

**PROTOCOLO DE TÉCNICAS DE ADQUISICIÓN Y PROCESAMIENTO DE  
SEÑALES PARA USO FORENSE EN INSTITUCIONES DE SEGURIDAD DEL  
ESTADO**

**JORGE CASTRO CALDERON  
DANNY ALEXANDER RUBIO**

**PROYECTO DE GRADO  
INGENIERÍA DE SONIDO X SEMESTRE**

**UNIVERSIDAD DE SAN BUENAVENTURA  
INGENIERÍA DE SONIDO  
BOGOTÁ, D. C.**

**2006**

**PROTOCOLO DE TÉCNICAS DE ADQUISICIÓN Y PROCESAMIENTO DE  
SEÑALES PARA USO FORENSE EN INSTITUCIONES DE SEGURIDAD DEL  
ESTADO**

**JORGE CASTRO CALDERON  
DANNY ALEXANDER RUBIO**

**TUTORES**

**Fis. LUIS JORGE HERRERA  
Ing. ESPERANZA CAMARGO**

**UNIVERSIDAD DE SAN BUENAVENTURA  
INGENIERÍA DE SONIDO  
BOGOTÁ, D. C.**

**2006**

## TABLA DE CONTENIDO

INTRODUCCIÓN.....	10
1. FORMULACIÓN DEL PROBLEMA.....	11
1.1 Antecedentes.....	11
1.2 Descripción del problema. ....	12
1.3 Justificación. ....	12
1.4 Objetivos de la investigación. ....	13
1.4.1 Objetivo general. ....	13
1.4.2 Objetivo específico. ....	13
1.5 Alcances y limitaciones del proyecto. ....	13
1.5.1 Alcances.....	13
1.5.2 Limitaciones.....	14
2. MARCO REFERENCIAL.....	15
2.1 Marco conceptual.....	15
2.1.1 Fundamentos del sonido.....	15
2.1.2 Características del sonido.....	16
2.1.3 Clasificación de las ondas.....	19
2.1.4 El decibelio.....	21
2.1.5 Espectro Audible.....	23
2.1.6 Ruido.....	24
2.1.7 Fuente de ruido y distorsiones en grabaciones.....	27
2.1.8 Análisis espectral de frecuencia.....	30
2.1.9 Formatos de grabación de audio.....	33
2.1.10 Digitalización.....	34
2.1.11 Formatos de archivo.....	40
2.2 Marco legal o normativo.....	42
2.3 Marco teórico.....	43
2.3.1 La voz humana.....	43
2.3.2 Clasificación de los sonidos de la voz.....	48

2.3.3	Material fonograbado dubitado e indubitado.....	56
2.3.4	Equipos utilizados en la captura de de la señal dubitada e indubitada.....	60
2.3.5	Biometría.....	62
2.3.6	Biometría de la voz.....	67
2.3.7	Acústica Forense.....	68
2.3.7.1	Reconocimiento de locutores.....	69
3.	METODOLOGÍA.....	74
3.1	Enfoque de la investigación.....	74
3.2	Campo temático del programa.....	74
3.3	Técnicas de recolección de información.....	74
3.4	Población y muestra.....	76
3.5	Hipótesis.....	76
3.6	Variables.....	77
3.6.1	Variables dependientes.....	77
3.6.2	Variables independientes.....	77
4.	PRESENTACIÓN Y ANÁLISIS DE RESULTADOS.....	78
4.1	Resultado de ingresos de material dubitado e indubitado.....	78
4.2	Resultados de la encuesta.....	87
4.3	Resultado de una mala digitalización de señales análogas.....	88
4.3.1	Errores de muestreo.....	88
4.3.2	Errores de resolución.....	91
4.3.3	Influencia de ruido en la interpretación de resultados.....	95
5.	DESARROLLO INGENIERIL.....	99
5.1	Protocolo.....	99
5.1.1	Métodos de grabación para la toma de muestra de voz.....	99
5.1.1.1	Material dubitado.....	99
5.1.1.1.2	Equipamiento material dubitado.....	100

5.1.1.2	Material indubitado.....	109
5.1.1.2.1	Condiciones acústicas para la sala.....	109
5.1.1.2.2	Equipamiento para la sala.....	110
5.1.2	Técnicas de grabación para material indubitado.....	130
5.1.3	Procesamiento del material indubitado.....	133
5.1.4	Método de análisis acústico del habla.....	137
5.1.4.1	Tipos de análisis.....	139
5.4.1.2	Espectrograma.....	140
5.4.1.3	LPC Frequency Response.....	141
5.4.1.4	FFT Power spectrum.....	145
5.4.1.5	Cepstrum.....	149
5.4.1.6	Formant History.....	149
5.4.1.7	Pitch Contour.....	151
5.4.1.8	Energy contour.....	153
6.	CONCLUSIONES.....	155
7.	RECOMENDACIONES.....	156
7.1	Equipo interdisciplinario.....	156
7.2	Tipos de análisis que intervienen en la Identificación del locutor.....	157
7.3	Banco de voz.....	159
	ANEXOS A.....	159
A.1	Tabla de ingreso de material dubitado e indubitado.....	159
A.2	Tabla de canales de comunicación.....	161
A.3	Tabla de tipo de formato digital dubitado.....	161
	GLOSARIO.....	163
	BIBLIOGRAFÍA.....	167

## ÍNDICE DE FIGURAS

<b>Figura 1.</b> Señal sinusoidal.....	16
<b>Figura 2.</b> Nivel de presión sonora.....	22
<b>Figura 3.</b> Ruido blanco.....	25
<b>Figura 4.</b> Ruido rosa.....	25
<b>Figura 5.</b> Ruido marrón.....	26
<b>Figura 6.</b> Banda de octava.....	30
<b>Figura 7.</b> Ancho de banda.....	33
<b>Figura 8.</b> Sistema fonatorio.....	45
<b>Figura 9.</b> Cuerdas vocales.....	45
<b>Figura 10.</b> Corte esquemático del aparato fonatorio humano.....	46
<b>Figura 11.</b> Modelo simplificado de producción de sonidos del habla a partir de una fuente y un filtro.....	47
<b>Figura 12.</b> Características acústicas de los sonidos del habla.....	48
<b>Figura 13.</b> Vocales abiertas, medias y cerradas.....	49
<b>Figura 14.</b> Vocales anteriores, centrales y posteriores.....	50
<b>Figura 15.</b> Vocales labializadas y no labializadas.....	50
<b>Figura 16.</b> Sonido directo y reflejado.....	59
<b>Figura 17.</b> Grafica de una Señal de voz en un espectrograma.....	67
<b>Figura 18.</b> Comportamiento de material ingresado en los 4 meses.....	81
<b>Figura 19.</b> Grabaciones de material dubitado (análogo Vs. digital).....	81
<b>Figura 20.</b> Grabaciones de material indubitado (análogo Vs. digital).....	82
<b>Figura 21.</b> Unidades de material dubitado según el canal de comunicación.....	83
<b>Figura 22.</b> Formatos de audio del material dubitado.....	84
<b>Figura 23.</b> Escala de calidad de material dubitado.....	86
<b>Figura 24.</b> Forma de onda en el oscilograma “que paso” grabado correctamente.....	88
<b>Figura 25.</b> Espectrograma de la frase “que paso” con un frecuencia de muestreo correcta.....	89

<b>Figura 26.</b> Forma de onda en el oscilograma de la frase “que paso” grabada incorrectamente.....	89
<b>Figura 27.</b> Espectrograma de la frase “que paso” con un frecuencia de muestreo incorrecta.....	90
<b>Figura 28.</b> Comparación de espectrogramas de la frase “que paso” adecuado e inadecuado respectivamente.....	91
<b>Figura 29.</b> Forma de onda en el oscilograma de la palabra “Aló” con una resolución correcta.....	92
<b>Figura 30.</b> Espectrograma de la palabra “Aló” con una resolución correcta.....	92
<b>Figura 31.</b> Forma de onda en el oscilograma de la palabra “Aló” con una incorrecta resolución.....	93
<b>Figura 32.</b> Espectrograma de la palabra “Aló” con una incorrecta resolución.....	93
<b>Figura 33.</b> Comparación de espectrogramas de la palabra “Aló”.....	94
<b>Figura 34.</b> Comparación de oscilograma (Digital vs. Análogo).....	95
<b>Figura 35.</b> Comparación de espectrogramas (Digital vs. Análogo).....	96
<b>Figura 36.</b> Análisis del Rango del tono de la señal digital vs. señal Análoga.....	96
<b>Figura 37.</b> Comparación de Energía (Digital vs. Análogo).....	97
<b>Figura 38.</b> Respuesta en frecuencia de SM-48.....	122
<b>Figura 39.</b> Directividad de SM-48.....	123
<b>Figura 40.</b> Configuración de montura AKG C 420.....	124
<b>Figura 41.</b> Respuesta en frecuencia AKG C 420.....	125
<b>Figura 42.</b> Diagrama polar AKG C 420.....	125
<b>Figura 43.</b> Respuesta en frecuencia de los HD 250 II.....	128
<b>Figura 44.</b> Forma de onda en oscilograma de la frase de una banda de rock...	140
<b>Figura 45.</b> Espectrograma de la frase “de una banda de rock”.....	141
<b>Figura 46.</b> Análisis LPC frequency Response de la frase “de una banda de rock”.....	142
<b>Figura 47.</b> Análisis LPC WaterFall de la palabra “deeee”.....	143
<b>Figura 48.</b> Análisis FFT Power Spectrum de la sílaba “deeee”.....	146
<b>Figura 49.</b> Análisis FFT WaterFall de la palabra “deeee”.....	147

<b>Figura 50.</b> Análisis FFT Power Spectrum Vs. análisis LTA Power Spectrum de la silaba “deeee” .....	148
<b>Figura 51.</b> Análisis histórico de formantes de la frase “deeee una banda de rock” .....	149
<b>Figura 52.</b> Análisis de los tonos de contorno de la frase “deeee una banda de rock” .....	151
<b>Figura 53.</b> Análisis energético de la frase “deeee una banda de rock” .....	153



## ÍNDICE DE TABLAS

<b>Tabla 1.</b> Frecuencia de muestreo.....	35
<b>Tabla 2.</b> Clasificación acústica de los sonidos del habla en función del la fuente y el filtro.....	49
<b>Tabla 3.</b> Clasificación de las consonantes de la lengua castellana según el lugar, el modo de articulación y la sonoridad.....	53
<b>Tabla 4.</b> Tipo de grabadora de bobina cerrada.....	61
<b>Tabla 5.</b> Ingreso total de material.....	79
<b>Tabla 6.</b> Ingreso material dubitado.....	80
<b>Tabla 7.</b> Ingreso de material análogo y digital.....	80
<b>Tabla 8.</b> Canales de comunicación.....	82
<b>Tabla 9.</b> Unidades de material dubitado según el tipo de formato.....	83
<b>Tabla 10.</b> Estado de la grabación del material dubitado (escala 1-5).....	85
<b>Tabla 11.</b> Resultado de la encuesta.....	87
<b>Tabla 12.</b> Comparación de tonos entres la señal digital y la señal análoga.....	97
<b>Tabla 13.</b> Comparación energética de la señal digital Vs. La señal análoga.....	98
<b>Tabla 14.</b> Manual de usuario Gnome 2M.....	106
<b>Tabla 15.</b> STC-H246 Tarjeta Profesional I / O.....	115
<b>Tabla 16.</b> Especificaciones técnicas de sonido STC-H246.....	117
<b>Tabla 17.</b> Información numérica de formantes del análisis histórico de la frase “deeee una banda de rock”.....	150

## INTRODUCCIÓN

Los avances en el reconocimiento de voz en los últimos años han aumentado gracias al desarrollo de sistemas automáticos y semiautomáticos empleados en el análisis forense del locutor y comparación de voces. Estos sistemas procesan y analizan señales para determinar patrones de la voz que caracterizan un individuo. Estos procesos efectuados sobre la señal, tienen una inherencia directa con los resultados de los análisis cuando son aplicados de manera inadecuada.

En Colombia, las entidades del estado encargadas del estudio forense de la voz se encuentran en un proceso de desarrollo. Sin embargo se presentan falencias en el manejo de señales desde el punto de vista ingenieril, que de un soporte científico a los resultados obtenidos.

La investigación realizada gracias a la colaboración del personal profesional que trabaja en estas áreas, plasma un estudio cuantitativo y cualitativo del material fonograbado que ingresa a su dependencia para ser analizados, determinando formatos de almacenamiento, canales de grabación, calidad de grabación, entre otros.

Debido a los resultados obtenidos por la investigación, este proyecto plantea un protocolo el cual contiene técnicas y herramientas adecuadas para la captura (técnicas de grabación), procesamiento (digitalización, filtrado) y análisis de señales de voz., mejorando la productividad y calidad en el análisis de voz.

## 1. FORMULACIÓN DEL PROBLEMA

### 1.1 ANTECEDENTES

Según las entidades de seguridad del estado, en Colombia no se ha implementado un protocolo para la toma de muestras de voz en el área forense. Sin embargo, a nivel internacional existen recomendaciones sobre técnicas de grabación y procesamiento de señales como la Universidad de Michigan en su publicación *Best Practices in the Acquisition, Processing, and Analysis of Acoustic Speech Signals*.

Encontramos también, recomendaciones sobre microfonía, calibración y cuartos acústicos por Svante Granqvist (Department of Speech, Music and Hearing TMH, Royal Institute of Technology KTH, Stockholm, Sweden) y Jan Švec Groningen (Voice Research Lab, Department of Biomedical Engineering, University of Groningen, The Netherlands) en su publicación *Microphones and Room Acoustics and Their Influence on Voice Signals*.

Para un estudio desde el punto de vista fonético-forense en lengua castellana el Dr. Carlos Delgado realizó una memoria para la obtención de su doctorado llamada *la identificación de locutores en el ámbito forense*.

En el tratamiento de señales y limpieza de ruido que impidan un análisis eficiente el departamento científico de Speech Technology Center (STC) brinda pautas generales para su aplicación en señales de baja calidad con presencia de ruido. Además, el STC cuenta con expertos que realizan publicaciones de investigaciones sobre el análisis de señales para uso forense dentro de los cuales se encuentra el Dr. Sergey Koval quien recientemente visitó el país compartiendo sus conocimientos y experiencia en esta rama de la Criminalística.

Existen diferentes organizaciones y asociaciones como la NCVS (National Center for Voice and Speech), LAIP (Laboratoire d'analyse informatique de la parole), IAFP (Internacional Association for Forensic Phonethics), entre otras que estudian la voz.

## 1.2 DESCRIPCIÓN DEL PROBLEMA

La principal problemática, según estudios realizados por las entidades de seguridad del estado, un 40% de las muestras de voz obtenidas por medio análogo o digital, no reúne los requisitos técnicos para realizar los estudios acústicos, fonéticos ni lingüísticos de comparación; aunque se cuenta con gran cantidad de material grabado de voces, en formato digital y análogo, su nivel de calidad es muy bajo razón por la que los resultados obtenidos en análisis y comparación no son los esperados.

