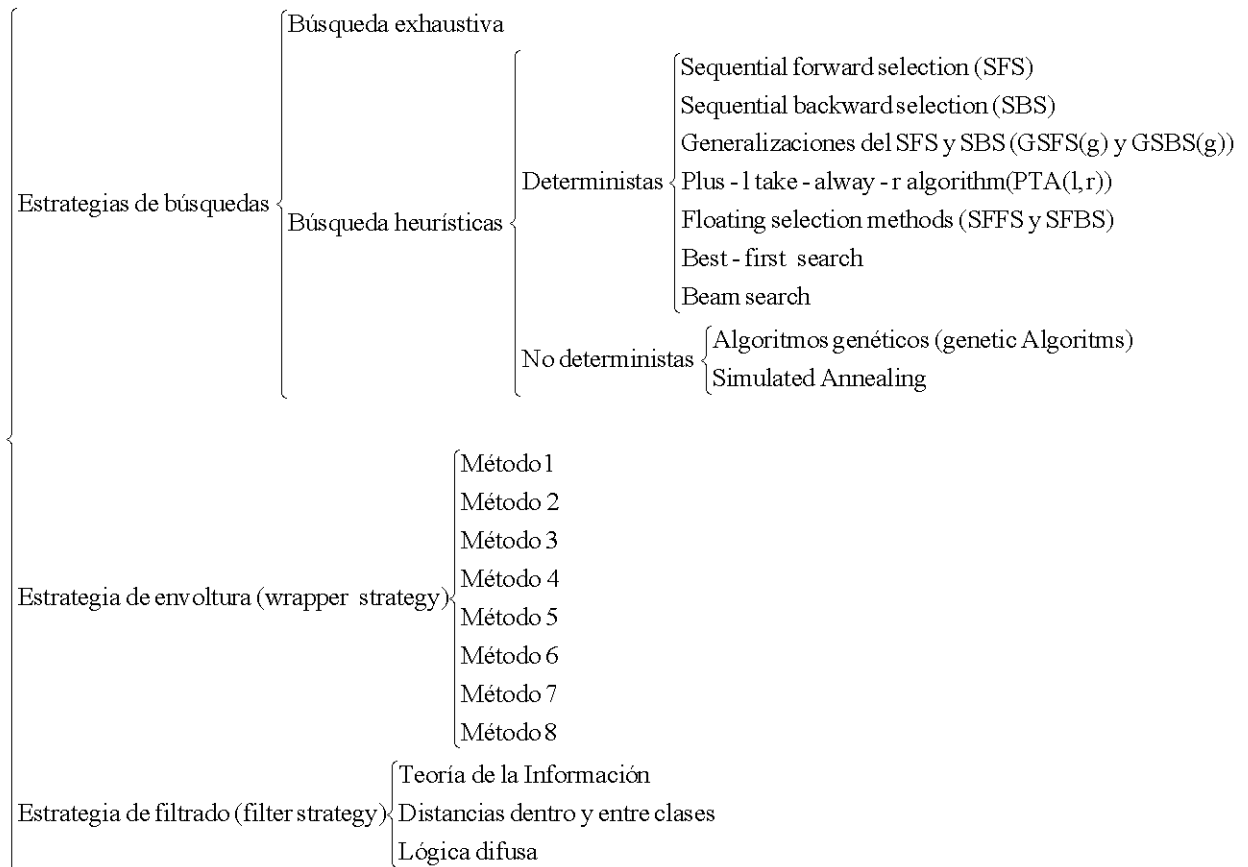
menor subconjunto de características formado por los mejores valorados, de tal forma que al incluir una característica más no implique un aumento en las prestaciones del clasificador.

Por otro lado, existen autores [5] que indican que el mejor subconjunto de m características no necesariamente está compuesto por las m mejores características individuales.

En este trabajo se han dividido los distintos métodos de ordenación de parámetros en tres grupos a los que hemos denominado: estrategias de búsqueda, estrategias de filtrado y estrategias de envoltura. Las estrategias de búsqueda y de envoltura utilizan el clasificador para obtener la función objetivo, a diferencia de las estrategias de filtrado en las que la función objetivo no utiliza información aportada por el clasificador. Por otro lado, las estrategias de filtrado y envoltura proporcionan una ordenación de cada uno de los parámetros en función de su importancia, mientras que las estrategias de búsqueda indican el subconjunto óptimo de parámetros.

En la siguiente figura se muestra un esquema de las distintas estrategias y a continuación se comentan cada una de ellas.

TABLA 5-1: CLASIFICACIÓN DE LAS TÉCNICAS DE SELECCIÓN DE CARACTERÍSTICAS



1. *Estrategias de búsquedas:* una vez fijado el sistema de clasificación a utilizar, este sistema pretende buscar por medio de distintos procedimientos [6] la selección de características de entrada que producen una mayor efectividad en la clasificación. Existen distintos métodos de búsqueda del subconjunto de características óptimo:

1.1. *Búsqueda exhaustiva (complete search):* consistiría en evaluar todas las posibles combinaciones de agrupaciones de características. Una vez evaluadas todas las combinaciones, se selecciona la agrupación de características que mejor efectividad produce evaluada sobre el sistema de clasificación. Existen distintas implementaciones: *depth-first*, *breadth-first*, y *branch and bound*.

1.2. *Búsqueda heurísticas (heuristic search):* se dividen en dos grupos: deterministas y no deterministas.

1.2.1. *Deterministas:* estos métodos se caracterizan por obtener los mismos resultados para los mismos datos.

1.2.1.1. *Sequential forward selection (SFS)*: se trata de un algoritmo de búsqueda del tipo *hill-climbing* el cual empieza con un conjunto de características vacío y van añadiendo características que proporcionan un aumento de la efectividad hasta que no quedan características que aumenten la efectividad. Presenta el inconveniente de que una vez seleccionada una determinada característica, ya no puede ser eliminada.

1.2.1.2. *Sequential backward selection (SBS)*: se trata de otro algoritmo de búsqueda del tipo *hill-climbing* en el cual se empieza con la totalidad del conjunto de características y se van descartando características que al eliminarlas no reducen la efectividad del sistema. De forma similar al SFS, una vez eliminada una determinada característica, ésta no puede volver a ser elegida.

1.2.1.3. *Generalizaciones del SFS y SBS (GSFS(g) y GSBS(g))*. Se evalúa un determinado número de características (g) a la vez, siendo en grupos posteriormente seleccionados y descartados.

1.2.1.4. *Plus-l take-alway-r algorithm (PTA(l,r))*: se seleccionan l características por medio de un algoritmo SFS y se eliminan r características por medio de un algoritmo SBS. Además, existe una variante, la versión generalizada $GPTA(l,r)$.

1.2.1.5. *Floating selection methods (SFFS y SFBS)*: es una variante del $PTA(l,r)$ en el que puede existir una vuelta atrás en las características seleccionadas.

1.2.1.6. *Best-first search*: no se visitan todas las posibles combinaciones.

1.2.1.7. *Beam search*: versión modificada del *Best-first search*. Presenta múltiples soluciones.

1.2.2. *No deterministas*: buscan evitar, en la función de evaluación de las características, los máximos locales. Estos procedimientos presentan elementos de aleatoriedad, lo que provoca que distintas realizaciones con los mismos datos no obtengan los mismos resultados.

1.2.2.1. *Algoritmos genéticos (genetic Algorithms)*: existen distintas implementaciones de algoritmos [2] [4] [5], diferenciándose éstos en los distintos criterios para la generación de las nuevas poblaciones: *parallel algorithms* (se basa en algoritmos *hill-climbing*), *Estimation of Distribution algorithm (EDA)*, etc.

1.2.2.2. *Simulated Annealing*. Se distingue de los algoritmos genéticos en que presenta una solución única, mientras que los algoritmos genéticos presentan múltiples soluciones.

2. *Estrategia de filtrado (filter strategy)*: también llamada *Model-free*, se estima la importancia de las distintas características en función únicamente de los datos disponibles del conjunto de entrenamiento y del correspondiente etiquetado de las muestras. Los métodos basados en el análisis del conjunto de entrenamiento se han agrupado en tres categorías [1], en función del tipo de técnicas que utilizan: conceptos de la Teoría de la Información, medida de distancias dentro y entre las clases y por último, lógica difusa (*fuzzy*).

A priori, estos métodos presentan la ventaja respecto al resto de las estrategias de que no hay que entrenar ningún sistema clasificador para aplicarlos. Por ello, su carga computacional suele ser mucho menor.

2.1. *Teoría de la Información*: dentro de este grupo de métodos, resulta interesante el método [22] basado en conceptos de Teoría de la Información cuyo algoritmo es el siguiente: haciendo uso del valor de la información mutua entre las diferentes características y el etiquetado de las distintas clases, se selecciona el conjunto de características que maximiza el resultado de la expresión “valor de la información mutua entre el conjunto de características seleccionadas respecto al etiquetado de las clases menos el valor de la información mutua del conjunto de características seleccionadas con respecto al conjunto de características no seleccionadas”.

Otras estrategias están basadas en el cálculo de la entropía de una entrada [25] o un nuevo método basado en el cálculo de ganancias [26], donde se relacionan parámetros relacionados con la información esperada. Otro ejemplo, es el método [27] el cual se basa en la medida llamada Distancia GD, una generalización de la Distancia de Mántaras, la cual permite recoger las interdependencias entre las entradas y definir una ordenación de las mismas según su importancia.

2.2. *Distancias dentro y entre clases*: uno de los métodos incluidos en este grupo es muy popular en la literatura sobre reconocimiento de patrones, y se basa en un algoritmo [28] usado en la distancia euclídea entre las características y ciertos puntos de referencia en el espacio de características. Otro método, característico de este grupo define la distancia entre dos entradas como la “Distancia Euclídea sopesada” [29].

Otro método del grupo, propone una ordenación de los parámetros en función de ciertas funciones objetivos [5] : variación intra-clase (varianza de las características dentro de la misma clase), variación inter-clase (separación entre clases) y la correlación entre características.

2.3. *Lógica difusa*: En este tipo de métodos la bibliografía es menos extensa. Este método [30] realiza un análisis de regiones difusas y para ello es necesario generar en primer lugar un conjunto de reglas difusas que delimiten dichas regiones.

3. *Estrategia de envoltura (wrapper strategy)*: esta estrategia busca maximizar el éxito en la clasificación, para ello tiene en cuenta, además de los datos disponibles, la arquitectura del clasificador que se utilizará y del algoritmo de aprendizaje utilizado [23] [24] . Ha de tenerse en cuenta la dependencia de las características con el clasificador en el que se van a evaluar. Existen distintas técnicas dependientes del clasificador utilizado. A continuación se describirán los métodos de selección de entradas basados en el análisis de una red neuronal entrenada de tipo *Multilayer Feedforward* [1] [8] , pero existen estudio orientados a la selección de parámetros para otro tipo de clasificadores como por ejemplo redes neuronales RBF [8] o Redes Gaussianas Condicionadas [9] .

Todos estos métodos, que hacen uso de redes neuronales entrenadas del tipo *Multilayer Feedforward*, están basados en el principio de que durante la fase de entrenamiento la red neuronal debería seleccionar la importancia de las entradas y este resultado se debería reflejar en la estructura final de la red, por ejemplo los pesos conectados a las unidades de entrada. Todos ellos proponen cómo seleccionar las características útiles, es decir, dan una ordenación en función de su importancia.

Algunos de ellos, definen la relevancia de una entrada i , S_i , siendo la manera de calcular dicha relevancia muy variada. Por ejemplo, utilizando la siguiente ecuación se puede calcular la relevancia S_i a partir de la relevancia s_{ij} de cada peso w_{ij} conectado a la unidad de entrada i y a la unidad de la capa oculta j . Una entrada es considerada más importante si su relevancia es mayor.

$$S_i = \sum_{j=1}^{Mn} s_{ij} \quad [\text{Ec.5-1}]$$

Se han agrupado [1] los diferentes métodos que se van a estudiar en función de la idea básica de medida de la relevancia, obteniéndose ocho grupos que son los siguientes.

3.1. Primer Grupo: En este primer grupo el criterio para definir la relevancia de los pesos está basado en la estimación directa de la magnitud de los pesos. Existen varios métodos basados en esta idea [10] .

Los métodos incluidos en este grupo proponen calcular la relevancia para una característica de entrada i como la suma de las relevancias de los pesos entre la capa de entrada y la capa oculta.

Dentro del primer grupo que se está estudiando, se encuentra un método basado en la estimación directa de las magnitudes de los pesos conectados entre las neuronas de la capa de entrada y la capa oculta de la red neuronal [10] . Se miden los valores absolutos de las magnitudes de los pesos w_{ij} entre la capa oculta y la de entrada y se obtiene la relevancia de cada uno.

3.2. Segundo Grupo: En este grupo el criterio de relevancia está basado en la estimación del incremento en el error cuadrático medio, cuando se podan los pesos, es decir, cuando se sustituye su valor por el valor cero.

Existen diversos métodos [10] [11] [12] donde cada uno de los distintos autores de dichos métodos propone una forma diferente en función de la variación de la función de error E de la red cuando se suprime un determinado peso, es decir se hace cero.

3.3. Tercer Grupo: Otros métodos definen la relevancia usando la varianza de los pesos w_{ij} conectados a una entrada i . Están basados en la idea de que una pequeña varianza hace que la entrada se comporte como un umbral y su importancia será menor. Existen métodos propuestos basados en la idea de encontrar las características más relevantes

capaces de discriminar objetos [13] [14] . Cada patrón o ejemplo de entrada en la red neuronal es representado por un punto del hiperespacio de características. Los patrones que pertenecen a las diferentes clases serán agrupados en diferentes regiones del espacio y separados por hiperplanos.

3.4. Cuarto Grupo: Otros métodos [10] [15] evalúan la contribución de una entrada a la salida teniendo en cuenta el valor de los pesos en toda la estructura de la red neuronal, proponiendo evaluar la contribución de una entrada en las salidas de la red teniendo en cuenta toda la estructura de la misma, es decir, todos los pesos.

3.5. Quinto Grupo: Hay varios métodos que usan la sensibilidad de las salidas con respecto a una entrada para definir su relevancia. Están basados en la idea de que una alta sensibilidad significa una gran variación de los valores de salida debido a un cambio en una entrada, por lo que, se puede suponer que la entrada es más importante.

En la referencia [9] se describe un método en el que para definir la importancia de las entradas se mide la sensibilidad de la salida respecto de las entradas. Otro método mide la relevancia de una unidad de entrada en función de la sensibilidad de las salidas respecto a las entradas [16] .

Otros métodos se basan en la medida del error en la clasificación correcta de un vector de entrada [17] [18] , midiendo la probabilidad a posteriori de una clase dado un vector de entrada. También se ha utilizado las sensibilidades logarítmicas [14] con el fin de evitar que aparezca el término de saturación.

Sexto Grupo: Algunos métodos se basan en la estimación del decremento de la eficiencia de la red cuando se sustituye la entrada por un valor fijo apropiado. La relevancia de una entrada se considera mayor si el decremento en la eficiencia aumenta. Se ha propuesto un método basado en el cálculo del incremento en el error cuadrático medio cuando se sustituye una característica de entrada i por el valor medio de los valores que esa característica toma en todos los patrones de entrada [19] . Otro método se basa en calcular el decremento en el porcentaje de aciertos en el test cuando se sustituye el valor de una entrada por su valor medio [18] .

3.6. *Séptimo Grupo*: En este grupo se propone un algoritmo de aprendizaje que simultáneamente selecciona las entradas de la red y el número de unidades de la capa oculta [31] [32] . Para ello, definen una función que establece la importancia de cada unidad dentro de la estructura de la red. Dicha definición está basada en conceptos de teoría de la información: el análisis de interdependencias.

3.7. *Octavo Grupo*: se propone un método [21] donde, partiendo del entrenamiento de redes *Multilayer Feedforward* basado en el algoritmo *Backpropagation*, el funcionamiento del algoritmo utiliza la diferencia (o error) entre la salida actual de la red y la deseada, propagando hacia atrás a lo largo de la red neuronal, de forma que durante el proceso de entrenamiento se van modificando los pesos. En el método se va acumulando durante el proceso de entrenamiento un error e_{it} (*accumulated error index*), que será el que después defina la relevancia de las entradas. Dentro de este grupo, se han propuesto otras estrategias las cuales se basan en calcular el efecto de una entrada en la salida de la red neuronal [22] , determinando el grado de participación que cada neurona de entrada tiene en una neurona de salida.

A modo de conclusión en este apartado, destacar que existe una tesis doctoral (Fernández-Redondo y Hernández-Espinosa, [1]) en la que se ha estudiado y comparado gran parte de los métodos incluidos en la estrategia de filtrado y de *wapper*. Como resultado los autores proponen, de forma general, la utilización de los métodos basados en estrategias de *wapper* frente a los métodos basados en la estrategia de filtrado. Esta conclusión es tomada en base a los mejores resultados obtenidos para distintas bases de datos [33] .

El método mejor valorado fue el propuesto por J. Utans, J. Moody y colaboradores [19] . Este método corresponde al grupo sexto de las estrategias de *Wapper*. En segundo lugar, y de forma aislada, se recomienda un método de grupo Teoría de la Información dentro de la estrategia de filtrado propuesto por Lorenzo, Hernández y Méndez [27] . En tercer lugar, se propone otro método que corresponde al grupo primero de las estrategias de Wrapper propuesto por I.V. Tetko, A.E.P. Villa y D.J. Livingstone [10] .

5.2 Método propuesto

En el contexto del reconocimiento de patrones, una base de datos está constituida por observaciones donde cada observación está compuesta por un conjunto concreto de medidas y en el que las diferentes observaciones son etiquetadas en diferentes unidades de clasificación. A partir de la base de datos es posible identificar, dentro del conjunto concreto de medidas del que están constituidas las observaciones, qué medidas tienen mayor relevancia o cuál es el conjunto óptimo de medidas, usando para ello una de las técnicas de selección de características descritas en el apartado anterior.

Las medidas son evaluadas en función de su comportamiento frente a un determinado clasificador o en función de la interacción de las medidas en función de las clases.

Sin embargo, resulta inusual en la bibliografía actual encontrar técnicas de selección de características que hagan uso de conocimientos a priori como la naturaleza de las características o su interpretación física.

En esta tesis doctoral, a modo de aportación, consideramos que en el contexto de la selección de medidas de calidad de la voz, la información de la naturaleza o la interpretación física de la medida puede resultar interesante ya que es posible identificar irrelevancias de medidas debido a su propia naturaleza o al proceso de creación de la base de datos, ya que es posible la existencia de determinadas medidas que no permiten obtener una generalización de los resultados obtenidos a bases de datos diferentes a las usadas.

Además, cabe la posibilidad, en determinadas aplicaciones, de necesitarse la información que aporta una o varias medidas, cuya importancia es manifiesta. Ésto es un factor a tener en cuenta ya que en determinados métodos de selección de características, para diferentes realizaciones, el conjunto de características relevantes obtenido no coincide. Este fenómeno es especialmente frecuente en bases de datos significativamente pequeñas y poco representativas del espacio muestral.

Por lo tanto, a tenor de lo expuesto, en este tesis doctoral se propone un método de selección de características, en el ámbito de las medidas de calidad de la voz, que combina el conocimiento a priori de las peculiaridades de la caracterización de la calidad de la voz y un método determinístico de selección de características.

En el contexto de la estimación de la calidad de la voz a partir de una grabación de voz, no existe un único fenómeno que permita diferenciar entre las distintas calidades de voz, y por tanto, es necesario cuantificar diferentes fenómenos físicos que intervienen en la caracterización de la calidad de la voz. Este hecho, se traduce en que resulta necesario tener en cuenta diferentes medidas de calidad para poder cuantificar la calidad de la voz de un locutor.

En este trabajo se han identificado cuatro fenómenos físicos cuantificables que permiten realizar una medición de la calidad de la voz, los cuales ya han sido definidos en el capítulo 3, y que han sido denominados como:

- **Estabilidad de la voz:** es la capacidad de un locutor de producir, durante la producción de un sonido sonoro sostenido, un flujo de aire con una intensidad constante para excitar las cuerdas vocales.
- **Riqueza espectral:** se define como la capacidad de generar, durante la fonación de un sonido sostenido, un movimiento periódico de las cuerdas vocales y de producir una excitación sonora al tracto vocal con gran cantidad de componentes espectrales.
- **Presencia de ruido:** se define como la presencia de ruido glótico en la voz, durante la fonación de un sonido sonoro sostenido, debido a carencias en la fase de cierre de las cuerdas vocales.
- **Comportamiento no lineal:** este fenómeno físico es ocasionado por un funcionamiento anómalo de las cuerdas vocales, debido a irregularidades en las masas de cuerdas vocales que intervienen en la fase de cierre de las cuerdas, existencia de asimetrías en el movimiento de las cuerdas y factores relacionados con la mucosa de dichas cuerdas.

Resulta necesario disponer de medidas de calidad de la voz de cada uno de los grupos con el fin de poder identificar la gran variedad de tipos de alteraciones de la calidad de la voz, ya que cualquier voz de baja calidad presenta al menos la alteración de una de las cuatro características.

En esta tesis doctoral se propone clasificar las diferentes medidas de calidad de voz en los cuatro grupos atendiendo al fenómeno físico que cuantifica. Una vez identificadas las diferentes medidas de calidad de voz pertenecientes a cada una de los diferentes grupos de características es necesario estudiar la relevancia de las medidas para cada grupo. El método de selección de características buscado, para estudiar la relevancia de las medidas de cada grupo, consiste en un método que permita obtener ordenadas las medidas de la calidad de la voz en función de su relevancia. Resulta interesante que tenga un carácter determinístico con el fin de poder generalizar los resultados obtenidos. Además, se busca un método de selección de características con un coste computacional razonable. El algoritmo seleccionado fue el *floating forward feature selection*.

En concreto, el método propuesto consiste en asignar las diferentes medidas de calidad de la voz en uno de los cuatro grupos de medidas que hemos identificado. Posteriormente, las medidas de cada grupo se ordenan en función de su relevancia por medio del método *floating forward feature selection*. Este método proporciona, para cada grupo de medidas, un conjunto óptimo de características, y es por ello que el proceso de evaluación de cada uno de los diferentes grupos de medidas se repite k iteraciones, y posteriormente se ordenan las características en función del número de ocasiones en las que para las diferentes k iteraciones estuvieron seleccionadas dentro del subconjunto óptimo.

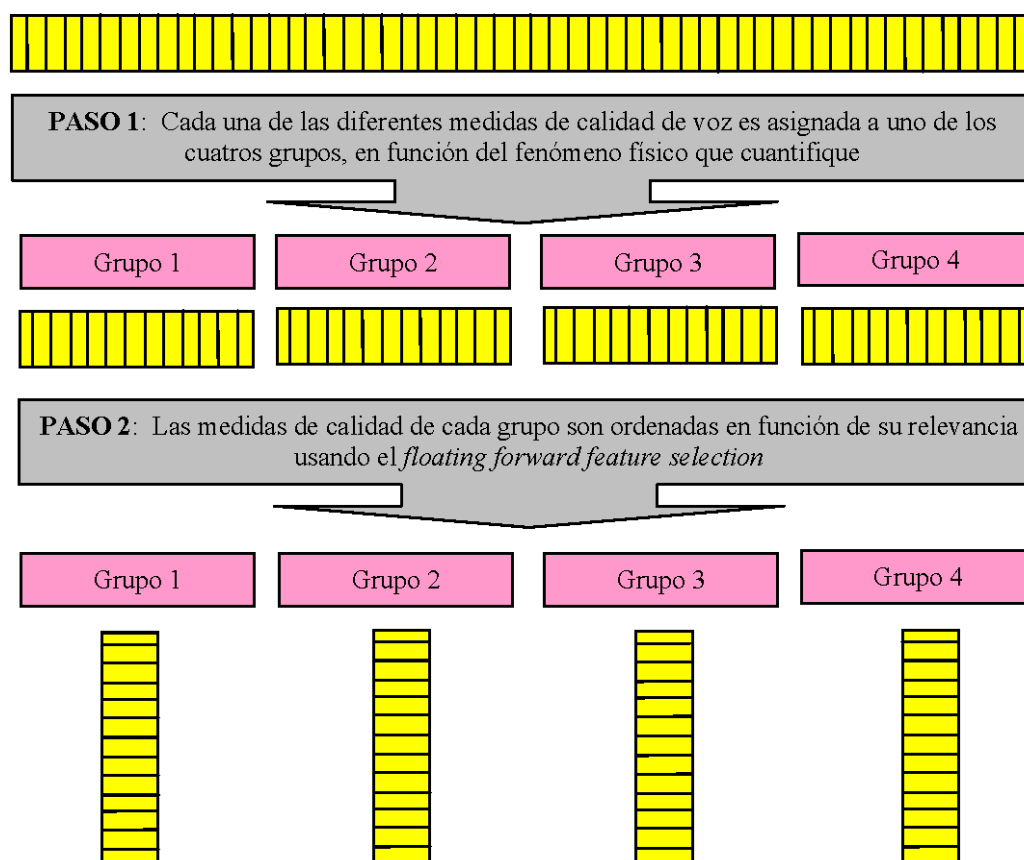


Figura 5-2: Esquema del procedimiento de selección de características de medidas de calidad de voz propuesto

Con este método de selección de características, en el contexto de las medidas de calidad de la voz, se obtiene cuatro listas de medidas de calidad. Las medidas incluidas en cada una de las listas coinciden en el fenómeno físico que cuantifican. Una cuestión que se plantea una vez presentado el procedimiento para seleccionar las medidas, es: ¿cuántas medidas de cada una de las listas de los cuatro grupos de fenómenos físicos son necesarias para cuantificar cada uno de los fenómenos físicos? . La presentación de las medidas de cada uno de los grupos y el número necesario de medidas de cada uno de los grupos es estudiado en los próximos cuatro capítulos.

En este trabajo se ha realizado una revisión de la bibliografía actual referente a las medidas de la calidad de la voz, clasificando cada una de las diferentes medidas propuestas en los diferentes artículos en alguno de los cuatro fenómenos físicos cuantificables identificados en este trabajo.

Existe en la bibliografía un conjunto de medidas que pese a que poseen un procedimiento de medida objetivo, el fenómeno que cuantifican presenta características subjetivas como son, medidas del ritmo de lectura en el que se contabiliza el número de palabras leídas por minuto, intervalos de tiempo en la lectura, medidas de la fluidez verbal, número de error en la lectura, etc, [37] . En este grupo de medidas se ha incluido la medida tiempo de fonación de un sonido sostenido y medidas similares como las mencionadas [34] [35] [36] . Este conjunto de medidas, debido a que entran a evaluar otras situaciones como el nivel cultural del locutor o la presencia de trastornos del aprendizaje, no han sido estudiadas en este trabajo por no considerarse medidas objetivas para una amplia muestra de la población.

No se han estudiado las medidas objetivas de calidad de la voz que presentan un formato vectorial como MFCC, LFCC, NFHE, GCI, ... principalmente debido a que a un vector resulta de gran complejidad asignarle una interpretación física, y no sería posible cumplir con uno de los objetivos de la tesis. Tampoco se ha estudiado el uso de parámetros de los modelos de cuerdas vocales como medidas de calidad de la voz, pese a lo interesante de poder asociar los valores de los parámetros del modelo a las características fisiológicas del sistema de producción de voz, debido a que la correcta obtención de los parámetros de estos modelos es una tarea aún no resuelta y sería necesario entrar a resolver cuestiones que actualmente están abiertas en el campo de la síntesis de voz, alejándose este objetivo de los planteados en esta tesis doctoral.

5.3 Referencias

- [1] Mercedes Fernández Redondo, Carlos A. Hernández Espinosa, *Hacia un Diseño Óptimo de la Arquitectura Multilayer Feedforward*. Tesis Doctoral. Universidad Jaime I, septiembre de 2001.
- [2] Peck, C.C., Dhawan, A.P., "SSME parameter model input selection using genetic algorithms". *IEEE Transactions on Aerospace and Electronic Systems*, vol.32, no.1, pp.199-212, January 1996.
- [3] Peck, C.C., Dhawan, A.P., "Genetic algorithm based input selection for a neural network function approximator with applications to SSME health monitoring". *Proceedings of the 1993 IEEE International Conference on Neural Networks*, vol.22, pp.1115-1122, 1993.
- [4] Inza, I., Merino, M., Larrañaga P., Quiroga J., Sierra B., Giralda M., "Feature subset selection by genetic algorithms and estimation of distribution algorithms. A case study in the survival of cirrhotic patients treated with TIPS". *Artificial Intelligence in Medicine*, vol.23, no.2, pp.187-205, october 2001.
- [5] Pablo Estévez Valencia, "Selección de Características para Clasificadores Neuronales", *Anales del Instituto de Ingeniería de Chile*, pp.65-74, diciembre de 1999.
- [6] Mineichi Kudo, Jack Sklansky, "Comparison of Algorithms that select Features for Pattern Classifiers". *Pattern Recognition*, vol.33, no.1, pp.25-41, 2000.
- [7] Carlos Hernández-Espinosa, Mercedes Fernández-Redondo, Pedro Gómez-Vilda, "Neural Network based Input Selection and Diagnosis of Pathologic Voices". *Neural Network World*, vol.11, no.1, pp.49-63, 2001.
- [8] Xiuju Fu and lipo Wang, "Rule Extraction Based on Data Dimensionality Reduction using RBF Neural Networks", *Proceedings of the International Joint Conference on Neural Networks*, vol.2, pp.1275-1280, 2003;
- [9] J.M. Peña, J.A. Lozano, P. Larrañaga and I. Inza, "Dimensionality Reduction in Unsupervised Learning of Conditional Gaussian Networks", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.23, no. 6, pp.590-603, 2001
- [10] Tetko, I.V., Villa, A.E.P., Livingstone, D.J., "Neural network studies. 2 Variable selection". *Journal of Chemical Information and Computer Sciences*, vol.36, no. 4. pp.794-803, 1996.
- [11] Cibas, T., Soulié, F.F., Gallinari, P., Raudys, S., "Variable selection with neural networks". *Neurocomputing*, vol.12, pp.223-248, 1996
- [12] Mao, J., Mohiuddin, K., Jain, A.K., "Parsimonious network design and feature selection through node pruning". *Proceedings of the 12th IAPR International Conference on Pattern Recognition*, vol.2, pp.622-624, 1994
- [13] Devena, L., "Automatic selection of the most relevant features to recognize objects". *Proceedings of the International Conference on Artificial Neural Networks*, vol.2, pp.1113-1116, 1994.
- [14] El-Deredy, W., Branston, N.M., "Identification of relevant features in HMR tumor spectra using neural networks". *Proceedings of the 4th International Conference on Artificial Neural Networks*, pp.454-458,1995
- [15] Tetko, I.V., Tanchuk, V.Y., Luik, A.I., "Simple heuristic methods for input parameter estimation in neural networks". *Proceedings of the IEEE International Conference on Neural Networks*, vol.1, pp.376-380, 1994.

- [16] Engelbrecht, AP., Cloete, I., "A sensitivity analysis algorithm for pruning feedforward neural networks". *Proceedings of the International Conference on Neural Networks*, vol.2, pp.1274-1277, 1996.
- [17] Priddy, K.L., Rogers, S.K., Ruck D.W., Tarr G.L., Kabrisky, M., "Bayesian selection of important features for feedforward neural networks". *Neurocomputing*, ; vol.5, no.23, pp.91-103,1993.
- [18] Sano, H., Nada, A., Iwahori, Y., Ishii, N., "A method of analyzing information represented in neural networks". *Proceedings of 1993 International Joint Conference on Neural Networks*, pp.2719-2722,1993.
- [19] Utans, J., Moody, J., et alt, "Input variable selection for neural networks: Application to predicting the U.S. business cycle". *Proceedings of IEEE/IAFE Computational Intelligence for Financial Engineering*, pp.118-122, 1995.
- [20] Lee, H., Mehrotra, K., Mohan, C. Ranka, S., "Selection procedures for redundant inputs in neural networks". *Proceedings of the World Congress on Neural Networks*, vol.1, pp.300-303, 1993.
- [21] Bowles, A., "Machine learns which features to select", *Proceedings of the 5th Australian Joint Conference on Artificial Intelligence*, pp.127-132, 1992.
- [22] Younes, B., Fabrice, B., "A neural network based variable selector". *Proceedings of the Artificial Neural Network in Engineering*, (ANNIE'95), pp.425-430, 1995.
- [23] Greene, R.A., "Neural net applications in the cognitive sciences: a case study". *Journal of Microcomputer Applications*, vol.16, pp.197-201, 1993.
- [24] Hammitt, A.M., Bartlett, E.B., "Determining functional relationships from trained neural networks". *Mathematical Computing Modelling*, vol.22, no.3, pp.83-103, 1995 .
- [25] Chi, Jabri, "Entropy based feature evaluation and selection technique". *Proceedings of 4th Australian Conference on Neural Networks*, (ACNN'93), 1993; 193-196.
- [26] Setiono, R., Liu, H., "Improving Backpropagation learning with feature selection". *Applied Intelligence: The International Journal of Artificial Intelligence, eural Networks, and Complex Problem-Solving Technologies*, vol.6, no.2, pp.129-139, 1996.
- [27] Lorenzo, Hernández, Méndez, "Attribute selection through a measurement based on information theory", *7^a Conferencia. de la Asociación Española para la Inteligencia Artificial*, (CAEPIA), pp.469-478, 1997.
- [28] Kira, K., Rendell, L.A., "The feature selection problem: Traditional methods and a new algorithm", *Proceedings of 10th National Conference on Artificial Intelligence*, pp.129-134, 1992.
- [29] Scherf, "A new approach to feature selection". *Proceedings of the 6th Conference on Artificial Intelligence in Medicine*, (AIME'97), pp.181-184, 1997
- [30] Thawonmas, R., Abe, S., "Feature reduction based on analysis of fuzzy regions". *Proceedings of the 1995 International Conference on Neural Networks*, vol.4, pp.2130-2133, 1995
- [31] Bartlett, E.B., "Self determination of input variable inportance using neural networks". *Neural, Parallel & Scientific Computations*, vol.2, pp.103-1141, 994.
- [32] Kwok, T.Y., Yeung, D.Y., "Constructive neural networks: some practical considerations". *Proceedings. of the 1994 IEEE International Conference on Neural Networks*, vol.1, pp.198-203, 1994.

- [33] Blake, C.L., Merz, C.J.. UCI Repository of machine learning databases [<http://www.ics.uci.edu/~mlearn/MLRepository.html>]. Irvine, CA: University of California, Department of Information and Computer Science, 1998
- [34] Godino-Llorente, J. I, Aguilera-Navarro, Santiago , Hernández-Espinosa, Carlos , Fernández-Redondo, Mercedes ,Gómez-Vilda, Pedro, "On the selection of meaningful speech parameters used by a pathologic/non pathologic voice register classifier". *Sixth European Conference on Speech Communication and Technology (EUROSPEECH'99)*, pp. 563-566, Budapest, Hungary.September-1999.
- [35] Hernández-Espinosa, C.; Fernández-Redondo, M.; Gómez-Vilda P. "Diagnosis of Vocal and Voice Disorders by the Speech Signal ". *IEEE-INNS-ENNS International Joint Conference on Neural Networks*, vol.4, pp. 253-258, 2000.
- [36] Jo, C.; Kim, K.; Kim, D.; Wang, S. "Screening of Pathological Voice from ARS using Neural Network". *Pathological voice; ARS; Neural Networks*. 2001.
- [37] Salvatore, A.P.; Thome, N.; Gorss, C.M.; Cannito, M.P., "Neural network approach to speech pathology", *42nd Midwest Symposium on Circuits and Systems*, 1999, vol.439 – 442, 999.

CAPÍTULO 6

Cuantificación de la estabilidad de la voz

La **estabilidad de la voz** es la capacidad de un locutor de producir, durante la producción de un sonido sonoro sostenido, un flujo de aire con una intensidad constante para excitar las cuerdas vocales.

Este fenómeno físico se cuantifica a partir de medidas de estabilidad de la señal de voz, es decir, las diferentes medidas de calidad de la voz que cuantifican el fenómeno físico, “estabilidad de la voz”, se basan en cuantificar la variación de la intensidad de la señal de voz a lo largo del tiempo, durante la fonación de un sonido sonoro sostenido.

A continuación se presentan las diferentes medidas de calidad de la voz, de este grupo, clasificadas en los diferentes dominios de representación. Además, se presenta una ordenación de dichas medidas en función de su relevancia. Por último, dicha ordenación de medidas es evaluada de forma incremental utilizando una batería de diferentes clasificadores.

6.1 Estado de la Técnica

En este apartado de la memoria, se presenta el conjunto de las principales medidas de calidad de la voz presentes en la bibliografía que han sido clasificadas en este trabajo, como medidas que cuantifican la estabilidad de la voz.

6.1.1 Dominio Temporal

6.1.1.1 Perturbación de la amplitud: *Shimmer*

La medida de la perturbación de la amplitud, también llamada “*shimmer*”, caracteriza la variación del valor del pico positivo máximo de cada trama de voz en el dominio temporal. Se determina mediante la variación de la amplitud máxima entre tramas adyacentes.

La perturbación de la amplitud, el “*shimmer*”, ha sido definida [14] [16] [18] , [21] como:

$$AP = \left(\frac{1}{N-1}\right) \left(\frac{1}{|SA_{max}|}\right) \sum_{i=1}^{N-1} |SA(i+1) - SA(i)| \quad [\text{Ec.6-1}]$$

donde AP es el valor de la perturbación de la amplitud a lo largo del tiempo, SA es el valor del pico positivo máximo de cada trama de voz y N es el número de tramas en las que se divide el audio que se analiza.

Existen trabajos como [14] , en los que se utiliza un estimador diferente para el *Shimmer*. Este estimador se calcula de la siguiente manera:

$$AP = \left(\frac{100\%}{N-1}\right) \sum_{i=2}^N \left| \frac{SA(i) - SA(i-1)}{SA(i)} \right| \quad [\text{Ec.6-2}]$$

De forma similar se ha definido una medida genérica denominada *relative average perturbation* (RAP) [2] la cual se define según:

$$RAP = \left(\frac{1}{N-2}\right) \sum_{i=1}^{N-2} \frac{[(P(i-1) + P(i) + P(i+1))/3 - P(i)]}{\frac{1}{N} \sum_{i=1}^N P(i)} \quad [\text{Ec.6-3}]$$

cuando esta medida se aplica al valor máximo($P(i)$) de la i -ésima trama, el valor de la medida cuantifica el mismo fenómeno físico que el *shimmer*. Esta medida ha sido utilizada en trabajos como [9] [13] [12] [10]. En estos trabajos, al valor de la RAP aplicado al valor máximo de la trama se la ha denominado *Amplitude Perturbation Quotient* (APQ).

Además, se ha definido la medida denominada intensidad [17], la cual cuantifica el mismo fenómeno que el *shimmer*. Esta medida se estima como el valor de la energía expresado en dB, para ventanas de corta duración, recomendándose en el trabajo citado ventanas de 10 milisegundos.

Otra medida, denotada *Fatr* [12] [13] y similar al *shimmer*, cuantifica la frecuencia con mayor intensidad dentro de las componentes de baja frecuencia de la señal moduladora de la modulación de amplitud presente en la señal de voz. Esta modulación de la señal de voz es posible obtenerse por medio de un filtrado pasobanda alrededor del primer formante. La medida *Fatr*, se obtiene por medio de técnicas lineales. Existen estudios [11] [6] [7] [8] donde se realiza un procesamiento no lineal de esta modulación de amplitud con el fin de obtener unas medidas de la calidad de la voz.

El *shimmer* se ha estudiado también en diferentes dominios, como es el caso en el que se estima a partir de la señal obtenida tras aplicar a la señal de voz un filtrado inverso [1].

6.1.1.2 STCCF (Short Time Cross Correlation Function)

La *short time cross correlation function* (STCCF) [21] [5] cuantifica el grado de similitud en la forma de onda entre dos periodos de *pitch* consecutivos. Esta cuantificación del grado de similitud se consigue mediante el valor máximo de la función de correlación cruzada entre dos periodos de *pitch* consecutivos.

Al valor medio del STCCF para un sonido sonoro sostenido se le denomina *mean waveform matching coefficient* (MWC) [21] [5] [14].

La perturbación a lo largo del tiempo de la STCCF resulta una medida útil para cuantificar la calidad de la voz.

Una alternativa a esta medida es la denominada *Cycle-to-cycle mainform variation* [24] [4] en la que se aplica la distancia de *Chebyshev* normalizada entre dos periodos de *pitch* consecutivos:

$$AP = \left(\frac{1}{N-1} \right) \sum_{i=1}^{N-1} \left| \frac{S^{i+1} - S^i}{S_m} \right| \quad [\text{Ec.6-4}]$$

siendo S^i la forma de onda de un periodo de *pitch*, N el número de periodos de *pitch* y S_m es un periodo de *pitch* promediado, el cual es definido según:

$$S_m = \left(\frac{1}{\max(S)} \right) \sum_{i=1}^{N-1} S^i(j), \quad \text{donde } j = 0 \dots MIN \quad [\text{Ec.6-5}]$$

donde $\max(S)$ es el valor máximo de la señal de voz y MIN es la longitud mínima de los diferentes ciclos.

6.1.2 Dominio espectral

En [24] se propone una medida de la calidad de la voz basada en la cuantificación del peso de la energía espectral del primer armónico, y denominada *Normalized First Harmonic Energy* (NFHE), estimándose como el cociente entre la energía concentrada en el primer armónico frente a la energía concentrada en el segundo y sucesivos armónicos.

6.1.3 Dominio del Modelo Inverso

6.1.3.1 Kurtosis de la señal residual

Esta medida cuantifica la estabilidad del flujo de aire que excita las cuerdas vocales. Para poder estimar esta medida para una grabación de voz, es necesario previamente realizar un filtrado inverso obteniéndose así la denominada señal residual.

La *Kurtosis* (γ_4) [22] [23] es un estadístico de cuarto orden, el cual es una medida estadística que describe el apuntamiento o achatamiento de una cierta distribución de probabilidad con respecto a una distribución normal. La *kurtosis* positiva indica una distribución relativamente apuntada, y la negativa indica una distribución relativamente achatada. La *Kurtosis* se define como:

$$\gamma_4 = E\left\{\left(\frac{x - \mu}{\sigma}\right)^4\right\} \quad [\text{Ec.6-6}]$$

dado que no es conocida la función densidad espectral de la señal residual, resulta necesario recurrir a un estimador de la *kurtosis* el cual se define de la siguiente manera:

$$\hat{\gamma}_4 = \frac{1}{N} \sum_{i=1}^N \left(\frac{x_i - \mu}{\sigma}\right)^4 \quad [\text{Ec.6-7}]$$

Esta medida ha sido empleada en diferentes trabajos [15] [1] para cuantificar la calidad de la voz.

6.1.3.2 Valor cuadrático medio de la señal residual

Esta medida cuantifica la variación de la energía de la señal residual [1], y se define como:

$$VCM = \frac{1}{L} \sum_{i=1}^L (x(i) - \mu_x)^2 \quad [\text{Ec.6-8}]$$

6.1.3.3 Forma de onda glótica

La utilidad de la señal del flujo glótico ha sido estudiada para la cuantificación de la calidad de la voz [1]. En dicho estudio se realiza una descripción concienzuda de la forma de onda del flujo glótico correspondiente a cada grupo de los distintos grupos acústicos en los que se puede dividir las enfermedades en las cuerdas vocales. En dicho estudio se proponen diferentes medidas de la forma de onda del flujo glótico, las cuales se indican a continuación:

- *Speech Quotient* (SQ): respecto a este parámetro se propone la utilización según la definición clásica (duración de la fase de apertura respecto a la duración de la fase de

cierre). Además se propone la utilización del SQ con la modificación de tomar los tiempos comprendidos entre el 90%-10% y el 10%-90% de la pendiente de subida y de la pendiente de bajada de la forma de onda glótica.

- *Closing Quotient* (CQ): se propone este parámetro utilizando para su cuantificación el tiempo comprendido entre el 10% y el 90% de la pendiente de bajada de la señal del flujo glótico.
- Valor de pico y margen dinámico de un periodo de la señal residual.

6.2 Análisis

Las medidas de calidad de la voz pertenecientes al grupo de medidas de “estabilidad de la voz” presentan especial sensibilidad a dos aspectos: al ruido, especialmente al ruido no estacionario, y al movimiento del locutor respecto al micrófono. Por este motivo puede resultar recomendable, la implementación de medidas que minimicen el efecto de variaciones severas de amplitud de una trama a otra y por este motivo resulta interesante las medidas que cuantifican la variación de energía de la señal de voz a largo plazo.

Determinadas medidas de calidad de la voz, que se estiman en el dominio temporal, necesitan información sobre el valor del *pitch*. Esta situación se convierte en un inconveniente frente a las muestras de voces de muy baja calidad de voz.

6.3 Aportaciones

Además de la identificación de las medidas de calidad de la voz que cuantifican la “estabilidad de la voz”, se han aportado nuevas medidas [18] , también utilizadas en otros trabajos como [19] y [20] .

Se ha propuesto el uso del estadístico Kurtosis, de forma novedosa, en el dominio temporal. La medida consiste en la cuantificación de la variación a lo largo del tiempo del valor del Kurtosis para las distintas tramas. La justificación de esta medida está en el intento de cuantificar la variación de la energía de la señal a lo largo del tiempo por medio de la cuantificación de la variación de la forma de la función densidad de probabilidad de los valores de amplitud que toman las distintas tramas.

Con el mismo fin, se ha propuesto el uso del valor cuadrático medio, en el dominio temporal, para cuantificar la variación de la energía de la trama, como mejora respecto a tomar el valor máximo de cada trama. El valor cuadrático medio se define para una trama de la siguiente forma:

$$VCM = \frac{1}{L} \sum_{i=1}^L (x(i) - \mu_x)^2 \quad [\text{Ec.6-9}]$$

donde VCM es el valor cuadrático medio de una trama, $x(n)$ es la trama de voz la cual tiene una longitud de L muestras, y μ_x es el valor medio de la trama.

6.4 Nomenclatura

En los siguientes apartados de este capítulo se presentan los resultados del estudio de la relevancia de las diferentes medidas de calidad que cuantifican la estabilidad de la voz. Para mayor comodidad a cada una de las medidas se le ha asignado un identificador. En la Tabla 6-1 se presentan los identificadores de las diferentes medidas.

TABLA 6-1: MEDIDAS DE CALIDAD DE LA VOZ PERTENECIENTES AL GRUPO I (ESTABILIDAD DE LA VOZ)

Identificador de la medida	Grupo I : “Estabilidad de la Voz”
1	Shimmer versión 1
2	Shimmer versión 2
3	Shimmer versión 3 (APQ)
4	Shimmer versión 4 (APQ)
5	Shimmer versión 5 (Fatr)
6	Media Short Time cross corelation function
7	Variación Short Time cross corelation function
8	Clycle-to-clycle mainform variation
9	Media Kurtosis de la señal residual
10	Variación Kurtosis de la señal residual
11	Media Amplitud máxima de Ug
12	Variación Amplitud máxima de Ug
13	Media Margen dinámico de Ug
14	Variación Margen dinámico de Ug
15	Media Valor cuadrático medio
16	Variación Valor cuadrático medio
17	Media kurtosis de las diferentes tramas
18	Variación kurtosis de las diferentes tramas

6.5 Valoración de la relevancia de las Medidas de Calidad de la Voz

En este apartado se presentan los resultados del proceso consistente en ordenar las diferentes medidas de calidad que cuantifican la estabilidad de voz en función de su relevancia. El estudio se ha realizado diferenciándose dos casos: la clasificación entre “voz sana – voz patológica” y la clasificación entre “voz sana – disfonía leve – disfonía moderada – disfonía severa”. Para cada caso, se ha estudiado las medidas de calidad de la voz diferenciando la vocal fruto de la fonación.

El procedimiento de obtención de la relevancia de cada medida ha sido descrito en el apartado 5.2. Para la estimación media de la relevancia de las diferentes medidas de la calidad ha sido evaluado 100 veces y posteriormente los resultados han sido promediados.

6.5.1 Estudio: Sano – Patológico

En este caso las muestras de voz, consistentes en vocales, han sido etiquetadas o como “voz sana” o como “voz patológica”.

TABLA 6-2: ORDEN EN FUNCIÓN DE LA RELEVANCIA PARA EL ESTUDIO
SANO – PATOLÓGICO

Orden de relevancia	GRUPO I Vocal 'A'		GRUPO I Vocal 'E'		GRUPO I Vocal 'I'		GRUPO I Vocal 'O'		GRUPO I Vocal 'U'	
	Id. Medida	Prob ¹ .	Id. Medida	Prob.	Id. Medida	Prob.	Id. Medida	Prob.	Id. Medida	Prob.
1	5	65%	4	74%	3	61%	18	72%	5	65%
2	4	61%	1	48%	6	55%	1	68%	4	61%
3	18	57%	8	41%	12	48%	9	58%	18	57%
4	6	53%	6	41%	5	48%	2	51%	6	53%
5	2	48%	5	41%	13	47%	12	45%	2	48%
6	9	44%	2	40%	10	45%	15	40%	9	44%
7	15	43%	18	39%	16	43%	4	39%	15	43%
8	8	43%	9	38%	11	43%	14	36%	8	43%
9	13	39%	3	35%	14	42%	16	35%	13	39%
10	11	39%	14	32%	1	41%	6	34%	11	39%
11	3	37%	13	29%	15	40%	3	34%	3	37%
12	12	36%	16	26%	2	40%	7	32%	12	36%
13	16	33%	12	26%	9	38%	10	27%	16	33%
14	10	29%	7	24%	7	36%	11	26%	10	29%
15	1	25%	15	22%	4	29%	5	25%	1	25%
16	14	23%	11	21%	8	26%	13	24%	14	23%
17	7	21%	10	21%	17	18%	8	18%	7	21%
18	17	14%	17	17%	18	16%	17	9%	17	14%

6.5.2 Estudio: Sano – Disfonía Leve – Disfonía Moderada – Disfonía Severa

En este caso las muestras de voz, al igual que en el apartado anterior consisten en vocales y han sido etiquetadas como: “voz sana”, “disfonía leve”, “disfonía moderada” o “disfonía severa”. Este etiquetado ha sido realizado atendiendo al grado de disfonía.

¹ Prob corresponde al porcentaje de ocasiones en las que, utilizando el método de selección de medidas de calidad de la voz propuesto, la medida fue seleccionada dentro del conjunto óptimo de características.

TABLA 6-3: ORDEN EN FUNCIÓN DE LA RELEVANCIA PARA EL ESTUDIO
SANO – DISFONÍA LEVE – DISFONÍA MODERADA – DISFONÍA SEVERA

Orden de relevancia	GRUPO I Vocal 'A'		GRUPO I Vocal 'E'		GRUPO I Vocal 'I'		GRUPO I Vocal 'O'		GRUPO I Vocal 'U'	
	Id. Medida.	Prob.	Id. Medida	Prob.	Id. Medida	Prob.	Id. Medida	Prob.	Id. Medida	Prob.
1	5	60%	4	86%	1	62%	1	71%	4	67%
2	3	51%	6	65%	15	59%	9	65%	5	65%
3	7	44%	8	63%	6	55%	15	60%	6	58%
4	1	44%	5	46%	4	53%	18	51%	18	54%
5	6	42%	9	45%	7	47%	2	51%	9	53%
6	10	40%	1	36%	12	45%	7	47%	15	52%
7	18	38%	2	35%	5	45%	6	41%	11	49%
8	2	38%	10	28%	3	45%	11	37%	12	40%
9	14	35%	3	26%	16	41%	3	37%	2	38%
10	4	34%	14	24%	14	41%	13	35%	3	36%
11	9	29%	16	23%	11	41%	4	33%	13	35%
12	17	26%	7	21%	13	40%	17	32%	16	34%
13	8	24%	12	20%	8	37%	16	31%	8	33%
14	15	23%	17	19%	10	35%	14	31%	1	33%
15	16	21%	18	18%	18	33%	12	30%	10	32%
16	13	21%	15	18%	2	33%	10	23%	7	29%
17	12	20%	13	17%	9	25%	5	23%	14	26%
18	11	19%	11	10%	17	24%	8	18%	17	22%

6.6 Evaluación de las Medidas de Calidad de la Voz

En este apartado se han evaluado las medidas de calidad que cuantifican la estabilidad de la voz de forma secuencial, incrementando el número de medidas, siguiendo el orden descendente marcado por la relevancia de dichas medidas, siendo evaluadas en diferentes clasificadores: redes neuronales probabilísticas (PNN), redes neuronales *feedforward* (NN) con una capa oculta, clasificadores basados en funciones de base radial, máquinas de vectores soporte (SVM) y clasificadores basados en los K vecinos más cercanos (KNN).

Para cada subconjunto de medidas (cuya dimensión está indicada en número de medidas) y cada vocal, se obtiene la tasa de éxito en la clasificación dado un determinado clasificador. Los resultados en cada caso son promediados 20 veces.

6.6.1 Estudio: Sano – Patológico

6.6.1.1 Red Neuronal Probabilística (PNN)

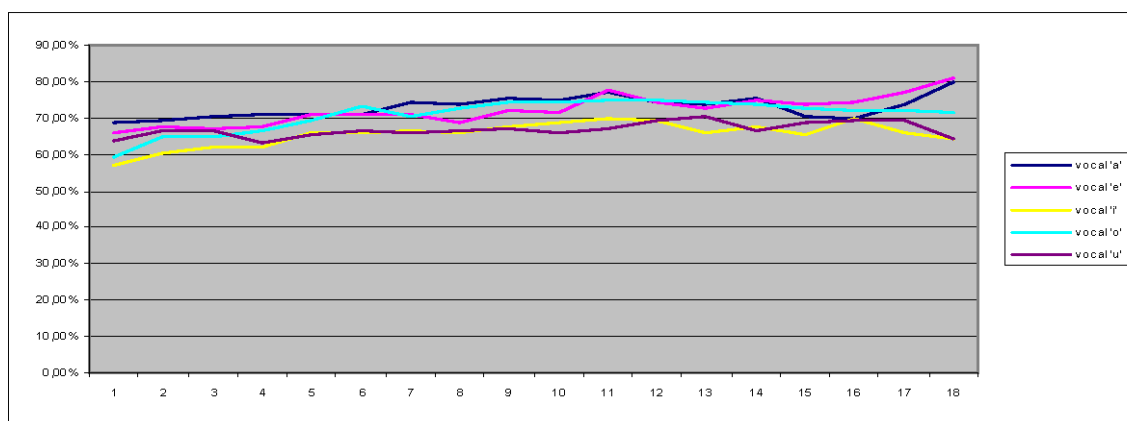


Gráfico 6-1: Evaluación secuencial de la relevancia para las medidas de calidad de la voz del grupo I usando el clasificador PNN en la clasificación entre voz sana y voz patológica

6.6.1.2 Red Neuronal *FeedForward* con 1 Capa Oculta (NN)

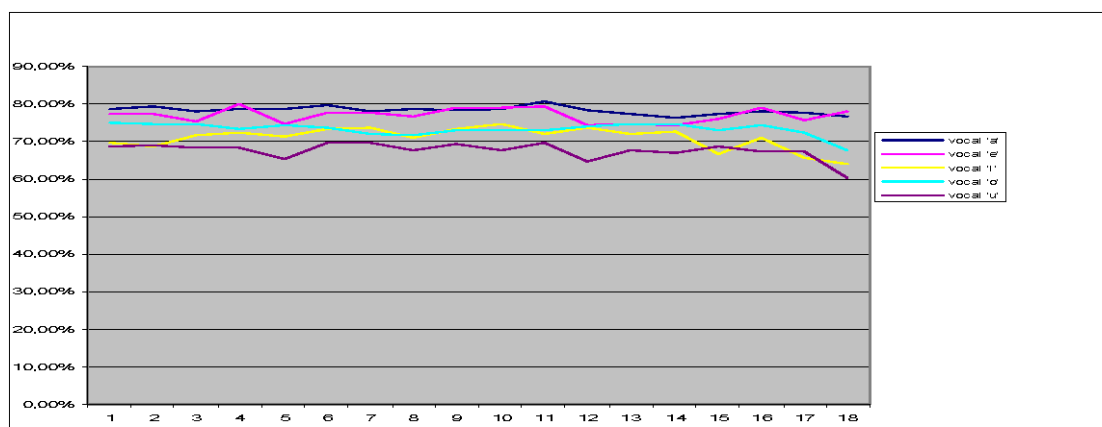


Gráfico 6-2: Evaluación secuencial de la relevancia para las medidas de calidad de la voz del grupo I usando el clasificador NN en la clasificación entre voz sana y voz patológica

6.6.1.3 Función de Base Radial (RBF)

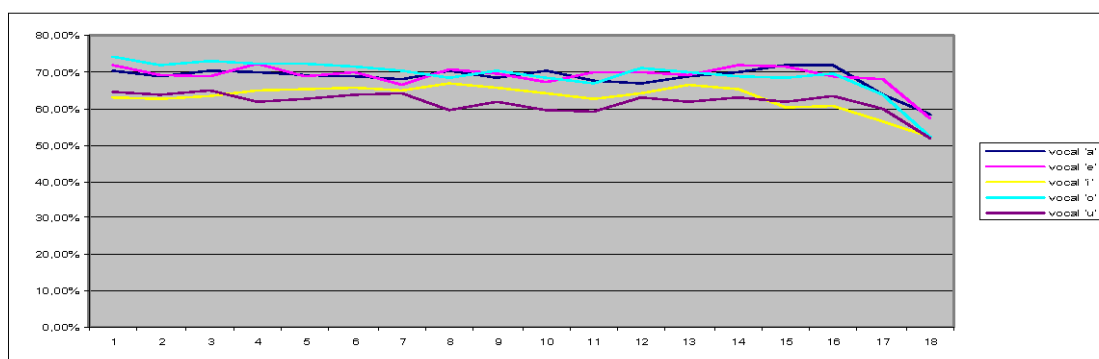


Gráfico 6-3: Evaluación secuencial de la relevancia para las medidas de calidad de la voz del grupo I usando el clasificador RBF en la clasificación entre voz sana y voz patológica

6.6.1.4 Máquinas de Vectores Soporte (SVM)

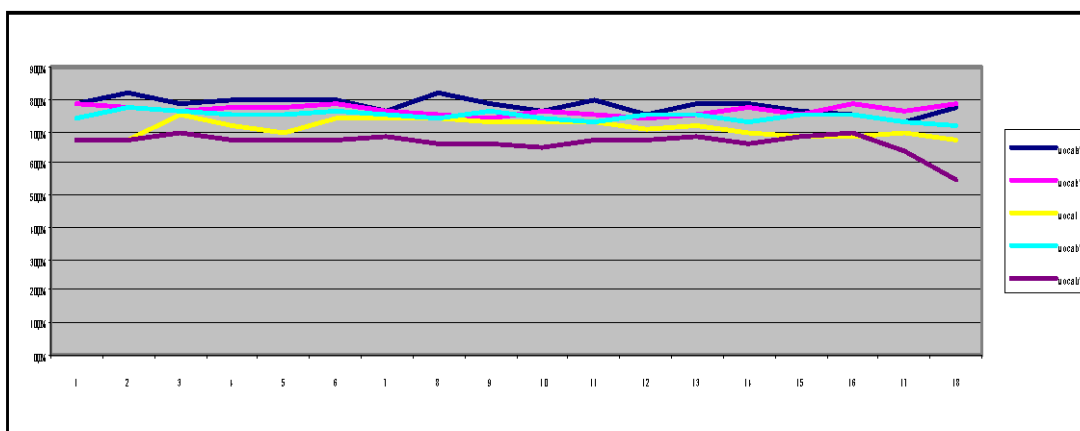


Gráfico 6-4: Evaluación secuencial de la relevancia para las medidas de calidad de la voz del grupo I usando el clasificador SVM en la clasificación entre voz sana y voz patológica

6.6.1.5 K vecinos más cercanos (KNN)

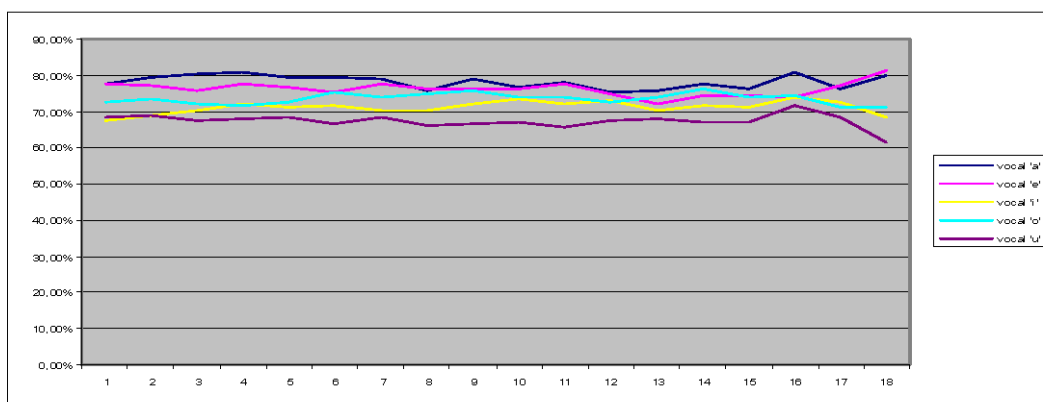


Gráfico 6-5: Evaluación secuencial de la relevancia para las medidas de calidad de la voz del grupo I usando el clasificador KNN en la clasificación entre voz sana y voz patológica

6.6.2 *Estudio: Sano – Disfonía Leve – Disfonía Moderada – Disfonía Severa*

6.6.2.1 Red Neuronal Probabilística (PNN)

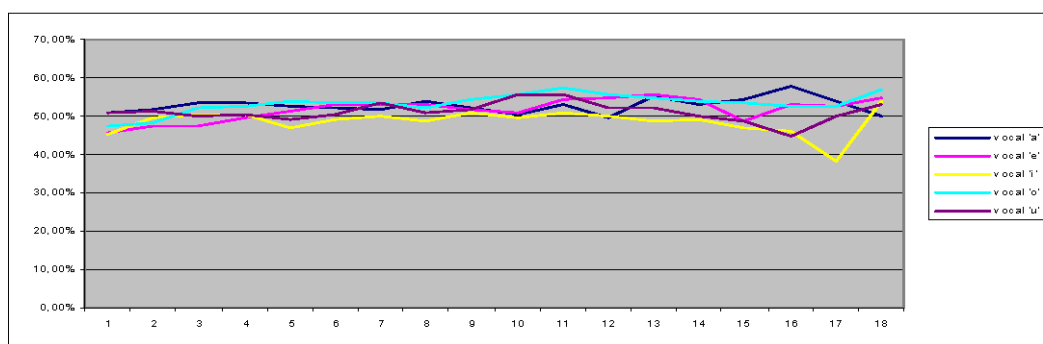


Gráfico 6-6: Evaluación secuencial de la relevancia para las medidas de calidad de la voz del grupo I usando el clasificador PNN en la clasificación entre distintos grados de disfonía

6.6.2.2 Red Neuronal *FeedForward* con 1 Capa Oculta (NN)

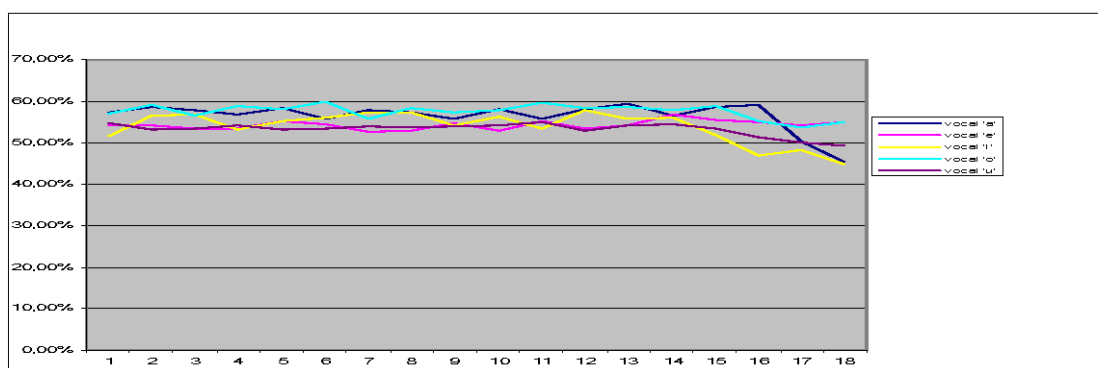


Gráfico 6-7: Evaluación secuencial de la relevancia para las medidas de calidad de la voz del grupo I usando el clasificador NN en la clasificación entre distintos grados de disfonía

6.6.2.3 Función de Base Radial (RBF)

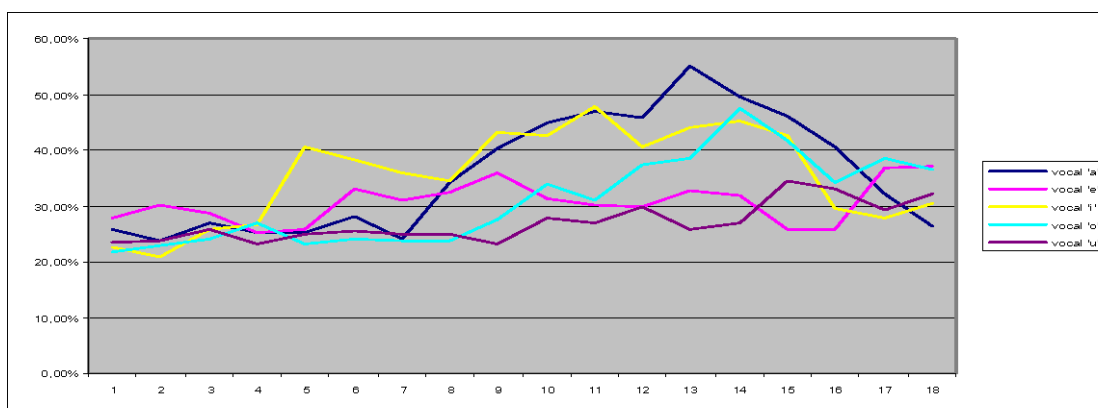


Gráfico 6-8: Evaluación secuencial de la relevancia para las medidas de calidad de la voz del grupo I usando el clasificador RBF en la clasificación entre distintos grados de disfonía

6.6.2.4 Máquinas de Vectores Soporte (SVM)

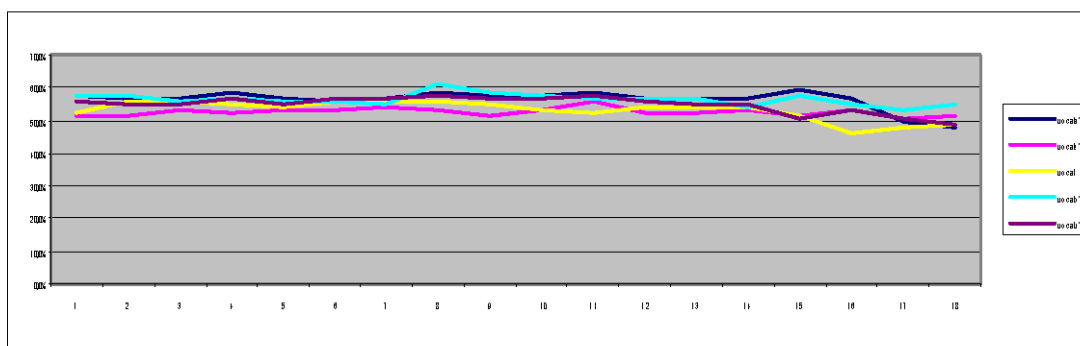


Gráfico 6-9: Evaluación secuencial de la relevancia para las medidas de calidad de la voz del grupo I usando el clasificador PNN en la clasificación entre distintos grados de disfonía

6.6.2.5 K vecinos más cercanos (KNN)

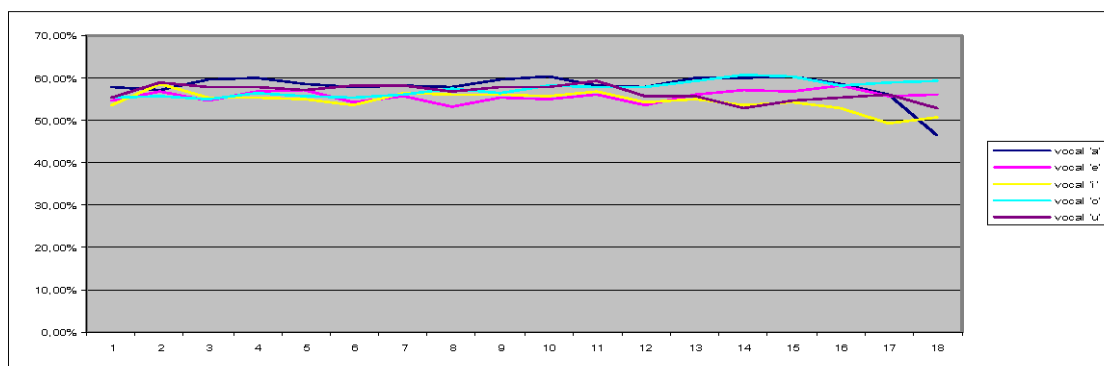


Gráfico 6-10: Evaluación secuencial de la relevancia para las medidas de calidad de la voz del grupo I usando el clasificador KNN en la clasificación entre distintos grados de disfonía

6.7 Evaluación

En la clasificación entre voz sana y voz patológica se observa que para la PNN existe un ligero incremento de la tasa de éxito a medida que se incrementa el número de medidas de calidad, estando acotado el rango entre el 60 y 80%. Para el RBF existe una relativa estabilidad de la tasa de éxito, estando acotado el éxito entre el 60 y 70%. El resto de los clasificadores, el NN, SVM y KNN, se muestran con una tasa de éxito totalmente estable a medida que se incrementa el número de medidas de calidad, estando el rango de éxito acotado entre el 70 y el 80%.

En el caso de la clasificación entre voz sana, disfonía leve, disfonía moderada y disfonía severa, se observa que para la PNN existe un ligero incremento de la tasa de éxito, acotada entre el 50-60%. En la RBF, existe significativos incrementos que van desde el 20% hasta el 50%. Sin embargo, el NN, SVM y el KNN, mantienen una tasa de éxito totalmente constante, entre el 50-60%, pese al incremento de medidas de calidad de la voz.

Se observa de forma general para los diferentes clasificadores y los distintos casos estudiados (clasificación entre voz sana y voz patológica, y clasificación entre voz sana, disfonía leve, disfonía moderada y disfonía severa) que la tasa de éxito en la clasificación no se incrementa de forma significativa cuando a la medida de calidad que presenta la mayor relevancia se evalúa con otras medidas más, principalmente para NN, SVM y RBF. En la interpretación de los resultados es necesario tener en cuenta la reducida dimensión de la base de datos utilizada para la evaluación, lo cual puede justificar las fluctuaciones existentes.

Por este motivo, en esta tesis doctoral, consideramos que con una única medida es posible cuantificar el fenómeno físico “estabilidad de la voz”, y que para cada vocal, la medida que presenta mayor relevancia es la medida de calidad que mejor cuantifica este fenómeno.

Por lo tanto es posible asociar, para cada vocal por separado, la cuantificación del fenómeno físico “estabilidad de la voz” a una única medida, la cual se ha identificado como la medida que presenta mayor relevancia.

6.8 Referencias

- [1] K. Marasek, "EGG and Voice Quality", <http://www.ims.uni-stuttgart.de/phonetik/EGG/frmst1.htm>, 1997. [Consulta 10/10/05].
- [2] Boyanov, B; Hadjitodorov, S. "Acoustic analysis of pathological voices. A voice analysis system for the screening of laryngeal diseases". *IEEE Engineering in Medicine and Biology Magazine*, vol.16,no.4, pp.74 – 82, 1997.
- [3] Feijoo S.; Hernandez C.; Carmelo R. "Multidimensional analysis of phonological degeneration in pathological voices". *IEEE International conference on Acoustics, Speech and Signal Processing*, ICASSP 86, vol.1, pp.673 – 676, 1986;
- [4] Feijoo, S, et al. "Acoustic evaluation of glottal cancer based on short-term stability measures", *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, vol.2, pp.675-676, 1989.
- [5] Fröhlich, M.; Michaelis, D.; Strube, H.; Kruse, E. "Acoustic voice quality description: Case studies for different regions of the hoarseness diagram". *Advances in Quantitative Laryngoscopy*, 2nd 'Round Table', Erlangen. pp. 143-150, 1997.
- [6] Gavidia-Ceballos, L.; Hansen, J.H.L. "Direct Speech Feature Estimation using an Iterative EM Algorithm for Vocal Fold Pathology Detection". *Biomedical Engineering*, IEEE Transactions on. 1996; 43(4):373 - 383.
- [7] Gavidia-Ceballos, L.; Hansen, J.H.L.; Kaiser, J.F. "Vocal fold pathology assessment using AM autocorrelation analysis of the Teager energy operator". *Proceedings of Fourth International Conference on Spoken Language*, ICSLP 96, vol.2, pp. 757-760, 1996.
- [8] Gavidia-Ceballos, L.; Hansen, J.H.L.; Riski, J.E., "A non-invasive speech processing method for analysis of vocal fold cancer based on an iterative EM algorithm". *Proceedings of the 16th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, vol.2, pp.1308–1309, 1994.
- [9] Godino-Llorente, J. I, et al. "On the selection of meaningful speech parameters used by a pathologic/non pathologic voice register classifier". *Sixth European Conference on Speech Communication and Technology (EUROSPEECH'99)*, pp.563-566, 1999
- [10] Godino-Llorente, J.I.; Aguilera-Navarro, S.; Gomez-Vilda, P. "Non supervised neural net applied to the detection of voice impairment". *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing*, ICASSP '00. vol.6, pp.3594-3597, 2000.
- [11] Hansen, J.H.L.; Gavidia-Ceballos, L.; Kaiser, J.F. "A nonlinear operator-based speech feature analysis method with application to vocal fold pathology assessment". *IEEE Transactions on Biomedical Engineering*,. Vol.45, no.3, pp.300-313. March 1998.
- [12] Hernández-Espinosa, C.; Fernández-Redondo, M.; Gómez-Vilda P. "Diagnosis of Vocal and Voice Disorders by the Speech Signal ". *IEEE-INNS-ENNS International Joint Conference on Neural Networks*. vol.4, pp.253-258, 2000
- [13] Jo, C.; Kim, K.; Kim, D.; Wang, S. "Screening of Pathological Voice from ARS using Neural Network". *International Workshop on MAVEBA*, pp.13-15, 2001.
- [14] Michaelis D.; Frohlich M.; Strube H. W. "Selection and combination of acoustic features for the description of pathologic voices". *Acoustical Society of America*, vol.103, no.3, pp.1628-1640, 1998.

- [15] O. Rosa, M.; Pereira, J.C.; Carvalho, A.C.P.L.F. "Evaluation of neural classifiers using statistic methods for identification of laryngeal pathologies". *Proceedings of Vth Brazilian Symposium on Neural Networks*, pp. 220 – 225, 1998.
- [16] Wallen, E.J.; Hansen, J.H.L."A screening test for speech pathology assessment using objective quality measures". *Proceedings of Fourth International Conference on Spoken Language*, ICSLP 96, vol.2, pp.776 – 779, 1996.
- [17] Wester, M. "Automatic Classification Of Voice Quality: Comparing Regression Models and Hidden Markov Models". *Proceedings of VOICEDATA98*, Symposium on Databases in Voice Quality Research and Education, pp.92-97, 1998.
- [18] Alonso, J. B.; de León, J.; Alonso, I.; Ferrer, M. A. "Automatic Detection of pathologies in the voice by Hos based parameters". *Eurasip journal on, Applied signal processing*, vol.4, pp.275-284, 2001.
- [19] Alonso, JB.; Carmona, C.; de Leon, J.; Ferrer, M.A. "Combining Neural Network and Hidden Markov models for automatic detection of pathologies". *16-th Biennial International EURASIP Conference Biosignal*. pp.466-469,2002.
- [20] Alonso, J. B.; de León, J.; Ferrer, M. A. "Automatic Detection of Pathologies in the voice by HOS based parameters". NSIP June 3-6, 2001, Baltimore, Maryland USA.
- [21] Dirk Michaelis y Hans Werner Strube, "Empirical Study to Test the Independence of Differen Voice Parameters on a Large Voice Database", *Procedings of EUROSPEECH '95*, vol.3,pp.1891-1893, 1995.
- [22] Jerry M. Mendel, "Tutorial on Higher-Order statistics (Spectra) in Signal Processing and System Theory: Theoretical Results and Some Applications", *Proceedings of the IEEE*, vol.79, no.3,1991.
- [23] Chrysostomos L. Nikias y Mysore R. Raghuveer, "Bispectrum estimation: A Digital Signal Processing Framework", *in Proceedings of the IEEE*, vol.75, no.7, 1987.
- [24] Stefan Hadjitodorov, Petar Mitev, "A computer system for acoustic analysis of pathological voices and laryngeal diseases screening", *Medical Engineering Physics*, vol .24, pp. 419-429, 2002.

CAPÍTULO 7

Cuantificación de la Riqueza Espectral

La **Riqueza Espectral**: se define como la capacidad de generar, durante la fonación de un sonido sostenido, un movimiento periódico de las cuerdas vocales y de producir una excitación sonora al tracto vocal con gran cantidad de componentes espectrales. Este fenómeno físico se cuantifica por medio del cálculo de la estabilidad de la frecuencia de *pitch* y por la presencia de armónicos con energía elevada en las diferentes bandas de frecuencias.

