



Revista INGENIERÍA UC

ISSN: 1316-6832

revistaing@uc.edu.ve

Universidad de Carabobo

Venezuela

Pino, Paulino Del; Díaz, José A.; Jiménez, Carlos; Rothman, Howard B.  
Identificación de algunos parámetros espectrales que determinan la calidad de la voz  
Revista INGENIERÍA UC, vol. 11, núm. 3, diciembre, 2004, pp. 7-16  
Universidad de Carabobo  
Valencia, Venezuela

Disponible en: <http://www.redalyc.org/articulo.oa?id=70711302>

- Cómo citar el artículo
- Número completo
- Más información del artículo
- Página de la revista en redalyc.org

redalyc.org

Sistema de Información Científica

Red de Revistas Científicas de América Latina, el Caribe, España y Portugal

Proyecto académico sin fines de lucro, desarrollado bajo la iniciativa de acceso abierto

## Identificación de algunos parámetros espectrales que determinan la calidad de la voz

*Paulino Del Pino<sup>(1)</sup>, José A. Díaz<sup>(1)</sup>, Carlos Jiménez<sup>(1)</sup>, Howard B. Rothman<sup>(3)</sup>*

<sup>(1)</sup> *Escuela de Ingeniería Eléctrica, Facultad de Ingeniería, Universidad de Carabobo, Valencia, Venezuela*

<sup>(2)</sup> *Department of Electrical and Computer Engineering, University of Florida*

<sup>(3)</sup> *Department of Communication Sciences and Disorders, University of Florida*

*Email: pdelpi@uc.edu.ve, jadiaz@uc.edu.ve, cjimenez@uc.edu.ve*

### Resumen

El trabajo que se presenta en este artículo tiene como objetivo principal contribuir en el proceso de clasificación de la calidad de la señal de voz, al aportar parámetros que permiten realizar la mencionada operación. Esta contribución se logra mediante el análisis de un conjunto de parámetros propuestos (PMR, SNR, SNRL, SNRM, SNRH), obtenidos mediante el análisis de las señales de voz en el dominio de la frecuencia, es decir, estudiando su espectro de frecuencias a través de la transformada discreta de Fourier. Como resultado de este trabajo se analizaron los espectros de magnitud de veinticuatro muestras de voces normales y patológicas, se propusieron cinco parámetros los cuales se analizaron estadísticamente obteniéndose como resultado tres parámetros significativos: PMR, SNRM y SNRH. Estos parámetros no se han encontrado en los reportes de investigaciones realizadas a pesar de existir un gran número de trabajos, en el sentido de obtener información para calificar la calidad de una señal de voz, por lo que se considera un aporte novedoso.

**Palabras clave:** Voz, calidad, espectro, DFT .

## Identification of some spectral parameters that determine voice quality

### Abstract

The main objective of this paper is to contribute in the process of classification of voice quality. This contribution is achieved by means of the analysis of a set of proposed parameters (PMR, SNR, SNRL, SNRM, SNRH), obtained at the frequency domain. These parameters were obtained from the frequency domain after transforming the signal using the DFT. As a result of this work twenty four voices sampling of pathological and good voices were analysed, several parameters were proposed and the results were statistically tested, which yielded three statistically significant parameters PMR, SNRM and SNRH. These parameters have not been found in the reports of investigations performed by others researchers in spite of the great number of articles in this field. Therefore this work can be considered a novel contribution.

**Keywords:** voice, quality, spectrum, DFT.

### 1. INTRODUCCIÓN

En gran medida los métodos de caracterización y determinación de la calidad de la voz siguen procesos donde predomina el carácter subjetivo, es decir, que dependen de la representación mental de lo que es percibido por la persona receptora de la información hablada, esto es, del oyente. Particularmente, cuando se trata de establecer la calidad de la voz en términos médicos, como voces sanas o patológicas, la calificación es lograda después de la realización de exámenes físicos al sistema fonador humano.

Es lógico pensar que desde el punto de vista espectral, las señales de voz tienen características particulares, y en cada caso la calidad de la voz estará asociada con estas propiedades espectrales, esto último sugiere la posibilidad de establecer parámetros en el dominio de la frecuencia que permitan la evaluación de estas señales, y con éstos realizar de una forma menos subjetiva y más útil la calificación de la voz o el diagnóstico a pacientes con problemas de fonación. Para lograr esto es imprescindible identificar que

## Identificación de algunos parámetros espectrales que determinan la calidad de la voz

parámetros del espectro son característicos a estas señales.

Se propone entonces la identificación de algunos parámetros espectrales que caracterizan y determinan la calidad de la voz, de tal manera de lograr una mejor aproximación en la clasificación de los mismos de acuerdo a sus propiedades, esto se realizó a través de un análisis sistemático mediante técnicas de procesado digital de señales. Estas técnicas aplicadas a las señales de voz digitalizada permitirán determinar las propiedades espectrales. Así se dispondrá de mayor información objetiva para la clasificación y determinación de la calidad de voz.

La propuesta está fundamentada en el análisis del dominio de la frecuencia, específicamente de los indicadores de ruido de un conjunto de muestras de señales de voz calificadas, unas como sanas y otras como patológicas a nivel clínico, con el objeto de identificar si existen parámetros en el dominio de la frecuencia que permitan la calificación anteriormente señalada.

### 2. BASE DE DATOS UTILIZADA

Se debe aclarar antes de describir la base de datos utilizada en los análisis, que trataremos con señales de voz representadas en el dominio del tiempo discreto, es decir digitalizadas, por lo que las técnicas asociadas serán las correspondientes a esta disciplina, todas ellas realizadas a través del computador como herramienta fundamental y de algoritmos previstos para el fin señalado.

Para el desarrollo de este trabajo se dispuso de una base de datos de muestras de voces sanas y de muestras de voces patológicas, la cual consta de 12 locutores sanos y 12 locutores patológicos. La muestra de voz de cada locutor corresponde a la emisión de la vocal "a" en forma sostenida durante un intervalo de tiempo de al menos tres segundos de duración.