¿Cómo mejorar los procedimientos utilizados en la captura, procesamiento y análisis de voz para la identificación de locutor en uso forense?

## 1.3 JUSTIFICACIÓN

Según la información del Ministerio de Defensa Nacional *Fondelibertad*, han ocurrido 227 de casos de secuestro extorsivo en el presente año de los cuales, gran parte de las exigencias de estos individuos al margen de la ley son realizados por medios de telefonía fija o móvil. Para el primer semestre del año 2006, el secuestro y la extorsión han disminuido con respecto al año anterior un 49% y 54% respectivamente, según información suministrada por CIC – DIJIN, Policía Nacional y la fundación Seguridad y Democracia.<sup>1</sup>

Sin embargo los índices de delitos continúan siendo elevados, afectando la seguridad del estado y perturbando la convivencia normativa en sociedad. Por lo tanto los organismos competentes en el área de seguridad nacional están en la necesidad de implementar técnicas y herramientas adecuadas para reconocimiento de locutor.

Con el fin de mejorar la eficiencia para interpretar el material fonograbado, se pretende optimizar los procedimientos en la captura, conversión análogo-digital, procesamiento, almacenamiento y finalmente el análisis de las muestras que se desean comparar.

<sup>1</sup> <http://www.seguridadydemocracia.org/docs/pdf/seguridadUrbana/primerSemestre2006.pdf>

Dentro de estos procedimientos cabe destacar la importancia de una base de datos automatizada para almacenamiento de muestras, que actualmente no existe, que tendrá como función la recolección y almacenamiento de registros de voz, con el propósito de identificar personas que involucren delitos como secuestro, extorsión, narcotráfico, terrorismo, subversión, etc. el cual servirá como soporte fundamental a las investigaciones que adelantan los grupos de acción unificada, las diferentes unidades operativas y de inteligencia del estado.

## 1.4 OBJETIVOS DE LA INVESTIGACIÓN

### 1.4.1 Objetivo General

Implementar técnicas basadas en herramientas computacionales existentes adecuadas para el reconocimiento de locutor aplicadas en acústica forense.

### 1.4.2 Objetivos Específicos

- Implementar técnicas adecuadas para el registro fonográfico.
- Procesar la señal para realizar filtrado digital, edición, control de nivel de ganancia y adecuación play back.
- Analizar muestras de voz mediante software especializado.
- Realizar un protocolo de captura y procesamiento de señal para análisis de locutor.

## 1.5 ALCANCES Y LIMITACIONES DEL PROYECTO

### 1.5.1 Alcances

El alcance principal del proyecto es la implementación en entidades de seguridad del estado a nivel nacional. Este trabajo será piloto en el desarrollo de la acústica forense en el país.

### 1.5.2 Limitaciones

Debido a que durante la realización de este proyecto se emplean software especializados con los cuales la universidad no cuenta, tenemos que recurrir a una entidad de seguridad del estado para su préstamo que es limitado, por lo cual el tiempo será una variable independiente, el cual puede alargar la ejecución del proyecto.

## 2. MARCO REFERENCIAL

### 2.1 MARCO CONCEPTUAL

Debido a que los profesionales que se desempeñan actualmente en el área de la acústica forense carecen de formación profesional en ingeniería de sonido, es indispensable brindarles la información básica necesaria para su análisis y estudio de este protocolo, ya que su perfil profesional esta enfocado en el análisis lingüístico y fonético.

#### 2.1.1 FUNDAMENTOS DE SONIDO

##### **SONIDO<sup>2</sup>**

El *sonido* se puede definir de varias formas. De todas ellas, la más intuitiva es la siguiente:

- Vibración mecánica que se propaga a través de un medio elástico y denso (habitualmente el aire) y, que es capaz de producir una sensación auditiva.
- Sensación auditiva producida por una vibración de carácter mecánico que se propaga a través de un medio elástico y denso.

##### GENERACIÓN Y PROPAGACIÓN DEL SONIDO

El elemento generador del sonido se denomina *fente sonora* y esta puede ser un tambor, una cuerda de un violín o simplemente las cuerdas vocales de una persona. La generación del sonido tiene lugar cuando dicha fuente entra en vibración y ésta se transmite por las partículas que conforman el aire.

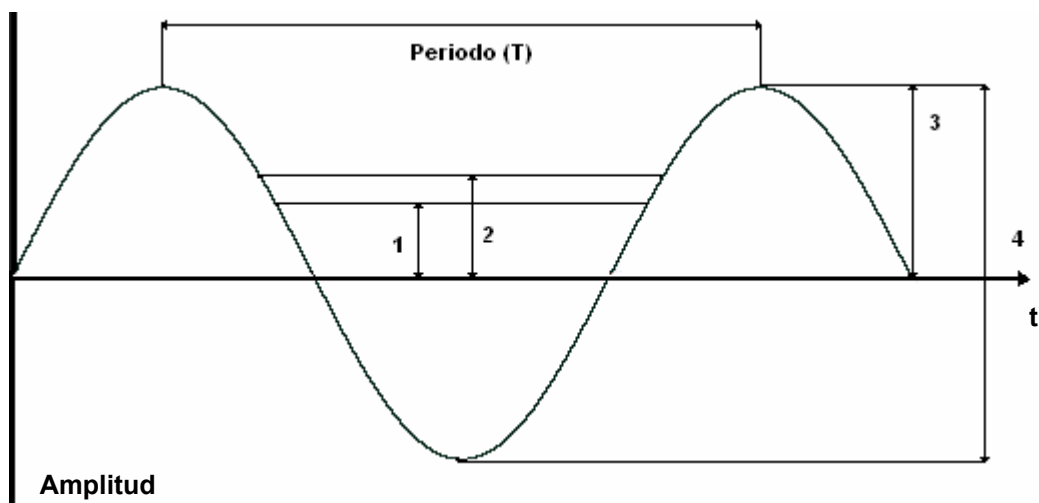
La partículas no se mueven con la *perturbación*, lo que ocurren es que oscilan alrededor de su posición de equilibrio y se produce la *propagación* de unas a otras. [Estas perturbaciones del medio producen las llamadas ondas sonoras que producen en el oído humano una sensación descrita como sonido].

<sup>2</sup> Delgado, Manuel. *Sistemas de radio y televisión* (2001).

### 2.1.2 CARACTERISTICAS DEL SONIDO<sup>3</sup>

Todas las *ondas sonoras* se caracterizan por una serie de parámetros cuyo cálculo es sencillo en señales que tienen *amplitud* y *frecuencia* fijas. Cuando se trata de ondas que no tienen una amplitud y una frecuencia constante en el tiempo es más difícil determinar los parámetros que la caracterizan.

La *forma de onda* más sencilla de caracterizar es con toda probabilidad la del *sonido periódico simple* o tono puro. En este tipo de onda se varía la amplitud de la señal manteniendo constante la frecuencia.



**Figura 1.** Señal sinusoidal

Extraída de *Sistemas de Radio y Televisión (Manuel Delgado)*

- |                     |                          |
|---------------------|--------------------------|
| 1. Amplitud media.  | 3. Amplitud de pico.     |
| 2. Amplitud eficaz. | 4. Amplitud pico a pico. |

#### AMPLITUD

Es el nivel que alcanzan las oscilaciones. La forma de expresar la amplitud de una señal puede hacerse por varios métodos. Según su amplitud de pico, su amplitud media o su amplitud eficaz.

<sup>3</sup>Delgado. *Sistemas de radio y televisión* (2001).



- *Amplitud de pico*: Es la amplitud máxima que alcanza la señal durante el periodo.
- *Amplitud media*: Es la energía media que tiene la señal.
- *Amplitud eficaz*: Es la raíz cuadrado del valor cuadrático medio de la amplitud.

## PERIODO

Es el tiempo en que tarde en finalizar una onda u oscilación. Se mide en segundos y se representa normalmente por la letra T.

## FRECUENCIA

La *frecuencia* de un sonido es el número ondas que hay un segundo. Se mide en *Hercios* (Hz) o en ciclos por segundo (c/s) y se representa por la letra f. La frecuencia también se puede definir como la inversa del periodo.

$$f=1/T$$

En general, los sonidos no son *sinusoidales* y su *amplitud* y *frecuencia* pueden variar con el tiempo.

[El oído humano es capaz de percibir vibraciones que están entre los 20 y los 20000Hz. La frecuencia de un sonido se incrementa al aumentar el número de ciclos por segundo. Por ejemplo, un sonido agudo como el producido por un silbato o una flauta, tienen componentes de alta frecuencia las cuales contienen miles de ciclos por segundo. Los sonidos graves, como los producidos por bajo eléctrico, tienen bajas frecuencias y pocos ciclos por segundo.

Cuando hablamos de *frecuencia fundamental* nos referimos a la frecuencia más baja que puede tener una onda sonora].

## VELOCIDAD DEL SONIDO

Es la velocidad a la que viajan las ondas sonoras a través de un medio determinado que depende *del índice de propagación* y de la *elasticidad del medio*. La *velocidad de propagación* es mayor en sólidos que en líquidos y éste a su vez

mayor que en los gases, cuanto más denso y menos elástico sea el medio, mayor será la velocidad del sonido a través de él. [texto adaptado de Delgado, sistemas de radio y televisión (2001)].

Un factor que afecta directamente a la velocidad del sonido es la *temperatura* ya que, a medida que ésta aumenta, la velocidad también se incrementa.

El sonido viaja en el aire a una temperatura de 20°C a una velocidad de 343m/s y a 0°C a una velocidad de 331m/s. La relación de la velocidad del sonido con la temperatura es:

$$c = 331 + 0,6T$$

Donde: c velocidad del sonido.

T temperatura en grados Celsius.

## LONGITUD DE ONDA

Se define como la distancia entre dos máximos o entre dos mínimos de presión consecutivos de una onda. Se indica con la letra ( $\lambda$ ). La relación entre la frecuencia y la *longitud de onda* es inversamente proporcional es decir, que a medida que la frecuencia aumente la longitud disminuye.

$$\lambda = \frac{c}{f}$$

Donde:

$\lambda$  es la longitud de onda de una onda sonora.

c es la velocidad de propagación de la onda.

f es la frecuencia en Hz.

## TONO<sup>5</sup>

El *tono* de un sonido está determinado por la *frecuencia* del mismo o su frecuencia fundamental. A través del tono podemos clasificar los sonidos en *graves* y *agudos*.

El tono es denominado también *altura* pero esta denominación es usada en psicoacústica para determinar la *percepción del tono*.

## TIMBRE<sup>6</sup>

Es la cualidad del sonido que permite distinguir un sonido de otro, es decir diferenciar de cual fuente proviene. Por ejemplo una nota producida por un piano es también producida por una guitarra, aunque ambas notas tengan una misma frecuencia (el mismo tono) el timbre nos permite saber a que instrumento pertenece cada nota.

El *timbre* se puede determinar por el número y la *intensidad de armónicos* que acompañan a *la frecuencia fundamental* cuando éste se emite.

## ARMÓNICOS<sup>7</sup>

Los armónicos de un sonido, son los múltiplos de la frecuencia fundamental para un determinado sonido. En la naturaleza carecemos de sonidos puros ya que todos tienen una serie de armónicos que lo acompañan.

La intensidad de los armónicos da al sonido un timbre característico. Por ejemplo cuando una onda fundamental esta acompañada hasta séptimo armónico se puede decir que el sonido es rico en armónicos.

### 2.1.3 CLASIFICACIÓN DE LAS ONDAS

- *Según su naturaleza*

#### ONDAS MECÁNICAS

Las ondas mecánicas para propagarse necesitan un medio elástico ya sea sólido, líquido o gaseoso. Las partículas que conforman el medio oscilan por la excitación producida por la onda. Esta oscilación ocurre en un punto fijo sin desplazar a las partículas de su posición inicial transmitiendo así el sonido de una partícula a otra. Esta transmisión es puramente energética.

<sup>5, 6 y 7</sup> Adaptado de texto Delgado, *sistemas de radio y televisión (2001)*.

## ONDAS ELECTROMAGNETICAS

Las ondas electromagnéticas no necesitan un medio para propagarse, es decir pueden propagarse en el vacío, esto debido a que las ondas estas son producidas por oscilaciones de un campo eléctrico en relación con un campo magnético.

- *Según la dirección de la perturbación*<sup>8</sup>

## ONDAS LONGITUDINALES

Una onda longitudinal es aquella en la que la vibración del medio (movimiento de las partículas que transporta la onda) es paralela a la dirección de propagación de la onda. Por ejemplo la voz.

## ONDAS TRANSVERSALES

Una onda transversal es aquella en la que la vibración de las partículas es perpendicularmente a la dirección de propagación de la onda, como por ejemplo las producidas en el agua.

- *Según su periodicidad:*

## ONDAS PERIÓDICAS (ONDAS SIMPLES)

El ejemplo más simple de una *onda periódica* es una *onda senoidal* y la perturbación (*fente sonora*) que las origina produce ciclos repetitivos. En la naturaleza, carecemos de ondas periódicas, pero estas son empleadas para el estudio y el comportamiento del sonido. Este tipo de ondas poseen una única frecuencia.

Además encontramos las *ondas complejas* que son las ondas formadas por la suma de varias ondas simples, este tipo de ondas poseen más de una componente de frecuencia y son el tipo de onda que encontramos en la naturaleza.

<sup>8</sup>Adaptado de Schaum, Daniel. *Teoría y problemas de física general* (1969).

## ONDAS NO PERIÓDICAS

En este tipo de ondas, la fuente sonora que las origina se da separadamente, es decir que su periodicidad no sigue ningún tipo de ciclo o patrón repetitivo.

## ONDAS ESTACIONARIAS<sup>9</sup>

La composición de dos ondas de la misma frecuencia y amplitud que se propagan en la misma dirección, pero en sentido contrario producen una onda estacionaria. Las ondas estacionarias se pueden generar en distintos cuerpos vibrantes como cuerdas o columnas de aire. Recibe el nombre de *nodo* todo el punto de la onda estacionaria tal, que el desplazamiento o elongación de la partícula es constantemente nula.

Se llama *vientre* [antinodo] un punto de la onda estacionaria en la que la elongación de la partícula en vibración es máxima.

### 2.1.4 EL DECIBELIO

Es la unidad de medida del *nivel de presión sonora* (dB) y se define como “El *decibelio* (dB) es una unidad de nivel que denota la relación entre dos cantidades que son directamente proporcionales en su potencia. El numero de decibeles que corresponde a esta relación es 10 veces el logaritmo (de base 10) de la razón de las dos cantidades”.<sup>10</sup>

Anteriormente el *nivel de presión sonora* era medido en Pascales (Pa). El rango de valores que una persona puede escuchar se encuentra entre 20 $\mu$ Pa que seria el nivel más bajo hasta 20Pa que es el nivel de presión sonora capaz de producir dolor.