A continuación, al igual que en el capítulo anterior, se presentan las diferentes medidas de calidad de la voz pertenecientes al grupo de “riqueza espectral”, tanto las presentes en la bibliografía actual como las propuestas en este trabajo, clasificadas en los diferentes dominios de representación. Además, se presenta una ordenación de dichas medidas en función de su relevancia. Por último, dicha ordenación de medidas es evaluada de forma incremental utilizando una batería de diferentes clasificadores.

7.1 Estado de la Técnica

En este apartado de la memoria, se presenta el conjunto de las principales medidas que cuantifican la riqueza espectral de la voz presentes en la bibliografía y que han sido clasificadas en este trabajo.

7.1.1 Dominio Espectral

7.1.1.1 Perturbación del Pitch: Jitter

Esta medida cuantifica la variación a lo largo del tiempo del valor del *pitch*, siendo esta la frecuencia a la que vibran las cuerdas vocales. Se define la perturbación del valor de *Pitch* [20] [3], “*jitter*”, como:

$$PP = \left(\frac{1}{N-1}\right) \left(\frac{1}{|SP_{max}|}\right) \sum_{i=1}^{N-1} |SP(i+1) - SP(i)| \quad [\text{Ec.7-1}]$$

donde *PP* es el valor de la perturbación del *pitch* a lo largo del tiempo, *SP* es el valor del *pitch* para cada trama de voz y *N* es el número de tramas en las que se divide el audio que se analiza.

Existen diferentes definiciones de esta medida tal como la definida según la siguiente expresión [18] :

$$PP = \left(\frac{100\%}{N-1}\right) \sum_{i=2}^N \left| \frac{SP(i) - SP(i-1)}{SP(i)} \right| \quad [\text{Ec.7-2}]$$

o una versión modificada [16] en la que se sustituye la sustracción del valor del *pitch* de la ventana anterior por el valor promediado de *pitch* de *K* ventanas, tal y como se describe en la siguiente expresión:

$$PP = \left(\frac{100\%}{N-K}\right) \sum_{i=\frac{k}{2}}^{N-\frac{K}{2}} \left| \frac{SP(i) - \frac{1}{k} \sum_{d=-\frac{k}{2}}^{\frac{k}{2}} SP(i+d)}{\frac{1}{k} \sum_{d=-\frac{k}{2}}^{\frac{k}{2}} SP(i+d)} \right| \quad [\text{Ec.7-3}]$$

De forma similar se ha definido una medida genérica denominada perturbación promedio relativa (RAP) [3] la cual se define según:

$$RAP = \left(\frac{1}{N-2} \right) \sum_{i=1}^{N-2} \frac{|[P(i-1) + P(i) + P(i+1)]/3 - P(i)|}{\frac{1}{N} \sum_{i=1}^N P(i)} \quad [\text{Ec. 7-4}]$$

Cuando esta medida se aplica a la frecuencia del *pitch* $P(i)$ de la i -ésima trama, el valor de la medida cuantifica el mismo fenómeno físico que el *Jitter*. Al valor de la RAP aplicado a la frecuencia del *pitch* se la ha denominado *pitch period perturbation quotient* (PPQ) [9] [14] [13] [11].

Este parámetro es ampliamente utilizado para la cuantificación de la calidad de la voz. Se ha estudiado en diferentes dominios, como en el caso en el que se estima el valor del *Jitter* a partir de la señal obtenida tras aplicar un filtro inverso a la señal de voz [2].

Existen trabajos [5] [11] [14] [13] [9] en los que a partir de la serie temporal que se obtiene al estimar la frecuencia de vibración de las cuerdas vocales para cada una de las diferentes tramas, se divide esta señal en dos: la señal correspondiente a las altas frecuencias HFPV (*high frequency pitch variation*) y la señal correspondiente a las bajas frecuencias denominada *Tremor*.

Otra medida, similar al *Jitter*, es la que cuantifica la frecuencia con mayor intensidad dentro de las componentes de baja frecuencia de la señal moduladora de la modulación en frecuencia presente en la señal de voz (Fftr) [13] [14]. Esta modulación de la señal de voz es posible obtenerse por medio de un filtrado pasobanda alrededor del primer formante. El parámetro Fftr, se obtiene por medio de técnicas lineales.

7.1.1.2 DUV (Degree of unvoiceness)

Esta medida cuantifica la presencia de periodos de tiempo durante la fonación de un sonido sostenido en los que las cuerdas no vibran o vibran de forma muy irregular.

Se calcula como la relación entre el número de tramas sordas y el número total de tramas durante la fonación de una vocal sostenida [3] [11] . El DUV se calcula mediante la siguiente ecuación:

$$DUV = \frac{N_{tramas\ sordas}}{N_{tramas}} \quad [Ec.7-5]$$

donde $N_{tramas\ sordas}$ es el número de tramas sordas y N_{tramas} es el número total de tramas.

7.1.1.3 SHR (Subharmonic-to-harmonic ratio)

Existen trabajos [19] en los que se indica la presencia de subarmónicos del *pitch*. Esta medida de la calidad de la voz se define como:

$$SHR = \frac{SS}{SH} \quad [Ec.7-6]$$

donde ambos términos se definen según las siguientes expresiones:

$$SH = \sum_{n=1}^N X(nf_o) \quad [Ec.7-7]$$

$$SS = \sum_{n=1}^N X((n - \frac{1}{2})f_o) \quad [Ec.7-8]$$

siendo X el espectro de la señal de voz.

7.1.2 *Dominio Cepstral*

7.1.2.1 Medidas basadas en los *cepstrum*

7.1.2.1.1 CE (*Cepstral Energy*)

Por medio de esta medida se estudia la información que proporcionan los *cepstrum* complejos acerca de la excitación de una voz sonora. Estudios realizados [4] [20] demuestran que la región de altas “*quefreny*” contiene información sobre el grado de periodicidad de la señal. Este grado de periodicidad se cuantifica por medio de la amplitud máxima del pico cepstral principal en el rango de las altas *quefreny*.

En la Figura 7-1 se muestra los *cepstrum* complejos de una trama de una muestra de voz de alta calidad (color azul) y una trama de voz de baja calidad (color

rojo). Ambas muestras corresponden a locutores masculinos. En dicha figura se puede apreciar como el valor máximo de los *cepstrum* complejos se obtiene para la muestra de voz de alta calidad.

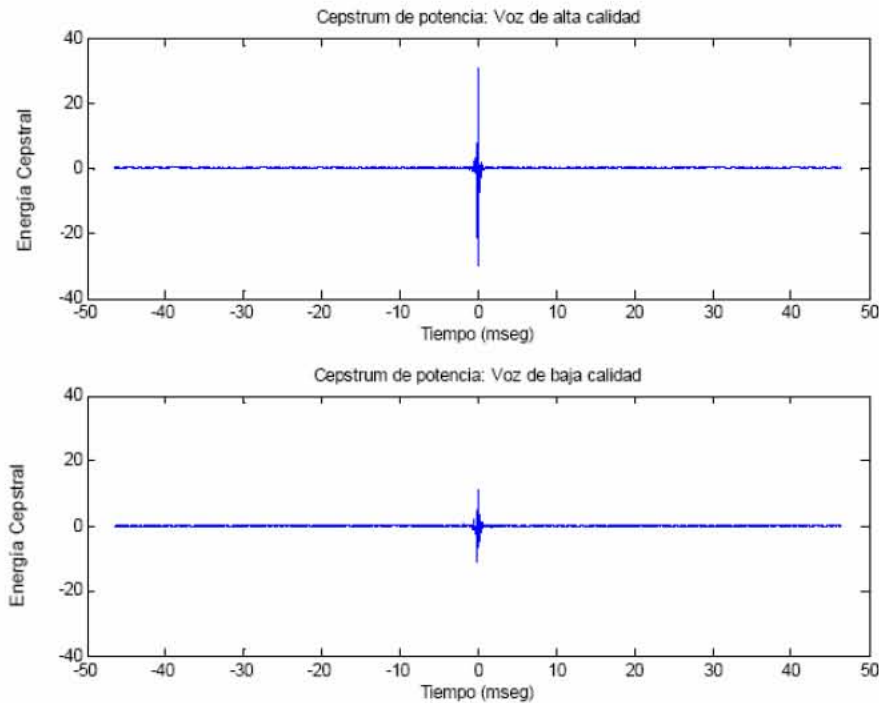


Figura 7-1: Cepstrum complejos de una muestra de voz sana y de una muestra de voz patológica.

Este parámetro, denominado *Cepstral Energy* (CE), se calcula de la siguiente manera:

$$CE = \frac{1}{N_{tramas}} \sum_{n=1}^{N_{tramas}} PC_n \quad [Ec.7-9]$$

donde PC_n es la amplitud máxima del pico cepstral principal de los *cepstrum* complejos de la trama n , y N_{tramas} el número total de tramas en las que se divide la señal de voz sonora que se pretende analizar.

7.1.2.1.2 PCEM (*Pitch energy cepstral measure*)

Esta medida cuantifica la capacidad del locutor de hacer vibrar las cuerdas vocales a la misma frecuencia a lo largo de la fonación. Esta medida calcula dicha capacidad en cortos periodos de tiempo. Además, cuantifica la relación porcentual de ruido existente en la voz. Su cálculo se realiza en el dominio de los *cepstrum* de potencia.

En la Figura 7-2 se muestra una representación de los *cepstrum* de potencia para una muestra de voz de alta calidad y para una muestra de voz de baja calidad.

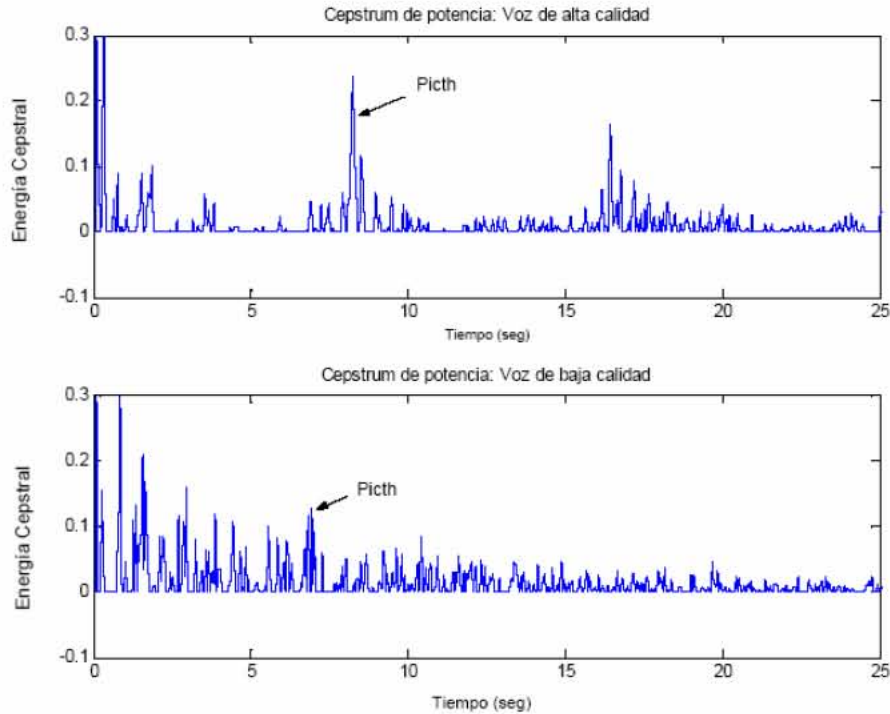


Figura 7-2: Representación en el dominio de los *cepstrum* de potencia de una muestra de voz de voz de alta calidad (superior) y para una muestra de voz de baja calidad (inferior)

La medida de energía cepstral de *pitch* PECM [3] evalúa la relación entre la energía *cepstral* concentrada en el pulso del *pitch* y la energía cepstral total de cada ventana, en el contexto de los *cepstrum* de potencia rectificados. Para localizar el pulso del *pitch* es necesario previamente haberlo localizado mediante un detector de *pitch*. El PECM se calcula para cada trama. La expresión del PECM es la siguiente:

$$PECM = \frac{E_{\text{pulso del pitch}}}{E_{\text{total de la trama}}} \quad [\text{Ec.7-10}]$$

donde la energía cepstral del pulso del *pitch* se define como:

$$E_{\text{pulso del pitch}} = \sum_{n=n_{\text{inicio}}+1}^{n_{\text{fin}}} c(n) \quad [\text{Ec.7-11}]$$

siendo: $c(n)$ los coeficientes cepstrales que forman el pulso del *pitch*, n_{inicio} es el número de la muestra después del último cruce por cero que se produce a la izquierda de la posición cepstral del *pitch*, y n_{fin} es el número de la muestra antes del primer cruce

por cero que se produce a la derecha de la posición cepstral del *pitch*. La energía cepstral total definida como:

$$E_{\text{total de la trama}} = \sum_{n=0}^L c(n) \quad [\text{Ec.7-12}]$$

donde L es el número total de coeficientes *cepstrum* de cada trama.

La medida útil para cuantificar la calidad de la voz es la perturbación a lo largo del tiempo de PCEM durante la fonación de un sonido sostenido. Esta medida no se puede estimar durante los fragmentos de tiempo en los que se produce una ausencia de vibración de las cuerdas vocales.

7.1.2.1.3 APR (*Aperiodic periodic ratio*)

La medida APR [25] proporciona una medida de la relación de la parte armónica con respecto al ruido. En este caso se relaciona la energía cepstral del *pitch* y del segundo armónico con respecto a la energía procedente del tracto vocal, los coeficientes *cepstrales* inferiores al pico de *pitch*.

$$APR = \frac{\sum_{n=\text{inicio\_pitch}}^{\text{fin\_pitch}} c(n) + \sum_{n=\text{inicio\_2º armónico}}^{\text{fin\_2º armónico}} c(n)}{\sum_{n=1}^{\text{inicio\_pitch}} c(n)} \quad [\text{Ec.7-13}]$$

7.1.2.2 Medidas basadas en los *Mel-Cepstrums*

En estudios referentes a reconocimiento de voz, como por ejemplo [1], se aplica una transformación en el dominio espectral con el fin de realizar una representación del espectro sobre una escala de frecuencias no lineal, la cual corresponde a una distribución de las frecuencias que se aproxima más a la percepción real de los sonidos por parte de las personas.

Se han distinguido diferentes transformaciones del eje frecuencial, como son la escala *Bark* y la escala *Mel*. Ambas se definen a continuación:

$$f_{Bark} = 13 \cdot \arctan\left(\frac{0.76f}{1000}\right) + 3.5 \cdot \arctan\left(\frac{f^2}{(7500)^2}\right) \quad [\text{Ec.7-14}]$$

$$f_{Mel} = 2595 \cdot \log_{10}\left(\frac{1+f}{700}\right) \quad [\text{Ec.7-15}]$$

Sobre esta transformación de la escala de frecuencias también se ha calculado los coeficientes *cepstrum*. Esta transformación de un espectro con escala no lineal, en el caso de la escala *Mel*, ha dado lugar a los denominados *Mel-Cepstrum*. Sobre este dominio modificado de los *cepstrum* se han estimado diferentes medidas de la calidad de la voz [15].

7.1.2.3 Medidas basadas en los *Delta-Cepstrums*

En el dominio de los *cepstrum*, con el fin de plasmar la variación a lo largo del tiempo de los diferentes coeficientes cepstrales, se han definido los *delta-Cepstrum* [1]. Para su cálculo, se realiza una derivada temporal de los *cepstrum* obtenidos de las diferentes tramas. Para cada componente cepstral se define su correspondiente componente en el dominio *delta-Cepstrum* según:

$$\dot{C}_i(n) = C_i(n) - C_i(n-1) \quad [\text{Ec.7-16}]$$

donde $C_i(n)$ corresponde a la componente cepstral i de la trama n , mientras que $C_i(n-1)$ a la componente cepstral i de la trama $n-1$.

Este dominio ha sido modelado para cuantificar la calidad de la voz [15].

7.1.2.4 Medidas basadas en los *Delta-Mel-Cepstrums* (MFCC)

A partir del dominio cepstral, combinando la transformación no lineal de las componentes frecuenciales según la escala *Mel*, y calculando los *cepstrum* derivados a partir de los *mel-cepstrum* se obtiene la señal denominada en la bibliografía MFCC. Esta señal ha sido utilizada para estudiar la calidad de la voz [10] [15].

7.1.3 Dominio del Modelo Inverso

En diferentes trabajos se han propuesto medidas de la calidad de la voz cuantificadas a partir de la forma de onda glótica, obtenida a partir de un filtrado inverso [2] . Estas medidas son:

- *Open quotient*
- *Speech quotient*
- *Closing quotient*
- Amplitud máxima de dU_g
- Amplitud mínima de dU_g
- Margen dinámico de dU_g
- Momento de pico normalizado
- Momento de excitación normalizado
- Momento de cierre normalizado
- Momento de apertura normalizado

Una medida similar a la PCEM (medida de energía cepstral del *pitch*) en el ámbito del modelo inverso, es la que cuantifica la variación de la amplitud del segundo pico de la señal obtenida de la función de autocorrelación normalizada de la señal residual [17] .

7.2 Análisis

Las medidas que cuantifican la estabilidad de la vibración de las cuerdas vocales durante la fonación de un sonido sonoro sostenido presentan especial sensibilidad al detector de *pitch*. Esta sensibilidad se incrementa frente a la presencia de ruidos no estacionarios y relaciones señal ruido muy bajas. Esto es posible minimizarlo por medio de detectores de *pitch* que minimicen la posibilidad de detectar frecuencias doble o mitad de *pitch* donde además resulta de interés la medida de la variabilidad del *pitch* a largo plazo, durante la fonación de un sonido sonoro sostenido.

Determinas medidas de calidad de la voz, necesitan la información del valor del *pitch*, para su estimación. Sin embargo, en algunos casos de fonaciones “sonoras” de muy baja calidad, no existe una clara vibración de las cuerdas vocales, lo que produce la imposibilidad de estimar la medida. Este hecho que aparentemente resulta muy dañino para el estudio de la voz, consideramos que es todo lo contrario. Si se cuantifica adecuadamente, los periodos de ausencia de vibración de las cuerdas durante una fonación “sonora” pueden ser muy adecuados para cuantificar la calidad de la voz.

Otro aspecto a destacar en las medidas de calidad que cuantifican la riqueza espectral, es que la información no está en la cuantificación de la cantidad de armónicos existentes en una fonación sonora sostenida sino que lo verdaderamente interesante para cuantificar la calidad de la voz es cuantificar la variación a lo largo del tiempo de la cantidad de armónicos. El que un locutor tenga más o menos riqueza espectral es simplemente una característica más de las peculiaridades de la voz de un locutor aunque sin embargo, niveles muy bajos de armónicos son indicios de una baja calidad de la voz.

Una representación de la señal de voz en el dominio de los *Cepstrum*, o cualquier otro dominio de representación derivado, permite visualizar, y consecuentemente cuantificar, con gran facilidad los niveles de riqueza espectral. Sin embargo, las medidas de calidad de la voz obtenidas a partir de filtrados inversos del tracto vocal, presentan la necesidad de muy buenos algoritmos de estimación de la forma de onda glótica, principalmente frente a la presencia de ruidos estacionarios o niveles de relación señal a ruido bajas, además de los casos en los que la señal de voz presenta bajos niveles de periodicidad.

7.3 Aportaciones

Además de la identificación de las medidas de calidad de la voz que cuantifican la “riqueza espectral”, se han realizado diferentes aportaciones de medidas para dicho fenómeno físico. Igualmente se ha clasificado a cada una de las diferentes medidas atendiendo al dominio de representación utilizado para su estimación.

7.3.1 Dominio Espectral

7.3.1.1 UPR (Unvoiced periodicity Ratio)

Esta medida complementa la medida denominada grado de sonoridad (DUV), la cual cuantifica la presencia de periodos de tiempo durante la fonación de un sonido sostenido en los que las cuerdas no vibran o vibran de forma muy irregular. La medida descrita en esta apartado (UPR) cuantifica la frecuencia de las ocurrencias de periodos de tiempo en los que ocurre el comportamiento irregular de las cuerdas vocales.

La medida UPR cuantifica si la producción de tramas sordas a lo largo de la fonación de un sonido sonoro se produce de forma consecutiva o de forma alternada. Esta medida, presentada en [21] , cuantifica el número de transiciones entre “trama sorda - trama sonora” o bien entre “trama sonora – trama sorda” respecto al número de tramas totales.

$$UPR = \frac{\text{número de transiciones}}{N_{\text{tramas}}} \quad [\text{Ec.7-17}]$$

7.3.1.2 GST (Grado de sonoridad por trama)

Esta medida de la calidad de la voz está basada en el funcionamiento del codificador paramétrico de voz *Multiband Excitation Vocoder* (MBE). Esta medida cuantifica la riqueza espectral de cada trama [24] . Para ello, identifica las componentes del espectro que pertenece al *pitch* y todos sus armónicos: ésto nos permite dividir el espectro de cada trama en diferentes regiones, cada una de ellas correspondientes a un armónico (véase la Figura 7-3). Posteriormente se etiqueta cada una de las diferentes bandas en “banda armónica” o “banda no armónica” en función de la energía máxima de la banda que supera un determinado umbral. Este etiquetado permite calcular la medida GST, la cual se define como el porcentaje de bandas del espectro de una trama que son etiquetadas como armónicas con respecto al número total de bandas en la trama.

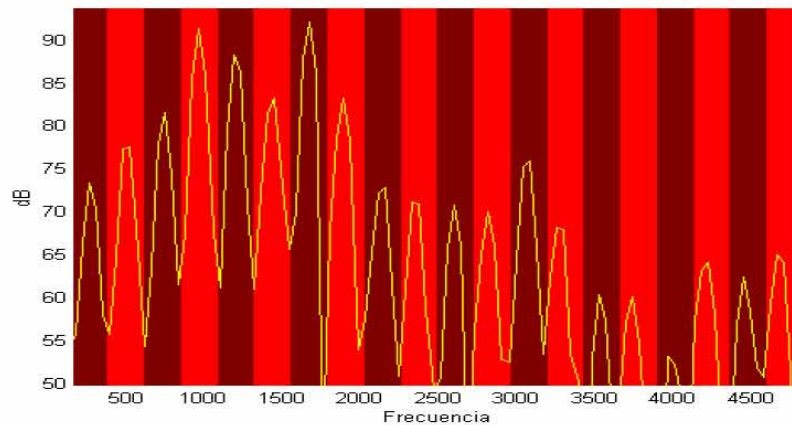


Figura 7-3: Representación de las diferentes bandas en las que se divide el espectro de una trama de voz en el proceso de estimación del grado de sonoridad por trama

7.3.2 Dominio Cepstral

7.3.2.1 Energía Cepstral del Segundo Armónico (SHECM)

Este parámetro, de forma similar al PCEM, cuantifica la capacidad del locutor de producir, de forma constante a lo largo del tiempo que dura la fonación de un sonido sonoro sostenido, un segundo armónico con estabilidad y con energía significativamente diferente respecto al componente de ruido de la señal. En la Figura 7-4 se aprecia la localización del primer armónico (el *pitch*) y el segundo armónico en el dominio de los *cepstrum* de potencia.

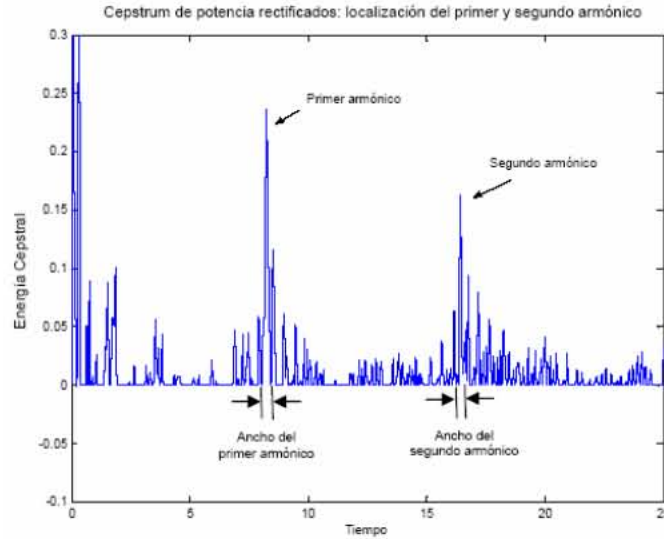


Figura 7-4: Localización del primer y segundo armónico en el dominio de los cepstrum reales

La medida SHCEM se define de forma similar al PECM [21], con el hecho diferencial de cuantificar la energía del pico cepstral centrado en el segundo armónico respecto a la energía total de la ventana de los *cepstrum*. Al igual que ocurre en el caso de la medida PECM, la perturbación a lo largo del tiempo de la medida SHCEM, durante la fonación de un sonido sonoro sostenido, resulta útil para medir la calidad de la voz.

7.3.2.2 Cepstrum derivados espacialmente

Con el fin de cuantificar las manifestaciones del fenómeno “riqueza espectral” que se obtienen de una representación cepstral de la señal de voz, se ha estudiado la derivada espacial de una ventana cepstral [21]. Estudiándose el comportamiento del primer armónico (*pitch*) y del segundo armónico en el dominio de la derivada espacial de los *cepstrum*. Para la obtención de la derivada espacial de los cepstrum se ha utilizado una aproximación a un diferenciador ideal mediante un filtro de fase lineal, el cual se define de la siguiente manera:

$$\dot{s}(n) \equiv \frac{d}{dt} s(n) \approx \sum_{m=-N_d}^{N_d} m \cdot s(n+m) \quad [\text{Ec.7-18}]$$

donde es práctica habitual un N_d igual a 5 muestras, para una frecuencia de muestreo 22050 Hz.

El resultado obtenido en la transformación de una trama de voz al dominio de los *cepstrum* derivados espacialmente se muestra en la Figura 7-5, donde además se indica la localización de los dos primeros picos positivos, que corresponden a las pendientes de subida del primer armónico y del segundo armónico, siendo proporcionales a los respectivos anchos de los pulsos.

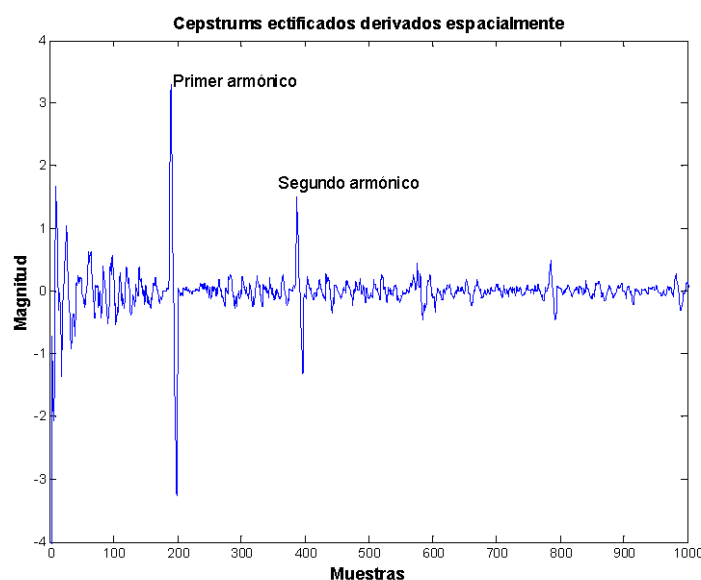


Figura 7-5: Localización del primer y segundo armónico en el dominio de los cepstrum derivados

Al igual que sucedía con el espectro, es significativo los cambios que presentan tramas adyacentes de tres periodos de *pitch* de duración en una representación cepstral, principalmente en el caso de las voces de baja calidad.

Las medidas de calidad que cuantifican la riqueza espectral de la voz propuestas son el valor de pico de primer armónico, el valor de pico del segundo armónico y la diferencia de amplitudes entre el primer y el segundo armónico, en una representación en el dominio de los cepstrum derivados espacialmente. Como medidas de calidad que cuantifican la riqueza espectral de la voz, resulta igual de interesante los valores que toman las magnitudes descritas, como la variación a lo largo del tiempo de los valores de dichas magnitudes.

7.4 Nomenclatura

En los siguientes apartados de este capítulo se presentan los resultados del estudio de la relevancia de las diferentes medidas de calidad que cuantifican la riqueza espectral de la voz. Para mayor comodidad, a cada una de las medidas se le ha asignado un identificador. En la Tabla 7-1 se presenta los identificadores de las diferentes medidas.

TABLA 7-1: MEDIDAS DE CALIDAD DE LA VOZ PERTENECIENTES AL GRUPO II (RIQUEZA ESPECTRAL)

Identificador de la medida	Grupo II : “Riqueza espectral”
1	Jitter versión 1
2	Jitter versión 2
3	Jitter versión 3 (PPQ)
4	Jitter versión 4 (Tremor)
5	Jitter versión 5 (HFPV)
6	Jitter versión 6 (Fftr)
7	media Relación Espectral Subarmónico-armónico
8	variación Relación Espectral Subarmónico-armónico
9	DUV
10	IPTS
11	Pico Cepstral Principal
12	variación Pico Cepstral Principal
13	Energía Cepstral de Pitch (PCEM)
14	variación Cepstral de Pitch (PCEM)
15	media Energía Cepstral del segundo armónico (SHECM)
16	variación Energía Cepstral del segundo armónico (SHECM)
17	media Valor Cepstral del primer pico de los cepstrum derivados
18	variación Valor Cepstral del primer pico de los cepstrum derivados
19	media Valor Cepstral del segundo pico de los cepstrum derivados
20	variación Valor Cepstral del segundo pico de los cepstrum derivados
21	media Valor Cepstral de la diferencia entre el segundo y el primer pico de los cepstrum derivados
22	variación Valor Cepstral de la diferencia entre el segundo y el primer pico de los cepstrum derivados
23	media Calculo de la Relación Armónico- Subarmónico
24	variación Calculo de la Relación Armónico- Subarmónico
25	media Open quotient
26	variación Open quotient
27	media Speech quotient
28	variación Speech quotient
29	media Closing quotient
30	variación Closing quotient
31	media Amplitud máxima de dUg
32	variación Amplitud máxima de dUg
33	media Amplitud mínima de dUg
34	variación Amplitud mínima de dUg
35	media Margen dinámico de dUg
36	variación Margen dinámico de dUg
37	media Momento de pico normalizado
38	variación Momento de pico normalizado

39	media Momento de excitación normalizado
40	variación Momento de excitación normalizado
41	media momento de cierre normalizado
42	variación momento de cierre normalizado
43	media momento de apertura normalizado
44	variación momento de apertura normalizado
45	media PECM de la señal Ug
46	variación PECM de la señal Ug
47	media PECM del segundo armónico de la señal Ug
48	variación PECM del segundo armónico de la señal Ug
49	media Skewness del la señal residual
50	media Grado de Sonoridad por Tramas (GST)
51	variación Grado de Sonoridad por Tramas (GST)
52	media Relación de la parte no periódica a la periódica (APR)
53	variación Relación de la parte no periódica a la periódica (APR)

7.5 Valoración de la relevancia de las Medidas de Calidad de la Voz

En este apartado se presentan los resultados del proceso consistente en ordenar las diferentes medidas de calidad de voz del grupo de medidas que cuantifican la riqueza espectral en función de su relevancia. El estudio se ha realizado diferenciándose dos casos la clasificación entre “voz sana – voz patológica” y la clasificación entre “voz sana – disfonía leve – disfonía moderada – disfonía severa”. Para cada caso, se ha estudiado las medidas de calidad de la voz diferenciando la vocal fruto de la fonación.

El procedimiento de obtención de la relevancia de cada medida ha sido descrito en el apartado 5.2. Para la estimación media de la relevancia de las diferentes medidas de la calidad ha sido evaluado 100 veces y posteriormente los resultados han sido promediados.

7.5.1 Estudio: Sano – Patológico

En este caso las muestras de voz, consistentes en vocales, han sido etiquetadas o como “voz sana” o como “voz patológica”.

TABLA 7-2: ORDEN EN FUNCIÓN DE LA RELEVANCIA PARA EL ESTUDIO
SANO – PATOLOGÍA

Orden de relevancia	GRUPO II Vocal 'A'		GRUPO II Vocal 'E'		GRUPO II Vocal 'I'		GRUPO II Vocal 'O'		GRUPO II Vocal 'U'	
	Id. Medida	Prob.	Id. Medida	Prob.	Id. Medida	Prob.	Id. Medida	Prob.	Id. Medida	Prob.
1	19	63%	19	70%	7	43%	49	68%	2	54%
2	17	53%	49	55%	23	41%	7	50%	15	39%
3	49	46%	17	40%	16	36%	23	42%	40	36%
4	25	40%	2	33%	22	34%	16	37%	13	34%
5	11	39%	7	32%	19	34%	5	34%	11	34%
6	2	39%	4	30%	51	32%	4	33%	5	34%
7	52	29%	23	27%	32	31%	2	32%	7	32%
8	50	29%	21	26%	13	30%	17	31%	49	31%
9	39	29%	11	25%	18	29%	32	29%	8	30%
10	7	25%	5	24%	3	29%	28	29%	27	29%
11	28	24%	13	22%	53	27%	51	28%	20	29%
12	5	24%	50	21%	12	27%	24	28%	44	28%
13	29	23%	32	21%	14	26%	53	27%	14	28%
14	26	23%	43	20%	8	26%	22	25%	12	28%
15	21	22%	44	17%	52	25%	14	25%	28	26%
16	43	21%	34	16%	4	25%	12	25%	24	26%
17	27	20%	12	16%	2	25%	8	25%	23	26%
18	45	19%	30	15%	24	24%	34	24%	18	26%
19	32	19%	22	15%	9	24%	21	23%	16	26%
20	4	18%	36	13%	21	23%	46	22%	47	25%
21	41	17%	31	13%	33	22%	36	21%	45	25%
22	34	17%	20	13%	5	22%	26	21%	50	24%
23	31	17%	15	13%	36	21%	40	20%	32	24%
24	35	16%	47	12%	1	21%	30	19%	19	24%
25	22	16%	38	12%	30	20%	20	19%	30	23%
26	47	15%	14	12%	26	20%	19	19%	51	22%
27	33	15%	28	11%	46	19%	18	18%	48	21%
28	23	15%	18	11%	40	19%	43	17%	46	21%
29	12	15%	16	11%	34	19%	31	16%	38	21%
30	36	14%	10	11%	31	19%	15	16%	36	21%
31	13	14%	9	11%	28	18%	11	16%	4	21%
32	30	13%	8	11%	20	18%	9	16%	53	20%
33	24	13%	3	11%	15	18%	48	15%	42	20%
34	40	12%	52	10%	11	18%	35	15%	26	20%
35	20	12%	42	10%	35	17%	33	15%	35	19%
36	48	11%	40	10%	48	16%	3	15%	22	19%
37	16	11%	35	10%	44	16%	42	14%	34	18%
38	1	11%	27	10%	10	16%	10	14%	31	18%
39	18	10%	25	10%	45	14%	1	14%	9	18%
40	14	10%	53	9%	42	14%	44	13%	3	18%
41	46	9%	46	9%	17	14%	13	12%	52	17%
42	3	9%	1	9%	50	13%	27	11%	33	17%
43	44	7%	45	8%	38	12%	29	10%	37	16%

44	9	7%	26	8%	37	12%	52	9%	43	15%
45	53	6%	24	8%	47	11%	25	9%	29	13%
46	42	6%	48	7%	27	10%	47	8%	17	13%
47	37	6%	33	7%	49	9%	50	7%	1	12%
48	8	6%	41	6%	43	9%	37	7%	39	10%
49	15	5%	39	6%	39	7%	45	6%	25	10%
50	51	4%	51	5%	29	4%	38	6%	10	10%
51	38	2%	37	5%	41	1%	41	4%	41	9%
52	10	1%	29	5%	25	1%	39	4%	21	8%
53	6	0%	6	0%	6	0%	6	0%	6	0%

7.5.2 Estudio: Sano – Disfonía Leve – Disfonía Moderada – Disfonía Severa

En este caso las muestras de voz, al igual que en el apartado anterior consisten en vocales y han sido etiquetadas con una de las siguientes etiquetas: “voz sana”, “disfonía leve”, “disfonía moderada” o “disfonía severa”. Este etiquetado ha sido realizado atendiendo al grado de disfonía.

TABLA 7-3: ORDEN EN FUNCIÓN DE LA RELEVANCIA PARA EL ESTUDIO
SANO – DISFONÍA LEVE – DISFONÍA MODERADA – DISFONÍA SEVERA

Orden de relevancia	GRUPO II Vocal ‘A’		GRUPO II Vocal ‘E’		GRUPO II Vocal ‘I’		GRUPO II Vocal ‘O’		GRUPO II Vocal ‘U’	
	Id. Medida	Prob.	Id. Medida	Prob.	Id. Medida	Prob.	Id. Medida	Prob.	Id. Medida	Prob.
1	19	66%	49	64%	7	64%	49	76%	2	68%
2	17	53%	17	50%	23	46%	23	53%	12	42%
3	7	43%	2	47%	9	45%	7	51%	40	40%
4	2	40%	19	44%	5	43%	11	42%	7	40%
5	23	39%	10	44%	12	41%	17	41%	32	38%
6	32	29%	21	43%	18	39%	2	41%	28	37%
7	25	29%	7	43%	22	38%	16	34%	14	36%
8	49	27%	4	40%	8	36%	8	34%	31	35%
9	3	27%	5	39%	16	35%	46	32%	30	35%
10	13	26%	52	37%	13	35%	43	31%	23	35%
11	9	26%	23	36%	3	35%	35	31%	5	35%
12	4	25%	50	34%	32	34%	21	31%	35	34%
13	45	24%	34	32%	4	34%	12	31%	8	33%
14	21	24%	12	29%	34	33%	9	31%	13	32%
15	28	23%	36	28%	14	32%	32	30%	18	31%
16	11	23%	25	28%	51	31%	4	30%	15	31%
17	50	22%	43	24%	46	31%	40	29%	9	31%
18	39	22%	42	23%	52	30%	14	29%	19	30%
19	12	22%	40	23%	20	30%	30	28%	4	30%
20	52	21%	16	23%	1	30%	28	28%	42	29%

21	34	20%	9	23%	44	29%	50	27%	24	29%
22	22	20%	44	22%	48	28%	44	27%	27	28%
23	18	20%	30	22%	31	28%	20	27%	51	27%
24	33	19%	8	22%	30	28%	15	27%	44	27%
25	5	19%	48	21%	26	28%	5	27%	33	27%
26	10	18%	26	21%	53	27%	26	26%	20	27%
27	31	17%	24	21%	24	27%	19	26%	53	26%
28	24	17%	22	21%	15	27%	24	25%	50	25%
29	30	16%	32	20%	36	26%	48	24%	46	25%
30	47	15%	20	20%	28	24%	18	24%	36	25%
31	44	15%	11	20%	33	22%	36	23%	22	25%
32	16	15%	38	19%	17	22%	34	23%	26	24%
33	8	15%	28	19%	19	21%	53	22%	3	24%
34	43	14%	18	19%	40	20%	31	22%	49	23%
35	40	14%	13	19%	35	20%	51	20%	48	22%
36	26	14%	46	18%	29	18%	25	20%	45	21%
37	53	13%	39	16%	50	17%	10	19%	16	21%
38	46	13%	35	16%	42	17%	45	18%	47	20%
39	36	13%	29	15%	10	17%	22	18%	34	20%
40	29	13%	53	14%	37	14%	42	17%	37	19%
41	14	13%	51	14%	21	14%	33	17%	10	19%
42	42	12%	33	14%	11	14%	13	17%	39	17%
43	15	12%	3	14%	45	13%	39	15%	11	15%
44	51	11%	1	14%	39	12%	29	15%	1	14%
45	48	11%	27	13%	2	11%	47	14%	29	13%
46	35	11%	14	13%	27	10%	38	14%	17	12%
47	20	11%	31	11%	47	9%	27	14%	38	8%
48	27	9%	47	9%	38	9%	3	14%	25	8%
49	1	6%	45	9%	49	7%	52	13%	21	7%
50	41	5%	15	7%	41	7%	1	13%	41	6%
51	38	3%	41	6%	25	6%	41	6%	52	4%
52	37	2%	37	5%	43	3%	37	4%	43	4%
53	6	0%	6	0%	6	0%	6	0%	6	4%

7.6 Evaluación de las Medidas de Calidad de la Voz

En este apartado se han evaluado las medidas de calidad que cuantifican la riqueza espectral de la voz de forma secuencial, incrementando el número de medidas, siguiendo el orden descendente marcado por la relevancia de dichas medidas y siendo evaluadas en diferentes clasificadores: redes neuronales probabilísticas (PNN), redes neuronales *feedforward* (NN) con una capa oculta, clasificadores basados en funciones de base radial, máquinas de vectores soporte (SVM) y clasificadores basados en los K vecinos más cercanos (KNN).

Para cada subconjunto de medidas (cuya dimensión está indicada en número de medidas) y cada vocal, se obtiene la tasa de éxito en la clasificación dado un determinado clasificador. Los resultados en cada caso son promediados 20 veces.

7.6.1 Estudio: Sano – Patológico

7.6.1.1 Red Neuronal Probabilística (PNN)

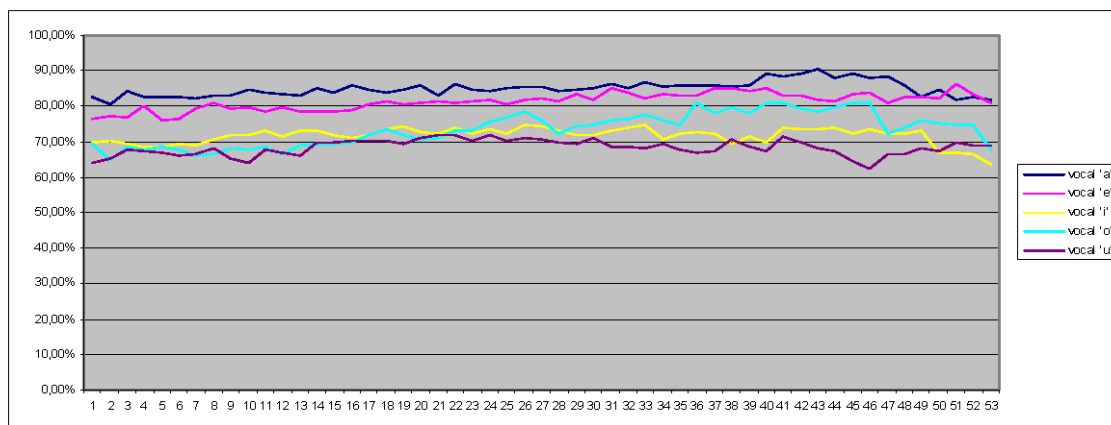


Gráfico 7-1: Evaluación secuencial de la relevancia para las medidas de calidad de la voz del grupo II usando el clasificador PNN en la clasificación entre voz sana y voz patológica

7.6.1.2 Red Neuronal *FeedForward* con 1 Capa Oculta (NN)

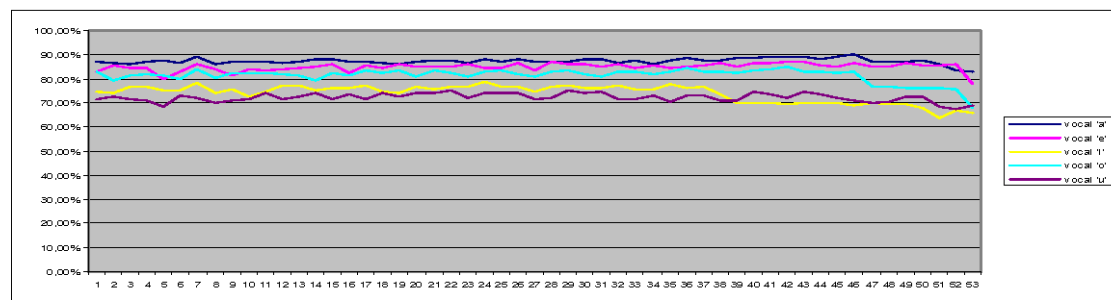


Gráfico 7-2: Evaluación secuencial de la relevancia para las medidas de calidad de la voz del grupo II usando el clasificador NN en la clasificación entre voz sana y voz patológica

7.6.1.3 Función de Base Radial (RBF)

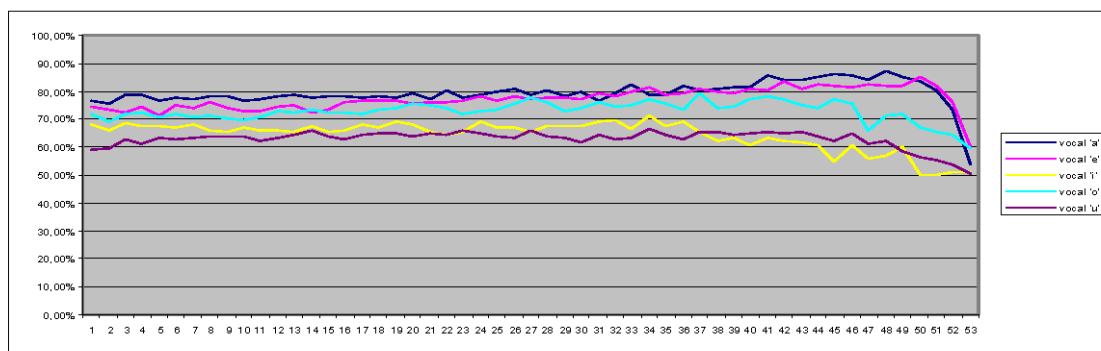


Gráfico 7-3 Evaluación secuencial de la relevancia para las medidas de calidad de la voz del grupo II usando el clasificador RBF en la clasificación entre voz sana y voz patológica

7.6.1.4 Máquinas de Vectores Soporte (SVM)

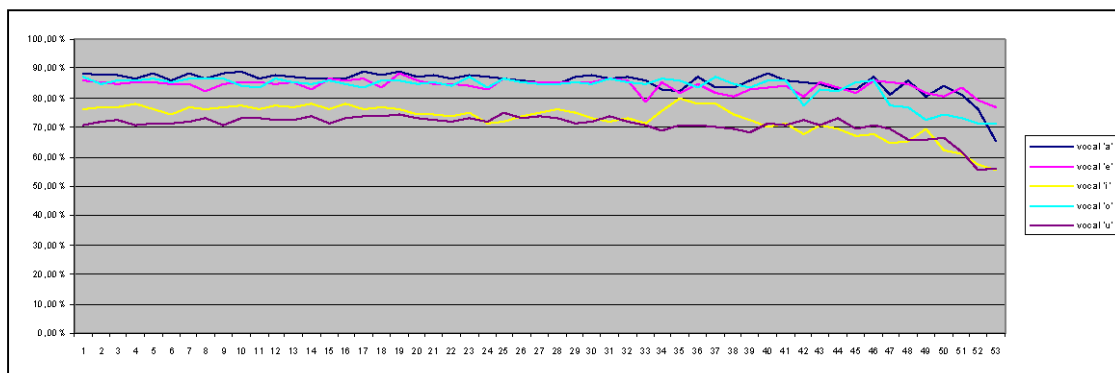


Gráfico 7-4 Evaluación secuencial de la relevancia para las medidas de calidad de la voz del grupo II usando el clasificador SVM en la clasificación entre voz sana y voz patológica

7.6.1.5 K vecinos más cercanos (KNN)

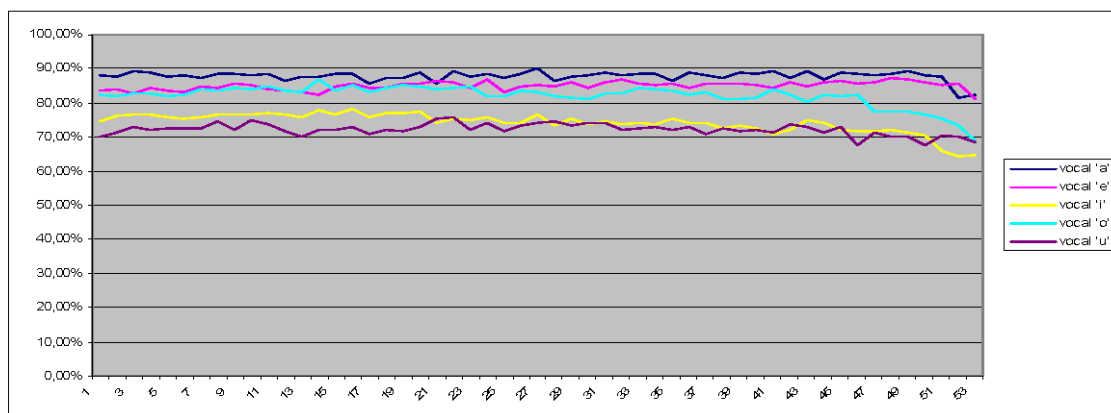


Gráfico 7-5: Evaluación secuencial de la relevancia para las medidas de calidad de la voz del grupo II usando el clasificador KNN en la clasificación entre voz sana y voz patológica

7.6.2 *Estudio: Sano – Disfonía Leve – Disfonía Moderada – Disfonía Severa*

7.6.2.1 Red Neuronal Probabilística (PNN)

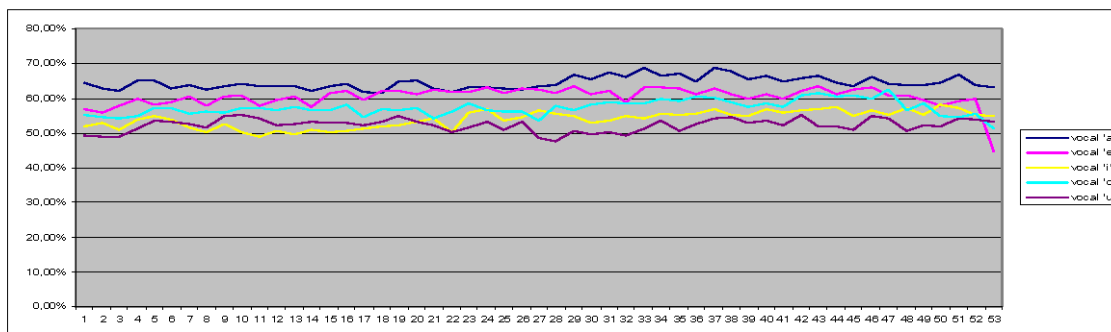


Gráfico 7-6: Evaluación secuencial de la relevancia para las medidas de calidad de la voz del grupo II usando el clasificador PNN en la clasificación entre distintos grados de disfonía

7.6.2.2 Red Neuronal FeedForward con 1 Capa Oculta (NN)

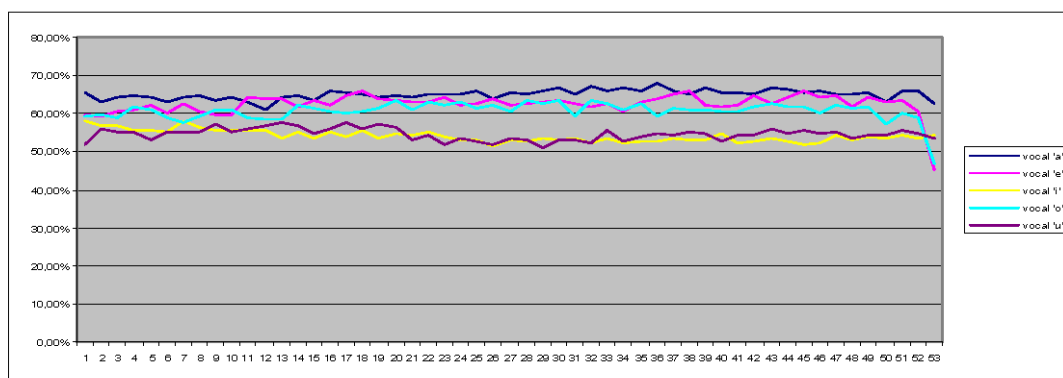


Gráfico 7-7: Evaluación secuencial de la relevancia para las medidas de calidad de la voz del grupo II usando el clasificador NN en la clasificación entre distintos grados de disfonía

7.6.2.3 Función de Base Radial (RBF)

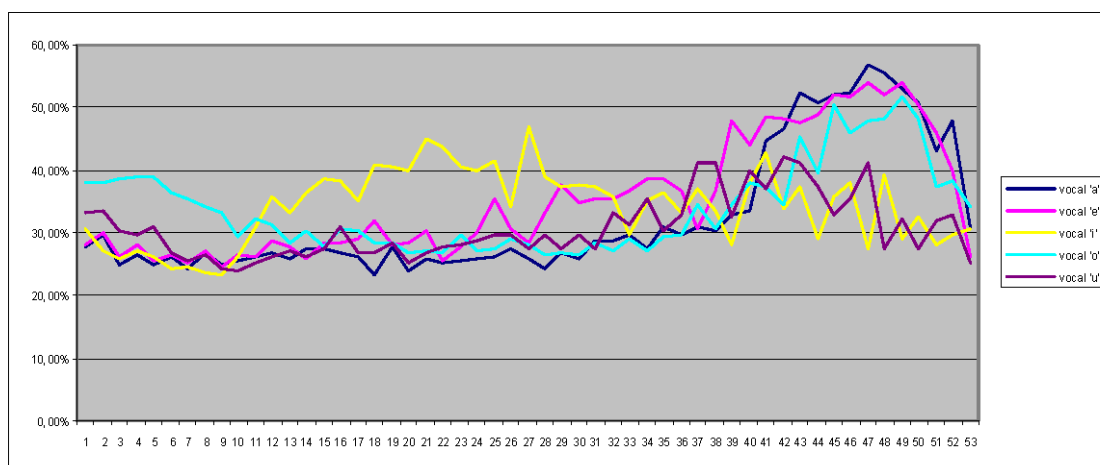


Gráfico 7-8: Evaluación secuencial de la relevancia para las medidas de calidad de la voz del grupo II usando el clasificador RBF en la clasificación entre distintos grados de disfonía

7.6.2.4 Máquinas de Vectores Soporte (SVM)

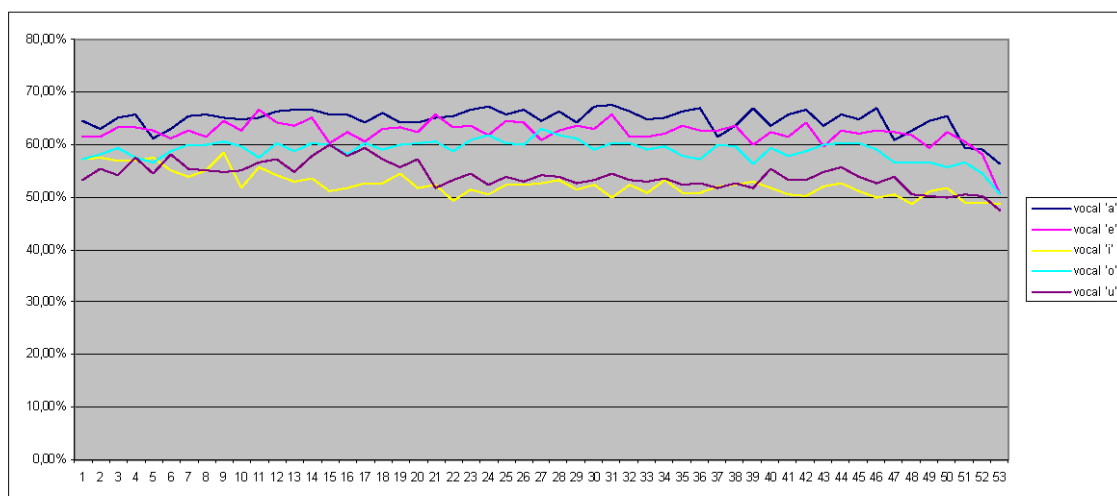


Gráfico 7-9: Evaluación secuencial de la relevancia para las medidas de calidad de la voz del grupo II usando el clasificador PNN en la clasificación entre distintos grados de disfonía

7.6.2.5 K vecinos más cercanos (KNN)

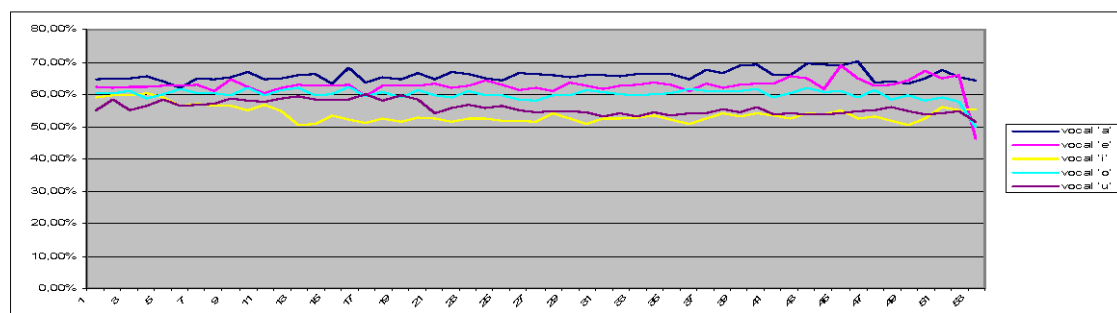


Gráfico 7-10: Evaluación secuencial de la relevancia para las medidas de calidad de la voz del grupo II usando el clasificador KNN en la clasificación entre distintos grados de disfonía

7.7 Evaluación

De forma muy similar a lo ocurrido con las medidas que cuantifican la estabilidad de la voz, en la clasificación entre voz sana y voz patológica se observa que para la PNN existe un ligero incremento de la tasa de éxito a medida que se incrementa el número de medidas de calidad, estando acotado el rango entre el 65 y 90%. Para el RBF existe una relativa estabilidad de la tasa de éxito, estando acotado el éxito entre el 60 y 85%. El resto de los clasificadores, el NN, SVM y KNN, se muestran con una tasa de éxito totalmente estable a medida que se incrementa el número de medidas de calidad, estando el rango de éxito acotado entre el 70 y el 90%.

En el caso de la clasificación entre voz sana, disfonía leve, disfonía moderada y disfonía severa, se observa que para la PNN existe un ligero incremento de la tasa de éxito, acotada entre el 50-65%. En la RBF, existe significativos incrementos que van desde el 25% hasta el 55%. Sin embargo, el NN, SVM y el KNN, mantienen una tasa de éxito totalmente constante, entre el 50-70%, pese al incremento de medidas de calidad de la voz.

Se observa de forma general para los diferentes clasificadores y los distintos casos estudiados (clasificación entre voz sana y voz patológica, y clasificación entre voz sana, disfonía leve, disfonía moderada y disfonía severa) que la tasa de éxito en la clasificación no se incrementa de forma significativa cuando a la medida de calidad que presenta la mayor relevancia se evalúa con otras medidas más, principalmente para NN, SVM y RBF. En la interpretación de los resultados es necesario tener en cuenta la reducida dimensión de la base de datos utilizada para la evaluación, lo cual puede justificar las fluctuaciones existentes.

Por este motivo, en esta tesis doctoral, consideramos que con una única medida es posible cuantificar el fenómeno físico “riqueza espectral”, y que para cada vocal, la medida que presenta mayor relevancia es la medida de calidad que mejor cuantifica este fenómeno.

Por lo tanto es posible asociar, para cada vocal por separado, la cuantificación del fenómeno físico “riqueza espectral” a una única medida, la cual se ha identificado como la medida que presenta mayor relevancia.

7.8 Referencias

- [1] Picone, J. W. , "Signal modeling techniques in speech recognition". *Proceedings of the IEEE* , Vol.81 , no. 9 , pp.1215–1247. 1993;
- [2] K. Marasek, "EGG and Voice Quality", <http://www.ims.uni-stuttgart.de/phonetik/EGG/frmst1.htm>, 1997. [Consulta 10/10/05].
- [3] Boyanov, B; Hadjitodorov, S. "Acoustic analysis of pathological voices. A voice analysis system for the screening of laryngeal diseases". *IEEE Engineering in Medicine and Biology Magazine*. Vol.16, no.4, pp.74 – 82, 1997.
- [4] Feijoo, S, et al. "Acoustic evaluation of glottal cancer based on short-term stability measures". *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, vol.2, pp.675-676, 1989;
- [5] Gabelman, B.; Alwan, A. "Analysis and synthesis of amplitude modulation components in pathological voices". *Proceedings 2002 IEEE Workshop on UCLA of Speech Synthesis*, pp. 51 – 54, 2002
- [6] Gavidia-Ceballos, L.; Hansen, J.H.L. "Direct Speech Feature Estimation using an Iterative EM Algorithm for Vocal Fold Pathology Detection". *IEEE Transactions on Biomedical Engineering*,. vol.43, no.4, pp.373 -383,1996.
- [7] Gavidia-Ceballos, L.; Hansen, J.H.L.; Kaiser, J.F. "Vocal fold pathology assessment using AM autocorrelation analysis of the Teager energy operator". *Proceedings of Fourth International Conference on Spoken Language, ICSLP 96*, vol.2, pp.757 – 760,1996.
- [8] Gavidia-Ceballos, L.; Hansen, J.H.L.; Riski, J.E, "A non-invasive speech processing method for analysis of vocal fold cancer based on an iterative EM algorithm". *Proceedings of the 16th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, vol.2, pp.1308 – 1309,1994
- [9] Godino-Llorente, J. I, et al. "On the selection of meaningful speech parameters used by a pathologic/non pathologic voice register classifier". *Sixth European Conference on Speech Communication and Technology, EUROSPEECH'99*, pp.563-566,1999.
- [10] Godino-Llorente, J.; Aguilera-Navarro, S.; Gómez-Vilda, P. "Automatic detection of voice impairments due to vocal misuse by means if gaussian mixture models", *Procedings of the 23rd Annual EMBS International conference*, October 25-28, Istanbul, Turkey. 2001; 1723 -1726.
- [11] Godino-Llorente, J.I.; Aguilera-Navarro, S.; Gomez-Vilda, P. "Non supervised neural net applied to the detection of voice impairment". *Proceedings. 2000 IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 6, pp.3594–3597, 2000
- [12] Hansen, J.H.L.; Gavidia-Ceballos, L.; Kaiser, J.F. "A nonlinear operator-based speech feature analysis method with application to vocal fold pathology assessment". *IEEE Transactions on Biomedical Engineering*, vol.45, no.3, pp.300-313, 1998
- [13] Hernández-Espinosa, C.; Fernández-Redondo, M.; Gómez-Vilda P. "Diagnosis of Vocal and Voice Disorders by the Speech Signal ". *IEEE-INNS-ENNS International Joint Conference on Neural Networks*, vol.4, pp.253-258. 2000
- [14] Jo, C.; Kim, K.; Kim, D.; Wang, S. "Screening of Pathological Voice from ARS using Neural Network". *International Workshop on MAVEBA*, pp.13-15, 2001.

- [15] Martínez César E.; Rufiner Hugo L. "Acoustic Analysis of Speech for Detection of Laryngeal Pathologies". *Proceedings of the Chicago 2000 World Congress IEEE EMBS*, pp.2369-2372, 2000.
- [16] Michaelis D.; Frohlich M.; Strube H. W. "Selection and combination of acoustic features for the description of pathologic voices". *Acoustical Society of America*. vol.103, no.3, pp.1628-1640, 1998.
- [17] O. Rosa, M.; Pereira, J.C.; Carvalho, A.C.P.L.F. "Evaluation of neural classifiers using statistic methods for identification of laryngeal pathologies". *Proceedings of Vth Brazilian Symposium on Neural Networks*, pp.220-225, 1998
- [18] Plante, F.; Kessler, H.; Cheetham, B.; Earis, J. "Speech monitoring of infective laryngitis". *Proceedings of Fourth International Conference on Spoken Language*., ICSLP 96, vol.2, pp.749-752, 1996
- [19] Sun, X. "Pitch determination and voice quality analysis using subharmonic-to-harmonic ratio". *Proceedings. of IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol1, pp.333-336, 2002.
- [20] Wallen, E.J.; Hansen, J.H.L."A screening test for speech pathology assessment using objective quality measures". *Proceedings of Fourth International Conference on Spoken Language*, ICSLP 96, vol .2, pp.776-779, 1996.
- [21] Alonso, J. B.; de León, J.; Alonso, I.; Ferrer, M. A. "Automatic Detection of pathologies in the voice by Hos based parameters". *Eurasip journal on Applied signal procesing*., vol.2001, no. 4, pp.275-284, 2001
- [22] Alonso, JB.; Carmona, C.; de Leon, J.; Ferrer, M.A. "Combining Neural Network and Hidden Markov models for automatic detection of pathologies". *16-th Biennial International EURASIP Conference Biosignal 2002*, pp.466-469, 2002
- [23] Alonso, J. B.; de León, J.; Ferrer, M. A. "Automatic Detection of Pathologies in the voice by HOS based parameters". *IEEE-EURASIP Workshop on Nonlinear Signal Processing* , NSIP 2001, June 3-6, 2001, Baltimore, Maryland USA.
- [24] Carmona, C., Alonso, J.B., de León, J., Ferrer, M.A. "Automatic detection of pathological voices using Multiband Excitation Vocoder parameters", *16-th Biennial International EURASIP Conference Biosignal 2002*, pp.120-123, 2002
- [25] Boyanov B., Doskov D., Mitev P., Hadjitodorov S., Teston B. , "New cepstral parameters for description of pathologic voice", *Comptes Rendus de L'Academie Bulgare des Sciences (Ann. of Bulgarian Academy of Sciences)*, vol.53, no.3, pp.41-44, 2000

CAPÍTULO 8

Cuantificación de la Presencia de Ruido

La **Presencia de Ruido** se define como la presencia de ruido glótico en la voz, durante la fonación de un sonido sonoro sostenido, debido a carencias en la fase de cierre de las cuerdas vocales. Este fenómeno físico se cuantifica por medio de la presencia de ruido no estacionario en la voz.

A continuación, y de forma similar a lo realizado para las medidas de calidad que cuantifican la estabilidad de la señal de voz y las que cuantifican su riqueza espectral, en este capítulo se presentan las diferentes medidas de calidad de la voz pertenecientes al grupo que cuantifican la presencia de ruido, tanto las presentes en la bibliografía actual como las propuestas en este trabajo, clasificadas en los diferentes dominios de representación. Además, se presenta una ordenación de dichas medidas en función de su relevancia. Por último, dicha ordenación de medidas es evaluada de forma incremental utilizando una batería de diferentes clasificadores.

8.1 Estado de la Técnica

En este apartado de la memoria, se presenta el conjunto de las principales medidas de calidad que cuantifican la presencia de ruido en la voz presentes en la bibliografía que han sido clasificadas en este trabajo, como medidas que cuantifican la presencia de ruido en la voz.

8.1.1 Dominio Temporal

8.1.1.1 TNI (Turbulence noise index)

El Índice de Turbulencia del Ruido [26] es una medida de calidad denominada (TNI) el cual se define como:

$$TNI = 100 * (1 - R_{\max}) \quad [\text{Ec. 8-1}]$$

donde $R_{\max}$ es el valor medio del máximo del factor de la correlación entre dos periodos consecutivos de la señal de voz.

8.1.2 Dominio Espectral

8.1.2.1 Relaciones entre las Energías de los Formantes

El conjunto de medidas de la calidad de la voz citadas en este apartado cuantifican las peculiaridades espectrales que presentan tanto las voces de alta calidad de voz como las voces de baja calidad. Estas características son medidas por medio de relaciones entre las energías concentradas alrededor de los diferentes formantes, principalmente son fruto de estudio las energías concentradas alrededor del primer y tercer formante. Es aconsejable para el cálculo de estas energías, no sólo calcular la energía que corresponde a la componente espectral del formante, sino, además, resulta conveniente utilizar una pequeña banda de frecuencias centrada en el primer formante o tercer formante, según corresponda.

$$F1 = \sum_{\omega=\omega_{F1}-\Delta\omega}^{\omega_{F1}+\Delta\omega} |H_{trama}(\omega)|^2 \quad [\text{Ec. 8-2}]$$

$$F3 = \sum_{\omega=\omega_{F3}-\Delta\omega}^{\omega_{F3}+\Delta\omega} |H_{trama}(\omega)|^2 \quad [\text{Ec. 8-3}]$$

ω_{F1} donde es la frecuencia del primer formante, ω_{F3} donde es la frecuencia del tercer formante y $\Delta\omega$ es un intervalo de frecuencias de aproximadamente 100 Hz.

Se ha estudiado como medida útil para cuantificar la presencia de ruido el cociente relacional entre la energía concentrada alrededor de F1 y la energía concentrada alrededor de F3 [5].

8.1.2.2 Relaciones entre Regiones de Energía

En diversos trabajos existen descripciones cualitativas del comportamiento del espectro en voces pertenecientes a locutores con ciertas patologías del sistema fonador [9].

Existen medidas de la calidad de la voz en las cuales se estiman relaciones de energía del espectro entre distintas regiones del espectro. En [2] se distingue entre dos regiones del espectro: rango de bajas frecuencias (70 - 1500 Hz) y el rango de altas frecuencias (1500 - 4000 Hz). Sin embargo, en [5] se realiza una distribución diferente del rango de frecuencias del espectro:

Región 0: desde 60 Hz hasta 400 Hz

Región 1: desde 400 Hz hasta 2 KHz

Región 2: desde 2 Khz hasta 5 KHz

Región 3: desde 5 Khz hasta 8 Khz

Las medidas descritas en este apartado se caracterizan por tratarse de medidas que relacionan niveles de energías de diferentes regiones. Las diferentes medidas presentan una formulación genérica de la siguiente forma:

$$L_i L_j = \frac{\sum_{\omega \in \text{región } i} |H_{trama}(\omega)|^2}{\sum_{\omega \in \text{región } j} |H_{trama}(\omega)|^2} \quad [\text{Ec. 8-4}]$$

donde $H_{trama}(\omega)$ es el espectro de la trama de voz que se analiza.

Las relaciones de energía más utilizadas para cuantificar la calidad de la voz son [5] : el cociente relacional entre la energía de la región 1 y la energía de la región 0 (L1L0) el cociente relacional entre la energía de la región 3 y la energía de la región 2 (L3L2).

En otros trabajos, se analiza la importancia de las relaciones entre la energía de cada región con respecto a la energía total del espectro. Este fenómeno se cuantifica según la siguiente expresión:

$$TE_{región_i} = \frac{\sum_{\omega=región\ i} |H_{trama}(\omega)|^2}{\sum_{\omega} |H_{trama}(\omega)|^2} \quad [\text{Ec.8-5}]$$

8.1.2.3 Modelado Tiempo-Frecuencia

La señal de voz es una variable en el tiempo. En el caso de realizar un estudio de la voz basado en un inventariado temporal de la señal, es posible de forma genérica realizar una descripción de la señal correspondiente a cada una de las diferentes tramas ($x(t)$) por medio de la siguiente expresión [17] :

$$x(t) = \sum_{n=0}^{\infty} a_n g_{\gamma_n}(t) \quad [\text{Ec.8-6}]$$

donde g_{γ_n} es una función genérica, la cual se define como combinación de escalado (s_n), translaciones (p_n) y modulaciones (f_n), tal y como se expresa en la ecuación:

$$g_{\gamma_n}(t) = \frac{1}{\sqrt{s_n}} g\left(\frac{t-p_n}{s_n}\right) \exp\{j(2\pi f_n t + \phi_n)\} \quad [\text{Ec.8-7}]$$

Para realizar la estimación de los parámetros p , s y f del modelo de señal de voz para cada una de las tramas se recurre a técnicas basadas en la descomposición adaptativa de señales basadas en el algoritmo MP [18] . En el algoritmo MP, se trata a la señal $x(t)$ como una señal no adaptativa la cual es descompuesta en una combinación lineal de funciones gaussianas $g_{\gamma_n}(t)$. Este algoritmo es iterativo. Las medidas útiles para cuantificar la calidad de la voz son los estadísticos que modelan el comportamiento

de los parámetros p , s y f del modelo de señal de voz. Un ejemplo de modelado tiempo-frecuencia es la transformada Wavelet.

8.1.2.4 FHN (Fundamental Harmonic Normalised Spectra)

Se propone una técnica novedosa para la representación del espectro. Se propone normalizar el espectro de las diferentes ventanas, tanto la magnitud del espectro como el eje de frecuencias [16]. En el caso de la magnitud del espectro la normalización se realiza respecto a su valor máximo y en el caso del eje de frecuencias la normalización se realiza respecto a la frecuencia del *pitch*.

Una vez realizada la transformación del espectro al espectro normalizado en función de la frecuencia de *pitch*, pueden aplicarse las diferentes medidas aplicables al espectro. Esta técnica presenta un inconveniente ya que en el caso de las voces de muy mala calidad, es posible la presencia de tramas sordas durante la fonación de un sonido sonoro sostenido y bajo esta circunstancia es imposible realizar la transformación debido a que no existe una frecuencia de *pitch*.

8.1.2.5 Distancia LPC

Al igual que en el caso anterior, esta medida cuantifica la variación a lo largo del tiempo del espectro de la señal de voz. Esta medida se basa en una técnica de estimación espectral paramétrica denominada predicción lineal, la cual modela la envolvente del espectro mediante un proceso autorregresivo (AR). Dada una señal $s(n)$, en nuestro caso la señal de voz, puede ser modelada de la siguiente manera [1]:

$$s(n) = \sum_{i=1}^{N_{LP}} a_{LP}(i)s(n-i) + e(n) \quad [\text{Ec.8-8}]$$

donde N_{LP} representa el número de coeficientes del modelo, o lo que es lo mismo, el orden del predictor. A $\{a_{LP}\}$ se les denomina los coeficientes de predicción lineal (LPC) y $e(n)$ representa el error en el modelo, es decir, la diferencia entre el valor predicho y el verdadero valor de la muestra. La secuencia de coeficientes $\{a_{LP}\}$ se calcula mediante algoritmos de mínimo error cuadrado.

La utilidad de los LPC radica en que se trata de un modelo paramétrico, con las ventajas en lo que a necesidades de almacenamiento se refiere, y además, presenta un espectro suavizado de la señal que representa.

La medida de calidad de la voz que se propone se estima por medio de una métrica entre los coeficientes LPC de una trama y los coeficientes LPC de la trama anterior. La métrica utilizada determina una medida de similitud la cual se define de la siguiente manera según [19] :

$$d(\vec{a}_t, \vec{a}_r) = \log\left(\frac{\vec{a}_t R_r \vec{a}_r^T}{\vec{a}_r R_r \vec{a}_r^T}\right) \quad [\text{Ec.8-9}]$$

donde $\vec{a}_t$ es el vector de coeficientes LPC de la trama que se analiza, $\vec{a}_r$ y R_r son el vector de coeficientes LPC y la matriz de covarianza de los coeficientes LPC de la trama anterior, respectivamente.

Otra medida similar es la definida por [19] :

$$d_e(\vec{a}_t, \vec{a}_r) = E(x_t) \log\left(\frac{\vec{a}_t R_r \vec{a}_r^T}{\vec{a}_r R_r \vec{a}_r^T}\right) \quad [\text{Ec.8-10}]$$

donde $E(x_t)$ es la energía de la trama que se analiza.

8.1.2.6 HNR (*Harmonic noise ratio*)

La relación Armónico Ruido (HNR) es una medida que se define como la relación entre la componente armónica y la componente de ruido del espectro de potencia de una trama de voz. Esta medida se ha utilizado en diversos estudios [2] [5] . Para realizar la separación entre las distintas componentes existen algunos métodos, tales como el método de d'Alessandro[22] , y una variante de este método propuesto en [21] .

En [23] se presenta diferentes medidas de HNR caracterizadas para una de las diferentes regiones espectrales:

HNR1: HNR para la región comprendida desde 60 hasta 400 Hz

HNR2: HNR para la región comprendida desde 400 hasta 2000 Hz

HNR3: HNR para la región comprendida desde 2000 hasta 5000 Hz

HNR4: HNR para la región comprendida desde 5000 hasta 8000 Hz

Además, resulta de interés las diferencias de niveles de HNR entre diferentes bandas de frecuencias: HNR4-HNR3, HNR3-HNR2, HNR2-HNR1.

Existen medidas similares a la HNR, como el caso de la medida denominada *degree of hoarseness* (DH) [2] o la medida denominada *voice turbulence index* (VTI) [6] [7].

8.1.2.7 NNE (Normalized Noise Energy)

Esta medida cuantifica la presencia de ruido glótico, el cual se comporta de forma no estacionaria. El parámetro NNE estima la relación entre la energía del ruido (componente de ruido) y la energía total de la señal (ambas medidas en dB) [3] [10] [11] [13], siendo descrita como:

$$NNE = 10 \log \left(\frac{\frac{1}{L} \sum_{\omega=1}^L |W(\omega)|^2}{\frac{1}{L} \sum_{\omega=1}^L |X(\omega)|^2} \right) \quad [\text{Ec. 8-11}]$$

donde el numerador es una estimación de la energía del ruido ($W(\omega)$) y el denominador es la energía de la señal de voz ($X(\omega)$). La energía del ruido es obtenida directamente del espectro por medio de técnicas iterativas. Los pasos en los que se dividen estas técnicas iterativas son los siguientes:

1. Para cada trama de voz se calcula su FFT de N muestras.
2. Las componentes espectrales armónicas del espectro de la trama son sustituidas por un valor nulo. Las componentes espectrales son localizadas por medio de un detector de *pitch*.
3. Se calcula la IFFT y se eliminan las muestras cuyo índice sea mayor de N.
4. Se vuelve a calcular la FFT y se sustituye las componentes espectrales del ruido por las mismas componentes espectrales de la trama original.
5. Los pasos 3 y 4 se repiten varias veces.