Las muestras de voz se grabaron en formato .WAV a 44100 muestras por segundo y con 16 bits por muestra. Se supone que durante el proceso de grabación se tomaron las precauciones necesarias para que las señales de audio no tuviesen un nivel de intensidad inadecuado ni presentar saturación.

Es importante en este punto señalar que la clasificación, como sana o patológica, de cada muestra de voz es conocida de antemano, es por ello que el objetivo principal del análisis es identificar sus diferencias y proponer sobre los resultados algún parámetro espectral, es decir, en el dominio de la frecuencia, que las diferencien de manera significativa.

En la Tabla 1 pueden observarse las características generales de cada archivo de muestra de voz obtenido como se describió anteriormente.

Tabla 1. Muestras de voz digitalizadas.

|                            | Nº  | Muestra de voz | Duración (s) | Nº de Muestras $F_s=44.100$ Muestras/s |
|----------------------------|-----|----------------|--------------|--|
| Muestras de voz sana       | b1  | 1101-1ah       | 5.61         | 247271                                 |
|                            | b2  | 1102-6ah       | 5.34         | 235424                                 |
|                            | b3  | 1104-5ah       | 5.79         | 255551                                 |
|                            | b4  | 1106-6ah       | 5.86         | 258264                                 |
|                            | b5  | 1108-8ah       | 5.38         | 237232                                 |
|                            | b6  | 1109-9ah       | 5.73         | 252879                                 |
|                            | b7  | 2101-6ah       | 5.89         | 259967                                 |
|                            | b8  | 2104-7ah       | 5.58         | 246002                                 |
|                            | b9  | 2106-5ah       | 5.43         | 239543                                 |
|                            | b10 | 2107ah         | 5.56         | 245619                                 |
|                            | b11 | 2201-6ah       | 5.62         | 124111                                 |
|                            | b12 | 2205-5ah       | 6.12         | 270135                                 |
| Muestras de voz patológica | m1  | 2209-9ahBAD    | 5.56         | 245132                                 |
|                            | m2  | 2204-4ahBAD    | 5.68         | 250755                                 |
|                            | m3  | 2207-7ahBAD    | 5.81         | 256284                                 |
|                            | m4  | 2208-7ahBAD    | 5.74         | 253205                                 |
|                            | m5  | 3101-2ahBAD    | 5.82         | 256664                                 |
|                            | m6  | 3102-5ahBAD    | 3.31         | 145837                                 |
|                            | m7  | 3103-5ahBAD    | 5.73         | 252657                                 |
|                            | m8  | 3105-5ahBAD    | 4.20         | 185080                                 |
|                            | m9  | 3106-5ahBAD    | 3.48         | 153521                                 |
|                            | m10 | 3201-9ahBAD    | 6.58         | 290181                                 |
|                            | m11 | 3203-4ahBAD    | 6.52         | 287721                                 |
|                            | m12 | 3207-3ahBAD    | 5.95         | 262387                                 |

### 3. METODOLOGÍA

El procedimiento realizado en este trabajo para la identificación de los parámetros espectrales que determinan la calidad de la voz puede resumirse de la forma siguiente:

1. Revisión de los modelos teóricos de producción de la señal de voz y de análisis de la misma en el dominio de la frecuencia.
2. Diseño de un experimento para evaluar las características espectrales de las señales a estudiar constituido por:
  - a. Descomposición y análisis en el dominio de la frecuencia (Transformada discreta de Fourier), de doce muestras calificadas como “sanas” y doce muestras calificadas como “patológicas”.
  - b. Identificación de parámetros del espectro de las muestras señaladas a objeto de establecer las diferencias apreciables visualmente y formulación de las hipótesis correspondientes.
  - c. Cálculo de los parámetros espectrales de las mismas.
  - d. Aplicación de Análisis de Varianza ANOVA para establecer la existencia de diferencias significativas entre los grupos de voces analizados en lo que respecta a los parámetros formulados en 3c.
3. Proposición de parámetros espectrales que determinan la calidad de la voz.

Para emprender la tarea de analizar en el dominio de la frecuencia las señales de voz de las muestras sanas y patológicas, es necesario diseñar el experimento con el que se tratará a cada muestra, el mismo consta de una inspección visual a su espectro de frecuencias y una audición de la misma a efectos de confirmar cualitativamente la clasificación de las mismas anteriormente señalada, es decir, como voz patológica y voz sana. De la observación derivada de esta inspección se establecieron un conjunto de condiciones para el análisis, se formuló una hipótesis en lo que respecta al parámetro, se determinaron varios parámetros y se evaluaron estadísticamente a los fines de verificar su validez. Puede entonces esquematizarse el experimento realizado según indica la Figura 1.

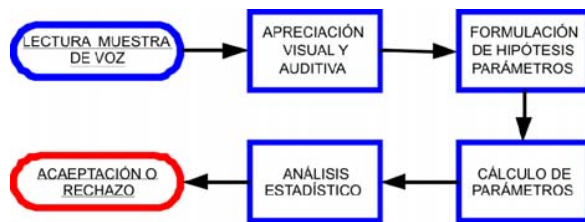


Figura 1. Proceso de análisis de las muestras de voz.

En la fase de apreciación visual y auditiva de

Tabla 2. Resolución espectral de cada señal de voz.