La diferencia que existe entre estos dos valores es de un millón, por lo que manejar cálculos numéricos con estas cifras tan grandes es muy complicado.

<sup>9</sup> Schaum. *Teoría y problemas de física general* (1969).

<sup>10</sup> Harris, Cryri. *Manual de medidas acústicas y control del ruido*.

Alexander Graham Bell desarrolló una relación logarítmica entre estos valores para facilitar el manejo y además, porque el oído no responde de manera lineal a los sonidos sino que tiene un comportamiento logarítmico.

## NIVEL DE PRESION SONORA

A medida que una onda viaja a través del aire, ésta produce variaciones de presión en el mismo. Estas variaciones se denominan *presión sonora* y se emplean para medir la fuerza de la onda a la que llamamos *nivel de presión sonora (SPL)*.

El nivel de presión sonora expresa en el aire 20 veces el logaritmo en base 10 de la relación entre una presión acústica y una de *presión de referencia* determinada ( $20\mu\text{Pa}$ ). Se determina por medio de la fórmula:

$$\text{SPL (dB)} = 20\text{Log} (p/p_0)$$

Donde:

SPL es el nivel de presión sonora (sound pressure level)

p es la presión sonora instantánea

$p_0$  es la presión de referencia ( $20\mu\text{Pa}$ ).

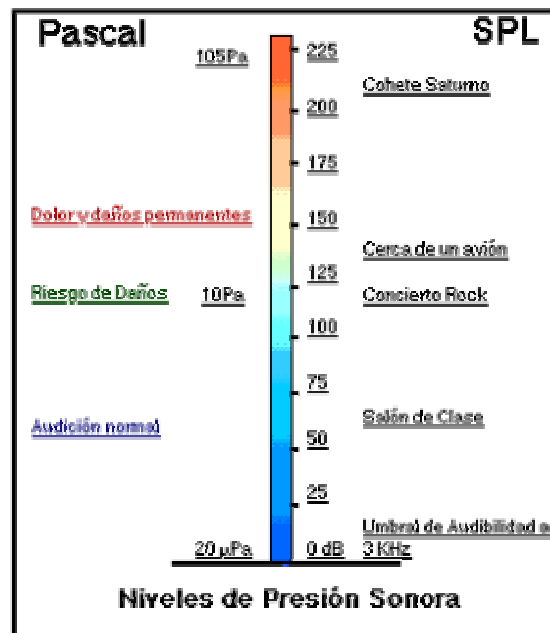


Figura 2. Niveles de presión sonora. Tomado de <http://www.labc.usb.ve/ec4514/audio/sonido>

## INTENSIDAD SONORA<sup>11</sup>

Se define con la letra  $I$  y es el valor medio de la *energía sonora* que atraviesa una unidad de área en el campo en la unidad de tiempo. Podemos ver que cuando un sonido es mas intenso mayor cantidad de energía transporta.

## POTENCIA SONORA<sup>12</sup>

La *potencia sonora* ( $W$ ) es una propiedad que caracteriza únicamente a las *fuentes de sonido* y es la energía que radia una fuente sonora en un instante determinado de tiempo dividido por la duración de dicho intervalo. La potencia sonora es independiente del entorno.

### 2.1.5 EL ESPECTRO AUDIBLE<sup>13</sup>

El *espectro audible* del oído humano es el considerado dentro del rango de frecuencias que va desde los 20 hasta los 20000Hz. Los sonidos cuya frecuencia se encuentre por encima de los 20000Hz se le denominan ultrasonidos, mientras que los sonidos cuya frecuencia sea inferior a la audible se le denominan infrasonidos. El espectro audible varía dependiendo de las características auditivas de la persona y va disminuyendo con la edad, es decir que a medida que pasan los años el oído pierde sensibilidad y sobre todo en altas frecuencias.

El espectro audible se puede dividir en tres rangos:

- Graves.
- Medios.
- Altos.

*Frecuencias graves:* Comprende las frecuencias que van desde los 20Hz hasta los 300Hz aproximadamente, como por ejemplo el bombo de una batería que contiene componentes de frecuencia importantes dentro de este rango.

<sup>11, 12 y 13</sup> Adaptado de Delgado (2001).

*Frecuencias medias:* Corresponde a las frecuencias que van desde los 300Hz hasta los 4000Hz aproximadamente, como por ejemplo la voz humana.

*Frecuencias Altas:* Corresponde a las frecuencias que van desde los 4000 hasta los 20000Hz aproximadamente, como por ejemplo el trinar de un ave.

Además de los sonidos que existen en la naturaleza con todas sus características y cualidades encontramos algunos que no son de nuestro agrado y nos producen cierto tipo de sensaciones, estos sonidos los llamamos ruido.

### **2.1.6 RUIDO**

El *ruido* es un *sonido no deseado* el cual produce una sensación de desagrado que interfiere en la *inteligibilidad de la palabra* y en la interpretación de los sonidos. Además, es un sonido complejo ya que es producido por movimientos vibratorios no periódicos y en general presentan componentes en la mayoría de las frecuencias comprendidas en el *espectro audible*.

En general el ruido además de ser molesto, tiene una gran importancia en mediciones acústicas.

Existen dos clasificaciones de ruido, según la intensidad y el periodo y según la frecuencia.

- *Según la frecuencia.*

Estos tipos de ruidos son emitidos por fuentes generadoras de ruido y son empleados en acústica para realizar ciertas mediciones como aislamiento acústico e insonorización, reverberación, etc.

Encontramos tres tipos:

-*Ruido blanco*

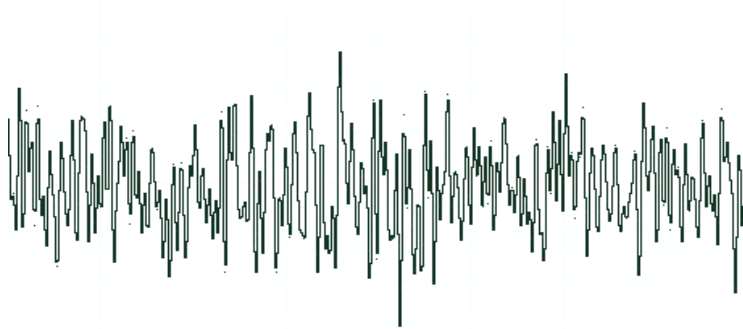
-*Ruido rosa*

-*Ruido marrón*



## RUIDO BLANCO

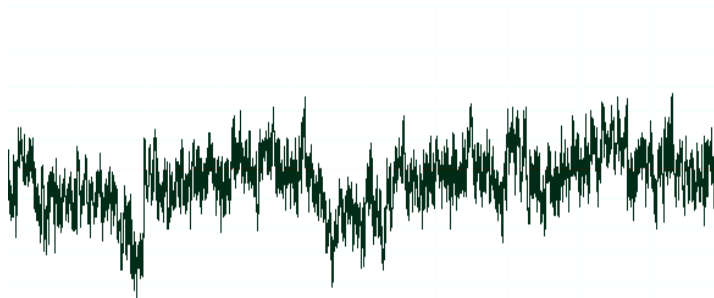
El *ruido blanco* es un ruido aleatorio que se caracteriza por tener una distribución constante de energía en el *espectro audible* (20Hz a de 20 KHz) para cada frecuencia. El ruido blanco ofrece una respuesta plana en frecuencia.



**Figura 3.** Ruido blanco.

## RUIDO ROSA

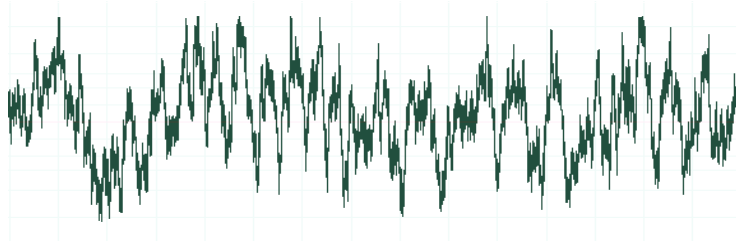
El ruido rosa es un ruido aleatorio el cual su intensidad decae 3dB por octava de frecuencia. El ruido rosa no ofrece una respuesta plana en frecuencia.



**Figura 4.** Ruido Rosa.

## RUIDO MARRÓN

El ruido marrón tiene una densidad espectral de potencia que disminuye con el cuadrado de la frecuencia. Este ruido es producto de realizar una integración al ruido blanco.



**Figura 5.** Ruido marrón.

- *Según la intensidad y el periodo.*

Estos tipos de ruidos los encontramos en nuestra vida diaria y día a día afectan nuestro oído sin percatarnos lo peligrosos que son.

#### RUIDO CONTINUÚO

El *ruido continuo* es aquel donde la intensidad del ruido permanece constante, es decir sufre fluctuaciones mínimas en un intervalo de tiempo determinado. Este ruido es producido por ejemplo, por circulación vehicular o maquinaria la cual no sufre procesos de apagado o interrupción.

#### RUIDO IMPULSIVO

El *ruido impulsivo* es aquel que su intensidad tiene un incremento rápido producto de un golpe o un impulso. Estos impulsos tienen corta duración, por lo general contienen una gran cantidad de energía y son producidos por explosiones, impactos, etc.

#### RUIDO INTERMITENTE

El *ruido intermitente* es aquel que su intensidad varía en un tiempo determinado. Estas variaciones pueden ser periódicas o aleatorias. Por ejemplo este tipo de ruido lo encontramos en el paso de los aviones, el paso de automóviles, etc.

## RELACIÓN SEÑAL RUIDO

La relación señal ruido es la diferencia que hay entre la señal de interés y el ruido de fondo de un determinado sistema. Esta relación es medida en decibeles (dB).

### 2.1.7 FUENTE DE RUIDO Y DISTORSIONES EN GRABACIONES<sup>14</sup>

Las fuentes de ruido pueden provenir por diferentes acontecimientos durante una grabación. Estos ruidos pueden ser impulsivos como por ejemplo golpes, puertas rozamiento, explosiones, etc. y continuos como maquinaria industrial, oficinas, ruido de circulación vehicular, ect. El ruido intermitente producido por paso vehicular, sonido del viento, mal canal de comunicación, etc.

Es necesario separar todos estos ruidos (ruido de banda ancha) que tiene componentes espectrales de cambio aleatorio en muchas frecuencias dentro de las grabaciones.

El ruido de banda angosta consiste de una serie de armónicos producidos por ruido de motor de grabación, tonos, obstáculos de sistema eléctrico, etc. son ruidos regulares de banda angosta.

La mayor parte del ruido puede ser asumido como *aditivo* a la señal útil, como por ejemplo, las fuentes de ruido acústico agregadas al ruido de la señal de voz en la misma habitación donde esta situado el micrófono o las fuentes electromagnéticas agregadas al ruido en las líneas de transmisión de sonido, líneas de grabación de sonido y líneas de reproducción.

En muchos casos la influencia de ruido aditivo se puede disminuir con medios técnicos, usualmente el ruido cambia en el tiempo un poco mas despacio que la voz, esto permite encontrar y remover solo componentes de ruido.

<sup>14</sup>Adaptado de Speech Technology Center, Scientific Department, *Cancelación de ruido y decodificación de texto de grabaciones de voz de baja calidad: Enfoque práctico (2002)*.

## INFLUENCIA DEL RUIDO Y DISTORSIONES EN LA PERCEPCIÓN DE VOZ

El ruido y las distorsiones impiden la percepción de voz debido al funcionamiento del oído humano. La diferencia entre una fuente de señal de voz distorsionada y una señal de voz clara es:

- La presencia de componentes de sonido extra adicionales.
- La ruptura del rendimiento perimétrico de señal de voz.
- la proporción de amplitudes y fases de diferentes componentes de voz.

Así cada imagen de voz de una grabación distorsionada cuando es oída es percibida es reconocida por el humano como:

- Ruido de fondo extra.
- Forma no estándar, inusual.

Esto resulta en el hecho que imágenes de sonido familiar no son reconocidas o son reconocidas con dificultad.

Una característica específica de la percepción del oído humano define los rasgos de reconocimiento de señales distorsionada, esto debido al carácter no lineal de sensibilidad de oído y a la selectividad de la señal ruidosa y la frecuencia.

Además, el oído en algunos casos se ve afectado por ciertos factores que dificultan el reconocimiento de una señal:

- Enmascaramiento de frecuencia y tiempo.
- Fatiga.
- Individualidad.
- Asociación de la organización de memoria de voz.
- Alucinaciones auditivas.

Una de las principales causas del mal reconocimiento de información de voz ruidosa es la presencia dentro de estas componentes extra que enmascaran la señal de voz útil. El oído se caracteriza por enmascarar las señales en frecuencia y tiempo.

El *enmascaramiento de frecuencia* es un fenómeno que ocurre cuando una señal de bajo nivel se vuelve inaudible (enmascarada) producto de una señal ruidosa (mascara) que se emite simultáneamente y que se encuentra a un valor cercano de frecuencia a la señal de bajo nivel. Existen dispositivos y programas que se introducen en la señal de voz que pueden enmascarar los componentes de voz débiles que están cerca de la frecuencia influenciando el proceso de correcto reconocimiento de señal.

El *enmascarado de tiempo* juega un papel importante en la percepción del oído humano. Esto toma lugar cuando dos sonidos aparecen dentro de un periodo pequeño de tiempo. El pico de amplitud de señal que aparece debilita las señales (por amplitud) y estas se vuelven inaudibles durante un tiempo después de haber sonado e incluso más adelante.

La información principal de la señal que llega al oído debe tener una pequeña variación con respecto a las típicas para voz. De lo contrario la inteligibilidad de voz es disminuida y un sonido es percibido como no natural.

Cuando el oído percibe una señal ruidosa por un tiempo determinado se produce *fatiga* en la audición, la sensibilidad y habilidad para reconocer disminuye considerablemente. Las *alucinaciones auditivas* son las producidas por imágenes de audición generadas por neuronas de oído no por señales de entrada.

La *individualidad* de audición se expresa en como una dispersión larga de la sensibilidad de audición y habilidad para reconocer la voz con ruido de persona a persona. Por ejemplo en un fragmento de ruido blanco un operador puede escuchar frases que no existen.