Existe una variante de la medida NNE en la que se realiza una estimación del ruido de forma adaptativa en función de la variación a corto plazo de la frecuencia de *pitch* [24] .

8.1.3 Dominio Cepstral

8.1.3.1 CHNR (*Cepstral Harmonic Noise Ratio*)

Esta medida cuantifica la relación armónico ruido en el dominio de los cepstrum de potencia. Se estima [4] [5] [12] [13] como la relación entre la energía total respecto a la energía del ruido, ambas medidas expresadas en dB. La técnica para el cálculo de CHNR difiere de la utilizada para NNE. La CHNR elimina el pico cepstral del *pitch* y los picos de todos sus armónicos con la finalidad de estimar la energía cepstral del ruido.

$$CHNR = 10 \log \left(\frac{\frac{1}{L} \sum_{n=1}^L |C_W(n)|}{\frac{1}{L} \sum_{n=1}^L |C_X(n)|} \right) \quad [\text{Ec.8-12}]$$

donde el numerador es la estimación de la energía cepstral del ruido y el denominador es la energía cepstral de la señal.

8.1.4 Dominio del modelo inverso

8.1.4.1 GNE (*Glotal-to-Noise Excitation Ratio*)

Esta medida caracteriza la cantidad de excitación de voz producida por las cuerdas vocales respecto a la cantidad de excitación de voz producida por turbulencias de ruido. Existen diferentes trabajos referentes a esta medida, encontrándose una amplia descripción en [13] , donde se proponen los siguientes pasos para la estimación de la medida GNE a partir de una grabación de voz:

1. Se realiza un filtrado inverso del tracto vocal mediante un modelo LPC.
2. Se aplican tres filtros pasobanda centrados en 500Hz, 1500Hz y 2500Hz, con un ancho de banda de 1000Hz.
3. A la salida de cada filtro se le calcula la envolvente de Hilbert de la señal.

4. El valor de GNE es el valor máximo de la correlación cruzada de las distintas combinaciones de las envolventes de Hilbert.

En la Figura 8-1 se muestra un esquema del método de cálculo de GNE.

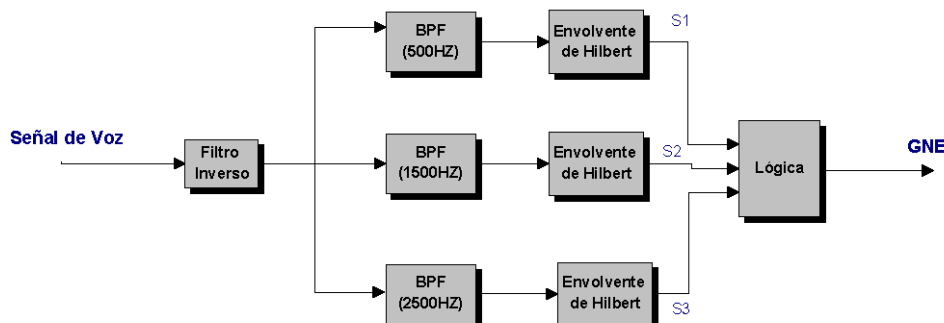


Figura 8-1: Esquema de bloques del cálculo del parámetro GNE

El bloque “lógica” realiza la siguiente operación matemática:

$$GNE = \max([xcorr(S1, S2), xcorr(S1, S3), xcorr(S2, S3)]) \quad [Ec.8-13]$$

Tanto los parámetros NNE como CHNR presentan el inconveniente de depender fuertemente de la estimación del *pitch* para su cálculo. Este inconveniente no lo presenta GNE.

8.1.4.2 NEP (Normalised Error Prediction)

NEP [15] es otra medida que permite estimar el nivel de ruido calculando la energía normalizada de señal residual obtenida de una predicción lineal. El error residual obtenido de la predicción lineal puede ser expresada por medio del cociente relacional entre la media geométrica del modelo del espectro y su media aritmética. Esta es una medida de dispersión espectral. Las voces de alta calidad presentan un espectro con picos y valles muy pronunciados y por lo tanto los valores que toma el espectro para las diferentes componentes espectrales presentan una gran dispersión. Por otro lado, el espectro de las voces de baja calidad no presenta significativas componentes armónicas y siendo la dispersión a lo largo del espectro menor, principalmente en altas frecuencias.

8.1.4.3 Rugosidad espectral de la señal residual

Se ha propuesto una medida de la rugosidad del espectro de la señal residual (SFR) [14] basandose dicha medida en que una señal residual tendente a un espectro plano indicaría una gran presencia de ruido en la excitación glótica. Esta medida se estima como la relación entre la media geométrica y la media aritmética del espectro de la señal residual.

Otra medida propuesta es una medida de lo plano que resulta el espectro del modelo del tracto vocal estimado para obtener la señal residual (SFF), el cual se define como la diferencia entre el cociente entre la media geométrica y la media aritmética del espectro de la señal de voz [25] .

Un espectro plano del modelo del tracto vocal implicaría una gran componente de ruido en la señal de voz. Además, se propone una medida de relación señal a ruido (SNR) como el valor cuadrático medio de la señal residual.

8.1.4.4 HNRR (*Harmonic to Noise Ratio from Speech*) y HNRS(*Harmonic to Noise Ratio from Residual*)

Se ha propuesto dos medidas que relacionan la componente armónica y la componente de ruido de una señal de voz [7] [8] . Para el cálculo de la medida HNRS se estima la señal de voz en el dominio de los *cepstrum* de potencia y posteriormente se realiza una separación de la componente armónica y la componente de ruido. Para medida HNRR se realiza en primer lugar un filtrado inverso, y posteriormente en el dominio de los *cepstrum* de potencia, se realiza la separación de la componente armónica y la componente de ruido. Ambas medidas se estiman como el cociente relacional entre la componente armónica y la componente de ruido. Para la separación de las componentes se aplica el método de d'Alessandro[22] .

8.2 Análisis

Las medidas de calidad que cuantifican la presencia de ruido en la voz obtenidas en el dominio temporal, presentan una muy alta sensibilidad a los niveles de ruido, quizás sea éste el motivo por el cual existen tan pocas medidas que cuantifiquen la presencia de ruido que se obtenga en el dominio temporal.

La presencia de ruido en la señal de voz, sea estacionario o no, implica gran dificultad para estimar determinadas medidas como, por citar alguna de ellas, las medidas de calidad que cuantifican la presencia de ruido que se estima por medio de la cuantificación de la energía alrededor del primer y tercer formante, ya que tienen el problema de que resulta muy difícil identificar con cierta garantía la región del espectro correspondiente al primer y tercer formante. Además, presenta la dificultad de estimar el *pitch*, y en casos de voces de muy baja calidad, se da el hecho de la inexistencia de una vibración de las cuerdas vocales. Este hecho tiene una importancia significativa, debido a que muchas medidas de calidad que cuantifican la presencia de ruido se basan en la información de *pitch*, por citar algún ejemplo, para separar la componente armónica de la componente de ruido.

Un aspecto interesante en las medidas de calidad que cuantifican niveles de energía o relaciones de energía, es que de forma general, lo importante no es los valores de energías que toman en un momento dado; lo importante es la variación de esos niveles de energía o relaciones de energía a lo largo del tiempo. De igual forma ocurre con las medidas de calidad de la voz que cuantifican las relaciones de señal a ruido, donde el interés está en la variación de los valores de las medidas a lo largo del tiempo.

8.3 Aportaciones

En este trabajo se han realizado diferentes aportaciones en lo referente a medidas de calidad que cuantifican la presencia de ruido, además de la identificación de las medidas que cuantifican este fenómeno físico. Las aportaciones realizadas se han hecho a partir de cuantificaciones de fenómenos manifestados claramente en el dominio espectral.

8.3.1 Dominio Espectral

8.3.1.1 Relaciones entre Regiones de Energía

En el apartado 8.1.2.2 se han presentado diferentes medidas de calidad de la voz basadas en relaciones de energía espectral entre las distintas regiones en las que previamente se ha dividido el espectro (*Región 0*: desde 60 Hz hasta 400 Hz, *Región 1*: desde 400 Hz hasta 2 KHz, *Región 2*: desde 2 KHz hasta 5 KHz y *Región 3*: desde 5 KHz hasta 8 KHz)

En el trabajo [20] se añade un nuevo rango de frecuencia debido a que se ha aumentado, respecto a los trabajos anteriores, el rango de frecuencias utilizado en el estudio:

Región 4: desde 8 KHz hasta 11025 KHz

Las medidas descritas en el apartado 8.1.2.2 se caracterizan por tratarse de medidas que relacionan los niveles de energías entre diferentes regiones. En el trabajo [20] se presenta el cociente relacional entre la energía de la región 4 y la energía de la región 1 (L4L1).

De igual forma, en el apartado 8.1.2.2 se analiza la importancia de las relaciones entre la energía de cada región con respecto a la energía total del espectro. En [20] , se presenta como medidas útiles de la calidad de la voz la perturbación a lo largo del tiempo del cociente relacional de energías entre la región 1 y la energía total, y también la perturbación a lo largo del tiempo del cociente relacional de energías entre la región 2 y la energía total.

8.3.1.2 Medidas de Distancia Espectral

Estas medidas miden la variación a lo largo del tiempo del espectro de la señal de voz, durante la fonación de un sonido sonoro sostenido, manifestándose especialmente significativa para los locutores con baja calidad de voz durante la fonación de un sonido sostenido.

Las medidas propuestas en este apartado cuantifican la regularidad de la secuencia de los respectivos espectros correspondientes a las diferentes tramas en las que se divide la señal de voz. Las medidas que se realizan son de dos tipos diferentes: la variación a lo largo del tiempo del espectro y la variación a lo largo del tiempo de la diferencia de fase, respecto a un modelo.

La medida de la variación del espectro se realiza por medio de la estimación de una distancia espectral, para cada trama, entre el espectro de la trama y el espectro de la trama anterior. Esta medida ha sido descrita en [20] por medio de la siguiente expresión:

$$D_{\text{módulo}} = \frac{1}{L} \sum_{\omega=0}^L \|T(\omega) - T_{\text{ref}}(\omega)\| \quad [\text{Ec.8-14}]$$

donde $D_{\text{módulo}}$ es el valor de la medida de la distancia de los módulos, $|T(w)|$ es el módulo de la transformada de Fourier de la trama de voz que se analiza y que se calcula por medio de la transformada rápida de Fourier (FFT), $|T_{\text{ref}}(w)|$ es el módulo de la transformada de Fourier del modelo de la trama anterior y L es la longitud de la FFT.

La otra medida citada, también descrita en [20], es una medida de distancia de fases la cual se calcula por medio de la siguiente expresión:

$$D_{\text{fase}} = \frac{1}{L} \sum_{\omega=0}^L |\varphi_T(\omega) - \varphi_{\text{ref}}(\omega)| \quad [\text{Ec.8-15}]$$

donde D_{fase} es el valor de la medida de distancia de fase, $\varphi_T(\omega)$ es el vector de fase, obtenido a partir de la FFT, de la trama de voz y $\varphi_{\text{ref}}(\omega)$ es el vector de fase de la trama anterior.

8.4 Nomenclatura

En los siguientes apartados de este capítulo se presentan los resultados del estudio de la relevancia de las diferentes medidas de calidad que cuantifican la presencia de ruido en la voz. Por mayor comodidad, a cada una de las medidas se le ha asignado un identificador. En la Tabla 8-1 se presenta los identificadores de las diferentes medidas.

TABLA 8-1: MEDIDAS DE CALIDAD DE LA VOZ PERTENECIENTES AL GRUPO III (PRESENCIA DE RUIDO)

Identificador de la medida	GRUPO III: "Presencia de Ruido"
1	media energía concentrada alrededor del 1er formante
2	variación energía concentrada alrededor del 1er formante
3	media energía concentrada alrededor del 3er formante
4	variación energía concentrada alrededor del 3er formante
5	media Relación de energía entre el la región 3 y la región 2
6	variación Relación de energía entre el la región 3 y la región 2
7	media Relación de energía entre el la región 4 y la región 1
8	variación Relación de energía entre el la región 4 y la región 1
9	media Relación de energía entre el la región 1 y la energía total
10	variación Relación de energía entre el la región 1 y la energía total
11	media Relación de energía entre el la región 2 y la energía total
12	variación Relación de energía entre el la región 2 y la energía total
13	media Relación de energía entre el la región 3 y la energía total
14	variación Relación de energía entre el la región 3 y la energía total
15	media Relación de energía entre el la región 4 y la energía total
16	variación Relación de energía entre el la región 4 y la energía total
17	media Distancia de módulos entre las transformadas de Fourier (FFT) de dos tramas consecutivas
18	variación Distancia de módulos entre las transformadas de Fourier (FFT) de dos tramas consecutivas
19	media Distancia de fase entre las transformadas de Fourier (FFT) de dos tramas consecutivas
20	variación Distancia de fase entre las transformadas de Fourier (FFT) de dos tramas consecutivas
21	media Distancia basada en los LPC
22	variación Distancia basada en los LPC
23	media Distancia de máxima verosimilitud
24	variación Distancia de máxima verosimilitud
25	media energía de ruido normalizada;
26	variación energía de ruido normalizada;
27	media Relación armónico ruido cepstral
28	variación Relación armónico ruido cepstral
29	media Relación Espectral Armónico Ruido
30	variación Relación Espectral Armónico Ruido
31	media Cálculo del Glotal-to-Noise Excitation (GNE)
32	variación Cálculo del Glotal-to-Noise Excitation (GNE)
33	media Normalilised error prediction (NEP)
34	variación Normalilised error prediction (NEP)
35	media valor cuadrático medio de la señal residual
36	variación valor cuadrático medio de la señal residual
37	media Relación Espectral armónico ruido de la señal residual
38	variación Relación Espectral armónico ruido de la señal residual
39	media Relación Espectral Subarmónico - armónico de al señal residual
40	variación Relación Espectral Subarmónico - armónico de al señal residual
41	Relación armónica -ruido cepstral (HNRS)
42	variación Relación armónica -ruido cepstral (HNRS)

8.5 Valoración de la relevancia de las Medidas de Calidad de la Voz

En este apartado se presentan los resultados del proceso consistente en ordenar las diferentes medidas de calidad que cuantifican la presencia de ruido en la voz en función de su relevancia. El estudio se ha realizado diferenciándose dos casos: la clasificación entre “voz sana – voz patológica” y la clasificación entre “voz sana – disfonía leve – disfonía moderada – disfonía severa”. Para cada caso, se ha estudiado las medidas de calidad de la voz diferenciando la vocal fruto de la fonación.

El procedimiento de obtención de la relevancia de cada medida ha sido descrito en el apartado 5.2. Para la estimación media de la relevancia de las diferentes medidas de la calidad ha sido evaluado 100 veces y posteriormente los resultados han sido promediados.

8.5.1 Estudio: Sano – Patológico

En este caso las muestras de voz, consistentes en vocales, han sido etiquetadas o como “voz sana” o como “voz patológica”.

TABLA 8-2: ORDEN EN FUNCIÓN DE LA RELEVANCIA PARA EL ESTUDIO SANO – PATOLÓGICO

Orden de relevancia	GRUPO III Vocal ‘A’		GRUPO III Vocal ‘E’		GRUPO III Vocal ‘I’		GRUPO III Vocal ‘O’		GRUPO III Vocal ‘U’	
	Id. Medida	Prob.	Id. Medida	Prob.	Id. Medida	Prob.	Id. Medida	Prob.	Id. Medida	Medida
1	39	77%	3	70%	3	92%	39	93%	39	74%
2	41	57%	1	70%	1	57%	15	53%	19	56%
3	4	53%	39	69%	39	51%	13	47%	5	53%
4	1	53%	15	55%	11	49%	17	45%	6	50%
5	3	51%	5	54%	13	48%	4	39%	29	49%
6	11	49%	41	53%	7	46%	2	39%	10	45%
7	8	49%	11	52%	17	44%	16	37%	25	43%
8	37	45%	17	49%	18	42%	1	36%	11	43%
9	29	39%	37	45%	9	41%	12	32%	34	42%
10	25	39%	13	44%	37	39%	35	30%	13	42%
11	7	37%	10	44%	35	35%	19	30%	32	40%
12	23	36%	25	41%	19	34%	31	28%	28	40%
13	17	35%	7	39%	16	34%	11	28%	8	40%
14	10	35%	31	38%	10	34%	37	27%	37	39%
15	6	35%	22	38%	4	34%	8	27%	7	39%

16	26	32%	12	37%	33	33%	28	26%	26	38%
17	15	31%	19	34%	15	33%	14	26%	17	38%
18	2	31%	8	33%	8	32%	18	25%	21	37%
19	28	30%	4	33%	25	31%	9	25%	23	35%
20	20	29%	21	32%	41	29%	30	24%	9	35%
21	33	27%	16	31%	31	27%	29	24%	1	34%
22	31	27%	24	30%	29	27%	27	24%	12	33%
23	13	27%	18	30%	23	25%	25	24%	15	31%
24	19	26%	14	29%	12	24%	7	23%	14	31%
25	38	25%	33	28%	5	24%	38	22%	2	31%
26	36	25%	23	28%	42	22%	33	22%	18	30%
27	16	25%	9	28%	26	22%	26	22%	4	30%
28	12	25%	32	27%	20	20%	22	22%	33	29%
29	40	24%	2	23%	14	20%	6	22%	30	29%
30	35	24%	29	21%	40	18%	40	21%	16	27%
31	30	24%	30	20%	2	18%	3	21%	22	26%
32	18	24%	26	20%	32	16%	41	20%	42	25%
33	14	23%	38	19%	28	16%	36	20%	3	25%
34	22	21%	6	19%	22	16%	42	19%	40	23%
35	34	20%	42	18%	6	16%	10	18%	24	23%
36	42	19%	40	17%	36	15%	5	18%	36	22%
37	5	19%	34	15%	21	15%	23	17%	20	21%
38	24	18%	35	14%	38	14%	21	16%	38	20%
39	21	17%	28	14%	30	14%	34	15%	41	18%
40	9	15%	36	13%	34	13%	32	15%	27	18%
41	32	13%	20	13%	24	12%	24	15%	31	10%
42	27	9%	27	8%	27	3%	20	15%	35	9%

8.5.2 Estudio: Sano – Disfonía Leve – Disfonía Moderada – Disfonía Severa

En este caso las muestras de voz, al igual que en el apartado anterior, consisten en vocales, siendo etiquetadas con una de las siguientes etiquetas: “voz sana”, “disfonía leve”, “disfonía moderada” o “disfonía severa”. Este etiquetado ha sido realizado atendiendo al grado de disfonía.

TABLA 8-3: ORDEN EN FUNCIÓN DE LA RELEVANCIA PARA EL ESTUDIO
SANO – DISFONÍA LEVE – DISFONÍA MODERADA – DISFONÍA SEVERA

Orden de relevancia	GRUPO III Vocal 'A'		GRUPO III Vocal 'E'		GRUPO III Vocal 'I'		GRUPO III Vocal 'O'		GRUPO III Vocal 'U'	
	Id. Medida	Prob.	Id. Medida	Prob.	Id. Medida	Prob.	Id. Medida	Prob.	Id. Medida	Prob.
1	39	72%	1	70%	3	86%	39	79%	39	67%
2	1	72%	41	65%	39	73%	17	77%	17	58%
3	41	69%	3	63%	7	60%	19	54%	13	51%
4	3	58%	39	62%	9	56%	3	53%	8	47%
5	4	55%	7	52%	11	47%	15	49%	37	46%
6	11	51%	5	52%	15	46%	4	48%	29	46%
7	10	51%	13	51%	1	44%	11	47%	25	44%
8	7	51%	37	50%	17	43%	16	44%	15	44%
9	37	47%	10	49%	13	42%	2	44%	6	41%
10	2	46%	15	48%	41	41%	26	42%	5	41%
11	16	44%	25	44%	25	41%	14	40%	32	40%
12	15	44%	22	44%	33	38%	42	39%	41	39%
13	20	41%	8	44%	28	37%	28	39%	19	39%
14	17	41%	29	41%	19	35%	18	39%	4	38%
15	8	41%	9	41%	26	34%	12	39%	23	37%
16	9	40%	21	40%	29	33%	38	38%	1	37%
17	18	39%	14	39%	12	33%	37	38%	28	36%
18	13	39%	11	38%	37	30%	30	36%	9	35%
19	42	38%	31	36%	18	30%	13	36%	7	35%
20	28	38%	19	35%	8	30%	35	34%	30	34%
21	26	38%	16	35%	42	28%	40	32%	10	34%
22	6	38%	23	34%	30	27%	29	30%	21	33%
23	40	37%	17	34%	38	26%	22	30%	12	33%
24	25	37%	35	33%	40	25%	41	29%	26	31%
25	38	35%	4	33%	4	25%	24	29%	14	31%
26	34	35%	33	32%	20	24%	27	28%	3	31%
27	14	33%	12	29%	31	23%	25	28%	38	30%
28	35	30%	32	27%	10	21%	1	28%	11	30%
29	30	30%	40	25%	5	21%	31	27%	18	29%
30	31	28%	20	25%	21	19%	5	25%	40	28%
31	29	27%	36	23%	23	18%	9	24%	2	28%
32	33	26%	24	23%	34	14%	8	24%	42	26%
33	19	25%	30	22%	16	14%	33	23%	16	26%
34	36	24%	28	22%	35	12%	32	23%	33	25%
35	23	22%	6	22%	22	12%	21	23%	24	24%
36	12	22%	18	21%	2	12%	7	20%	36	22%
37	32	21%	34	20%	32	10%	6	19%	27	20%
38	5	21%	26	20%	24	10%	23	18%	22	20%
39	22	19%	42	18%	36	7%	36	16%	20	19%
40	21	13%	38	18%	14	7%	20	16%	34	16%
41	24	11%	2	15%	6	6%	10	16%	31	10%
42	27	8%	27	3%	27	5%	34	15%	35	5%

8.6 Evaluación de las Medidas de Calidad de la Voz

En este apartado se han evaluado las medidas de calidad que cuantifican la presencia de ruido en la voz de forma secuencial, incrementando el número de medidas, siguiendo el orden descendente marcado por la relevancia de dichas medidas y siendo evaluadas en diferentes clasificadores: redes neuronales probabilísticas (PNN), redes neuronales *feedforward* (NN) con una capa oculta, clasificadores basados en funciones de base radial, máquinas de vectores soporte (SVM) y clasificadores basados en los K vecinos más cercanos (KNN).

Para cada subconjunto de medidas (cuya dimensión está indicada en número de medidas) de cada grupo de medidas, y para cada vocal, se obtiene la tasa de éxito en la clasificación dado un determinado clasificador. Los resultados en cada caso son promediados 20 veces.

8.6.1 Estudio: Sano – Patológico

8.6.1.1 Red Neuronal Probabilística (PNN)

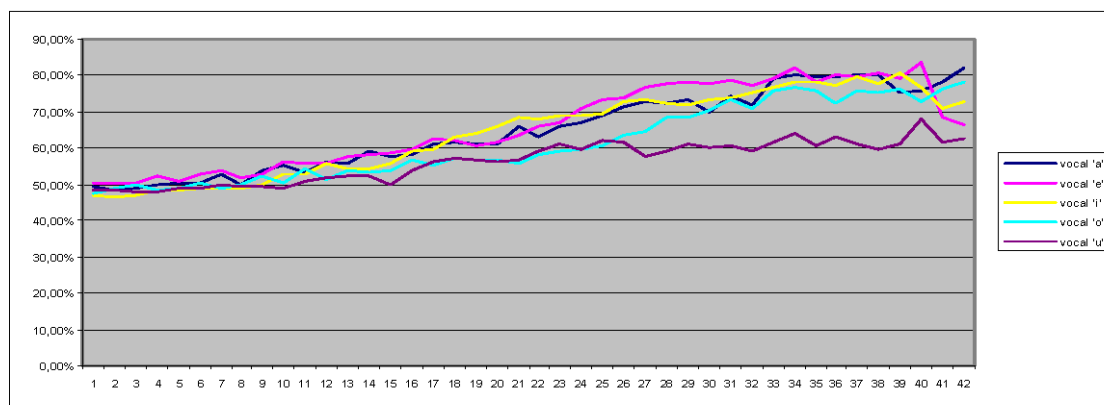


Gráfico 8-1: Evaluación secuencial de la relevancia para las medidas de calidad de la voz del grupo III usando el clasificador PNN en la clasificación entre voz sana y voz patológica

8.6.1.2 Red Neuronal *FeedForward* con 1 Capa Oculta (NN)

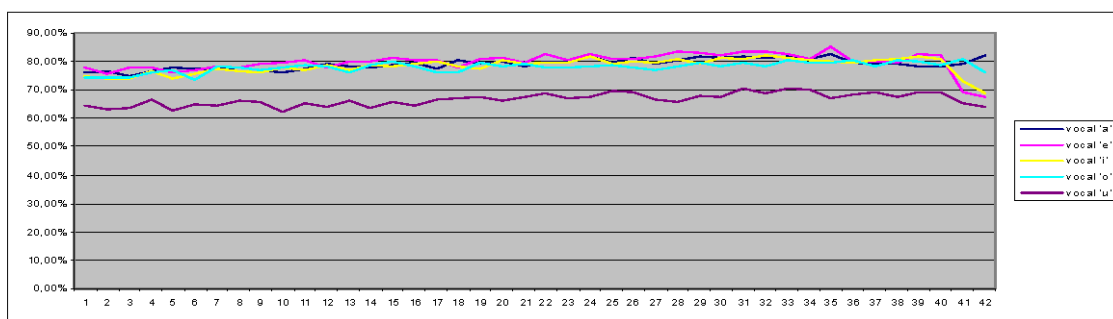


Gráfico 8-2: Evaluación secuencial de la relevancia para las medidas de calidad de la voz del grupo III usando el clasificador NN en la clasificación entre voz sana y voz patológica

8.6.1.3 Función de Base Radial (RBF)

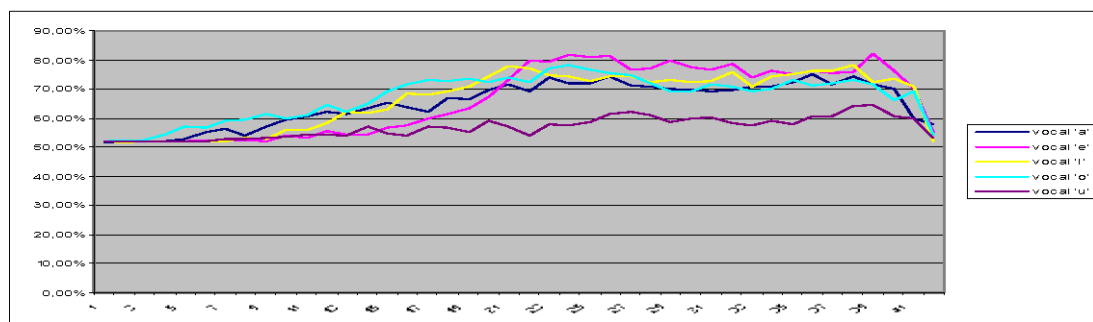


Gráfico 8-3 Evaluación secuencial de la relevancia para las medidas de calidad de la voz del grupo III usando el clasificador RBF en la clasificación entre voz sana y voz patológica

8.6.1.4 Máquinas de Vectores Soporte (SVM)

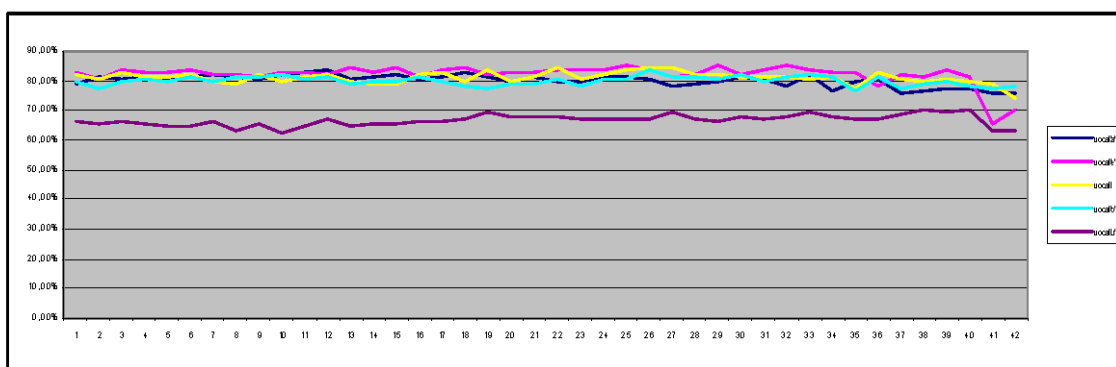


Gráfico 8-4 Evaluación secuencial de la relevancia para las medidas de calidad de la voz del grupo III usando el clasificador SVM en la clasificación entre voz sana y voz patológica

8.6.1.5 K vecinos más cercanos (KNN)

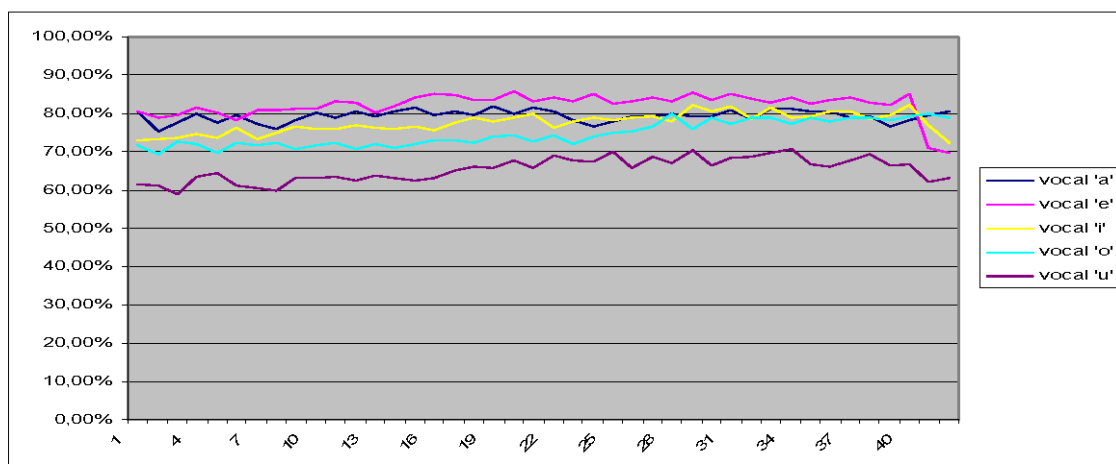


Gráfico 8-5: Evaluación secuencial de la relevancia para las medidas de calidad de la voz del grupo III usando el clasificador KNN en la clasificación entre voz sana y voz patológica

8.6.2 Estudio: Sano – Disfonía Leve – Disfonía Moderada – Disfonía Severa

8.6.2.1 Red Neuronal Probabilística (PNN)

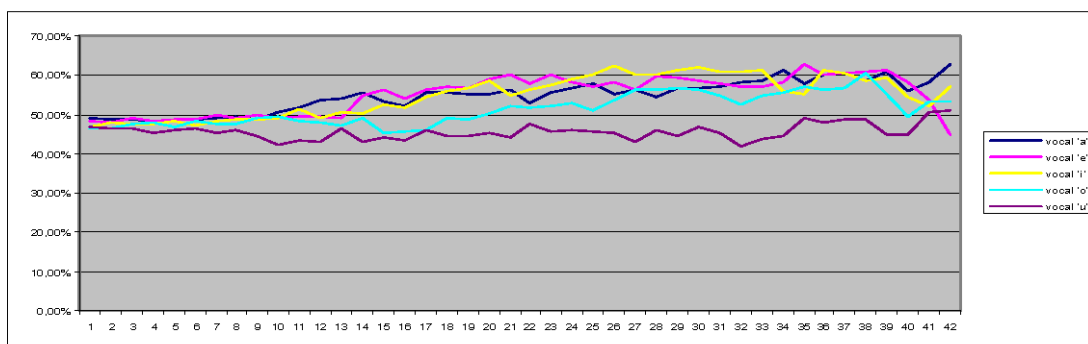


Gráfico 8-6: Evaluación secuencial de la relevancia para las medidas de calidad de la voz del grupo III usando el clasificador PNN en la clasificación entre distintos grados de disfonía

8.6.2.2 Red Neuronal FeedForward con 1 Capa Oculta (NN)

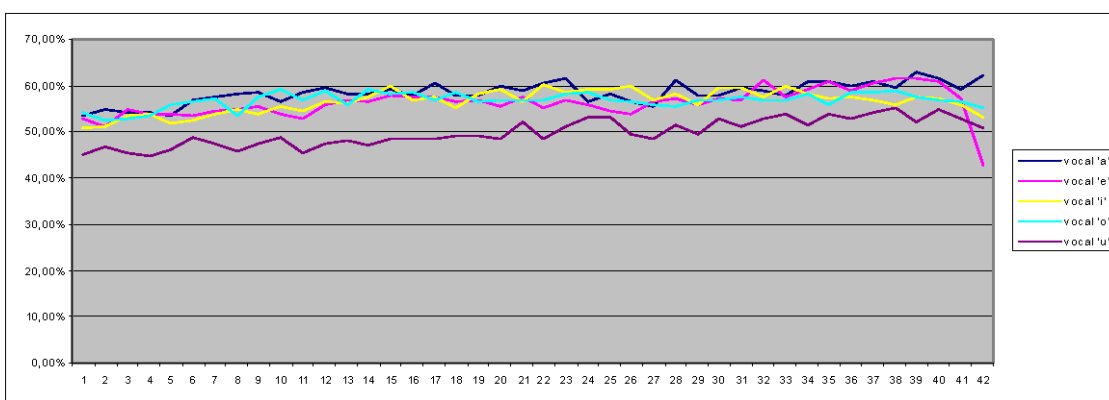


Gráfico 8-7: Evaluación secuencial de la relevancia para las medidas de calidad de la voz del grupo III usando el clasificador NN en la clasificación entre distintos grados de disfonía

8.6.2.3 Función de Base Radial (RBF)

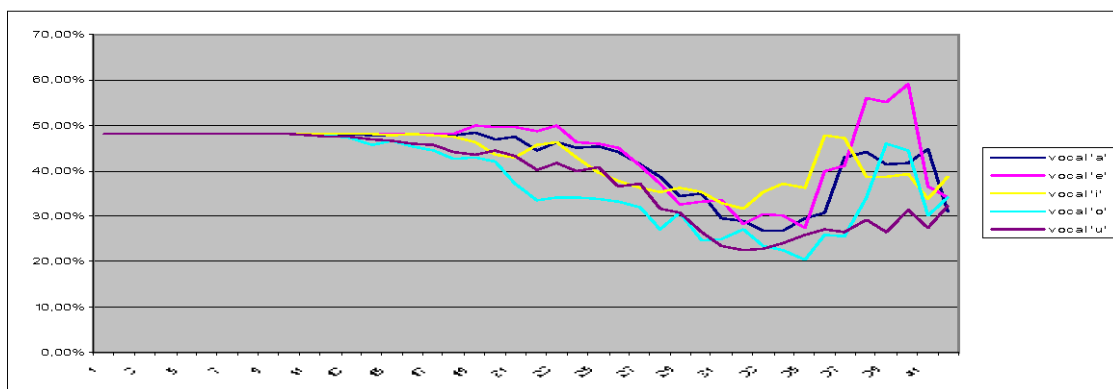


Gráfico 8-8: Evaluación secuencial de la relevancia para las medidas de calidad de la voz del grupo III usando el clasificador RBF en la clasificación entre distintos grados de disfonía

8.6.2.4 Máquinas de Vectores Soporte (SVM)

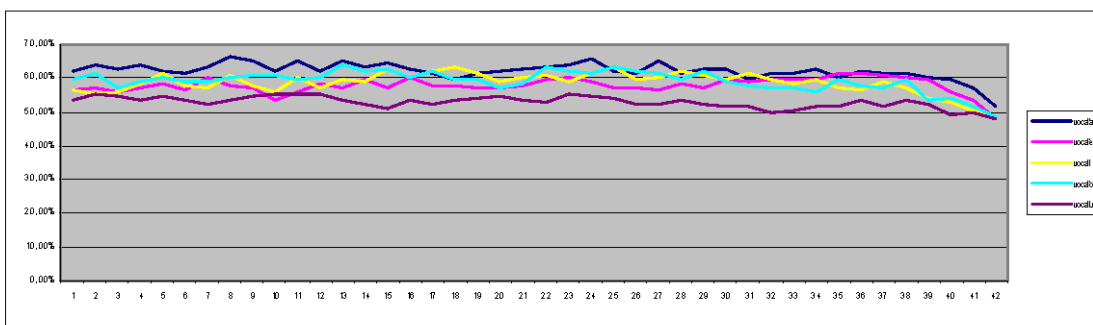


Gráfico 8-9: Evaluación secuencial de la relevancia para las medidas de calidad de la voz del grupo III usando el clasificador PNN en la clasificación entre distintos grados de disfonía

8.6.2.5 K vecinos más cercanos (KNN)

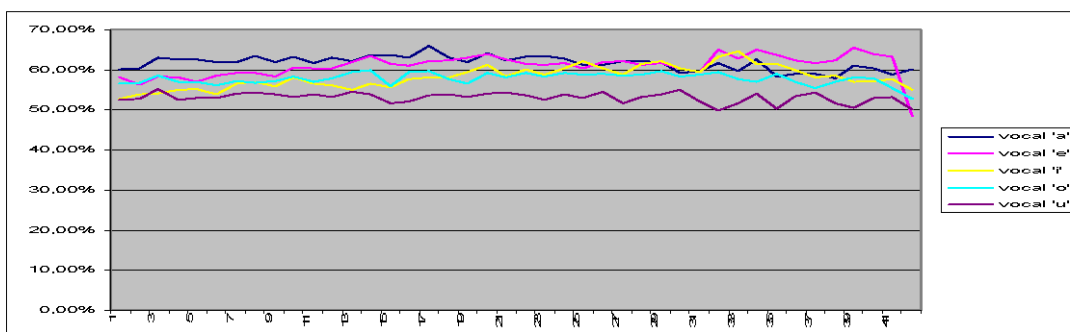


Gráfico 8-10: Evaluación secuencial de la relevancia para las medidas de calidad de la voz del grupo III usando el clasificador KNN en la clasificación entre distintos grados de disfonía

8.7 Evaluación

En este grupo de medidas existen algunas diferencias respecto a los resultados obtenidos para los anteriores conjuntos de medidas.

En la clasificación entre voz sana y voz patológica se observa que para la PNN existe un gran incremento de la tasa de éxito a medida que se incrementa el número de medidas de calidad, estando acotado el rango entre el 50 y 80%. Para el RBF existe igualmente gran inestabilidad de la tasa de éxito, estando acotado el éxito entre el 50 y 80%. El resto de los clasificadores, el NN, SVM y KNN, se muestran con una tasa de éxito totalmente estable a medida que se incrementa el número de medias de calidad, estando el rango de éxito acotado entre el 60 y el 80%.

En el caso de la clasificación entre voz sana, disfonía leve, disfonía moderada y disfonía severa, se observa que para la PNN existe un significativo incremento de la tasa de éxito, acotada entre el 45-60%. En la RBF, existe significativas oscilaciones que van desde el 50% hasta el 20%. Sin embargo, una vez más, el NN, SVM y el KNN, mantienen una tasa de éxito totalmente constante, entre el 55-65%, pese al incremento de medidas de calidad de la voz.

Se observa de forma general para los diferentes clasificadores y los distintos casos estudiados (clasificación entre voz sana y voz patológica, y clasificación entre voz sana, disfonía leve, disfonía moderada y disfonía severa) que la tasa de éxito en la clasificación no se incrementa de forma significativa cuando a la medida de calidad que presenta la mayor relevancia se evalúa con otras nuevas medidas, principalmente para NN, SVM y RBF. En la interpretación de los resultados es necesario tener en cuenta la reducida dimensión de la base de datos utilizada para la evaluación, lo cual puede justificar las fluctuaciones existentes.

Por lo tanto es posible asociar, para cada vocal por separado, la cuantificación del fenómeno físico “presencia de ruido” a una única medida, la cual se ha identificado como la medida que presenta mayor relevancia.

8.8 Referencias

- [1] Picone, J. W. , "Signal modeling techniques in speech recognition", *Proceedings of the IEEE* , vol. 81 , no.9 , pp.1215–1247, 1993
- [2] Boyanov, B; Hadjitodorov, S. "Acoustic analysis of pathological voices. A voice analysis system for the screening of laryngeal diseases". *IEEE Engineering in Medicine and Biology Magazine*. vol.16, no.4, pp.74 – 82, 1997
- [3] Feijoo, S, et al. "Acoustic evaluation of glottal cancer based on short-term stability measures". *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, vol.2, pp.675-676, 1989
- [4] Fröhlich, M.; Michaelis, D.; Strube, H.; Kruse, E. " Acoustic voice quality description: Case studies for different regions of the hoarseness diagram". *Advances in Quantitative Laryngoscopy*, 2nd 'Round Table', pp.143-150, 1997 .
- [5] Frohlich, M.; Michaelis, D.; Werner Strube, H. "Acoustic "Breathiness measures" in the description of pathologic voices". *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, vol.2, pp.937-940, 1998
- [6] Godino-Llorente, J.I.; Aguilera-Navarro, S.; Gomez-Vilda, P. "Non supervised neural net applied to the detection of voice impairment". *Proceedings. 2000 IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP '00*, vol.6, pp.3594-3597, 2000.
- [7] Jo, C.; Kim, K.; Kim, D.; Wang, S. "Screening of Pathological Voice from ARS using Neural Network". *International Workshop on MAVEBA*, pp.13-15, 2001.
- [8] Jo, Cheol-Woo; Kim, Dae-Hyun. "Classification of pathological voice into normal/benign/malignant state", *Sixth European Conference on Speech communication and Technology*, (EUROSPEECH'99). pp.571-574, 1999.
- [9] Kang, G.; Lin, J.; Yang, Q.; Cai, G. "A method of spectra-analysis for abnormal voice signals", *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, vol.6, pp.1887-1888, 1989
- [10] Kasuya, H.; Ogawa, S.; Kikuchi, Y. "An adaptive comb filtering method as applied to acoustic analyses of pathological voice". *IEEE International Conference on Acoustics, Speech, and Signal Processing*, ICASSP '86. vol.11, pp.669–672, 1986
- [11] Manfredi, M; Brusaglioni, P. "Pitch and Noise Estimation in Pathological Speech Signals". *International Conference on Information Systems, Analysis and Synthesis*, vol.5, pp.412-416, 2001
- [12] Michaelis D.; Frohlich M.; Strube H. W. "Selection and combination of acoustic features for the description of pathologic voices". *Acoustical Society of America*, vol.103, no.3, pp.1628-1640, 1998.
- [13] Michaelis, D.; Gramss, T.; Strube, H. W. "Glottal to noise excitation ratio - a new measure for describing pathological voices". *Acustica / acta acustica* (83), vol.83, pp.700-706, 1997.
- [14] O. Rosa, M.; Pereira, J.C.; Carvalho, A.C.P.L.F. "Evaluation of neural classifiers using statistic methods for identification of laryngeal pathologies", *Proceedings of Vth Brazilian Symposium on Neural Networks*, pp.220-225, 1998
- [15] Plante, F.; Kessler, H.; Cheetham, B.; Earis, J. "Speech monitoring of infective laryngitis", *Proceedings of Fourth International Conference on Spoken Language*, ICSLP 96, vol.2, pp.749-752, 1996

- [16] Ritchings, R.T.; McGillion, M.A.; Conroy, G.V.; Moore, C.J. "Objective Assessment of pathological voice quality". *IEEE International Conference on Conference Proceedings of Systems, Man, and Cybernetics, SMC '99*. vol.6, pp.340-345, 1999.
- [17] Umapathy, K.; Krishnan, S.; Parsa, V.; Jamieson, D. "Time-frequency modeling and classification of pathological voices". *Proceedings of 24th Annual Conference and the Annual Fall Meeting of the Biomedical Engineering Society Engineering in Medicine and Biology, 2002, EMBS/BMES Conference 2002*. vol.1, pp.116 – 117, 2002.
- [18] Umapathy, K.; Krishnan, S.; Parsa, V.; Jamieson, D.G. "Discrimination of pathological voices using an adaptive time-frequency approach". *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol.4, pp.3852-3855, 2002
- [19] Wallen, E.J.; Hansen, J.H.L. "A screening test for speech pathology assessment using objective quality measures". *Proceedings of Fourth International Conference on Spoken Language, 1996, ICSLP 96*, vol.2, pp.776-779, 1996.
- [20] Alonso, J. B.; de León, J.; Alonso, I.; Ferrer, M. A. "Automatic Detection of pathologies in the voice by Hos based parameters". *Eurasip journal on Applied signal procesing*, vol.2001, no.4, pp.275-284, 2001
- [21] Raphael Ahn y Harvey Holmes, "Harmonic-Plus-Noise Decomposition and its Application in Voiced/Invoiced Classification", *IEEE TENCON – Speech and Image Technologies for Computing and Telecommunications*, pp.587-590,1997.
- [22] B.Yegnanarayana, C.d'Alessandro, V.Darsinos,"An Iterative Algorithm for Decomposition of Speech Signals into Periodic and Aperiodic Components", *IEEE transaction on Speech and Audio Processing*, vol.6, no.1, pp.1-10,1998.
- [23] Mirjam Wester, Automatic "Cassification of Voice Quality: Comparing Regresión Models and Hidden Markov Models", *Proceedings of Symposium on Databases in Voice Quality Research and Education, VOICEDATA98*, pp 92-97, 1998.
- [24] Claudia Manfredi, "Adaptive Noises Energy Estimation in Pathological Speech Signals", *IEEE Transations on Biomedical Engineering*, vol.47,no.11, pp.1538-1543, 2000.
- [25] Marcelo de Oliveira Rosa, Carlos Pereira, Marcos Grellet, "Adaptive Estimation of Residue Signal for Voice Pathology Diagnosis", *IEEE Transactions on biomedical Engineering*, vol.47, no. 1, pp.96-104, 2000.
- [26] Stefan Hadjitodorov, Petar Mitev, "A computer system for acoustic analysis of pathological voices and laryngeal diseases screening", *Medical Engineering Physics*, vol.24, pp.419-429, 2002

CAPÍTULO 9

Cuantificación del comportamiento no lineal

Las manifestaciones en la señal de voz de un **comportamiento no lineal** están ocasionadas por un funcionamiento anómalo de las cuerdas vocales, debido, principalmente, a irregularidades en las masas de cuerdas vocales que intervienen en la fase de cierre de cuerdas, existencia de asimetrías en el movimiento de las cuerdas y factores relacionados con la mucosa de las cuerdas. Estos fenómenos son cuantificables por medio de la identificación de comportamientos no lineales en la señal de voz.

A continuación, y de forma similar a lo realizado en los tres últimos capítulos, se presentan las diferentes medidas de calidad de la voz pertenecientes al grupo que cuantifica la presencia de un “comportamiento no lineal”, tanto las presentes en la bibliografía actual como las propuestas en este trabajo, clasificadas en los diferentes dominios de representación. Además, se presenta una ordenación de dichas medidas en función de su relevancia. Por último, dicha ordenación de medidas es evaluada de forma incremental utilizando una batería de diferentes clasificadores.

9.1 Estado de la Técnica

En este apartado de la memoria, se presenta el conjunto de las principales medidas de calidad que cuantifican la presencia de un comportamiento no lineal de la voz presentes en la bibliografía que han sido clasificadas en este trabajo, como medidas que cuantifican el comportamiento no lineal de la voz.

9.1.1 Modulaciones angulares

Existen estudios que indican que las características de las resonancias del modelo de producción de voz pueden cambiar rápidamente tanto en amplitud como en frecuencia, e incluso en un periodo de *pitch* [23] [34]. Mediante un operador diferencial no lineal *Teager* (operador TEO) se pueden detectar modulaciones de amplitud (AM) y modulaciones de frecuencia (FM) alrededor de una resonancia del tracto vocal, es decir, alrededor de un formante [23] [34] [35] [36]. El operador *Teager* es considerado un estimador de energía de alta resolución. El operador TEO es utilizado en varias áreas del procesamiento de señal como son la detección de voz patológica [37] [38], y la detección del estrés por medio de la voz [39].

En el modelo propuesto en los trabajos [23] y [34], se representa una resonancia de voz $R(t)$ como una señal de AM-FM:

$$R(t) = a(t) \cos \left\{ 2\pi \left[f_c t + \int_0^t q(\tau) d\tau \right] + \theta \right\} \quad [\text{Ec.9-1}]$$

donde $q(t)$ es la señal moduladora de frecuencia y $a(t)$ es la señal moduladora de amplitud en el tiempo. La frecuencia instantánea del formante se define como $f_i(t) = f_c + q(t)$, donde f_c es la frecuencia de un formante (polo en el modelo del tracto vocal). Con la finalidad de, a partir de la resonancia, remodelar sus variaciones de amplitud $|a(t)|$ y su frecuencia instantánea $f_i(t)$, se aplica un algoritmo de separación de energías (ESA) a la señal resonante $R(t)$, obtenida ésta de filtrar la señal de voz alrededor del formante que se estudia.

El operador TEO se define, para tiempo continuo, como:

$$\psi[s(t)] = [\dot{s}(t)^2 - (s(t)\ddot{s}(t))] \quad [\text{Ec.9-2}]$$

donde $\dot{s}(t) = ds/dt$, mientras que para tiempo discreto se define como:

$$\psi[s(n)] = [s^2(n) - (s(n+1)s(n-1))] \quad [\text{Ec.9-3}]$$

Mediante el algoritmo ESA se obtienen las siguientes ecuaciones que nos permiten calcular $|a(n)|$ y $f_i(n)$:

$$f_i(n) \approx \frac{1}{4\pi T} \arccos \left\{ 1 - \frac{\psi[x(n+1) - x(n-1)]}{2\psi[x(n)]} \right\}$$

$$a(n) \approx \frac{2\psi[x(n)]}{\sqrt{\psi[x(n+1) - x(n-1)]}} \quad [\text{Ec.9-4}]$$

Los errores que se cometen en estos estimadores son debidos a las diferencias que existen entre la versión analógica y la versión discreta del estimador $\psi[x(n)]$, los cuales han de ser tenidos en cuenta. Esto se realiza fijando un umbral en el operador TEO de tal forma que no se obtengan valores negativos [40] [41] .

Se proponen, en los trabajos [23] y [34] , cuatro medidas para el estudio de la calidad de la voz:

La primera medida es la variación de la señal $q(t)$ a lo largo del tiempo, véase [Ec.9-1].

La segunda medida, corresponde a una medida vectorial, para la cual es necesario estimar la envolvente de la función de autocorrelación normalizada de la señal moduladora de AM $|a(n)|$. A partir de la autocorrelación normalizada de la señal moduladora de AM se modela la envolvente mediante un polinomio cúbico de la siguiente forma:

$$p(X) = C_0 + C_1X + C_2X^2 + C_3X^3 \quad [\text{Ec.9-5}]$$

donde X es el eje de tiempo de la función de autocorrelación normalizada, $p(X)$ es la envolvente y C_0 es igual a uno. La segunda medida de calidad de la voz, una medida vectorial, está constituida por los valores de C_1 , C_2 y C_3 .

La tercera medida de calidad de voz es la energía contenida en la función de autocorrelación normalizada respecto a -1 .

Y la cuarta medida de calidad es la pendiente de la función de autocorrelación en la franja comprendida entre 0 y 5 milisegundos.

El esquema general del proceso se muestra en la Figura 9-1.

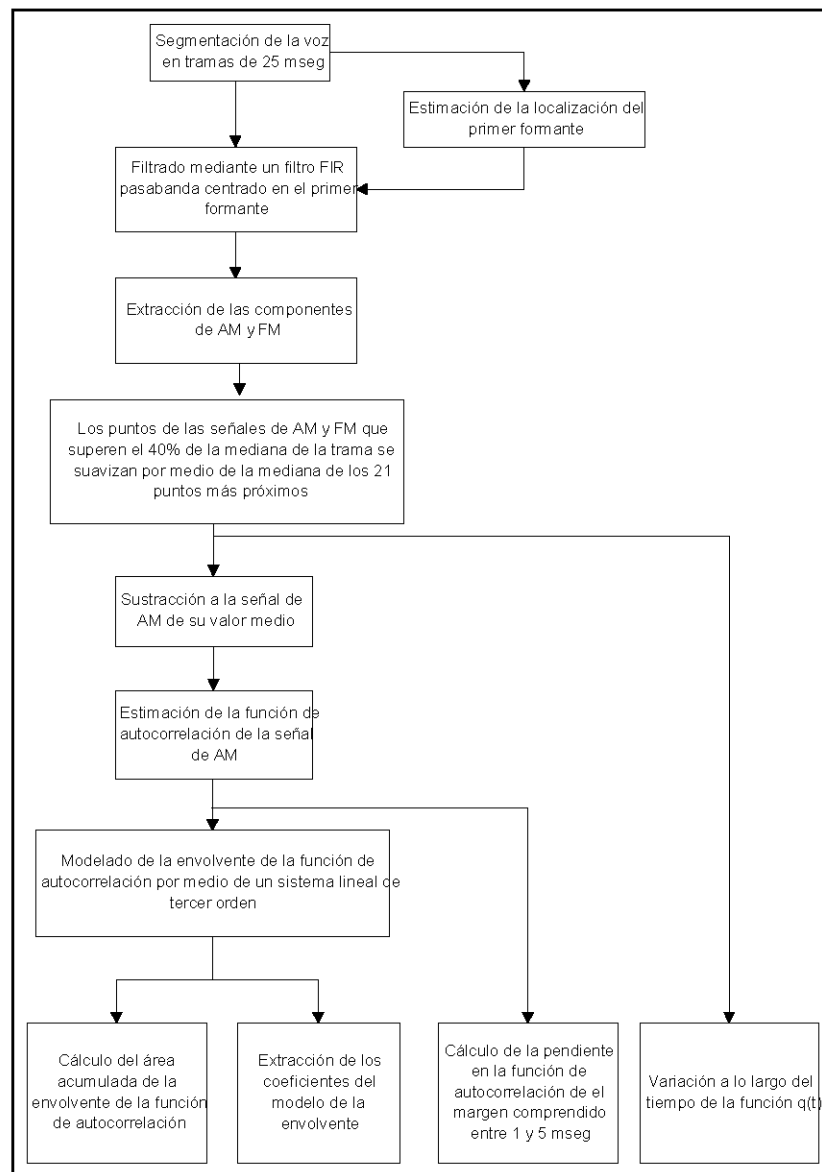


Figura 9-1: Esquema de la estimación de las medidas de calidad de la voz basadas en modulaciones angulares

A continuación se presentan distintas figuras en las que se muestran las señales que intervienen en el algoritmo. En la Figura 9-2 se muestra la respuesta impulsiva del filtro pasobanda, centrado en el primer formante, que se utiliza en el algoritmo. En la misma Figura 9-2 se muestra el espectro de potencia de una trama, de 25 milisegundos, en la que se aprecia además la envolvente espectral calculada mediante un filtro LPC.

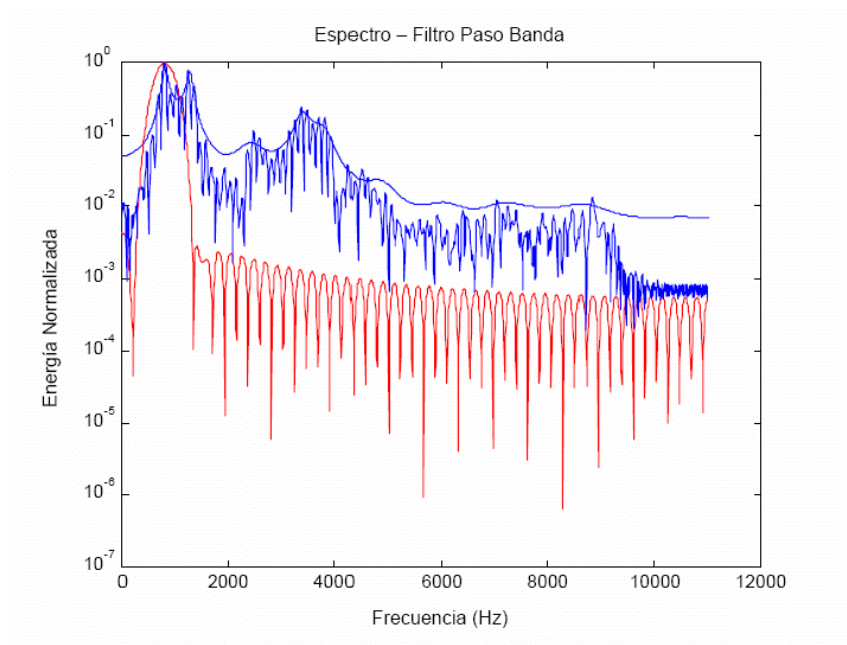


Figura 9-2: Espectro de una trama de voz (azul) y la respuesta del filtro pasobanda centrado en el primer formante (rojo). En el espectro se puede apreciar la envolvente espectral calculada mediante coeficientes LPC

En las Figura 9-3 y Figura 9-4 se muestra las señales moduladoras de AM y FM, obtenidas de una muestra de voz de alta calidad (voz sana) y una muestra baja calidad (voz patológica). Se aprecia en las gráficas una mayor periodicidad en las señales de AM y FM de la muestra de voz de alta calidad.

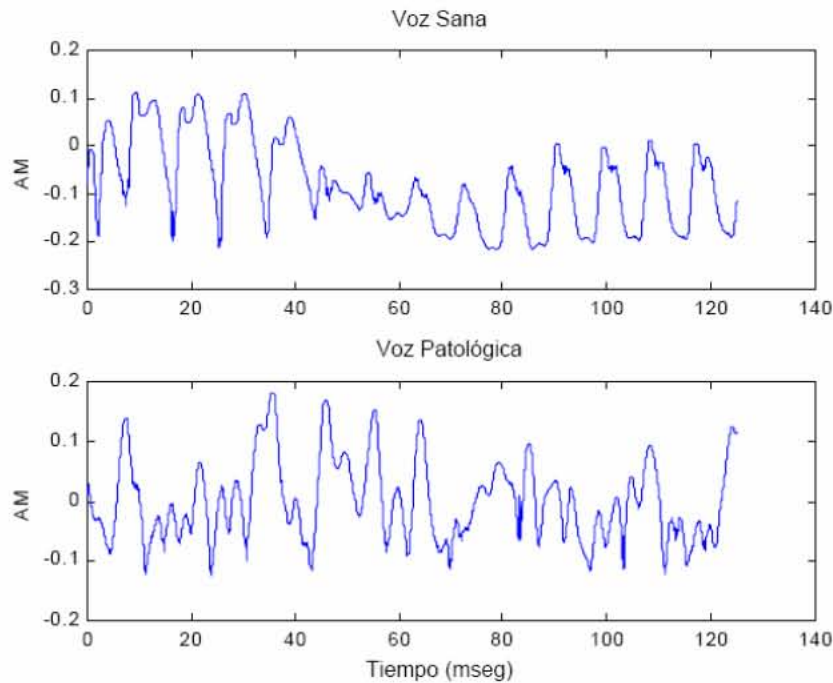


Figura 9-3: Señal moduladora de AM de una señal de voz de alta calidad (voz sana) y de una señal de voz de baja calidad (voz patológica)

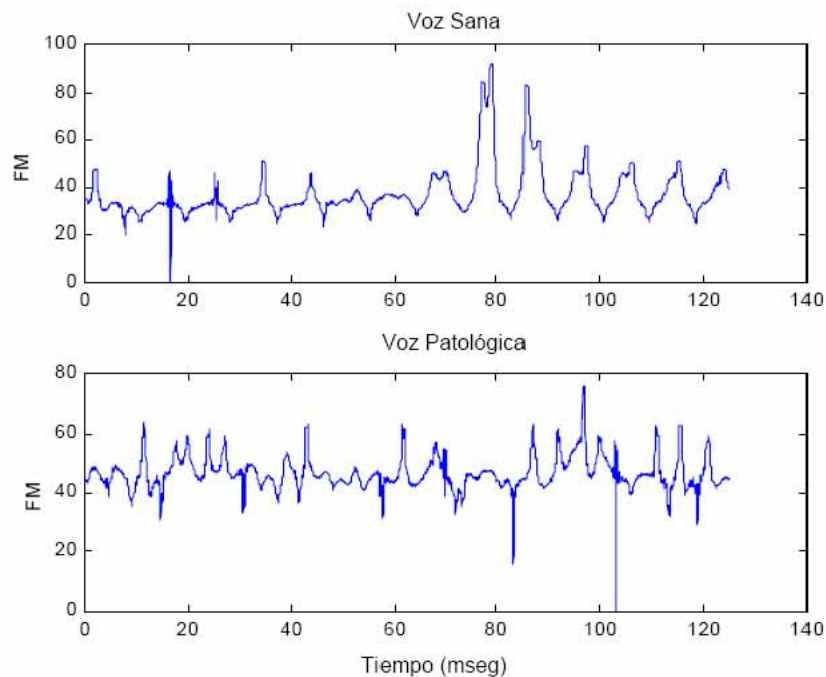


Figura 9-4: Señal moduladora de FM de una señal de voz de alta calidad (voz sana) y de una señal de voz de baja calidad (voz patológica)

Las figuras Figura 9-5 y Figura 9-6 muestran la función de autocorrelación normalizada de la señal moduladora de AM y su envolvente, modelada mediante un

polinomio cúbico, tanto para una muestra de voz de alta calidad (voz sana) como para una muestra de voz de baja calidad (voz patológica).

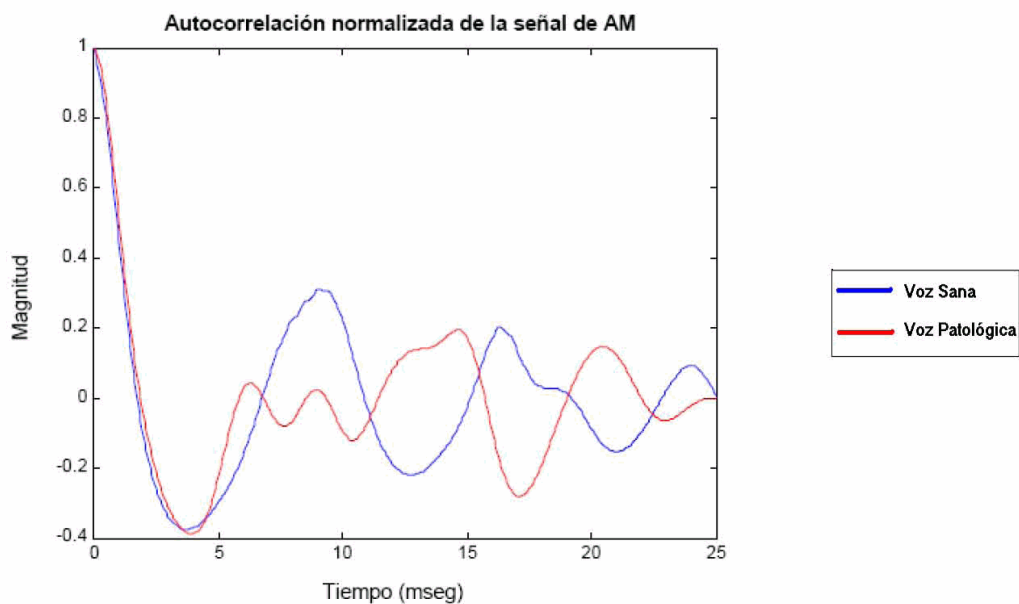


Figura 9-5: Funciones de autocorrelación normalizada de la señal moduladora de AM para voz de alta calidad (azul) y para voz de baja calidad (roja)

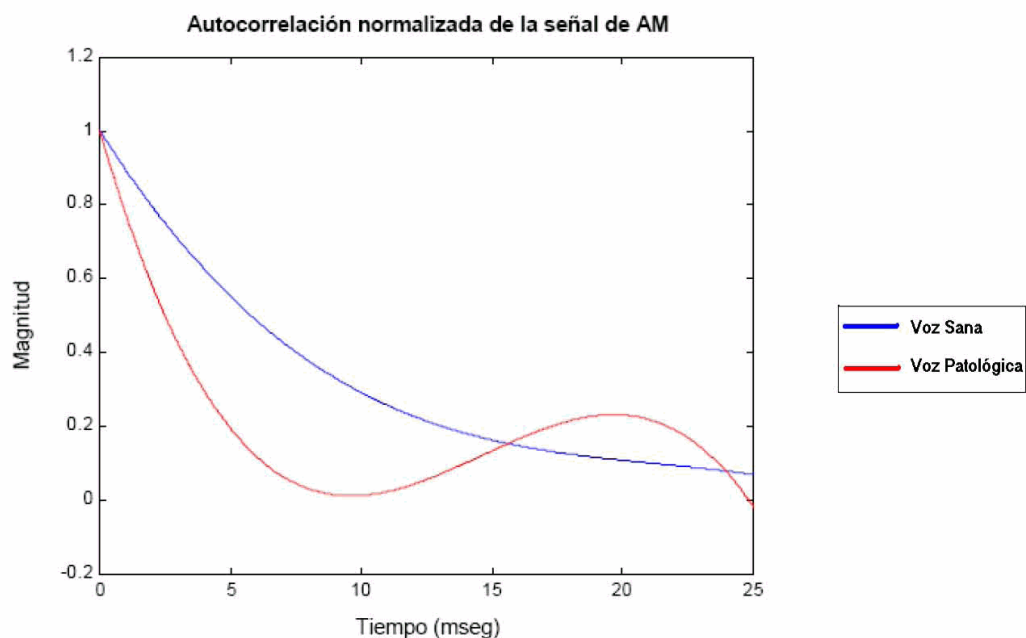


Figura 9-6: Envolventes de las funciones de autocorrelación normalizada de la señal moduladora de AM para voz de alta calidad (azul) y para voz de baja calidad (roja)

9.1.2 Estadísticos de Orden Superior (*High order statistic* (HOS))

Se ha indicado la posibilidad de utilizar el biespectro, para el reconocimiento de locutores y para la identificación automática de patologías laríngeas a partir de grabaciones de voz [11]. Descripciones más detalladas sobre los estadísticos de orden superior pueden encontrarse en [42] y [43].

9.1.3 Sistemas dinámicos no lineales deterministas: Sistemas Caóticos

La Teoría del Caos es utilizada para comprender mejor e interpretar el comportamiento de las observaciones de los sistemas dinámicos complejos, resultando interesante para la predicción y control de la evolución temporal de dichas series temporales [44].

Los sistemas dinámicos determinísticos describen la evolución temporal de un sistema en un espacio de estados $\Gamma \subset \mathbb{R}^d$, tal que un determinado caso de una evolución temporal puede ser descrita por medio de ecuaciones diferenciales ordinarias:

$$\dot{x}(t) = F(x(t)) \quad [\text{Ec.9-6}]$$

o en un sistema discreto donde $t = n\Delta t$, por medio de:

$$x_{n+1} = F(x_n) \quad [\text{Ec.9-7}]$$

Desafortunadamente, el vector del espacio de estados se puede deducir solamente para los sistemas absolutamente simples, y como cualquier persona puede imaginarse, el sistema dinámico subyacente en el proceso de producción de voz es muy complejo. Sin embargo, según lo establecido por "*embedding theorem*" [44], es posible reconstruir un espacio del estado equivalente al original, denominado espacio de fases. Además, un vector del espacio de estados formado por las muestras de la observación retrasadas en el tiempo (nuestro caso, las muestras de voz) podría ser una opción apropiada:

$$\mathbf{s}_n = [s(n), s(n-T), \dots, s(n-(d-1)T)]^T \quad [\text{Ec.9-8}]$$

donde $s(n)$ es la señal de voz, d es la dimensión del vector del espacio de estados y T es el retardo (*delay*).

Finalmente, el vector reconstruido del estado-espacio dinámico, $s_{n+1} = F(s_n)$, puede ser modelado a través de los modelos locales o globales, por medio de modelos altamente polinomiales. Esta representación es denominada *atractor* (*attractor*).

En la Figura 9-7 se muestra un ejemplo de representación en el espacio de fases de una fonación sostenida de una muestra de voz para distintas calidades de voz.

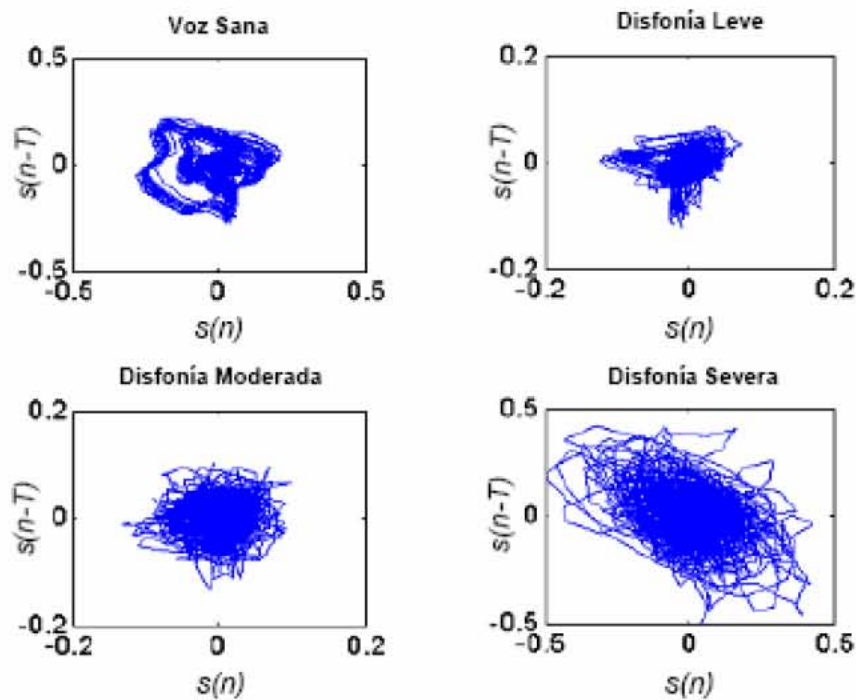


Figura 9-7: Representación en el espacio de fases de una fonación sostenida de una muestra de voz para distintas calidades de voz

Las señales de voz de baja calidad manifiestan una representación en el espacio de fases, el atractor, más dispersas que las voces de alta calidad.

En el campo de la dinámica caótica aplicada a la señal de voz, los trabajos realizados hasta ahora se basan o bien en la modelización mediante técnicas no lineales del proceso de generación de la voz o bien en la extracción de medidas basadas en la dinámica caótica. En general los parámetros más estudiados han sido los exponentes de Lyapunov y las dimensiones del atractor.

Existen pocas experiencias de la aplicación de modelos dinámicos no lineales, o características extraídas de ellos, aplicadas a la cuantificación de la calidad de la voz,

destacando la dimensión de correlación como parámetro con el que se puede distinguir voces sanas de patológicas [48] [49] . Además, se ha propuesto la utilización de la dimensión fractal [50] , en este último trabajo donde los autores concluyen que mediante este parámetro se obtiene una clara separación entre voces sanas y patológicas, e incluso entre diferentes tipos de patologías.

9.1.4 Modelos No Lineales de Cuerdas Vocales

La dinámica de las cuerdas vocales ha sido extensamente estudiada durante décadas y se han desarrollado varios modelos que simulan su comportamiento, pudiendo obtenerse a partir de éstos, la forma de onda del flujo glotal. Los principales modelos desarrollados se pueden dividir en cuatro categorías principales:

- Modelos mecánicos, donde la fuente glotal se representan a partir de osciladores mecánicos. Dentro de este grupo se encuentran el modelo de una masa (Flanagan y Landgraf [59]), el modelo de dos masas (Ishizaka y Flanagan [60]) y el modelo de múltiples masas (Titze [61]).
- Modelo continuo, donde las cuerdas vocales se representan como un medio continuo y deformable.
- Modelo de Ribbon [52] , con cuatro parámetros, donde la vibración de las cuerdas vocales se representa a través del movimiento de una cinta.
- Modelos *Body-cover* [62] . En estos modelos, la cuerda vocal se divide en el cuerpo (*body*), formado principalmente por la masa muscular, y los tejidos que la cubren (*cover*).

Los modelos más destacados son los modelos mecánicos y en concreto el modelo de dos masas, pues tiene un alto grado de eficiencia, sin implicar unos cálculos demasiado engorrosos.

9.1.4.1 Modelo de una masa

En el modelo de una masa la vibración de las cuerdas vocales se simula a partir de un sistema formado por una sola masa, un resorte y un pistón, como se puede ver en la siguiente figura:

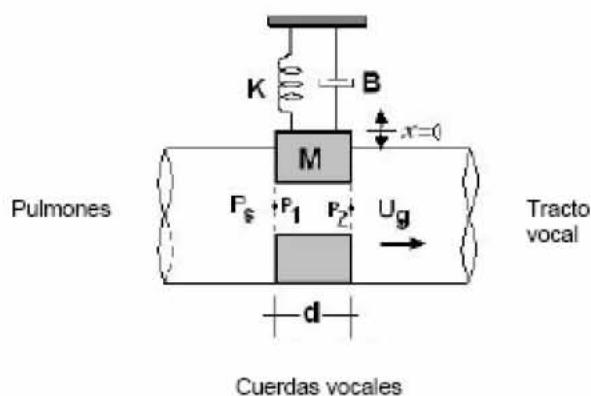


Figura 9-8 : Modelo de cuerdas vocales “Modelo de una masa”

El modelo de una masa[53] [58] se trata de un modelo simple con baja carga computacional que considera la interacción fuente-tracto.

9.1.4.2 Modelo de dos masas (Ishizaka y Flanagan)

El modelo de dos masas permite una simulación realista de las propiedades de la glotis. En este modelo, las cuerdas vocales se dividen en una masa superior y una masa inferior, debido a la división anatómica y funcional entre la mucosa y la base de la cuerda vocal. Cada parte está formada por un oscilador compuesto de una masa, un resorte y un pistón. Los resortes representan las propiedades elásticas de las cuerdas vocales, mientras que el pistón representa las fuerzas disipativas, como la viscosidad y la fricción. Además existe una interacción entre las dos masas, la cual se representa mediante una rigidez de acoplo, (K_c o S_c).

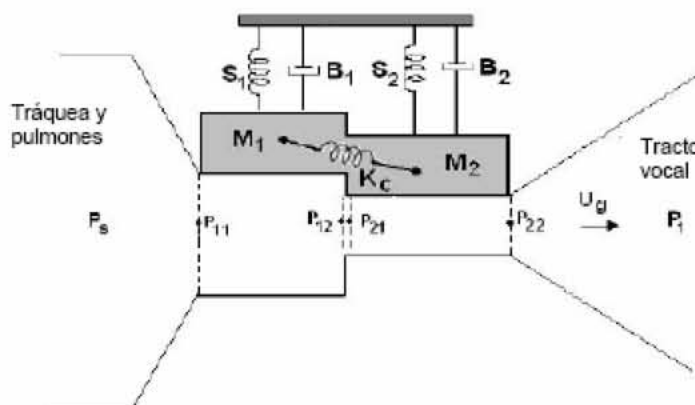


Figura 9-9 : Modelo de cuerdas vocales “Modelo de dos masas”

En el modelo, los resortes tienen una característica no lineal para ajustarse a la rigidez existente en las cuerdas vocales. Durante el cierre de la glotis, existe una fuerza de contacto que da lugar a una deformación adicional.

En [52] se incluye una tabla con los valores sugeridos por Ishizaka y Flanagan para las constantes fisiológicas de una voz normal. Para simular las distintas calidades de la voz se deben ajustar los parámetros adecuadamente.

El modelo de dos masas proporciona una simulación realista de las propiedades glotales, permitiendo reproducir voz natural con una carga computacional razonable.

9.1.4.3 Modelo de múltiples masas

El modelo de dos masas no es capaz de modelar los distintos modos de vibración longitudinal observados en la fonación humana. Titze, en un intento de mejorar el modelo anterior, propuso un modelo de 16 masas compuesto por dos filas de 8 masas cada una. Las masas de la fila superior representan principalmente la mucosa, mientras que las masas de la fila inferior el ligamento vocal y el músculo vocalis (la base de las cuerdas vocales). Las fuerzas T_m y T_v representan las tensiones longitudinales. Las constantes de los resortes de ambas filas, aumentan de forma no lineal respecto a la elongación de las cuerdas vocales.

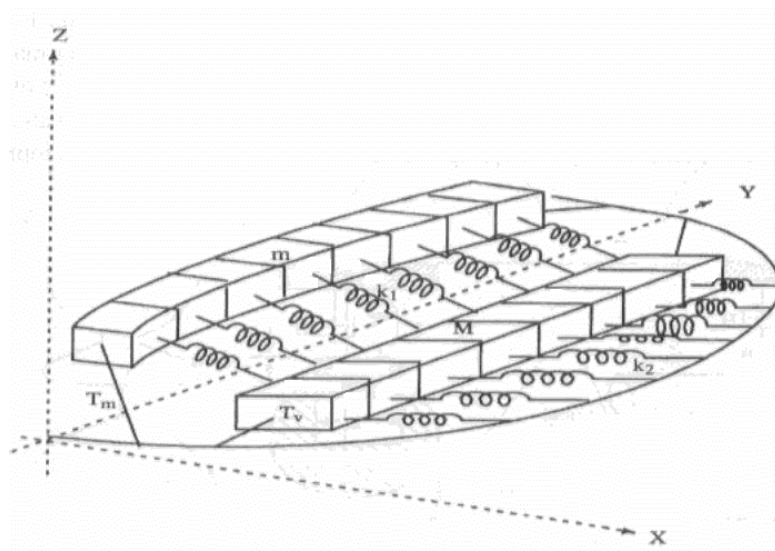


Figura 9-10 : Modelo de cuerdas vocales “Modelo de múltiples masas”

El modelo de múltiples masas se trata de un modelo complejo con una alta carga computacional. Se puede regular con parámetros que tienen una correlación fisiológica directa, consiguiendo aumentar la naturalidad de los sonidos.