|   | Nº  | Muestra de voz | Nº de Muestras<br>Fs=44.10<br>Muestras/s | Resolución Espectral<br>(Fs/N)<br>Hz |
|---|-----|----------------|--|--------------------------------------|
| Espectros de Muestras de Voz Sanas      | b1  | 1101-1ah       | 247271                                   | 0.178                                |
|   | b2  | 1102-6ah       | 235424                                   | 0.187                                |
|   | b3  | 1104-5ah       | 255551                                   | 0.173                                |
|   | b4  | 1106-6ah       | 258264                                   | 0.171                                |
|   | b5  | 1108-8ah       | 237232                                   | 0.186                                |
|   | b6  | 1109-9ah       | 252879                                   | 0.174                                |
|   | b7  | 2101-6ah       | 259967                                   | 0.170                                |
|   | b8  | 2104-7ah       | 246002                                   | 0.179                                |
|   | b9  | 2106-5ah       | 239543                                   | 0.184                                |
|   | b10 | 2107ah         | 245619                                   | 0.180                                |
|   | b11 | 2201-6ah       | 124111                                   | 0.178                                |
|   | b12 | 2205-5ah       | 270135                                   | 0.163                                |
| Espectros de Muestras de voz Patológica | m1  | 2209-9ahBAD    | 245132                                   | 0.180                                |
|   | m2  | 2204-4ahBAD    | 250755                                   | 0.176                                |
|   | m3  | 2207-7ahBAD    | 256284                                   | 0.172                                |
|   | m4  | 2208-7ahBAD    | 253205                                   | 0.174                                |
|   | m5  | 3101-2ahBAD    | 256664                                   | 0.172                                |
|   | m6  | 3102-5ahBAD    | 145837                                   | 0.302                                |
|   | m7  | 3103-5ahBAD    | 252657                                   | 0.175                                |
|   | m8  | 3105-5ahBAD    | 185080                                   | 0.238                                |
|   | m9  | 3106-5ahBAD    | 153521                                   | 0.287                                |
|   | m10 | 3201-9ahBAD    | 290181                                   | 0.152                                |
|   | m11 | 3203-4ahBAD    | 287721                                   | 0.153                                |
|   | m12 | 3207-3ahBAD    | 262387                                   | 0.168                                |

cada muestra de audio, se escribió un programa para la presentación del espectro en frecuencias y simultáneamente la reproducción del sonido de cada archivo.

Si bien es común decir que la señal de voz no es estacionaria, aquí se ha considerado lo contrario, tomando en cuenta que se trata de la emisión de una vocal en forma sostenida durante un tiempo relativamente corto en lo que respecta al esfuerzo físico de quien lo emite, por lo que se asume que las variaciones estadísticas de la misma son despreciables, esto permite determinar el espectro de la señal tomando la DFT del vector de muestras de voz digitalizada en toda su extensión, es decir, considerando todos los puntos y no haciendo estimaciones mediante enventanado durante un segmento corto de la señal.

Las ventajas que se obtienen de la consideración anterior, en lo que respecta al objetivo de este trabajo, es tener un espectro lo más detallado posible utilizando la mejor resolución disponible que en este caso es Fs/N, donde N es el número de muestras de la

## Identificación de algunos parámetros espectrales que determinan la calidad de la voz

señal de voz, o bien el número de elementos del vector de muestras de voz digitalizada, y  $F_s$  es la frecuencia de muestreo con la que se digitalizó el archivo.

Los resultados relativos a la resolución de cada espectro calculado de la manera antes señalada se muestran en la Tabla 2.

A continuación, y a manera de ilustración, la Figura 2 muestra el espectro de la señal b1, correspondiente al archivo 1101-1ah de las muestras de voz sanas y al archivo 2202-9ahBAD, correspondiente a una de las muestras de voz patológicas.

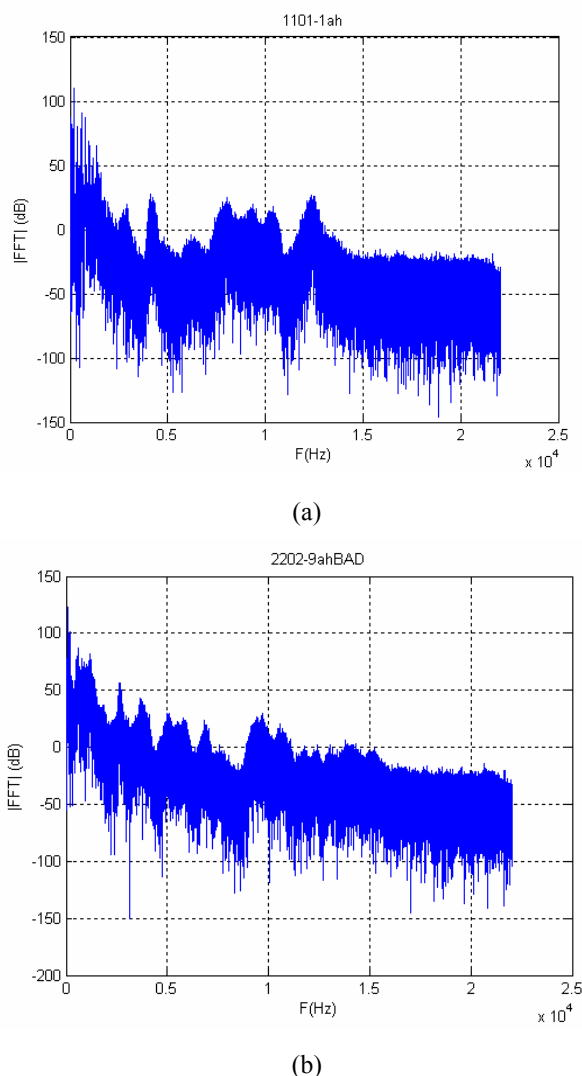


Figura 2. Espectros de las señales: a) b1: 1101-1ah (sana) y b) m1: 2202-9 ah (patológica).

La observación cuidadosa de los espectros de

las señales de voz estudiadas evidencia algunos aspectos que se repiten para todas las señales analizadas, estos aspectos ayudan a establecer las condiciones de análisis para el desarrollo del trabajo, estos aspectos se pueden enumerar como:

1. La presencia de ruido en todo el espectro ya que de lo contrario sólo se apreciarían los armónicos de la fuente de excitación (glotis) modificados por la respuesta del tracto vocal, esta característica se manifiesta en todos los espectros examinados.

2. En promedio, desde los 2.5 KHz en adelante el ruido es de tal magnitud que oscurece completamente a los armónicos de las señales estudiadas. Cualquier análisis que se pretenda realizar en este rango se vería seriamente comprometido ya que es imposible discriminar armónicos de ruido en esta banda. Por la razón antes expuesta se considerará para el estudio como rango de frecuencias sensibles a examen el comprendido entre 0 y 2.5 KHz.