## 2.1.8 ANÁLISIS ESPECTRAL DE FRECUENCIA

Para realizar un análisis de espectral de frecuencia es necesario conocer:

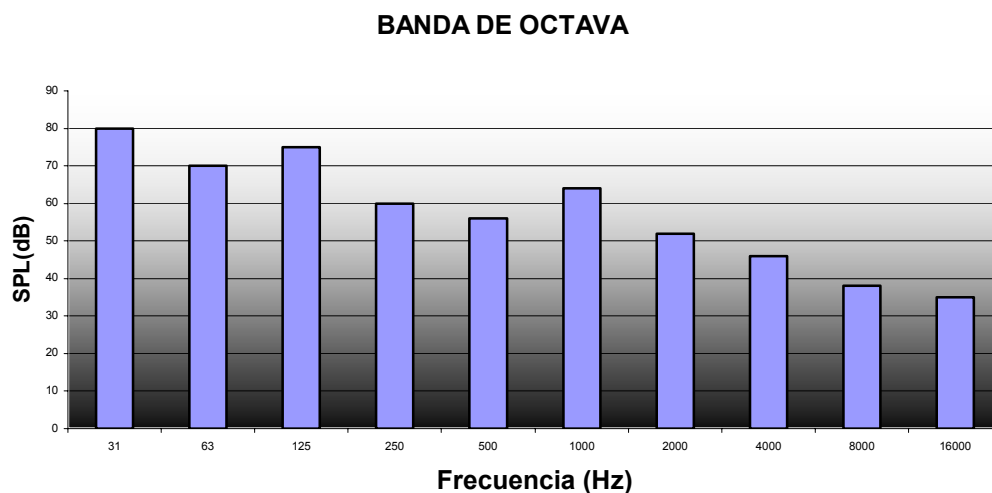
### OCTAVA

Es el intervalo que existe entre una frecuencia fundamental y el doble de su valor. Por ejemplo, si tenemos una frecuencia de 250Hz su octava es 500Hz, es decir que tienen una relación de 2:1.

### BANDA DE OCTAVA

*Banda de octava* es el segmento del espectro frecuencial cuyo ancho es de una octava y está comprendida por el *rango audible de frecuencias* para el oído humano. Una *banda de tercio de octava* es el segmento del *espectro frecuencial* cuyo ancho es de un tercio (1/3) de la banda de octava, es decir que encontramos una subdivisión de la banda de octava permitiendo un análisis mas detallado del sonido.

La gama de frecuencias se separa en bandas de octava por conveniencia, en donde cada banda específica se representa por su frecuencia de centro que cubre la mayoría de la gama de las frecuencias audibles.: 63Hz, 125Hz, 250Hz, 500Hz, 1KHz, 2 KHz, 4 KHz, 8 KHz, y 16 KHz,



**Figura 6.** Banda de octava.

## ANCHO DE BANDA

El *ancho de banda* es el tamaño de un rango de frecuencias en el que se concentra la mayor parte de la potencia de la señal, entonces el ancho de banda (que se crea a través de un filtro) es la diferencia entre las frecuencias en las que su atenuación al pasar a través de dicho filtro se mantiene igual o inferior a 3dB comparada con la frecuencia central de pico.

Una señal periódica simple de una sola frecuencia tiene un ancho de banda mínimo. Por ejemplo, la señal periódica tiene componentes en varias frecuencias, su ancho de banda es mayor.

Las señales generadas en la realidad ya sea la voz, señales de radio y televisión o datos informáticos (sistemas electrónicos) son señales que no son periódicas y varían en el tiempo, pero con la suma de muchas señales periódicas de diferentes frecuencias se pueden caracterizar, que es el principio en el cual se basa el Análisis de Fourier.

## FILTROS

Los filtros son mecanismos de discriminación para una señal determinada de entrada para obtener ciertas variaciones deseadas en su salida. Encontramos filtros digitales y análogos.

Los filtros analógicos se emplean para procesar señales analógicas o continuas en el tiempo. Los filtros digitales se usan para procesar señales digitales en el tiempo. Estos filtros digitales tienen una entrada la cual puede ser una señal analógica o digital y una salida la cual puede ser otra señal analógica o digital, en ese proceso de filtrado podemos afectar o cambiar factores como la amplitud, frecuencia o fase dependiendo de las características del filtro.

## TIPOS DE FILTROS

Existen tipos de filtros según la parte del espectro de información frecuencial que atenúan y permiten pasar:

- Filtro pasa alto.
- Filtro pasa bajo.
- Filtro pasa banda.
- Banda eliminada
- Multibanda
- Pasa todo
- Resonador
- Oscilador
- Filtro peine (combfilter)
- Filtro ranura (notchfilter)

Según su orden:

- Primer orden
- Segundo orden
- Etc.

Según el tipo de respuesta ante una entrada unitaria:

- FIR (Finite Impulse Response)
- IIR (Infinite Impulse Response)
- TIIR (Truncated Infinite Impulse Response)

A la hora de entrar a estudiar y a analizar el ruido vemos que el espectro de frecuencia se debe dividir en *bandas de octava* que por medio de *filtros*, permiten pasar el ruido entre frecuencias máximas y mínimas, donde podemos evaluar la composición de ruido. Estos filtros actúan rechazando las frecuencias cuyos valores sean superiores o inferiores a los límites establecidos.



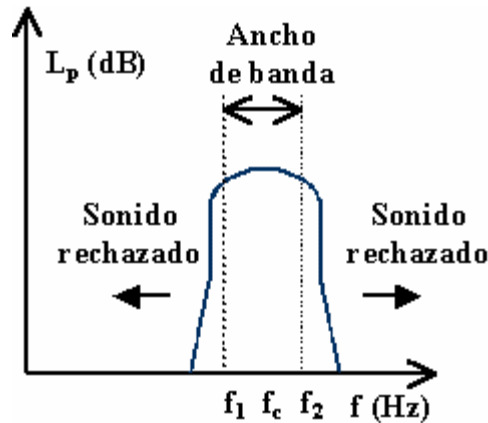


Figura 7. Ancho de banda. Tomado de [www.ehu.es/acústica/español/ruido/genes/genes.html](http://www.ehu.es/acústica/español/ruido/genes/genes.html)

### 2.1.9 FORMATOS DE GRABACION DE AUDIO

A la hora de grabar una señal es muy importante conocer en que tipo de formato debo almacenar el audio, ya que existen varios tipos de formatos que comprimen la señal lo que produce pérdidas de información.

Wave y AIFF (Audio Interchange File Format)

Estos son formatos de audio no comprimidos y almacenan aproximadamente 10Mb por minuto, lo que indica que son archivos de gran tamaño, entonces necesitamos un espacio suficiente en donde guardar dicha información. Estos tipos de archivos pueden ser leídos en cualquier parte del mundo por el reproductor más sencillo pues fueron estandarizados

Además, son muy flexibles lo que facilita el tratamiento del sonido ya que permite ser comprimido y grabado en distintas calidades y tamaños (11025, 22050, 44100Hz).

MP3

Formato de audio con capacidad de compresión media que afecta la calidad en cuanto a información mas no en percepción, ya que su algoritmo se basa en la forma como el oído humano escucha, pues las frecuencias que no audibles son suprimidas (por encima de 20khz y debajo de 20hz). Este formato almacena

aproximadamente 1MB por minuto, este valor de almacenamiento puede variar debido a que el MP3 puede crearse a varios índices de compresión.

## WINDOWS MEDIA

Formato WMA (Windows Media Audio), es el formato de competencia al formato de compresión MP3. Estos son ligeramente más pequeños que los que MP3 y una calidad de sonido similar, pero no todos los reproductores pueden leerlo.

### 2.1.10 DIGITALIZACIÓN<sup>15</sup>

La *señal digital* es la representación de un conjunto de datos binarios de una onda sonora (análoga). Un sistema de audio digital consiste en un *transductor* (micrófono) que convierte la onda de presión en una señal eléctrica analógica.

Esta señal analógica sufre procesos de filtrado, compresión, amplificación, que están determinadas por las características del transductor (micrófono).

Tras el procesamiento analógico la señal, ésta se convierte a datos binarios (digital) a través de 3 procesos:

- *Muestreo.*
- *Cuantificación.*
- *Codificación.*

El *muestreo* o comúnmente llamado *sampling* es la toma de un conjunto de valores discretos de la amplitud de onda en la señal analógica. Estas muestras son intervalos fijos en el tiempo (periódicamente).

Las muestras deben efectuarse en el mismo intervalo y con la misma duración.

El número de muestras que se toman por segundo de la señal es lo que llamamos *frecuencia de muestreo* (tasa de muestreo) o *frecuencia de muestreo*.

<sup>15</sup>Adaptado de *Introducción al audio digital*. Juan Ignacio Arribas (2001-2002) <http://www.lpi.tel.uva.es/>

La calidad de una señal digital es directamente proporcional al número de muestras, es decir al valor de la *frecuencia de muestreo*. Cuanto mayor sea el número de muestras, más confiable y exacto será la calidad de la señal resultante.

Según el *teorema de muestreo* de Nyquist-Shannon, para poder replicar con exactitud la forma de una onda es necesario que la *frecuencia de muestreo* sea como mínimo el doble de la frecuencia a muestrear.

En la siguiente tabla podemos ver las frecuencias de muestreo en diferentes aplicaciones.

FRECUENCIA DE MUESTREO	APLICACIONES
<b>AUDIO</b>	
8,000 Hz	Telefonia
11,025 Hz	Toma de muestra de voz(hombre y mujer)
22,050 Hz	Radio.
32,000 Hz	Video digital en formato mini DV.
44,100 Hz	CD,Formatos MPEG-1 (VCD, SVCD, MP3).
48,000 Hz	Sonido en televisión digital, DVD, sistemas DAT.
96,000 ó 192,400 Hz	Audio de alta definición para DVD y BD-ROM
<b>VIDEO</b>	
50 Hz	Video PAL.
60 Hz	Video NTSC.

**Tabla 1.** Frecuencias de muestreo.

La *cuantificación* es la asignación de valores discretos a cada una de las muestras, las muestras de amplitud continua son transformadas en sucesiones de valores de amplitudes discretas. El valor de cada muestra de la señal se representa dentro de un conjunto finito de valores posibles según los niveles de voltaje medidos.

En este proceso la señal sufre una pérdida de información ya que la señal ya no es idéntica a la original.

La *codificación* es la asignación de una secuencia de bits a cada valor de tensión eléctrica analógica ya cuantificada al sistema binario, mediante códigos preestablecidos. Esta secuencia intenta reproducir la forma de la onda de la señal de entrada.

La *resolución digital* viene determinada por el número de bits. Cuanto mayor sea la resolución, más precisa será la representación. Por ejemplo, con una resolución de 8 bits, el rango de variación de la señal se divide en 256 subintervalos, mientras que con una resolución de 16 bits lo hace en 65536 subintervalos, cuya amplitud será, por consiguiente, mucho menor.

El audio digital para el consumo masivo por ejemplo el CD tiene una resolución de 16 bits. En sistemas de audio profesional se utilizan 20 bits y aún 24 bits.

¿Porqué digitalizar la señal de voz?<sup>16</sup>

- Poder trabajar con software de análisis de voz.
- Permite rápido procesamiento y separación de locutores.
- Encriptación.
- Posibilidad de almacenarla y transmitirla.
- Control de errores, regeneración.
- Mejor relación señal-ruido.
- Mayor privacidad y seguridad de la información.

<sup>16</sup> Kittsteiner, Escobar, *Algoritmos de compresión* [http://electricidad.uvipro.cl/teleco/guia\\_teleco/comp-sen.ppt](http://electricidad.uvipro.cl/teleco/guia_teleco/comp-sen.ppt)

Después de ver el proceso de digitalización la pregunta más frecuente a la hora de tomar una muestra de voz son<sup>17</sup>:

1. ¿Cuál debe ser la frecuencia de muestreo?
2. ¿Qué sucede cuando no seleccionamos la frecuencia de muestreo adecuadamente?
3. ¿Cuál debe ser la resolución?
4. ¿Qué relación hay entre el ruido de digitalización y la resolución?
5. ¿Que formato de compresión se deben usar?

1. A la hora de tomar una muestra de voz es muy importante definir que *frecuencia de muestreo* debemos utilizar. El rango de frecuencia abarcado por la voz comprende aproximadamente entre los 100 y los 7.000Hz, pero la información relevante para uso forense se encuentra en un ancho de banda entre los 300 y 4.500Hz.

Según este *ancho de banda*, aplicando el teorema de Nyquist la frecuencia de muestreo (debe ser mayor que el doble de la máxima frecuencia presente en la señal) óptima para la voz es de **11.025Hz**. Por ejemplo en grabaciones telefónicas donde *el canal de comunicación* solo permite una frecuencia máxima de 4000Hz, podríamos decir que bastaría con una frecuencia de muestreo de 8000Hz.

2. Cuando no seleccionamos la *frecuencia de muestreo* adecuadamente, es decir que no se cumple el *teorema* de Nyquist al intentar reconstruir la señal se generan frecuencias espurias (todas las frecuencias indeseadas que puedan resultar de cualquiera de los procesos involucrados en la generación, amplificación o modulación) que no estaban presentes originalmente. Supongamos.

<sup>17</sup>Las preguntas son adaptadas del FAQ de *audio digital* por Federico Miyara. [www.eie.fceia.unr.edu.ar/~acustica/audigfaq.htm](http://www.eie.fceia.unr.edu.ar/~acustica/audigfaq.htm)

Ahora, si queremos muestrear una señal que contiene además un ruido de 30 kHz. Si utilizamos la frecuencia normalizada de 44,1 kHz, a pesar que el ruido es inaudible por ser superior al límite del oído humano (20 kHz), cuando recuperamos la señal aparece un ruido en la frecuencia de 14.1 kHz (44,1 kHz - 30 kHz), que se encuentra dentro del espectro audible, este tipo de frecuencias se denominan *frecuencias alias*.

*Aliasing* es el efecto que produce una distorsión cuando se utiliza una frecuencia menor a la establecida por el teorema de Nyquist. El *aliasing* impide recuperar correctamente la señal cuando las muestras de ésta se obtienen a intervalos de tiempo demasiado largos. La forma de la onda recuperada presenta pendientes muy abruptas.

La solución para eliminar el *aliasing* es incluir *filtros* pasa bajos, que eliminan todas las frecuencias que sobrepasan la frecuencia correspondiente a la mitad con la que se llevó a cabo el muestreo elegida en la señal de entrada. Entonces todas las frecuencias que queden por encima de la frecuencia de muestreo seleccionada son eliminadas.

3. *La resolución digital* que debemos emplear para la toma de muestra de voz es de **16 bits**. El bit en un sistema se describe como su exactitud potencial para representar una onda de audio grabada o procesada por unidad de tiempo. Cada bit representa incrementar en 6 dB el rango dinámico de nuestra señal, por lo que 16 bits representan 96 dB y 24 bits representa grabar con un rango dinámico de 144 dB.

4. La relación señal/ruido (S/R) es la mejor manera de evaluar el ruido de cualquier sistema incluidos los de audio digital. La máxima relación señal ruido que puede obtenerse para el audio digital es aproximadamente  $6 \cdot n$ , donde  $n$  es la resolución en bits.

Ahora un sistema digital de 16 bits admite una S/R de  $6 \cdot 16 = 96$  dB. Por otra parte en señales analógicas debido a sus características, la relación S/R tiende a ser

menor que el valor anteriormente mencionado, por ejemplo 90 dB. Es necesario tener en cuenta el rango dinámico del oído, que en promedio en personas de escucha normal se encuentra en 100 dB.

##### 5. Que formato de compresión se deben usar en el análisis acústico del habla

Las *técnicas de compresión* son herramientas fundamentales con que se cuenta para alcanzar un adecuado balance entre capacidad de almacenamiento y procesamiento requerido.