9.1.4.4 Modificaciones de los modelos mecánicos

Otros autores [54] [55] consideran el modelo de una masa como el formado por una masa para cada una de las cuerdas vocales, como se observa en la siguiente figura:

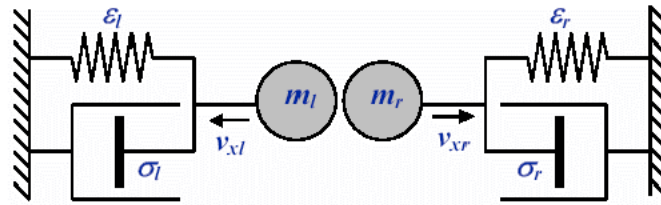


Figura 9-11 : Modelo de cuerdas vocales “Modelo de una masa”

donde m_l es la masa de la cuerda izquierda y m_r es la masa de la cuerda derecha. Este modelo expresa la elasticidad y viscosidad de las cuerdas vocales a través de los parámetros ϵ y σ , respectivamente. Las masas se mueven a la derecha y a la izquierda con velocidades v_{xr} y v_{xl} . Si la excitación es suficiente, ambas cuerdas colisionan y rebotan.

Debido a la simplicidad del modelo, éste no permite explicar un fenómeno muy importante, el efecto de vibración de la mucosa (*mucosal wave effect*), que se produce en las cuerdas vocales debido a la distribución de masas en los tejidos de las cuerdas [56]. Ésto se puede describir como un segundo modo de vibración, superpuesto al primer modo, el anterior. La explicación de este fenómeno se encuentra utilizando el modelo de dos masas, que se muestra en la siguiente figura:

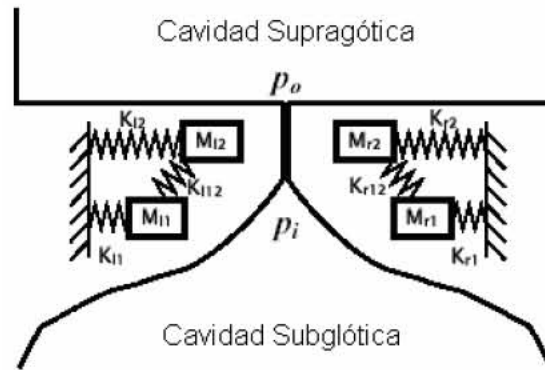


Figura 9-12 : Modelo de cuerdas vocales “Modelo de dos masas modificado”

El modelo incluye dos masas por cada cuerda: M_{l1} y M_{r1} para la primera cuerda y M_{l2} y M_{r2} para la segunda. Además hay tres resortes por cuerda (K_{l1} , K_{l2} , K_{l12} , y K_{r1} , K_{r2} , K_{r12}) que representan el comportamiento elástico de ambas masas y su acoplamiento. Las pérdidas de viscosidad se representan a través de los parámetros R_{l1} , R_{l2} , R_{r1} y R_{r2} .

Utilizando este modelo y a través de técnicas de filtrado inverso se puede estimar la presencia de vibración de la mucosa [54]. En este artículo se llega a un resultado interesante: cuando se estudia una voz patológica, la apertura glotal y la apertura glotal diferencial se comportan como el primer modo de vibración (modelo de una masa) y la vibración de la mucosa no está presente.

A partir del modelo de dos masas de Ishizaka y Flanagan, se han derivado muchos estudios de los que han surgido nuevos modelos. Por ejemplo, en [57] se propone un modelo basado aproximadamente en este modelo pero reduciendo el número de parámetros. Primero se estima la forma de onda de excitación glotal, junto con los parámetros del filtro del tracto vocal, utilizando técnicas de filtrado inverso. La forma de onda estimada es utilizada para identificar el modelo glotal no lineal, que está formado por dos bloques configurados en lazo cerrado: un filtro resonante de segundo orden, H_{res} , sintonizado respecto a la frecuencia de pitch, y una función de regresión, f , cuyos coeficientes son estimados a través de técnicas de identificación no lineales.

9.2 Análisis

Las técnicas para cuantificar la calidad de la voz a partir de representaciones en dominios no lineales presentan gran interés debido la capacidad potencial de manifestar información que en otros dominios no ésta presente o es difícilmente cuantificable.

Los estadísticos de alto orden, por un lado presentan todo una gama de nuevas opciones a la hora de cuantificar la presencia de alinealidades o la posibilidad de separar la componente armónica de la componente de ruido. Sin embargo, por otro lado, y de forma general, la utilización de técnicas basadas en estadísticos de orden superior presenta el inconveniente de tener un coste computacional significativo. Ésto obliga a la utilización de algoritmos optimizados y estrategias que minimicen el coste computacional.

De igual forma, las técnicas para cuantificar la calidad de la voz basadas en características obtenidas a partir de modelos dinámicos no lineales con comportamiento caótico de la señal de voz, presentan un gran potencial, teniendo en cuenta los buenos resultados mostrados en otros campos. Sin embargo, también presentan el inconveniente de que si los modelos están perfectamente sintonizados (dimensión del espacio de fases y el *delay*) probablemente no sea posible obtener las máximas prestaciones.

Debido a que la estimación de los parámetros de los diferentes modelos de las cuerdas vocales no es trivial, y dado que las técnicas a utilizar son más propias del área de la síntesis de voz, en esta tesis doctoral no se han tenido en cuenta estos modelos del funcionamiento de las cuerdas vocales, pese a que a partir de ellos sería posible cuantificar la calidad de la voz de un locutor.

9.3 Aportaciones

9.3.1 Estadísticos de Orden Superior (HOS)

En el trabajo [25] se proponen nuevas medidas de calidad que cuantifican la presencia de un comportamiento no lineal de la voz las cuales se estiman a partir de los Estadísticos de Alto Orden (HOS). La mayoría de las medidas son estimadas directa o indirectamente a partir del biespectro de una trama de voz.

9.3.1.1 Perturbación del Índice de Bicoherencia

La señal de voz presenta acoplos cuadráticos de fase. Este fenómeno es cuantificable por medio de la rugosidad del índice de bicoherencia o un balance de energía del índice de bicoherencia integrado.

El índice de bicoherencia es definido [26] [27] como:

$$b(\omega_1, \omega_2) \equiv \frac{B(\omega_1, \omega_2)}{P(\omega_1)P(\omega_2)P(\omega_1 + \omega_2)} \quad [\text{Ec.9-9}]$$

donde $P(\omega_1)$ es el espectro de potencia y $B(\omega_1, \omega_2)$ es el biespectro. El biespectro se define de la siguiente manera:

$$B(\omega_1, \omega_2) = \sum_{\tau_1=-\infty}^{\infty} \sum_{\tau_2=-\infty}^{\infty} C_{3,x}(\tau_1, \tau_2) \times e^{-j \sum_{i=1}^2 \omega_i \tau_i} \quad [\text{Ec.9-10}]$$

donde $C_{3,x}(\tau_1, \tau_2)$ es el *cumulant* de tercer orden, y es definido como:

$$C_{3,x}(\tau_1, \tau_2) = E\{x(t)x(t + \tau_1)x(t + \tau_2)\} \quad [\text{Ec.9-11}]$$

La primera medida de la calidad de la voz propuesta es la rugosidad del índice de bicoherencia (IB), definiéndose como:

$$IB = \left(\frac{1}{N-1} \right) \left(\frac{1}{|b_{\max}|} \right) \sum_{w_1=0}^{\dim_x} \sum_{w_2=0}^{\dim_y-1} |b(w_1, w_2+1) - b(w_1, w_2)| \quad [\text{Ec.9-12}]$$

donde $b(w_1, w_2)$ es el índice de bicoherencia, $\dim_x$ y $\dim_y$ son sus dimensiones, y N es el producto entre $\dim_x$ y $\dim_y$. En Figura 9-13 y Figura 9-14 se muestra una representación del índice de bicoherencia para una trama de voz de alta calidad y una trama de voz de baja calidad, respectivamente.

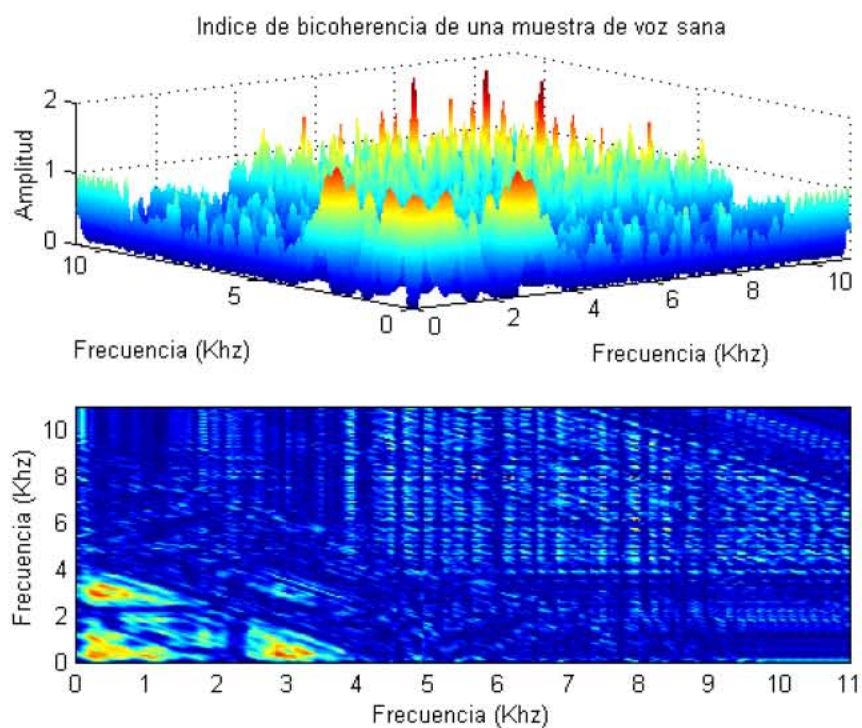


Figura 9-13: Índice de bicoherencia de una trama de voz de alta calidad

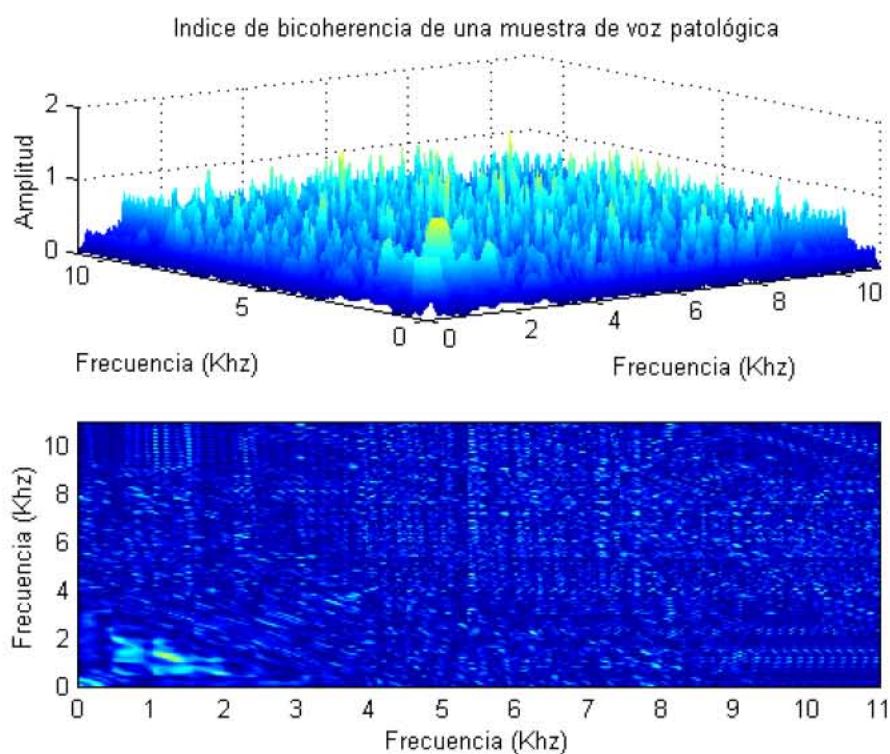


Figura 9-14: Índice de bicoherencia de una trama de voz de baja calidad

El potencial de la medida IB para cuantificar la calidad de la voz se justifica en el hecho de que la señal de voz de alta calidad presenta una mayor cantidad de acoplos cuadráticos de fase.

9.3.1.2 Índice de Bicoherencia Integrado

Es posible obtener una versión unidimensional del índice de bicoherencia, tal y como se propone en trabajos como [28] [29] [30] , denominándose índice de bicoherencia integrado.

$$b_{uni}(\omega) = \sum_{\omega_1=0}^{dim\_y} b(\omega, \omega_1) \quad [Ec.9-13]$$

donde $b(\omega_1, \omega_2)$ es el índice de bicoherencia, dim_y es una de las dos dimensiones del índice de bicoherencia y $b_{uni}(\omega)$ es el índice de bicoherencia integrado.

Un fenómeno que se observa cuando se representa el índice de bicoherencia integrado de una trama de voz es la presencia de mayor energía espectral en las bajas frecuencias en las voces de alta calidad respecto a las voces de baja calidad. Este fenómeno puede ser observado en la Figura 9-15, donde se representa el índice de bicoherencia integrado para una trama de voz de alta calidad y para una trama de voz de baja calidad.

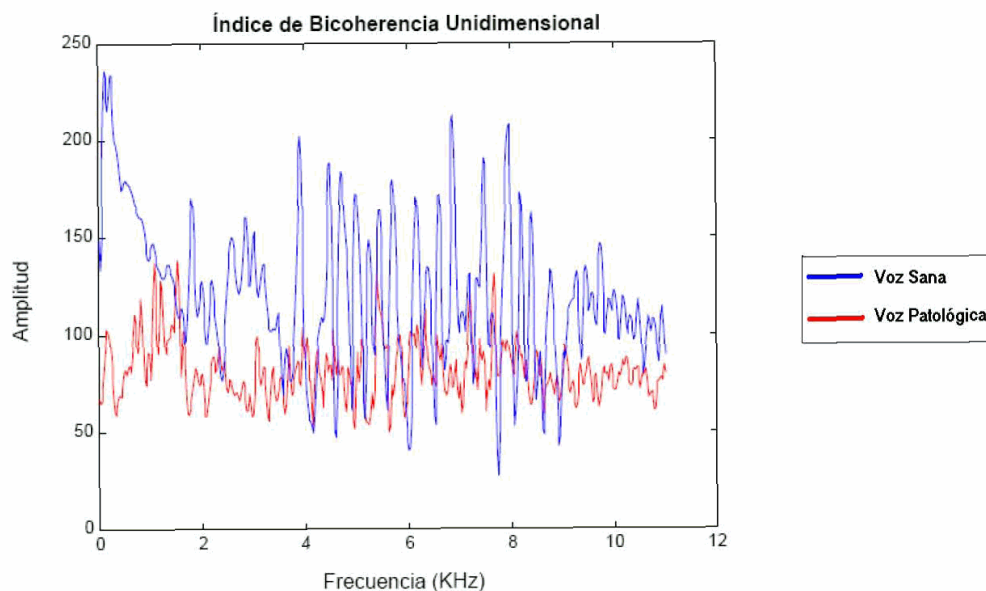


Figura 9-15: Índice de bicoherencia integrado para una trama de voz de alta calidad y para una trama de voz de baja calidad.

Dividiendo el índice de bicoherencia integrado en dos regiones, alrededor de la frecuencia 5300 Hz, es posible identificar una región superior y una región inferior. A partir de estas dos regiones, en [25] se proponen diferentes medidas de la calidad de la voz: la relación entre la energía de la región superior y la región inferior respecto a la energía total. Las diferentes energías son definidas como:

$$Energía_{superior} = \frac{\sum_{\omega_i = \text{umbral}}^{\omega_{sup}} b_{uni}(\omega_i)}{\sum_{\omega_i = 0}^{\omega_{sup}} b_{uni}(\omega_i)} \quad [\text{Ec.9-14}]$$

$$Energía_{inferior} = \frac{\sum_{\omega_i = 0}^{\omega_{sup}} b_{uni}(\omega_i)}{\sum_{\omega_i = 0}^{\text{umbral}} b_{uni}(\omega_i)} \quad [\text{Ec.9-15}]$$

Las medidas propuestas para la cuantificación de la calidad de la voz son:

Valor de la Energía de la Región Superior del Índice de Bicoherencia Integrado (**HFEUBI**): Es el valor de la medida “ $Energía_{superior}$ ” tomada para cada trama de voz.

Valor de la Energía de la Región Inferior del Índice de Bicoherencia Integrado (**LFEUBI**): Es el valor de la medida “ $Energía_{inferior}$ ” tomada para cada trama de voz.

Perturbación de la Energía de la Región Inferior de Índice de Bicoherencia Integrado (**ILFEUBI**): Es la variación a lo largo del tiempo de valor de la medida “ $Energía_{inferior}$ ” tomada para cada trama de voz. El valor de la perturbación es calculado por medio de la siguiente expresión:

$$ILFEUBI = \left(\frac{1}{N-1}\right) \left(\frac{1}{LFEUBI_{max}}\right) \sum_{i=1}^{N-1} |LFEUBI(i+1) - LFEUBI(i)| \quad [Ec.9-16]$$

siendo N el número de tramas utilizadas para la evaluación.

La capacidad de estas medidas para cuantificar la calidad de la voz puede ser atribuida a que existe, en las voces de alta calidad, un mayor número de acoplos cuadráticos de fase en las altas frecuencias respecto a las bajas frecuencias.

9.3.1.3 Cuantificación del Ruido Glótico

En un determinado grupo de voces de baja calidad, en particular un conjunto de voces patológicas en las que existe un incorrecto cierre glótico, existe una gran cantidad de ruido debido a la presencia de dicha anomalía.

En [25] se propone que este ruido tiene dos componentes diferenciables; una componente que sigue una función densidad de probabilidad simétricamente distribuida y otra componente que sigue una función densidad de probabilidad asimétricamente distribuida, siendo posible detectar el ruido simétricamente distribuido por medio de la sustracción espectral entre el espectro calculado por medio de la FFT y el espectro calculado por medio del biespectro.

El procedimiento se basa en que el espectro obtenido a partir de la FFT aporta información de la componente armónica y la componente del ruido (tanto la componente del ruido que presenta una función densidad de probabilidad simétricamente distribuida como la componente del ruido que presenta una función densidad de probabilidad asimétricamente distribuida). Sin embargo, el espectro obtenido a partir del biespectro [31] ofrece información de la componente armónica y

sólo una de las componentes del ruido (sólo el ruido que presenta una función densidad de probabilidad asimétricamente distribuida), ya que el biespectro elimina la componente de ruido simétricamente distribuida. Por lo tanto, la diferencia espectral entre el espectro calculado por medio de la FFT y el espectro calculado por medio del biespectro nos devuelve el ruido simétricamente distribuido.

En [25] se asocia este ruido estimado al ruido añadido a la voz debido a la falta de cierre glótico, siendo este caracterizable como un proceso estacionario. Esta suposición de que el ruido es estacionario es corroborada en [32] y [33], siendo más cierta en el caso de voces de baja calidad y en particular en ciertas voces patológicas, donde se manifiesta una mayor cantidad de ruido blanco (simétricamente distribuido) en la excitación glótica debido al incorrecto cierre glótico. Este fenómeno se muestra en la Figura 9-16, principalmente en las altas frecuencias para una muestra de voz de baja calidad.

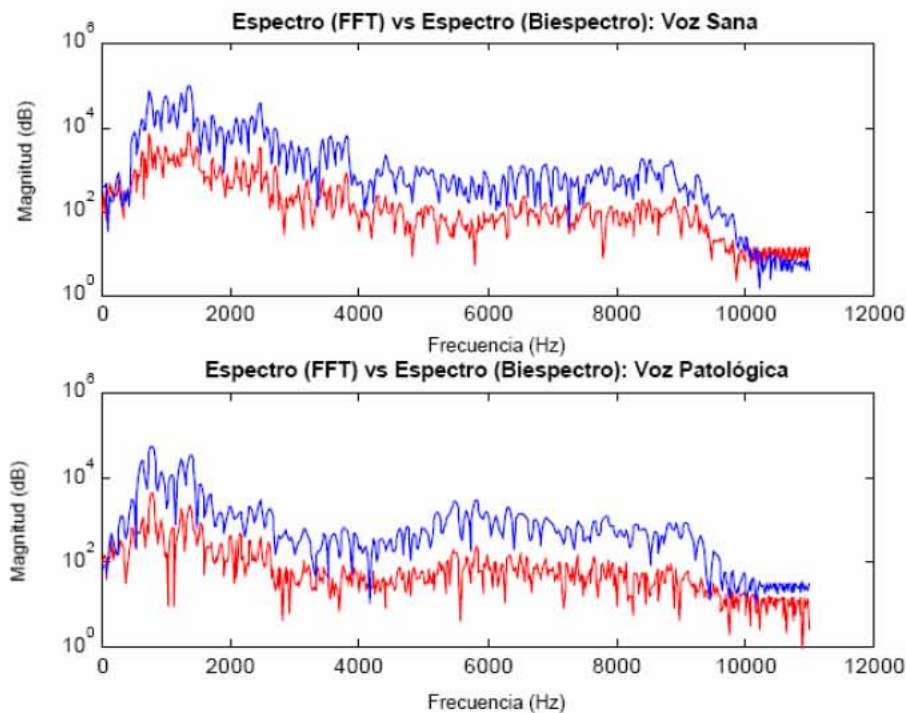


Figura 9-16: Superposición de la representación de la estimación del espectro de potencia calculado por medio de la FFT (azul) y por medio del biespectro (rojo), para una trama de voz de alta calidad (voz sana) y para una trama de baja calidad (voz patológica)

Las medidas de la calidad de la voz propuestas son: la perturbación del ruido estimado por el módulo del biespectro (MN) y la perturbación del ruido estimado por la misma técnica, teniendo en cuenta la información de la fase (MPN). La estimación de las medidas es a partir de las siguientes expresiones:

$$Ruido_{módulo} = \sum \left(\left| Espectro_{FFT}(\omega) \right| - \left| Espectro_{Biespectro}(\omega) \right| \right)^2 \quad [Ec.9-17]$$

$$MN = \left(\frac{1}{N-1} \right) \left(\frac{1}{Ruido_{módulo\_max}} \right) \sum_{i=1}^{N-1} \left| Ruido_{módulo}(i+1) - Ruido_{módulo}(i) \right| \quad [Ec.9-18]$$

$$Ruido_{módulo-fase} = \sum \left(Espectro_{FFT}(\omega) - Espectro_{Biespectro}(\omega) \right)^2 \quad [Ec.9-19]$$

$$MPN = \left(\frac{1}{N-1} \right) \left(\frac{1}{Ruido_{módulo\_fase\_max}} \right) \sum_{i=1}^{N-1} \left| Ruido_{módulo\_fase}(i+1) - Ruido_{módulo\_fase}(i) \right| \quad [Ec.9-20]$$

donde $Espectro_{FFT}$ es el espectro de una trama de voz y $Espectro_{Biespectro}(\omega)$ es el espectro de la misma trama. N es el número de tramas en las que se divide la muestra de voz.

9.3.1.4 Kurtosis

Algunas muestras de voces patológicas presentan gran variación de sus estadísticos siendo este fenómeno cuantificable por medio de la kurtosis. La kurtosis se define como:

$$Kurtosis = E \left\{ \left(\frac{x - \mu}{\sigma} \right)^4 \right\} \quad [Ec.9-21]$$

En [25] se propone una medida de la calidad de la voz a partir del kurtosis, la variación a lo largo de tiempo del valor del kurtosis de las diferentes tramas.

$$IK = \left(\frac{1}{N-1} \right) \left(\frac{1}{Kurtosis_{max}} \right) \sum_{i=1}^{N-1} \left| Kurtosis(i+1) - Kurtosis(i) \right| \quad [Ec.9-22]$$

9.3.2 Sistemas dinámicos no lineales deterministas: Sistemas Caóticos

En el trabajo [45] se ha realizado un estudio exhaustivo de diferentes medidas de calidad que cuantifican la presencia de un comportamiento no lineal de la voz obtenidas a partir de una representación de la señal de voz en el dominio del espacio de fases. Además existe otro trabajo [46], en el que se ha estudiado el tamaño óptimo de la ventana temporal, para alguna de las medidas de calidad de la voz. Para estos estudios se ha utilizado el *TISEAN package* [47].

9.3.2.1 Estimador del mínimo de la información mutua

La función de información mutua es una medida de la cantidad de información que poseemos sobre el valor de $s(n + \tau)$ si conocemos el valor de $s(n)$. Si se crea un histograma de la distribución de probabilidad de los datos, la información mutua promedio se define como:

$$I(\tau) = \sum_{i,j} p_{ij}(\tau) \ln \left[\frac{p_{ij}(\tau)}{p_i p_j(\tau)} \right] \quad [\text{Ec.9-23}]$$

donde $p_{ij}(\tau)$ es la probabilidad de que $s(n)$ esté en el bin i y que $s(n + \tau)$ esté en el bin j y p_i es la probabilidad de que la señal tenga un valor dentro del bin i -ésimo.

El mínimo de la función es:

$$I_{\min} = \min(I(\tau)) \quad [\text{Ec.9-24}]$$

Este parámetro cuantifica el valor de la información mutua en el momento de tiempo en el que la señal de voz $s(n)$ y su versión desplazada τ muestras, $s(n + \tau)$, son lo suficientemente independientes (es decir, no tan independientes como para estar completamente relacionados). En ese mínimo $s(n + \tau)$ agrega máxima información al conocimiento que tenemos de $s(n)$, es decir, la redundancia es menor.

9.3.2.2 Estimación de la dimensión de correlación (estimador *Takens*)

La dimensión de correlación cuantifica la organización espacial del atractor donde a mayor complejidad, mayor dimensión de correlación.

La dimensión de correlación se define como sigue:

$$D_2 = \frac{d \ln C(\varepsilon)}{d \ln \varepsilon} \quad [\text{Ec.9-25}]$$

siendo $C(\varepsilon)$ la suma de correlación sobre un conjunto de N puntos del espacio de fase:

$$C(\varepsilon) = \lim_{N \rightarrow \infty} \frac{1}{N(N-1)} \sum_{i=1}^N \sum_{\substack{j=1 \\ j \neq i}}^N \Theta(\varepsilon - \|X_i - X_j\|) \quad [\text{Ec.9-26}]$$

donde Θ es la función *Heaviside* o función escalón, $\Theta(x) = 0$ si $x \leq 0$ y $\Theta(x) = 1$ si $x > 0$. La función escalón cuenta el número de puntos que se encuentran dentro del círculo de radio ε y centro X_i . $C(\varepsilon)$ representa la fracción promedio de puntos dentro de ε .

El estimador *Takens-Theiler* de la dimensión de correlación realiza la siguiente operación sobre la suma de correlación:

$$D_{TT}(\varepsilon) = \frac{C(\varepsilon)}{\int_0^\varepsilon \frac{C(\varepsilon')}{\varepsilon'} d\varepsilon'} \quad [\text{Ec.9-27}]$$

En una representación de los valores de $C(\varepsilon)$ para diferentes valores de escala ε y de dimensión m , la suma de correlación ha de ser independiente de ε y además D_{TT} converge si aumentamos m . Finalmente, a partir de esta representación se obtienen diferentes curvas (cada curva representa un valor de dimensión m), siendo cada una de ellas función de ε . El dibujo de las curvas obtenidas tiene una parte plana en un rango de ε (concretamente el intermedio). El valor de esta parte plana será el estimador de la dimensión de correlación.

9.3.2.3 Estimador de la media de la entropía de correlación (*Grassberg-Procaccia*)

La entropía de correlación cuantifica la cantidad de desorden del sistema. Se define como sigue:

$$H_2(m, \varepsilon) = -\ln \sum_{i_1, i_2, \dots, i_m} p_{i_1, i_2, \dots, i_m}^2 \quad [\text{Ec.9-28}]$$

$$h_2 = \sup_I \lim_{m \rightarrow \infty} \left(\frac{1}{m} H_2(m, \varepsilon) \right) \quad [\text{Ec.9-29}]$$

o de forma equivalente:

$$h_2 = \sup_I \lim_{m \rightarrow \infty} (H_2(m+1, \varepsilon) - H_2(m, \varepsilon)) \quad [\text{Ec.9-30}]$$

como se observa, es la entropía diferencial de Rényi de orden 2.

La medida de calidad de la voz se estima por medio del algoritmo de *Grassber-Procaccia*, para $m=2$ y calculando su valor medio:

$$h_{2\text{MEDIA}} = \frac{1}{\text{length}(h_2)} \sum_{i=\varepsilon_{\min}}^{\varepsilon_{\max}} h_2(i) \quad [\text{Ec.9-31}]$$

9.3.2.4 Estimador de la media de la entropía Rényi de orden 1

La entropía de Rényi de bloque m de orden $q=1$ se define de la siguiente manera:

$$H_1(m, \varepsilon) = \sum_{i_1, i_2, \dots, i_m} p_{i_1, i_2, \dots, i_m} \ln(p_{i_1, i_2, \dots, i_m}) \quad [\text{Ec.9-32}]$$

Se estima por medio de una partición a diferencia de *Grasberg-procaccia*, que realiza un recubrimiento. Se estimó su valor medio para $m=2$.

$$H_{1\text{MEDIO}} = \frac{1}{L_{H_1}} \sum_{i=\varepsilon_{\min}}^{\varepsilon_{\max}} H_1(i) \quad [\text{Ec.9-33}]$$

donde L_{H_1} equivale al número de elementos de H_1 .

9.3.2.5 Estimador de la media de la entropía Rényi orden 2

La entropía de Rényi de bloque m de orden $q=2$ se define de la siguiente manera:

$$H_2(m, \varepsilon) = - \sum_{i_1, i_2, \dots, i_m} \ln(p_{i_1, i_2, \dots, i_m}^2) \quad [\text{Ec.9-34}]$$

De forma similar, se estima por medio de una partición a diferencia de *Grasberg-procaccia*, que realiza un recubrimiento, estimando su valor medio para $m=2$.

$$H_{2\text{MEDIO}} = \frac{1}{L_{H_2}} \sum_{i=\varepsilon_{\min}}^{\varepsilon_{\max}} H_2(i) \quad [\text{Ec.9-35}]$$

donde L_{H_2} equivale al número de elementos de H_2 .

9.3.2.6 Estimador de la entropía *Shannon* de la vocal

La entropía *Shannon* se define de la siguiente forma:

$$H = - \sum_i p_i \ln p_i \quad [\text{Ec.9-36}]$$

donde p_i es la probabilidad de que en el intervalo i -ésimo en el que se ha dividido los valores del espacio de fase se encuentre un determinado valor. Para el cálculo de la medida de la calidad basada en el valor de la entropía *Shannon*, se utiliza la totalidad de las muestras de la vocal.

9.4 Nomenclatura

En los siguientes apartados de este capítulo se presentan los resultados del estudio de la relevancia de las diferentes medidas de calidad que cuantifican la presencia de un comportamiento no lineal de la voz. Para mayor comodidad, a cada una de las medidas se le ha asignado un identificador. En la Tabla 9-1 se presentan los identificadores de las diferentes medidas.

TABLA 9-1: MEDIDAS DE CALIDAD DE LA VOZ PERTENECIENTES AL GRUPO IV (COMPORTAMIENTO NO LINEAL)

Identificador de la medida	GRUPO IV: “Comportamiento no lineal”
1	media Energía de la envolvente de AM
2	variación Energía de la envolvente de AM
3	media Pendiente de la envolvente de AM
4	variación Pendiente de la envolvente de AM
5	media 1er coeficiente de la envolvente de AM
6	variación 1er coeficiente de la envolvente de AM
7	media 2o coeficiente de la envolvente de AM
8	variación 2o coeficiente de la envolvente de AM
9	media 3o coeficiente de la envolvente de AM
10	variación 3o coeficiente de la envolvente de AM
11	media Perturbación del índice de bicoherencia
12	variación Perturbación del índice de bicoherencia
13	media Área superior del biespectro integrado
14	variación Área superior del biespectro integrado
15	media Área inferior del biespectro integrado
16	variación Área inferior del biespectro integrado
17	media cuantificación del ruido estacionario (magnitud y fase)
18	variación cuantificación del ruido estacionario (magnitud y fase)
19	media cuantificación del ruido estacionario (sólo magnitud)
20	variación cuantificación del ruido estacionario (sólo magnitud)
21	media Mínima información mutua
22	variación Mínima información mutua
23	media Mínima dimensión de la correlación a partir del estimador Taken
24	variación Mínima dimensión de la correlación a partir del estimador Taken
25	media Mínima dimensión de la correlación sólo para m=2
26	variación Mínima dimensión de la correlación sólo para m=2
27	media Medida de la entropía (H1) para m=2
28	variación Medida de la entropía (H1) para m=2
29	media Medida de la entropía (H2) para m=2
30	variación Medida de la entropía (H2) para m=2
31	Entropía de Shanon

9.5 Valoración de la Relevancia de las Medidas de Calidad de la Voz

En este apartado se presentan los resultados del proceso consistente en ordenar las diferentes medidas de calidad que cuantifican la presencia de un comportamiento no lineal de la voz en función de su relevancia. El estudio se ha realizado diferenciándose dos casos: la clasificación entre “voz sana – voz patológica” y la clasificación entre “voz sana – disfonía leve – disfonía moderada – disfonía severa”. Para cada caso, se ha estudiado las medidas de calidad de la voz diferenciando la vocal fruto de la fonación.

El procedimiento de obtención de la relevancia de cada medida ha sido descrito en el apartado 5.2. Para la estimación media de la relevancia de las diferentes medidas

de la calidad ha sido evaluado 100 veces y posteriormente los resultados han sido promediados.

9.5.1 Estudio: Sano – Patológico

En este caso las muestras de voz, consistentes en vocales, han sido etiquetadas o como “voz sana” o como “voz patológica”.

TABLA 9-2: ORDEN EN FUNCIÓN DE LA RELEVANCIA PARA EL ESTUDIO
SANO – PATOLÓGICO

Orden de relevancia	GRUPO IV Vocal ‘A’		GRUPO IV Vocal ‘E’		GRUPO IV Vocal ‘I’		GRUPO IV Vocal ‘O’		GRUPO IV Vocal ‘U’	
	Id. Medida	Prob.	Id. Medida	Prob.	Id. Medida	Prob.	Id. Medida	Prob.	Id. Medida	Prob.
1	15	70%	13	80%	15	80%	13	73%	27	76%
2	13	68%	15	74%	13	77%	17	68%	21	61%
3	18	49%	20	54%	30	57%	19	65%	15	60%
4	11	44%	17	49%	11	57%	15	62%	14	58%
5	4	42%	3	47%	20	53%	5	51%	13	56%
6	27	40%	1	47%	27	51%	1	49%	2	56%
7	21	40%	11	41%	21	46%	21	48%	3	54%
8	20	40%	18	39%	10	46%	18	48%	25	49%
9	14	39%	27	37%	1	45%	12	47%	11	43%
10	2	37%	19	37%	31	44%	10	46%	31	42%
11	31	36%	2	36%	22	44%	16	45%	7	40%
12	22	36%	23	33%	18	42%	11	40%	16	39%
13	1	36%	21	33%	6	41%	14	36%	10	39%
14	3	31%	4	33%	3	40%	2	34%	5	39%
15	12	29%	25	32%	28	39%	31	31%	1	37%
16	28	28%	5	29%	8	38%	9	30%	4	36%
17	24	27%	12	27%	2	37%	3	29%	26	35%
18	17	27%	26	26%	23	34%	23	27%	9	35%
19	16	27%	7	26%	29	33%	20	27%	29	34%
20	6	26%	29	25%	14	30%	7	27%	28	34%
21	9	25%	14	24%	16	29%	26	26%	23	34%
22	26	24%	24	23%	12	29%	24	26%	22	34%
23	23	24%	16	23%	24	28%	8	26%	24	33%
24	7	23%	9	23%	26	27%	22	23%	19	32%
25	5	21%	28	22%	4	27%	4	23%	18	32%
26	29	20%	22	19%	17	26%	28	22%	6	30%
27	8	18%	6	18%	7	26%	27	22%	12	29%
28	30	17%	31	15%	19	24%	6	19%	20	26%
29	25	17%	30	14%	25	23%	25	15%	8	26%
30	10	17%	10	14%	9	19%	30	14%	30	25%
31	19	16%	8	13%	5	15%	29	14%	17	25%

9.5.2 Estudio: Sano – Disfonía Leve – Disfonía Moderada – Disfonía Severa

En este caso las muestras de voz, al igual que en el apartado anterior consisten en vocales y han sido etiquetadas con una de las siguientes etiquetas: “voz sana”, “disfonía leve”, “disfonía moderada” o “disfonía severa”. Este etiquetado ha sido realizado atendiendo al grado de disfonía.

TABLA 9-3: ORDEN EN FUNCIÓN DE LA RELEVANCIA PARA EL ESTUDIO
SANO – DISFONÍA LEVE – DISFONÍA MODERADA – DISFONÍA SEVERA

Índice	GRUPO IV Vocal ‘A’		GRUPO IV Vocal ‘E’		GRUPO IV Vocal ‘I’		GRUPO IV Vocal ‘O’		GRUPO IV Vocal ‘U’	
	Medida	Prob.	Medida	Prob.	Medida	Prob.	Medida	Prob.	Medida	Prob.
1	11	74%	13	84%	15	74%	17	78%	27	68%
2	13	69%	15	77%	13	72%	15	69%	21	63%
3	1	67%	1	66%	11	71%	13	69%	15	56%
4	15	63%	20	62%	30	64%	21	67%	13	55%
5	3	60%	18	60%	27	52%	31	58%	14	53%
6	21	49%	3	59%	23	51%	19	55%	6	51%
7	31	47%	9	46%	28	46%	7	48%	19	47%
8	22	47%	11	44%	29	45%	1	48%	24	43%
9	18	43%	21	42%	21	45%	5	46%	11	42%
10	30	41%	31	38%	4	42%	11	45%	2	42%
11	26	40%	23	38%	31	39%	3	44%	8	41%
12	2	38%	4	37%	20	39%	12	42%	3	41%
13	12	37%	19	36%	1	38%	28	39%	23	40%
14	24	34%	7	36%	3	37%	18	35%	20	40%
15	9	34%	6	35%	18	36%	14	35%	1	40%
16	20	33%	25	34%	12	36%	9	33%	16	39%
17	4	33%	17	34%	9	36%	29	31%	5	37%
18	28	32%	2	32%	26	34%	27	31%	4	37%
19	19	31%	26	30%	17	34%	30	30%	28	36%
20	17	30%	8	29%	10	34%	24	27%	29	35%
21	23	29%	27	27%	8	34%	4	27%	25	35%
22	8	29%	5	27%	5	34%	20	26%	9	35%
23	10	28%	29	26%	2	33%	2	26%	26	34%
24	29	27%	12	26%	14	30%	16	25%	17	33%
25	14	27%	30	25%	22	29%	25	23%	31	32%
26	6	27%	24	25%	19	29%	23	23%	30	32%
27	27	26%	14	23%	16	29%	10	22%	18	26%
28	16	25%	10	23%	7	29%	26	17%	12	25%
29	5	19%	28	19%	6	29%	8	17%	10	25%
30	25	17%	16	16%	25	21%	6	17%	7	25%
31	7	15%	22	15%	24	17%	22	15%	22	24%

9.6 Evaluación de las Medidas de Calidad de la Voz

En este apartado se han evaluado las medidas de calidad que cuantifican la presencia de un comportamiento no lineal de la voz de forma secuencial, incrementando el número de medidas, siguiendo el orden descendente marcado por la relevancia de dichas medidas y siendo evaluadas en diferentes clasificadores: redes neuronales probabilísticas (PNN), redes neuronales *feedforward* (NN) con una capa oculta, clasificadores basados en funciones de base radial, máquinas de vectores soporte (SVM) y clasificadores basados en los K vecinos más cercanos (KNN).

Para cada subconjunto de medidas (cuya dimensión está indicada en número de medidas) de cada grupo y para cada vocal se obtiene la tasa de éxito en la clasificación dado un determinado clasificador. Los resultados en cada caso son promediados 20 veces.

9.6.1 Estudio: Sano – Patológico

9.6.1.1 Red Neuronal Probabilística (PNN)

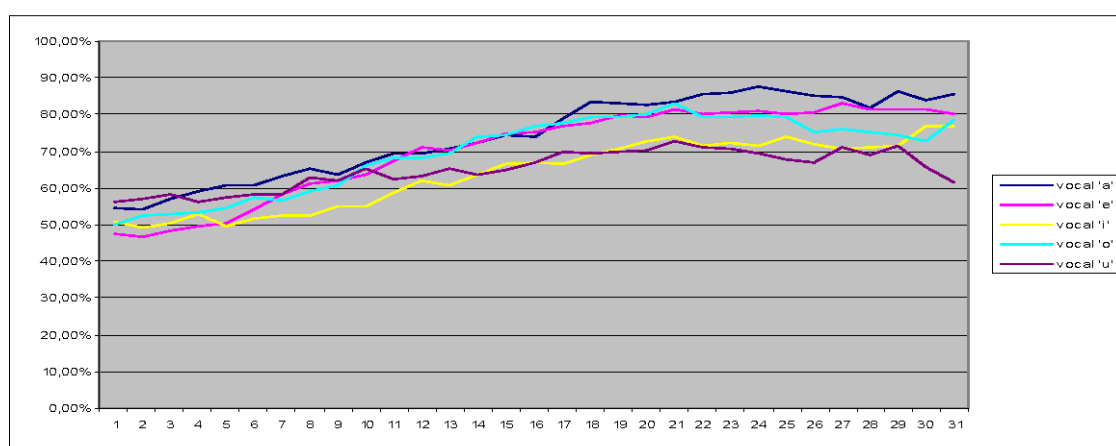


Gráfico 9-1: Evaluación secuencial de la relevancia para las medidas de calidad de la voz del grupo IV usando el clasificador PNN en la clasificación entre voz sana y voz patológica

9.6.1.2 Red Neuronal *FeedForward* con 1 Capa Oculta (NN)

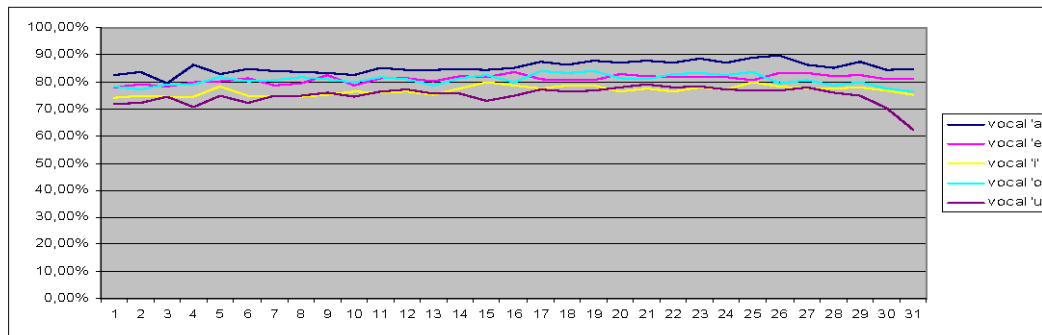


Gráfico 9-2: Evaluación secuencial de la relevancia para las medidas de calidad de la voz del grupo IV usando el clasificador NN en la clasificación entre voz sana y voz patológica

9.6.1.3 Función de Base Radial (RBF)

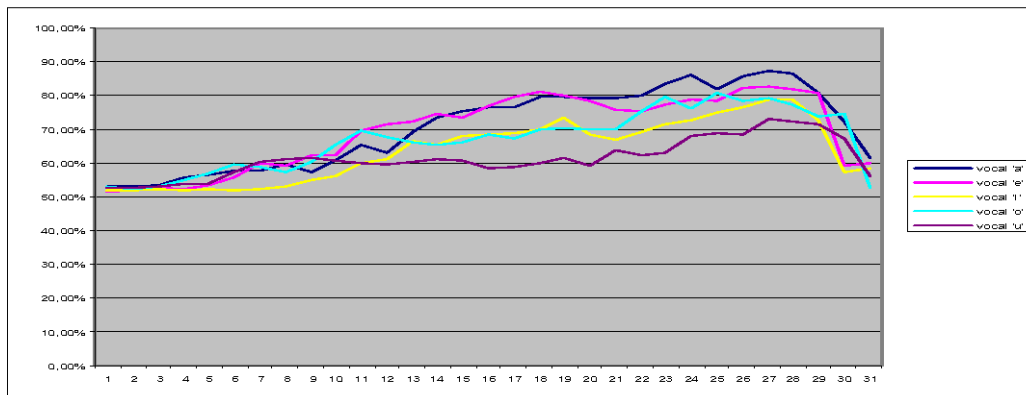


Gráfico 9-3 Evaluación secuencial de la relevancia para las medidas de calidad de la voz del grupo IV usando el clasificador RBF en la clasificación entre voz sana y voz patológica

9.6.1.4 Máquinas de Vectores Soporte (SVM)

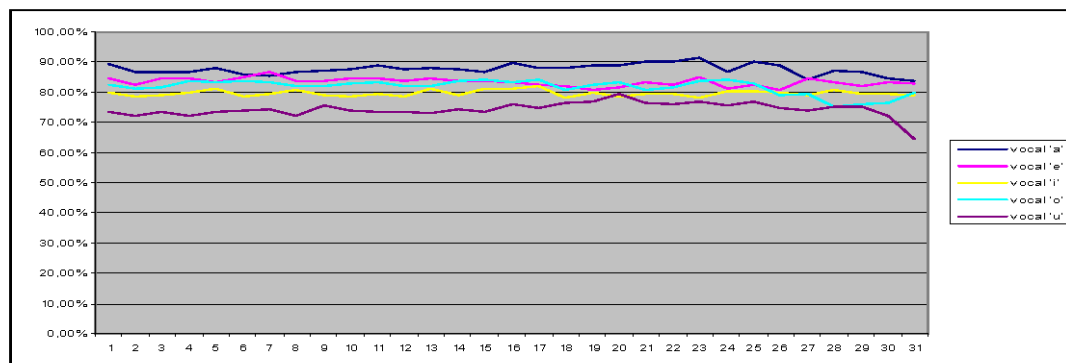


Gráfico 9-4 Evaluación secuencial de la relevancia para las medidas de calidad de la voz del grupo IV usando el clasificador SVM en la clasificación entre voz sana y voz patológica

9.6.1.5 K vecinos más cercanos (KNN)

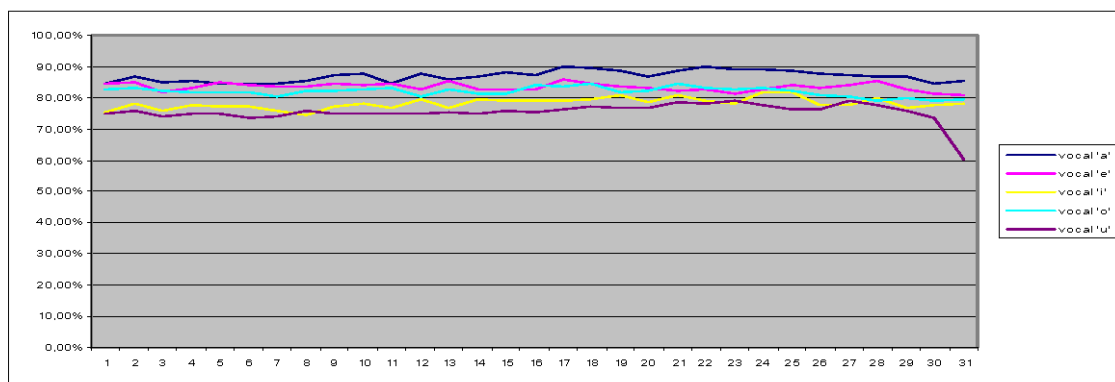


Gráfico 9-5: Evaluación secuencial de la relevancia para las medidas de calidad de la voz del grupo IV usando el clasificador KNN en la clasificación entre voz sana y voz patológica

9.6.2 *Estudio: Sano – Disfonía Leve – Disfonía Moderada – Disfonía Severa*

9.6.2.1 Red Neuronal Probabilística (PNN)

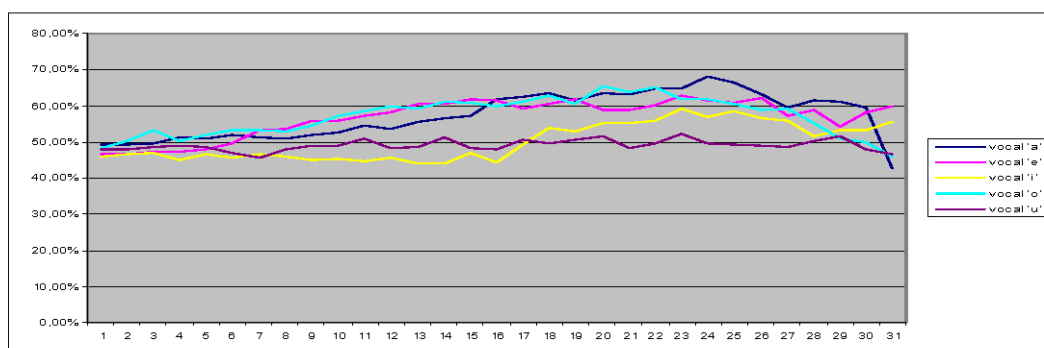


Gráfico 9-6: Evaluación secuencial de la relevancia para las medidas de calidad de la voz del grupo IV usando el clasificador PNN en la clasificación entre diferentes grados de disfonía

9.6.2.2 Red Neuronal FeedForward con 1 Capa Oculta (NN)

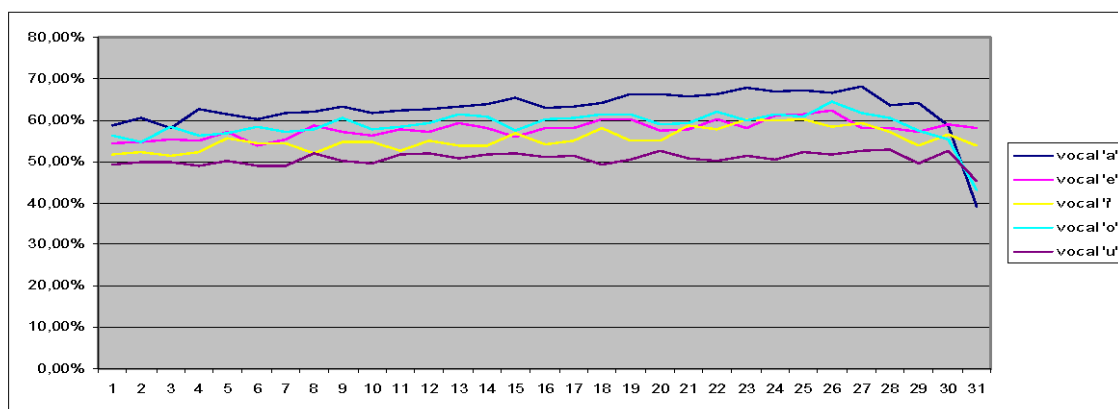


Gráfico 9-7: Evaluación secuencial de la relevancia para las medidas de calidad de la voz del grupo IV usando el clasificador NN en la clasificación entre diferentes grados de disfonía

9.6.2.3 Función de Base Radial (RBF)

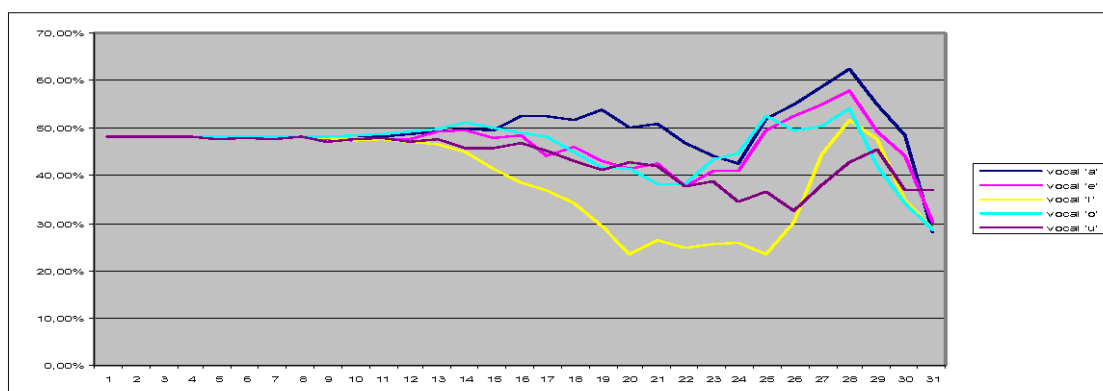


Gráfico 9-8: Evaluación secuencial de la relevancia para las medidas de calidad de la voz del grupo IV usando el clasificador PNN en la clasificación entre diferentes grados de disfonía

9.6.2.4 Máquinas de Vectores Soporte (SVM)

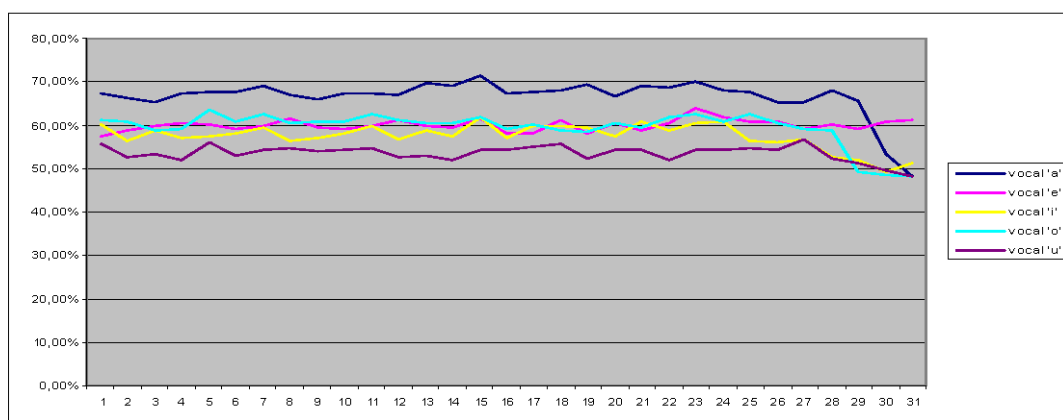


Gráfico 9-9: Evaluación secuencial de la relevancia para las medidas de calidad de la voz del grupo IV usando el clasificador PNN en la clasificación entre diferentes grados de disfonía

9.6.2.5 K vecinos más cercanos (KNN)

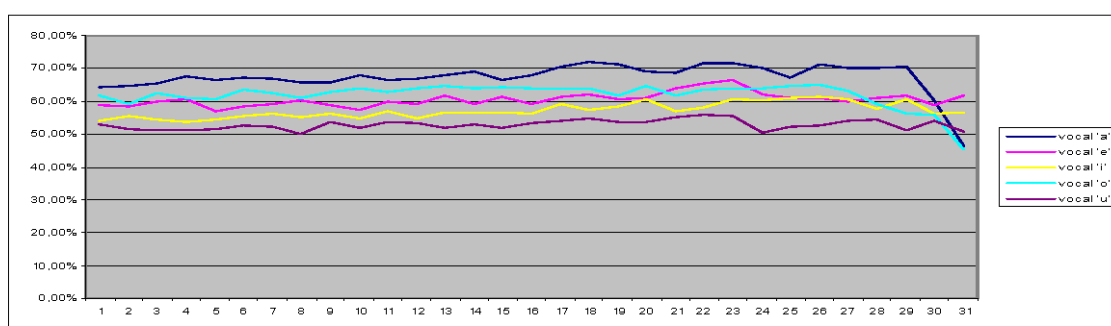


Gráfico 9-10: Evaluación secuencial de la relevancia para las medidas de calidad de la voz del grupo IV usando el clasificador KNN en la clasificación entre diferentes grados de disfonía

9.7 Evaluación

Este grupo de medidas presenta un comportamiento muy similar a las medidas que cuantifican el comportamiento no lineal.

En la clasificación entre voz sana y voz patológica se observa que para la PNN existe un gran incremento de la tasa de éxito a medida que se incrementa el número de medidas de calidad, estando acotado el rango entre el 50 y 85%. Para el RBF existe igualmente gran inestabilidad de la tasa de éxito, estando acotado el éxito entre el 50 y 85%. El resto de los clasificadores, el NN, SVM y KNN, se muestran con una tasa de

éxito totalmente estable a medida que se incrementa el número de medidas de calidad, estando el rango de éxito acotado entre el 75 y el 90%.

En el caso de la clasificación entre voz sana, disfonía leve, disfonía moderada y disfonía severa, se observa que para la PNN existe un significativo incremento de la tasa de éxito, acotada entre el 50-65%. En la RBF, existe significativas oscilaciones que van desde el 25% hasta el 60%. Sin embargo, una vez más, el NN, SVM y el KNN, mantienen una tasa de éxito totalmente constante, entre el 50-70%, pese al incremento de medidas de calidad de la voz.

Se observa de forma general para los diferentes clasificadores y los distintos casos estudiados (clasificación entre voz sana y voz patológica, y clasificación entre voz sana, disfonía leve, disfonía moderada y disfonía severa) que la tasa de éxito en la clasificación no se incrementa de forma significativa cuando la medida de calidad que presenta la mayor relevancia se evalúa con otras medidas más, principalmente para NN, SVM y RBF. En la interpretación de los resultados es necesario tener en cuenta la reducida dimensión de la base de datos utilizada para la evaluación, lo cual puede justificar las fluctuaciones existentes.

Por tanto es posible asociar, para cada vocal por separado, la cuantificación del fenómeno físico “cuantificación del comportamiento no lineal” a una única medida, la cual se ha identificado como la medida que presenta mayor relevancia.

9.8 Referencias

- [1] Robert F. Orlikoff, R. J. Baken, "Curing Diagnosis: Improving the Taxonomy of Phonatory Dysfunction", *Sixth Conference on Advances in Quantitative Laryngology*. Hamburg, Germany, April 2003
- [2] Ping Yu, Maurice Ouaknine, Joana Revis, Antonine Giovanni, "Objective Voice Analysis for Dysphonic Patients: A Multiparametric Protocol Including Acoustic and Aerodynamic Measurements", *Journal of Voice*, vol.15, no. 4, pp.529-542, 2001.
- [3] Wilden, H. Herzel, G Peters, G. Tembrock, "Subharmonics, Biphonation and Deterministic Chaos in Mammal Vocalization", *Bioacoustics: The International Journal of Animal Sound and its Recording*, vol. 9, pp.171-196, 1998.
- [4] Jürgen Neubauer, Patrick Eysholdt, Ulrich Eysholdt, Hanspeter Herzel, "Spatio-temporal analysis of irregular vocal fold oscillations: Biphonation due to desynchronization of spatial models", *Journal of Acoustical Society of America*, vol.110, no.6. pp. 3179-3192, 2001
- [5] Xuejing Sun and Yi Xu, "Perceived Pitch of Synthesized Voice with Alternate Cycles", vol.16, no.4, pp.443-459, 1995.
- [6] Titze, IR., *Principles of Voice Production*, Englewood Cliffs, NJ: Prentice-Hall, Inc, 1994.
- [7] Svec JG, Schutte HK, Miller DG., "A subharmonic vibratory pattern in normal vocal folds", *Journal of Speech, Language, and Hearing Research*, vol.39, no.1, pp. 135-143, 1996
- [8] Nuñez Batalla F, Suarez Nieto, C., Muñoz Pinto, C., Baragaño Rio, L., Álvarez Zapico, M.J., Martínez Ferreras, A., "Estudio Espectrográfico de la disfonía: Subarmónicos", *Acta Otorrinolaring. Esp.*, vol.52, no.1, pp.52-56, 2000.
- [9] C. Michael Haben, Karen Kost, George Papagiannis, "Lateral Phase Mucosal Asymmetries in the Clinical Voice Laboratory", *Journal of Voice*, vol.17, no.1, pp.3-11, 2003
- [10] Stephane Ayache, Maurice Ouaknine, Philippe Dejonkere, Pierre Prindere, Antoine Giovanni, "Experimental Study of the Effects of Surface Mucus Viscosity on the Glottic Cycle", *Journal of Voice*, vol.18, no.1, 2004
- [11] Boyan Boyanov, Stefan Hadjitodorov y Ognian Krumov, "Analysis of Voiced Speech by means of Bispectrum", *Communications and Signal Processing*, COMSIG-90, pp.2267-2268, 1990
- [12] J. W. A. Fackrell y S. McLaughlin, "The Higher-Order Statistics of Speech Signals", *Techniques for Speech Processing and their Application*, pp.711-716, 1994
- [13] Stanley Wenndt y Sanyogita Shamsunder, "Bispectrum Features for Robust Speaker Identification", *Acoustics, Speech and Signal Processing*, ICASSP-97, vol.2, pp.1095-1098, 1997
- [14] Ralph Fulchiero y Andreas S. Spanias, "Speech Enhancement Using the Bispectrum", *Acoustics, Speech and Signal Processing*, ICASSP-93, vol.2, pp.367-370, 1993
- [15] S. Seetharaman y M. E. Jernigan, "Speech Signal Reconstruction Based on Higher Order Spectra", *Acoustics, Speech and Signal Processing*, ICASSP-88, vol.1, pp.703-706, 1988
- [16] K. K. Paliwal y M. M. Sondhi, "Recognition of Noisy Speech Using Cumulant-Based Linear Prediction Analysis", *Acoustics, Speech and Signal Processing*, ICASSP-91, vol.1, pp.429-432, 1991

- [17] Asunción Moreno, Sergio Tortla, Josep Vidal y J.A.R. Fonollosa, "New Hos-Based Parameter estimation Methods for Speech Recognition in Noisy Environments", *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing*, ICASSP '95, vol.1, pp. 429-432, 1995
- [18] Asunción Moreno y Miquel Rutllán, "Integrated Polispectrum on Recognition", *Spoken Language*, ICSLP-96, vol.2, pp.1281-1284, 1996
- [19] Matthias Frohlich, Dirk Michaelis y Hans Werner Srube, "Acoustic 'Breathiness Measures' in the Description of Pathologic Voices", *Acoustics, Speech and Signal Processing*, ICASSP-98, vol.2, pp.937-940,1998
- [20] T. Koizumi, S. Taniguchi, and S. Hiromitsu, "Glottal source-vocal tract interaction," *Journal of the Acoustical Society of America*, November; vol.78, pp. 1541-1547, 1985
- [21] D. M. Brookes and P. A. Naylor, "Speech production modelling with variable glottal reflection coefficient," in *International Conference on Acoustics, Speech and Signal Processing*, pp.671 – 674, 1988.
- [22] H. M. Teager and S. M. Teager, "Evidence of nonlinear sound production mechanisms in the vocal tract," in *Proceedings of the NATO Advanced Study Institute on Speech Production and Modelling*, pp.241 – 261, 1989
- [23] Hansen, J.H.L.; Gavidia-Ceballos, L.; Kaiser, J.F. "A nonlinear operator-based speech feature analysis method with application to vocal fold pathology assessment". *IEEE Transactions on Biomedical Engineering*, vol.45, no. 3, pp. 300 – 313, 1998
- [24] Michael Banbrook, *Nonlinear analysis of speech from a synthesis perspective*, PhD Thesis, The University of Edinburgh, 1996.
- [25] Alonso, J. B.; de León, J.; Alonso, I.; Ferrer, M. A. "Automatic Detection of pathologies in the voice by Hos based parameters". *Eurasip journal on Applied signal procesing*, vol. 2001, no. 4, pp.275-284, 2001
- [26] Jerry M. Mendel, "Tutorial on Higher-Order statistics (Spectra) in Signal Processing and System Theory: Theoretical Results and Some Applications", in *Proceedings of the IEEE*, vol.79, no. 3, pp.278-305, 1991
- [27] Chrysostomos L. Nikias and Mysore R. Raghuveer, "Bispectrum estimation: A Digital Signal Processing Framework", in *Proceedings of the IEEE*, vol.75, no.7, pp.869-791, 1987
- [28] Steve Elgar and T. Guza, "Statistics of Bicoherence", *IEEE Transactions on Acoustics, Speech, and Signal Processing*, , vol.36, no.10, pp.1667-1668, 1988
- [29] Asunción Moreno, Sergio Tortla, Josep Vidal and J.A.R. Fonollosa, "New Hos-Based Parameter estimation Methods for Speech Recognition in Noisy Environments", *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing*, ICASSP'95, vol.1, pp. 429-432, 1995.
- [30] Asunción Moreno and Miquel Rutllán, "Integrated Polispectrum on Recognition", *Spoken Language*, ICSLP-96, vol.2, pp.1281-1284, 1996
- [31] Gopal Sundaramoorthy, M.R. Raghuveer and Soheil A. Dianat, "Bispectral Reconstruction of Signals in Noise: Amplitude Reconstruction Issues", *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol.38, no.7,1990
- [32] Liliana Gavidia-Ceballos and John H.L. Hansen, "Direct Speech Feature estimation Usiong an Iterative EM Algorithm for Vocal Fold Pathology Detection", *IEEE Transactions on Biomedical Engineering*, vol. 43, no. 4, 1996
- [33] Liliana Gavidia-Ceballos, John H.L. Hansen and John E. Riski, "A Non-invasive Speech Processing Method For Analysis of Vocal Fold Cancer Based on an Iterative Em Algorithm", *16th Annual International Conference of the IEEE of Engineering in Medicine and Biology Society*, vol.2, pp.1308-1309,1996

- [34] Liliana Gavidia-Ceballos, John H.L. Hansen y F. James F. Kaiser, "Vocal Fold Pathology Assessment using AM Autocorrelation Analysis of the Teager Energy Operator", *Spoken Language*, ICSLP-96, vol.2, pp.757-760, 1996
- [35] Shan Lu y Peter C. Doerschuk, "Demodulators for AM-FM models of Speech Signals: a Comparison", *Acoustics, Speech and Signal Processing*, ICASSP-96, vol.1, pp.263-266, 1996.
- [36] Shan Lu y Peter C. Doerschuk, "Nonlinear Modeling and Processing of Speech Based on Sums of AM-FM Formant Models", *IEEE Transactions on Signal Processing*, vol.44, no. 4, pp. 773-774, 1996
- [37] Douglas A. Cairns, John H. L. Hansen, "A Noninvasive Technique for Detecting Hypernasal Speech Using a Nonlinear Operator", *IEEE Transactions on Signal Processing*, vol.43, no.1, pp. 35-44, 1996
- [38] Douglas A. Cairns, John H.L. Hansen, y James F. Kaiser, "Recent Advances in Hypernasal Speech Detection Using the Nonlinear Teager Energy Operator", *Proceedings of International Conference on Spoken Language Processing*, ICSLP'96, vol.2, pp.780-783, 1996.
- [39] Guojun Zhou, John H.L. Hansen, y James F. Kaiser, "Classification of Speech Under Stress Based on Features Derived from the Nonlinear Teager Energy Operator", *Acoustics, Speech and Signal Processing*, ICASSP-98, vol.1, pp.549-552, 1998
- [40] Alan C. Bovik y Petros Maragos, "Conditions for Positivity of an Energy Operator", *IEEE Transactions on Signal Processing*, vol.42, no. 2, 1994
- [41] James F. Kaiser, "Some Useful Properties of Teager's Energy Operators", *Acoustics, Speech and Signal Processing*, ICASSP-93, vol.3, pp.149-152, 1993
- [42] Jerry M. Mendel, "Tutorial on Higher-Order statistics (Spectra) in Signal Processing and System Theory: Theoretical Results and Some Applications", in *Proceedings of the IEEE*, vol.79, no.3, pp. 278-305, 1991
- [43] Chrysostomos L. Nikias y Mysore R. Raghuvver, "Bispectrum estimation: A Digital Signal Processing Framework", in *Proceedings of the IEEE*, vol.75, no.7, pp. , 1987
- [44] Holger Kantz and Thomas Schreiber, *Nonlinear Time Series Analysis*, Cambridge Nonlinear Science Series 7, Cambridge University Press 1997.
- [45] Patricia Henríquez Rodríguez, *Aplicación de la Teoría Del Caos en la Detección de Patologías Laringeas*, PFC (directores: Miguel Ángel Ferrer Ballester y Jesús Bernardino Alonso Hernández), Escuela Técnica Superior de Ingenieros de Telecomunicación, Universidad de Las Palmas de Gran Canaria, Noviembre 2005.
- [46] Jesús B. Alonso, Fernando Díaz de María, Carlos M. Travieso, Miguel A. Ferrer, "Using Nonlinear features for voice disorder detection", in *3rd International Conference on Non-linear speech processing, Barcelona*, pp.94-106, 2005
- [47] R. Hegger, H. Kantz y T. Schreiber, "Practical implementation of nonlinear time series methods: The TISEAN package", *CHAOS* 9, 413 (1999) [<http://www.mpi-pks-dresden.mpg.de/~tisean/>]
- [48] Yu Zhang y Jiang J.J., "Nonlinear dynamic analysis in signal typing of pathological human voices", *Electronics Letters*, vol.39, no.13, pp.1021-1023, 2003
- [49] Jiang J.J. y Yu Zhang, "Nonlinear dynamic analysis of speech from pathological subjects", *Electronics Letters*, vol.38, no.6, pp.294-295, 2002
- [50] Accardo A., Fabbro F. y Mumolo E. , "Analysis of normal and pathological voices via short-time fractal dimension", *Proceedings of the Annual International Conference of the IEEE of Engineering in Medicine and Biology Society*, vol.14., no.3, pp.1270-1271, 1992.

- [51] Petros Maragos, Alexander G. Dimakis, y Iasonas Kokkinos, "Some advances in nonlinear speech modeling usig modulations, fractals, and chaos", 14th International Conference on Publication of Digital Signal Processing, DSP 2002, vol. 2002, no.1, pp.325- 332, 2002
- [52] D. G. Childers, *Speech Processing and Synthesis Toolboxes*, John Wiley & Sons, Inc. 2000.
- [53] K. Marasek, "EGG and Voice Quality", <http://www.ims.uni-stuttgart.de/phonetik/EGG/frmst1.htm> , 1997. [Consulta 10/10/05]
- [54] P. Gómez, V. Nieto, R. Martínez, A. Alvarez, F. Rodríguez, V. Rodellar, "Detection of the Mucosal Wave in the Speech Signal for Larynx Pathology Characterization", *Proceedings of the International Symposium on Information and Communication Technologies*, Ed. Trinity College Dublin, Ireland, pp 33–39, Sept. 2003,
- [55] P. Gómez, J. I. Godino, F. Rodríguez, F. Díaz, V. Nieto, A. Alvarez, V. Rodellar, "Evidence of vocal cord pathology form the mucosal wave cepstral contents", *Proceedings of International Conference on Acoustics, Speech, and Signal Processing (ICASSP'04)*, Montreal, Canada, pp.437-440, 2004.
- [56] P. Gómez, F. Díaz, R. Martínez, J. I. Godino, A. Alvarez, F. Rodríguez, V. Rodellar, "Precise Reconstruction of the mucosal wave for voice pathology detection and characterization", *Proceedings of the European Signal Processing Conference (EUSIPCO'04)*, pp. 2263-2266, Vienna, 2004.
- [57] C. Drioli, F. Avanzini, "Model based synthesis and transformation of voiced sounds", *COST-G6 Conference on Digital Audio Effects (DAFx-00)*, Verona, pp 45-48, December 2000.
- [58] C. Drioli, "A flow waveform adaptative mechanical glottal model", *Center for Speech Technology*, KTH, Stockholm, TMH-QPSR, Vol 43, pp 69-79, 2002
- [59] J.L. Flanagan, I.L. Landgraf, "Self-oscillating source for vocal tract synthesizers", *IEEE Transation of Audio and Electroacustics*, vol.16, pp. 57-64, 1968
- [60] K. Ishizaka, J. L. Flanagan, "Synthesis of voiced sounds from a two-mass model of the vocal cords", *Bell System Tech. J*, vol.51, no. 6, pp. 1233-1268, 1977.
- [61] I. R. Titze, "The human vocal cords: A mathematical model", Part 1. *Phonetica*, vol.28, pp. 129-170, 1973.
- [62] P. Gómez, J. I. Godino, A. Álvarez, R. Martínez, V. Nieto, V. Rodellar, "Evidence of Glottal Source Spectral Features found in Vocal Fold Dynamics", *International Conference on Acoustics, Speech and Signal Processing (ICASSP'05)*, 2005

CAPÍTULO 10

Aplicaciones

En los capítulos anteriores se han estudiado las diferentes medidas de calidad de la voz, presentando una ordenación en función de su relevación, y se ha identificado el tamaño óptimo del conjunto de características por medio de una evaluación secuencial frente a distintos clasificadores y para diferentes escenarios: voz sana – voz patológica y voz sana – disfonía leve – disfonía moderada – disfonía severa.

A partir de los resultados obtenidos es posible montar con mayor facilidad diferentes aplicaciones en el ámbito de la medida de la calidad de la voz. Las aplicaciones que mayor interés presentan son los sistemas de clasificación automáticos, en estos la voz es adquirida, convertida en un vector de medidas de calidad y posteriormente se evalúa en el clasificador dicho vector lo que nos permite evaluar la valía de la parametrización, ésto es, la valía de las medidas de calidad de la voz.

El hecho de haber propuesto un subconjunto reducido de medidas de calidad de la voz, basado en cuantificar cuatro fenómenos físicos por medio de una única medida

de calidad de la voz para cada fenómeno físico, permite, por un lado, dar una interpretación a los valores de las diferentes medidas de calidad de la voz, y por otro lado, optimizar el proceso de entrenamiento de los sistemas automáticos de clasificación orientados a medir la calidad de la voz.

En este apartado de la memoria se presentan varias aplicaciones en el contexto del reconocimiento automático de la calidad de la voz. Se presenta el estudio de tres aplicaciones: un estimador automático de la presencia de patologías en el sistema fonador a partir de la fonación de las cinco vocales castellanas, un estimador automático del grado de disfonía igualmente desde la fonación de las cinco vocales castellanas y un sistema automático para estimar la presencia de patologías en el sistema fonador a partir de habla continua adquirida a partir de un canal telefónico. Además, se presenta un estudio estadístico sobre los rangos de normalidad de los valores para las diferentes medidas que cuantifican los distintos fenómenos físicos implicados en la caracterización de la calidad de la voz.

10.1 Sistema de detección Automática de trastornos del sistema fonador

Los autores del trabajo [1] comparten la opinión de que resulta imposible, excepto en algunos casos identificar el trastorno del sistema fonador presente en un locutor usando únicamente una grabación de voz. Esta opinión es generalizada por varios autores.

El principal motivo es que, en determinados casos, las características acústicas de la fonación de dos locutores con diferentes trastornos del sistema fonador pueden coincidir, e incluso en una inspección visual de la laringe pueden existir dudas sobre la naturaleza del trastorno. Por otro lado, también es frecuente la coexistencia de más de una del sistema fonador.

Sin embargo, son comunes los trabajos dirigidos a identificar la existencia o no de un trastorno en el sistema fonador teniendo múltiples muestras de estos trabajos localizados en las referencias de los capítulos 6, 7, 8 y 9. Una característica común en estos trabajos es la de no existir una base de datos de voces común que permita comparar los resultados obtenidos por los diferentes autores.

La primera aplicación que se presenta, es un sistema de detección automático de lapresencia de trastornos del sistema fonador, sigue el esquema general de un reconocedor de voz el cual se muestra en la Figura 10-1.

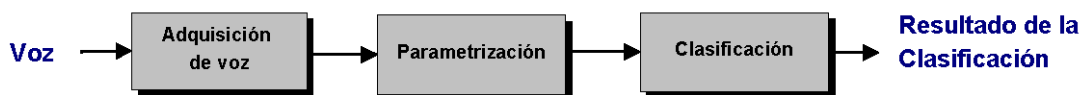


Figura 10-1 :Esquema general de sistema del detección automático de trastornos del sistema fonador

El bloque “adquisición de voz” tiene la finalidad de digitalizar la señal acústica de voz de un locutor, utilizándose como fonación en el sistema implementado las cinco vocales castellanas de forma sostenida. Además de digitalizar la señal de audio, en este bloque se suele realizar una discriminación entre la parte de la señal de audio que corresponde a voz y la parte de la señal de audio que corresponde a ruido. Este proceso es necesario debido a que en la mayoría de las aplicaciones de reconocimiento de voz, la señal que corresponde al audio entre distintos golpes de voz pertenece a ruido y su estudio no resulta útil. Además, en este bloque se realiza una segmentación de la voz en diversas tramas cuya duración se ha fijado a 3 periodos de *pitch* en el caso de la fonación de un sonido sonoro o de 30 milisegundos en el caso de una fonación sorda.