3. Dado que en el dominio de la frecuencia la señal de ruido es un elemento esencial en la definición de los parámetros de calidad de voz tal y como se comentó en el capítulo anterior, no se realizarán operaciones invasivas en la señal, esto es, el análisis se realizará conservando el espectro de la señal tan original como sea posible, de manera que cualquier parámetro de calidad esté relacionado con el contenido de la señal de ruido de la misma.

4. Los análisis se llevaron a cabo identificando tres bandas de frecuencias a saber: Baja (L): 50 a 500Hz, Media (M): 500 a 1500 Hz, Alta (H) : 1500 a 2500 Hz. Con estas condiciones de análisis presentes se examinaron las señales de voz para establecer las hipótesis de los parámetros de calidad respectivos.

Como ya se ha comentado, el examen de las señales de voz estudiadas indica muy poca definición de los armónicos a partir de los 2.5 KHz, sin embargo si examinamos la señal en la banda comprendida entre los 0 y 3KHz, podemos observar que las señales de voz sanas tienen una mejor definición de los armónicos que las señales de voz patológicas, esta es una propiedad que se verifica, en mayor o menor grado, para todas las señales analizadas. A manera de ilustrar este aspecto inspeccionemos de nuevo visualmente las señales 1101ah (voz sana) y 2202-ahBAD (Patológica), ver Figuras 3 y 4.

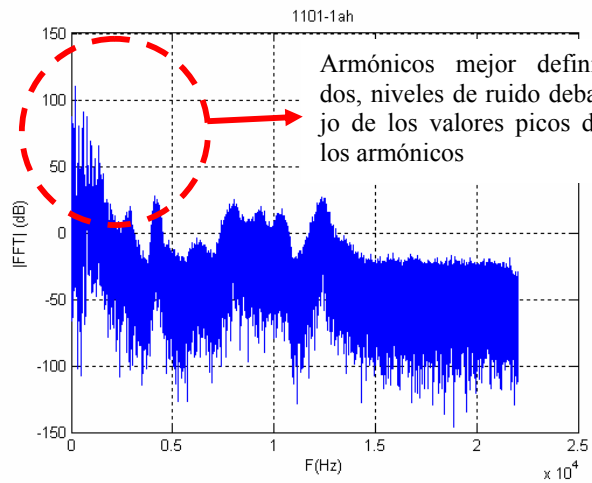


Figura 3. Señal de voz b1: 1101-1ah.

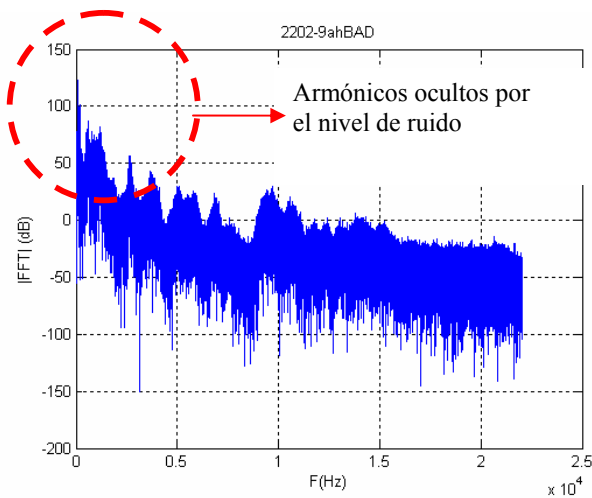
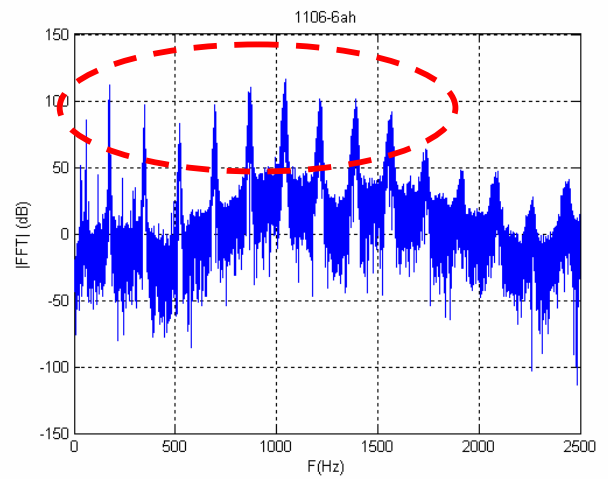


Figura 4. Señal de voz m1: 2202-9ahBAD.

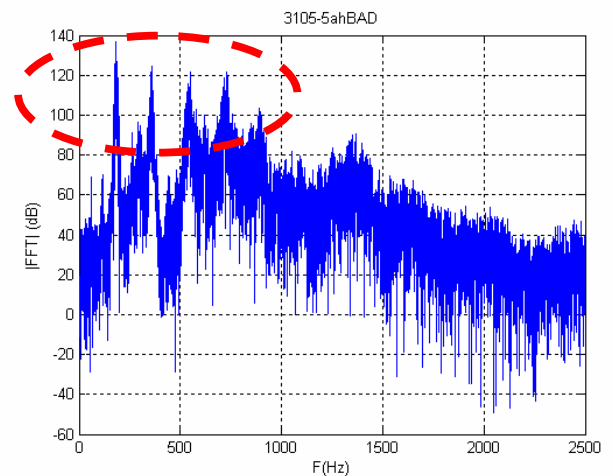


Figura 5. Señales de voz exploradas en el rango de 0 a 2.5 KHz correspondientes a 1106-ah y 3105-5ahBAD.

Para precisar un poco más respecto a la relación entre los armónicos de la señal y el ruido, y teniendo presente que la banda de frecuencias dentro de las cuales todavía pueden observarse armónicos de la señal es de 0 a 2.5 KHz, representaremos de nuevo dos señales: una sana por ejemplo 1106-6ah, y otra patológica, 3105-5ahBAD, en el rango de frecuencias señalado (ver Figura 5).

Otra observación importante es la relativa al contenido de ruido alrededor de los armónicos. En la Figura 6 puede observarse la señal de voz sana 1101-1ah y la señal de voz patológica 3105-5ahBAD, alrededor de los tres primeros armónicos.