Las *técnicas de compresión* suministran una importante reducción en la capacidad de almacenamiento, y a la vez requieren un procesado importante tanto para compresión como para la descompresión. Las técnicas más sencillas brindan reducciones moderadas con poco procesamiento.

Las técnicas más avanzadas simplifican aquellos elementos irrelevantes para la sensación sonora, debido a que éstas, analizan la respuesta del oído con respecto a la señal (formato MP3) logrando valores de compresión mucho mayor.<sup>18</sup>

Pero en el *análisis acústico forense de la voz*, estos valores discriminados en estas técnicas, son de una gran importancia ya que representan gran contenido de información frecuencial para los sistemas biométricos.

“Las señales de voz pueden tener un rango dinámico superior a los 60 dB, por lo que para conseguir una alta calidad de voz se deben usar un elevado número de niveles de reconstrucción. Sin embargo, interesa que la resolución del cuantificador sea mayor en las partes de la señal de menor amplitud que en las de mayor amplitud. Por tanto, en la cuantificación lineal se desperdician niveles de reconstrucción y, consecuentemente, ancho de banda. Esto se puede mejorar incrementando la distancia entre los niveles de reconstrucción conforme aumenta la amplitud de la señal”.<sup>19</sup>

<sup>18</sup> Adaptado de López. *Formatos de audio*.Pdf

<sup>19</sup> Kittsteiner, *Algoritmos de compresión*.

Al ser cuantificada uniformemente a la salida la señal será expandida, invirtiendo la compresión (técnica de compresión).

#### MU-LAW<sup>20</sup>

Es un algoritmo de Ley Mu empleado para las telecomunicaciones, un estándar de compresión de señales análogas para la modulación por códigos de pulsos (PCM) para mejorar o modificar el rango dinámico de una señal análoga previo a ser digitalizada.

Esta *encriptación* es usada para la voz, ya que la compresión posee un rango dinámico ancho, haciendo que no se produzca una encriptación lineal eficiente.

La percepción de intensidad es logarítmica, lo que hace la Ley Mu es reducir el rango dinámico de la señal, aumentando la eficiencia de compresión, teniendo como resultado una información con una relación SEÑAL-RUIDO mucho mejor que la que entrega una encriptación lineal. La ley  $\mu$  utiliza 15 segmentos y no 8 como PCM sin optimización. La ley A-law utilizada 13 segmentos.

Existen distintos tipos de compresión pero los más utilizados y recomendado para el uso forense son PCM (Pulse Code Modulation) y ADPCM (Adaptative differential Pulse Code Modulation). Estos tipos manejan una *codificación diferencial*: en lugar de representar las muestras de la señal, se almacenan la diferencia entre muestras consecutivas, que suele ser pequeña para la señal de audio.

#### 2.1.11 Formatos de archivo<sup>21</sup>

Los *formatos de archivo* indican la estructura con la que el audio es almacenado.

Se distingue entre dos grupos:

<sup>20</sup> Adaptado de *Algoritmos de compresión*.

<sup>21</sup> López, Alberto. Formatos de audio digital adaptado de

[http://pi.tel.uva.es/~nacho/docencia/ing\\_ond\\_1/trabajos\\_01\\_02/formatos\\_audio\\_digital/html/frames.htm](http://pi.tel.uva.es/~nacho/docencia/ing_ond_1/trabajos_01_02/formatos_audio_digital/html/frames.htm)



- *Formatos autodescriptivos*: Nos muestran una clara información sobre los parámetros de codificación en algún punto del archivo.
- *Formatos sin cabecera o tipo raw*: La información de codificación son es fija.

Los *Formatos de archivo autodescriptivos* permiten la elección entre varias codificaciones como por ejemplo la que se especifica en una *cabecera*.

La *cabecera* es un número mágico que es un valor fijo que admite identificar el archivo como un archivo del formato buscado.

Esta *cabecera* contiene la información sobre la *tasa de muestreo*, el *número de bits* por muestra y otro tipo de información como por ejemplo *copyright*.

Los *Formatos de fichero sin cabecera o tipo raw* precisan un solo esquema de codificación y no permiten variación de los parámetros, no obstante en algunos casos nos muestra la tasa de muestreo, aunque muchas veces no se puede conocer de ninguna forma la tasa de muestreo, a menos que se reproduzca el sonido.

Gracias a los avances tecnológicos en cuanto al sonido, una nueva clase de tecnología se viene aplicando años atrás. Este tipo de tecnología que siempre esta en avance son los sistemas de identificación y control mediante medidas biométricas. Sistemas informáticos capaces de detectar intrusos y autenticar a los usuarios mediante huellas dactilares, identificaciones de retina o de iris, reconocimiento de patrones en las manos, etc. Esta ciencia se denomina la biometría.

## 2.2 MARCO LEGAL O NORMATIVO

En Colombia no existen leyes o normas que nos regulen el desarrollo de este proyecto, pero si leyes que regulen el manejo y la obtención de información de personas como los contemplados en el artículo 15 de la CONSTITUCIÓN POLÍTICA DE COLOMBIANA el cual dicta que “Todas las personas tienen derecho a su intimidad personal y familiar y a su buen nombre, y el Estado debe respetarlos y hacerlos respetar. De igual modo tiene derecho a conocer, actualizar y rectificar las informaciones que se hayan recogido sobre ellas en bancos de datos y en archivos de entidades públicas y privadas. En la recolección, tratamiento y circulación de datos se respetarán la libertad y demás garantías consagradas en la Constitución. La correspondencia y demás formas de comunicación privada son inviolables. Sólo pueden ser interceptadas o registradas mediante orden judicial, en los casos y con las formalidades que establezca la ley”.

Por otra parte en la recomendación de equipos se tiene en cuenta la normativa *DIN 45500* la cual hace referencia a las reglas que establecen las demandas mínimas que un producto debe satisfacer antes de que pueda ser llamado “de alta fidelidad”. Estas reglas no sólo incluyen estándares si no también describen las medidas sobre las cuales los datos técnicos deben ser basados.

## 2.3 MARCO TEORICO

### 2.3.1 LA VOZ HUMANA

#### DESCRIPCIÓN ANATÓMICA DEL SISTEMA FONATORIO<sup>22</sup>

La voz humana se produce voluntariamente por medio del aparato fonatorio. Está formado por los pulmones como fuente de energía en la forma de un flujo de aire, la laringe, que contiene las cuerdas vocales, la faringe, las cavidades oral (o bucal) y nasal y una serie de elementos articulatorios: los labios, los dientes, el alvéolo, el paladar, el velo del paladar y la lengua (figura 8).

Las cuerdas vocales son, en realidad dos membranas dentro de la laringe orientadas de adelante hacia atrás (figura 9). Por delante se unen en el *cartilago tiroides* (que puede palparse sobre el cuello, inmediatamente por debajo de la unión con la cabeza; en los varones suele apreciarse como una protuberancia conocida como *nuez de Adán*).

Por detrás, cada una está sujeta a uno de los dos *cartílagos aritenoides*, los cuales pueden separarse voluntariamente por medio de músculos.

La abertura entre ambas cuerdas se denomina *glotis*.

Cuando las cuerdas vocales se encuentran separadas, la glotis adopta una forma triangular. El aire pasa libremente y prácticamente no se produce sonido. Es el caso de la respiración.

Cuando la glotis comienza a cerrarse, el aire que la atraviesa proveniente de los pulmones experimenta una turbulencia, emitiéndose un ruido de origen aerodinámico conocido como *aspiración* aunque en realidad acompaña a una espiración o exhalación. Esto sucede en los sonidos denominados “aspirados” (como la h inglesa).

<sup>22</sup>Miyara, Federico. *La voz humana* adaptado de <http://www.eie.fceia.unr.edu.ar/~acustica/biblio/fonatori.pdf>

Al cerrarse más, las cuerdas vocales comienzan a vibrar a modo de lengüetas, produciéndose un sonido tonal, es decir periódico.

La frecuencia de este sonido depende de varios factores:

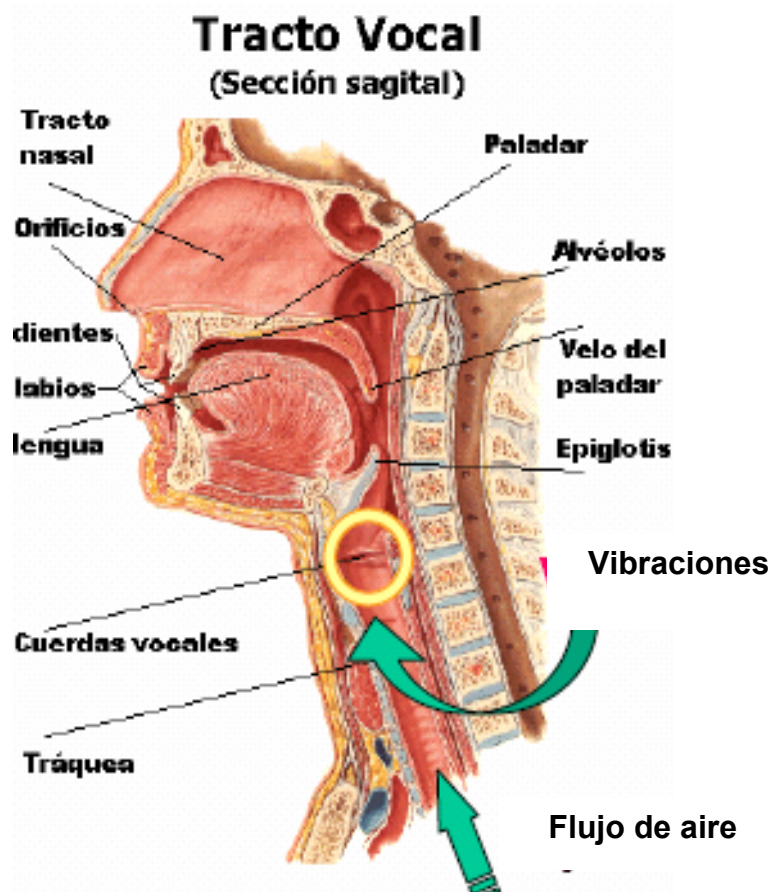
- El tamaño y la masa de las cuerdas vocales.
- La tensión que se les aplique
- La velocidad del flujo del aire proveniente de los pulmones.

A mayor tamaño, menor frecuencia de vibración, lo cual explica por qué en los varones, cuya glotis es en promedio mayor que la de las mujeres, la voz es en general más grave.

A mayor tensión la frecuencia aumenta, siendo los sonidos más agudos. Así, para lograr emitir sonidos en el registro extremo de la voz es necesario un mayor esfuerzo vocal.

En los hombres los valores de abertura y cierre de están aproximadamente en 100Hz y 250Hz en las mujeres.

También aumenta la frecuencia (a igualdad de las otras condiciones) al crecer la velocidad del flujo de aire, razón por la cual al aumentar la intensidad de emisión se tiende a elevar espontáneamente el tono de voz.



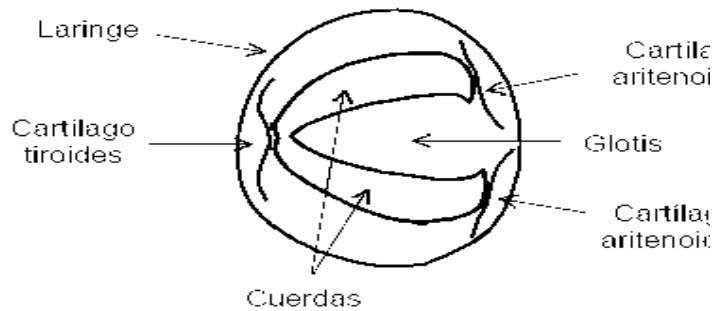
**Figura 8.** Sistema fonatorio.

Tomado de <http://gavab.escet.urjc.es/recursos/JOrtega04.pdf>



**Figura 9.** Cuerdas vocales.

Tomado de [gavab.escet.urjc.es](http://gavab.escet.urjc.es)



**Figura 10.** Corte esquemático del aparato fonatorio humano.  
Tomado de *La voz humana* Federico Miyara.

Finalmente, es posible obturar la glotis completamente. En ese caso no se produce sonido. Sobre la glotis se encuentra la *epiglotis*, un cartílago en la faringe que permite tapar la glotis durante la deglución para evitar que el alimento ingerido se introduzca en el tracto respiratorio. Durante la respiración y la fonación (emisión de sonido) la epiglotis está separada de la glotis permitiendo la circulación del flujo de aire. Durante la deglución, en cambio, la laringe ejecuta un movimiento ascendente de modo que la glotis apoya sobre la epiglotis.

La porción que incluye las cavidades faríngea, oral y nasal junto con los elementos articulatorios se denomina genéricamente *cavidad supraglótica*, en tanto que los espacios por debajo de la laringe, es decir la tráquea, los bronquios y los pulmones, se denominan *cavidades infraglóticas*.

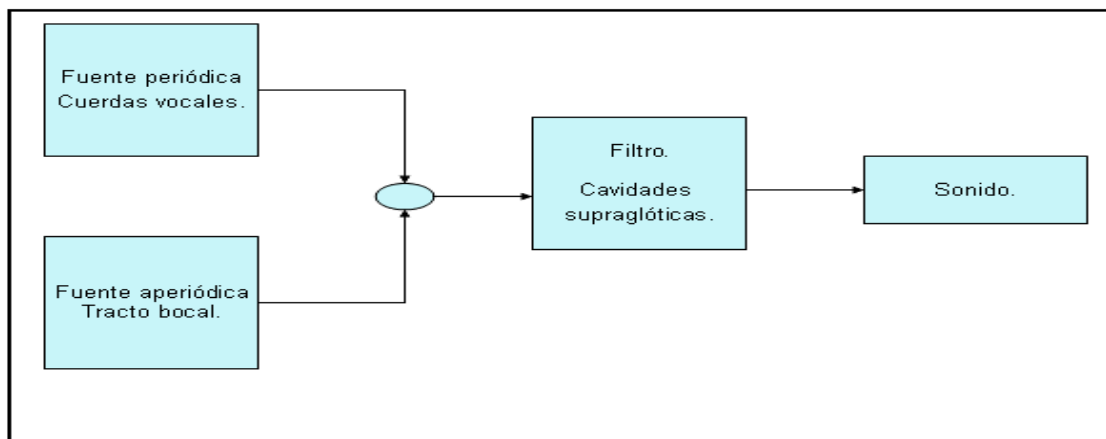
Varios de los elementos de la cavidad supraglótica se controlan a voluntad, permitiendo modificar dentro de márgenes muy amplios los sonidos producidos por las cuerdas vocales o agregar partes distintivas a los mismos, e inclusive producir sonidos propios. Todo esto se efectúa por dos mecanismos principales: el *filtrado* y la *articulación*.

El *filtrado* actúa modificando el espectro del sonido. Tiene lugar en las cuatro cavidades supraglóticas principales: la faringe, la cavidad nasal, la cavidad oral y la cavidad labial.