En el bloque “parametrización” se realiza una cuantificación de la calidad de la voz por medio de las diferentes medidas de calidad de las distintas tramas en las que se dividió la señal de voz. Esta cuantificación nos debe permitir descubrir características diferenciales entre las distintas unidades de clasificación, en nuestro caso o voz sana o voz patológica. En este bloque se convierte cada trama de voz en un vector de parámetros (un vector de medidas). Para la estimación de las medidas en el dominio espectral cada trama es enventanada usando una ventana *hamming*.

Una vez que se dispone de un vector de medidas para cada una de las tramas en la que se dividió la fonación, se estima un único vector de medidas por fonación. Para ello existen medidas de calidad que, bien promedian determinadas cuantificaciones de una determinada característica acústica, o bien evalúan su evolución temporal a lo largo de la fonación.

En el bloque “clasificación” se realiza una clasificación automática del vector de medidas entre las distintas unidades de clasificación. Es decir, a partir de una evaluación del vector de medidas obtenido en el bloque “parametrización” se realiza la identificación de la unidad de clasificación correspondiente donde en el caso que nos ocupa, cada vector de medidas es clasificado o bien en voz sana o bien en voz patológica.

10.1.1 Adquisición de Voz

El audio utilizado en el sistema de detección automático de trastornos del sistema fonador consiste en la grabación de las cinco vocales sostenidas, de aproximadamente 2,5 segundos de duración cada vocal, con pausas entre vocales de aproximadamente 2 segundos.. El audio es adquirido con una frecuencia de muestreo de 22050 Hz y con una resolución de 16 bits por muestra.

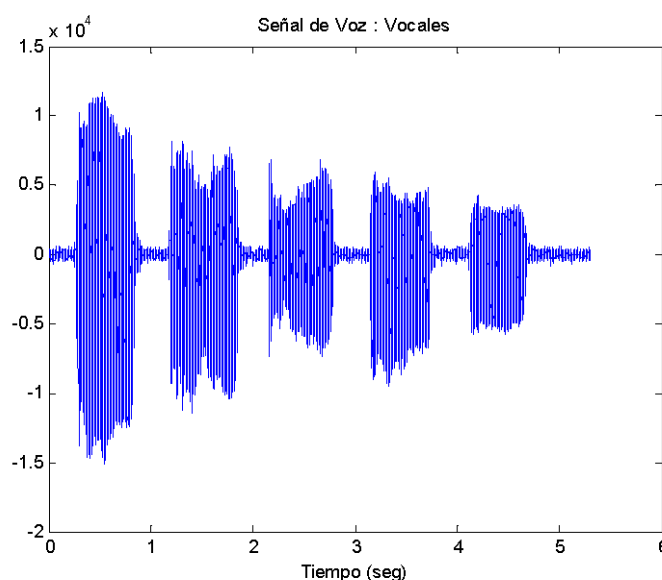


Figura 10-2: Señal de voz registrada correspondiente a la fonación de cinco vocales de forma sostenida

Para la fase de entrenamiento y evaluación del sistema se ha utilizado la base de datos creada para este trabajo y descrita en el capítulo 4.

10.1.2 Parametrización

Han sido analizados los resultados obtenidos en los capítulos 6, 7, 8 y 9, principalmente los resultados de la evaluación secuencial de las medidas de la calidad de la voz, es decir, se evalúan las medidas de calidad en un determinado clasificador incrementando el número de medidas a evaluar en función de una ordenación previa atendiendo a la relevancia de las medidas.

Se ha observado, de forma general, que la tasa de éxito en la clasificación no se incrementa significativamente con el número de atributos, pese a que existe una cierta variabilidad en la tasa de éxito, esto ha sido interpretado en este trabajo como una consecuencia de la proporción entre el número de observaciones y los posibles órdenes de magnitud del espacio de características.

Como consecuencia de esta observación se ha decidido utilizar, para la fase de parametrización, sólo el atributo más significativo de cada uno de los diferentes grupos de medidas de calidad de la voz, para cada una de las respectivas vocales. Por lo tanto cada vocal se parametriza por medio de cuatro medidas de calidad de la voz, es decir una de cada grupo de medidas.

TABLA 10-1: MEDIDAS DE CALIDAD DE LA VOZ UTILIZADAS EN LA PARAMETRIZACIÓN DEL SISTEMA DE DETECCIÓN AUTOMÁTICO DE TRASTORNOS DEL SISTEMA FONADOR

Vocal 'a'	<i>Grupo I</i>	(4) Shimmer versión 4 (APQ)
	<i>Grupo II</i>	(19) media Valor Cepstral del primer pico de los cepstrum derivados
	<i>Grupo III</i>	(39) media Relación Espectral Subarmónico - armónico de la señal residual
	<i>Grupo IV</i>	(15) media Área inferior del biespectro integrado
Vocal 'e'	<i>Grupo I</i>	(4) Shimmer versión 4 (APQ)
	<i>Grupo II</i>	(19) media Valor Cepstral del primer pico de los cepstrum derivados
	<i>Grupo III</i>	(3) media energía concentrada alrededor del 3er formante
	<i>Grupo IV</i>	(13) media Área superior del biespectro integrado
Vocal 'i'	<i>Grupo I</i>	(3) Shimmer versión 3 (APQ)
	<i>Grupo II</i>	(7) media Relación Espectral Subarmónico -armónico
	<i>Grupo III</i>	(3) media energía concentrada alrededor del 3er formante
	<i>Grupo IV</i>	(15) media Área inferior del biespectro integrado
Vocal 'o'	<i>Grupo I</i>	(18) Variación kurtosis de las diferentes tramas
	<i>Grupo II</i>	(49) media Skewness de la señal residual
	<i>Grupo III</i>	(39) media Relación Espectral Subarmónico - armónico de la señal residual
	<i>Grupo IV</i>	(13) media Área superior del biespectro integrado
Vocal 'u'	<i>Grupo I</i>	(5) Shimmer versión 5 (Fatr)
	<i>Grupo II</i>	(2) Jitter versión 2
	<i>Grupo III</i>	(39) media Relación Espectral Subarmónico - armónico de la señal residual
	<i>Grupo IV</i>	(27) media Medida de la entropía (H1) para m=2

10.1.3 Clasificación

En la fase de clasificación se ha utilizado una red de clasificadores constituida por cinco clasificadores donde cada clasificador evalúa la fonación correspondiente a una determinada vocal. A cada clasificador lo hemos denominado “clasificador vocal”. Los resultados de los cinco clasificadores son evaluados en una lógica en la cual se obtiene el resultado final de la clasificación.

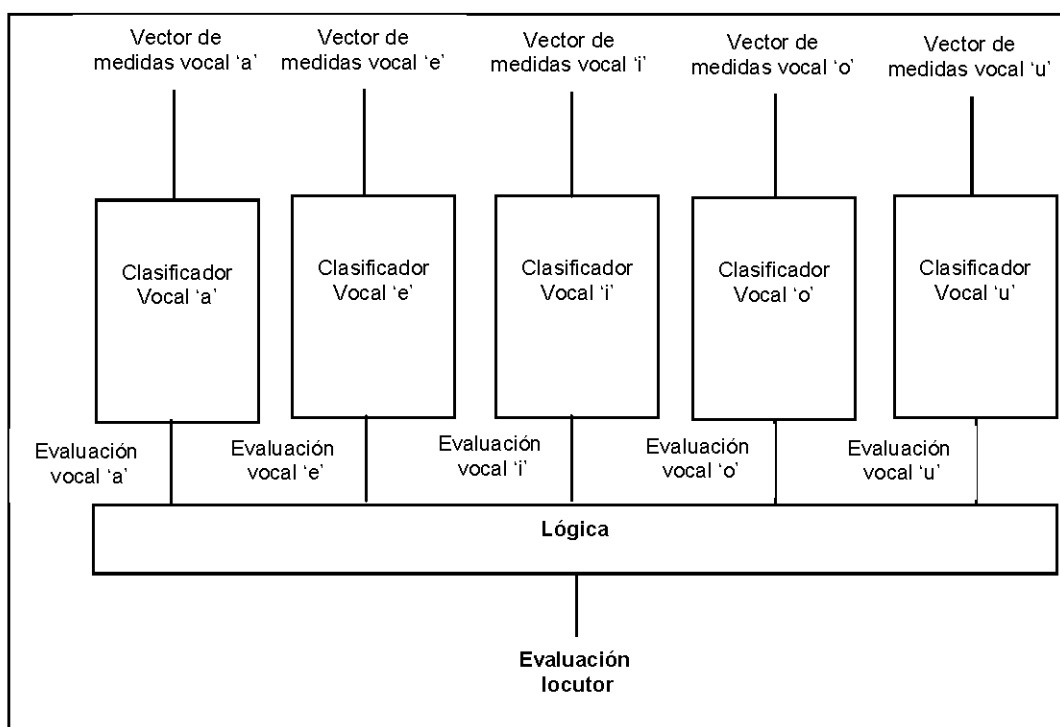


Figura 10-3: Esquema de la red de clasificadores del sistema de detección automático de trastornos del sistema fonador

Para la implementación de los clasificadores vocales se han estudiado diferentes sistemas de clasificación como son: redes neuronales probabilísticas (PNN), redes neuronales *feedforward* (NN), clasificadores basados en funciones de base radial (RBF), máquinas de vectores soporte (SVM) y el clasificador de los k vecinos más cercanos (KNN).

Para la implementación de la “lógica” se han evaluado diferentes estrategias de combinación de clasificadores optándose por un sistema basado en el más votado. Esto es, cada clasificador evalúa su correspondiente vocal entre “vocal sana” o “vocal patológica”, y posteriormente el “lógica” etiqueta la fonación como “sana” si existe

mayor número de vocales “sanas” y en caso contrario la fonación es etiquetada como “patológica”.

Las evaluaciones de los clasificadores son repetidas 20 veces y posteriormente se promedian. Las tasas medias de éxito obtenidas en las diferentes simulaciones son presentadas en la Tabla 10-2,

TABLA 10-2: TASAS DE ÉXITO (%) EN LA CLASIFICACIÓN OBTENIDAS CON DIFERENTES CLASIFICADORES EN EL SISTEMA DE DETECCIÓN AUTOMÁTICO DE TRASTORNOS DEL SISTEMA FONADOR USANDO LA BASE DE DATOS PROPIA

Clasificador		Clasificador vocal ‘a’	Clasificador vocal ‘e’	Clasificador vocal ‘i’	Clasificador vocal ‘o’	Clasificador vocal ‘u’	Tasa Final
PNN	Sin PCA	80.95	83.02	72.79	81.79	66.93	88.72
	Con PCA	80.59	84.79	74.41	79.94	65.10	88.97
NN	Sin PCA	88.21	83.78	80.42	86.84	69.24	89.55
	Con PCA	88.69	84.60	80.36	86.63	65.71	89.89
RBF	Sin PCA	86.94	84.01	73.80	81.38	58.99	87.96
	Con PCA	77.27	74.22	69.82	72.50	60.99	83.02
SVM	Sin PCA	84.34	81.47	80.41	86.66	70.98	90.05
	Con PCA	73.13	86.69	79.83	84.20	64.57	87.04
KNN	Sin PCA	84.66	81.81	75.16	84.20	68.14	87.64
	Con PCA	84.20	80.97	73.79	82.52	61.37	87.02

10.1.4 Resultados

El clasificador vocal que mejor resultados globales presenta en la detección automática de trastornos del sistema fonador es el SVM, con una tasa media de éxito en la clasificación del 90,05%.

Las matrices de confusión obtenidas en los distintos sistemas de clasificación son mostradas a continuación, mostrándose los resultados en porcentajes de éxito en la clasificación.

NN

Sin transformación del espacio de características						Con transformación del espacio de características (PCA)					
Salidas	Entradas					Salidas	Entradas				
		Sano	Disf. Leve	Disf. Moderada	Disf. Severa			Sano	Disf. Leve	Disf. Moderada	Disf. Severa
	Sano	91.00	20.44	13.82	1.43		Sano	88.78	17.80	9.97	1.33
	Patológico	9.00	79.56	86.18	98.57		Patológico	11.22	82.20	90.03	98.67

SVM

Sin transformación del espacio de características						Con transformación del espacio de características (PCA)					
Salidas	Entradas					Salidas	Entradas				
		Sano	Disf. Leve	Disf. Moderada	Disf. Severa			Sano	Disf. Leve	Disf. Moderada	Disf. Severa
	Sano	90.50	21.53	11.07	1.24		Sano	85.06	18.83	12.96	2.49
	Patológico	9.50	78.47	88.93	98.76		Patológico	14.94	81.17	87.04	97.51

RFB

Sin transformación del espacio de características						Con transformación del espacio de características (PCA)					
Salidas	Entradas					Salidas	Entradas				
		Sano	Disf. Leve	Disf. Moderada	Disf. Severa			Sano	Disf. Leve	Disf. Moderada	Disf. Severa
	Sano	96.11	35.56	21.83	3.38		Sano	86.25	35.18	19.03	6.96
	Patológico	3.89	64.44	78.17	96.62		Patológico	13.75	64.82	80.97	93.04

KNN

Sin transformación del espacio de características						Con transformación del espacio de características (PCA)					
Salidas	Entradas					Salidas	Entradas				
		Sano	Disf. Leve	Disf. Moderada	Disf. Severa			Sano	Disf. Leve	Disf. Moderada	Disf. Severa
	Sano	90.45	19.93	22.05	2.55		Sano	88.77	25.57	17.46	1.67
	Patológico	9.55	80.07	77.95	97.45		Patológico	11.23	74.43	82.54	98.33

PNN

Sin transformación del espacio de características						Con transformación del espacio de características (PCA)					
Salidas	Entradas					Salidas	Entradas				
		Sano	Disf. Leve	Disf. Moderada	Disf. Severa			Sano	Disf. Leve	Disf. Moderada	Disf. Severa
	Sano	90.75	22.90	14.11	2.83		Sano	91.09	25.62	13.48	1.29
	Patológico	9.25	77.10	85.89	97.17		Patológico	8.91	74.38	86.52	98.71

10.1.5 Test del sistema

En los apartados anteriores, partiendo de una base de datos se ha buscado la parametrización y el sistema de clasificación que optimizan la tasa de éxito. Con el fin de validar los resultados en la fase de parametrización, hemos evaluado el sistema usando otra base de datos y optimizando nuevamente el sistema de clasificación, sin modificar la parametrización.

Como base de datos hemos utilizado la base de datos *Voice Disorders Database Model 4337* descrita en el capítulo 4, con ciertas modificaciones. La fonación utilizada es la vocal ‘a’ de forma sostenida. La base de datos original presentaba 657 muestras de voces patológicas y 53 de voces sanas. En el caso de intentar entrenar un sistema de reconocimiento automático donde la cantidad de muestras utilizadas de cada clase es diferente produciría, por ejemplo en un problema biclásico, que el clasificador no generalizase bien en el aprendizaje y mostrase cierta tendencia a etiquetar erróneamente una clase, la que menor número de muestras presenta, frente a la otra. Con el fin de aumentar el número de muestras a utilizar en el entrenamiento, se han añadido 85 muestras más de voz sana. Con esta modificación se obtiene un balance final de la base de datos de 657 muestras de voces patológicas y 138 muestras de voces sanas.

En la fase de entrenamiento se ha dividido la base de datos de la siguiente manera: se ha utilizado un 50% de las muestras de voces sanas para el entrenamiento y el 50% de las muestras restantes para el test, mientras que respecto a las muestras de voces

patológicas se ha utilizado en la fase de entrenamiento el mismo número de muestras patológicas que muestras sanas, utilizando para la fase de test el resto de las muestras.

Se han utilizado diferentes sistemas de clasificación como son: PNN, NN, RBF, SVM y KNN. La evaluación obtenida de cada uno de los diferentes sistemas de clasificación es fruto de promediar los resultados de 20 evaluaciones.

Los resultados obtenidos se muestran a continuación.

TABLA 10-3: TASAS DE ÉXITO (%) EN LA CLASIFICACIÓN OBTENIDAS CON DIFERENTES CLASIFICADORES EN EL SISTEMA DE DETECCIÓN AUTOMÁTICO DE TRASTORNOS DEL SISTEMA FONADOR USANDO LA BASE DE DATOS *VOICE DISORDERS DATABASE MODEL 4337*

Clasificador		Clasificador vocal 'a'
PNN	Sin PCA	78.21
	Con PCA	79.31
NN	Sin PCA	84.36
	Con PCA	83.65
RBF	Sin PCA	84.53
	Con PCA	76.93
SVM	Sin PCA	85.31
	Con PCA	83.78
KNN	Sin PCA	85.38
	Con PCA	83.43

Las matrices de confusión obtenidas en los distintos sistemas de clasificación son mostradas a continuación, mostrándose los resultados en porcentajes de éxito en la clasificación.

NN

Sin transformación del espacio de características			
Entradas			
Salidas		Sano	Patológica
	Sano	86.52	17.79

Con transformación del espacio de características (PCA)			
Entradas			
Salidas		Sano	Patológica
	Sano	87.90	20.60

SVM

Sin transformación del espacio de características			
Salidas	Entradas		
		Sano	Patológica
	Sano	91.01	20.38
	Patológico	8.99	79.62

Con transformación del espacio de características (PCA)			
Salidas	Entradas		
		Sano	Patológica
	Sano	87.25	19.68
	Patológico	12.75	80.32

RFB

Sin transformación del espacio de características			
Salidas	Entradas		
		Sano	Patológica
	Sano	93.04	23.97
	Patológico	6.96	76.03

Con transformación del espacio de características (PCA)			
Salidas	Entradas		
		Sano	Patológica
	Sano	86.74	32.88
	Patológico	13.26	67.12

KNN

Sin transformación del espacio de características			
Salidas	Entradas		
		Sano	Patológica
	Sano	89.71	18.94
	Patológico	10.29	81.06

Con transformación del espacio de características (PCA)			
Salidas	Entradas		
		Sano	Patológica
	Sano	87.97	21.10
	Patológico	12.03	78.90

PNN

Sin transformación del espacio de características			
Salidas	Entradas		
		Sano	Patológica
	Sano	77.68	21.24
	Patológico	22.32	78.76

Con transformación del espacio de características (PCA)			
Salidas	Entradas		
		Sano	Patológica
	Sano	80.43	21.80
	Patológico	19.57	78.20

Por lo tanto, a tenor de los resultados se observa que no existen grandes variaciones en las tasas de éxito obtenidas utilizando la base de datos *Voice Disorders Database Model 4337* y la base de datos propias, descritas en el capítulo 4, pese a la diferencia significativa en el número de muestras de cada una de ellas.

10.2 Cuantificador del Grado de Disfonía

Con objeto de valorar la calidad de la voz, la Sociedad Otorrinolaringológica Japonesa (Comité para las Pruebas de la Función Fonatoria) propuso un sistema de valoración acústica subjetiva que se conoce con el nombre del acrónimo GRABS, que corresponde, en inglés, a las palabras: G - Grade (grado), R - Rough (rasposa), A - Astenic (asténica), B - Breathy (Aérea), S - Strain (constreñida). Cada uno de los apartados se valora en una escala de 4 puntos (0 = normal, 1 = ligero, 2 = moderado, 3 = extremo).

El atributo G indica el grado general de afectación de la voz; el 0 correspondería a una voz normal; el 3 correspondería a una voz muy patológica. *El atributo R* indica si la voz es rasposa, ronca o con rozamiento. Se emplea para expresar la irregularidad o defecto de la vibración (rozamiento), e incluso la ausencia de vibración. *El atributo A* sería el grado de astenia o fatiga y debilidad de la voz. Correspondería a la incapacidad para fonar durante periodos largos de tiempo sin cambiar el timbre vocal. Cada persona tiene su límite, pero oscila entre 80 y 120 minutos de habla continua. El tono se hace más grave y la dicción es más monótona, perdiendo flexibilidad. Tiene su origen en el uso prolongado de la voz y en ciertas enfermedades neurológicas como la miastenia, esclerosis múltiple, ELA o Parkinson. *El atributo B* correspondería a la sensación de aire en la voz, voz aérea o voz soplada, que se produce por escape de aire entre las cuerdas vocales, sea cual fuese la causa que la determina. Si por alguna circunstancia no hay cierre glótico, parte del aire espirado se pierde de forma turbulenta entre las cuerdas, lo cual percibimos como voz aérea. *El atributo S* lo tienen las voces constreñidas, tensas o duras. Esta interpretación psicoacústica corresponde al fenómeno de la hiperfonación o tensión excesiva de la laringe. La contracción excesiva es el origen de muchas lesiones de las cuerdas vocales, pero también la consecuencia porque el paciente trata de superar el efecto de la lesión presionando excesivamente ambas hemilaringes.

Los atributos R y B se asocian en mayor grado a lesiones orgánicas en donde predomina la disminución de la vibración (R) y el defecto de cierre (B), mientras que los atributos A y S se asocian más, en principio, a lesiones funcionales, en donde encontramos cansancio vocal (A) y emisión hiperfónica (S).

Hasta la actualidad, son muy pocas las experiencias que han abordado la tarea de obtener valoraciones de este tipo mediante sistemas automáticos. La ventaja de automatizar este tipo de valoraciones estriba en la independencia del criterio particular del especialista, de su entrenamiento, de su cansancio, de su estrés, etc.

El *National Center for Voice and Speech* (www.ncvs.org) recomendó que la evaluación de la voz patológica debe comenzar con la realización de un análisis espectrográfico para clasificarla con el fin de determinar los métodos de estudio más idóneos para cada caso en particular. Así, una voz con el suficiente grado de periodicidad (tipo 1) debería ser analizada por medio de los parámetros de perturbación a corto plazo (jitter, shimmer, NNE y HNR). Si la voz presenta aperiodicidad, subarmónicos y roturas de voz (tipo 2) no puede ser analizada de forma fiable con los parámetros antes mencionados observándose que una perturbación mayor del 5% nos indica que estamos ante una voz de este tipo la cual debe ser estudiada por un método de clasificación perceptual (GRBAS) y por un método visual como el espectrograma. Las voces caóticas (tipo 3) sólo pueden ser estudiadas por el momento, según el NCVS, mediante métodos perceptuales, en tanto no se apliquen de forma práctica las dimensiones que estudian los fenómenos caóticos (dimensión fractal, exponente de Lyapunov, etc). Por tanto, la clasificación GRBAS es un método imprescindible para la valoración de la voz patológica, dado que es necesario aplicarla en un importante número de pacientes en los cuales las medidas de la perturbación a corto plazo no son fiables o no se pueden calcular.

En el trabajo [4] , los autores han descrito los tipos de fonación según la siguiente clasificación:

- *Modal* o voz normal
- *Breathy Voice*: con gran cantidad de flujo de aire.
- *Creaky Voice*: con muy baja frecuencia y usualmente con irregularidades espaciadas en el tiempo.
- *Harsh voice*: con una frecuencia fundamental normal pero irregularidades o ruido en el espectro.
- *Tense* o *Strained voice*: con baja tasa de flujo de aire (a menudo se la describe como voz metálica).

Además, apuntan los autores que estos tipos de voces generan nuevos tipos fruto de la combinación de los tipos iniciales.

Ha sido propuesta una alternativa a la escala GRABS, la escala RHB (Roughness, Breathiness y Hoarseness), con cuatro niveles de cuantificación en su uso clínico y destacando su dependencia con el sexo del locutor. El mismo autor propone en [5] la dependencia de las medidas de la calidad de la voz con las vocales.

En la literatura actual existen diferentes trabajos que pretenden por medio de técnicas objetivas cuantificar la calidad de la voz. Para ello realizan un procesado digital de una grabación de voz emitiendo una valoración de la calidad de dicha voz. Un ejemplo de ello es el trabajo [1], donde se realiza una cuantificación de la calidad de la voz por medio de modelos ocultos de Markov (HMM), donde se evalúan tres escalas (*Roughness*, *Breathiness* y *Degree of Deviance*) cuantificadas en cinco niveles.

En [17] se propone el uso de diferentes medidas de calidad de voz, para locutores femeninos, basadas en el dominio espectral y temporal para realizar una clasificación del grado de la calidad de la voz identificando las siguientes unidades de clasificación: *Breathy*, *Hoarse*, and *Rough*. Para el análisis de las señales de voz sólo tienen en cuenta cinco medidas de la calidad de la voz: el logaritmo del *Shimmer* (LOGSHIM), la relación entre la amplitud del pico cepstral correspondiente al armónico principal obtenida por medio de la autocorrelación y obtenida por medio de una predicción lineal (CPP/EXP), relaciones de energía espectral (DFTR), el inverso de la raíz cuadrada de la frecuencia fundamental (INVSQRTFo), y el inverso de la raíz cuadrada del sigma *pitch* (INVSQRTSIG). En el trabajo citado se ha realizado un estudio de diferentes clasificadores obteniéndose diferentes tasas de éxitos:

- Normal vs. *Breathy*: con un modelo de dos variables (CPP/EXP y DFTR) se obtiene una correcta clasificación de 87.8%, destacando la contribución de la variación CPP/EXP.
- Normal vs. *Hoarse*: con un modelo de cuatro variables (LOGSHIM, CPP/EXP, DFTR y INVSQRTSIG) se obtiene una correcta clasificación de 97.4%, destacando la contribución de LOGSHIM.

- Normal vs. *Rouge*: con un modelo de cuatro variables (LOGSHIM, CPP/EXP, INVSQRTFo y INVSQRTSIG) se obtiene una correcta clasificación de 93.4%, destacando la contribución de LOGSHIM.
- Breathy vs. *Hoarse*: con un modelo de una sola variable (LOGSHIM) se obtiene una correcta clasificación de 84.5%.
- Breathy vs. *Rouge*: con un modelo de dos variables (LOGSHIM y INVSQRTFo) se obtiene una correcta clasificación de 78.6%, contribuyendo por igual ambas medidas.
- *Hoarse* vs. *Rouge*: con un modelo de tres variables (LOGSHIM, INVSQRTFo y DFTR) se obtiene una correcta clasificación de 88%, destacando la contribución de la variación DFTR.

Existen trabajos en la línea de relacionar medidas objetivas con medidas subjetivas, como en el caso de [2], donde se realiza un estudio de 6 medidas (medidas acústicas y medidas aerodinámicas) en las que se estudia la correlación de las medidas con la escala GRABS: destaca la correlación de las medidas aerodinámicas con las medidas acústicas, sin embargo no se tiene en cuenta los problemas de unicidad de la escala GRABS con distintos evaluadores.

De forma similar, en [20] se presentan los resultados de un trabajo en el que se investiga la correlación entre parámetros acústicos (Jitter, Shimmer, HNR, perturbación de Fo, presencia de subarmónicos y falta de armónicos) y la escala GRABS. Además se estudia la coincidencia intra e interobservador en el proceso de valoración de la calidad de la voz por medio de la escala GRABS.

Otro ejemplo es el trabajo [18], en el que se determina la correlación existente entre la escala subjetiva GRABS y la escala objetiva MDVP (*Multi-Dimensional Voice Program*) de *Kay Elemetric*. En el estudio estadístico realizado usando regresión multivariante se detectó una correlación entre las medias relativas al ruido y componentes de la escala GRABS. El *Roughness* está únicamente correlado con NHR, el *Breathiness* está sólo correlado con el índice de fonación suave SPI, y la *Aesthenia* sólo con SPI. Concluyendo entoces que el ruido puede ser la única característica acústica percible de la voces disfónicas.

Por último, se encuentra en la bibliografía trabajos sobre la interpretación de los valores de medidas objetivas, como en el caso de [21], donde se presenta un estudio sobre la correlación entre las medidas de la calidad de la voz calculadas por medio del sistema MDVP (*Multi-Dimensional Voice Program*) de la empresa KAY Elemetric. En este estudio se indican los valores de normalidad de las diferentes medidas.

10.2.1 Sistema automático de medida del grado de disfonía propuesto

El esquema del sistema automático propuesto para medida del grado de disfonía sigue un esquema idéntico al sistema de detección automático de trastornos del sistema fonador descrito en el apartado 10.1.

En líneas generales el esquema consiste en, a partir de una fonación de las cinco vocales castellanas, segmentar cada golpe de voz en tramas de duración de bien tres periodos de *pitch* en las tramas sonoras o bien 30 milisegundos en las tramas sordas. Cada trama es parametrizada, convirtiendo cada una de ellas en un vector de medidas de calidad de la voz. A partir de la secuencia de vectores de calidad de la voz se obtiene un vector promedio el cual es evaluado. Cada vector promedio correspondiente a una determinada vocal es evaluado por separado en un clasificador vocal, es decir, cada vocal es evaluada en un clasificador vocal diferente. Posteriormente, las salidas de los diferentes clasificadores son evaluadas en una lógica de control la cual emite una valoración sobre la calidad de la voz analizada.

En la fase de parametrización, se han analizado los resultados de la evaluación incremental de las medidas de calidad de la voz según su orden de relevancia para cada uno de los cuatro grupos, sobre diferentes clasificadores. Al igual que ocurría en el caso de la identificación automática de trastornos en la voz, se observa que a medida que aumenta el número de medidas de calidad de la voz no se observa un incremento significativo de la tasa de éxito en la clasificación para los diferentes clasificadores, observándose un gran variabilidad la cual es atribuida al relativo reducido número de observaciones y al incremento en la relación entre la dimensionalidad del espacio de características y el número de observaciones. Consecuencia de esta observación, es que en este trabajo se ha optado por utilizar en la parametrización la medida de la calidad de la voz con mayor relevancia de cada uno de los grupos de medidas de calidad. Las

medidas utilizadas en la parametrización para cada una de las diferentes vocales son descritas en la Tabla 10-4.

TABLA 10-4: MEDIDAS DE CALIDAD DE LA VOZ UTILIZADAS EN LA PARAMETRIZACIÓN DEL SISTEMA AUTOMÁTICO DE MEDIDA DEL GRADO DE DISFONÍA

Vocal 'a'	<i>Grupo I</i>	(5) Shimmer versión 5 (Fatr)
	<i>Grupo II</i>	(19) media Valor Cepstral del primer pico de los cepstrum derivados
	<i>Grupo III</i>	(39) media Relación Espectral Subarmónico - armónico de al señal residual
	<i>Grupo IV</i>	(11) media Perturbación del índice de bicoherencia
Vocal 'e'	<i>Grupo I</i>	(4) Shimmer versión 4 (APQ)
	<i>Grupo II</i>	(49) media Skewness de la señal residual
	<i>Grupo III</i>	(1) media energía concentrada alrededor del 1er formante
	<i>Grupo IV</i>	(13) media Área superior del biespectro integrado
Vocal 'i'	<i>Grupo I</i>	(1) Shimmer versión 1
	<i>Grupo II</i>	(7) media Relación Espectral Subarmónico -armónico
	<i>Grupo III</i>	(3) media energía concentrada alrededor del 3er formante
	<i>Grupo IV</i>	(15) media Área inferior del biespectro integrado
Vocal 'o'	<i>Grupo I</i>	(1) Shimmer versión 1
	<i>Grupo II</i>	(49) media Skewness de la señal residual
	<i>Grupo III</i>	(39) media Relación Espectral Subarmónico - armónico de al señal residual
	<i>Grupo IV</i>	(17) medida cuantificación del ruido estacionario (magnitud y fase)
Vocal 'u'	<i>Grupo I</i>	(4) Shimmer versión 4 (APQ)
	<i>Grupo II</i>	(2) Jitter versión 2
	<i>Grupo III</i>	(39) media Relación Espectral Subarmónico - armónico de al señal residual
	<i>Grupo IV</i>	(27) media Medida de la entropía (H1) para m=2

Las tasas medias de éxito obtenidas en las diferentes simulaciones, se muestran en la Tabla 10-5, donde se muestran los resultados de los diferentes clasificadores vocales y el resultado de la combinación de dichos resultados. Las evaluaciones de los clasificadores son obtenidas por medio del promediado de los resultados obtenidos tras 20 evaluaciones.

Se ha utilizado una política de combinación lineal, presentando los mejores resultados la política basada en la mayoría. Es decir, se etiqueta la muestra de voz con la clase más votada, y en caso de empate se aplica el criterio de máxima probabilidad. Este último criterio consiste en estimar cuál ha sido la clase más votada entre las diferentes tramas evaluadas.

TABLA 10-5: TASAS DE ÉXITO (%) EN LA CLASIFICACIÓN OBTENIDAS CON DIFERENTES CLASIFICADORES EN EL SISTEMA DE DETECCIÓN AUTOMÁTICO DE TRASTORNOS DEL SISTEMA FONADOR

Clasificador		Clasificador vocal 'a'	Clasificador vocal 'e'	Clasificador vocal 'i'	Clasificador vocal 'o'	Clasificador vocal 'u'	Tasa Final
PNN	Sin PCA	0.5717	0.6081	0.5323	0.5944	0.4817	0.6721
	Con PCA	0.5527	0.6160	0.5444	0.5837	0.4730	0.6524
NN	Sin PCA	0.6330	0.6454	0.5968	0.6221	0.5438	0.6784
	Con PCA	0.5939	0.6220	0.5741	0.5836	0.5489	0.6709
RBF	Sin PCA	0.4459	0.4742	0.5061	0.4431	0.3023	0.4984
	Con PCA	0.4701	0.3773	0.4145	0.4467	0.2406	0.4030
SVM	Sin PCA	0.5780	0.6186	0.5619	0.5910	0.5480	0.6191
	Con PCA	0.5878	0.6253	0.5659	0.5918	0.5243	0.6031
KNN	Sin PCA	0.6071	0.6103	0.5294	0.5925	0.4660	0.6693
	Con PCA	0.5710	0.5935	0.5231	0.5806	0.4536	0.6588

Las matrices de confusión obtenidas con los diferentes clasificadores son mostradas a continuación, mostrándose los resultados en porcentajes de éxito en la clasificación.

NN

Sin transformación del espacio de características					
Salidas	Entradas				
		Sano	Disf. Leve	Disf. Moderada	Disf. Severa
	Sano	94.84	32.20	17.41	4.10
	Disfonia Leve	1.90	33.32	20.26	14.98
	Disfonia Moderada	1.86	22.87	45.00	36.65
	Disfonia Severa	1.40	11.62	17.33	44.26

Con transformación del espacio de características (PCA)					
Salidas	Entradas				
		Sano	Disf. Leve	Disf. Moderada	Disf. Severa
	Sano	94.98	34.94	21.69	5.39
	Disfonia Leve	1.63	29.71	14.38	11.53
	Disfonia Moderada	2.78	21.99	43.76	36.27
	Disfonia Severa	0.62	13.36	20.17	46.81

SVM

Sin transformación del espacio de características					
Salidas	Entradas				
		Sano	Disf. Leve	Disf. Moderada	Disf. Severa
	Sano	98.96	77.10	52.82	28.17
	Disfonia Leve	0.14	5.92	2.94	2.33
	Disfonia Moderada	0.15	8.91	27.17	25.03
	Disfonia Severa	0.74	8.07	17.07	44.47

Con transformación del espacio de características (PCA)					
Salidas	Entradas				
		Sano	Disf. Leve	Disf. Moderada	Disf. Severa
	Sano	99.13	75.73	59.24	33.63
	Disfonia Leve	0.14	6.98	3.75	2.42
	Disfonia Moderada	0.44	11.86	24.71	29.90
	Disfonia Severa	0.28	5.44	12.29	34.04

RBF

Sin transformación del espacio de características					
Salidas	Entradas				
		Sano	Disf. Leve	Disf. Moderada	Disf. Severa
	Sano	58.28	8.48	4.18	3.99
	Disfonia Leve	30.70	58.26	39.17	33.90
	Disfonia Moderada	8.61	20.98	35.54	28.36
	Disfonia Severa	2.41	12.28	21.11	33.75

Con transformación del espacio de características (PCA)					
Salidas	Entradas				
		Sano	Disf. Leve	Disf. Moderada	Disf. Severa
	Sano	40.05	5.63	3.45	3.25
	Disfonia Leve	38.85	61.34	45.17	36.62
	Disfonia Moderada	13.00	17.72	23.91	18.44
	Disfonia Severa	8.11	15.32	27.46	41.69

KNN

Sin transformación del espacio de características						Con transformación del espacio de características (PCA)					
Salidas	Entradas					Salidas	Entradas				
		Sano	Disf. Leve	Disf. Moderada	Disf. Severa			Sano	Disf. Leve	Disf. Moderada	Disf. Severa
	Sano	88.85	21.36	12.71	2.84		Sano	88.44	14.95	9.87	3.73
	Disfonia Leve	4.56	47.12	16.68	16.17		Disfonia Leve	5.03	45.80	22.68	18.28
	Disfonia Moderada	5.01	23.61	46.46	35.79		Disfonia Moderada	4.38	23.22	44.99	33.00
	Disfonia Severa	1.59	7.91	24.15	45.21		Disfonia Severa	2.15	16.04	22.46	44.98

PNN

Sin transformación del espacio de características						Con transformación del espacio de características (PCA)					
Salidas	Entradas					Salidas	Entradas				
		Sano	Disf. Leve	Disf. Moderada	Disf. Severa			Sano	Disf. Leve	Disf. Moderada	Disf. Severa
	Sano	91.20	29.13	12.48	2.79		Sano	89.22	25.15	14.77	3.45
	Disfonia Leve	2.54	39.77	20.99	17.85		Disfonia Leve	3.77	43.38	19.15	18.73
	Disfonia Moderada	5.34	23.00	48.50	32.30		Disfonia Moderada	5.37	22.22	43.01	34.24
	Disfonia Severa	0.92	8.11	18.03	47.05		Disfonia Severa	1.64	9.24	23.07	43.58

El clasificador vocal que mejor resultados globales presenta en la cuantificación automática de la calidad de la voz, según el grado de disfonía, es el de las redes neuronales, con una tasa media de éxito en la clasificación de 67.84%.

10.3 Sistema de detección automática de patologías del sistema fonador en habla continua

Partiendo de una grabación del teléfono de un locutor, por medio de las técnicas subjetivas basadas en la experiencia, es posible identificar las características del locutor como por ejemplo el origen lingüístico, el nivel cultural, y el uso repetido de ciertas expresiones lingüísticas.

En este apartado se presentan los resultados de aplicar las técnicas basadas en las medidas de calidad de la voz para caracterizar a un locutor [26] . Esta técnica es objetiva y se basa en el comportamiento del sistema del fonador.

En esta aplicación se presenta un estudio de un sistema que permite a partir de una grabación de voz obtenida a través de una comunicación telefónica obtener diferentes medidas de la calidad de la voz del locutor. Estas medidas objetivas de la voz, permiten cuantificar ciertas peculiaridades del sistema fonador del locutor sin necesidad de grabaciones previas.

A continuación se muestra el esquema del sistema propuesto.



Ilustración 10-1: Esquema de un sistema de medida de calidad de la voz por medio de voz telefónica

10.3.1 Medidas de la calidad de la voz

En el proceso de producción del habla se produce voz sorda y voz sonora. En las voces sonoras se produce vibración de las cuerdas vocales mientras que no existe vibración en la voz sorda.

Las medidas de calidad de la voz, identificadas en apartados anteriores, que son utilizadas en los entornos médicos como otorrinolaringología, foniatría y logopedia, son estudiadas sobre sonidos sonoros sostenidos, vocales principalmente. Ésto es debido a que durante la fonación de un sonido se produce el movimiento de las cuerdas vocales, lo cual provoca que cualquier anomalía en las cuerdas vocales se refleje en la calidad de la fonación.

A continuación se muestra un fragmento de un monólogo telefónico de un locutor con alta calidad de voz y un locutor con baja calidad. Aparentemente en el

dominio del tiempo no se aprecian diferencias significativas y sin embargo en el dominio de la frecuencia se aprecia, entre otras características, una baja presencia de componente armónica en las señales de baja calidad de voz.

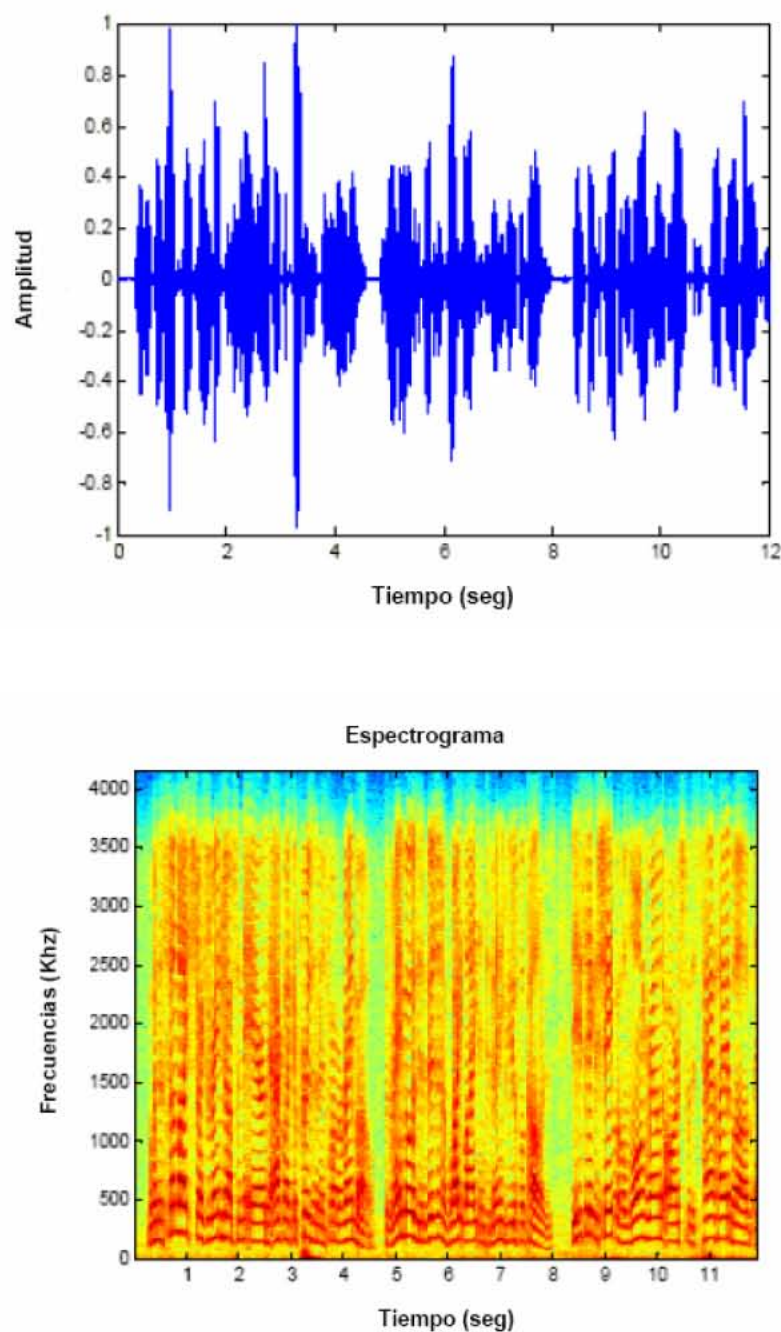


Figura 10-4: Muestra de habla de un locutor de alta calidad de voz

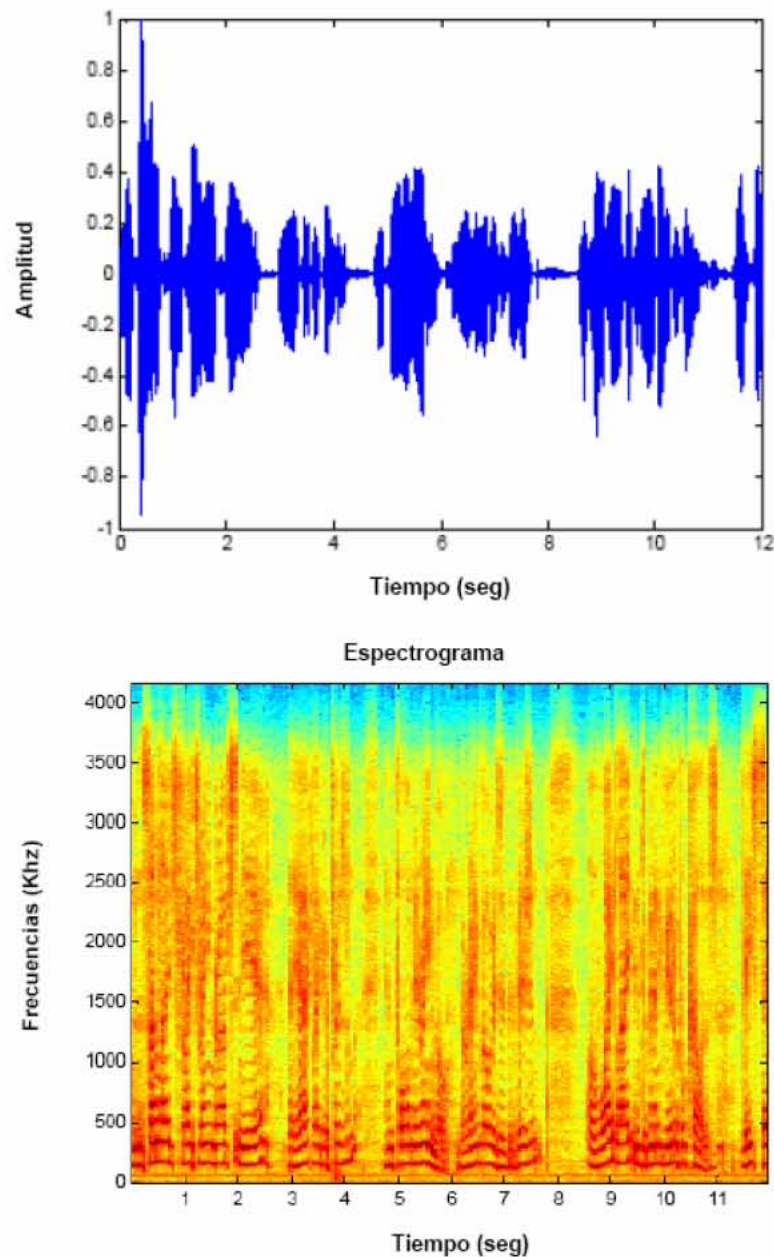


Figura 10-5: Muestra de habla de un locutor de baja calidad de voz

Para poder medir la calidad de la voz a partir de una muestra de conversación es necesario inicialmente identificar los diferentes golpes de voz donde posteriormente es necesario segmentar cada golpe de voz diferenciando entre voz sorda y voz sonora. A partir de cada segmento de voz sonora ha de tomarse una muestra de voz a partir de las cuales se realizará la medida de calidad de la voz. La muestra de voz sonora ha sido tomada manteniendo una guarda de 60 milisegundos para garantizar la transición entre voz sorda y voz sonora. Las muestras de voz utilizadas tienen una duración de 3

periodos de *pitch*, con el fin de que las medidas sean independientes del sexo del locutor. Posteriormente las medidas de calidad de voz obtenidas de las diferentes muestras de voz sonora son promediadas. En la Figura 10-6 se muestra un esquema general del proceso.

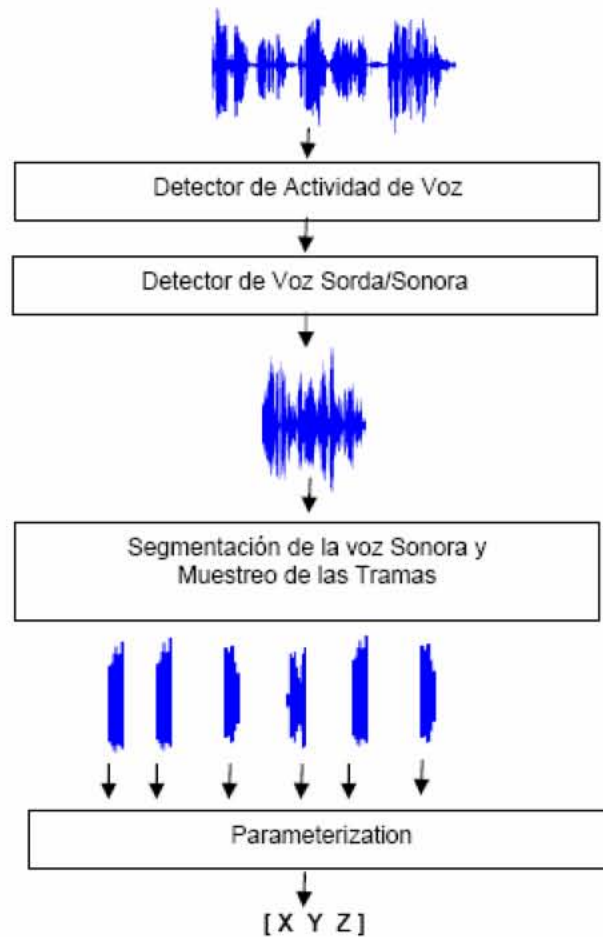


Figura 10-6: Esquema general de un sistema para la caracterización de la voz de un locutor a partir de habla continua

De entre las medidas existentes en los cuatro grupos de medidas de calidad de voz, no resulta coherente utilizar las medidas pertenecientes a tres de ellos. No resultan interesantes las medidas de estabilidad de voz debido a que en el proceso de habla continuo se diferencia un sonido sonoro sostenido en que existe una variación continua en el tiempo del tracto vocal y de la intensidad de la excitación glótica debido al proceso de articulación y entonación. Tampoco resulta interesante el grupo de medidas de calidad de voz que cuantifican el comportamiento no lineal, ya que al adquirir la señal de voz a través de la red telefónica, no existen garantías de la linealidad del canal. Por

otro lado, no resultan útiles las medidas de presencia de ruido debido al posible ruido aditivo del canal.

Por lo tanto se ha estudiado la utilidad de las medidas de la calidad de la voz que miden la riqueza espectral para caracterizar la voz de un locutor a partir de una comunicación telefónica. Las medidas utilizadas son presentadas en la Tabla 10-6.

TABLA 10-6: MEDIDAS DE CALIDAD ESTUDIADAS EN EL SISTEMA PARA LA CARACTERIZACIÓN DE LA VOZ DE UN LOCUTOR A PARTIR DE HABLA CONTINUA

Índice	Descripción
1	(11) Pico Cepstral Principal
2	(13) Energía Cepstral de Pitch
3	(15) Energía Cepstral del segundo armónico
4	(17) Valor Cepstral del primer pico de los cepstrum derivados
5	(19) Valor Cepstral del segundo pico de los cepstrum derivados
6	(21) Valor Cepstral de la diferencia entre el segundo y el primer pico de los cepstrum derivados
7	(52) Relación de la parte no periódica a la periódica
8	(50) Grado de Sonoridad por Tramas (GST)

No se han utilizado todas las medidas identificadas en el grupo de medidas que cuantifican la calidad de la voz, ya que para algunas de ellas el valor de la medidas depende de la variación de alguna característica a lo largo del tiempo, y no resulta coherente utilizarlas teniendo en cuenta que se trata de habla continua.

El la Tabla 10-7 se muestra la ordenación de las medidas de la calidad de la voz, de la Tabla 10-6, en función de su relevancia, donde se han ordenado las medidas de calidad de la voz de mayor a menor relevancia.

TABLA 10-7: ORDENACIÓN DE LAS MEDIDAS DE CALIDAD ESTUDIADAS EN EL SISTEMA PARA LA CARACTERIZACIÓN DE LA VOZ DE UN LOCUTOR A PARTIR DE HABLA CONTINUA EN FUNCIÓN DE SU RELEVANCIA

Índice	Descripción
1	(11) Pico Cepstral Principal
4	(17) Valor Cepstral del primer pico de los cepstrum derivados
6	(21) Valor Cepstral de la diferencia entre el segundo y el primer pico de los cepstrum derivados
8	(50) Grado de Sonoridad por Tramas (GST)
3	(15) Energía Cepstral del segundo armónico
2	(13) Energía Cepstral de Pitch
7	(52) Relación de la parte no periódica a la periódica
5	(19) Valor Cepstral del segundo pico de los cepstrum derivados

Con el fin de evaluar el sistema se ha implementado un sistema automático de etiquetado de voces sanas y voces patológicas.

Se ha utilizado la base de datos *Voice Disorders Database Model 4337*, desarrollada por el *Massachusetts Eye and Ear Infirmary Voice and Speech Lab*. Se han utilizado 50 grabaciones de una frase del pasaje *Rainbow*, 50 muestras de pacientes con una amplia variedad de patologías del sistema fonador de carácter orgánico, neurológico, traumático y psicológico, y 50 muestras de voces sanas. Las diferentes muestras de voz han sido previamente filtradas para simular las señales telefónicas (4KHz de ancho de banda).

El sistema ha sido evaluado de forma incremental, de tal forma que, comenzando con la característica más relevante, se va evaluando el sistema con diferentes clasificadores, y una vez evaluado, se incrementa añadiendo una medida de calidad de más en el vector de parametrización, volviendo a repetirse la evaluación en los diferentes clasificadores y repitiéndose este proceso hasta evaluar todas las medidas de calidad de la voz. El resultado se presenta en la Tabla 10-8.

TABLA 10-8: TASAS DE ÉXITO (%) CON EVALUCIÓN INCREMENTAL COMENZANDO CON LA CARACTERÍSTICA MÁS RELEVANTE EN EL PROBLEMA DE LA CUANTIFICACIÓN AUTOMÁTICA DE LA CALIDAD DE LA VOZ

Medidas Evaluadas	NN	RBF	PNN	SVM	KNNC
1	55.63	55.31	56.88	59.40	56.88
1 4	80.94	81.88	77.50	80.94	78.13
1 4 6	77.19	79.69	77.81	83.44	83.13
1 4 6 8	76.56	82.50	77.19	78.75	82.19
1 4 6 8 3	74.06	80.00	68.44	70.31	73.44
1 4 6 8 3 2	77.81	80.31	71.56	68.75	77.81
1 4 6 8 3 2 7	72.81	78.75	69.06	72.50	75.31
1 4 6 8 3 2 7 5	70.94	75.00	68.75	66.56	72.81

Analizando los resultados utilizando un clasificador SVM se obtiene una tasa de éxito en la clasificación automática de 83.44%. Esta tasa de éxito se obtiene únicamente con 3 medidas de calidad de la voz.

Por lo tanto, con tres medidas de calidad de la voz es posible caracterizar determinadas características de un locutor a partir de una grabación de voz obtenida de una línea telefónica. Por medio de estas medidas, es posible identificar características de un locutor telefónico, del que a priori no es conocida ninguna característica, permitiéndose obtener mayor información para tareas de identificación en diferentes ámbitos como por ejemplo para el forense en tareas policiales.

10.4 Protocolo de Evaluación

A continuación se presenta un protocolo de medidas de la calidad clínica de la voz a partir de grabaciones de audio. Este protocolo, pretende medir la calidad de la voz a partir de medidas objetivas, con el fin de eliminar la subjetividad que introduce una valoración auditiva de un profesional de la medicina.

La utilización de medidas objetivas no es una novedad en el ámbito clínico. A modo de introducción se presenta una revisión de algunas de las experiencias clínicas presentes en la bibliografía actual. Posteriormente se presenta el protocolo propuesto en esta tesis doctoral.

10.4.1 Revisión de Experiencias Clínicas

En este apartado se presentan los resultados de diferentes trabajos clínicos en los que se utilizan las medidas de calidad de la voz como herramientas para la evaluación de la voz frente a diferentes circunstancias como la aplicación de un determinado tratamiento o el sometimiento a una intervención quirúrgica, entre otras. A continuación se presentan los resultados de diferentes estudios clínicos.

Angeles Rovirosa y colaboradores [8] muestran los resultados de un interesante estudio en el que se evalúan los resultados de un tratamiento de radioterapia sobre cuerdas con carcinomas T1 para un grupo de pacientes de control. En el estudio se cuantifica por medio de medidas objetivas (valor de la frecuencia fundamental, *Jitter*, *Shimmer* y relación armónico ruido), la variación de la calidad de la voz respecto al estado de partida de los pacientes. En las conclusiones de este trabajo se analizan los resultados, observándose que para este tipo de tratamientos, determinadas medidas de calidad de la voz varían más que otras. Los pacientes tienen mayor frecuencia fundamental, *jitter*, *shimer* y relación armónico ruido y sin embargo sólo la frecuencia fundamental y el *jitter* presentan significativas diferencias frente al resto de las medidas en diferentes condiciones de fonación (sonidos sostenidos, lectura o canto). El *Shimmer* sólo muestra diferencias significativas durante la fonación de un sonido sostenido, mientras que la relación armónico ruido presenta cambios significativos durante la lectura oral y el canto.

Adriana Vélez Feijó y colaboradores [9] estudian el efecto de medidas objetivas de la calidad de la voz en un grupo de paciente con *Esclerosis Múltiple* (SM) y otro grupo de locutores sin patologías. El estudio concluye con el resultado de que la frecuencia fundamental es mayor para pacientes con SM, mostrándose una significativa varianza de la frecuencia principalmente para mujeres. Sin embargo el *Jitter* es mayor en pacientes de SM hombres.

Dana M. Hartl y colaboradores [10] han estudiado el efecto de las medidas objetivas de la calidad de la voz en pacientes con parálisis unilateral de cuerda vocal (UVFP) con el fin de eliminar la componente subjetiva. Se ha comparado un grupo de locutores con UVFP y un grupo de locutores sin la patología. En el estudio se indica que

tanto las medidas de *Jitter* y *Shimmer* están incrementadas en los locutores que presentan UVFP, mientras que la relación armónico ruido y la prominencia del pico cepstral principal disminuyen. Además, se aprecia un incremento en la energía espectral de media y alta frecuencia, a la vez, que las bajas frecuencias presentan un decremento de la pendiente espectral.

D. Heman-Ackah y colaboradores [12] han estudiado la correlación entre las medidas subjetivas de calidad de la voz respecto a las objetivas. A partir de un grupo de locutores con voces pre y postoperatorias de una muestra de pacientes con un cuadro de parálisis en el nervio laríngeo recurrente unilateral, se ha estudiado la correlación entre medidas objetivas con respecto a medida subjetivas de calidad de la voz: grado de disfonía, *breathiness* y *roughness*. Se estudiaron las siguientes medidas objetivas: la prominencia del pico cepstral principal, la relación señal ruido, el cociente de perturbación de amplitud, la perturbación promedio relativa y el cociente de perturbación de *pitch*. El grado de disfonía y el *breathiness* presentan un alto grado de correlación con la prominencia del pico central principal, sin embargo se obtienen bajos niveles de correlación con el resto de las medidas objetivas. El trabajo concluye en que la prominencia del pico central principal es un buen indicador del grado de disfonía.

M. Rosique y colaboradores [13] realizan un estudio de la fonación de tres grupos de locutores: locutores con laringectomía total y voces erigimofónicas, locutores con laringectomía total con fistuloplastia y prótesis fonatoria, y locutores normales. En el trabajo se han estudiado las características del tracto vocal: energías, frecuencias y anchos de banda, de los cuatro formantes de las cinco vocales castellanas en determinadas frases leídas. En el estudio, en función de los resultados, se afirma que la voz traqueoesofágica no se aproxima más a la voz normal que una voz esofágica, pese a que contradice lo publicado en la bibliografía referente a este tema. Los autores achacan esta situación a la calidad de las medidas empleadas en el estudio.

F. Núñez Batalla y colaboradores [14] presentan un estudio sobre la interrelación entre la astenia vocal y la predicción a partir de medidas acústicas que cuantifican la hipofunción vocal. En este trabajo se propone un sistema basado en siete medidas de calidad de voz, procesadas por medio de regresión logarítmica, para la predicción de la astenia, demostrándose que, tras estudiar diferentes medidas de calidad de la voz como

Jitter, Shimmer, relación armónica ruido (HNR), energía de ruido normalizada (NNE), una cuantificación discreta sobre la falta de armónicos y una cuantificación de la presencia de subarmónicos. Aplicando técnicas de regresión logarítmica sobre la pérdida de componentes armónicos en las porciones medias y altas del espectro es posible la predicción de la astenia vocal.

J.C. Casado Morente y colaboradores [15] han estudiado el comportamiento de un conjunto de medidas de calidad de la voz: frecuencia fundamental, *Jitter*, *Shimmer* y ruido glótico (NNE, HNR, SNR), en determinados grupos de locutores: locutores afectados por nódulos, afectados por pólipos y normales. Concluyéndose en dicho trabajo que los pacientes con una disfonía funcional muestran unos resultados patológicos representados por un descenso de la frecuencia fundamental, un aumento de las perturbaciones de la frecuencia (*Jitter*), un aumento de las perturbaciones de la intensidad (*Shimmer*) y un aumento del ruido glótico (NNE, HNR, SNR). Dentro de los pacientes disfónicos se detectaron mayores alteraciones en el grupo de las disfonías por pólipos vocales.

Michael P. Cannito y colaboradores [16] han estudian la sensibilidad de las amplitudes espectrales en casos de voces con disfonía espasmódica abducida, estudiándose el efecto sobre el primer (H1) y el segundo (H2) armónico, y sobre el primer, el segundo y el tercer formante (A1, A2, A3). Como conclusiones destacan la significativa utilidad de una medida de la calidad de la voz basada en la diferencia entre la amplitud espectral del primer armónico con respecto a la amplitud espectral del segundo armónico.

Kristine Tanner y colaboradores [19] han estudio la evolución de las características espectrales de un grupo de pacientes con disfonías funcionales, antes y después de la terapia. En este trabajo se estudian los cuatro primeros momentos (media, desviación típica, *skewness* y *kurtosis*) de la función de densidad de probabilidad del espectro de la señal registrada. En este trabajo, a modo de conclusión, se propone la utilidad del momento de primer orden (la media) como medida objetiva que plasma la evolución de la calidad de voz de pacientes con disfonías funcionales que se someten a terapia.

J. Damborenea Tajada y colaboradores [22] han estudiado la variación de las medidas de la calidad de la voz para locutores sanos diferenciándolos entre fumadores y no fumadores, observándose que la frecuencia fundamental desciende y que el *Shimmer* y el *Jitter* se incrementan en el caso de los fumadores. Sin embargo no se apreciaron en el estudio características significativas en las medidas HNR y NNE.

Preciado López y colaboradores [23] han estudiado estadísticamente cuál de la cinco vocales del español es la más adecuada en el análisis de la voz, concluyéndose que medidas de la calidad de la voz como el *Jitter*, *Shimmer* y el nivel de ruido calculados a partir de la vocal “e” expresan mejor la evolución de la calidad de la voz tras ser intervenidos quirúrgicamente o seguir un tratamiento ortofónico.

Núñez Batalla y colaboradores [24] se ponen en práctica el protocolo de *Yanagihara* donde las características acústicas de la voz se dividen en: componentes de ruido en el formante principal de cada vocal, componentes de ruido en altas frecuencias (por encima de 3Khz), y pérdida de componentes armónicas en altas frecuencias. Con la progresión de la severidad estos tres aspectos se hacen más manifiestos, identificándose cuatro grados:

Grado I: Las componentes armónicas se mezclan con componentes de ruido, principalmente en la región de los formantes vocales.

Grado II: Las componentes de ruido predominan sobre las componentes armónicas del segundo formante. Además, existen ligeros componentes de ruido en altas frecuencias (por encima de 3 Khz).

Grado III: El segundo formante se reemplaza en su totalidad por el ruido. El comportamiento del ruido en altas frecuencias intensifica su energía y expande su rango.

Grado IV: El primer formante pierde sus componentes periódicas y se observan componentes de ruido. En las altas frecuencias se intensifica aún más el ruido.

En este trabajo, los autores concluyen añadiendo que esta estrategia para objetivar la calidad de voz necesita ser complementada con otros procedimientos con el fin de diagnosticar, para lo cual propone el uso de la laringoscopia.

A. Nieto y colaboradores [25] indican que las lesiones por abuso vocal en las cuerdas altera la producción de voz aumentando la presencia de ruido en el sonido. Indicando que HNR y el análisis espectrográfico son dos posibilidades para la cuantificación de la alteración. El estudio evaluó a un conjunto locutores afectados por abuso vocal antes y después de la intervención quirúrgica concluyéndose que la HNR aumenta tras la intervención quirúrgica. Al mismo tiempo, la mejora en el cierre glótico determinó una mejora en los distintos parámetros de análisis acústico.

Y como último ejemplo, Jilua C. Selby y colaboradores [11] han estudiado la calidad de la voz antes y después del tratamiento de locutores con reflujo laringofaríngeos. En el estudio se indica la existencia de una pequeña pero significativa mejora en la percepción de la voz de los locutores tras el tratamiento. Las medidas de la calidad de la voz no han manifestado cambios excepto en la relación señal ruido.

10.4.2 Protocolo Propuesto

En este apartado se propone un protocolo de evaluación de la calidad clínica de la voz, a partir de una grabación de la fonación sostenida de una vocal, que consiste en evaluar de forma objetiva, por medio de un sistema informático, cuatro aspectos de la calidad de la voz: la estabilidad de la voz, la riqueza espectral, la presencia de ruido y el comportamiento no lineal. Para ello se asigna el valor de una determinada medida a cada una de los diferentes aspectos que caracterizan la calidad de la voz.

Además de dicho protocolo, se presenta un estudio preliminar sobre el valor de normalidad de las diferentes medidas. El estudio estadístico se ha realizado con el software SPSS.

10.4.2.1 Protocolo de grabación

Las características del proceso de grabación son las siguientes:

- La grabación consiste en la fonación de forma sostenida de la vocal “a” de forma sostenida, durante al menos 2 segundos. Ha de ser una fonación no susurrada.

- La grabación de audio se puede realizar con un PC de propósito general equipado con una tarjeta de sonido convencional. La grabación se realiza con una frecuencia de muestreo de 22050 Hz y una resolución de 16 bits por muestra.
- Se podrá utilizar un micrófono convencional de sobremesa, que al menos presente un ancho de banda de 11 Khz. Se recomienda los micrófonos de sobremesa, a una distancia de 15 centímetros del locutor.

10.4.2.2 Medidas de Calidad de la Voz

En el protocolo de cuantificación de la calidad de la voz, se indica la necesidad de medir cuatro características de la voz, las cuales se citan a continuación:

- **Estabilidad de la voz:** es la capacidad de un locutor de producir, durante la producción de un sonido sonoro sostenido, un flujo de aire con una intensidad constante para excitar las cuerdas vocales.
- **Riqueza espectral:** se define como la capacidad de generar, durante la fonación de un sonido sostenido, un movimiento periódico de las cuerdas vocales y de producir una excitación sonora al tracto vocal con gran cantidad de componentes espectrales.
- **Presencia de ruido:** se define como la presencia de ruido glótico en la voz, durante la fonación de un sonido sonoro sostenido, debido a carencias en la fase de cierre de las cuerdas vocales.
- **Comportamiento no lineal:** este fenómeno físico es ocasionado por un funcionamiento anómalo de las cuerdas vocales, debido a irregularidades en las masas de cuerdas vocales que interviene en la fase de cierre de cuerdas, existencia de asimetrías en el movimiento de las cuerdas y factores relacionados con la mucosa de las cuerdas.

Cada una de las diferentes medidas, puede ser cuantificada por medio de diferentes medidas o combinaciones de las mismas. En este trabajo se propone las siguientes

medidas para cuantificar cada uno de los diferentes aspectos que caracterizan la calidad de la voz²:

TABLA 10-9 : MEDIDAS PARA CUANTIFICAR LOS DISTINTOS FENÓMENOS FÍSICOS IMPLICADOS EN LA MEDIDA DE LA CALIDAD DE LA VOZ.

Estabilidad de la voz	(GI.4) Shimmer versión 4 (APQ)
Riqueza espectral	(GII.19) Media Valor Cepstral del primer pico de los cepstrum derivados
Presencia de ruido	(GIII.39) Media Relación Espectral Subarmónico - armónico de al señal residual
Comportamiento no lineal	(GIV.15) Media Área inferior del biespectro integrado

10.4.2.3 *Estadísticos descriptivos de las medidas de calidad*

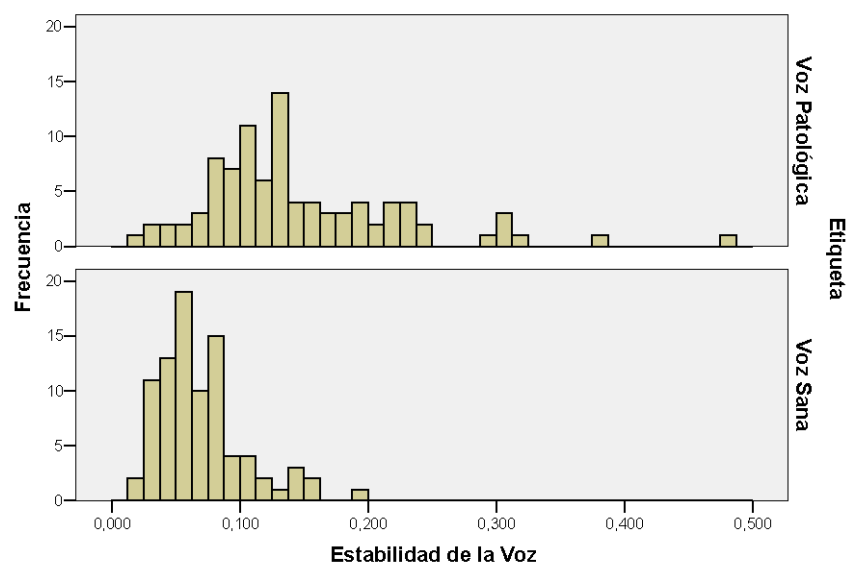
Se ha realizado una descripción del comportamiento de un conjunto de estadísticos con el fin de caracterizar las diferentes medidas de calidad propuestas en el protocolo. En la interpretación de los resultados, es necesario tener presente el reducido número de muestras de voces que se dispone donde el estudio se ha realizado con la base de datos propia.

Las diferentes medidas se han estudiado, utilizando como muestras de alta calidad voces sanas y como muestras de voces de baja calidad, voces de locutores que presentan alguna patología en el sistema fonador. Para cada una de las medidas se muestran dos representaciones: el histograma de frecuencia y la representación de tallo y hoja, y diferentes estadísticos descriptivos: media, mediana, desviación típica, skewness, kurtosis, varianza, rango, máximos, mínimos, etc.

² Esta asociación entre aspectos característicos de la voz y medidas para cuantificarlos, depende de la fonación. En el caso del protocolo que se presenta, se asume que la fonación consiste en la fonación de la vocal “a” de forma sostenida.

Estabilidad de la voz

Histograma de Frecuencia

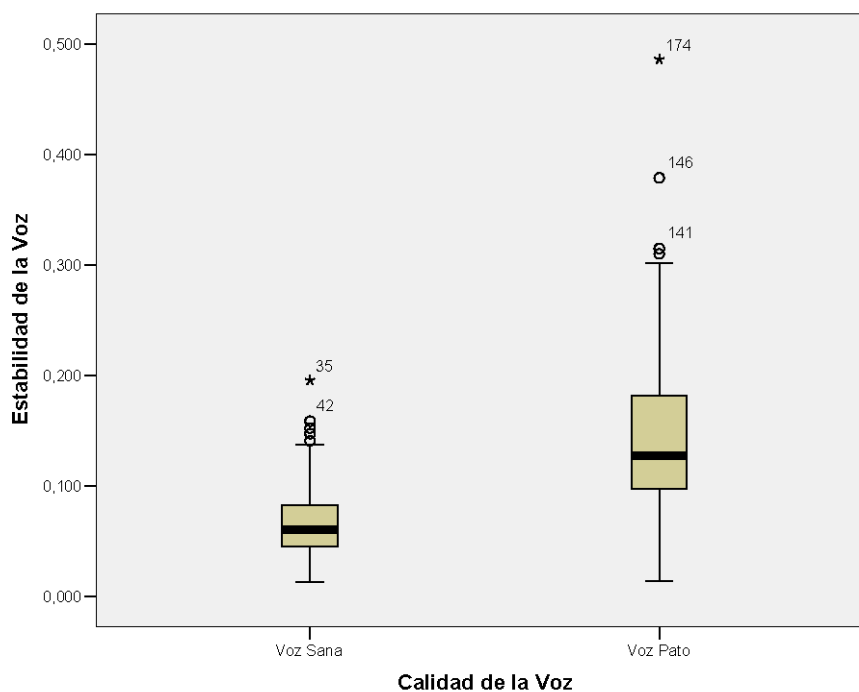


Resumen del procesamiento de los casos

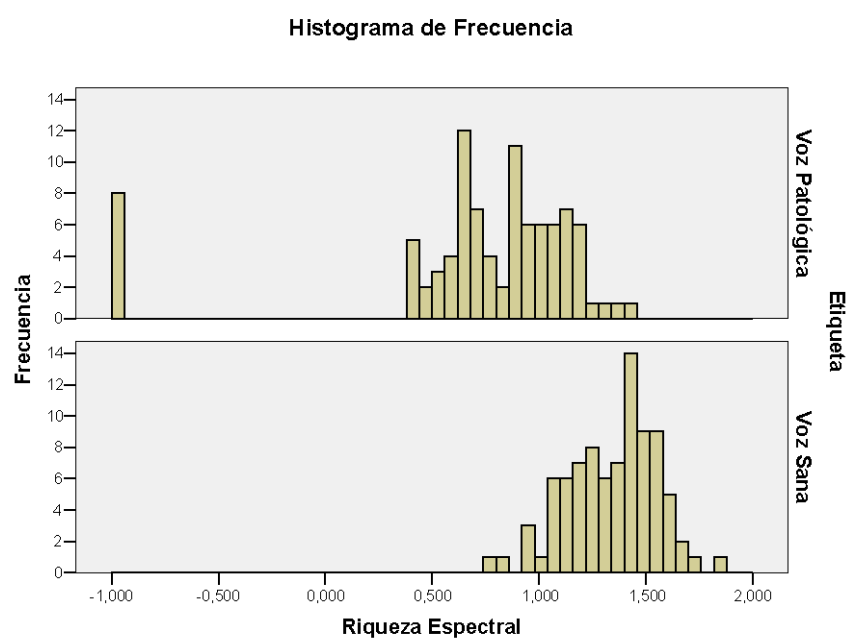
		Casos					
		Válidos		Perdidos		Total	
		N	Porcentaje	N	Porcentaje	N	Porcentaje
Estabilidad de la Voz	Voz Sana	87	100,0%	0	,0%	87	100,0%
	Voz Patológica	93	100,0%	0	,0%	93	100,0%

Descriptivos

Calidad de la Voz				Estadístico	Error típ.
Estabilidad de la Voz	Voz Sana	Media		,06878	,003671
		Intervalo de confianza para la media al 95%	Límite inferior	,06149	
			Límite superior	,07608	
		Media recortada al 5%		,06614	
		Mediana		,06066	
		Varianza		,001	
		Desv. típ.		,034238	
		Mínimo		,013	
		Máximo		,196	
		Rango		,183	
		Amplitud intercuartil		,038	
		Asimetría		1,246	
		Curtosis		1,849	
	Voz Patológica	Media		,14582	,008176
		Intervalo de confianza para la media al 95%	Límite inferior	,12958	
			Límite superior	,16205	
		Media recortada al 5%		,14011	
		Mediana		,12752	
		Varianza		,006	
		Desv. típ.		,078842	
		Mínimo		,014	
		Máximo		,486	
		Rango		,472	
		Amplitud intercuartil		,090	
		Asimetría		1,471	
		Curtosis		3,383	



Riqueza espectral



En el histograma se observa una característica peculiar: para algunos casos las muestras de voz patológica presentan valores de la medida igual a -1. Esto es debido, a que en determinadas fonaciones no existe vibración de las cuerdas vocales, lo cual

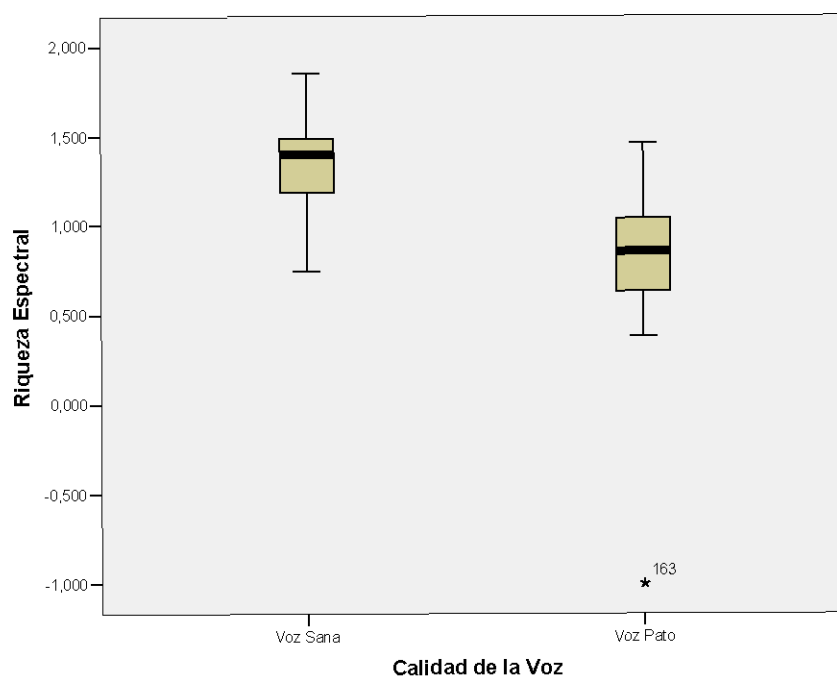
produce que sea imposible estimar el valor de la medida. En estos casos, se le asigna el valor -1 a la medida.