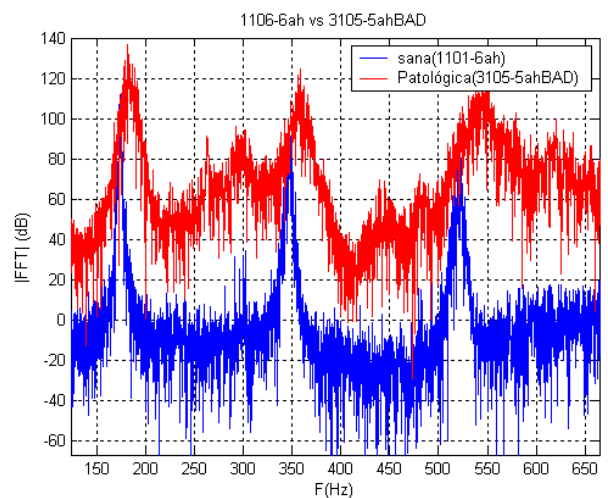


Figura 6. Primeros tres armónicos de la señal de voz sana (1101-1ah) y de la señal de voz patológica (3105-5ahBAD).

Es importante señalar que las diferencias observadas para las parejas de voces sanas y patológicas de las Figuras 3 a 6, se repiten en mayor o menor grado, desde el punto de vista de la inspección visual para todas las parejas analizadas en este trabajo. Sólo se han utilizado algunas para ilustrar dichas propiedades y así justificar los parámetros que en hipótesis podrían resultar indicadores de la presencia de una voz sana o patológica.

Puede asegurarse que la influencia del ruido es fundamental en la calidad espectral de las señales de voz, es decir, este parámetro puede definir de manera importante cualquier medición que sobre la calidad de la voz se pretenda realizar.

Es importante señalar que las fuentes de este ruido son muy diversas, esto es: variaciones de amplitud, variaciones de frecuencia, turbulencias en el tracto vocal, etc. De ahí que resulte muy difícil su evaluación tomando en cuenta que es ruido de banda ancha, es decir, tiene componentes en todas las frecuencias del espectro de voz.

#### 4. FORMULACIÓN DE HIPÓTESIS PARA LOS PARÁMETROS ESPECTRALES

La bibliografía consultada [15-19], reporta parámetros de calidad de voz que se determinan realizando operaciones combinadas en el dominio del tiempo y de la frecuencia, lo que significa descomposición de la señal de voz en numerosas operaciones de estimación de los mismos, en este trabajo se obtuvieron parámetros en el dominio de la frecuencia sin mayores transformaciones que las que pueden derivar de un filtro de suavizado, es decir la hipótesis principal es que los parámetros que determinan la calidad de la voz pueden ser extraídos del espectro de frecuencias de la misma.

Sobre la base de las características advertidas de la inspección visual y auditiva se formularon las siguientes hipótesis para los parámetros que potencialmente definen la calidad de la voz.

1. **Parámetro PMR:** Debido a la presencia de ruido en todo el espectro de frecuencias de la señal de voz, se observa un incremento en el nivel de amplitud del ruido de las voces patológicas en relación con las voces sanas. Por esta razón se define como PMR a la relación existente entre el valor pico del

espectro de la señal de voz y su valor medio, se formula la hipótesis de que esta relación es indicadora de la calidad de la voz y se espera que este indicador sea mayor para las señales de voz sanas.

2. **Parámetro SNR:** Se define como la relación de la Energía de Señal (medida en las componentes en frecuencias que se extienden desde 50 Hz hasta 2500 Hz) a la energía de Ruido (medida desde 1000 Hz hasta 2500 Hz) de las señales de voz. Se formula como hipótesis que esta relación es diferente para señales de voz sanas que para las señales de voz patológicas, se fundamenta en el hecho observado de que los armónicos de la señal son oscurecidos de manera más pronunciada en las señales de voz patológicas que en las señales de voz sanas.

Dado que la forma en que la distribución de energía de señal y la energía de ruido varía entre las diferentes bandas de frecuencias, como se observó anteriormente, definimos los siguientes parámetros:

3. **SNRL:** Relación entre la energía de la Señal (medida entre los 50 Hz y 500 Hz) a la energía de ruido medida en el mismo rango de frecuencias.
4. **SNRM:** Relación entre la energía de la Señal (medida entre los 500 Hz y 1500 Hz) a la energía de ruido medida en el mismo rango de frecuencias.
5. **SNRH:** Relación entre la energía de la Señal (medida entre los 1500 Hz y 2500 Hz) a la energía de ruido medida en el mismo rango de frecuencias.

Se formuló como hipótesis que estos parámetros son diferentes para las señales de voz sanas que para las patológicas.

#### 5. RESULTADOS Y ANÁLISIS

##### 5.1. Relación valor pico a valor medio, PMR.

Para este caso se calculó el PMR para cada señal de voz sana y patológica en forma original y filtrada, el filtro utilizado fue un filtro de mediana de seis puntos ya que la idea es suavizar las señales de voz y averiguar cuál es la influencia del mismo en el PMR. Los resultados obtenidos para cada una de las señales de voz se resumen en la Tabla 3. Se puede observar claramente el efecto del suavizado de las señales de voz al arrojar valores de PMR de señales filtradas

mayores que los correspondientes a las señales originales debido al mayor contenido de ruido. En la Tabla 3, se puede apreciar que la media del PMR de las voces sanas es superior a la media del PMR de las voces patológicas.

Tabla 3. Valores medios de PMR para las señales de voz sanas y patológicas.

| Parámetro | VALOR MEDIO |          |          |          |
|-----------|-------------|----------|----------|----------|
|           | ORIGINAL    |          | FILTRADA |          |
|           | Buenas      | Malas    | buenas   | malas    |
| PMR       | 461,7243    | 219.7036 | 546.8758 | 250.8199 |

Estadísticamente el hecho de que se tengan dos medias muestrales diferentes no indica que las muestras sean significativamente diferentes, a fin de investigar si la media PMR de las señales de voz sanas y la media de PMR de las señales de voz patológicas son significativamente diferentes, se aplicó el análisis de varianza ANOVA para confirmar si efectivamente el parámetro PMR representa un indicador apropiado para concluir si se está en presencia de una voz sana o patológica.