Las mismas constituyen resonadores acústicos que enfatizan determinadas bandas frecuenciales del espectro generado por las cuerdas vocales, conduciendo al concepto de *formantes*, es decir una serie de picos de resonancia ubicados en frecuencias o bandas de frecuencia que, según veremos, son bastante específicas para cada tipo de sonido.

## EL MODELO DE LA FUENTE Y EL FILTRO<sup>23</sup>

El sistema fonador puede considerarse una fuente, el lugar donde se produce la corriente de aire necesario para la producción del sonido y un filtro debido al sistema de cavidades que por el fenómeno de la resonancia, modifican las características de la fuente.



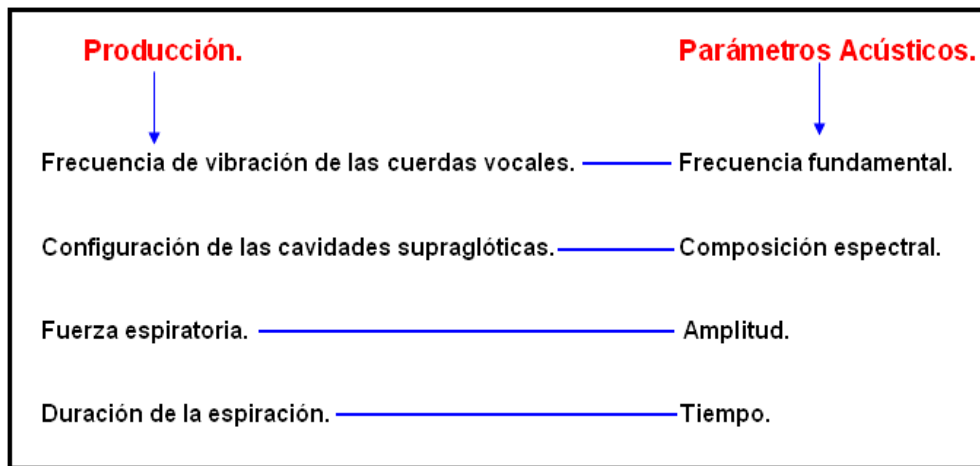
**Figura 11.** Modelo simplificado de producción de sonidos del habla a partir de una fuente y un filtro.  
Adaptado de [http://liceu.uab.es/~joaquim/phonetics/fon\\_anal\\_acus/fon\\_acust.html#clasificacion\\_acustica](http://liceu.uab.es/~joaquim/phonetics/fon_anal_acus/fon_acust.html#clasificacion_acustica).

La *articulación* es una modificación principalmente a nivel temporal de los sonidos, y está directamente relacionada con la emisión de los mismos y con los fenómenos transitorios que los acompañan. Está caracterizada por el lugar del tracto vocal en que tiene lugar, por los elementos que intervienen y por el modo en que se produce, factores que dan origen a una clasificación fonética de los sonidos.

<sup>23</sup> Listerri, Joaquín. *Phonetics analysis acoustic* (2006).

Adaptado de [http://liceu.uab.es/~joaquim/phonetics/fon\\_anal\\_acus/fon\\_acust.html#clasificacion\\_acustica](http://liceu.uab.es/~joaquim/phonetics/fon_anal_acus/fon_acust.html#clasificacion_acustica).

En la figura 11 se presentan las características acústicas de los sonidos del habla y el modo de producción



**Figura 12.** Características acústicas de los sonidos del habla. Adaptado de *Phonetics analysis acoustic*. Llsterri (2006).

### 2.3.2 CLASIFICACIÓN DE LOS SONIDOS DE LA VOZ<sup>24</sup>

Los sonidos emitidos por el aparato fonatorio pueden clasificarse de acuerdo con diversos criterios que tienen en cuenta los diferentes aspectos del fenómeno de emisión.

Estos criterios son:

- Según su carácter vocálico o consonántico.
- Según su oralidad o nasalidad
- Según su carácter tonal (sonoro) o no tonal (sordo)
- Según el lugar de articulación
- Según el modo de articulación
- Según la posición de los órganos articulatorios
- Según la duración

<sup>24</sup>Miyara, Federico. *La voz humana*.



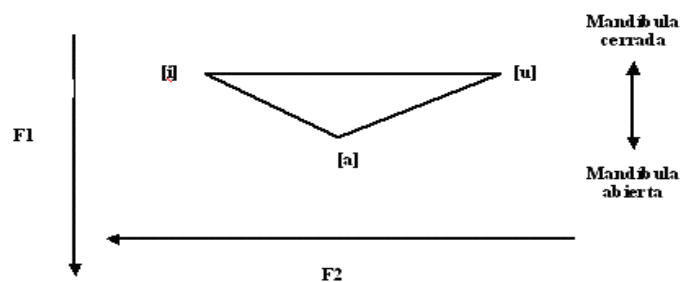
Fuente	Filtro	Clase de sonidos
Periódica	Fijo, oral	vocales orales
Aperiódica continua	Fijo, oral	fricativas sordas
Aperiódica impulsional	Variable, oral	oclusivas sordas
Aperiódica continua + periódica	Fijo, oral	fricativas sonoras
Aperiódica impulsional + periódica	Variable, oral	oclusivas sonoras
Aperiódica impulsional + periódica	Variable, nasal	nasales
Aperiódica continua + periódica	Variable, oral	laterales y vibrantes

**Tabla 2.** Clasificación acústica de los sonidos del habla en función de la fuente y el filtro Adaptado de LANDERCY, A.- RENARD, R. (1977) *Éléments de phonétique*. Mons / Bruxelles: Centre International de Phonétique Appliquée / Didier.

## VOCALES Y CONSONANTES

Desde un punto de vista mecano-acústico, las *vocales* son los sonidos emitidos por la sola vibración de las cuerdas vocales sin ningún obstáculo o constricción entre la laringe y las aberturas oral y nasal. Dicha vibración se genera por el principio del oscilador de relajación, donde interviene una fuente de energía constante en la forma de un flujo de aire proveniente de los pulmones. Son siempre sonidos de carácter tonal (cuasiperiódicos), y por consiguiente de espectro discreto.

### Clasificación de las vocales en abiertas, medias y cerradas



**Figura 13.** Vocales abiertas, medias y cerradas. Tomado de *Phonetics analysis acoustic*. Listerri (2006)

Clasificación de las vocales en anteriores, localización de la constricción (constricción palatal), centrales (constricción faríngea) y posteriores (constricción velar)

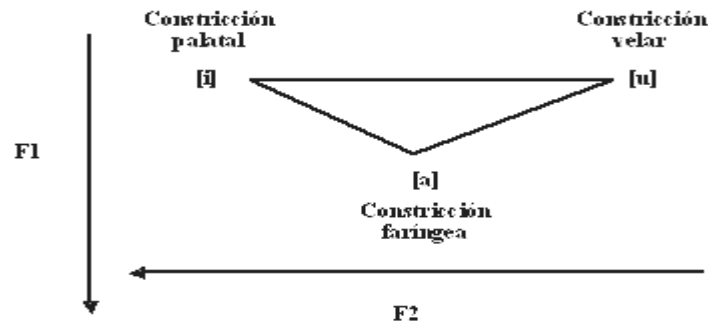


Figura 14. Vocales anteriores, centrales y posteriores.  
Tomado de *Phonetics analysis acoustic*. Listerri (2006)

Posición de los labios Clasifica las vocales en labializadas y no labializadas

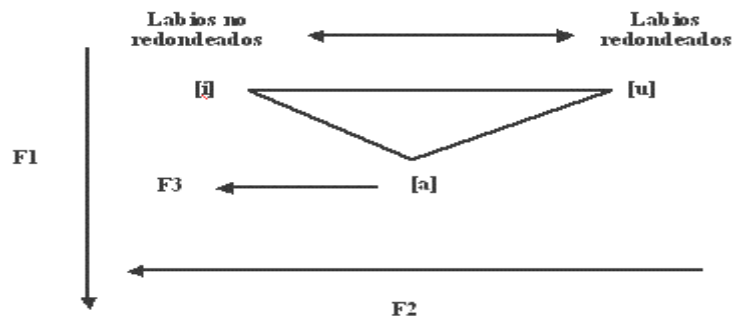


Figura 15. Vocales labializadas y no labializadas.  
Tomado de *Phonetics analysis acoustic*. Listerri (2006)

Las *consonantes*, por el contrario, se emiten interponiendo algún obstáculo formado por los elementos articulatorios. Los sonidos correspondientes a las consonantes pueden ser tonales o no, dependiendo de si las cuerdas vocales están vibrando o no. Funcionalmente, en el castellano las vocales pueden constituir palabras completas, no así las consonantes.

## ORALIDAD Y NASALIDAD

Los fonemas en los que el aire pasa por la cavidad nasal se denominan *nasales*, en tanto que aquéllos en los que sale por la boca se denominan *orales*. La diferencia principal está en el tipo de resonador principal por encima de la laringe (cavidad nasal y oral, respectivamente). En castellano son nasales sólo las consonantes “m”, “n”, “ñ”.

## TONALIDAD

Los fonemas en los que participa la vibración de las cuerdas vocales se denominan *tonales* o, también, *sonoros*. La tonalidad lleva implícito un espectro cuasiperiódico.

Como se puntualizó anteriormente, todas las vocales son tonales, pero existen varias consonantes que también lo son: “b”, “d”, “m”, etc. Aquellos fonemas producidos sin vibraciones glotales se denominan *sordos*. Varios de ellos son el resultado de la turbulencia causada por el aire pasando a gran velocidad por un espacio reducido, como las consonantes “s”, “z”, “j”, “f”.

## LUGAR Y MODO DE ARTICULACIÓN (CONSONANTES)

La *articulación* es el proceso mediante el cual alguna parte del aparato fonatorio interpone un obstáculo para la circulación del flujo de aire. Las características de la articulación permitirán clasificar las consonantes.

Los órganos articulatorios son los labios, los dientes, las diferentes partes del paladar (alvéolo, paladar duro, paladar blando o velo), la lengua y la glotis. Salvo la glotis, que puede articular por sí misma, el resto de los órganos articula por oposición con otro.

Según el lugar o punto de articulación se tienen fonemas:

- *Bilabiales*: oposición de ambos labios
- *Labiodentales*: oposición de los dientes superiores con el labio inferior
- *Linguodentales*: oposición de la punta de la lengua con los dientes superiores
- *Alveolares*: oposición de la punta de la lengua con la región alveolar
- *Palatales*: oposición de la lengua con el paladar duro
- *Velares*: oposición de la parte posterior de la lengua con el paladar blando
- *Glotales*: articulación en la propia glotis.

A su vez, para cada punto de articulación ésta puede efectuarse de diferentes modos, dando lugar a fonemas:

- *Oclusivos*: la salida del aire se cierra momentáneamente por completo.
- *Fricativos*: el aire sale atravesando un espacio estrecho.
- *Africados*: oclusión seguida por fricación.
- *Laterales*: la lengua obstruye el centro de la boca y el aire sale por los lados.
- *Vibrantes*: la lengua vibra cerrando el paso del aire intermitentemente.
- *Aproximantes*: La obstrucción muy estrecha que no llega a producir turbulencia.

Los fonemas oclusivos (correspondientes a las consonantes “b” inicial o postnasal, “c”, “k”, “d”, “g” inicial, postnasal o postlateral, “p”, “t”) también se denominan a veces *explosivos*, debido a la liberación repentina de la presión presente inmediatamente antes de su emisión. Pueden ser sordos o sonoros, al igual que los fricativos (“b” postvocálica, postlateral y postvibrante, “g” postvocálica y post vibrante, “f”, “j”, “h” aspirada, “s”, “y”, “z”). Sólo existe un fonema africado en castellano, correspondiente a la ch”. Los laterales (“l”, “ll”) a veces se denominan líquidos, y son siempre sonoros.

Los dos fonemas vibrantes del castellano (consonantes “r”, “rr”) difieren en que en uno de ellos (“r”) se ejecuta una sola vibración y es intervocálico, mientras que en el otro (“rr”) es una sucesión de dos o tres vibraciones de la lengua.

Lugar de articulación	Modo de articulación								
	Oral							Nasal	
	Oclusiva		Fricativa		Africada	Lateral	Vibrante	Aproximante	Sonora
	Sorda	Sonora	Sorda	Sonora	Sorda	Sonora	Sonora	Sonora	
Bilabial	p	b, v		b, v				w	m
Labiodental			f						
Linguodental			z	d					
Alveolar	t	d	s	y	ch	l	r, rr		n
Palatal				(y)	(ch)	ll		i	ñ
Velar	k	g	j	g					
Glotal			h						

**Tabla 3.** Clasificación de las consonantes de la lengua castellana según el lugar y el modo de articulación y la sonoridad. Tomado de Federico Miyara, *La voz humana*.

Finalmente, los fonemas Aproximantes (la “i” y la “u” cerradas que aparecen en algunos diptongos) son a veces denominados semivocales, pues en realidad suenan como vocales. Pero exhiben una diferencia muy importante: son de corta duración y no son prolongables.

## LA RESONANCIA<sup>25</sup>

La *resonancia* es la modificación de la amplitud de los *armónicos de un sonido complejo* ejercida por la cavidad en la que se produce la vibración de las moléculas de aire.

Las cavidades pequeñas refuerzan la amplitud de los *armónicos de frecuencia alta* por esto percibimos un sonido agudo y las cavidades grandes refuerzan la amplitud de los *armónicos de frecuencia baja* obteniendo sonidos percibidos como graves.

<sup>25</sup> Listerri, Joaquín. *Phonetics analysis acoustic* (2006).

En las cavidades supraglóticas la amplitud de los armónicos de la onda procedente de la glotis depende de la posición adoptada para la articulación del sonido. Especialmente en el caso de los sonidos periódicos complejos, se refuerza la amplitud de grupos de armónicos situados alrededor de una determinada frecuencia, configurándose así los formantes.

## FORMANTES<sup>26</sup>

Las *formantes* son las *resonancias* producidas por *el tracto vocal* debido a las dimensiones y la velocidad de propagación del sonido, aparece en promedio 1 formante por cada kHz. El tracto vocal *filtra* la onda glotal: amplifica cada componente de frecuencia con una determinada ganancia.

El filtrado del tracto vocal considerando las pérdidas por radiación tiene una caída en altas frecuencias 6 dB / década.

## ESTACIONARIEDAD DE LA VOZ<sup>27</sup>

La señal de voz es *estacionaria a trozos* durante la pronunciación de un fonema es cuasiestacionaria debido a los cambios de velocidad del tracto vocal y cambios en la velocidad de las cuerdas vocales. La voz sufre ondas estacionarias durante 20 – 40 ms. La velocidad de pronunciación es de 5-20 fonemas / seg. El Análisis de estos “trozos de voz estacionarios” se llama ventanas.

## INTELIGIBILIDAD<sup>28</sup>

La inteligibilidad del habla viene definida por la claridad con que una persona se expresa, de modo que su habla sea bien comprensible a su interlocutor.