Resumen del procesamiento de los casos

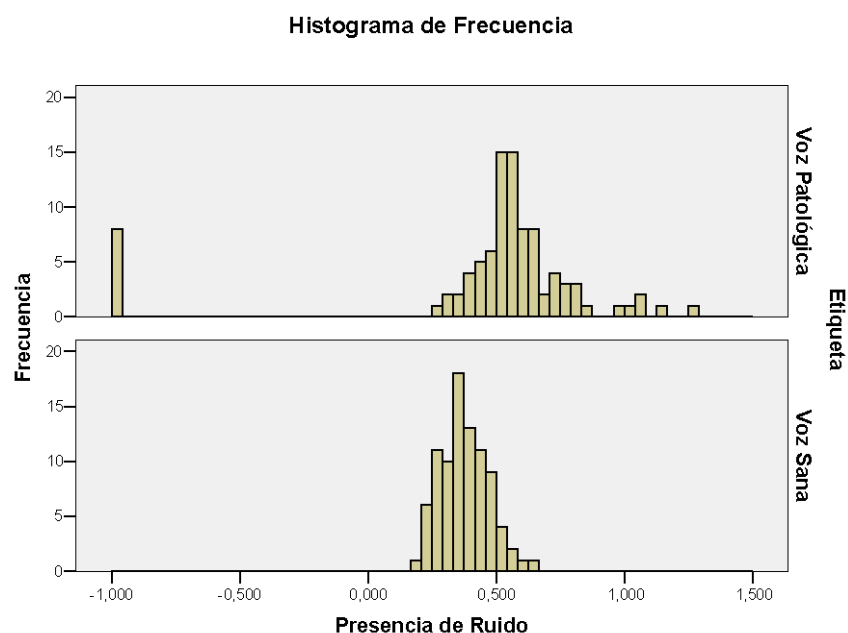
		Casos					
		Validos		Perdidos		Total	
		N	Porcentaje	N	Porcentaje	N	Porcentaje
Riqueza Espectral	Voz Sana	87	100,0%	0	,0%	87	100,0%
	Voz Patológica	93	100,0%	0	,0%	93	100,0%

Descriptivos

Calidad de la Voz				Estadístico	Error típ.
Riqueza Espectral	Voz Sana	Media		1,33998	,022567
		Intervalo de confianza para la media al 95%	Límite inferior	1,29512	
			Límite superior	1,38484	
		Media recortada al 5%		1,34537	
		Mediana		1,39320	
		Varianza		,044	
		Desv. típ.		,210494	
		Mínimo		,747	
		Máximo		1,850	
		Rango		1,103	
		Amplitud intercuartil		,309	
		Asimetría		-,419	,258
		Curtosis		,026	,511
	Voz Patológica	Media		,69886	,059699
		Intervalo de confianza para la media al 95%	Límite inferior	,58030	
			Límite superior	,81743	
		Media recortada al 5%		,75808	
		Mediana		,85539	
		Varianza		,331	
		Desv. típ.		,575721	
		Mínimo		-1,000	
		Máximo		1,457	
		Rango		2,457	
		Amplitud intercuartil		,417	
		Asimetría		-2,114	,250
		Curtosis		4,108	,495



Presencia de ruido



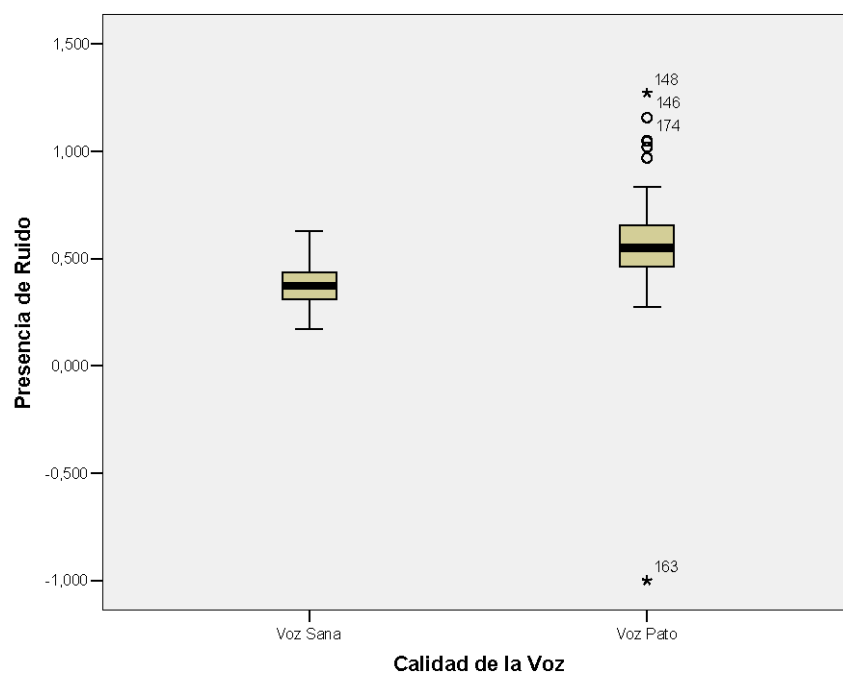
Al igual que en la medida anterior, en el histograma se observa que, en algunos casos las muestras de voces patológicas presentan valores de la medida igual a -1. Esto es debido, a que en determinadas fonaciones no existe vibración de las cuerdas vocales, lo cual produce que sea imposible estimar el valor de la medida, asignándosele el valor -1 a la medida.

Resumen del procesamiento de los casos

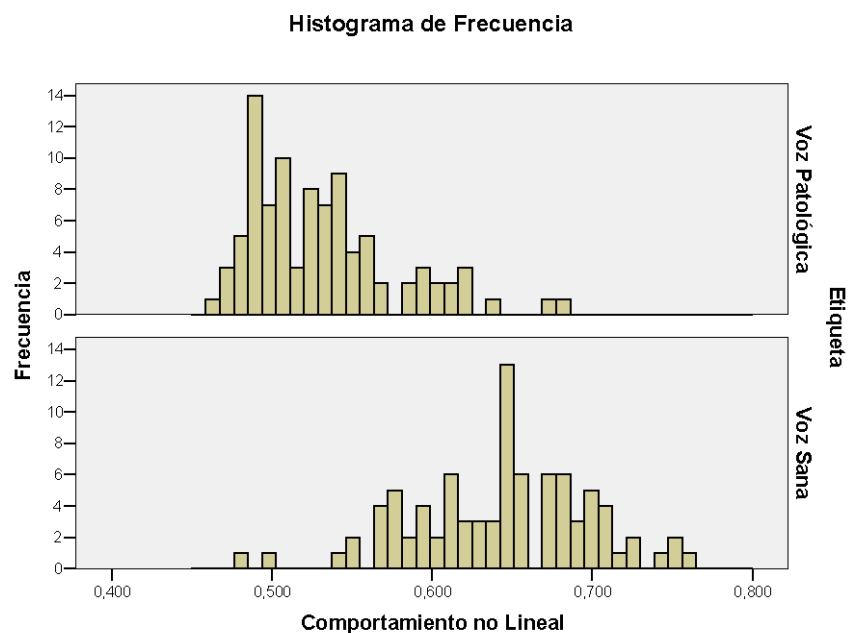
		Casos					
		Validos		Perdidos		Total	
	Calidad de la Voz	N	Porcentaje	N	Porcentaje	N	Porcentaje
Presencia de Ruido	Voz Sana	87	100,0%	0	,0%	87	100,0%
	Voz Patológica	93	100,0%	0	,0%	93	100,0%

Descriptivos

Calidad de la Voz				Estadístico	Error tip.
Presencia de Ruido	Voz Sana	Media		,37508	,009990
		Intervalo de confianza para la media al 95%	Limite inferior	,35522	
			Limite superior	,39494	
		Media recortada al 5%		,37277	
		Mediana		,37313	
		Varianza		,009	
		Desv. tip.		,093181	
		Mínimo		,174	
		Máximo		,629	
		Rango		,455	
		Amplitud intercuartil		,130	
		Asimetría		,326	
		Curtosis		-,148	
	Voz Patológica	Media		,46139	,050042
		Intervalo de confianza para la media al 95%	Limite inferior	,36200	
			Limite superior	,56078	
		Media recortada al 5%		,50621	
		Mediana		,54881	
		Varianza		,233	
		Desv. tip.		,482591	
		Mínimo		-1,000	
		Máximo		1,273	
		Rango		2,273	
		Amplitud intercuartil		,194	
		Asimetría		-2,267	
		Curtosis		4,962	



Comportamiento no lineal

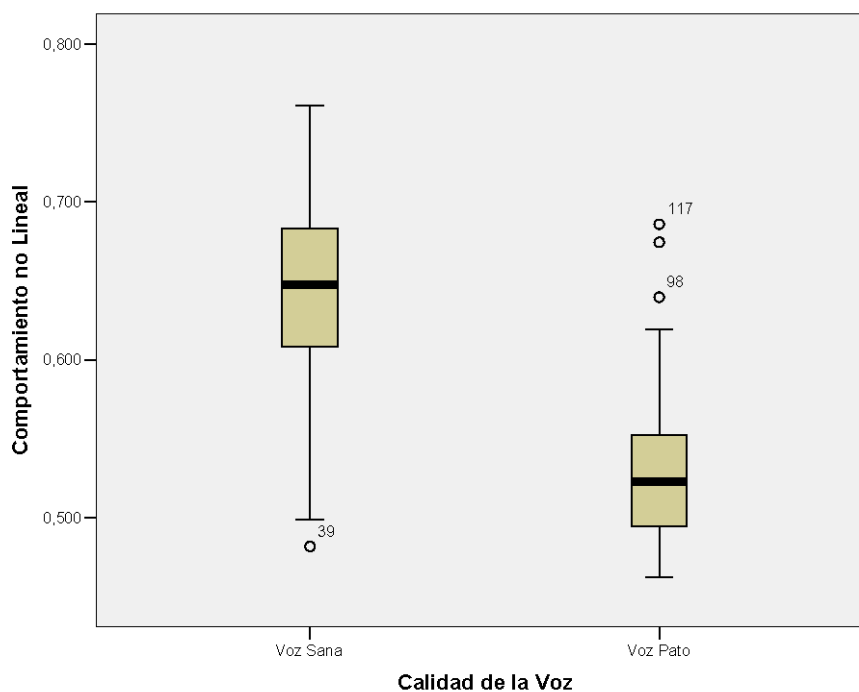


Resumen del procesamiento de los casos

		Casos					
		Válidos		Perdidos		Total	
	Calidad de la Voz	N	Porcentaje	N	Porcentaje	N	Porcentaje
Comportamiento no Lineal	Voz Sana	87	100,0%	0	,0%	87	100,0%
	Voz Patológica	93	100,0%	0	,0%	93	100,0%

Descriptivos

Calidad de la Voz				Estadístico	Error tip.
Comportamiento no Lineal	Voz Sana	Media		,64269	,006012
		Intervalo de confianza para la media al 95%	Límite inferior	,63074	
			Límite superior	,65464	
		Media recortada al 5%		,64351	
		Mediana		,64780	
		Varianza		,003	
		Desv. tip.		,056076	
		Mínimo		,482	
		Máximo		,761	
		Rango		,280	
		Amplitud intercuartil		,077	
		Asimetría		-,288	
		Curtosis		,071	
	Voz Patológica	Media		,53157	,004778
		Intervalo de confianza para la media al 95%	Límite inferior	,52208	
			Límite superior	,54105	
		Media recortada al 5%		,52831	
		Mediana		,52251	
		Varianza		,002	
		Desv. tip.		,046074	
		Mínimo		,462	
		Máximo		,686	
		Rango		,224	
		Amplitud intercuartil		,059	
		Asimetría		1,130	,250
		Curtosis		1,115	



10.4.2.4 Recomendaciones de Uso

Una voz de calidad anormal presenta al menos uno de los valores correspondientes a la cuantificación de los cuatro fenómenos físicos fuera de los rangos de normalidad. Este protocolo de evaluación cuantifica la calidad de la voz permitiendo identificar calidades de voz anómalas de diferentes orígenes. A medida que la calidad de la voz de un locutor empeora, aumentará el número de fenómenos físicos y la desviación de los valores de la cuantificación de dichos fenómenos respecto a los valores de normalidad.

Se ha estimado los rangos de normalidad de las cuatro medidas de calidad de la voz, utilizando la base de datos propia y tomando el 95% central de los valores de la población de muestras sanas. Para la estimación de estos rangos se ha realizado un pequeño estudio: se ha realizado el test de *kolmogorov-Smirnov* de los valores de las medidas para la población correspondiente a voces sanas, representando la distribución normal resultante sobre el histograma de frecuencias en cada caso. Una vez que se han obtenido los parámetros de la estimación del modelo normal de cada medida (media muestral $\bar{r}$ y desviación típica $\sigma(\bar{r})$), se identifica el rango de normalidad como:

$$RANGO = [\bar{r} - 2\sigma(\bar{r}), \bar{r} + 2\sigma(\bar{r})] \quad [\text{Ec.10-1}]$$

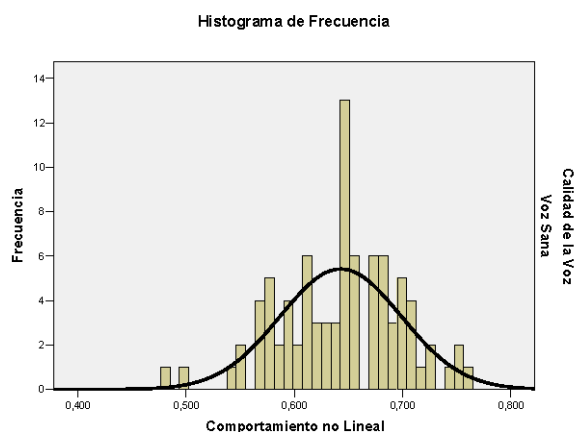
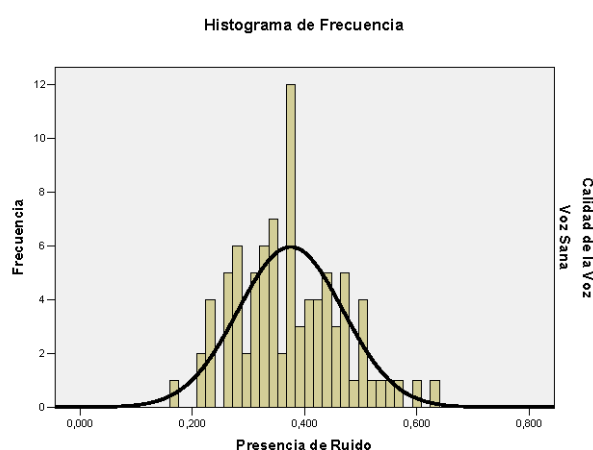
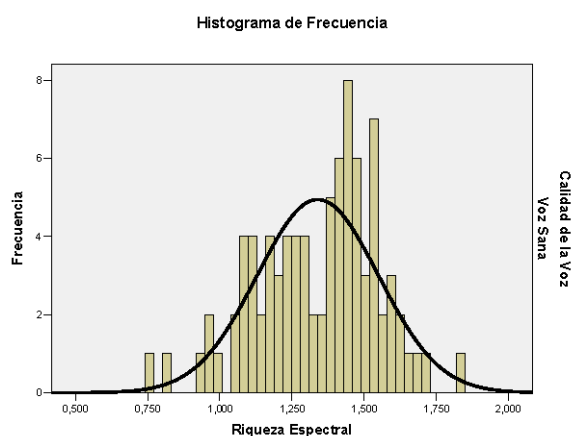
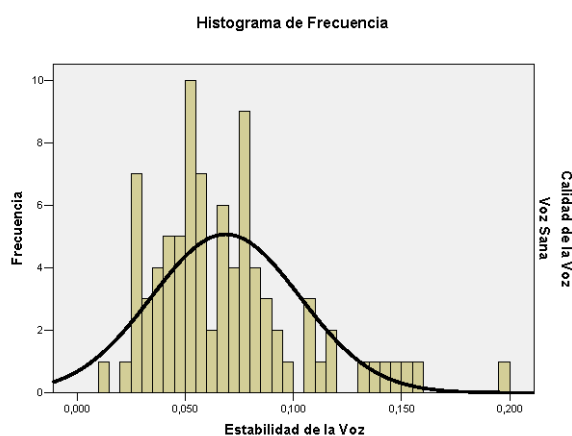
Prueba de Kolmogorov-Smirnov para una muestra

		Estabilidad de la Voz	Riqueza Espectral	Presencia de Ruido	Comportamiento no lineal
N		87	87	87	87
Parámetros normales ^{a,b}	Media	,06878	1,33998	,37508	,64269
	Desviación típica	,034238	,210494	,093181	,056076
Diferencias más extremas	Absoluta	,109	,106	,070	,081
	Positiva	,109	,048	,070	,039
	Negativa	-,090	-,106	-,029	-,081
Z de Kolmogorov-Smirnov		1,017	,990	,652	,753
Sig. asintót. (bilateral)		,252	,280	,789	,622

a. La distribución de contraste es la Normal.

b. Se han calculado a partir de los datos.

Se observa que para el caso de las medidas de presencia de ruido y comportamiento no lineal la significancia es relativamente baja, siendo este hecho atribuible al número reducido de muestras.



Por lo tanto, el protocolo propone cuantificar la calidad de la voz por medio de cuatro características, cada una de las cuales se cuantifica estimando el valor de una determinada medida de calidad de la voz. Utilizando la fonación de la vocal “a” los rangos de normalidad son los siguientes.

TABLA 10-10 : RANGOS DE NORMALIDAD DE LAS DIFERENTES MEDIDAS DE LA CALIDAD DE LA VOZ.

Estabilidad de la voz	0.000.....0.136
Riqueza espectral	0.919.....1.759
Presencia de ruido	0.189.....0.561
Comportamiento no lineal	0.530.....0.754

10.5 Referencias

- [1] Mirjam Wester, Automatic Classification of Voice Quality: Comparing Regression Models and Hidden Markov Models, *Proceedings of of Symposium on Databases in Voice Quality Research and Education, VOICEDATA98*, pp 92-97, 1998.
- [2] Robert F. Orlikoff, R. J. Baken, "Curing Diagnosis: Improving the Taxonomy of Phonatory Dysfunction", *Sixth Conference on Advances in Quantitative Laryngology*. Hamburg, Germany, April 2003.
- [3] Ping Yu, Maurice Ouaknine, Joana Revis, Antonine Giovanni, "Objective Voice Analysis for Dysphonic Patients: A Multiparametric Protocol Including Acoustic and Aerodynamic Measurements", *Journal of Voice*, vol.15, no.4, pp.529-542, 2001.
- [4] Elisabeth Zetterholm, "Auditory and Acoustic Analysis of Voice Quality Variations in Normal Voices", *Proceedings of the XIVth International Congress of Phonetic Sciences, ICPhS-99*, pp.973-976, 1999;
- [5] Jacques Koreman y Manfred Pützer, "Finding Correlates of Vocal Fold Adduction Deficiencies", *Phonus 3*, Institute of Phonetics, University of the Saarland, vol. 3, pp.155-178, 1997.
- [6] Xuejing Sun and Yi Xu, "Perceived Pitch of Synthesized Voice with Alternate Cycles", *Journal of Voice*, vol.16, no.4, pp.443-459, 1999
- [7] Stefan Hadjitodorov, Petar Mitev, "A computer system for acoustic analysis of pathological voices and laryngeal diseases screening", *Medical Engineering Physics*, vol.24, pp. 419-429, 2002
- [8] Angeles Rovirosa, y colaboradores, "Acoustic Analysis after Radiotherapy in T1 Vocal Cord Carcinoma: a New Approach to the Analysis of Voice Quality", *International Journal of Radiation Oncology Biol. Phys.* , vol.47, no.1, pp.73-79, 2000
- [9] Adriana Vélez Feijó, María Alice Parente, Mara Behlau, Sergio Haussen, María Cecília De Vecino, Beatriz Castellar de Faria Martignago, "Acoustic Analysis of Voice in Multiple Sclerosis Patients", *Journal of Voice*, vol.18, no. 3, 2004
- [10] Dana M. Hartl, Stéphane Hans, Jaqueline Vaissière, Marc Riquet, Daniel F. Brasnu, "Objective Voice Analysis Before and After Onset of Unilateral Vocal Fold Paralysis", *Journal of Voice*, vol.15, no. 3, pp.351-361, 2001
- [11] Jilua C. Selby, Harvey R. Gilbert, J.W.Lerman, "Perceptual and Acoustic Evaluation of Individuals with Laryngopharyngeal Reflux Pre- and Post-treatment", *Journal of Voice*, vol.17, no. 4, pp. 557-570, 2003
- [12] Yolanda D. Heman-Ackah, Deirdre D. Michael, George S. Goding Jr., The "Relationship Between Cepstral Peak Prominence and Selected Parameter of Dysphonia", *Journal of Voice*, vol.16, no. 1, pp.20-27, 2002
- [13] M. Rosique, J.L. Ramón, M. Canteras, L. Rosique, "Análisis Discriminante Aplicado a los Formantes de las Vocales Castellanas en la Fonación con Prótesis y Erigmo fonía tras Laringuectomía Total", *Acta Otorrinolaringol. Esp.* , vol.54, pp.361-366, 2003
- [14] F. Núñez Batalla, P. Corte Santos, B. Señaris González, N. Rodríguez Prado, C. Suarez Nieto, "Evaluación Espectral Cuantitativa de la Hipofunción Vocal", *Acta otorrinolaringol. Esp.* , vol.55, pp.327-333, 2004
- [15] J.C. Casado Morente, J. A. Adrián Torres, M. Conde Jiménez, D. Piédrola Maroto, V. Povedano Rodríguez, E. Muñoz Gomariz, E. Cantillo Baños, A. Jurado Ramos, "Estudio Objetivo de la Voz en Población Normal y en la Disfonía por Nódulos y Pólipos Vocales", *Acta Otorrinolaringol Esp.* vol.52, pp.476-482, 2001

- [16] Michael P. Cannito, Eugene H. Buder, Lesya B. Chorna, "Spectral Amplitude Measures of Adductor Spasmodic Dysphonic Speech", *Journal of Voice*, vol.19, no.3, pp. 391-410, 2005
- [17] Shaheen N. Awan and Nelson Roy, "Acoustic Prediction of Voice Type in Women with Functional Dysphonia", *Journal of Voice*, vol.19, no.2, pp.268-282, 2005
- [18] Tarika Bhuta, Linda Patrick, James D. Garnett, "Perceptual Evaluation of Voice Quality and its Correlation with Acoustic Measurements", *Journal of Voice*, vol.18, no.3, 2004
- [19] Kristine Tanner, Nelson Roy, Andrea Ash, and H. Buder, "Spectral Moments of the Long-term Average Spectrum: Sensitive Indices of Voice Change After Therapy?", *Journal of Voice*, vol.19, no.2, pp 221-222, 2005.
- [20] F Nuñez Batalla, P. Corte Santos, G. Sequeiros Santiago, B. Señaris González, C. Suárez Nieto, "Evaluación Perceptual de la disfonía: Correlación con los Parámetros Acústicos y Fiabilidad", *Acta Otorrinolaringol. Esp.*, vol.55, pp.282-287, 2004
- [21] J. González, T. Cervera, J. L. Miralles, "Análisis Acústico de la voz: Fiabilidad de un conjunto de parámetros multidimensionales", *Acta Otorrinolaringol. Esp*; vol.53, pp.256-268, 2002
- [22] Damborenea Tajada J., Fernández Liesa, R., Llorente Arenas E, Naya Galvez, M. J., Marín Garrido, C., Rueda Gormedino P, Ortiz García, A., "Efecto del Consumo de Tabaco en el análisis Acústico de la Voz", *Acta Otorrinolaring Esp*; vol.50, no.6, pp.448-452, 1999
- [23] Preciado López, J.A., Calzada Uriondo M.G., Zabaleta López M, García Cano, F. J., "Variabilidad en el análisis digital de la voz según vocal analizada, en pacientes normales y pacientes disfónicos", *Acta Otorrinolaring Esp*, vol.51, no.7, pp.618-628, 2000
- [24] Núñez Batalla F, Suarez Nieto, C., Maldonado Fernández, M., Baragaño Río, L., Muñoz Pinto, C., "Aplicaciones del espectrograma de banda estrecho en fonocirugía", *Acta Otorrinolaring Esp*, vol.51, no.3, pp.243-245, 2000
- [25] A. Nieto, I Cobeta, F.J. Gamboa, J. Montojo, A Vegas, "La relación Armónico/ruido y el análisis espectrográfico en la patología de abuso vocal", *Acta Otorrinolaring Esp*, vol.47, no.5, pp. 370-376, 1996
- [26] Jesús B Alonso, Carlos M. Travieso, José de León, Miguel A. Ferrer, "Characterization of a speaker from a telephone recording using speech quality Detection", *39th IEEE International Carnahan Conference on security Technology*, Las Palmas de Gran Canaria, october 2005.

CAPÍTULO 11

Conclusiones y Líneas Futuras

En el contexto de la estimación de la calidad de la voz a partir de una grabación de voz, cualquiera de los diferentes fenómenos físicos que caracterizan dicha calidad de la voz se manifiesta en diferentes dominios aunque sin embargo, generalmente se manifiesta de forma preferencial en uno de ellos. Resulta de interés identificar el dominio preferencial en el que se manifiesta cada uno de los distintos fenómenos, permitiendo aumentar la capacidad de cuantificar dicho fenómeno y por lo tanto dando lugar a la posibilidad de cuantificar la calidad de la voz de forma más efectiva.

No existe un único fenómeno que permita diferenciar entre las distintas calidades de voz, para poder documentar dicha calidad de voz es necesario tener en cuenta un conjunto de fenómenos físicos que intervienen en la caracterización de la calidad de la voz de un locutor.

En este trabajo se han identificado cuatro fenómenos físicos cuantificables que permiten realizar una medición de la calidad de la voz, los cuales se han denominado en este trabajo como:

- **Estabilidad de la voz:** es la capacidad de un locutor de producir, durante la fonación de un sonido sonoro sostenido, un flujo de aire para excitar las cuerdas vocales con una intensidad constante. Esta medida se calcula a partir de medias de estabilidad de la señal de voz.
- **Riqueza espectral:** se define como la capacidad de generar, durante la fonación de un sonido sostenido, un correcto movimiento periódico de las cuerdas vocales y un correcto golpeo de las cuerdas en cuanto a la intensidad. Una voz con alta riqueza espectral provoca una frecuencia de vibración de las cuerdas vocales (*pitch*) estable a lo largo del tiempo en la fonación de sonidos sonoros sostenidos, además de una excitación sonora al tracto vocal con gran cantidad de componentes armónicas en las diferentes bandas de frecuencias.
- **Presencia de ruido:** se define como un incorrecto cierre glótico dando esto lugar a la presencia de ruido glótico en la señal de voz, el cual se caracteriza por resultar ruido del tipo no estacionario.
- **Comportamiento no lineal:** esta característica de la voz es ocasionada por un funcionamiento anómalo de las cuerdas vocales, debido a irregularidades en las masas de las cuerdas vocales que intervienen en la fase de cierre de cuerdas, la existencia de asimetrías en el movimiento de las cuerdas y factores relacionados con la mucosa de las cuerdas. Estos fenómenos son cuantificables por medio de caracterizaciones no lineales de la señal de voz.

El cuarto fenómeno físico cuantificable, el comportamiento no lineal, resulta una aportación novedosa dentro de los sistemas convencionales de evaluación de la calidad de la voz.

Fruto de los resultados obtenidos en los distintos trabajos presentes en la literatura, es posible identificar la coincidencia en que no existe una única medida que permita documentar la calidad de una voz de un locutor. Como resultado de este trabajo, se ha identificado cuatro características de la voz donde, cualquier voz de calidad anormal presenta una desviación respecto al patrón de normalidad de al menos una de las cuatro características. Además, resulta de gran interés, la nueva línea de trabajo que pretende identificar características no lineales de la señal de voz como un elemento más de los que caracteriza la calidad de voz.

Existen múltiples medidas asociadas a cada uno de los cuatro grupos de medidas (estabilidad de la voz, riqueza espectral, presencia de ruido y comportamiento no lineal), las cuales presentan diferentes resultados con el fin de discriminar entre voces sanas y voces patológicas. Al no existir una base de datos de referencia no resulta útil comparar los resultados presentados por los diferentes autores. Este hecho pone de manifiesto la necesidad de disponer de bases de datos de voces comunes con diferentes calidades de voz, lo suficientemente amplia y balanceada para dotar de mayor valía los estudios en la línea de este trabajo.

Obtener unas medidas eficaces de la calidad de la voz siempre ha sido una tarea no exenta de complejidad debido a la fuerte componente subjetiva de dichas medidas. En este trabajo, se ha combinado la experiencia clínica con la precisión de técnicas propias de la minería de datos. Se ha estudiado la relevancia de las diferentes medidas de la calidad de la voz, utilizando un procedimiento mixto en el que se combina la información a priori sobre las características cuantificables relativas a la calidad de la voz y técnicas heurísticas de selección de atributos.

En el estudio de la selección de medidas de la calidad de la voz se ha buscado que los resultados sean lo más independientes que sea posible de la base de datos utilizada. Este hecho se ha conseguido agrupando a priori las diferentes medidas de calidad de la voz en cuatro grupos en función del fenómeno físico que cuantifican, y posteriormente aplicando técnicas de selección de características a los diferentes grupos de medidas.

Esta estrategia permite solventar la carencia de una gran base de datos de voces con diferentes calidades de voz, permitiendo identificar las medidas de calidad de voz más relevantes a partir de bases de datos más modestas y minimizar el efecto de, entre otros, factores como: la cantidad de muestras registradas de diferentes patologías del sistema fonador, la cantidad de muestras de voces representativas de cada clase, la cantidad de muestras de voces con diferentes grados de disfonía, el volumen de muestras, el número de repeticiones, etc. Por tanto, la técnica propuesta permite obtener resultados que pueden estimarse independientemente de la base de datos.

Como resultado de este procedimiento de selección de medidas de calidad, se propone un nuevo protocolo de medida de la calidad de la voz a partir de grabaciones acústicas. En este protocolo, a partir de cuatro medidas de calidad de la voz por vocal, es posible obtener una correlación entre el valor numérico de la medida y el fenómeno físico que cuantifica, permitiendo detectar desviaciones respecto de los márgenes de normalidad.

Atendiendo a los resultados del proceso de estimación en la relevancia de las diferentes medidas de calidad de la voz, es posible identificar las mediadas de calidad óptimas para cada vocal, así como cuales son las vocales más apropiadas para la evaluación de la calidad de la voz.

Con la intención de validar la utilidad de las medidas de calidad de voz propuestas se han implementado diferentes sistemas de clasificación automáticos: un discriminador entre voz sana y voz patológica para voz sostenida, un discriminador entre voz sana y diferentes grados de disfonía para voz sostenida y un discriminador entre voz sana y voz patológica para habla continua. Atendiendo a los resultados obtenidos en las diferentes aplicaciones, y teniendo en cuenta que la fase de clasificación es realizada únicamente a partir de las medidas de la calidad de la voz, puede entenderse que las aplicaciones implementadas validan la calidad de las medidas, y por consiguiente del procedimiento de estimación de la relevancia de las diferentes medidas de la calidad de la voz.

El sistema discriminador entre voz sana y voz patológica para voz sostenida fue implementado utilizando dos bases de datos diferentes. Con una de las base de datos se realizó el estudio de la relevancia de las medidas de calidad de la voz, siendo independiente la segunda base de datos del proceso de selección de características. Los resultados obtenidos empleando una y otra base de datos son muy similares, lo cual nos hace entender que el subconjunto de medidas de calidad de la voz propuesto en este trabajo resulta independiente de la base de datos.

Además, atendiendo a los resultados obtenidos en los diferentes sistemas de clasificación automática, y teniendo en cuenta que las distintas grabaciones de voz de las bases de datos fueron adquiridas en condiciones acústicas realistas, ésto apoya la

posibilidad de implementar sistemas que funcionen en condiciones acústicas realistas propios de entornos reales de trabajo.

11.1 Líneas Futuras

Como líneas futuras de trabajo se plantea la necesidad de estudios clínicos que demuestren la utilidad en entornos hospitalarios de este trabajo dentro del ámbito de la otorrinolaringología, logopedia y foniatría. Estos estudios abrirán la puerta a la aplicación del protocolo propuesto en aplicaciones como la valoración de una intervención quirúrgica, la documentación de la evolución de un tratamiento, o su aplicación en la documentación médico legal, e incluso para su aplicación en la telemedicina.

A partir de bases de datos mayores de las utilizadas en este trabajo, será posible implementar sistemas de clasificación automática entre voz sana y voz patológica, o incluso sistemas que de forma automática a partir de una grabación de voz implemente un medidor del grado de disfonía. Estos sistemas pueden resultar de interés para realizar un primer *screening* en atención primaria o en evaluación logopédica.

Se abre la puerta a sistemas de síntesis de voz de alta calidad que tengan en cuenta aspectos como, además de la necesaria inteligibilidad del sistema, la naturalidad, estando este concepto íntimamente ligado a las medidas de calidad de la voz. Estos sistemas son aplicables a nuevas generaciones de interfases máquina-hombre.

Cabe la posibilidad de avanzar en la investigación aplicando el conocimiento adquirido en medidas de calidad de la voz en la detección de estados anímicos como tristeza, enfado, o alegría. Con estos sistemas será posible percibir diferentes aspectos del lenguaje no verbal.

Y por último, por citar algunas posibilidades más, es posible aplicar el protocolo de medidas de calidad de la voz, a la evaluación de sistemas de transmisión-recepción de voz con pérdidas, permitiendo evaluar objetivamente la calidad de los sistemas de comunicación los cuales habitualmente se evalúan por medio de técnicas subjetivas.



ANEXO A

Resultados de la Selección de Características

En este apartado de la memoria se presentan los resultados de las simulaciones realizadas en esta tesis doctoral a partir de los cuales fue posible el diseño de diferentes aplicaciones, las cuales se presentan en el capítulo 10, y se obtuvieron las conclusiones de este trabajo, descritas en el capítulo 11.

En este capítulo se presenta, en el apartado 12.1, las diferentes medidas de la calidad de la voz, agrupadas en los cuatro grupos de medidas: estabilidad de la voz (Grupo I), riqueza espectral (Grupo II), presencia de ruido (Grupo III) y comportamiento no lineal (Grupo IV).

En el apartado 12.2, se presentan los resultados de la ordenación de las características de los diferentes grupos de medidas en función del grado de relevancia. Diferenciando entre la fonación analizada, es decir la vocal estudiada, y la unidad de clasificación: “voz sana – voz patológica” o “voz sana – disfonía leve – disfonía moderada – disfonía severa”.

Por último, en el apartado 12.3, se presenta los resultados de un estudio sobre la variación de la tasa de éxito en función del número de medidas de calidad de la voz utilizadas en el evaluación. En el estudio se ha diferenciado, entre vocal, grupo de medidas y clasificador utilizado.

Para la realización de las simulaciones ha sido empleado el entorno matemático Matlab, la versión 6.1, y las siguientes toolbox:

- Pattern Recognition Tools (PRTOOLS 3.0)[1]
- TISEAN package [2]
- Higher-Order Spectral Analysis Toolbox (Version 2.0.3) [3]
- The Genetic Algorithm Optimization Toolboz GAOT [4]

12.1 Medidas de Calidad

En este apartado de la memoria se indican las medidas de la calidad de la voz utilizadas en este estudio, así como su clasificación en uno de los cuatro grupos de medidas propuestos. Cada medida de calidad de voz se identifica por un índice en uno de los cuatro grupos de medidas.

TABLA 12-1: MEDIDAS DE CALIDAD DE LA VOZ PERTENECIENTES AL GRUPO I (ESTABILIDAD DE LA VOZ)

Identificador de la medida	Grupo I : “Estabilidad de la Voz”
1	Shimmer versión 1
2	Shimmer versión 2
3	Shimmer versión 3 (APQ)
4	Shimmer versión 4 (APQ)
5	Shimmer versión 5 (Fatr)
6	Media Short Time cross correlation function
7	Variación Short Time cross correlation function
8	Clycle-to-clycle mainform variation
9	Media Kurtosis de la señal residual
10	Variación Kurtosis de la señal residual
11	Media Amplitud máxima de Ug
12	Variación Amplitud máxima de Ug
13	Media Margen dinámico de Ug
14	variación Margen dinámico de Ug
15	media Valor cuadrático medio
16	variación Valor cuadrático medio
17	Media kurtosis de las diferentes tramas
18	Variación kurtosis de las diferentes tramas

TABLA 12-2: MEDIDAS DE CALIDAD DE LA VOZ PERTENECIENTES AL GRUPO II (RIQUEZA ESPECTRAL)

Identificador de la medida	Grupo II : “Riqueza espectral”
1	Jitter versión 1
2	Jitter versión 2
3	Jitter versión 3 (PPQ)
4	Jitter versión 4 (Tremor)
5	Jitter versión 5 (HFPV)
6	Jitter versión 6 (Fftr)
7	media Relación Espectral Subarmónico-armónico
8	variación Relación Espectral Subarmónico -armónico
9	DUV
10	IPTS
11	Pico Cepstral Principal
12	variación Pico Cepstral Principal
13	Energía Cepstral de Pitch (PCEM)
14	variación Cepstral de Pitch (PCEM)
15	media Energía Cepstral del segundo armónico (SHECM)
16	variación Energía Cepstral del segundo armónico (SHECM)
17	media Valor Cepstral del primer pico de los cepstrum derivados
18	variación Valor Cepstral del primer pico de los cepstrum derivados
19	media Valor Cepstral del segundo pico de los cepstrum derivados
20	variación Valor Cepstral del segundo pico de los cepstrum derivados
21	media Valor Cepstral de la diferencia entre el segundo y el primer pico de los cepstrum derivados
22	variación Valor Cepstral de la diferencia entre el segundo y el primer pico de los cepstrum derivados
23	media Calculo de la Relación Armónico- Subarmónico
24	variación Calculo de la Relación Armónico- Subarmónico
25	media Open quotient
26	variación Open quotient
27	media Speech quotient
28	variación Speech quotient
29	media Closing quotient
30	variación Closing quotient
31	media Amplitud máxima de dUg
32	variación Amplitud máxima de dUg
33	media Amplitud mínima de dUg
34	variación Amplitud mínima de dUg
35	media Margen dinámico de dUg
36	variación Margen dinámico de dUg
37	media Momento de pico normalizado
38	variación Momento de pico normalizado
39	media Momento de excitación normalizado
40	variación Momento de excitación normalizado
41	media momento de cierre normalizado
42	variación momento de cierre normalizado
43	media momento de apertura normalizado
44	variación momento de apertura normalizado
45	media PECM de la señal Ug
46	variación PECM de la señal Ug
47	media PECM del segundo armónico de la señal Ug
48	variación PECM del segundo armónico de la señal Ug
49	media Skewness de la señal residual
50	media Grado de Sonoridad por Tramas (GST)
51	variación Grado de Sonoridad por Tramas (GST)

52	media Relación de la parte no periódica a la periódica (APR)
53	variación Relación de la parte no periódica a la periódica (APR)

TABLA 12-3: MEDIDAS DE CALIDAD DE LA VOZ PERTENECIENTES AL GRUPO III (PRESENCIA DE RUIDO)

Identificador de la medida	GRUPO III: "Presencia de Ruido"
1	media energía concentrada alrededor del 1er formante
2	variación energía concentrada alrededor del 1er formante
3	media energía concentrada alrededor del 3er formante
4	variación energía concentrada alrededor del 3er formante
5	media Relación de energía entre la región 3 y la región 2
6	variación Relación de energía entre la región 3 y la región 2
7	media Relación de energía entre la región 4 y la región 1
8	variación Relación de energía entre la región 4 y la región 1
9	media Relación de energía entre la región 1 y la energía total
10	variación Relación de energía entre la región 1 y la energía total
11	media Relación de energía entre la región 2 y la energía total
12	variación Relación de energía entre la región 2 y la energía total
13	media Relación de energía entre la región 3 y la energía total
14	variación Relación de energía entre la región 3 y la energía total
15	media Relación de energía entre la región 4 y la energía total
16	variación Relación de energía entre la región 4 y la energía total
17	media Distancia de módulos entre las transformadas de Fourier (FFT) de dos tramas consecutivas
18	variación Distancia de módulos entre las transformadas de Fourier (FFT) de dos tramas consecutivas
19	media Distancia de fase entre las transformadas de Fourier (FFT) de dos tramas consecutivas
20	variación Distancia de fase entre las transformadas de Fourier (FFT) de dos tramas consecutivas
21	media Distancia basada en los LPC
22	variación Distancia basada en los LPC
23	media Distancia de máxima verosimilitud
24	variación Distancia de máxima verosimilitud
25	media energía de ruido normalizada;
26	variación energía de ruido normalizada;
27	media Relación armónico ruido cepstral
28	variación Relación armónico ruido cepstral
29	media Relación Espectral Armónico Ruido
30	variación Relación Espectral Armónico Ruido
31	media Calculo del Glotal-to-Noise Excitation (GNE)
32	variación Calculo del Glotal-to-Noise Excitation (GNE)
33	media Normalilised error prediction (NEP)
34	variación Normalilised error prediction (NEP)
35	media valor cuadrático medio de la señal residual
36	variación valor cuadrático medio de la señal residual
37	media Relación Espectral armónico ruido de la señal residual
38	variación Relación Espectral armónico ruido de la señal residual
39	media Relación Espectral Subarmónico - armónico de al señal residual
40	variación Relación Espectral Subarmónico - armónico de al señal residual
41	Relación armónica -ruido cepstral (HNRS)
42	variación Relación armónica -ruido cepstral (HNRS)

TABLA 12-4: MEDIDAS DE CALIDAD DE LA VOZ PERTENECIENTES AL GRUPO IV (COMPORTAMIENTO NO LINEAL)

Identificador de la medida	GRUPO IV: “Comportamiento no lineal”
1	media Energía de la envolvente de AM
2	variación Energía de la envolvente de AM
3	media Pendiente de la envolvente de AM
4	variación Pendiente de la envolvente de AM
5	media 1er coeficiente de la envolvente de AM
6	variación 1er coeficiente de la envolvente de AM
7	media 2o coeficiente de la envolvente de AM
8	variación 2o coeficiente de la envolvente de AM
9	media 3o coeficiente de la envolvente de AM
10	variación 3o coeficiente de la envolvente de AM
11	media Perturbación del índice de bicoherencia
12	variación Perturbación del índice de bicoherencia
13	media Área superior del biespectro integrado
14	variación Área superior del biespectro integrado
15	media Área inferior del biespectro integrado
16	variación Área inferior del biespectro integrado
17	media cuantificación del ruido estacionario (magnitud y fase)
18	variación cuantificación del ruido estacionario (magnitud y fase)
19	media cuantificación del ruido estacionario (sólo magnitud)
20	variación cuantificación del ruido estacionario (sólo magnitud)
21	media Mínima información mutua
22	variación Mínima información mutua
23	media Mínima dimensión de la correlación a partir del estimador Taken
24	variación Mínima dimensión de la correlación a partir del estimador Taken
25	media Mínima dimensión de la correlación sólo para m=2
26	variación Mínima dimensión de la correlación sólo para m=2
27	media Medida de la entropía (H1) para m=2
28	variación Medida de la entropía (H1) para m=2
29	media Medida de la entropía (H2) para m=2
30	variación Medida de la entropía (H2) para m=2
31	Entropía de Shanon

12.2 Valoración de la Relevancia de las Medidas de Calidad de la Voz

En este apartado se presentan los resultados del proceso consistente en ordenar las diferentes medidas de calidad de voz en función de su relevancia, diferenciándose los cuatro grupos de medidas de calidad de la voz descritos en apartados anteriores. El estudio se ha realizado diferenciándose dos casos: la clasificación entre “voz sana – voz patológica” y la clasificación entre “voz sana – disfonía leve – disfonía moderada – disfonía severa”. Para cada caso, se ha estudiado las medidas de calidad de la voz diferenciando la vocal fruto de la fonación.

El procedimiento de obtención de la relevancia de cada medida ha sido descrito en el apartado 5.2. Para la estimación media de la relevancia de las diferentes medidas de la calidad ha sido evaluado 100 veces y posteriormente los resultados promediados.

12.2.1 Estudio: Sano – Patológico

En este caso las muestras de voz, consisten en vocales, han sido etiquetada o como “voz sana” o como “voz patológica”.

TABLA 12-5: ORDEN EN FUNCIÓN DE LA RELEVANCIA PARA EL ESTUDIO SANO – PATOLOGÍA DE LA VOCAL “A”

Orden de relevancia	GRUPO I		GRUPO II		GRUPO III		GRUPO IV	
	Id. Medida	Prob.	Id. Medida	Prob.	Id. Medida	Prob.	Id. Medida	Prob.
1	4	61%	19	63%	39	77%	15	70%
2	5	55%	17	53%	41	57%	13	68%
3	1	45%	49	46%	4	53%	18	49%
4	17	43%	25	40%	1	53%	11	44%
5	6	43%	11	39%	3	51%	4	42%
6	8	38%	2	39%	11	49%	27	40%
7	10	37%	52	29%	8	49%	21	40%
8	9	35%	50	29%	37	45%	20	40%
9	7	33%	39	29%	29	39%	14	39%
10	3	32%	7	25%	25	39%	2	37%
11	16	30%	28	24%	7	37%	31	36%
12	2	29%	5	24%	23	36%	22	36%
13	18	27%	29	23%	17	35%	1	36%
14	14	27%	26	23%	10	35%	3	31%
15	12	26%	21	22%	6	35%	12	29%
16	13	23%	43	21%	26	32%	28	28%
17	15	21%	27	20%	15	31%	24	27%
18	11	17%	45	19%	2	31%	17	27%
19			32	19%	28	30%	16	27%
20			4	18%	20	29%	6	26%
21			41	17%	33	27%	9	25%
22			34	17%	31	27%	26	24%
23			31	17%	13	27%	23	24%
24			35	16%	19	26%	7	23%
25			22	16%	38	25%	5	21%
26			47	15%	36	25%	29	20%
27			33	15%	16	25%	8	18%
28			23	15%	12	25%	30	17%
29			12	15%	40	24%	25	17%
30			36	14%	35	24%	10	17%
31			13	14%	30	24%	19	16%
32			30	13%	18	24%		
33			24	13%	14	23%		

34			40	12%	22	21%		
35			20	12%	34	20%		
36			48	11%	42	19%		
37			16	11%	5	19%		
38			1	11%	24	18%		
39			18	10%	21	17%		
40			14	10%	9	15%		
41			46	9%	32	13%		
42			3	9%	27	9%		
43			44	7%				
44			9	7%				
45			53	6%				
46			42	6%				
47			37	6%				
48			8	6%				
49			15	5%				
50			51	4%				
51			38	2%				
52			10	1%				
53			6	0%				

TABLA 12-6: ORDEN EN FUNCIÓN DE LA RELEVANCIA PARA EL ESTUDIO SANO – PATOLOGÍA DE LA VOCAL “E”

Orden de relevancia	GRUPO I		GRUPO II		GRUPO III		GRUPO IV		
	Id.	Medida	Prob.	Id.	Medida	Prob.	Id.	Medida	Prob.
1	4	74%	19	70%	3	70%	13	80%	
2	1	48%	49	55%	1	70%	15	74%	
3	8	41%	17	40%	39	69%	20	54%	
4	6	41%	2	33%	15	55%	17	49%	
5	5	41%	7	32%	5	54%	3	47%	
6	2	40%	4	30%	41	53%	1	47%	
7	18	39%	23	27%	11	52%	11	41%	
8	9	38%	21	26%	17	49%	18	39%	
9	3	35%	11	25%	37	45%	27	37%	
10	14	32%	5	24%	13	44%	19	37%	
11	13	29%	13	22%	10	44%	2	36%	
12	16	26%	50	21%	25	41%	23	33%	
13	12	26%	32	21%	7	39%	21	33%	
14	7	24%	43	20%	31	38%	4	33%	
15	15	22%	44	17%	22	38%	25	32%	
16	11	21%	34	16%	12	37%	5	29%	
17	10	21%	12	16%	19	34%	12	27%	
18	17	17%	30	15%	8	33%	26	26%	
19			22	15%	4	33%	7	26%	
20			36	13%	21	32%	29	25%	
21			31	13%	16	31%	14	24%	
22			20	13%	24	30%	24	23%	
23			15	13%	18	30%	16	23%	
24			47	12%	14	29%	9	23%	
25			38	12%	33	28%	28	22%	

26			14	12%	23	28%	22	19%
27			28	11%	9	28%	6	18%
28			18	11%	32	27%	31	15%
29			16	11%	2	23%	30	14%
30			10	11%	29	21%	10	14%
31			9	11%	30	20%	8	13%
32			8	11%	26	20%		
33			3	11%	38	19%		
34			52	10%	6	19%		
35			42	10%	42	18%		
36			40	10%	40	17%		
37			35	10%	34	15%		
38			27	10%	35	14%		
39			25	10%	28	14%		
40			53	9%	36	13%		
41			46	9%	20	13%		
42			1	9%	27	8%		
43			45	8%				
44			26	8%				
45			24	8%				
46			48	7%				
47			33	7%				
48			41	6%				
49			39	6%				
50			51	5%				
51			37	5%				
52			29	5%				
53			6	0%				

TABLA 12-7: ORDEN EN FUNCIÓN DE LA RELEVANCIA PARA EL ESTUDIO SANO – PATOLOGÍA DE LA VOCAL ‘I’

Orden de relevancia	GRUPO I		GRUPO II		GRUPO III		GRUPO IV	
	Id. Medida	Prob.	Id. Medida	Prob.	Id. Medida	Prob.	Id. Medida	Prob.
1	3	61%	7	43%	3	92%	15	80%
2	6	55%	23	41%	1	57%	13	77%
3	12	48%	16	36%	39	51%	30	57%
4	5	48%	22	34%	11	49%	11	57%
5	13	47%	19	34%	13	48%	20	53%
6	10	45%	51	32%	7	46%	27	51%
7	16	43%	32	31%	17	44%	21	46%
8	11	43%	13	30%	18	42%	10	46%
9	14	42%	18	29%	9	41%	1	45%
10	1	41%	3	29%	37	39%	31	44%
11	15	40%	53	27%	35	35%	22	44%
12	2	40%	12	27%	19	34%	18	42%
13	9	38%	14	26%	16	34%	6	41%
14	7	36%	8	26%	10	34%	3	40%
15	4	29%	52	25%	4	34%	28	39%
16	8	26%	4	25%	33	33%	8	38%
17	17	18%	2	25%	15	33%	2	37%

18	18	16%	24	24%	8	32%	23	34%
19			9	24%	25	31%	29	33%
20			21	23%	41	29%	14	30%
21			33	22%	31	27%	16	29%
22			5	22%	29	27%	12	29%
23			36	21%	23	25%	24	28%
24			1	21%	12	24%	26	27%
25			30	20%	5	24%	4	27%
26			26	20%	42	22%	17	26%
27			46	19%	26	22%	7	26%
28			40	19%	20	20%	19	24%
29			34	19%	14	20%	25	23%
30			31	19%	40	18%	9	19%
31			28	18%	2	18%	5	15%
32			20	18%	32	16%		
33			15	18%	28	16%		
34			11	18%	22	16%		
35			35	17%	6	16%		
36			48	16%	36	15%		
37			44	16%	21	15%		
38			10	16%	38	14%		
39			45	14%	30	14%		
40			42	14%	34	13%		
41			17	14%	24	12%		
42			50	13%	27	3%		
43			38	12%				
44			37	12%				
45			47	11%				
46			27	10%				
47			49	9%				
48			43	9%				
49			39	7%				
50			29	4%				
51			41	1%				
52			25	1%				
53			6	0%				

TABLA 12-8: ORDEN EN FUNCIÓN DE LA RELEVANCIA PARA EL ESTUDIO SANO – PATOLOGÍA DE LA VOCAL “O”

Orden de relevancia	GRUPO I		GRUPO II		GRUPO III		GRUPO IV					
	Id.	Medida	Prob.	Id.	Medida	Prob.	Id.	Medida	Prob.			
1		18	72%		49	68%		39	93%		13	73%
2		1	68%		7	50%		15	53%		17	68%
3		9	58%		23	42%		13	47%		19	65%
4		2	51%		16	37%		17	45%		15	62%
5		12	45%		5	34%		4	39%		5	51%
6		15	40%		4	33%		2	39%		1	49%
7		4	39%		2	32%		16	37%		21	48%
8		14	36%		17	31%		1	36%		18	48%
9		16	35%		32	29%		12	32%		12	47%

10	6	34%	28	29%	35	30%	10	46%
11	3	34%	51	28%	19	30%	16	45%
12	7	32%	24	28%	31	28%	11	40%
13	10	27%	53	27%	11	28%	14	36%
14	11	26%	22	25%	37	27%	2	34%
15	5	25%	14	25%	8	27%	31	31%
16	13	24%	12	25%	28	26%	9	30%
17	8	18%	8	25%	14	26%	3	29%
18	17	9%	34	24%	18	25%	23	27%
19			21	23%	9	25%	20	27%
20			46	22%	30	24%	7	27%
21			36	21%	29	24%	26	26%
22			26	21%	27	24%	24	26%
23			40	20%	25	24%	8	26%
24			30	19%	7	23%	22	23%
25			20	19%	38	22%	4	23%
26			19	19%	33	22%	28	22%
27			18	18%	26	22%	27	22%
28			43	17%	22	22%	6	19%
29			31	16%	6	22%	25	15%
30			15	16%	40	21%	30	14%
31			11	16%	3	21%	29	14%
32			9	16%	41	20%		
33			48	15%	36	20%		
34			35	15%	42	19%		
35			33	15%	10	18%		
36			3	15%	5	18%		
37			42	14%	23	17%		
38			10	14%	21	16%		
39			1	14%	34	15%		
40			44	13%	32	15%		
41			13	12%	24	15%		
42			27	11%	20	15%		
43			29	10%				
44			52	9%				
45			25	9%				
46			47	8%				
47			50	7%				
48			37	7%				
49			45	6%				
50			38	6%				
51			41	4%				
52			39	4%				
53			6	0%				

TABLA 12-9: ORDEN EN FUNCIÓN DE LA RELEVANCIA PARA EL ESTUDIO SANO – PATOLOGÍA DE LA VOCAL “U”

Orden de relevancia	GRUPO I		GRUPO II		GRUPO III		GRUPO IV	
	Id. Medida	Prob.	Id. Medida	Prob.	Id. Medida	Prob.	Id. Medida	Prob.
1	5	65%	2	54%	39	74%	27	76%
2	4	61%	15	39%	19	56%	21	61%
3	18	57%	40	36%	5	53%	15	60%
4	6	53%	13	34%	6	50%	14	58%
5	2	48%	11	34%	29	49%	13	56%
6	9	44%	5	34%	10	45%	2	56%
7	15	43%	7	32%	25	43%	3	54%
8	8	43%	49	31%	11	43%	25	49%
9	13	39%	8	30%	34	42%	11	43%
10	11	39%	27	29%	13	42%	31	42%
11	3	37%	20	29%	32	40%	7	40%
12	12	36%	44	28%	28	40%	16	39%
13	16	33%	14	28%	8	40%	10	39%
14	10	29%	12	28%	37	39%	5	39%
15	1	25%	28	26%	7	39%	1	37%
16	14	23%	24	26%	26	38%	4	36%
17	7	21%	23	26%	17	38%	26	35%
18	17	14%	18	26%	21	37%	9	35%
19			16	26%	23	35%	29	34%
20			47	25%	9	35%	28	34%
21			45	25%	1	34%	23	34%
22			50	24%	12	33%	22	34%
23			32	24%	15	31%	24	33%
24			19	24%	14	31%	19	32%
25			30	23%	2	31%	18	32%
26			51	22%	18	30%	6	30%
27			48	21%	4	30%	12	29%
28			46	21%	33	29%	20	26%
29			38	21%	30	29%	8	26%
30			36	21%	16	27%	30	25%
31			4	21%	22	26%	17	25%
32			53	20%	42	25%		
33			42	20%	3	25%		
34			26	20%	40	23%		
35			35	19%	24	23%		
36			22	19%	36	22%		
37			34	18%	20	21%		
38			31	18%	38	20%		
39			9	18%	41	18%		
40			3	18%	27	18%		
41			52	17%	31	10%		
42			33	17%	35	9%		
43			37	16%				
44			43	15%				
45			29	13%				
46			17	13%				
47			1	12%				

48			39	10%				
49			25	10%				
50			10	10%				
51			41	9%				
52			21	8%				
53			6	0%				

12.2.2 Estudio: Sano – Disfonía Leve – Disfonía Moderada – Disfonía Severa

En este caso las muestras de voz, al igual que en el apartado anterior consisten en vocales y han sido denominadas con una de las siguientes etiquetas: “voz sana”, “disfonía leve”, “disfonía moderada” o “disfonía severa”. Este etiquetado ha sido realizado atendiendo al grado de disfonía.

TABLA 12-10: ORDEN EN FUNCIÓN DE LA RELEVANCIA PARA EL ESTUDIO SANO – DISFONÍA LEVE – DISFONÍA MODERADA – DISFONÍA SEVERA DE LA VOCAL “A”

Orden de relevancia	GRUPO I		GRUPO II		GRUPO III		GRUPO IV	
	Id. Medida	Prob.	Id. Medida	Prob.	Id. Medida	Prob.	Id. Medida	Prob.
1	5	60%	19	66%	39	72%	11	74%
2	3	51%	17	53%	1	72%	13	69%
3	7	44%	7	43%	41	69%	1	67%
4	1	44%	2	40%	3	58%	15	63%
5	6	42%	23	39%	4	55%	3	60%
6	10	40%	32	29%	11	51%	21	49%
7	18	38%	25	29%	10	51%	31	47%
8	2	38%	49	27%	7	51%	22	47%
9	14	35%	3	27%	37	47%	18	43%
10	4	34%	13	26%	2	46%	30	41%
11	9	29%	9	26%	16	44%	26	40%
12	17	26%	4	25%	15	44%	2	38%
13	8	24%	45	24%	20	41%	12	37%
14	15	23%	21	24%	17	41%	24	34%
15	16	21%	28	23%	8	41%	9	34%
16	13	21%	11	23%	9	40%	20	33%
17	12	20%	50	22%	18	39%	4	33%
18	11	19%	39	22%	13	39%	28	32%
19			12	22%	42	38%	19	31%
20			52	21%	28	38%	17	30%
21			34	20%	26	38%	23	29%
22			22	20%	6	38%	8	29%
23			18	20%	40	37%	10	28%
24			33	19%	25	37%	29	27%
25			5	19%	38	35%	14	27%
26			10	18%	34	35%	6	27%

27			31	17%	14	33%	27	26%
28			24	17%	35	30%	16	25%
29			30	16%	30	30%	5	19%
30			47	15%	31	28%	25	17%
31			44	15%	29	27%	7	15%
32			16	15%	33	26%		
33			8	15%	19	25%		
34			43	14%	36	24%		
35			40	14%	23	22%		
36			26	14%	12	22%		
37			53	13%	32	21%		
38			46	13%	5	21%		
39			36	13%	22	19%		
40			29	13%	21	13%		
41			14	13%	24	11%		
42			42	12%	27	8%		
43			15	12%				
44			51	11%				
45			48	11%				
46			35	11%				
47			20	11%				
48			27	9%				
49			1	6%				
50			41	5%				
51			38	3%				
52			37	2%				
53			6	0%				

TABLA 12-11: ORDEN EN FUNCIÓN DE LA RELEVANCIA PARA EL ESTUDIO SANO – DISFONÍA LEVE – DISFONÍA MODERADA – DISFONÍA SEVERA DE LA VOCAL “E”

Orden de relevancia	GRUPO I		GRUPO II		GRUPO III		GRUPO IV	
	Id.	Medida	Prob.	Id.	Medida	Prob.	Id.	Medida
1	4	86%	49	64%	1	70%	13	84%
2	6	65%	17	50%	41	65%	15	77%
3	8	63%	2	47%	3	63%	1	66%
4	5	46%	19	44%	39	62%	20	62%
5	9	45%	10	44%	7	52%	18	60%
6	1	36%	21	43%	5	52%	3	59%
7	2	35%	7	43%	13	51%	9	46%
8	10	28%	4	40%	37	50%	11	44%
9	3	26%	5	39%	10	49%	21	42%
10	14	24%	52	37%	15	48%	31	38%
11	16	23%	23	36%	25	44%	23	38%
12	7	21%	50	34%	22	44%	4	37%
13	12	20%	34	32%	8	44%	19	36%
14	17	19%	12	29%	29	41%	7	36%
15	18	18%	36	28%	9	41%	6	35%
16	15	18%	25	28%	21	40%	25	34%
17	13	17%	43	24%	14	39%	17	34%

18	11	10%	42	23%	11	38%	2	32%
19			40	23%	31	36%	26	30%
20			16	23%	19	35%	8	29%
21			9	23%	16	35%	27	27%
22			44	22%	23	34%	5	27%
23			30	22%	17	34%	29	26%
24			8	22%	35	33%	12	26%
25			48	21%	4	33%	30	25%
26			26	21%	33	32%	24	25%
27			24	21%	12	29%	14	23%
28			22	21%	32	27%	10	23%
29			32	20%	40	25%	28	19%
30			20	20%	20	25%	16	16%
31			11	20%	36	23%	22	15%
32			38	19%	24	23%		
33			28	19%	30	22%		
34			18	19%	28	22%		
35			13	19%	6	22%		
36			46	18%	18	21%		
37			39	16%	34	20%		
38			35	16%	26	20%		
39			29	15%	42	18%		
40			53	14%	38	18%		
41			51	14%	2	15%		
42			33	14%	27	3%		
43			3	14%				
44			1	14%				
45			27	13%				
46			14	13%				
47			31	11%				
48			47	9%				
49			45	9%				
50			15	7%				
51			41	6%				
52			37	5%				
53			6	0%				

TABLA 12-12: ORDEN EN FUNCIÓN DE LA RELEVANCIA PARA EL ESTUDIO SANO – DISFONÍA LEVE – DISFONÍA MODERADA – DISFONÍA SEVERA DE LA VOCAL “I”

Orden de relevancia	GRUPO I		GRUPO II		GRUPO III		GRUPO IV	
	Id. Medida	Prob.	Id. Medida	Prob.	Id. Medida	Prob.	Id. Medida	Prob.
1	1	62%	7	64%	3	86%	15	74%
2	15	59%	23	46%	39	73%	13	72%
3	6	55%	9	45%	7	60%	11	71%
4	4	53%	5	43%	9	56%	30	64%
5	7	47%	12	41%	11	47%	27	52%
6	12	45%	18	39%	15	46%	23	51%
7	5	45%	22	38%	1	44%	28	46%
8	3	45%	8	36%	17	43%	29	45%

9	16	41%	16	35%	13	42%	21	45%
10	14	41%	13	35%	41	41%	4	42%
11	11	41%	3	35%	25	41%	31	39%
12	13	40%	32	34%	33	38%	20	39%
13	8	37%	4	34%	28	37%	1	38%
14	10	35%	34	33%	19	35%	3	37%
15	18	33%	14	32%	26	34%	18	36%
16	2	33%	51	31%	29	33%	12	36%
17	9	25%	46	31%	12	33%	9	36%
18	17	24%	52	30%	37	30%	26	34%
19			20	30%	18	30%	17	34%
20			1	30%	8	30%	10	34%
21			44	29%	42	28%	8	34%
22			48	28%	30	27%	5	34%
23			31	28%	38	26%	2	33%
24			30	28%	40	25%	14	30%
25			26	28%	4	25%	22	29%
26			53	27%	20	24%	19	29%
27			24	27%	31	23%	16	29%
28			15	27%	10	21%	7	29%
29			36	26%	5	21%	6	29%
30			28	24%	21	19%	25	21%
31			33	22%	23	18%	24	17%
32			17	22%	34	14%		
33			19	21%	16	14%		
34			40	20%	35	12%		
35			35	20%	22	12%		
36			29	18%	2	12%		
37			50	17%	32	10%		
38			42	17%	24	10%		
39			10	17%	36	7%		
40			37	14%	14	7%		
41			21	14%	6	6%		
42			11	14%	27	5%		
43			45	13%				
44			39	12%				
45			2	11%				
46			27	10%				
47			47	9%				
48			38	9%				
49			49	7%				
50			41	7%				
51			25	6%				
52			43	3%				
53			6	0%				

TABLA 12-13: ORDEN EN FUNCIÓN DE LA RELEVANCIA PARA EL ESTUDIO SANO – DISFONÍA LEVE – DISFONÍA MODERADA – DISFONÍA SEVERA DE LA VOCAL “O”

Orden de relevancia	GRUPO I		GRUPO II		GRUPO III		GRUPO IV	
	Id. Medida	Prob.	Id. Medida	Prob.	Id. Medida	Prob.	Id. Medida	Prob.
1	1	71%	49	76%	39	79%	17	78%
2	9	65%	23	53%	17	77%	15	69%
3	15	60%	7	51%	19	54%	13	69%
4	18	51%	11	42%	3	53%	21	67%
5	2	51%	17	41%	15	49%	31	58%
6	7	47%	2	41%	4	48%	19	55%
7	6	41%	16	34%	11	47%	7	48%
8	11	37%	8	34%	16	44%	1	48%
9	3	37%	46	32%	2	44%	5	46%
10	13	35%	43	31%	26	42%	11	45%
11	4	33%	35	31%	14	40%	3	44%
12	17	32%	21	31%	42	39%	12	42%
13	16	31%	12	31%	28	39%	28	39%
14	14	31%	9	31%	18	39%	18	35%
15	12	30%	32	30%	12	39%	14	35%
16	10	23%	4	30%	38	38%	9	33%
17	5	23%	40	29%	37	38%	29	31%
18	8	18%	14	29%	30	36%	27	31%
19			30	28%	13	36%	30	30%
20			28	28%	35	34%	24	27%
21			50	27%	40	32%	4	27%
22			44	27%	29	30%	20	26%
23			20	27%	22	30%	2	26%
24			15	27%	41	29%	16	25%
25			5	27%	24	29%	25	23%
26			26	26%	27	28%	23	23%
27			19	26%	25	28%	10	22%
28			24	25%	1	28%	26	17%
29			48	24%	31	27%	8	17%
30			18	24%	5	25%	6	17%
31			36	23%	9	24%	22	15%
32			34	23%	8	24%		
33			53	22%	33	23%		
34			31	22%	32	23%		
35			51	20%	21	23%		
36			25	20%	7	20%		
37			10	19%	6	19%		
38			45	18%	23	18%		
39			22	18%	36	16%		
40			42	17%	20	16%		
41			33	17%	10	16%		
42			13	17%	34	15%		
43			39	15%				
44			29	15%				
45			47	14%				
46			38	14%				

47			27	14%				
48			3	14%				
49			52	13%				
50			1	13%				
51			41	6%				
52			37	4%				
53			6	0%				

TABLA 12-14: ORDEN EN FUNCIÓN DE LA RELEVANCIA PARA EL ESTUDIO SANO – DISFONÍA LEVE – DISFONÍA MODERADA – DISFONÍA SEVERA DE LA VOCAL “U”

Orden de relevancia	GRUPO I		GRUPO II		GRUPO III		GRUPO IV	
	Id. Medida	Prob.	Id. Medida	Prob.	Id. Medida	Prob.	Id. Medida	Prob.
1	4	67%	2	68%	39	67%	27	68%
2	5	65%	12	42%	17	58%	21	63%
3	6	58%	40	40%	13	51%	15	56%
4	18	54%	7	40%	8	47%	13	55%
5	9	53%	32	38%	37	46%	14	53%
6	15	52%	28	37%	29	46%	6	51%
7	11	49%	14	36%	25	44%	19	47%
8	12	40%	31	35%	15	44%	24	43%
9	2	38%	30	35%	6	41%	11	42%
10	3	36%	23	35%	5	41%	2	42%
11	13	35%	5	35%	32	40%	8	41%
12	16	34%	35	34%	41	39%	3	41%
13	8	33%	8	33%	19	39%	23	40%
14	1	33%	13	32%	4	38%	20	40%
15	10	32%	18	31%	23	37%	1	40%
16	7	29%	15	31%	1	37%	16	39%
17	14	26%	9	31%	28	36%	5	37%
18	17	22%	19	30%	9	35%	4	37%
19			4	30%	7	35%	28	36%
20			42	29%	30	34%	29	35%
21			24	29%	10	34%	25	35%
22			27	28%	21	33%	9	35%
23			51	27%	12	33%	26	34%
24			44	27%	26	31%	17	33%
25			33	27%	14	31%	31	32%
26			20	27%	3	31%	30	32%
27			53	26%	38	30%	18	26%
28			50	25%	11	30%	12	25%
29			46	25%	18	29%	10	25%
30			36	25%	40	28%	7	25%
31			22	25%	2	28%	22	24%
32			26	24%	42	26%		
33			3	24%	16	26%		
34			49	23%	33	25%		
35			48	22%	24	24%		
36			45	21%	36	22%		
37			16	21%	27	20%		

38			47	20%	22	20%		
39			34	20%	20	19%		
40			37	19%	34	16%		
41			10	19%	31	10%		
42			39	17%	35	5%		
43			11	15%				
44			1	14%				
45			29	13%				
46			17	12%				
47			38	8%				
48			25	8%				
49			21	7%				
50			41	6%				
51			52	4%				
52			43	4%				
53			6	4%				

12.3 Evaluación de las Medidas de Calidad de la Voz

En este apartado se han evaluado las medidas de calidad de la voz de forma secuencial, siguiendo el orden descendente marcado por la relevancia de la medida, siendo ademásb evaluadas en diferentes clasificadores: redes neuronales probabilísticas (PNN), redes neuronales *feedforward* (NN) con una capa oculta, clasificadores basados en funciones de base radial, máquinas de vectores soporte (SVM) y clasificadores basados en los K vecinos más cercanos (KNN).

Para cada subconjunto de medidas de cada grupo y cada vocal, se obtiene la tasa de éxito en la clasificación dado un determinado clasificador. Los resultados en cada caso son promediados 20 veces.