Resumiendo el procedimiento podemos decir que básicamente se establece como hipótesis de nulo el que las medias de los dos grupos de voces son iguales, esto es hipótesis de nulo, y se determinan los valores de **F** y de **p** que permiten decidir si se rechaza o no la hipótesis de nulo, todo esto para un determinado valor de significancia que normalmente se establece en 5%. La Tabla 4 y la Figura 7 muestran el resultado del análisis de varianza al parámetro PMR.

De la Tabla 4, puede observarse que el valor de **p** es de 0.00052581 y  $F > F_{\text{crítico}}$  para  $\alpha = 0.05$ , por lo que podemos rechazar la hipótesis de nulo, esto permite concluir que los PMR de cada grupo son significativamente diferentes.

Se aprecia, en la Figura 7, los parámetros característicos de la estadística de los grupos observados (voces sanas y voces patológicas), esto es: datos ubicados en los respectivos percentiles para cada grupo, medianas, valores máximos y mínimos, etc. Aquí de nuevo se muestra claramente que los grupos son diferentes en lo que respecta al PMR

Para evaluar el comportamiento del PMR en el

Tabla 4. Análisis de Varianza de los PMR de las señales de voz sana y voz patológica.

Resumen Análisis de varianza de un factor

| Grupos    | Cuenta | Suma      | Promedio   | Varianza   |
|-----------|--------|-----------|------------|------------|
| Columna 1 | 12     | 5540,6919 | 461,724325 | 27836,3529 |
| Columna 2 | 12     | 2636,4426 | 219,70355  | 14883,8336 |

Análisis de varianza

| Origen de las variaciones | Suma de cuadrados | Grados de libertad | Promedio de los Cuadrados | F       | P      | Valor crítico para F |
|---------------------------|-------------------|--------------------|---------------------------|---------|--------|----------------------|
| Entre Grupos              | 351444,3          | 1                  | 351444,3                  | 16,4533 | 0,0005 | 4,3009               |
| Dentro de los Grupos      | 469922,0          | 22                 | 21360,0                   |         |        |                      |
| Total                     | 821366,3          | 23                 |                           |         |        |                      |

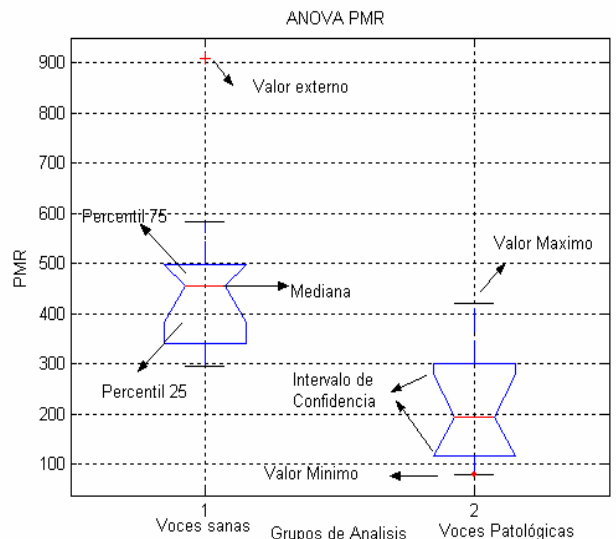


Figura 7. Gráfico tipo BOXPLOT del análisis de varianza de PMR.

rango en frecuencias de cada señal de voz, se calcularon los PMR de cada muestra en segmentos o ventanas de 500Hz con un solapamiento de 400 Hz, que se obtiene desplazando la ventana de cálculo de PMR 100Hz y se graficó el PMR de cada muestra de voz contra todas las voces patológicas a efectos de evaluar y comparar los comportamientos en toda la banda. Los resultados obtenidos de esta evaluación se pueden observar en las Figuras 8 y 9 para los casos de las voces sanas 1104-5ah vs. todas las voces patológicas



para 2107-ah vs. todas las voces patológicas, estas muestras se seleccionaron como ejemplo para ilustrar el comportamiento en frecuencia que tiene el parámetro PMR. Se debe aclarar que este mismo comportamiento se evidencia para todos las voces sanas y patológicas.

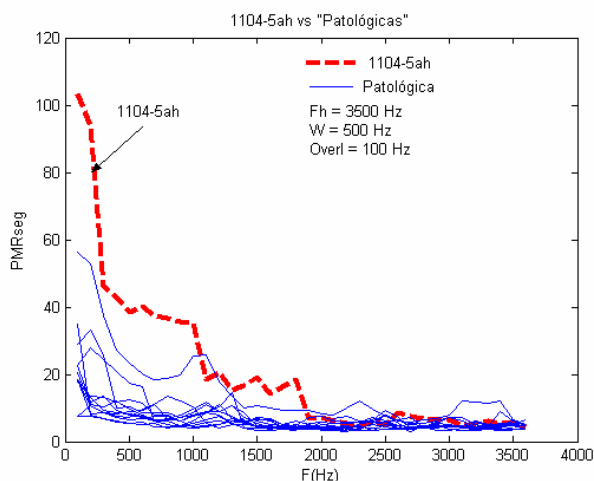


Figura 8. PMR 2106-5ah vs .Patológicas.

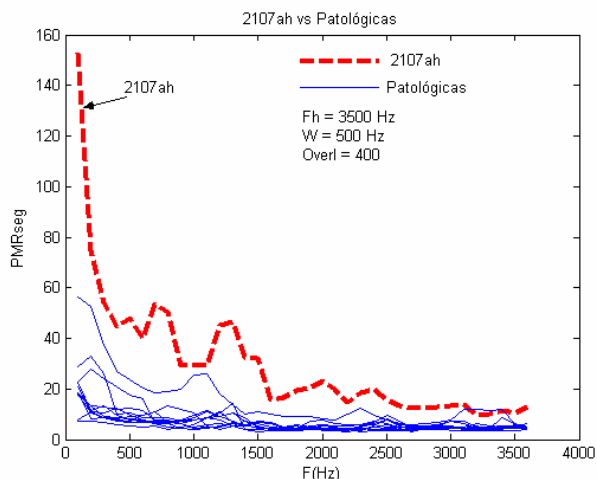


Figura 9. PMR 2107-5ah vs. Patológicas.