<sup>26</sup> y <sup>27</sup> Adatado de De la Torre Vega, Ángel. *Procesamiento de señales de voz* (2002)

<http://ceres.ugr.es/~atv/Documents/Docencia/voz.ppt>

<sup>28</sup> Collado, Concejero, Tapias, Castellanos, Hernández. *Medida de la inteligibilidad del habla: comparación entre humanos y sistemas automáticos de reconocimiento.*

[http://www.tid.es/documentos/boletin/numero21\\_1.pdf](http://www.tid.es/documentos/boletin/numero21_1.pdf)

Los términos inteligibilidad y articulación se refieren a la misma cuestión. El término inteligibilidad se entiende como la comprensión de palabras, frases o de un mensaje, (Sorkin & Kantowitz, 1987). Pero se pueden encontrar multitud de formulaciones, como la de la ANSI considerando la inteligibilidad como unidades de habla que son identificadas.

Más extensamente, la inteligibilidad de un sistema de comunicación de habla es la propiedad (dados unos oyentes entrenados) de recibir e identificar habla reproducida, tanto por hablantes entrenados como por un habla codificada, cuando el hablante o habla codificado y el oyente están comunicados por un sistema de comunicación de habla. (ANSI S3.2-1989 -ASA 85-1989).

#### MEDIDA DE LA INTELIGIBILIDAD<sup>29</sup>

La inteligibilidad se mide a través del *índice de articulación*, que indica el porcentaje de aciertos en la comprensión de una cantidad de emisiones vocales. Hay tres tipos de índices: el *índice de articulación silábico*, en el que se hace escuchar al sujeto cierta cantidad de sílabas sueltas sin sentido, el *índice de articulación de palabras*, para el cual se utilizan palabras en general de dos sílabas, y el *índice de articulación de frases*, que utiliza frases completas. En todos los casos las emisiones son *fonéticamente balanceadas*, es decir, los fonemas aparecen en la misma proporción que en el habla normal.

Se observa que a partir de 35 a 40 dBA de ruido ambiente los índices de articulación empiezan a disminuir. También se observa que el índice silábico es menor que el de palabras y éste que el de frases. Es decir, en una condición dada, se entienden más las frases completas que las palabras o las sílabas sueltas. Esto se debe a que la variedad de sílabas sueltas es mucho mayor que la de palabras o frases. Virtualmente cualquier combinación de consonantes y vocales da una sílaba válida, pero no cualquier combinación de sílabas es una palabra válida.

<sup>29</sup> Miyara, Federico. *El ruido y la inteligibilidad de la palabra* (2004).

<http://www.eie.fceia.unr.edu.ar/~acustica/biblio/inteligibilidad.pdf>

Por ejemplo, “caza” podría confundirse con “cafa”, pero como ésta no es una palabra válida, el sujeto se decide por “caza”. Análogamente, no toda combinación de palabras de una frase correcta, tanto en su sintaxis como en su sentido. Al decir “ese hombre caza conejos”, la palabra “caza” podría confundirse con “casa”, “caja”, “cafa”, “taza”, “tasa”, “tafa”, etc. Sin embargo, la única palabra válida y que da sentido a la frase es “caza”.

En una frase, las vocales tienen su principal contenido en frecuencias por debajo de los 1000Hz., dando la impresión de sonoridad.

La boca humana las produce con gran ángulo de apertura, pudiendo, en un recinto, incidir sobre superficies duras (paredes, cielorraso) creando fácilmente reverberación. Las consonantes, con un espectro de frecuencias por encima de los 1000Hz. y menor nivel que las vocales, proveen la articulación.

La boca humana produce estos sonidos con un ángulo de apertura pequeño, por lo que son direccionales. [Adaptado de Ruffa, Francisco. Inteligibilidad.ppt]

Teniendo en cuenta todas las características fisiológicas que afectan la voz en su reproducción, es necesario considerar que factores externos como la grabación, digitalización y procesamiento de las señales, inciden en la calidad de las muestras de voz.

Según la obtención de la muestra, a nivel forense se denomina material dubitadas e indubitadas.

### **2.3.3 MATERIAL FONOGRAFICO DUBITADO E INDUBITADO**

#### **MATERIAL DUBITADO**

Es aquel donde la voz de una persona es obtenida mediante grabaciones. Este material se adquiere del individuo por medio de grabadoras de bolsillo, interceptaciones telefónicas (radiofrecuencia, telefonía satelital, celular, fija, etc.), video grabaciones, a través de micrófonos ocultos, etc.



Cada uno de estos canales de comunicación posee características específicas que difieren la calidad de unas a otras, como por ejemplo el *ancho de banda*, la *relación señal-ruido*, calidad del equipo, etc.

Una característica muy importante de este tipo de material es, que al ser grabado en un lugar *no controlado acústicamente*, presentan problemas de inteligibilidad ocasionados por agentes externos tales como:

- Ruido vehicular.
- Ruido ambiental.
- Reverberación.
- Factores climáticos.
- Equipo de grabación no adecuado.
- Técnicas inadecuadas de grabación.
- Otros.

#### MATERIAL INDUBITADO

Voz obtenida de la persona involucrada en la causa y de quien se sospecha pertenece la prueba dubitada. Este material se adquiere *directamente* del individuo por medio de una grabación realizada en un recinto destinado para este fin.

La *calidad de una grabación sonora* dependerá del recinto en que se realice, ya que, un recinto que no esté *acondicionado acústicamente* producirá ecos y reverberación en la grabación que no son posibles de *eliminar* por medio de procesamiento electrónico de la señal.

A la hora de diseñar un recinto acústicamente apto para la toma de muestra de voz, es necesario tener en cuenta los principales problemas que pueden afectar la calidad de la grabación los cuales son: *aislamiento* y *acondicionamiento acústico*.

En primer lugar, los recintos diseñados para este fin, deben tener un *aislamiento acústico* para evitar ruidos que interfieran en la muestra. Por ejemplo ruido aéreo, vehicular, flujo de personas, de maquinaria, oficinas, etc.

Este aislamiento debe cumplir con ciertas leyes fundamentales para su diseño.

En segundo lugar, es conveniente tener un  *acondicionamiento acústico, y térmico,* con sistemas de climatización internos controlados, ya que no pueden aportar ningún nivel sonoro que afecten la grabación.

Dentro de un recinto, el sonido tiene un comportamiento que dependerá de las *dimensiones y materiales* de las superficies que lo conforman. Interiormente fenómenos como la reflexión, eco, tiempo de reverberación, respuesta en frecuencia del recinto afectan la señal en su dominio temporal y frecuencial.

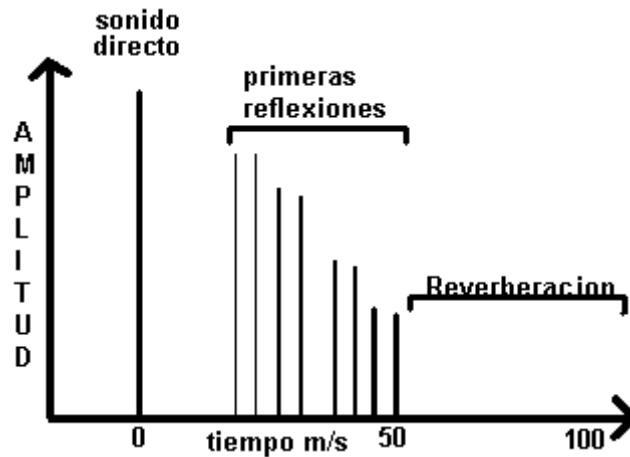
La propagación del sonido en un recinto cerrado se presenta de dos formas diferentes:

- Sonido directo.
- Sonido reflejado.

En el *sonido directo* la energía depende exclusivamente de la *distancia de separación* entre *la fuente* (voz) y *el receptor* (micrófono).

El *sonido reflejado* es la energía proveniente de las reflexiones del sonido (camino recorrido por la onda sonora) y del *coeficiente de absorción* de las superficies del recinto.

Cuanto mayor sea la distancia recorrida y mas absorbente sean los materiales del recinto, menor será la energía que tiene el sonido directo y el sonido reflejado.



**Figura 16.** Sonido directo y reflejado, Tomado de <http://www.codepixel.com/tutoriales/sonido51>.

Cuando se analiza el sonido reflejado (*reflexión del sonido*) de un recinto, observamos dos zonas con características muy diferentes. La primera zona está compuesta por las primeras reflexiones o *reflexiones tempranas* y la segunda formada por las reflexiones tardías.

Todas las reflexiones que llegan al receptor dentro de los primeros 50ms desde que llega el *sonido directo*, el oído las integra y no las percibe como diferentes una de la otra. Para la voz humana, estas reflexiones favorecen la inteligibilidad ya que aumenta la *sonoridad*.

Cuando las reflexiones sufren un retardo mayor a 50ms el oído percibe una repetición del sonido directo, esto es lo se denomina *eco*.

Un fenómeno derivado de la reflexión del sonido es *la reverberación* que consiste en la prolongación del sonido una vez que la *fente sonora* ha cesado, esto debido a las ondas reflejadas que viajan dentro del recinto.

Los factores nombrados con anterioridad afectan la inteligibilidad de la palabra, siendo esta importante para realizar un análisis audio perceptivo dentro de las etapas de la identificación del locutor. Por esta razón controlarlos es indispensable.

### 2.3.4 EQUIPOS UTILIZADOS EN LA CAPTURA DE LA SEÑAL DUBITADA E INDUBITADA

Para grabar la señal dubitada e indubitada existen diferentes dispositivos los cuales pueden ser análogos o digitales.

#### GRABACIÓN MAGNÉTICA<sup>30</sup>

Para grabar una señal análoga es necesario contar con los equipos basados en el principio de *grabación de magnetismo retenido*.

Cuando se aplica un campo magnético sobre un objeto metálico como el hierro o el acero, las partículas que componen el metal adoptan las propiedades del *campo magnético*. Este principio de grabación lo podemos encontrar en cintas de carrete abierto, VHS, cassette y microcassete. Estos dos últimos son los más utilizados en el área de acústica forense.

Las cintas magnéticas que conforman dichos dispositivos, deben tener un cuidado especial ya que, son sensibles a cambios de temperatura, humedad, fragilidad, campos magnéticos externos y por sus características de fabricación son propensos a rupturas, lo cual produce una variación o pérdida en la información que es de difícil recuperación comprometiendo la autenticidad de la cinta.

Los equipos destinados a la grabación de la señal análoga deben cumplir con ciertos parámetros de calidad que no afecten la señal de la voz. Estos parámetros son el ancho de banda del canal de grabación, relación señal-ruido, respuesta en frecuencia, etc.

<sup>30</sup> Adaptado de Delgado (2001).

En cuanto al tipo de grabadora, existen varios modelos para todos los propósitos. Son básicamente dos tipos de grabadora, a saber: de bobina abierta o de bobina cerrada.

Las grabadoras de bobina abierta son poco utilizadas en la actualidad ya que el avance de la tecnología a permitido la reducción del tamaño de las grabadoras, además que el precio y el mantenimiento son costosos por la escasez de producción de las mismas; por otra parte vale destacar que calidad de grabación es excelente.

De otra forma existen las grabadora de bobina cerrada siendo estas las que utilizan los casetes que todos conocemos y las podemos dividir en tres subtipos: tipo bolsillo, tipo sobremesa y tipo microcinta.

<b>Grabadora de bobina cerrada</b>	<b>Calidad.</b>
Tipo bolsillo	Buena
Tipo sobremesa	Buena
Tipo microcinta.	Baja.

**Tabla 4.** Tipos de grabadora bobina cerrada.

## GRABACIÓN DIGITAL

La grabación digital se produce por la conversión análogo-digital de una onda sonora (señal análoga). La señal al ser grabada, sufre una transformación de información que se traduce en códigos binarios.

Durante el proceso de conversión análogo-digital, la señal sufre tres procesos importantes los cuales son: muestreo cuantificación y codificación.

Existen tres métodos de grabación digital. La grabación magnética digital el cual almacena la información en cinta magnética (DAT). La grabación óptica digital es la efectuada por un soporte vía óptica, a través de un láser que lee la información.

(CD). La grabación magneto-óptico digital emplea un sistema combinado de grabación magnética con reproducción óptica de los (disco duro).

Ventajas de la grabación digital frente a la grabación análoga:

- La señal digital permite la reproducción infinita sin pérdidas de calidad.
- Facilidad para el procesamiento de la señal.
- La señal digital es menos sensible que la analógica a las interferencias.
- La señal digital tiene una mejor relación señal-ruido.
- En caso de pérdida de cierta cantidad de información, la señal digital puede ser reconstruida gracias a señales.

Desventajas de la grabación digital:

- La transmisión de señales digital requiere una sincronización precisa entre los tiempos del reloj de transmisor, con respecto a los del receptor. Un desfase, por mínimo que sea, cambia por completo la señal.
- La señal digital requiere mayor ancho de banda para ser transmitida que la analógica.
- Se necesita una conversión analógica-digital previa y una decodificación posterior, en el momento de la recepción.
- Existen valores de señales análogas en los cuales los equipos digitales no pueden alcanzar, debido a la estandarización de 96 dB de rango dinámico en señales digitales.

### **2.3.5 BIOMETRÍA**

“La biometría son métodos automatizados para la identificación o verificación de individuos mediante el uso de características físicas o del comportamiento de la persona. Esta tecnología se basa en la premisa de que cada persona es única y posee rasgos distintivos que pueden ser utilizados para identificarla”.<sup>31</sup>

<sup>31</sup>Vargas, Enrique. *Biometría la nueva clase secreta* (2001).

La biometría es aplicada en dos tipos de características:

Las características físicas:

- Las huellas digitales (geometría de la mano).
- El iris de los ojos.
- Exploración de la retina.
- Reconocimiento facial (2D o 3D, por termografía o luz visible).
- Análisis de ADN.

Las características de comportamiento:

- *Identificación por la voz.*
- Identificación por la escritura (la firma).
- Análisis del patrón de marcha.
- Dinámica de pulsación en teclado.

## SISTEMAS BIOMÉTRICOS<sup>32</sup>

El funcionamiento de los sistemas biométricos en general actúa de manera similar y consta de dos procedimientos. El primer procedimiento radica en que la persona debe ser registrada en un sistema de almacenamiento de datos. Durante este proceso, el sistema captura el rasgo característico de la persona, como por ejemplo la voz y lo procesa para crear una representación electrónica llamada modelo de referencia.

Dicho modelo de referencia debe ser almacenado en una base de datos dentro del sistema, como por ejemplo una tarjeta inteligente o en un medio magnético seguro, en el cual pueda hacerse uso dicha información.

El segundo procedimiento a seguir depende de dos factores:

- Si el sistema biométrico consiste en verificar la identidad de la persona.
- Si el sistema biométrico consiste en identificar a la persona.

<sup>32</sup> Adaptado de Vargas (2001).

Si el sistema biométrico es empleado para verificación, la persona le informa al sistema cual es su identidad ya ingresando alguna clave especial o sea presentando una tarjeta de identificación.