12.3.1 Estudio: Sano – Patológico

12.3.1.1 Red Neuronal Probabilística (PNN)

Grupo I

Núm. de Medidas	vocal 'a'	vocal 'e'	vocal 'i'	vocal 'o'	vocal 'u'
1	68,70%	65,83%	56,76%	59,35%	63,61%
2	69,07%	67,50%	60,46%	64,81%	66,48%
3	70,19%	66,85%	62,04%	64,81%	66,76%
4	71,02%	67,41%	61,94%	66,39%	63,43%
5	71,02%	70,74%	65,74%	69,54%	65,56%
6	71,20%	70,83%	65,83%	72,96%	66,30%
7	74,26%	71,20%	66,67%	70,28%	66,11%
8	73,70%	68,89%	65,83%	72,50%	66,76%
9	75,74%	72,04%	67,41%	74,07%	67,22%
10	75,00%	71,67%	68,52%	74,26%	65,93%
11	77,13%	77,50%	69,91%	75,09%	67,22%
12	74,44%	74,17%	69,07%	75,09%	69,44%
13	73,98%	72,41%	65,93%	74,17%	70,46%
14	75,65%	74,72%	67,59%	73,89%	66,48%
15	70,37%	73,52%	65,28%	72,78%	68,89%
16	69,81%	74,35%	69,81%	72,22%	69,07%
17	73,70%	77,41%	66,11%	72,13%	69,35%
18	80,00%	80,83%	64,26%	71,48%	64,17%

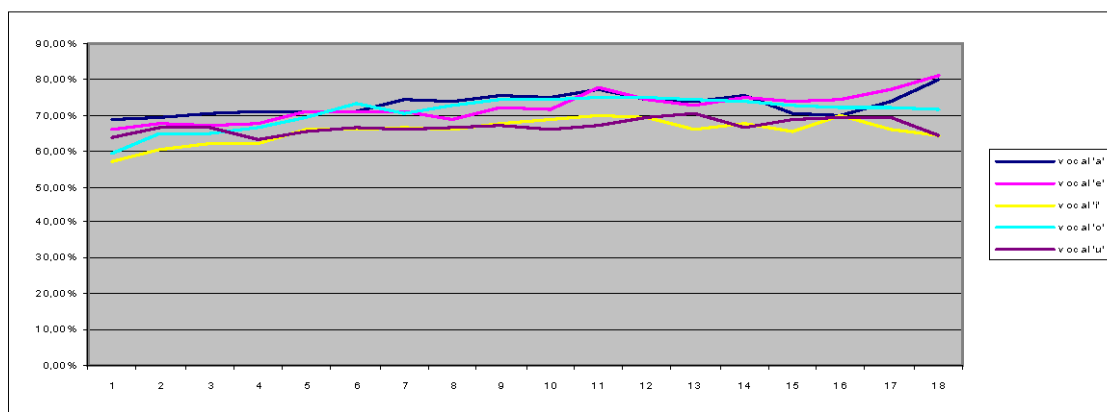


Gráfico 12-1: Evaluación secuencial de la relevancia para las medidas de calidad de la voz del grupo I usando el clasificador PNN en la clasificación entre voz sana y voz patológica

Grupo II

Núm. de Medidas	vocal 'a'	vocal 'e'	vocal 'i'	vocal 'o'	vocal 'u'
1	82,59%	76,57%	70,00%	69,44%	64,07%
2	80,65%	77,22%	70,19%	64,91%	65,28%
3	84,44%	77,04%	69,26%	68,61%	67,59%
4	82,50%	80,19%	68,70%	67,96%	67,50%
5	82,69%	76,11%	68,52%	68,70%	67,04%
6	82,69%	76,30%	69,44%	67,59%	66,11%
7	82,31%	79,44%	69,17%	66,02%	66,57%
8	83,06%	80,93%	70,65%	66,67%	68,24%
9	83,15%	79,44%	71,85%	68,24%	65,37%
10	84,63%	79,72%	72,04%	67,69%	63,98%
11	83,98%	78,33%	73,15%	68,61%	67,69%
12	83,33%	79,72%	71,57%	66,48%	67,13%
13	82,96%	78,33%	73,33%	69,17%	66,20%
14	85,19%	78,61%	72,96%	69,54%	69,81%
15	83,70%	78,61%	72,04%	69,35%	69,91%
16	85,83%	78,98%	71,20%	69,72%	70,37%
17	84,91%	80,74%	72,04%	71,85%	70,28%
18	83,89%	81,57%	73,61%	73,70%	70,37%
19	84,72%	80,65%	74,54%	71,94%	69,35%
20	86,11%	81,11%	72,69%	70,74%	71,02%
21	83,24%	81,30%	72,31%	71,20%	71,94%
22	86,30%	81,02%	74,17%	73,06%	71,85%
23	84,63%	81,30%	72,13%	73,33%	70,09%
24	84,26%	81,94%	73,52%	75,65%	71,85%
25	85,19%	80,65%	72,31%	77,04%	70,28%
26	85,56%	81,85%	74,91%	78,33%	71,11%
27	85,74%	82,04%	74,44%	75,83%	70,74%
28	84,26%	81,48%	73,24%	72,13%	69,72%
29	84,81%	83,43%	71,85%	74,54%	69,26%
30	85,09%	81,94%	71,94%	74,63%	71,11%
31	86,48%	85,28%	73,24%	75,93%	68,43%
32	85,00%	83,70%	73,80%	76,30%	68,80%
33	86,67%	82,41%	74,72%	77,78%	68,15%
34	85,74%	83,61%	70,65%	76,02%	69,44%
35	86,02%	83,06%	72,31%	75,00%	67,59%
36	86,11%	83,15%	72,59%	80,93%	66,94%
37	86,02%	85,00%	72,50%	78,24%	67,22%
38	85,74%	85,00%	69,26%	79,81%	70,83%
39	85,83%	84,44%	71,48%	78,15%	68,70%
40	89,17%	85,19%	70,00%	80,83%	67,22%
41	88,33%	83,15%	74,07%	81,02%	71,30%
42	89,17%	83,24%	73,43%	79,17%	70,00%
43	90,56%	81,76%	73,61%	78,43%	68,15%
44	88,06%	81,57%	73,80%	79,81%	67,31%
45	89,35%	83,61%	72,22%	80,93%	64,35%
46	87,96%	83,70%	73,43%	80,93%	62,50%
47	88,24%	81,02%	72,22%	72,41%	66,67%
48	86,02%	82,59%	72,13%	74,07%	66,57%

49	82,78%	82,59%	73,24%	76,11%	68,06%
50	84,81%	82,13%	66,76%	75,37%	67,31%
51	81,76%	86,30%	66,76%	74,91%	69,63%
52	82,59%	83,43%	66,57%	74,63%	68,98%
53	81,94%	81,02%	63,80%	67,87%	69,17%

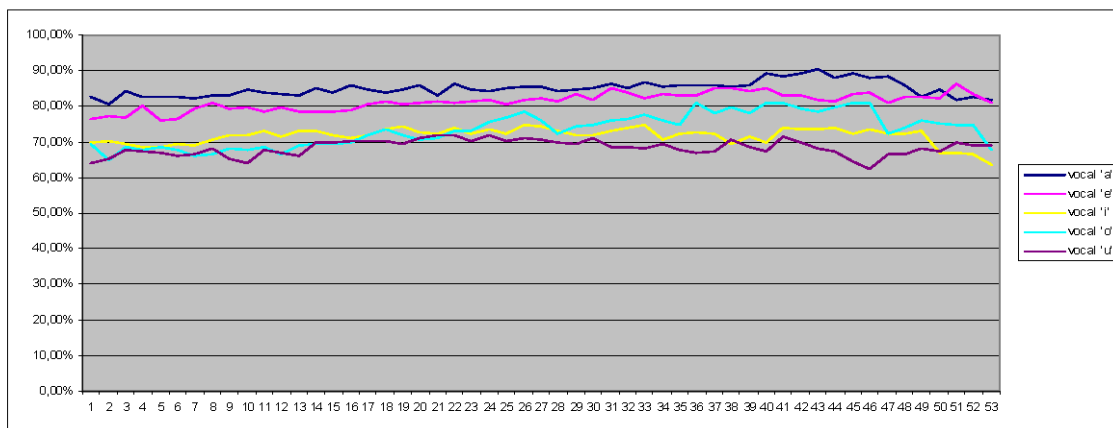


Gráfico 12-2: Evaluación secuencial de la relevancia para las medidas de calidad de la voz del grupo II usando el clasificador PNN en la clasificación entre voz sana y voz patológica

Grupo III

Núm. de Medidas	vocal 'a'	vocal 'e'	vocal 'i'	vocal 'o'	vocal 'u'
1	49,44%	50,46%	47,04%	47,50%	48,61%
2	48,61%	50,19%	46,30%	48,70%	48,61%
3	48,98%	50,19%	46,94%	49,91%	47,96%
4	49,81%	52,13%	48,33%	48,52%	48,15%
5	50,28%	51,11%	48,24%	48,98%	48,89%
6	50,46%	52,87%	48,98%	50,37%	48,70%
7	53,06%	53,70%	48,98%	48,89%	49,72%
8	50,09%	51,94%	48,89%	49,91%	49,63%
9	53,70%	52,78%	49,81%	52,22%	49,35%
10	55,28%	56,48%	52,87%	50,46%	48,70%
11	53,33%	55,83%	52,69%	54,44%	50,74%
12	56,11%	55,65%	55,65%	51,57%	51,94%
13	55,93%	57,59%	54,44%	53,70%	52,22%
14	58,98%	58,24%	54,44%	53,52%	52,50%
15	57,59%	58,80%	55,93%	53,89%	49,72%
16	58,33%	59,44%	59,07%	56,85%	53,61%
17	60,93%	62,59%	59,63%	55,19%	56,39%
18	61,67%	62,22%	62,87%	57,04%	57,22%
19	61,20%	60,83%	64,26%	56,94%	56,94%
20	61,02%	61,39%	66,11%	56,85%	56,20%
21	66,11%	63,70%	68,33%	55,56%	56,76%
22	63,33%	65,83%	68,15%	58,15%	59,26%
23	65,83%	67,04%	68,98%	59,17%	60,93%
24	67,04%	70,74%	69,17%	59,81%	59,44%
25	68,98%	73,43%	69,44%	60,56%	62,04%

26	71,20%	73,89%	72,69%	63,61%	61,39%
27	72,87%	76,57%	73,52%	64,35%	57,69%
28	72,41%	77,87%	72,22%	68,43%	58,98%
29	73,52%	78,24%	71,85%	68,61%	61,30%
30	69,72%	77,69%	73,43%	70,37%	60,00%
31	74,44%	78,98%	73,80%	73,15%	60,46%
32	72,13%	77,31%	75,28%	71,02%	59,35%
33	79,17%	79,17%	76,76%	75,74%	61,57%
34	80,28%	82,04%	78,15%	76,85%	64,17%
35	79,91%	78,24%	78,24%	75,93%	60,46%
36	79,81%	80,28%	77,04%	72,22%	63,24%
37	80,19%	79,91%	79,91%	75,65%	61,11%
38	80,37%	80,65%	77,69%	75,09%	59,44%
39	75,46%	79,07%	80,74%	76,30%	61,02%
40	75,83%	83,43%	76,57%	72,96%	67,96%
41	78,15%	68,33%	70,83%	76,20%	61,57%
42	82,41%	66,39%	72,69%	78,24%	62,69%

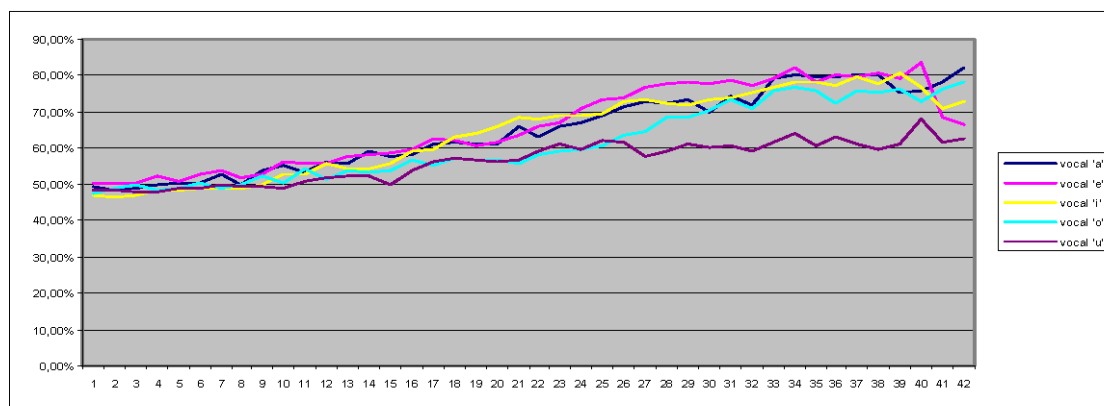


Gráfico 12-3: Evaluación secuencial de la relevancia para las medidas de calidad de la voz del grupo III usando el clasificador PNN en la clasificación entre voz sana y voz patológica

Grupo IV

Núm. de Medidas	vocal 'a'	vocal 'e'	Vocal 'i'	Vocal 'o'	vocal 'u'
1	54,54%	47,59%	50,93%	50,19%	56,39%
2	53,98%	46,57%	49,17%	52,59%	56,85%
3	57,22%	48,15%	50,56%	52,78%	58,15%
4	59,07%	49,54%	52,96%	53,33%	56,11%
5	60,56%	50,46%	49,44%	54,72%	57,31%
6	60,83%	54,17%	51,67%	57,41%	58,43%
7	63,24%	58,15%	52,59%	56,48%	58,43%
8	65,09%	61,11%	52,41%	59,07%	62,96%
9	63,52%	61,85%	55,09%	60,93%	61,94%
10	66,85%	63,61%	55,00%	66,02%	65,19%
11	69,26%	67,50%	58,52%	68,15%	62,41%
12	69,44%	71,20%	62,13%	68,06%	63,15%
13	70,56%	70,19%	60,93%	69,54%	65,19%

14	72,41%	72,41%	63,70%	73,98%	63,70%
15	74,54%	74,81%	66,57%	74,44%	64,91%
16	73,98%	75,28%	67,13%	76,94%	67,13%
17	79,07%	77,04%	66,48%	77,50%	70,00%
18	83,43%	77,59%	68,89%	79,44%	69,35%
19	82,96%	79,63%	70,56%	79,26%	69,91%
20	82,59%	79,17%	72,59%	80,19%	70,28%
21	83,61%	81,57%	73,98%	83,06%	72,78%
22	85,65%	80,37%	71,30%	79,44%	71,20%
23	85,83%	80,46%	72,50%	79,26%	70,83%
24	87,59%	81,11%	71,67%	79,91%	69,26%
25	86,39%	80,37%	74,17%	79,26%	67,87%
26	85,28%	80,74%	72,04%	75,28%	66,85%
27	84,91%	83,15%	70,56%	75,93%	71,11%
28	81,85%	81,30%	71,11%	75,28%	68,98%
29	86,20%	81,20%	71,57%	74,26%	71,57%
30	83,89%	81,39%	76,94%	72,87%	65,65%
31	85,56%	80,37%	76,67%	78,52%	61,48%

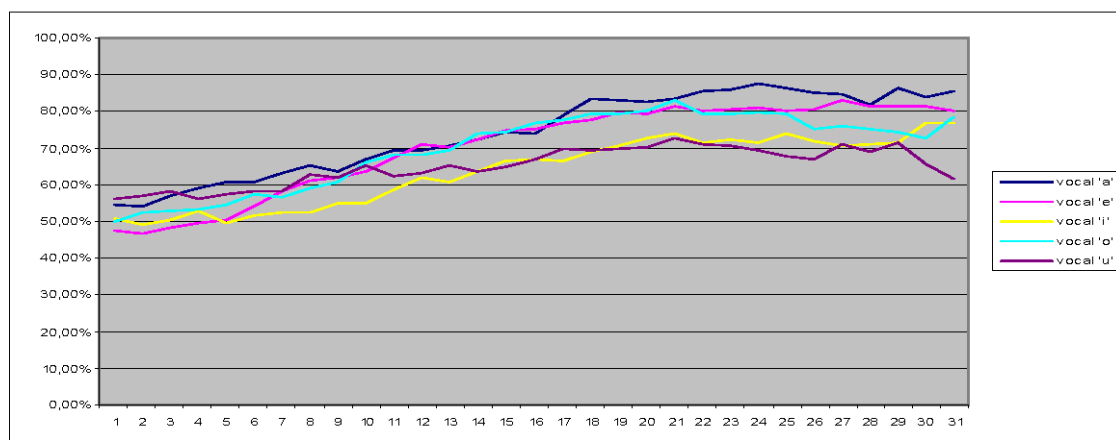


Gráfico 12-4: Evaluación secuencial de la relevancia para las medidas de calidad de la voz del grupo IV usando el clasificador PNN en la clasificación entre voz sana y voz patológica

12.3.1.2 Red Neuronal *FeedForward* con 1 Capa Oculta (NN)

Grupo I

Núm. de Medidas	vocal 'a'	vocal 'e'	vocal 'i'	vocal 'o'	vocal 'u'
1	78,52%	77,50%	69,72%	74,91%	68,80%
2	79,26%	77,22%	68,61%	74,63%	68,89%
3	78,06%	75,46%	71,76%	74,63%	68,43%
4	78,70%	80,09%	72,22%	73,24%	68,43%
5	78,70%	74,72%	71,39%	74,17%	65,37%
6	79,72%	77,78%	73,33%	73,80%	69,81%
7	78,15%	77,59%	73,80%	72,13%	69,54%
8	78,80%	76,76%	71,02%	71,57%	67,69%

9	78,33%	79,07%	73,43%	73,06%	69,35%
10	78,70%	78,89%	74,63%	72,96%	67,78%
11	80,65%	79,26%	72,04%	73,06%	69,72%
12	78,24%	74,17%	73,61%	73,89%	64,81%
13	77,22%	74,63%	71,85%	74,54%	67,59%
14	76,20%	74,26%	72,69%	74,63%	66,85%
15	77,50%	76,02%	66,67%	72,96%	68,52%
16	77,96%	78,98%	71,11%	74,35%	67,31%
17	77,69%	75,74%	65,65%	72,31%	67,50%
18	76,57%	77,96%	63,98%	67,78%	60,46%

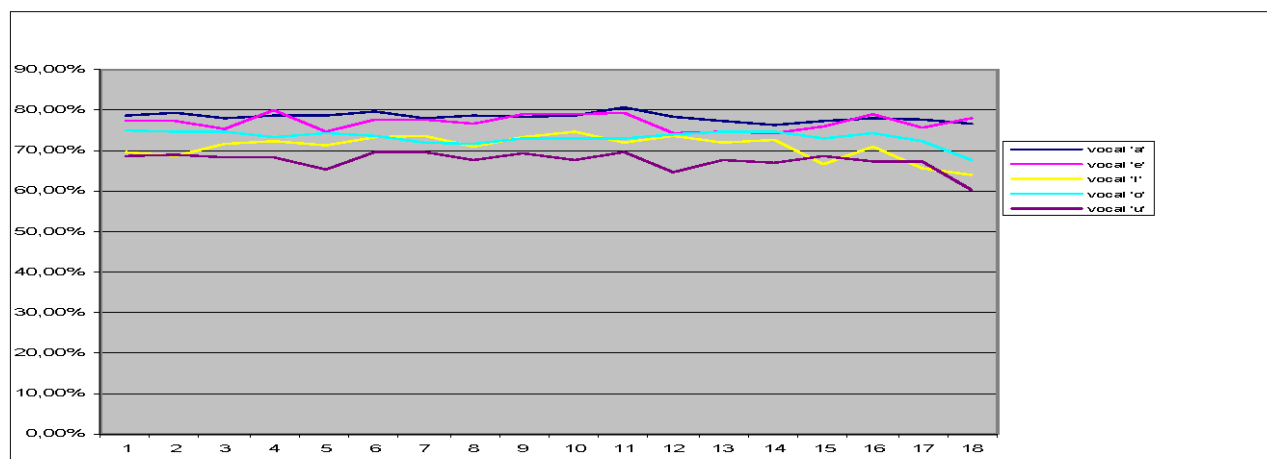


Gráfico 12-5: Evaluación secuencial de la relevancia para las medidas de calidad de la voz del grupo I usando el clasificador NN en la clasificación entre voz sana y voz patológica

Grupo II

Núm. de Medidas	vocal 'a'	Vocal 'e'	vocal 'i'	Vocal 'o'	vocal 'u'
1	86,85%	82,87%	74,54%	83,06%	71,39%
2	86,67%	85,37%	73,89%	79,07%	72,31%
3	85,83%	84,54%	76,67%	81,11%	71,57%
4	87,22%	84,35%	76,85%	81,85%	71,20%
5	87,50%	80,00%	75,28%	81,48%	68,52%
6	86,48%	82,87%	74,91%	80,00%	72,87%
7	88,98%	85,83%	78,15%	83,70%	71,85%
8	85,83%	84,07%	73,89%	80,28%	69,91%
9	87,13%	81,30%	75,65%	82,50%	70,83%
10	87,04%	84,07%	72,69%	82,41%	71,48%
11	87,04%	83,33%	74,63%	82,31%	74,26%
12	86,30%	84,17%	77,31%	81,94%	71,57%
13	87,04%	84,26%	77,04%	81,57%	72,69%
14	87,87%	84,91%	75,37%	79,17%	73,98%
15	87,87%	85,93%	76,30%	82,41%	71,76%
16	87,04%	82,13%	76,30%	81,57%	73,52%
17	86,85%	85,65%	77,22%	83,33%	71,76%
18	86,76%	84,26%	74,81%	82,31%	74,35%

19	85,93%	85,83%	74,35%	83,43%	72,50%
20	87,13%	85,00%	76,94%	80,65%	74,07%
21	87,31%	85,09%	75,46%	83,61%	74,26%
22	87,41%	85,00%	76,76%	82,31%	75,09%
23	86,48%	85,93%	76,85%	81,02%	72,22%
24	87,87%	84,26%	78,61%	82,87%	74,07%
25	86,94%	84,44%	76,57%	83,43%	74,26%
26	88,06%	86,39%	76,67%	82,04%	74,17%
27	86,94%	83,33%	74,54%	81,02%	71,30%
28	87,04%	87,04%	76,85%	82,87%	72,13%
29	87,13%	85,83%	77,13%	83,52%	75,00%
30	87,87%	85,83%	76,11%	81,76%	73,89%
31	87,96%	84,72%	75,93%	80,74%	74,63%
32	86,57%	85,83%	77,04%	82,87%	71,30%
33	87,78%	84,35%	75,46%	83,06%	71,76%
34	86,11%	85,46%	75,74%	81,85%	73,06%
35	87,78%	84,44%	77,50%	83,15%	70,46%
36	88,43%	84,72%	76,02%	84,44%	72,96%
37	87,69%	85,56%	76,94%	82,87%	73,15%
38	87,69%	86,67%	73,33%	83,06%	71,02%
39	88,52%	84,72%	70,09%	82,13%	71,11%
40	88,61%	86,57%	70,09%	83,33%	74,44%
41	89,07%	86,48%	70,19%	83,70%	73,52%
42	89,17%	87,13%	69,44%	84,72%	72,13%
43	89,26%	87,13%	69,81%	83,15%	74,44%
44	88,15%	85,46%	70,09%	82,69%	73,61%
45	88,89%	84,81%	70,00%	82,50%	71,85%
46	90,19%	86,57%	68,98%	82,96%	70,83%
47	87,13%	84,81%	70,00%	76,85%	70,09%
48	87,04%	85,19%	69,63%	76,67%	70,46%
49	86,94%	86,76%	69,26%	76,11%	72,41%
50	87,31%	85,28%	68,06%	76,20%	72,50%
51	86,20%	85,65%	63,61%	76,11%	68,33%
52	83,61%	86,02%	66,94%	75,74%	67,31%
53	82,69%	77,96%	65,65%	67,78%	69,07%

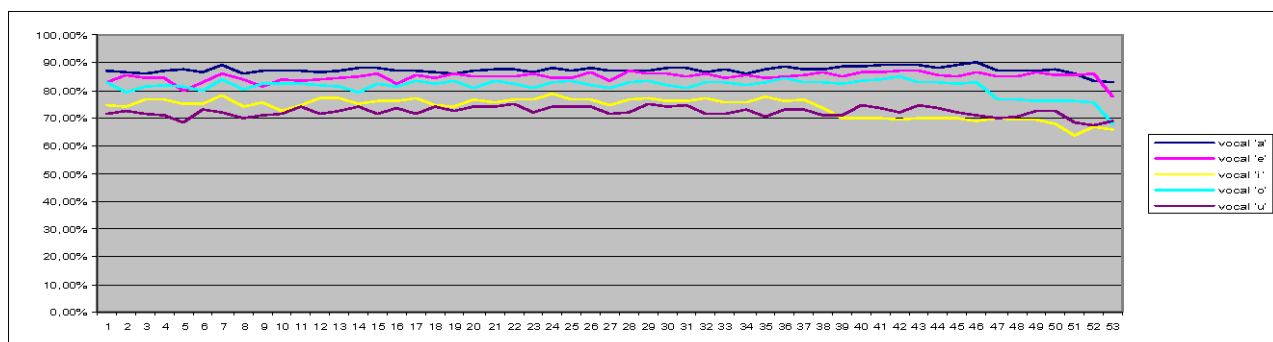


Gráfico 12-6: Evaluación secuencial de la relevancia para las medidas de calidad de la voz del grupo II usando el clasificador NN en la clasificación entre voz sana y voz patológica

Grupo III

Núm. de Medidas	vocal 'a'	vocal 'e'	vocal 'i'	vocal 'o'	vocal 'u'
1	76,11%	77,78%	75,00%	74,63%	64,26%
2	76,67%	75,93%	73,80%	74,63%	63,33%
3	74,81%	77,78%	73,98%	74,35%	63,70%
4	76,67%	78,06%	76,76%	76,11%	66,76%
5	77,69%	76,11%	73,80%	76,85%	62,69%
6	77,31%	77,04%	75,93%	73,52%	64,81%
7	77,59%	78,33%	77,41%	78,43%	64,44%
8	77,69%	77,69%	76,57%	77,96%	66,39%
9	77,22%	79,26%	76,30%	77,22%	65,93%
10	76,20%	79,63%	77,87%	77,78%	62,41%
11	77,41%	80,28%	77,04%	78,80%	65,37%
12	79,07%	77,87%	78,70%	78,15%	64,07%
13	78,24%	80,19%	77,04%	76,30%	66,11%
14	77,96%	80,09%	78,61%	78,61%	63,52%
15	78,80%	81,30%	78,24%	80,09%	65,65%
16	79,63%	80,65%	79,26%	78,52%	64,35%
17	77,50%	80,56%	80,00%	76,02%	66,48%
18	80,37%	77,69%	78,52%	76,02%	67,04%
19	78,98%	81,02%	77,50%	79,72%	67,69%
20	80,19%	81,20%	80,56%	78,33%	66,39%
21	78,52%	79,44%	79,26%	79,17%	67,31%
22	80,09%	82,59%	79,44%	78,06%	68,70%
23	80,00%	80,46%	79,63%	77,78%	66,94%
24	81,94%	82,50%	81,85%	78,15%	67,41%
25	79,63%	81,11%	79,07%	78,89%	69,54%
26	81,20%	80,74%	80,09%	77,78%	69,17%
27	79,07%	81,57%	79,81%	76,94%	66,85%
28	80,56%	83,33%	80,74%	78,33%	65,83%
29	81,57%	82,96%	79,35%	79,81%	67,78%
30	81,20%	82,22%	81,48%	78,24%	67,50%
31	81,67%	83,61%	80,83%	79,44%	70,37%
32	81,39%	83,43%	82,04%	78,15%	68,80%
33	81,57%	82,69%	81,39%	80,56%	70,46%
34	81,11%	81,02%	80,37%	79,72%	69,91%
35	82,50%	85,19%	80,09%	79,81%	66,94%
36	79,44%	80,00%	79,72%	80,56%	68,33%
37	79,07%	80,19%	80,65%	78,33%	69,17%
38	79,35%	79,72%	80,93%	80,37%	67,59%
39	78,15%	82,69%	81,94%	80,19%	69,07%
40	78,33%	82,31%	81,11%	78,80%	69,17%
41	79,07%	69,44%	72,96%	81,02%	65,28%
42	82,13%	67,50%	68,89%	76,20%	64,07%

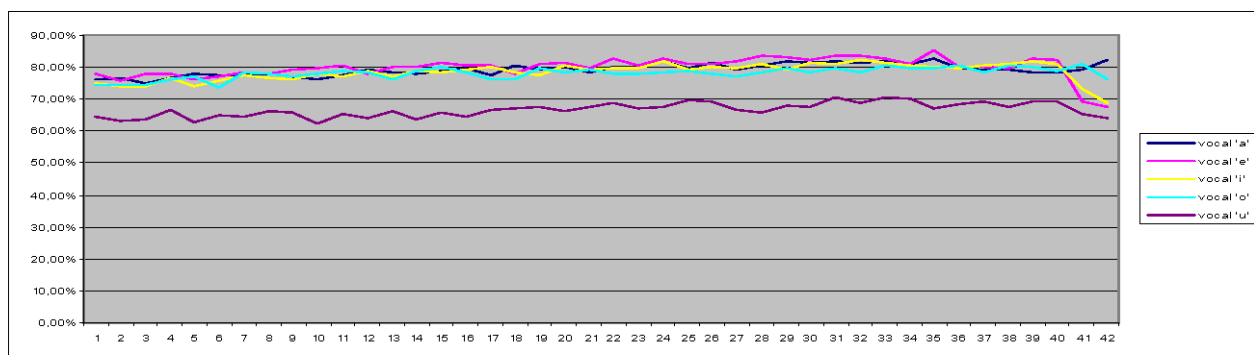


Gráfico 12-7: Evaluación secuencial de la relevancia para las medidas de calidad de la voz del grupo III usando el clasificador NN en la clasificación entre voz sana y voz patológica

Grupo IV

Núm. de Medidas	vocal 'a'	vocal 'e'	vocal 'i'	vocal 'o'	vocal 'u'
1	82,50%	78,06%	73,98%	78,43%	72,04%
2	83,70%	79,17%	75,00%	77,31%	72,22%
3	79,63%	78,24%	74,35%	78,98%	74,54%
4	86,48%	79,81%	74,63%	79,07%	70,56%
5	83,06%	80,37%	78,24%	81,67%	74,81%
6	84,91%	81,20%	75,00%	80,19%	72,31%
7	84,17%	78,52%	74,81%	80,74%	75,00%
8	83,80%	79,35%	74,44%	81,76%	74,91%
9	83,24%	82,59%	75,46%	80,93%	76,20%
10	82,59%	78,89%	76,48%	79,91%	74,44%
11	85,28%	81,20%	76,20%	81,67%	76,39%
12	84,35%	81,48%	76,30%	80,83%	77,31%
13	84,54%	80,19%	75,46%	78,61%	75,83%
14	84,63%	82,22%	77,69%	81,20%	75,83%
15	84,44%	81,76%	80,00%	82,59%	73,06%
16	85,00%	83,70%	78,52%	80,00%	75,00%
17	87,59%	81,02%	77,59%	84,17%	77,13%
18	86,48%	80,74%	78,52%	83,33%	76,57%
19	87,69%	80,56%	78,52%	84,07%	76,94%
20	87,04%	82,96%	76,48%	81,30%	77,96%
21	87,87%	82,22%	77,50%	81,11%	79,17%
22	87,13%	81,85%	76,57%	82,78%	77,87%
23	88,52%	81,76%	77,78%	83,24%	78,15%
24	87,22%	81,76%	77,13%	82,50%	77,04%
25	88,98%	80,74%	79,72%	83,70%	76,85%
26	89,91%	83,15%	78,24%	79,63%	76,85%
27	86,30%	83,43%	77,59%	80,74%	77,96%
28	85,19%	82,13%	77,59%	78,70%	76,02%
29	87,41%	82,59%	77,87%	80,00%	74,91%
30	84,54%	80,93%	76,76%	77,41%	70,37%
31	84,81%	81,02%	75,46%	76,30%	62,50%

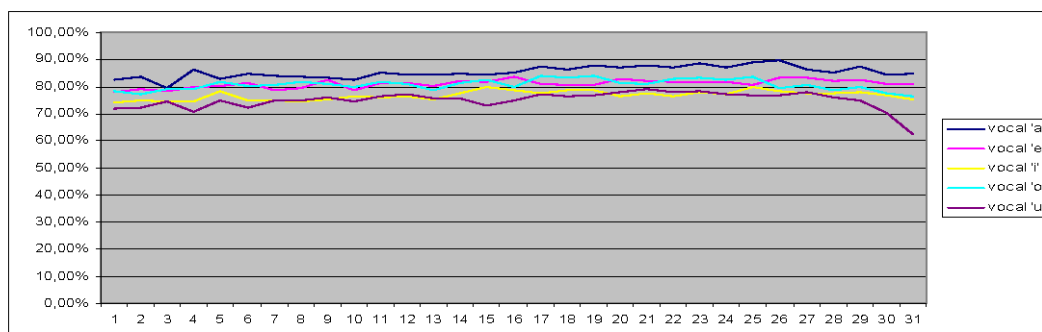


Gráfico 12-8: Evaluación secuencial de la relevancia para las medidas de calidad de la voz del grupo IV usando el clasificador NN en la clasificación entre voz sana y voz patológica

12.3.1.3 Función de Base Radial (RBF)

Grupo I

Núm. de Medidas	vocal 'a'	Vocal 'e'	Vocal 'i'	vocal 'o'	vocal 'u'
1	70,46%	71,76%	63,06%	74,17%	64,63%
2	68,70%	69,07%	62,69%	72,04%	63,61%
3	70,46%	68,89%	63,24%	72,96%	65,00%
4	70,09%	72,31%	64,81%	72,41%	61,94%
5	69,07%	68,70%	65,46%	72,22%	62,78%
6	68,98%	69,91%	65,65%	71,57%	63,80%
7	67,87%	66,48%	64,81%	70,46%	63,98%
8	70,28%	70,56%	67,04%	68,43%	59,35%
9	68,24%	69,44%	65,74%	70,37%	61,85%
10	70,37%	67,41%	64,26%	68,33%	59,54%
11	67,78%	69,91%	62,59%	66,85%	59,07%
12	66,76%	70,09%	64,07%	70,93%	63,15%
13	68,89%	69,35%	66,57%	70,09%	61,94%
14	70,09%	71,94%	65,19%	68,80%	63,15%
15	71,76%	71,39%	60,46%	68,33%	61,67%
16	72,04%	68,80%	60,83%	69,54%	63,24%
17	63,70%	68,06%	56,48%	63,80%	60,09%
18	58,24%	57,31%	52,31%	52,31%	51,67%

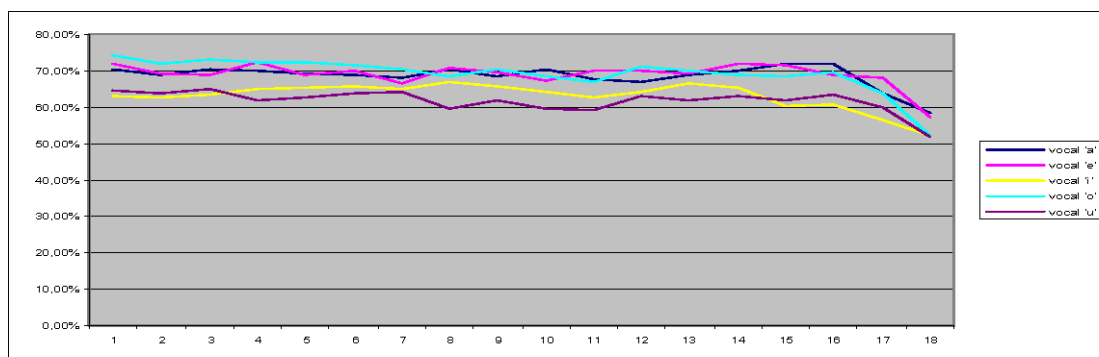


Gráfico 12-9: Evaluación secuencial de la relevancia para las medidas de calidad de la voz del grupo I usando el clasificador RBF en la clasificación entre voz sana y voz patológica

Grupo II

Núm. de Medidas	vocal 'a'	vocal 'e'	vocal 'i'	vocal 'o'	vocal 'u'
1	76,67%	74,26%	68,06%	71,57%	59,26%
2	75,65%	73,43%	66,20%	69,07%	59,72%
3	78,61%	72,13%	68,70%	71,67%	62,78%
4	78,61%	74,72%	67,31%	72,41%	61,20%
5	76,48%	71,39%	67,31%	70,74%	63,24%
6	77,78%	75,09%	67,04%	72,04%	62,69%
7	77,31%	73,89%	68,15%	70,93%	63,24%
8	78,43%	75,93%	66,20%	71,30%	63,70%
9	78,24%	74,07%	65,65%	70,00%	63,61%
10	76,57%	73,06%	67,13%	69,72%	63,80%
11	77,31%	73,06%	65,93%	70,83%	62,50%
12	78,06%	74,26%	65,83%	72,87%	63,15%
13	78,70%	74,81%	65,46%	72,50%	64,44%
14	77,87%	72,22%	67,41%	73,33%	65,74%
15	78,06%	73,61%	65,65%	72,31%	63,61%
16	78,24%	76,20%	65,74%	72,50%	62,87%
17	77,59%	76,39%	68,24%	71,85%	64,17%
18	78,33%	76,67%	66,94%	73,52%	65,00%
19	77,41%	76,85%	68,89%	74,17%	65,09%
20	79,07%	75,74%	68,33%	75,28%	63,70%
21	76,94%	75,83%	65,28%	75,00%	64,81%
22	80,37%	76,30%	64,26%	74,17%	64,35%
23	77,50%	76,39%	65,19%	71,57%	65,74%
24	78,89%	78,24%	69,26%	72,69%	64,72%
25	79,54%	76,67%	66,85%	73,43%	63,70%
26	80,65%	78,06%	66,85%	75,74%	63,15%
27	78,98%	76,94%	65,46%	77,78%	66,11%
28	80,28%	77,78%	67,69%	75,93%	63,89%
29	78,24%	77,78%	67,59%	72,78%	63,15%
30	79,54%	76,94%	67,69%	73,89%	61,48%
31	76,76%	79,07%	69,17%	76,20%	64,35%
32	79,44%	77,96%	69,72%	74,26%	62,96%
33	82,41%	80,00%	66,30%	74,91%	63,33%

34	78,61%	81,39%	71,02%	77,04%	66,30%
35	78,98%	78,61%	67,31%	75,28%	64,54%
36	81,94%	79,26%	69,26%	73,52%	62,69%
37	80,56%	80,74%	65,28%	79,17%	65,28%
38	81,02%	79,81%	62,31%	73,80%	65,56%
39	81,30%	79,07%	63,24%	74,26%	64,17%
40	81,57%	80,83%	60,46%	77,31%	64,91%
41	85,46%	80,37%	63,24%	78,06%	65,19%
42	84,26%	83,33%	62,31%	76,94%	64,72%
43	84,17%	80,65%	61,76%	75,09%	65,56%
44	84,91%	82,50%	60,56%	73,89%	63,98%
45	86,11%	81,67%	54,63%	77,13%	62,50%
46	85,74%	81,39%	60,83%	75,56%	64,91%
47	84,26%	82,41%	55,93%	66,11%	61,39%
48	87,22%	82,04%	57,04%	71,39%	62,04%
49	85,00%	82,13%	59,91%	72,04%	58,52%
50	83,52%	84,91%	50,00%	67,22%	56,39%
51	80,46%	81,85%	50,19%	65,28%	55,28%
52	73,15%	76,30%	50,83%	64,17%	53,98%
53	53,80%	60,19%	50,83%	59,54%	50,28%

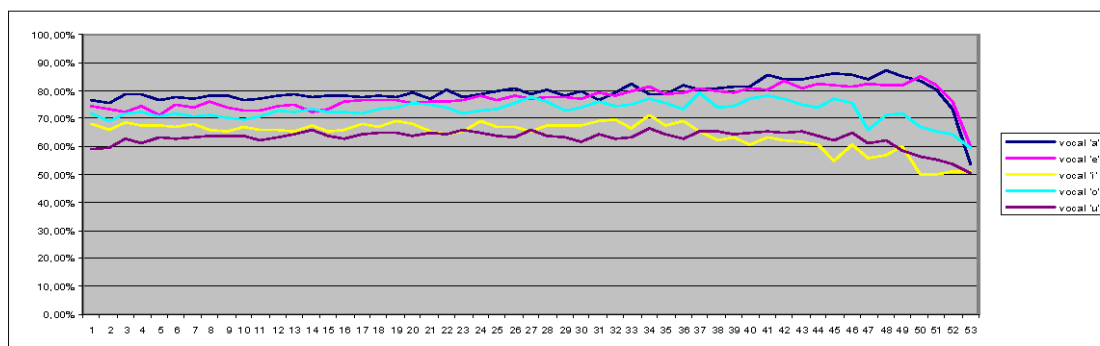


Gráfico 12-10 Evaluación secuencial de la relevancia para las medidas de calidad de la voz del grupo II usando el clasificador RBF en la clasificación entre voz sana y voz patológica

Grupo III

Núm. de Medidas	vocal 'a'	vocal 'e'	Vocal 'i'	vocal 'o'	vocal 'u'
1	51,76%	51,85%	51,85%	51,85%	52,13%
2	51,76%	51,85%	51,76%	52,59%	52,04%
3	52,04%	51,94%	51,85%	52,50%	51,85%
4	51,94%	52,13%	51,85%	54,44%	52,22%
5	52,96%	51,94%	51,85%	57,31%	51,85%
6	55,28%	52,31%	51,85%	56,57%	52,04%
7	56,39%	52,59%	51,94%	58,98%	52,96%
8	53,89%	52,59%	53,06%	59,44%	52,96%
9	56,94%	51,94%	52,96%	61,30%	53,15%
10	59,44%	54,17%	56,11%	60,00%	53,61%
11	60,46%	53,24%	55,83%	61,11%	54,26%
12	62,41%	55,37%	58,33%	64,63%	54,54%

13	61,48%	54,44%	62,41%	62,13%	54,07%
14	63,24%	54,26%	61,76%	64,91%	57,04%
15	65,46%	56,67%	62,96%	69,07%	54,63%
16	63,70%	57,41%	68,61%	71,76%	54,17%
17	62,22%	59,81%	68,15%	73,24%	57,04%
18	66,76%	61,39%	69,44%	72,59%	56,76%
19	66,57%	63,52%	70,65%	73,61%	55,00%
20	69,54%	67,13%	74,44%	72,41%	59,07%
21	71,67%	73,33%	77,96%	73,89%	57,31%
22	69,17%	79,81%	77,13%	72,41%	54,17%
23	73,80%	79,35%	74,63%	77,22%	58,06%
24	72,13%	81,67%	74,54%	78,24%	57,50%
25	72,13%	81,11%	72,59%	76,57%	58,61%
26	74,54%	81,48%	74,26%	75,56%	61,30%
27	71,11%	76,85%	74,17%	74,63%	62,04%
28	70,93%	77,04%	72,22%	71,85%	61,20%
29	70,09%	79,72%	73,15%	69,17%	58,61%
30	69,54%	77,59%	72,41%	69,07%	60,00%
31	69,17%	76,67%	72,69%	71,48%	60,19%
32	69,54%	78,61%	75,93%	70,65%	58,15%
33	70,65%	74,07%	70,93%	69,17%	57,59%
34	70,65%	76,30%	74,54%	70,00%	58,98%
35	72,31%	75,09%	75,28%	73,24%	57,87%
36	75,19%	76,30%	76,20%	71,39%	60,56%
37	71,67%	75,56%	76,20%	71,85%	60,74%
38	74,54%	75,74%	78,43%	73,24%	64,35%
39	71,76%	82,13%	72,50%	71,57%	64,63%
40	70,19%	76,20%	73,52%	66,30%	60,56%
41	59,81%	69,91%	71,02%	69,26%	60,00%
42	58,06%	55,09%	52,13%	54,44%	53,33%

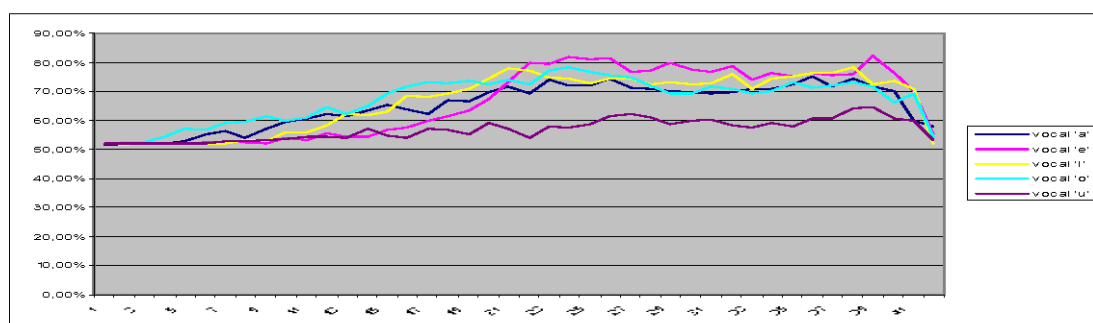


Gráfico 12-11 Evaluación secuencial de la relevancia para las medidas de calidad de la voz del grupo III usando el clasificador RBF en la clasificación entre voz sana y voz patológica

Grupo IV

Núm. de Medidas	vocal 'a'	Vocal 'e'	Vocal 'i'	vocal 'o'	vocal 'u'
1	53,24%	51,57%	52,04%	52,96%	52,87%
2	52,96%	51,76%	52,04%	52,22%	52,78%
3	53,52%	52,87%	52,13%	52,96%	52,96%
4	55,74%	52,50%	52,04%	55,09%	53,70%
5	56,48%	53,52%	52,31%	56,85%	53,98%
6	57,78%	55,74%	51,76%	59,44%	57,50%
7	57,59%	60,09%	52,22%	58,80%	60,28%
8	59,44%	59,35%	53,06%	57,41%	61,20%
9	57,22%	62,13%	54,81%	60,56%	61,48%
10	60,65%	62,13%	56,30%	65,46%	60,93%
11	65,56%	69,54%	60,19%	69,72%	60,19%
12	63,24%	71,39%	61,20%	67,69%	59,44%
13	69,17%	72,50%	66,67%	66,30%	60,37%
14	73,52%	74,63%	65,56%	65,56%	61,02%
15	75,46%	73,33%	68,15%	66,30%	60,65%
16	76,48%	76,94%	68,61%	68,61%	58,52%
17	76,48%	79,72%	68,89%	67,50%	58,98%
18	79,63%	81,30%	70,09%	70,09%	60,09%
19	79,54%	80,09%	73,43%	70,56%	61,39%
20	79,35%	78,61%	68,43%	70,19%	59,26%
21	79,17%	75,74%	66,76%	69,91%	63,98%
22	80,09%	75,46%	69,07%	75,28%	62,31%
23	83,33%	77,22%	71,57%	79,72%	62,96%
24	86,02%	78,80%	72,87%	76,02%	68,15%
25	81,85%	78,61%	74,81%	80,93%	68,80%
26	85,93%	82,22%	76,39%	78,61%	68,43%
27	87,31%	82,69%	78,89%	79,26%	73,24%
28	86,48%	82,04%	78,98%	77,13%	72,50%
29	80,65%	80,83%	72,69%	73,98%	71,67%
30	72,22%	59,35%	57,50%	74,63%	67,31%
31	61,48%	60,09%	58,43%	52,87%	56,02%

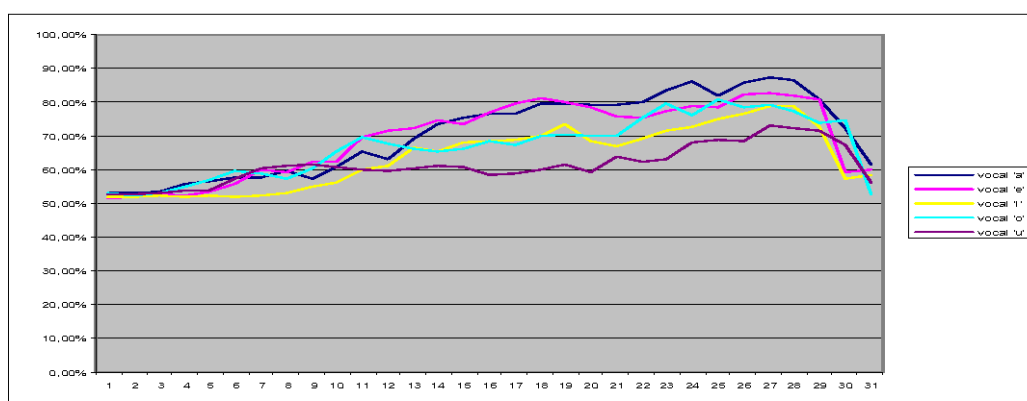


Gráfico 12-12 Evaluación secuencial de la relevancia para las medidas de calidad de la voz del grupo IV usando el clasificador RBF en la clasificación entre voz sana y voz patológica

12.3.1.4 Máquinas de Vectores Soporte (SVM)

Grupo I

Núm. de Medidas	vocal 'a'	Vocal 'e'	Vocal 'i'	vocal 'o'	vocal 'u'
1	79,17%	78,52%	67,69%	74,07%	67,41%
2	81,57%	77,87%	66,94%	77,04%	67,78%
3	79,07%	76,48%	75,28%	76,39%	69,54%
4	79,81%	77,87%	71,76%	75,28%	66,94%
5	79,63%	77,87%	70,00%	75,37%	67,78%
6	80,19%	78,06%	74,17%	76,30%	67,31%
7	76,76%	76,85%	73,80%	75,56%	68,33%
8	81,48%	75,74%	74,26%	74,54%	66,39%
9	78,52%	74,44%	73,43%	76,76%	66,57%
10	76,39%	76,48%	72,50%	74,35%	64,44%
11	79,44%	75,19%	73,06%	73,24%	67,50%
12	75,74%	73,80%	70,74%	74,72%	67,22%
13	79,17%	75,46%	71,94%	74,81%	68,61%
14	79,17%	77,41%	69,91%	72,50%	66,02%
15	76,39%	74,91%	68,43%	74,81%	68,24%
16	75,46%	78,15%	67,96%	74,81%	69,72%
17	72,78%	76,76%	68,98%	72,69%	64,17%
18	77,04%	78,61%	67,31%	72,22%	55,00%

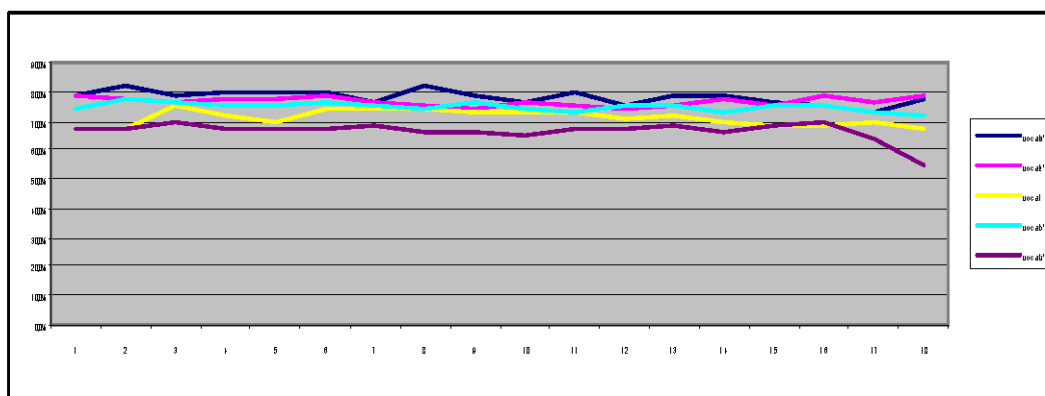


Gráfico 12-13: Evaluación secuencial de la relevancia para las medidas de calidad de la voz del grupo I usando el clasificador SVM en la clasificación entre voz sana y voz patológica

Grupo II

Núm. de Medidas	vocal 'a'	vocal 'e'	vocal 'i'	vocal 'o'	vocal 'u'
1	88,33%	85,83%	76,48%	87,04%	70,46%
2	87,69%	85,37%	76,57%	84,81%	71,76%
3	88,06%	84,91%	76,76%	85,74%	72,69%
4	86,76%	85,37%	78,15%	85,93%	70,65%
5	88,52%	85,28%	76,30%	86,48%	71,57%
6	86,02%	84,54%	74,35%	85,37%	71,11%
7	88,15%	84,81%	77,13%	86,76%	72,04%
8	86,85%	82,59%	76,11%	86,39%	72,87%
9	88,15%	84,72%	76,57%	86,30%	70,65%
10	88,98%	85,46%	77,22%	83,89%	73,06%
11	86,48%	85,28%	76,30%	83,52%	72,87%
12	88,06%	84,72%	77,31%	86,85%	72,59%
13	87,50%	85,37%	76,85%	85,56%	72,31%
14	86,48%	83,06%	78,06%	84,63%	73,89%
15	86,76%	86,30%	76,11%	86,02%	71,20%
16	86,76%	86,02%	78,15%	84,63%	73,43%
17	88,80%	86,67%	76,48%	83,43%	73,98%
18	87,96%	83,43%	76,85%	86,11%	73,80%
19	89,07%	88,61%	76,48%	86,11%	74,26%
20	87,22%	85,83%	74,44%	84,72%	73,06%
21	87,59%	84,91%	74,17%	85,65%	72,59%
22	86,76%	85,00%	73,98%	84,26%	72,13%
23	87,69%	84,35%	74,91%	87,04%	73,06%
24	87,41%	82,78%	71,20%	83,70%	71,85%
25	86,67%	86,57%	72,22%	86,85%	75,19%
26	85,93%	85,09%	73,61%	85,28%	73,24%
27	85,46%	85,46%	75,28%	84,91%	73,70%
28	84,91%	85,65%	76,02%	84,91%	72,87%
29	87,41%	85,37%	74,72%	85,65%	71,48%
30	87,87%	85,65%	73,06%	85,00%	72,04%
31	86,48%	86,30%	71,94%	86,48%	73,61%
32	86,94%	85,83%	72,87%	85,09%	72,13%
33	86,02%	78,80%	71,48%	84,81%	70,83%
34	83,15%	85,37%	75,83%	86,67%	69,17%
35	82,31%	81,76%	79,81%	85,74%	70,56%
36	87,41%	84,72%	78,06%	83,61%	70,65%
37	83,33%	81,67%	77,78%	87,41%	69,91%
38	83,33%	80,19%	74,63%	85,00%	69,63%
39	85,93%	82,96%	72,78%	83,33%	68,52%
40	88,33%	83,43%	70,00%	85,93%	71,11%
41	85,93%	84,35%	71,11%	85,74%	70,93%
42	85,46%	80,74%	67,78%	77,69%	72,59%
43	84,91%	85,19%	70,83%	83,06%	70,93%
44	83,06%	83,24%	69,54%	82,04%	73,33%
45	82,87%	81,57%	66,94%	85,09%	69,44%
46	87,04%	85,83%	67,59%	85,93%	70,65%
47	80,83%	85,37%	64,54%	77,31%	69,81%
48	86,20%	84,72%	65,46%	76,76%	65,65%

49	80,19%	81,48%	69,26%	72,69%	65,74%
50	84,17%	80,19%	62,50%	74,54%	66,57%
51	81,39%	83,24%	61,20%	72,87%	61,57%
52	76,02%	78,98%	57,59%	71,11%	55,19%
53	65,00%	76,76%	55,37%	71,30%	55,93%

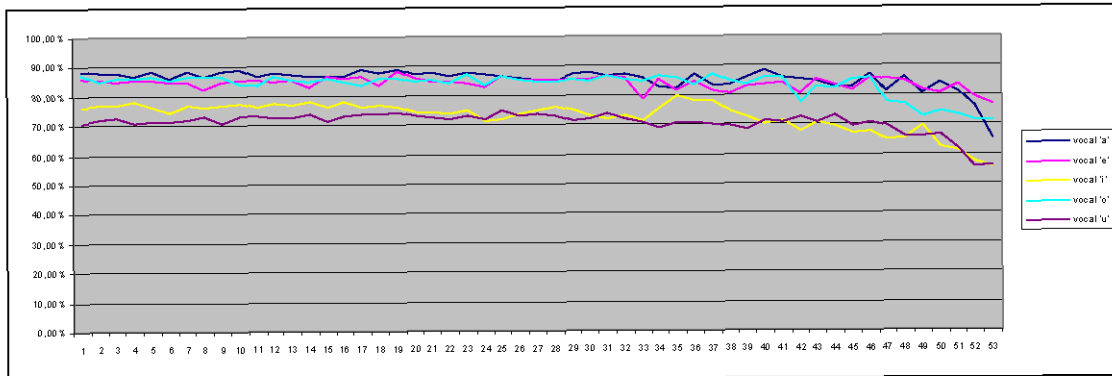


Gráfico 12-14 : Evaluación secuencial de la relevancia para las medidas de calidad de la voz del grupo II usando el clasificador SVM en la clasificación entre voz sana y voz patológica

Grupo III

Núm. de Medidas	vocal 'a'	vocal 'e'	Vocal 'i'	vocal 'o'	vocal 'u'
1	79,07%	82,78%	82,22%	79,44%	66,57%
2	81,30%	80,74%	80,28%	77,59%	65,19%
3	80,19%	83,52%	82,59%	80,00%	66,30%
4	80,93%	82,78%	81,30%	80,83%	65,65%
5	80,46%	83,06%	81,39%	80,09%	64,91%
6	81,20%	83,70%	81,94%	81,20%	65,00%
7	82,31%	82,50%	79,63%	79,63%	66,57%
8	81,48%	82,50%	79,07%	81,11%	63,15%
9	80,74%	81,67%	81,85%	81,11%	65,28%
10	81,85%	82,87%	79,91%	82,41%	62,69%
11	83,06%	82,78%	81,20%	80,28%	64,91%
12	83,98%	82,04%	81,85%	81,48%	66,76%
13	80,56%	84,26%	79,44%	79,26%	64,54%
14	80,93%	83,06%	78,70%	79,81%	65,19%
15	82,04%	84,26%	79,07%	79,54%	65,37%
16	80,83%	81,30%	82,50%	81,11%	65,93%
17	81,39%	83,61%	83,06%	79,81%	66,11%
18	83,06%	84,17%	79,54%	78,52%	66,94%
19	81,20%	82,31%	83,61%	77,31%	69,17%
20	79,35%	82,69%	79,54%	79,07%	68,15%
21	81,39%	82,78%	81,20%	79,26%	67,96%
22	79,63%	83,61%	84,44%	80,37%	68,15%
23	79,54%	83,43%	80,56%	78,24%	66,76%
24	81,67%	83,61%	82,22%	80,56%	66,76%

25	81,30%	84,91%	83,70%	80,83%	67,04%
26	80,37%	83,52%	84,44%	83,33%	66,85%
27	78,06%	81,57%	84,72%	81,02%	69,72%
28	79,26%	81,94%	81,94%	81,02%	67,31%
29	79,91%	85,00%	82,13%	80,46%	66,57%
30	81,11%	82,04%	81,11%	81,85%	68,24%
31	80,56%	83,33%	81,20%	79,63%	66,76%
32	77,87%	85,46%	81,20%	81,57%	67,69%
33	82,04%	83,61%	80,28%	81,76%	69,81%
34	76,94%	82,96%	81,11%	81,30%	67,78%
35	80,09%	82,59%	77,59%	76,94%	66,94%
36	80,83%	78,43%	82,59%	81,20%	67,13%
37	76,02%	82,31%	80,65%	77,04%	68,80%
38	76,48%	81,48%	79,91%	79,26%	70,00%
39	77,41%	83,61%	80,56%	79,81%	69,54%
40	77,59%	81,20%	80,09%	78,06%	69,91%
41	75,46%	65,83%	79,17%	77,41%	63,33%
42	75,83%	70,37%	73,98%	78,24%	62,96%

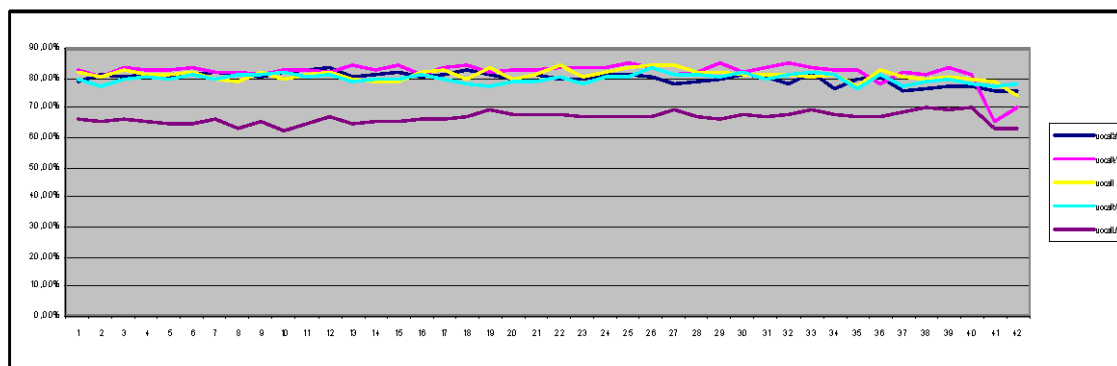


Gráfico 12-15 : Evaluación secuencial de la relevancia para las medidas de calidad de la voz del grupo III usando el clasificador SVM en la clasificación entre voz sana y voz patológica

Grupo IV

Núm. de Medidas	vocal 'a'	Vocal 'e'	Vocal 'i'	vocal 'o'	vocal 'u'
1	89,26%	84,81%	79,91%	82,50%	73,61%
2	86,85%	82,69%	78,52%	81,02%	72,31%
3	86,85%	84,72%	79,26%	81,76%	73,70%
4	86,85%	84,44%	79,72%	83,61%	72,41%
5	87,87%	83,52%	81,20%	83,43%	73,33%
6	85,93%	84,91%	78,43%	83,61%	73,89%
7	85,65%	86,85%	79,35%	83,24%	74,54%
8	86,94%	83,80%	80,74%	81,94%	72,31%
9	87,13%	83,89%	78,89%	81,85%	75,65%
10	87,69%	84,54%	78,43%	82,96%	73,98%
11	88,89%	84,72%	79,54%	83,33%	73,70%

12	87,59%	83,61%	78,70%	82,13%	73,70%
13	87,87%	84,54%	81,11%	81,94%	73,15%
14	87,41%	83,89%	79,26%	83,80%	74,17%
15	86,85%	83,80%	81,11%	83,98%	73,43%
16	89,54%	83,15%	81,11%	83,15%	76,02%
17	87,87%	82,31%	81,85%	84,07%	74,91%
18	87,96%	82,04%	78,24%	80,83%	76,48%
19	88,80%	80,93%	79,91%	82,50%	77,04%
20	88,89%	81,76%	78,43%	83,43%	79,54%
21	90,19%	83,33%	79,54%	80,93%	76,30%
22	90,19%	82,59%	79,54%	81,48%	76,11%
23	91,39%	85,00%	78,24%	83,80%	77,04%
24	86,85%	81,11%	80,28%	84,35%	75,65%
25	90,09%	82,41%	80,46%	82,96%	76,76%
26	88,98%	80,65%	80,00%	79,26%	74,72%
27	83,98%	84,63%	78,98%	79,54%	73,98%
28	87,04%	83,43%	80,93%	75,09%	75,19%
29	86,67%	82,22%	79,54%	76,20%	75,37%
30	84,63%	83,33%	79,44%	76,39%	72,31%
31	83,70%	83,06%	78,98%	80,00%	64,72%

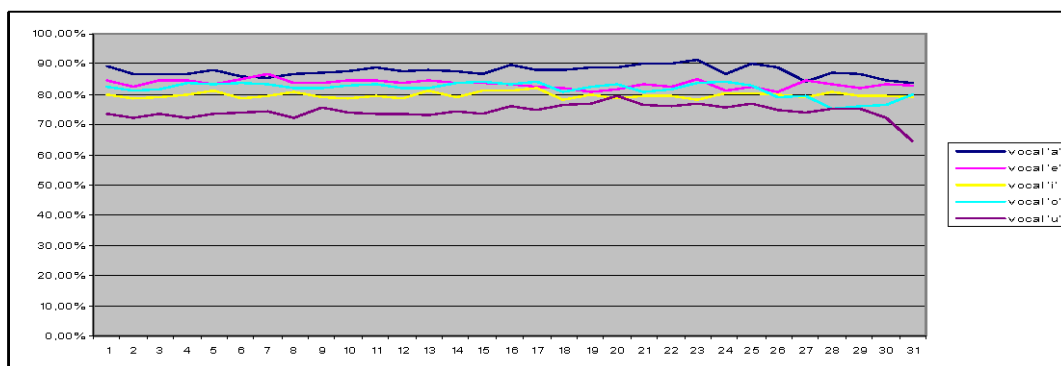


Gráfico 12-16 : Evaluación secuencial de la relevancia para las medidas de calidad de la voz del grupo IV usando el clasificador SVM en la clasificación entre voz sana y voz patológica

12.3.1.5 K vecinos más cercanos (KNN)**Grupo I**

Núm. de Medidas	vocal 'a'	vocal 'e'	vocal 'i'	Vocal 'o'	vocal 'u'
1	77,59%	77,69%	67,31%	72,59%	68,43%
2	79,63%	77,04%	68,89%	73,33%	68,89%
3	80,37%	75,93%	70,28%	72,04%	67,50%
4	80,74%	77,50%	72,13%	71,76%	68,06%
5	79,54%	76,85%	71,20%	72,78%	68,43%
6	79,54%	75,28%	71,48%	75,09%	66,67%
7	78,89%	77,50%	70,46%	73,98%	68,24%
8	75,83%	76,20%	70,28%	74,72%	66,30%
9	78,89%	76,20%	71,94%	75,56%	66,76%
10	76,76%	76,11%	73,61%	74,07%	66,94%
11	78,06%	77,50%	72,04%	73,70%	65,65%
12	75,37%	75,00%	72,96%	72,41%	67,50%
13	75,93%	72,31%	70,19%	73,80%	67,78%
14	77,41%	74,35%	71,57%	76,20%	67,22%
15	76,39%	74,54%	71,20%	74,07%	66,94%
16	80,65%	73,98%	73,98%	74,26%	71,57%
17	76,11%	76,94%	72,69%	71,30%	68,61%
18	80,00%	81,39%	68,43%	71,39%	61,39%

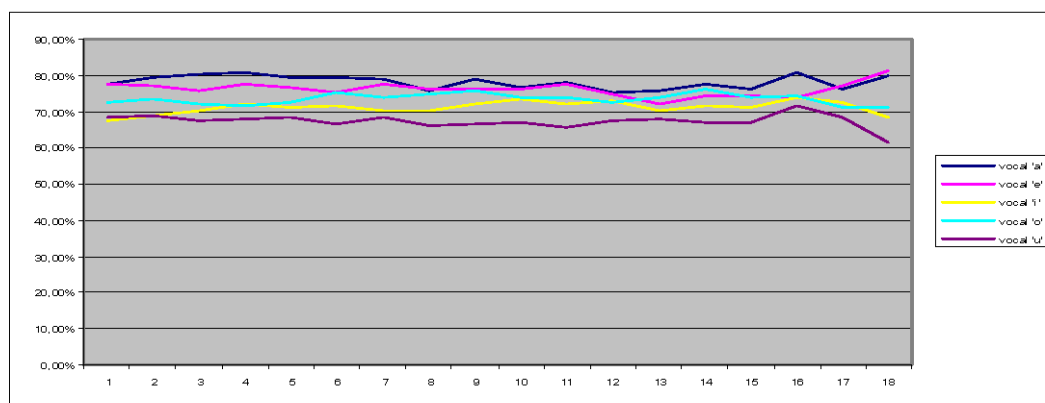


Gráfico 12-17: Evaluación secuencial de la relevancia para las medidas de calidad de la voz del grupo I usando el clasificador KNN en la clasificación entre voz sana y voz patológica

Grupo II

Núm. de Medidas	vocal 'a'	vocal 'e'	vocal 'i'	vocal 'o'	vocal 'u'
1	88,24%	83,52%	74,44%	82,41%	69,91%
2	87,69%	83,98%	76,20%	82,04%	71,11%
3	89,44%	82,87%	76,76%	82,69%	72,96%
4	89,07%	84,35%	76,48%	82,59%	72,22%
5	87,69%	83,61%	75,65%	82,13%	72,50%
6	88,24%	83,33%	75,46%	82,50%	72,69%
7	87,31%	84,72%	75,83%	83,89%	72,59%
8	88,52%	84,44%	76,76%	83,80%	74,44%
9	88,33%	85,74%	76,67%	84,44%	72,13%
10	88,15%	85,37%	76,57%	84,07%	75,19%
11	88,61%	84,17%	77,22%	84,91%	73,80%
12	86,67%	83,70%	76,57%	83,52%	71,76%
13	87,78%	83,15%	75,83%	83,24%	70,19%
14	87,78%	82,22%	77,69%	86,94%	71,94%
15	88,43%	84,91%	76,48%	83,52%	72,22%
16	88,70%	85,56%	78,43%	85,28%	72,96%
17	85,83%	84,44%	75,93%	83,06%	70,83%
18	87,41%	84,63%	76,85%	84,26%	72,22%
19	87,22%	85,83%	76,94%	85,28%	71,85%
20	88,80%	85,74%	77,41%	84,72%	72,78%
21	85,74%	86,30%	74,07%	84,07%	75,56%
22	89,44%	85,93%	75,28%	84,54%	75,65%
23	87,87%	84,35%	75,19%	84,91%	72,04%
24	88,33%	86,94%	75,74%	82,13%	73,98%
25	87,50%	83,24%	74,07%	82,04%	71,57%
26	88,70%	85,00%	73,98%	83,70%	73,33%
27	90,19%	85,19%	76,48%	83,15%	73,98%
28	86,48%	84,81%	73,24%	82,04%	74,44%
29	87,69%	85,93%	75,28%	81,67%	73,33%
30	88,24%	84,35%	73,80%	81,30%	73,98%
31	88,98%	86,11%	74,54%	82,69%	73,98%
32	88,15%	86,76%	73,89%	82,96%	72,22%
33	88,52%	85,56%	74,07%	84,26%	72,50%
34	88,33%	85,09%	73,70%	84,07%	72,96%
35	86,67%	85,46%	75,37%	83,43%	72,13%
36	88,80%	84,35%	74,07%	82,50%	73,15%
37	88,15%	85,65%	74,17%	83,15%	71,02%
38	87,50%	85,46%	72,50%	81,30%	72,41%
39	88,80%	85,56%	73,24%	81,02%	71,57%
40	88,43%	85,28%	72,69%	81,67%	72,04%
41	89,35%	84,35%	70,83%	83,89%	71,11%
42	87,31%	86,02%	72,22%	82,50%	73,89%
43	89,54%	84,72%	75,09%	80,19%	73,06%
44	86,76%	86,20%	74,17%	82,41%	71,39%
45	88,98%	86,30%	72,31%	82,04%	72,87%
46	88,43%	85,83%	71,57%	82,31%	67,69%
47	88,15%	86,11%	71,76%	77,31%	71,48%
48	88,43%	87,31%	71,94%	77,31%	70,28%

49	89,26%	87,04%	71,48%	77,41%	70,00%
50	88,24%	85,93%	70,65%	76,48%	67,69%
51	87,59%	85,28%	65,83%	75,37%	70,65%
52	81,67%	85,46%	64,44%	73,24%	70,19%
53	82,41%	81,11%	64,91%	68,70%	68,52%

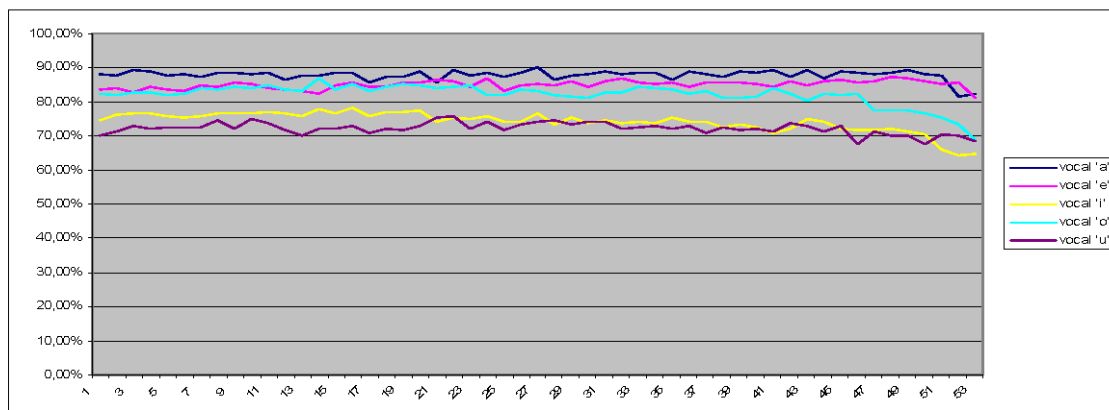


Gráfico 12-18: Evaluación secuencial de la relevancia para las medidas de calidad de la voz del grupo II usando el clasificador KNN en la clasificación entre voz sana y voz patológica

Grupo III

Núm. de Medidas	vocal 'a'	vocal 'e'	vocal 'i'	vocal 'o'	vocal 'u'
1	80,65%	80,65%	73,06%	71,76%	61,67%
2	75,46%	78,80%	73,43%	69,35%	61,20%
3	77,59%	79,63%	73,52%	72,78%	58,98%
4	79,81%	81,57%	74,72%	71,94%	63,33%
5	77,69%	80,37%	73,61%	69,63%	64,44%
6	79,63%	78,33%	76,39%	72,31%	61,11%
7	77,22%	81,02%	73,33%	71,67%	60,65%
8	76,11%	80,93%	74,91%	72,22%	60,00%
9	78,33%	81,20%	76,76%	70,74%	63,06%
10	80,37%	81,30%	75,93%	71,85%	63,15%
11	78,80%	83,24%	76,11%	72,22%	63,43%
12	80,74%	82,78%	77,13%	70,74%	62,50%
13	79,35%	80,37%	76,48%	72,04%	63,70%
14	80,65%	82,04%	75,83%	71,02%	63,15%
15	81,57%	84,35%	76,76%	72,04%	62,59%
16	79,54%	85,19%	75,65%	73,15%	63,06%
17	80,65%	85,00%	77,78%	72,87%	65,09%
18	79,72%	83,70%	79,07%	72,50%	66,20%
19	81,85%	83,70%	77,87%	74,17%	65,65%
20	79,81%	85,93%	78,98%	74,26%	67,69%
21	81,67%	83,33%	80,00%	72,78%	65,65%
22	80,56%	84,17%	76,48%	74,44%	69,17%
23	78,33%	83,06%	78,06%	72,04%	67,69%

24	76,57%	85,09%	78,89%	73,98%	67,41%
25	77,87%	82,41%	78,24%	75,09%	70,19%
26	79,26%	83,06%	79,07%	75,19%	65,74%
27	79,26%	84,26%	79,44%	76,76%	68,61%
28	80,00%	83,15%	77,87%	79,91%	67,13%
29	79,26%	85,56%	82,31%	76,11%	70,28%
30	79,44%	83,52%	80,56%	79,07%	66,48%
31	80,83%	85,09%	82,04%	77,22%	68,33%
32	78,52%	83,80%	78,61%	78,80%	68,61%
33	81,39%	82,78%	81,48%	79,07%	69,63%
34	81,30%	84,07%	78,98%	77,22%	70,83%
35	80,65%	82,41%	79,26%	78,80%	66,85%
36	80,46%	83,43%	80,65%	77,96%	66,20%
37	78,89%	84,17%	80,74%	78,89%	67,87%
38	79,44%	82,78%	78,52%	78,89%	69,54%
39	76,76%	82,13%	79,63%	78,33%	66,57%
40	78,33%	85,09%	82,13%	79,17%	66,94%
41	79,54%	70,93%	77,13%	79,91%	62,22%
42	80,46%	69,81%	72,22%	79,07%	63,15%

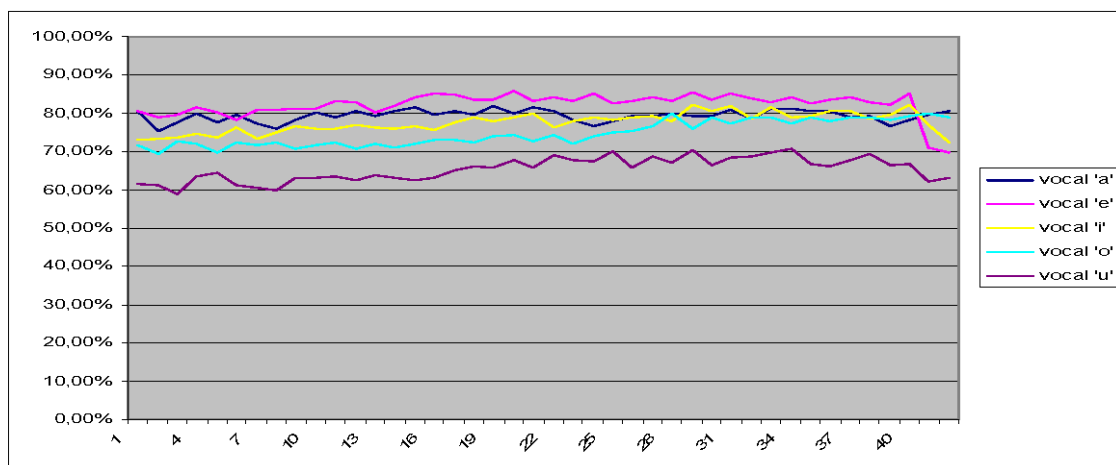


Gráfico 12-19: Evaluación secuencial de la relevancia para las medidas de calidad de la voz del grupo III usando el clasificador KNN en la clasificación entre voz sana y voz patológica

Grupo IV

Núm. de Medidas	vocal 'a'	vocal 'e'	vocal 'i'	vocal 'o'	vocal 'u'
1	84,35%	84,54%	75,37%	82,78%	75,00%
2	86,76%	84,81%	77,87%	83,33%	76,02%
3	85,00%	81,85%	75,74%	82,41%	73,80%
4	85,46%	83,15%	77,59%	81,67%	75,09%
5	84,54%	84,91%	77,04%	81,67%	75,00%
6	84,63%	84,07%	77,31%	81,94%	73,61%
7	84,26%	83,61%	75,83%	80,37%	74,17%
8	85,19%	83,52%	74,35%	82,31%	75,65%
9	87,31%	84,63%	77,31%	82,31%	74,72%
10	87,59%	84,07%	77,96%	82,78%	75,09%
11	84,35%	84,26%	76,85%	83,24%	74,91%
12	87,78%	82,50%	79,54%	80,56%	74,81%
13	85,74%	85,37%	76,76%	82,69%	75,28%
14	86,94%	82,87%	79,54%	81,11%	74,91%
15	87,96%	82,78%	79,07%	81,20%	75,74%
16	87,31%	82,69%	78,89%	83,80%	75,56%
17	89,81%	85,83%	78,98%	83,70%	76,39%
18	89,63%	84,63%	79,44%	84,44%	77,22%
19	88,52%	83,43%	81,02%	81,76%	76,57%
20	86,67%	83,24%	78,61%	82,13%	76,94%
21	88,52%	82,31%	80,65%	84,26%	78,61%
22	89,81%	82,87%	78,89%	83,33%	78,06%
23	88,98%	81,48%	78,06%	82,69%	79,07%
24	89,17%	82,87%	81,76%	82,96%	77,50%
25	88,52%	83,80%	81,67%	82,31%	76,39%
26	87,50%	83,06%	77,50%	81,02%	76,30%
27	87,41%	83,80%	77,59%	80,56%	79,17%
28	86,94%	85,46%	79,81%	79,17%	77,69%
29	86,76%	82,87%	76,76%	79,91%	75,65%
30	84,44%	81,39%	77,69%	78,89%	73,70%
31	85,56%	80,83%	78,24%	79,54%	60,37%

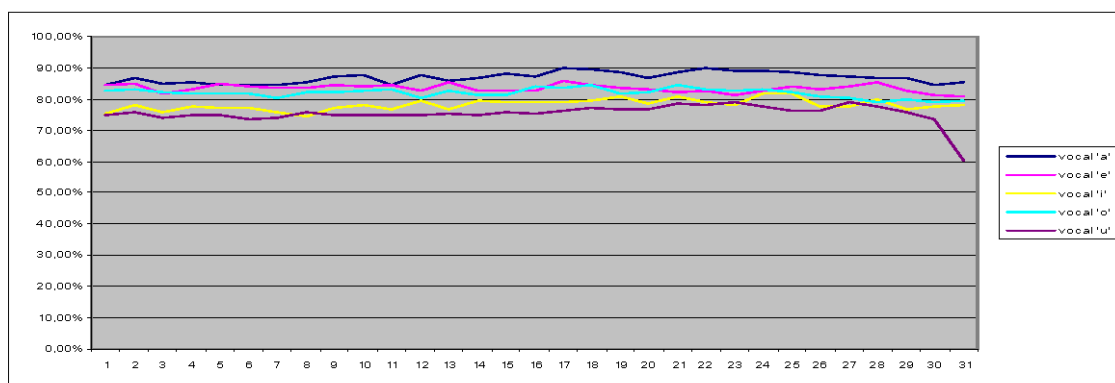


Gráfico 12-20: Evaluación secuencial de la relevancia para las medidas de calidad de la voz del grupo IV usando el clasificador KNN en la clasificación entre voz sana y voz patológica

12.3.2 Estudio: Sano – Disfonía Leve – Disfonía Moderada – Disfonía Severa

12.3.2.1 Red Neuronal Probabilística (PNN)

Grupo I

Núm. de Medidas	vocal 'a'	vocal 'e'	vocal 'i'	Vocal 'o'	vocal 'u'
1	50,74%	45,46%	45,09%	47,31%	50,93%
2	51,57%	47,50%	49,63%	48,43%	51,20%
3	53,61%	47,31%	50,46%	52,31%	50,09%
4	53,33%	49,63%	50,37%	52,41%	50,46%
5	52,78%	51,20%	46,76%	53,70%	49,26%
6	52,22%	52,96%	49,26%	53,52%	50,56%
7	51,67%	53,24%	49,91%	53,61%	53,43%
8	53,70%	53,15%	48,70%	52,13%	50,93%
9	52,13%	51,76%	50,83%	54,35%	51,94%
10	50,65%	50,83%	49,63%	55,74%	55,65%
11	52,87%	54,35%	50,74%	57,22%	55,65%
12	49,35%	55,00%	49,81%	55,65%	52,31%
13	55,28%	55,56%	48,89%	54,63%	52,04%
14	53,24%	54,26%	49,07%	53,98%	50,00%
15	54,26%	48,80%	46,94%	53,52%	48,80%
16	57,87%	52,87%	45,93%	52,78%	44,91%
17	53,80%	52,78%	38,15%	52,50%	50,09%
18	50,09%	54,81%	53,98%	57,04%	53,24%

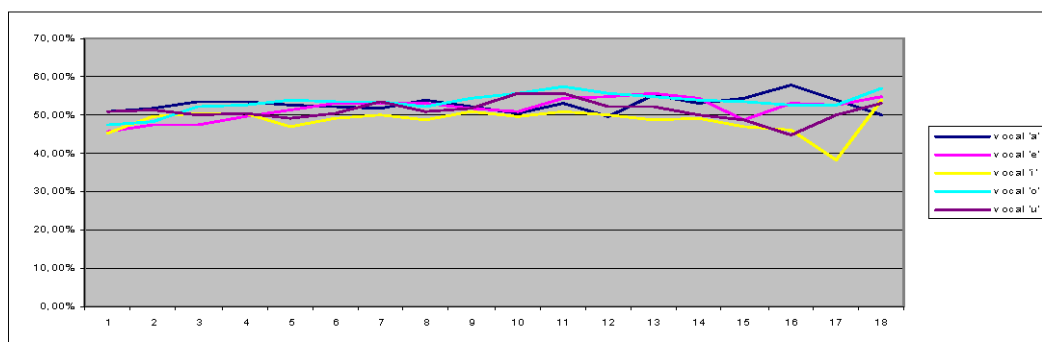


Gráfico 12-21: Evaluación secuencial de la relevancia para las medidas de calidad de la voz del grupo I usando el clasificador PNN en la clasificación entre diferentes grados de disfonía