Todas las gráficas del parámetro PMR demuestran, que en general este parámetro es mayor en las muestras de voces sanas que en las muestras de voz patológicas, en toda la banda de frecuencias hasta los 2500 Hz, a partir de este extremo comienzan a tener valores similares, por tanto no se considera necesario realizar una exploración segmentada en banda del mismo.

Puede observarse en las gráficas la existencia

de una señal, clasificada como patológica, cuyos valores de PMR, en muchos casos, tiene un comportamiento similar al de las señales de voz clasificadas como sanas, la señal en cuestión es la identificada como m2: 2404-4ahBAD, una prueba de audición arroja que la misma tiene un sonido aceptable.

## 5.2. SNR, SNRL, SNRM, SNRH

Para el caso del cálculo de SNR, SNRL, SNRM, SNRH, evidentemente se requiere discriminar entre señal y ruido, esta tarea es altamente complicada toda vez que el ruido es una señal aleatoria y de banda ancha razón por la cual se decidió aplicar un filtro de mediana a fin de suavizar lo mejor posible las señales de voz examinadas.

Después de varios ensayos se determinó que la mejor operación del filtro se obtenía con  $N = 6$ . La señal filtrada es restada de la señal original para tener una estimación del ruido y las energías son calculadas mediante el teorema de Parseval.

Los resultados del análisis de varianza de cada uno de estos parámetros obtenidos se resumen en la Tabla 5.

Tabla 5. Análisis de Varianza de SNR, SNRL, SNRM y SNRH.

| Parámetro | p       | F          | F <sub>crítico</sub> | Decisión                           |
|-----------|---------|------------|----------------------|------------------------------------|
| SNR       | 0.2383  | 1.46944875 | 4.3009436            | No se rechaza la hipótesis de nulo |
| SNRL      | 0.5155  | 0.43696    | 4.300946             | No se rechaza la hipótesis de nulo |
| SNRM      | 0.02771 | 5.55755    | 4.300946             | Se rechaza la hipótesis de nulo    |
| SNRH      | 0.03832 | 4.85601    | 4.300946             | Se rechaza la hipótesis de nulo    |

De acuerdo a lo señalado anteriormente para el caso de PMR, en relación con el análisis de varianza ANOVA, tenemos que para el caso de SNR Y SNRL no puede ser rechazada la hipótesis de nulo mientras que en los casos de SNRM y SNRH sí, por lo tanto en estos dos últimos parámetros se encuentra que la diferencia de los grupos analizados es significativa.

Si realizamos un examen del parámetro SNR en función de la frecuencia, ver Figuras 10 (a) y (b), se

puede apreciar para los dos casos de las figuras (dos muestras de voces sanas contra las patológicas) que el SNR tiene valores distintivos en la banda media (500 a 1500 Hz) y alta (1500 a 2500 Hz) en relación con la banda baja (50 a 500 Hz) este comportamiento refuerza la hipótesis de evaluar SNR diferenciando la banda en que se determina, esto es: SNRL, SNRM y SNRH, como se estableció en las hipótesis anteriormente señaladas. Se debe agregar que el comportamiento descrito para estas muestras tomadas como ejemplo se repite para todas las demás señales de voz.

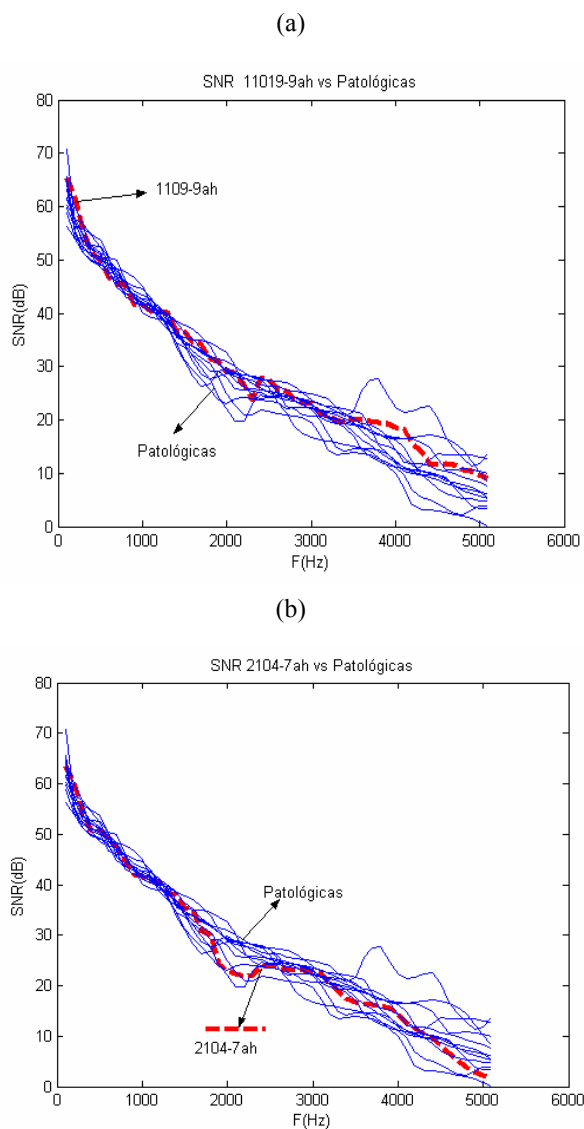


Figura 10. SNR en el rango 0 a 5 KHz para (a) 1109-9ah vs Patológicas y (b) 2104-7ah vs Patológicas.

## 6. CONCLUSIONES

En el presente trabajo se propusieron cinco parámetros para determinar la calidad de la voz y el análisis estadístico arrojó que los parámetros PMR, SNRM, y SNRH son significativos. Es importante destacar que los resultados se han obtenido con independencia del sexo y edad del locutor por lo que su aplicación es general. Estos parámetros no se han encontrado en los reportes de investigaciones realizadas a pesar de existir un gran número de trabajos realizados en esta área, por lo que se considera un aporte novedoso.