Grupo II

Núm. de Medidas	vocal 'a'	vocal 'e'	vocal 'i'	vocal 'o'	vocal 'u'
1	64,44%	56,76%	51,85%	55,28%	49,35%
2	62,96%	55,93%	52,78%	54,63%	48,89%
3	62,04%	57,87%	50,83%	54,07%	48,89%
4	65,09%	59,72%	53,80%	55,00%	51,39%
5	65,09%	58,06%	54,81%	57,04%	53,43%
6	62,78%	58,70%	53,89%	57,31%	53,15%
7	63,80%	60,56%	51,67%	55,65%	52,50%
8	62,59%	57,96%	50,28%	56,11%	51,57%
9	63,43%	60,65%	52,41%	56,02%	54,91%
10	64,17%	60,83%	50,28%	57,22%	55,37%
11	63,52%	57,87%	48,89%	57,31%	54,17%
12	63,43%	59,54%	50,74%	56,48%	52,13%
13	63,43%	60,37%	49,72%	57,41%	52,50%
14	62,31%	57,59%	51,02%	56,57%	53,06%
15	63,43%	61,48%	50,19%	56,39%	52,96%
16	63,98%	62,22%	50,74%	58,33%	52,96%
17	61,76%	59,35%	51,20%	54,63%	52,31%
18	61,39%	62,13%	52,04%	56,94%	53,33%
19	64,81%	62,13%	52,13%	56,57%	55,00%
20	65,28%	61,11%	53,33%	57,04%	53,15%
21	62,78%	62,59%	54,07%	54,26%	52,22%
22	61,67%	61,76%	50,19%	56,20%	50,28%
23	63,15%	61,76%	55,74%	58,61%	51,67%
24	63,06%	63,06%	56,94%	56,39%	53,33%
25	62,87%	61,48%	53,43%	56,20%	50,83%
26	62,41%	62,87%	54,63%	56,30%	53,06%
27	63,61%	62,59%	56,67%	53,61%	48,61%
28	63,80%	61,57%	55,56%	57,69%	47,59%
29	66,85%	63,52%	54,91%	56,39%	50,65%
30	65,56%	61,02%	52,78%	58,24%	49,72%
31	67,50%	62,04%	53,43%	58,70%	50,19%
32	66,02%	58,70%	54,91%	58,43%	49,35%
33	68,61%	63,15%	54,07%	58,52%	51,11%
34	66,39%	63,06%	55,56%	59,72%	53,61%
35	67,04%	62,69%	55,37%	59,17%	50,56%
36	64,91%	61,11%	55,46%	60,56%	52,50%
37	68,80%	62,78%	56,94%	60,09%	54,35%
38	67,87%	61,02%	55,19%	58,89%	54,44%
39	65,37%	59,91%	54,81%	57,41%	52,78%
40	66,48%	61,30%	56,85%	58,43%	53,52%
41	64,81%	59,72%	55,83%	57,41%	52,22%
42	65,74%	62,13%	56,57%	60,93%	55,19%
43	66,48%	63,61%	56,94%	61,57%	51,76%
44	64,44%	61,11%	57,41%	60,37%	52,04%
45	63,33%	62,50%	54,72%	60,93%	50,83%
46	66,02%	63,06%	56,67%	59,91%	55,00%
47	63,98%	60,74%	55,19%	62,41%	54,35%
48	63,89%	60,83%	57,04%	56,48%	50,65%

49	63,80%	59,44%	55,09%	58,61%	52,13%
50	64,54%	57,87%	58,06%	54,81%	51,76%
51	66,94%	59,26%	57,13%	54,44%	54,07%
52	63,70%	59,81%	55,37%	55,46%	53,80%
53	63,24%	44,54%	55,00%	51,39%	53,33%

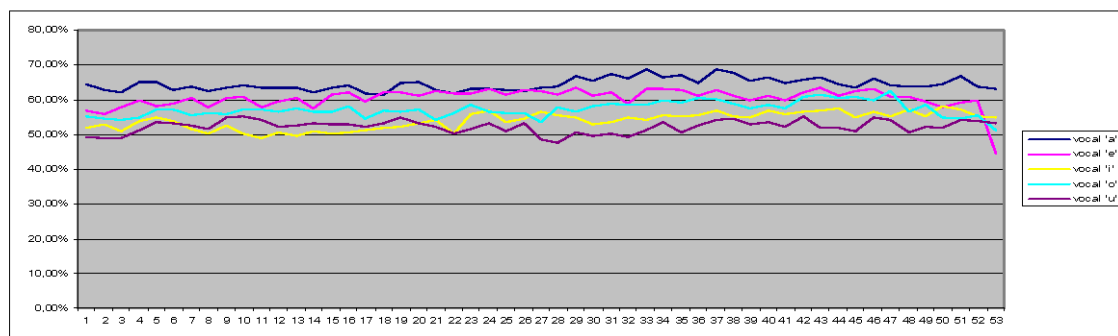


Gráfico 12-22: Evaluación secuencial de la relevancia para las medidas de calidad de la voz del grupo II usando el clasificador PNN en la clasificación entre diferentes grados de disfonía

Grupo III

Núm. de Medidas	vocal 'a'	vocal 'e'	vocal 'i'	vocal 'o'	vocal 'u'
1	48,89%	48,43%	46,85%	46,57%	46,67%
2	48,70%	47,96%	47,78%	46,39%	46,30%
3	48,61%	48,89%	47,59%	47,41%	46,30%
4	48,24%	48,33%	47,69%	47,87%	45,28%
5	48,61%	48,80%	48,43%	46,94%	46,02%
6	48,61%	48,61%	47,22%	48,33%	46,48%
7	48,98%	49,72%	48,24%	47,50%	45,37%
8	49,35%	48,98%	48,52%	47,50%	46,02%
9	48,52%	50,00%	48,52%	49,17%	44,44%
10	50,56%	49,07%	49,07%	49,44%	42,04%
11	51,76%	49,35%	51,39%	48,43%	43,33%
12	53,80%	49,35%	49,07%	47,87%	43,15%
13	53,98%	48,98%	50,56%	47,22%	46,48%
14	55,56%	54,81%	50,37%	48,89%	43,06%
15	53,15%	56,30%	52,31%	45,19%	44,07%
16	52,22%	54,17%	51,76%	45,83%	43,43%
17	55,65%	56,20%	54,54%	46,02%	46,11%
18	55,37%	57,04%	56,02%	49,07%	44,44%
19	55,00%	56,57%	56,57%	48,61%	44,54%
20	55,09%	58,80%	58,70%	50,09%	45,19%
21	56,30%	60,28%	54,63%	51,94%	44,26%
22	52,78%	57,87%	56,20%	51,76%	47,59%
23	55,46%	60,09%	57,50%	51,94%	45,83%
24	56,57%	58,24%	58,89%	52,87%	46,20%
25	57,69%	57,22%	60,00%	51,02%	45,83%
26	55,00%	58,15%	62,50%	53,52%	45,19%
27	56,39%	56,20%	60,09%	56,39%	42,87%

28	54,26%	59,72%	60,00%	56,30%	46,02%
29	56,76%	59,26%	61,39%	56,76%	44,44%
30	56,76%	58,70%	61,85%	56,30%	46,67%
31	57,04%	57,78%	60,83%	54,63%	45,19%
32	58,06%	56,94%	60,93%	52,69%	41,94%
33	58,52%	57,13%	61,11%	54,63%	43,70%
34	61,11%	58,33%	55,93%	55,46%	44,54%
35	57,69%	62,69%	55,00%	56,94%	49,17%
36	60,19%	60,19%	61,11%	56,48%	47,96%
37	60,65%	60,56%	60,65%	56,57%	48,61%
38	58,70%	60,93%	58,61%	60,37%	48,70%
39	60,74%	61,20%	59,26%	55,09%	44,91%
40	55,93%	58,15%	54,35%	49,54%	45,00%
41	58,24%	53,80%	52,22%	53,43%	50,65%
42	62,69%	44,81%	56,94%	53,33%	50,83%

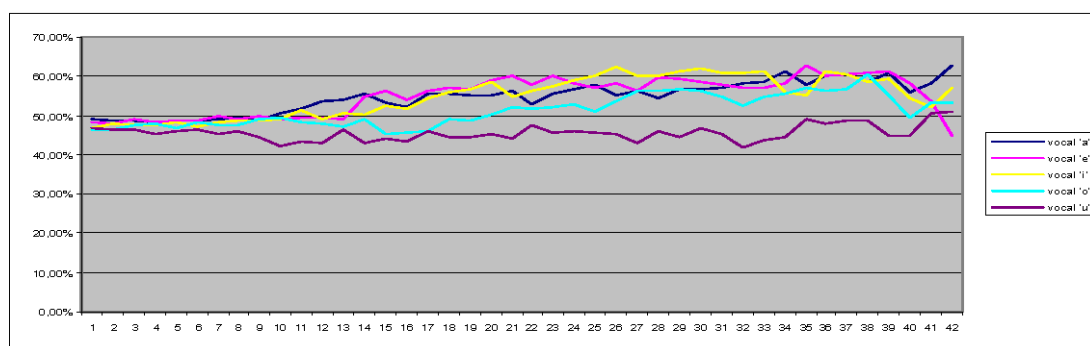


Gráfico 12-23: Evaluación secuencial de la relevancia para las medidas de calidad de la voz del grupo III usando el clasificador PNN en la clasificación entre diferentes grados de disfonía

Grupo IV

Núm. de Medidas	vocal 'a'	vocal 'e'	vocal 'i'	vocal 'o'	vocal 'u'
1	49,35%	46,67%	45,83%	48,15%	48,06%
2	49,17%	46,48%	46,48%	50,28%	48,06%
3	49,54%	47,13%	47,04%	53,15%	48,52%
4	51,39%	47,13%	45,09%	50,19%	48,80%
5	51,02%	48,06%	46,57%	52,04%	48,70%
6	51,85%	49,72%	45,46%	53,06%	46,94%
7	51,39%	53,33%	46,67%	53,33%	45,74%
8	50,83%	53,52%	45,93%	52,96%	47,87%
9	51,94%	55,74%	44,91%	54,44%	48,80%
10	52,41%	55,74%	45,19%	57,04%	48,98%
11	54,44%	57,04%	44,54%	58,61%	50,83%
12	53,52%	58,33%	45,65%	59,72%	48,24%
13	55,56%	60,65%	43,89%	59,26%	48,52%
14	56,57%	60,56%	43,98%	61,30%	51,30%
15	57,22%	61,67%	46,85%	60,83%	48,33%
16	61,94%	61,39%	44,44%	59,72%	48,06%
17	62,50%	59,07%	49,35%	61,20%	50,46%

18	63,52%	60,37%	53,98%	62,69%	49,72%
19	61,39%	61,94%	52,87%	60,37%	50,74%
20	63,52%	58,98%	55,19%	65,46%	51,57%
21	63,06%	58,98%	55,28%	63,89%	48,15%
22	64,91%	60,19%	55,74%	65,09%	49,63%
23	64,63%	62,78%	59,17%	61,94%	52,22%
24	68,06%	61,48%	56,76%	61,85%	49,72%
25	66,57%	60,74%	58,43%	60,65%	49,17%
26	63,15%	62,31%	56,39%	58,70%	48,89%
27	59,63%	57,22%	55,83%	59,17%	48,61%
28	61,39%	58,89%	51,67%	55,37%	50,09%
29	61,30%	54,17%	53,06%	51,02%	51,67%
30	59,44%	58,06%	53,15%	50,00%	47,78%
31	42,78%	59,91%	55,65%	46,11%	46,76%

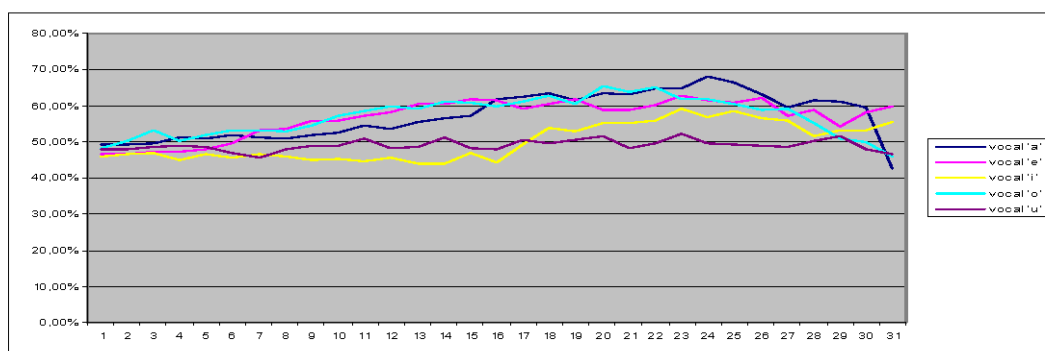


Gráfico 12-24: Evaluación secuencial de la relevancia para las medidas de calidad de la voz del grupo IV usando el clasificador PNN en la clasificación entre diferentes grados de disfonía

12.3.2.2 Red Neuronal *FeedForward* con 1 Capa Oculta (NN)

Grupo I

Núm. de Medidas	vocal 'a'	vocal 'e'	vocal 'i'	Vocal 'o'	vocal 'u'
1	57,22%	54,26%	51,57%	57,04%	54,72%
2	58,52%	54,07%	56,48%	59,17%	53,06%
3	57,69%	53,52%	56,85%	56,57%	53,43%
4	56,85%	53,43%	53,24%	58,89%	54,17%
5	58,43%	55,19%	55,19%	58,15%	53,24%
6	55,83%	54,35%	56,02%	60,00%	53,33%
7	57,87%	52,69%	57,31%	55,83%	53,98%
8	57,22%	52,78%	57,22%	58,43%	53,61%
9	55,74%	54,81%	54,17%	57,22%	53,80%
10	58,15%	52,96%	56,20%	57,78%	54,17%
11	55,65%	55,28%	53,33%	59,63%	55,09%
12	57,96%	53,33%	57,69%	58,33%	52,87%
13	59,35%	54,26%	55,65%	58,61%	54,26%
14	56,57%	56,67%	55,93%	57,69%	54,35%

15	58,52%	55,37%	51,94%	58,89%	53,43%
16	59,07%	55,00%	47,04%	55,19%	51,30%
17	50,28%	54,07%	48,15%	53,61%	50,09%
18	45,46%	55,00%	44,91%	55,09%	49,17%

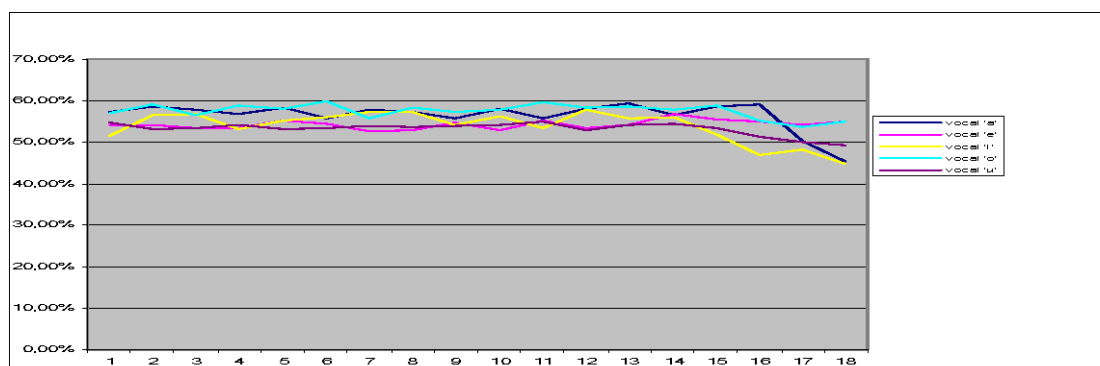


Gráfico 12-25: Evaluación secuencial de la relevancia para las medidas de calidad de la voz del grupo I usando el clasificador NN en la clasificación entre diferentes grados de disfonía

Grupo II

Núm. de Medidas	vocal 'a'	vocal 'e'	vocal 'i'	vocal 'o'	vocal 'u'
1	65,37%	59,54%	57,87%	59,44%	51,85%
2	62,87%	59,35%	56,94%	59,54%	55,93%
3	64,44%	60,37%	56,85%	58,89%	55,28%
4	64,72%	60,93%	55,56%	61,57%	55,19%
5	64,44%	62,31%	55,46%	60,83%	53,15%
6	62,96%	59,91%	55,19%	58,70%	55,09%
7	64,17%	62,69%	57,96%	57,59%	55,28%
8	64,81%	60,37%	56,20%	59,44%	55,28%
9	63,33%	59,72%	55,65%	61,11%	57,31%
10	64,07%	59,72%	55,65%	60,93%	55,19%
11	62,87%	64,07%	55,46%	58,70%	56,11%
12	60,83%	63,80%	55,65%	58,33%	56,76%
13	64,07%	63,89%	53,43%	58,52%	57,69%
14	64,63%	61,94%	55,28%	62,04%	56,76%
15	63,61%	63,33%	53,33%	61,30%	54,91%
16	65,93%	62,04%	55,00%	60,46%	56,02%
17	65,56%	64,54%	53,98%	60,19%	57,41%
18	65,09%	65,74%	55,65%	60,65%	55,93%
19	64,07%	63,98%	53,43%	61,48%	57,04%
20	64,63%	63,43%	54,72%	63,33%	56,48%
21	64,17%	62,96%	54,35%	60,93%	52,96%
22	65,28%	62,96%	55,09%	63,15%	54,44%
23	65,19%	64,07%	53,70%	62,31%	51,94%
24	65,28%	62,13%	52,87%	62,96%	53,43%
25	66,02%	62,69%	52,87%	61,30%	52,59%
26	63,98%	63,98%	51,39%	62,04%	51,76%
27	65,56%	62,22%	53,15%	60,65%	53,33%

28	65,28%	62,50%	52,69%	63,33%	52,96%
29	66,02%	62,96%	53,61%	62,78%	51,02%
30	66,76%	63,52%	52,87%	63,43%	53,15%
31	65,28%	62,50%	53,43%	59,26%	53,06%
32	67,04%	61,76%	52,22%	63,24%	52,22%
33	66,02%	62,59%	53,61%	62,50%	55,56%
34	66,94%	60,56%	52,41%	61,02%	52,78%
35	65,74%	62,96%	52,59%	62,69%	53,89%
36	67,78%	63,70%	52,59%	59,35%	54,72%
37	65,93%	65,28%	53,43%	61,30%	54,35%
38	65,19%	66,02%	52,96%	61,11%	55,09%
39	66,85%	62,22%	53,06%	60,93%	54,81%
40	65,37%	61,94%	54,72%	60,46%	52,69%
41	65,56%	62,04%	52,13%	60,56%	54,35%
42	65,00%	64,81%	52,50%	61,67%	54,26%
43	66,57%	62,78%	53,43%	62,59%	56,02%
44	66,39%	64,26%	52,78%	61,57%	54,54%
45	65,56%	65,83%	51,94%	61,57%	55,37%
46	66,02%	64,35%	52,31%	60,00%	54,63%
47	65,19%	64,54%	54,26%	62,13%	55,19%
48	65,09%	61,57%	52,96%	61,30%	53,52%
49	65,56%	64,26%	53,80%	61,57%	54,26%
50	62,87%	62,87%	53,33%	57,22%	54,44%
51	66,02%	63,43%	54,35%	60,09%	55,65%
52	65,93%	60,56%	53,43%	58,70%	54,63%
53	62,41%	45,09%	54,17%	46,76%	53,43%

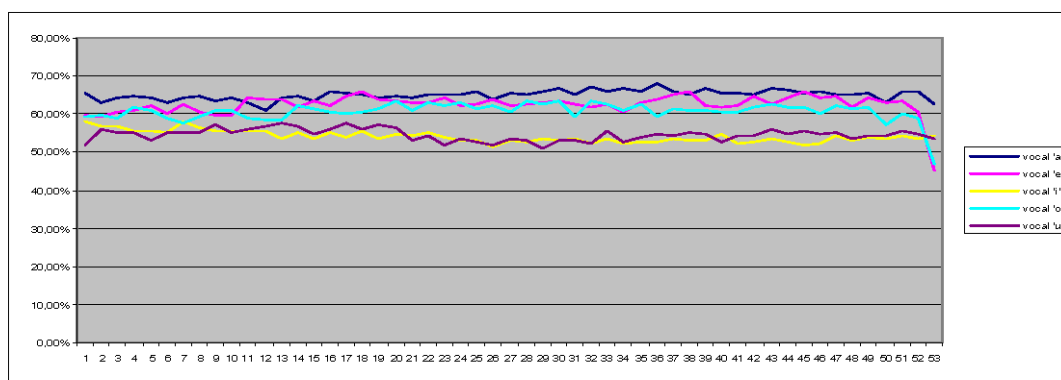


Gráfico 12-26: Evaluación secuencial de la relevancia para las medidas de calidad de la voz del grupo II usando el clasificador NN en la clasificación entre diferentes grados de disfonía

Grupo III

Núm. de Medidas	vocal 'a'	vocal 'e'	vocal 'i'	vocal 'o'	vocal 'u'
1	53,61%	52,69%	50,83%	54,07%	45,00%
2	54,91%	51,02%	51,11%	52,50%	46,94%
3	54,26%	55,00%	53,61%	52,78%	45,37%
4	54,35%	53,80%	53,70%	53,52%	44,72%
5	53,43%	53,98%	51,67%	55,93%	46,02%
6	56,85%	53,52%	52,41%	56,39%	48,70%
7	57,59%	54,54%	53,70%	57,22%	47,50%
8	58,06%	54,81%	54,72%	53,61%	45,93%
9	58,52%	55,37%	53,80%	57,50%	47,59%
10	56,39%	53,98%	55,37%	59,07%	48,70%
11	58,61%	52,96%	54,63%	56,94%	45,56%
12	59,44%	55,74%	56,39%	58,89%	47,41%
13	58,33%	56,76%	56,20%	55,74%	47,96%
14	58,06%	56,39%	57,69%	59,17%	47,13%
15	59,26%	57,87%	60,00%	58,24%	48,43%
16	57,87%	57,69%	57,04%	58,52%	48,61%
17	60,65%	57,59%	57,59%	56,94%	48,52%
18	57,87%	56,57%	55,09%	58,43%	49,17%
19	57,96%	56,85%	58,06%	56,57%	49,17%
20	59,91%	55,65%	59,07%	56,94%	48,52%
21	58,89%	57,59%	56,67%	56,94%	52,13%
22	60,46%	55,28%	60,37%	56,94%	48,43%
23	61,48%	57,04%	58,52%	58,15%	51,30%
24	56,57%	55,93%	59,07%	58,52%	53,15%
25	58,06%	54,54%	59,07%	56,76%	53,33%
26	56,48%	53,70%	59,81%	56,48%	49,63%
27	55,46%	56,39%	56,85%	56,02%	48,33%
28	61,39%	57,22%	58,33%	55,65%	51,57%
29	57,78%	55,74%	55,83%	56,76%	49,54%
30	57,78%	57,22%	59,44%	56,94%	52,69%
31	59,54%	57,04%	59,63%	57,59%	51,02%
32	58,89%	61,20%	57,59%	56,76%	52,87%
33	57,96%	57,50%	59,81%	56,76%	53,70%
34	60,93%	59,26%	58,15%	58,15%	51,57%
35	60,93%	60,93%	57,31%	55,83%	53,70%
36	59,81%	58,98%	57,59%	58,61%	52,78%
37	61,02%	60,74%	57,04%	58,61%	54,17%
38	59,72%	61,48%	56,02%	58,80%	55,09%
39	62,96%	61,57%	57,50%	57,50%	52,31%
40	61,57%	60,93%	57,31%	56,76%	55,00%
41	59,35%	57,22%	55,74%	56,67%	52,96%
42	62,31%	42,78%	53,24%	55,09%	50,93%

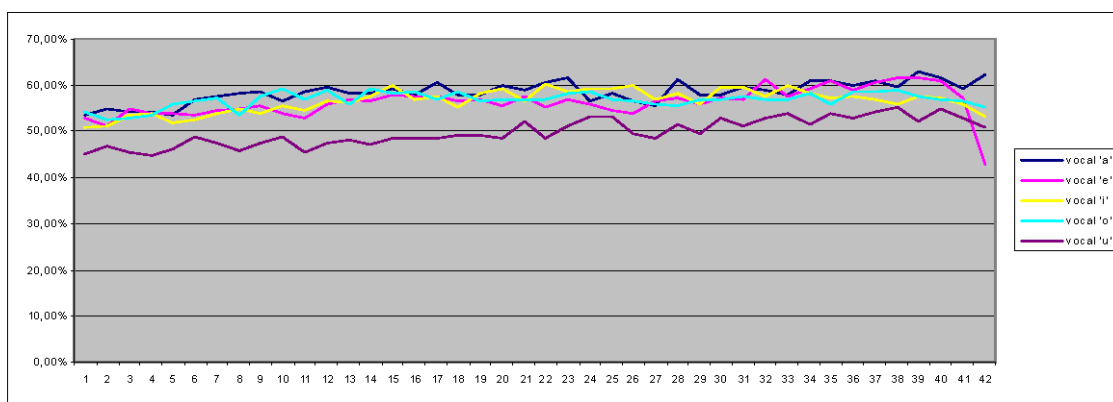


Gráfico 12-27: Evaluación secuencial de la relevancia para las medidas de calidad de la voz del grupo III usando el clasificador NN en la clasificación entre diferentes grados de disfonía

Grupo IV

Núm. de Medidas	vocal 'a'	vocal 'e'	vocal 'i'	vocal 'o'	vocal 'u'
1	58,70%	54,54%	51,67%	56,30%	49,17%
2	60,65%	54,63%	52,41%	54,81%	49,81%
3	58,24%	55,37%	51,30%	58,43%	49,81%
4	62,59%	55,19%	52,41%	56,30%	49,07%
5	61,48%	57,04%	55,74%	56,85%	50,09%
6	60,28%	53,80%	54,35%	58,43%	48,89%
7	61,76%	55,46%	54,44%	57,13%	48,89%
8	61,94%	58,80%	52,04%	57,78%	51,94%
9	63,24%	57,13%	54,72%	60,56%	50,19%
10	61,67%	56,39%	54,63%	57,87%	49,44%
11	62,50%	57,78%	52,50%	58,33%	51,85%
12	62,78%	57,31%	55,09%	59,44%	52,13%
13	63,24%	59,26%	53,98%	61,30%	50,83%
14	63,89%	58,24%	53,89%	60,83%	51,76%
15	65,46%	55,93%	56,76%	57,50%	52,13%
16	62,87%	58,24%	54,17%	60,37%	51,02%
17	63,24%	58,06%	55,09%	60,65%	51,48%
18	64,07%	60,19%	58,06%	61,48%	49,35%
19	66,30%	60,09%	55,19%	61,57%	50,37%
20	66,20%	57,41%	54,91%	58,89%	52,50%
21	65,65%	57,87%	58,70%	59,35%	50,83%
22	66,20%	60,09%	57,78%	62,13%	50,19%
23	67,69%	58,24%	60,19%	59,91%	51,30%
24	67,04%	61,20%	60,00%	61,57%	50,56%
25	67,31%	61,39%	60,37%	60,83%	52,41%
26	66,76%	62,31%	58,52%	64,44%	51,85%
27	68,24%	57,96%	59,26%	61,76%	52,59%
28	63,52%	58,15%	57,22%	60,46%	52,87%
29	64,07%	57,31%	53,98%	57,50%	49,54%
30	58,70%	58,98%	56,67%	55,37%	52,59%
31	39,17%	58,24%	53,89%	43,24%	45,28%

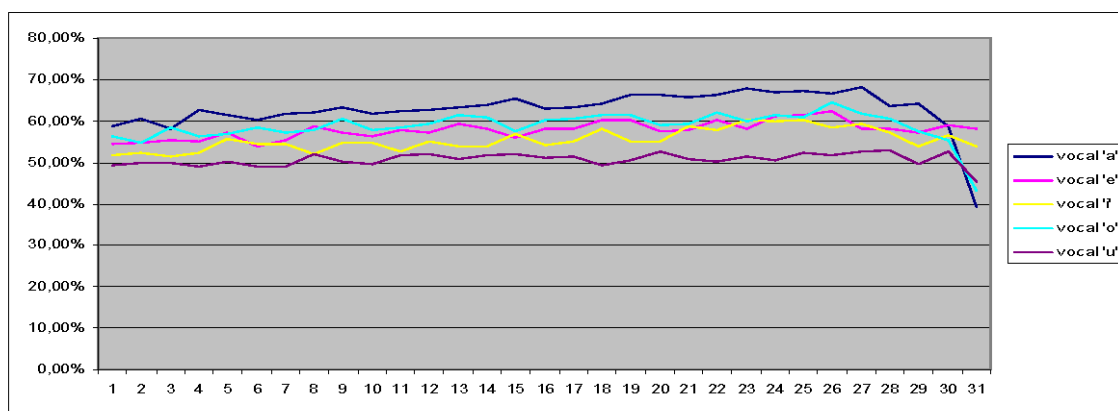


Gráfico 12-28: Evaluación secuencial de la relevancia para las medidas de calidad de la voz del grupo IV usando el clasificador NN en la clasificación entre diferentes grados de disfonía

12.3.2.3 Función de Base Radial (RBF)

Grupo I

Núm. de Medidas	vocal 'a'	vocal 'e'	vocal 'i'	Vocal 'o'	vocal 'u'
1	25,83%	27,69%	22,50%	21,85%	23,43%
2	23,89%	30,19%	20,74%	22,78%	23,80%
3	26,85%	28,61%	25,74%	24,07%	25,74%
4	25,19%	25,28%	26,57%	27,04%	23,24%
5	25,09%	25,83%	40,65%	23,33%	25,00%
6	28,15%	33,15%	38,15%	23,98%	25,46%
7	24,17%	31,02%	36,02%	23,70%	24,81%
8	34,07%	32,50%	34,44%	23,80%	25,00%
9	40,28%	35,93%	43,06%	27,59%	23,15%
10	44,81%	31,39%	42,50%	33,80%	27,69%
11	46,85%	30,09%	47,69%	31,02%	26,85%
12	45,93%	30,00%	40,65%	37,41%	29,72%
13	55,09%	32,87%	43,98%	38,43%	25,93%
14	49,63%	31,85%	45,28%	47,41%	26,94%
15	46,20%	25,93%	42,50%	41,85%	34,54%
16	40,65%	25,83%	29,54%	34,07%	33,06%
17	32,04%	36,76%	27,96%	38,61%	29,26%
18	26,39%	37,22%	30,56%	36,39%	32,31%

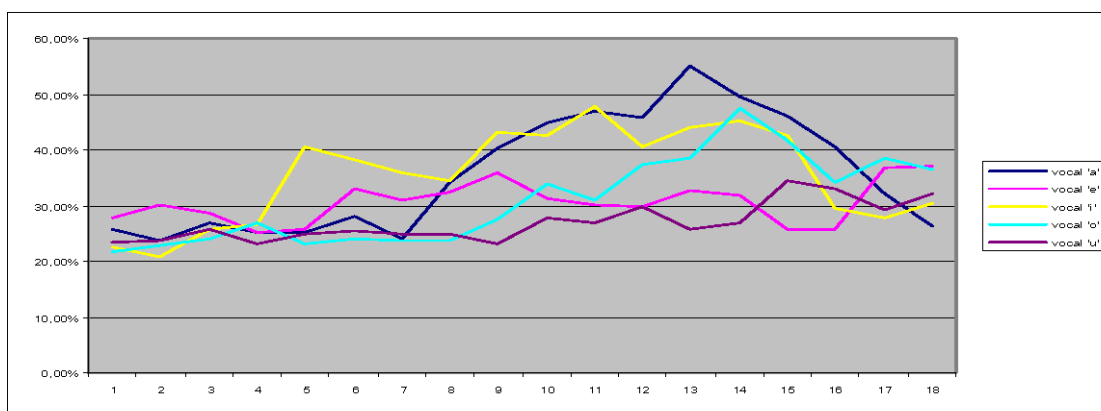


Gráfico 12-29: Evaluación secuencial de la relevancia para las medidas de calidad de la voz del grupo I usando el clasificador RBF en la clasificación entre diferentes grados de disfonía

Grupo II

Núm. de Medidas	vocal 'a'	vocal 'e'	vocal 'i'	vocal 'o'	vocal 'u'
1	27,78%	28,24%	30,74%	38,06%	33,33%
2	29,81%	30,00%	27,13%	37,87%	33,43%
3	25,00%	26,20%	25,83%	38,52%	30,28%
4	26,48%	28,24%	27,04%	38,80%	29,81%
5	25,00%	25,56%	26,11%	38,98%	31,02%
6	26,20%	26,57%	24,26%	36,30%	26,94%
7	24,17%	25,19%	24,44%	35,46%	25,46%
8	26,85%	27,22%	23,70%	34,07%	26,57%
9	25,00%	24,54%	23,24%	33,15%	24,26%
10	25,65%	26,57%	26,30%	29,35%	24,07%
11	26,11%	26,11%	30,83%	32,13%	25,19%
12	26,76%	28,70%	35,65%	31,39%	26,30%
13	25,93%	27,69%	33,33%	28,52%	27,13%
14	27,41%	25,93%	36,39%	30,46%	26,20%
15	27,41%	28,43%	38,61%	27,69%	27,59%
16	26,67%	28,43%	38,24%	30,65%	30,83%
17	26,02%	29,07%	35,19%	30,46%	26,94%
18	23,33%	32,04%	40,74%	28,52%	26,76%
19	27,87%	28,06%	40,56%	28,52%	28,33%
20	23,98%	28,52%	39,81%	26,94%	25,19%
21	25,93%	30,37%	44,91%	27,04%	26,85%
22	25,28%	25,65%	43,70%	26,94%	27,87%
23	25,65%	27,87%	40,46%	29,54%	28,06%
24	25,83%	30,09%	39,91%	27,22%	28,80%
25	26,20%	35,46%	41,57%	27,31%	29,81%
26	27,59%	30,74%	34,07%	28,98%	29,81%
27	25,93%	28,52%	46,76%	27,96%	27,41%
28	24,26%	33,06%	38,98%	26,57%	29,81%
29	26,67%	37,69%	37,41%	26,76%	27,50%
30	25,83%	34,72%	37,59%	26,57%	29,54%
31	28,70%	35,56%	37,41%	28,52%	27,50%

32	28,61%	35,46%	35,83%	27,22%	33,24%
33	29,63%	36,67%	29,63%	29,07%	31,39%
34	27,59%	38,52%	35,00%	27,04%	35,37%
35	31,11%	38,70%	36,30%	29,26%	30,46%
36	29,72%	36,57%	33,33%	29,81%	32,78%
37	30,83%	30,56%	36,94%	34,35%	41,30%
38	30,19%	36,76%	33,43%	30,56%	41,02%
39	32,96%	47,96%	28,24%	34,54%	32,41%
40	33,43%	44,17%	37,87%	38,06%	39,91%
41	44,81%	48,43%	42,87%	37,22%	37,04%
42	46,67%	48,33%	33,70%	34,35%	42,04%
43	52,41%	47,69%	37,22%	45,19%	41,02%
44	50,65%	48,89%	28,98%	39,72%	37,22%
45	52,13%	51,94%	35,74%	50,28%	32,87%
46	52,31%	51,85%	37,87%	46,11%	35,37%
47	56,76%	53,89%	27,50%	47,87%	41,20%
48	55,65%	51,94%	39,26%	48,15%	27,31%
49	52,87%	53,80%	28,98%	51,67%	32,22%
50	50,65%	50,37%	32,50%	48,33%	27,59%
51	43,24%	46,11%	28,15%	37,31%	31,94%
52	47,78%	40,00%	29,81%	38,33%	32,96%
53	30,65%	26,20%	30,56%	34,26%	25,28%

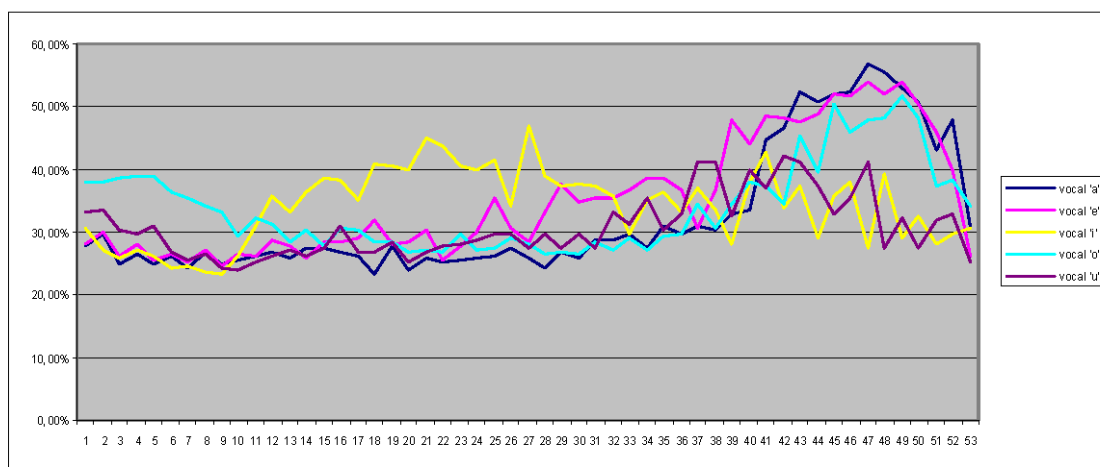


Gráfico 12-30: Evaluación secuencial de la relevancia para las medidas de calidad de la voz del grupo II usando el clasificador RBF en la clasificación entre diferentes grados de disfonía

Grupo III

Núm. de Medidas	vocal 'a'	vocal 'e'	vocal 'i'	vocal 'o'	vocal 'u'
1	48,15%	48,15%	48,15%	48,15%	48,15%
2	48,15%	48,15%	48,15%	48,15%	48,15%
3	48,15%	48,15%	48,15%	48,15%	48,15%
4	48,15%	48,15%	48,15%	48,15%	48,15%
5	48,15%	48,15%	48,15%	48,15%	48,15%
6	48,15%	48,15%	48,15%	48,15%	48,15%
7	47,96%	48,15%	48,15%	48,15%	48,15%
8	47,96%	48,15%	48,15%	48,15%	48,15%
9	48,06%	48,15%	48,15%	48,15%	48,06%
10	47,96%	48,15%	48,15%	48,15%	48,15%
11	47,87%	48,15%	48,15%	47,78%	47,69%
12	47,87%	48,15%	48,15%	47,69%	47,41%
13	47,96%	48,15%	48,15%	47,22%	47,41%
14	47,87%	48,15%	47,96%	45,65%	46,76%
15	47,69%	48,15%	47,87%	46,48%	46,57%
16	48,06%	48,15%	47,96%	45,46%	45,93%
17	47,96%	48,15%	47,69%	44,44%	45,65%
18	47,78%	48,15%	47,50%	42,69%	44,07%
19	48,24%	49,81%	46,11%	43,06%	43,52%
20	46,94%	49,72%	43,43%	41,85%	44,44%
21	47,59%	49,72%	42,78%	37,22%	43,15%
22	44,35%	48,61%	45,56%	33,61%	40,28%
23	46,39%	49,81%	46,20%	34,07%	41,57%
24	45,19%	46,20%	42,87%	34,17%	39,91%
25	45,37%	46,02%	39,44%	33,89%	40,74%
26	44,26%	44,91%	37,69%	33,15%	36,57%
27	41,48%	41,02%	36,11%	31,85%	37,22%
28	38,80%	37,13%	35,37%	26,94%	31,57%
29	34,54%	32,50%	36,11%	30,83%	30,74%
30	35,09%	33,24%	35,37%	24,72%	26,48%
31	29,44%	33,52%	32,87%	25,00%	23,52%
32	28,80%	28,43%	31,67%	27,04%	22,59%
33	26,67%	30,56%	35,19%	23,43%	22,78%
34	26,76%	30,19%	37,13%	22,41%	23,98%
35	29,44%	27,41%	36,20%	20,28%	26,02%
36	30,65%	39,91%	47,78%	25,93%	26,94%
37	42,96%	41,02%	47,13%	25,46%	26,57%
38	44,17%	56,02%	38,52%	34,17%	29,35%
39	41,39%	55,00%	38,80%	45,83%	26,48%
40	41,67%	58,98%	39,35%	44,54%	31,20%
41	44,63%	36,48%	33,80%	30,19%	27,41%
42	30,93%	33,98%	38,52%	33,98%	32,41%

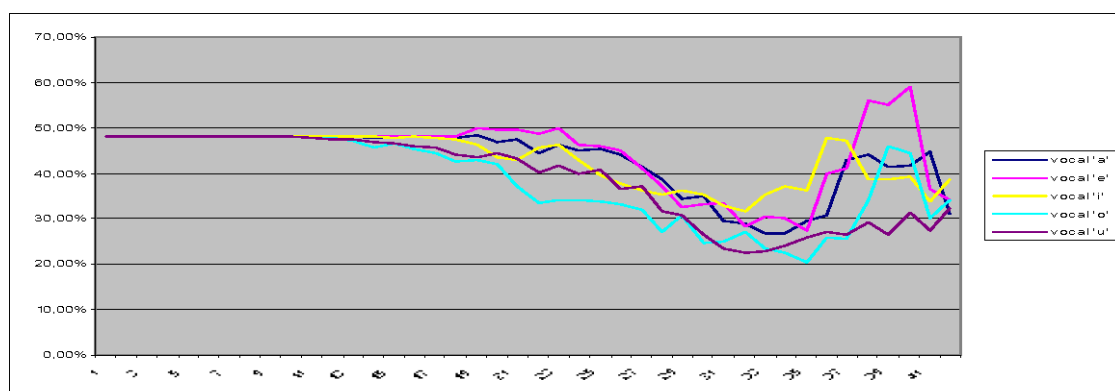


Gráfico 12-31: Evaluación secuencial de la relevancia para las medidas de calidad de la voz del grupo III usando el clasificador RBF en la clasificación entre diferentes grados de disfonía

Grupo IV

Núm. de Medidas	vocal 'a'	vocal 'e'	vocal 'i'	vocal 'o'	vocal 'u'
1	48,15%	48,15%	48,15%	48,15%	48,15%
2	48,15%	48,15%	48,15%	48,15%	48,15%
3	48,15%	48,15%	48,15%	48,15%	48,15%
4	48,15%	48,15%	48,15%	48,15%	48,06%
5	48,15%	48,15%	48,15%	48,15%	47,78%
6	48,15%	48,15%	48,15%	48,15%	47,87%
7	48,15%	48,15%	48,15%	48,15%	47,59%
8	48,15%	48,06%	48,06%	48,15%	48,06%
9	48,15%	47,69%	47,87%	48,24%	47,13%
10	48,43%	47,50%	47,50%	48,33%	47,69%
11	48,24%	47,78%	47,78%	48,61%	47,87%
12	48,70%	47,78%	47,04%	49,17%	47,04%
13	49,17%	49,26%	46,67%	49,81%	47,59%
14	49,72%	49,54%	45,00%	51,11%	45,83%
15	49,44%	47,87%	41,39%	50,00%	45,65%
16	52,59%	48,52%	38,43%	48,89%	46,76%
17	52,59%	44,07%	36,85%	48,06%	45,19%
18	51,67%	45,93%	34,07%	45,00%	42,96%
19	53,80%	43,15%	29,26%	41,76%	41,11%
20	50,09%	41,48%	23,33%	41,57%	42,69%
21	50,83%	42,50%	26,48%	38,15%	41,94%
22	46,94%	37,78%	24,81%	38,15%	37,59%
23	44,07%	41,02%	25,56%	43,24%	38,89%
24	42,41%	40,93%	25,74%	44,81%	34,35%
25	52,04%	49,44%	23,43%	52,50%	36,57%
26	54,81%	52,50%	30,28%	49,54%	32,69%
27	58,80%	55,00%	44,54%	50,46%	37,96%
28	62,41%	57,87%	51,57%	54,17%	42,87%
29	54,81%	49,35%	47,69%	42,13%	45,46%
30	48,52%	44,17%	34,72%	34,07%	36,76%
31	27,96%	30,28%	28,80%	28,43%	36,85%

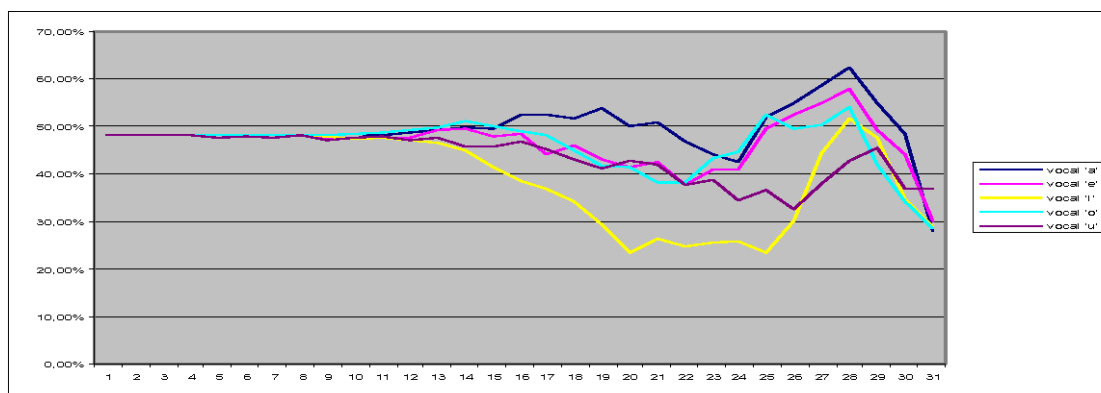


Gráfico 12-32: Evaluación secuencial de la relevancia para las medidas de calidad de la voz del grupo IV usando el clasificador PNN en la clasificación entre diferentes grados de disfonía

12.3.2.4 Máquinas de Vectores Soporte (SVM)

Grupo I

Núm. de Medidas	vocal 'a'	vocal 'e'	vocal 'i'	Vocal 'o'	vocal 'u'
1	57,78%	51,11%	51,85%	57,78%	55,74%
2	57,04%	51,67%	55,74%	57,22%	55,37%
3	57,04%	52,78%	56,11%	56,11%	55,37%
4	58,52%	52,04%	55,37%	57,04%	56,30%
5	56,67%	53,15%	54,26%	55,56%	55,19%
6	56,11%	52,96%	56,11%	55,74%	56,67%
7	57,04%	54,07%	55,93%	55,37%	57,04%
8	58,15%	53,52%	55,93%	60,74%	57,22%
9	57,78%	51,30%	54,63%	58,70%	56,30%
10	57,78%	52,96%	53,15%	57,96%	56,67%
11	58,52%	56,11%	52,59%	57,04%	57,78%
12	56,30%	52,22%	53,89%	56,30%	55,74%
13	57,04%	52,59%	54,07%	57,04%	54,63%
14	56,67%	53,52%	54,26%	54,26%	54,81%
15	59,81%	51,11%	51,11%	57,22%	50,93%
16	57,04%	53,15%	46,11%	55,37%	53,15%
17	49,44%	50,74%	48,15%	52,96%	50,19%
18	47,96%	51,67%	48,33%	54,81%	48,89%

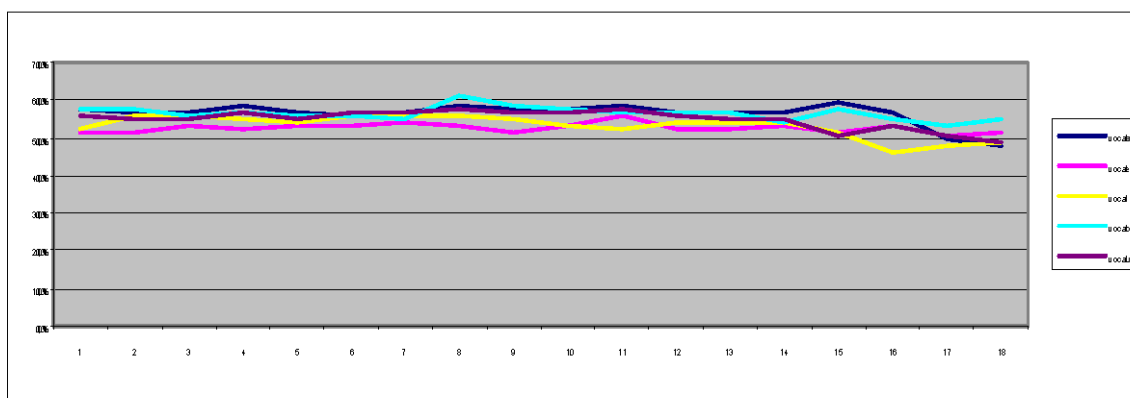


Gráfico 12-33: Evaluación secuencial de la relevancia para las medidas de calidad de la voz del grupo I usando el clasificador PNN en la clasificación entre diferentes grados de disfonía

Grupo II

Núm. de Medidas	vocal 'a'	vocal 'e'	vocal 'i'	vocal 'o'	vocal 'u'
1	64,63%	61,30%	57,04%	57,04%	53,15%
2	62,96%	61,48%	57,41%	58,15%	55,37%
3	65,19%	63,15%	56,85%	59,44%	54,07%
4	65,74%	63,33%	56,85%	57,41%	57,41%
5	61,11%	62,59%	57,59%	56,48%	54,44%
6	62,96%	61,11%	55,19%	58,70%	57,96%
7	65,37%	62,59%	53,89%	59,81%	55,37%
8	65,56%	61,30%	55,00%	60,00%	55,19%
9	65,00%	64,44%	58,33%	60,56%	54,63%
10	64,81%	62,59%	51,67%	59,63%	55,19%
11	65,00%	66,48%	55,74%	57,41%	56,48%
12	66,30%	64,07%	54,07%	60,19%	57,04%
13	66,67%	63,52%	52,78%	58,70%	54,81%
14	66,48%	65,00%	53,52%	60,19%	57,78%
15	65,74%	60,37%	51,11%	60,00%	59,81%
16	65,74%	62,22%	51,85%	58,15%	57,78%
17	64,26%	60,56%	52,59%	60,19%	59,26%
18	66,11%	62,96%	52,59%	59,07%	57,04%
19	64,26%	63,15%	54,44%	59,81%	55,56%
20	64,26%	62,22%	51,67%	60,19%	57,04%
21	65,19%	65,74%	52,41%	60,56%	51,85%
22	65,37%	63,15%	49,26%	58,70%	53,33%
23	66,67%	63,52%	51,48%	60,74%	54,44%
24	67,22%	61,67%	50,37%	61,85%	52,22%
25	65,74%	64,44%	52,22%	60,19%	53,89%
26	66,48%	64,26%	52,41%	60,00%	52,78%
27	64,63%	60,74%	52,59%	62,96%	54,26%
28	66,30%	62,78%	53,33%	61,67%	53,89%
29	64,26%	63,70%	51,30%	61,11%	52,59%
30	67,22%	62,96%	52,41%	58,89%	53,33%
31	67,41%	65,74%	49,81%	60,19%	54,44%

32	66,30%	61,30%	52,41%	60,37%	53,15%
33	64,81%	61,48%	50,93%	58,89%	52,78%
34	65,00%	62,04%	53,15%	59,63%	53,52%
35	66,30%	63,52%	50,74%	57,78%	52,22%
36	66,85%	62,59%	50,74%	57,04%	52,59%
37	61,48%	62,78%	52,04%	60,00%	51,67%
38	63,70%	63,52%	52,22%	59,63%	52,59%
39	66,85%	60,00%	52,78%	56,30%	51,67%
40	63,52%	62,22%	51,67%	59,44%	55,37%
41	65,74%	61,30%	50,37%	57,78%	53,15%
42	66,48%	64,07%	50,19%	58,70%	53,15%
43	63,70%	59,63%	52,04%	60,00%	54,81%
44	65,56%	62,78%	52,59%	60,37%	55,56%
45	64,81%	62,04%	51,11%	60,19%	53,89%
46	66,85%	62,59%	49,81%	59,07%	52,59%
47	60,74%	62,41%	50,56%	56,48%	53,70%
48	62,78%	61,85%	48,52%	56,48%	50,56%
49	64,63%	59,44%	51,11%	56,67%	50,19%
50	65,37%	62,22%	51,67%	55,74%	50,00%
51	59,44%	60,56%	48,89%	56,67%	50,37%
52	58,89%	57,96%	48,89%	54,44%	50,19%
53	56,30%	50,37%	48,70%	50,37%	47,41%

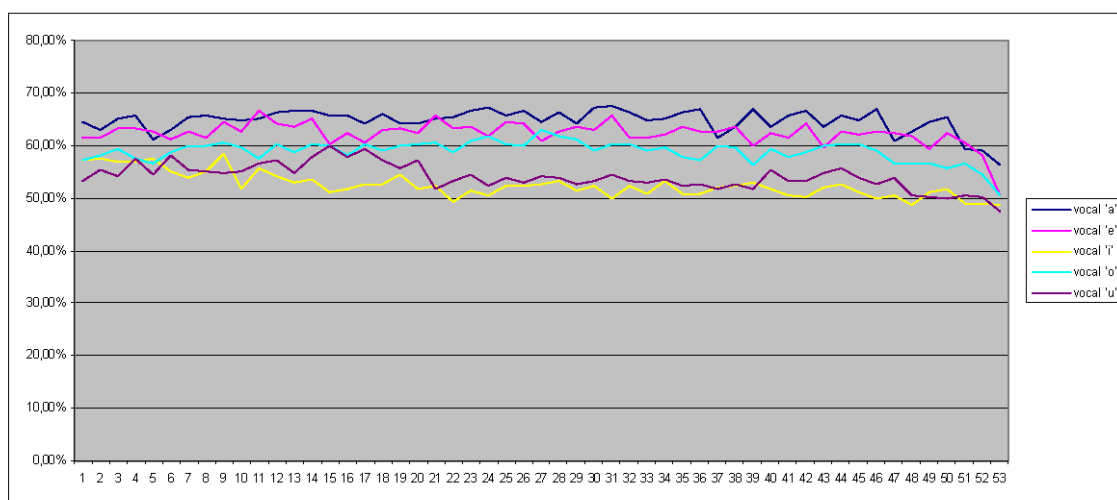


Gráfico 12-34: Evaluación secuencial de la relevancia para las medidas de calidad de la voz del grupo II usando el clasificador PNN en la clasificación entre diferentes grados de disfonía

Grupo III

Núm. de Medidas	vocal 'a'	vocal 'e'	vocal 'i'	vocal 'o'	vocal 'u'
1	61,85%	56,67%	56,48%	59,81%	53,70%
2	63,70%	57,22%	55,56%	61,67%	55,56%
3	62,41%	55,93%	56,11%	57,41%	54,44%
4	63,70%	57,04%	58,15%	58,89%	53,70%
5	62,22%	58,15%	61,48%	60,00%	54,81%
6	61,67%	56,30%	57,78%	58,89%	53,52%
7	63,33%	60,37%	56,85%	59,07%	52,41%
8	66,30%	57,78%	60,56%	60,37%	53,70%
9	65,37%	56,85%	57,78%	60,93%	54,63%
10	62,22%	53,52%	56,11%	60,93%	55,00%
11	64,81%	56,11%	60,19%	59,63%	55,56%
12	62,22%	58,33%	57,22%	60,37%	55,19%
13	65,00%	56,85%	59,81%	63,70%	53,70%
14	63,52%	59,44%	58,89%	62,22%	52,41%
15	64,44%	57,04%	62,41%	62,78%	50,93%
16	62,59%	60,00%	60,37%	60,00%	53,52%
17	61,30%	57,96%	61,85%	62,04%	52,22%
18	59,63%	57,78%	63,33%	59,26%	53,15%
19	61,30%	57,41%	61,30%	59,44%	53,89%
20	61,85%	57,41%	58,89%	57,22%	54,44%
21	62,41%	57,78%	60,00%	58,52%	53,15%
22	63,15%	59,81%	60,93%	62,96%	52,78%
23	63,70%	60,37%	59,07%	62,04%	55,00%
24	65,56%	58,70%	61,48%	61,30%	54,44%
25	62,22%	57,41%	63,33%	63,15%	54,07%
26	61,67%	57,22%	59,63%	61,85%	52,22%
27	65,19%	56,48%	60,00%	61,11%	52,22%
28	61,30%	58,15%	62,04%	59,26%	53,15%
29	62,78%	57,41%	60,93%	62,04%	52,04%
30	62,59%	59,26%	59,63%	58,70%	51,67%
31	59,26%	59,07%	61,30%	57,96%	51,85%
32	61,67%	59,44%	59,44%	57,22%	50,00%
33	61,11%	59,81%	58,15%	57,04%	50,37%
34	62,41%	59,63%	59,44%	55,93%	51,30%
35	60,37%	61,30%	57,41%	59,63%	51,67%
36	61,85%	61,67%	56,30%	57,78%	53,15%
37	61,11%	60,93%	59,07%	56,85%	51,67%
38	61,48%	60,37%	57,22%	59,26%	53,52%
39	60,00%	59,44%	53,89%	53,15%	52,22%
40	59,81%	55,93%	52,96%	54,07%	48,89%
41	57,41%	53,33%	50,56%	51,30%	49,81%
42	51,67%	48,15%	48,70%	48,33%	48,15%

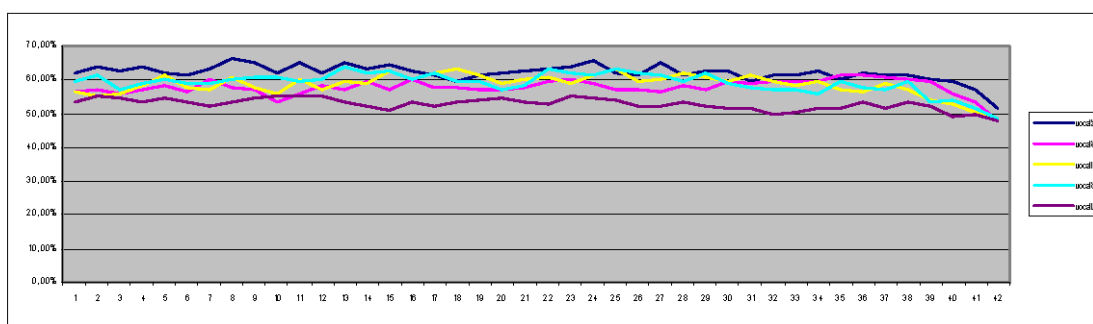


Gráfico 12-35: Evaluación secuencial de la relevancia para las medidas de calidad de la voz del grupo III usando el clasificador PNN en la clasificación entre diferentes grados de disfonía

Grupo IV

Núm. de Medidas	vocal 'a'	vocal 'e'	vocal 'i'	vocal 'o'	vocal 'u'
1	67,41%	57,41%	60,19%	61,11%	55,74%
2	66,30%	58,89%	56,48%	60,93%	52,78%
3	65,19%	59,81%	58,89%	58,89%	53,33%
4	67,41%	60,37%	57,04%	59,26%	51,85%
5	67,59%	60,19%	57,41%	63,70%	56,11%
6	67,78%	59,07%	57,96%	60,74%	52,96%
7	69,07%	59,81%	59,44%	62,41%	54,44%
8	67,04%	61,48%	56,30%	60,56%	54,81%
9	66,11%	59,63%	57,22%	60,93%	53,89%
10	67,22%	59,26%	58,15%	60,93%	54,44%
11	67,41%	59,81%	59,81%	62,41%	54,63%
12	67,04%	61,30%	56,85%	61,30%	52,59%
13	69,81%	60,00%	58,70%	60,37%	52,96%
14	68,89%	59,44%	57,59%	60,37%	52,04%
15	71,48%	61,48%	61,85%	61,85%	54,26%
16	67,41%	57,96%	57,22%	59,07%	54,44%
17	67,59%	58,15%	59,81%	60,19%	55,19%
18	68,15%	61,11%	59,81%	58,70%	55,74%
19	69,44%	57,96%	59,63%	58,52%	52,22%
20	66,67%	60,56%	57,41%	60,56%	54,44%
21	68,89%	58,70%	60,74%	59,63%	54,26%
22	68,70%	60,56%	58,70%	62,04%	52,04%
23	70,19%	64,07%	60,37%	62,41%	54,44%
24	67,96%	61,85%	60,93%	60,74%	54,26%
25	67,59%	60,93%	56,48%	62,41%	54,81%
26	65,37%	60,93%	56,11%	60,37%	54,26%
27	65,19%	59,26%	56,67%	59,26%	56,85%
28	68,15%	60,19%	52,59%	58,70%	52,22%
29	65,56%	59,26%	52,04%	49,26%	51,30%
30	53,33%	60,74%	49,07%	48,70%	49,63%
31	48,15%	61,11%	51,30%	48,15%	48,15%

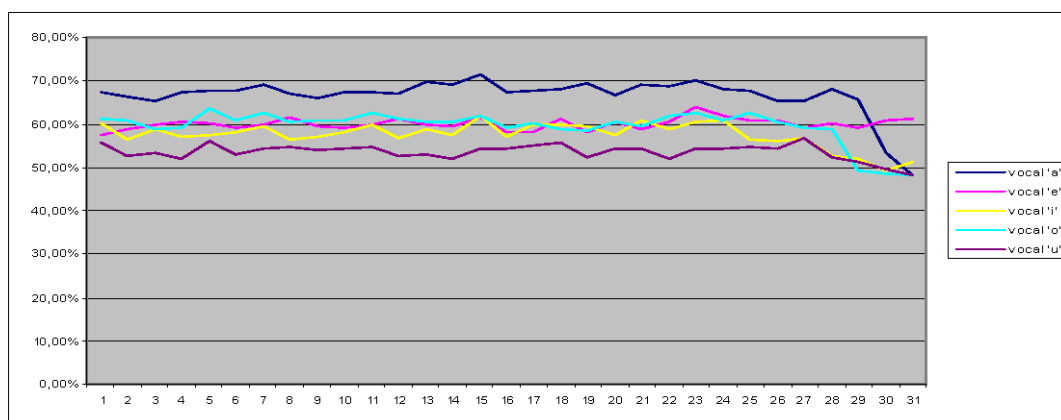


Gráfico 12-36: Evaluación secuencial de la relevancia para las medidas de calidad de la voz del grupo IV usando el clasificador PNN en la clasificación entre diferentes grados de disfonía

12.3.2.5 K vecinos más cercanos (KNN)

Grupo I

Núm. de Medidas	vocal 'a'	vocal 'e'	vocal 'i'	Vocal 'o'	vocal 'u'
1	57,87%	54,63%	53,61%	55,37%	55,19%
2	57,04%	56,67%	58,52%	55,56%	58,89%
3	59,72%	54,63%	55,46%	55,09%	57,87%
4	60,09%	56,76%	55,37%	56,30%	57,78%
5	58,52%	57,13%	55,09%	55,83%	57,13%
6	57,96%	54,26%	53,61%	55,19%	58,24%
7	58,33%	55,65%	56,30%	56,11%	58,06%
8	57,87%	53,24%	56,11%	57,50%	56,76%
9	59,81%	55,37%	56,11%	56,30%	57,96%
10	60,28%	54,91%	55,65%	58,24%	57,96%
11	58,24%	55,93%	56,76%	57,69%	59,35%
12	57,78%	53,43%	54,44%	57,78%	55,74%
13	60,09%	55,93%	54,91%	59,26%	55,56%
14	60,09%	57,04%	53,61%	60,56%	52,96%
15	60,28%	56,67%	54,26%	60,46%	54,81%
16	58,70%	58,24%	52,69%	58,15%	55,46%
17	55,93%	55,74%	49,44%	59,07%	56,02%
18	46,57%	55,93%	50,65%	59,44%	52,69%

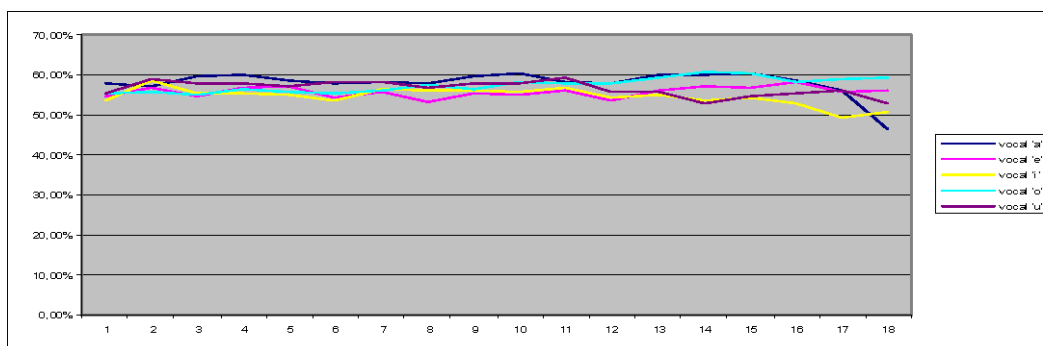


Gráfico 12-37: Evaluación secuencial de la relevancia para las medidas de calidad de la voz del grupo I usando el clasificador KNN en la clasificación entre diferentes grados de disfonía

Grupo II

Núm. de Medidas	vocal 'a'	vocal 'e'	vocal 'i'	vocal 'o'	vocal 'u'
1	64,63%	62,41%	58,89%	60,46%	55,09%
2	64,91%	61,94%	59,72%	60,37%	58,33%
3	65,00%	62,22%	59,54%	61,11%	55,19%
4	65,46%	62,41%	60,37%	58,61%	56,30%
5	63,89%	62,78%	58,89%	59,91%	58,52%
6	62,04%	62,69%	56,39%	61,57%	56,30%
7	64,91%	62,96%	57,13%	60,28%	56,85%
8	64,44%	61,02%	56,39%	60,19%	56,94%
9	65,09%	64,54%	56,48%	59,63%	58,61%
10	66,76%	62,41%	55,19%	62,04%	57,96%
11	64,54%	60,28%	56,85%	59,63%	57,78%
12	64,81%	61,85%	54,81%	61,39%	58,61%
13	65,74%	62,96%	50,65%	62,04%	59,35%
14	66,11%	62,59%	50,74%	59,81%	58,24%
15	63,43%	62,69%	53,52%	59,91%	58,43%
16	68,15%	62,96%	52,22%	62,31%	58,43%
17	63,70%	59,44%	51,11%	59,35%	60,00%
18	65,09%	62,50%	52,50%	60,65%	58,06%
19	64,54%	62,59%	51,57%	58,89%	59,72%
20	66,57%	62,69%	52,78%	61,30%	58,52%
21	64,72%	63,33%	52,31%	59,63%	54,26%
22	66,76%	62,04%	51,57%	59,17%	55,65%
23	66,39%	62,78%	52,41%	61,02%	56,76%
24	64,81%	64,35%	52,50%	59,63%	55,74%
25	64,17%	62,96%	51,76%	59,81%	56,48%
26	66,67%	61,39%	51,85%	58,43%	55,00%
27	66,39%	62,04%	51,57%	58,15%	54,44%
28	65,93%	60,83%	54,07%	59,72%	54,81%
29	65,28%	63,61%	52,50%	59,72%	54,91%
30	65,74%	62,69%	50,83%	61,20%	54,44%
31	66,02%	61,48%	52,41%	60,65%	53,06%
32	65,65%	62,69%	52,59%	60,00%	53,98%
33	66,20%	62,87%	52,69%	59,72%	53,24%

34	66,39%	63,52%	53,43%	59,91%	54,44%
35	66,11%	62,87%	52,04%	60,28%	53,52%
36	64,72%	60,93%	50,83%	61,48%	54,17%
37	67,69%	63,24%	52,31%	60,83%	54,17%
38	66,67%	62,13%	54,26%	61,02%	55,37%
39	68,70%	62,87%	53,15%	60,83%	54,44%
40	69,26%	63,24%	54,07%	61,57%	55,93%
41	65,83%	63,43%	53,52%	59,17%	53,61%
42	65,93%	65,46%	52,41%	60,28%	53,98%
43	69,63%	65,00%	53,98%	61,94%	53,61%
44	69,07%	61,76%	53,89%	60,74%	53,70%
45	68,98%	68,80%	55,00%	60,93%	54,07%
46	70,19%	65,00%	52,50%	59,17%	54,72%
47	63,70%	62,78%	53,06%	61,39%	55,09%
48	63,80%	62,87%	51,85%	58,43%	55,93%
49	63,33%	64,35%	50,56%	59,72%	54,72%
50	64,81%	67,22%	52,50%	58,15%	53,61%
51	67,69%	64,91%	56,20%	59,17%	53,98%
52	65,09%	66,02%	55,00%	57,69%	54,63%
53	64,17%	46,11%	55,37%	49,91%	51,39%

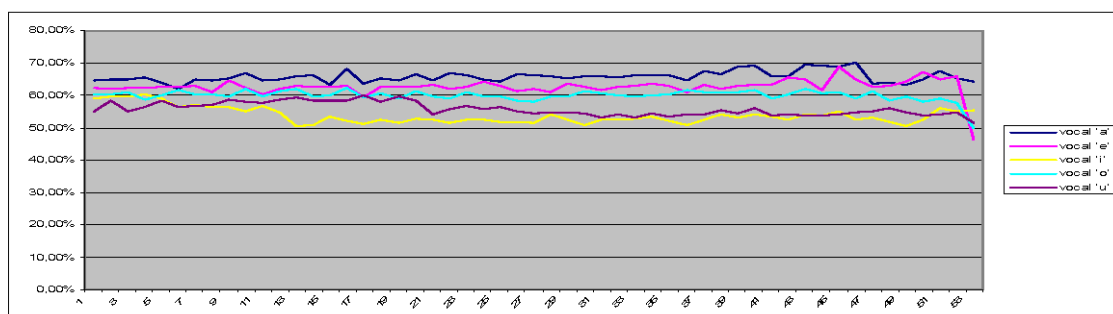


Gráfico 12-38: Evaluación secuencial de la relevancia para las medidas de calidad de la voz del grupo II usando el clasificador KNN en la clasificación entre diferentes grados de disfonía

Grupo III

Núm. de Medidas	vocal 'a'	vocal 'e'	vocal 'i'	vocal 'o'	vocal 'u'
1	60,09%	58,15%	52,69%	56,57%	52,50%
2	60,28%	56,39%	53,98%	56,76%	52,87%
3	63,15%	58,33%	54,17%	58,70%	55,19%
4	62,59%	58,24%	55,09%	57,04%	52,50%
5	62,69%	57,04%	55,19%	56,94%	52,96%
6	62,04%	58,61%	53,98%	56,11%	52,96%
7	61,85%	59,17%	56,67%	57,31%	54,17%
8	63,52%	59,26%	56,94%	56,76%	54,26%
9	61,85%	58,33%	56,02%	57,22%	53,98%
10	63,24%	60,56%	58,15%	58,43%	53,15%
11	61,67%	60,37%	56,48%	57,13%	53,80%

12	63,06%	60,37%	56,20%	57,87%	53,24%
13	62,22%	61,85%	54,91%	59,54%	54,54%
14	63,70%	63,61%	56,67%	60,00%	53,89%
15	63,70%	61,48%	55,74%	55,83%	51,76%
16	62,96%	61,02%	57,69%	59,54%	52,13%
17	65,93%	62,13%	58,06%	59,81%	53,70%
18	63,15%	62,41%	58,06%	57,69%	53,80%
19	61,94%	63,15%	59,54%	56,57%	53,24%
20	64,26%	64,07%	61,20%	59,17%	54,07%
21	62,50%	62,59%	58,43%	58,06%	54,44%
22	63,24%	61,57%	60,09%	59,26%	53,61%
23	63,61%	61,30%	58,89%	58,43%	52,50%
24	62,78%	61,67%	60,46%	59,35%	53,80%
25	61,30%	60,37%	62,22%	58,80%	52,96%
26	61,20%	62,04%	60,28%	59,07%	54,54%
27	62,13%	62,22%	58,89%	58,70%	51,76%
28	62,22%	61,30%	61,57%	58,80%	53,33%
29	62,04%	61,67%	62,13%	59,72%	53,98%
30	59,35%	60,37%	60,46%	58,43%	55,09%
31	59,81%	59,44%	59,44%	58,80%	52,04%
32	61,67%	65,19%	63,24%	59,54%	49,91%
33	59,72%	62,78%	64,72%	57,78%	51,76%
34	62,69%	65,09%	61,39%	56,94%	54,07%
35	58,33%	63,80%	61,39%	58,98%	50,28%
36	58,98%	62,41%	60,00%	57,04%	53,43%
37	59,07%	61,67%	58,06%	55,56%	54,35%
38	57,87%	62,50%	58,61%	57,13%	51,76%
39	61,11%	65,46%	57,13%	58,24%	50,46%
40	60,37%	63,98%	57,31%	57,87%	53,06%
41	58,80%	63,33%	57,59%	55,56%	53,33%
42	60,19%	48,43%	54,91%	52,69%	50,00%

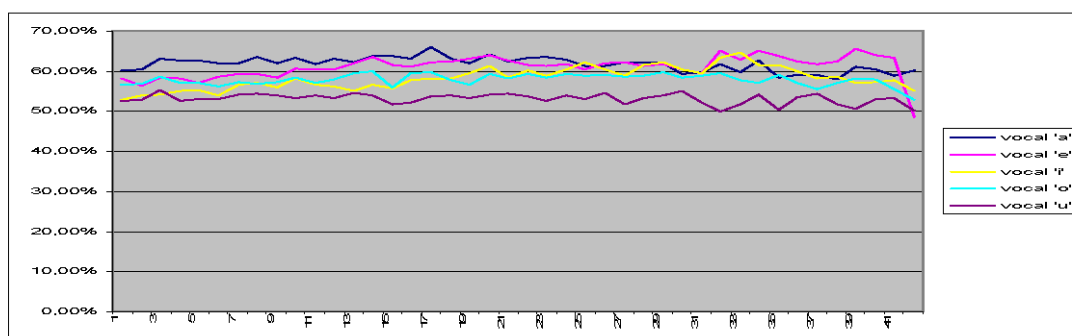


Gráfico 12-39: Evaluación secuencial de la relevancia para las medidas de calidad de la voz del grupo III usando el clasificador KNN en la clasificación entre diferentes grados de disfonía

Grupo IV

Núm. de Medidas	vocal 'a'	vocal 'e'	vocal 'i'	vocal 'o'	vocal 'u'
1	64,26%	58,98%	54,07%	61,85%	53,15%
2	64,63%	58,52%	55,37%	59,17%	51,48%
3	65,56%	60,00%	54,44%	62,59%	51,02%
4	67,50%	60,65%	53,52%	61,11%	51,20%
5	66,48%	57,04%	54,26%	60,74%	51,39%
6	67,04%	58,61%	55,65%	63,52%	52,69%
7	66,76%	59,07%	56,39%	62,59%	52,13%
8	65,83%	60,37%	55,19%	61,02%	50,00%
9	65,93%	58,98%	56,30%	62,69%	53,61%
10	67,78%	57,50%	54,81%	63,80%	51,94%
11	66,48%	59,91%	56,85%	62,78%	53,61%
12	66,85%	59,26%	54,91%	63,80%	53,24%
13	67,96%	61,57%	56,57%	64,81%	51,94%
14	69,17%	59,35%	56,57%	63,98%	53,06%
15	66,57%	61,39%	56,48%	64,35%	52,04%
16	67,87%	59,17%	56,20%	63,98%	53,33%
17	70,37%	61,30%	59,07%	63,80%	54,07%
18	72,13%	61,94%	57,41%	63,89%	54,72%
19	71,11%	60,65%	58,61%	61,76%	53,52%
20	68,89%	61,11%	60,74%	64,81%	53,70%
21	68,70%	63,80%	57,13%	61,57%	55,28%
22	71,48%	65,28%	58,06%	63,70%	55,74%
23	71,48%	66,48%	60,56%	63,80%	55,56%
24	70,19%	62,13%	60,37%	64,07%	50,46%
25	67,04%	60,93%	61,02%	64,81%	52,13%
26	71,11%	60,83%	61,30%	65,00%	52,69%
27	70,00%	60,00%	60,46%	63,33%	54,07%
28	70,09%	61,11%	57,69%	59,26%	54,26%
29	70,56%	61,76%	60,46%	56,11%	51,20%
30	60,37%	58,70%	56,11%	55,83%	53,98%
31	46,48%	61,67%	56,48%	45,19%	50,93%

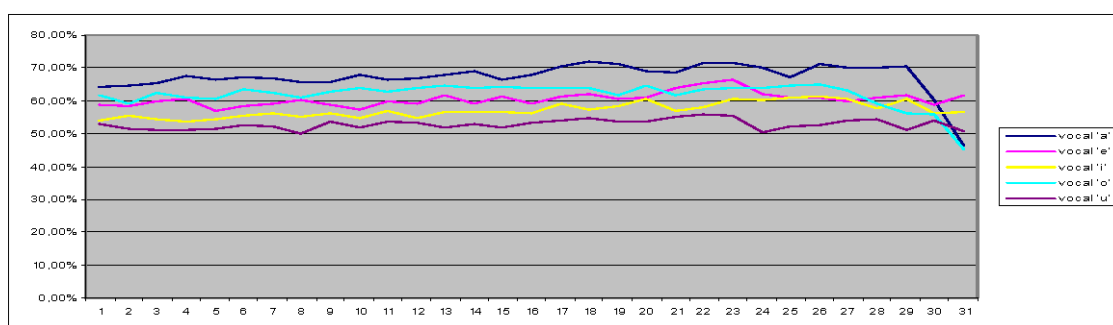


Gráfico 12-40: Evaluación secuencial de la relevancia para las medidas de calidad de la voz del grupo IV usando el clasificador KNN en la clasificación entre diferentes grados de disfonía

12.4 Referencias

- [1] <http://www.prtools.org/> [consulta realizada el 10/03/06].
- [2] R. Hegger, H. Kantz, and T. Schreiber, *Practical implementation of nonlinear time series methods: The TISEAN package*, [CHAOS](#) **9**, 413 (1999) <http://www.ie.ncsu.edu/mirage/GAToolBox/gaot/> [consulta realizada el 10/03/06]
- [3] <http://www.mathtools.net/> [consulta realizada el 10/03/06]
- [4] Chris Houck, Jeff Joines, a, Mike Kay, "A Genetic Algorithm for Function Optimization: A Matlab Implementation", NCSU-IE TR 95-09, 1995.