La importancia de los resultados obtenidos tiene que ver con el hecho de que fueron extraídos sin realizar mayores operaciones a la señal de voz que no sea la de obtener la DFT y a lo sumo un proceso de suavizado de la misma a través de un filtro de mediana. La mayor parte de los trabajos investigados utilizan procedimientos en el dominio del tiempo combinados con el dominio de la frecuencia, inventanado, filtrado severo de la señal y otras transformaciones.

La tarea de análisis, centrada en la observación de la señal de voz ruidosa y su apreciación auditiva, resulta complicada debido a la naturaleza de banda ancha del ruido que se esparce sobre todo el espectro de la señal de voz, sin embargo, permitió advertir patrones de comportamiento espectral relativos a la señal de ruido que se pudo pasar del dominio subjetivo al dominio cuantitativo a través de los parámetros encontrados.

Particular atención merece el parámetro PMR, cuyo análisis de varianza ANOVA lo muestra como un parámetro altamente diferenciador de los grupos de voces analizadas (voces sanas y voces patológicas) al obtener un valor de  $p$  significativamente pequeño ( $p = 0.00052581$ ) y de  $F > F_{\text{crítico}}$ ; además tener un comportamiento consistente en la banda de frecuencias observada.

En el caso de SNRM y SNRH, los valores de  $p$  obtenidos del análisis de varianza son más conservadores ( $p = 0.02771$  y  $p = 0.03831$ ), con  $F$  cercano a  $F_{\text{crítico}}$ ; pueden considerarse también diferenciadores de los grupos de voces analizadas en una primera aproximación, requiriendo tal vez confirmación a través de PMR.

Se recomienda incrementar el número de muestras de manera significativa para dar mayor solidez a los resultados desde del punto de vista estadístico, es importante conocer a priori las calificaciones que reciben estas señales desde el punto de vista de su calidad.

Se recomienda aumentar también el número de voces sanas y patológicas para que a través de análisis estadísticos producir umbrales para estos parámetros que permitan decidir cuando se está en presencia de una voz sana o patológica.

La calificación de las señales de voz utilizadas como sanas y patológicas era conocida de antemano por lo que se desconoce la patología asociada a cada una, se recomienda realizar los análisis pertinentes para relacionar los valores de los parámetros aportados en este trabajo con las correspondientes patologías.

De igual manera se recomienda el análisis estadístico correspondiente a fin de relacionar estos parámetros con otros obtenidos a través de los procedimientos que señala la literatura investigada, esto con el fin de identificar si existen patologías comunes que puedan ser diagnosticadas por ambos enfoques.

## REFERENCIAS

- [1] D. G. Childers. "Speech Processing and Syntesis Toolboxes". Ed. John Wiley & Sons, Inc. 2000.
- [2] Deller John, Proakis John y Hansen John. "Discrete-Time Processing of Speech Signal". Ed. Prentice Hall. 1987.
- [3] Haykin Simon, Van Been Barry. "Señales y Sistemas". 1<sup>ra</sup> Edición. Ed. Limusa Wiley. 2001.
- [4] Leon Cohen. "Time-Frequency Analysis". Ed. Prentice-Hall PTR. 1995.
- [5] Lyman Ott, R. "An Introduction to Statical Methods end data Analysis". 4<sup>ta</sup> Edición. Ed. Duxbury Press 1992.
- [6] Mintra Sanjit. "Digital Signal Processing". Ed. Mc Graw Hill, 2<sup>da</sup> Edición, 1998.
- [7] Millar, R. Irwin, Freund John E, Jonson R. "Probabilidad y Estadística para Ingenieros". 4<sup>ta</sup> Edición, Ed. Prentice Hall, 1992.
- [8] Oppenheim Alan, Willsky A, Hammid S. "Señales y Sistemas". 2<sup>da</sup> Edición, Prentice Hall. 1998.
- [9] Pelton, Gordon E. "Voice Processing". Ed. Mc Graw Hill. 1992
- [10] Proakis Jhon y Manolakis Dimitris. "Tratamiento Digital de Señales". Ed. Prentice Hall, 3<sup>ra</sup> Edición, 1998.
- [11] Madisetti, Vijay K.; Williams, Dougl's B. (editores). "The Digital Signal Processing Handbook". CRC-Press - IEEE Press. Boca Raton, FL, USA, 1997. Section X.
- [12] Smith Steven. "The Scientist and Engineer's Guide to Digital Signal Processing". California Technical Publishing, San Diego, California. 2<sup>da</sup> Edición. 1999.
- [13] Sondhi M. M. "Model for wave propagation in a lossy vocal tract". Journal of Acoustical Society of America. Vol. 56, N° 5, Mayo 1974, pp. 1070-1075.
- [14] Madisetti, Vijay K.; Williams, Dougl's B. (editores). "The Digital Signal Processing Handbook". CRC-Press - IEEE Press. Boca Raton, FL, USA, 1997. Section X.
- [15] Murphy P.J. "A cepstrum-Based Harmincs to Noise Ratio in Speech Signals". Department of Electronics and Computer Engineering. Unicersity of Limerick. Ireland.
- [16] Yegnanarana B, d' Alessandro Christophe, Darsinos Varsilis. "An Iterative Algoritm for Decomposition of Speech Signals into Periodic and Aperiodic Components". IEEE Transactions on Speech and Audio Processing, vol 6, n° 1, January 1998.
- [17] Boersma Paul. "Accurate Short - Term Analysis of the Fundamental Frequency and the Harmonics to Noise Ratio of a Sample Sound". UInstitute of Phonetic Sciencies. University of Amsterdam, Proceedings 17 (1993), 97-110.
- [18] Cheol-Woo Jo, Dae-Hyun Kim. "Classification of Pathological voice into Normal/Benign/Malignant State". Department of Control & Instrumentation Engineering, Changwon National University. Korea.
- [19] Deliyski, Dimitar D. "Acoustic Model and Evaluation of Pathological Voice Production". Kay Elemetrics Corp., Dept of Research and Development. USA.