Entonces, el sistema procede a la captura el rasgo característico de la persona (la voz por ejemplo) y la procesa para crear una representación electrónica llamada modelo en vivo. Después, el sistema compara el modelo en vivo con el modelo de referencia de la persona. Si ambos modelos coinciden la verificación es exitosa. De no serlos, la verificación es fallida.

Si el sistema biométrico consiste en la identificación, la persona no le informa al sistema biométrico cual es su identidad. El sistema solo captura el rasgo característico de la persona y lo procesa para crear el modelo en vivo. Luego el sistema procede a comparar el modelo en vivo con un conjunto de modelos de referencia que se encuentran almacenados en una base de datos con registros de otras personas para determinar la identidad de dicha persona.

En caso dado que la en verificación o la identificación sea exitosa, el sistema biométrico le concede ciertos privilegios a la persona, por ejemplo permiso para entrar a un área restringida o tener acceso a su cuenta de banco. Pero en el caso que la comparación no sea exitosa, los privilegios para dicha persona son negados.

## APLICACIONES DE SISTEMAS BIOMÉTRICOS <sup>33</sup>

Algunas aplicaciones comunes en la verificación son:

- Control de identidad por las autoridades.
- Control de acceso a un sistema informático.

<sup>33</sup> Pérez, Juan Carlos. *Reconocimiento de formas en biometría* (2005).



- Identificación en votaciones.
- Utilización de dispositivos (celulares, automóviles, etc.)
- Control de acceso a un recinto.
- Utilización de servicios (cajeros automáticos, transporte público, etc.).
- Cobro de servicios (pago a distancia, comercio electrónico, etc.)

Algunas aplicaciones comunes en la identificación:

- *Identificación forense por medio de toma de muestras de la voz.*
- Cobro automático sin interacción del usuario.
- Identificación forense de huellas dactilares.
- Detección de sujetos relacionados con terrorismo, delincuencia, etc.
- Control de fronteras.

#### IDENTIFICACIÓN POSITIVA Y NEGATIVA<sup>34</sup>

Además de la teoría básica de biometría surge también, una teoría moderna la cual define el segundo procedimiento (verificación o identificación) dependiendo si la función del sistema biométrico consiste en identificación positiva o identificación negativa.

Cuando hablamos de una identificación positiva nos referimos a que el sistema biométrico entra a comprobar si la identidad de la persona está registrada en el sistema. La persona hace una reclamación positiva de identidad al sistema biométrico, es decir, la persona alega que está registrada en el sistema.

Entonces, el sistema responde automáticamente comparando el modelo en vivo con varios modelos de referencia. Si la persona es identificada, se le conceden ciertos privilegios, de lo contrario son negados.

<sup>34</sup> Adaptado de Vargas (2001).

Ahora, si la función del sistema biométrico es la identificación negativa, la cual consiste en probar que la identidad de la persona no está registrada en el sistema, por ejemplo un sistema de verificación para personas que ingresan a una entidad bancaria, verificar que éstas no se encuentren en una lista de delincuentes.

La persona le hace una reclamación negativa de identidad al sistema biométrico, el cual responde comparando automáticamente el modelo en vivo con varios modelos de referencia. Si la identidad de la persona no está registrada, el sistema le concede ciertos privilegios, por ejemplo permitirle entrar al banco.

Pero si el sistema biométrico reconoce a la persona, este le niega dichos privilegios y dependiendo de los antecedentes de la persona puede una alerta en el caso que se deba tomar medidas más radicales como detener la persona.

Existen ciertos parámetros sobre los cuales un sistema biométrico trabaja como: <sup>35</sup>

- *Universalidad*: La inmensa mayoría de la población ha de poseer ese rasgo.
- *Facilidad de medida*: El rasgo debe poder medirse cuantitativamente de forma segura y sencilla.
- *Unicidad*: Debe ser distinto para cada persona.
- *Permanencia*: Debe ser invariante en el tiempo.
- *Aceptabilidad*: por parte de los usuarios.
- *Resistencia al engaño*: con elementos artificiales o incluso con partes del cuerpo del usuario legítimo.
- *Buenas prestaciones*: y bajo consumo de recursos.

[Para nuestro interés enfocaremos el estudio de los sistemas biométricos en el Reconocimiento de locutores por la voz].

<sup>35</sup> Pérez, (2005).

### 2.3.6 BIOMETRÍA DE VOZ

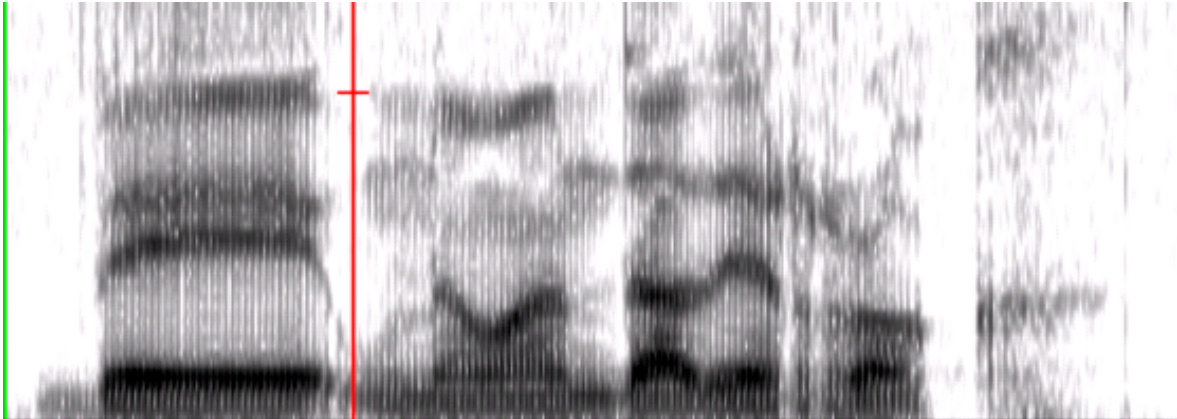


Figura 17. Grafica de una señal de voz en un espectrograma.

La voz es una de las características de que utilizan los sistemas biométricos para identificar a las personas. Al contestar el teléfono escuchamos una voz e inmediatamente reconocemos a alguien, podemos comprender la riqueza de esta característica como método de reconocimiento. Al igual que los sistemas basados en el rostro, el reconocimiento mediante la voz es muy aceptado entre sus usuarios. [Adaptado de texto *Biometría la nueva clase secreta*. Vargas (2001)]

Como este sistema biométrico es fundamentalmente usado para confirmar la identidad, esta confirmación debe cumplir con los parámetros anteriormente mencionados los cuales tienen una viabilidad ya sea baja o alta:

Viabilidad biométrica de la voz: <sup>36</sup>

- Facilidad de medida: *Alta*. Comodidad.
- Universalidad: *Alta*.
- Permanencia: *Baja*. Varía a menudo.
- Unicidad: *Baja*. Por la gran variabilidad.

<sup>36</sup> Pérez, (2005).

- Prestaciones: *Buenas*.
- Aceptabilidad: *Alta*. Natural.
- Resistencia al engaño: *Baja*. Grabaciones.

Mediante el análisis de los sonidos que emitimos, estos sistemas también crean modelos de la anatomía de la tráquea, cuerdas vocales y cavidades, ya que no basta con la forma en que hablamos y el tono de voz que tenga la persona, muchos de estos sistemas operan independientemente del idioma o el acento de la persona.

Para un funcionamiento óptimo del sistema se requiere de una sala especial, sin ecos, ruidos externos y con una buena acústica.

De esta manera las interferencias externas se reducen al máximo. Existen dos modelos para estos sistemas. En algunos casos este tipo de sistema biométrico puede fallar en caso de enfermedad que afecte a la voz del usuario, o, simplemente, si existe un cambio de ánimo en éste.

### **2.3.7 LA ACÚSTICA FORENSE**

“La acústica forense es una parte de la criminalística que engloba la aplicación de técnicas desarrolladas por la ingeniería acústica para el esclarecimiento de los delitos y la averiguación de la identidad de quienes los cometen. No existe más límite, en la selección y uso de esas técnicas, que el impuesto por la casuística pericial. Por tanto, los apartados que se exponen en este trabajo responden a los requerimientos judiciales por orden de importancia o, al menos, de frecuencia en las peticiones de informes periciales”.<sup>37</sup>

<sup>37</sup> Molina, José Juan. *La acústica forense* (2005)

[www.uned.es/investigacion/publicaciones/Cuadernillo\\_octubre200503.pdf](http://www.uned.es/investigacion/publicaciones/Cuadernillo_octubre200503.pdf)

### 2.3.7.1 RECONOCIMIENTO DE LOCUTORES POR LA VOZ<sup>38</sup>

El reconocimiento de locutores por medio de la voz fue una de las primeras tecnologías que recibió nombre propio dentro de la ciencia forense y el campo policial en la década de los sesenta en los Estados Unidos.

Esta tecnología de reconocimiento se le llamó VOICEPRINT y sirvió como base para realizar investigaciones de habla y música relacionadas con sistemas de comunicaciones. Además, esta técnica implementaba un instrumento de medida al que conocemos con el nombre de espectrógrafo.

El espectrógrafo fue usado por Lawrence G. Kersta en los Laboratorios Bell alrededor de 1941 para lograr identificar por medio de la voz a las personas. El espectrógrafo generaba un gráfico de la señal de voz teniendo en cuenta la información temporal, de frecuencia y energética a la que se denominó ESPECTROGRAMA o SONOGRAMA.

Lawrence G, empezó partiendo de una hipótesis en la cual expresaba que la voz de cada persona es tan única como la huella dactilar, pudiendo determinar lo que él llamaba HUELLA ACÚSTICA utilizando el análisis espectrográfico.

Al comprender que la voz tenía características propias y únicas para cada persona surgió el pensar en el mecanismo fisiológico que produce esa señal acústica, así como en el proceso de aprendizaje del habla. Kersta sostuvo que la individualidad de la voz se fundamentaba en los anteriores procesos.

<sup>38</sup> Adatado de Molina. *La acústica forense* (2005).

En general, Lawrence G afirma que partes específicas del tracto vocal como cavidades vocales y partes articuladoras son las que determinan la individualidad de la voz. Las cavidades vocales actúan como resonadores y se caracterizan por sus dimensiones y forma de acoplarse, demostrando que es de una probabilidad remota que dos personas tengan propiedades idénticas.

Las partes articuladoras (labios, dientes, músculos mandibulares, lengua, paladar blando, etc.) son controladas dinámicamente por cada persona y permiten que se puedan emitir los sonidos consonánticos y vocálicos. Entonces la probabilidad de hallara dos personas con un patrón dinámico idéntico en el uso de dichas partes articuladores es demasiado remota.

Además, encontramos que los filólogos tienen grandes aportes en el reconocimiento de locutores por la voz. La fonética, sintaxis, morfología, lexicología, dialectología, sociolingüística o la psicolingüística, entre otras, aportan ciertos conocimientos en la acústica forense.

Algunos de los aspectos de informaciones muy útiles que aportan filólogos son:

- La determinación del origen geográfico de un hablante.
- Edad.
- Ámbito social en el que se desenvuelve.
- El descubrimiento de rasgos de su personalidad.
- Hábitos de procedencia patológica o aprendida que lo caractericen de alguna manera.
- Etc.

También los filólogos han intentado aplicar sus conocimientos para resolver el problema de la identificación de la voz. La fonética articulatoria y, posteriormente,

la fonética acústica, han partido del estudio del mecanismo de producción del habla para explicar la naturaleza del lenguaje humano. Para esas ciencias, ha sido también esencial descubrir los átomos del lenguaje denominados rasgos distintivos. Por tanto, estudian lo que hay de común en todos los hablantes.

Muchos filólogos han aplicado sus conocimientos para detectar semejanzas y desemejanzas entre voces de diferentes locutores partiendo de la teoría de la unicidad de la voz para encontrar algunas relaciones, sus estudios suelen seguir la siguiente pauta:

- Determinación y medición de los parámetros acústicos más invariables en un locutor.
- Utilización de la teoría fonética a la hora de realizar un estudio auditivo de las voces.
- Análisis espectrográfico y estudio de los rasgos segmentales y suprasegmentales.
- Utilización de sus conocimientos de sintaxis, semántica, morfología, lexicología, dialectología, sociolingüística o psicolingüística para subrayar características que contribuyan a individuar al locutor.

Aunque la identidad de un locutor está directamente relacionada con las características fisiológicas y comportamiento del locutor (hábitos lingüísticos, entonación de las frases, etc.) variaciones en los tamaños de las cavidades del tracto vocal dan lugar a diferencias en las resonancias.

Variaciones en el tamaño y elasticidad de las cuerdas vocales se traducen en cambios en el valor medio de la frecuencia fundamental de los sonidos sonoros.

Variaciones en el paladar blando y el tamaño de la cavidad nasal producen diferencias espectrales en los sonidos nasales; la presencia de variaciones anatómicas atípicas (disposición y estado de los dientes, estructura del paladar,

etc.), pueden dar lugar a sonidos atípicos o una nasalidad anormal. El problema de la detección de las características presentes en la señal de voz que individualicen al hablante no ha sido un problema fácil.

Los expertos en procesado de señal de habla subrayan que la información inherente al locutor se encuentra completamente imbricada con el resto de las informaciones presentes en la señal de voz, y en cierto modo, enmascaradas por ella.

Si analizamos el proceso de la señal de voz captada por nuestro oído o por un transductor:

- Paso de flujo de aire sobre el tracto vocal del locutor
- Radiación de dicho sonido al exterior
- Propagación acústica hasta el oído o un transductor.
- Si es un transductor el paso de la señal de voz a través de sistemas eléctricos y electrónicos con sus respuestas en frecuencia y fase (amplificadores, filtros, canales telefónicos, conversiones analógico/digitales y digitales/analógicas, procesos de codificación, etc.)

Podemos ver que toda esa cadena de elementos introduce sus propios efectos, claramente perjudiciales para la identificación. Sin embargo, aún con todo lo anterior, los seres humanos somos capaces de identificar a las personas.

Existen factores que no tienen relación con el aparato fonatorio de la persona que afectan a la variabilidad de la voz:



En el primer factor encontramos una variabilidad propia de la señal de voz derivada de factores directamente atribuibles al locutor, ya sean voluntarios o involuntarios denominados factores intrínsecos:

- Edad.
- Estado emocional.
- Estado físico.
- Estar sometido a estrés.
- Velocidad de articulación.
- Tipo de habla -leída, susurrada, conversacional, etc.

En el segundo factor encontramos los intrínsecos de circunstancias externas al locutor:

- Dispositivos de adquisición y transmisión de la señal de voz.
- Ancho de banda.
- Distorsión de canal.
- Reverberación.
- Variabilidad debida al paso del tiempo.
- Ruido aditivo.
- etc.

### 3. METODOLOGÍA

#### 3.1 ENFOQUE DE LA INVESTIGACIÓN

Este proyecto está basado en una investigación de tipo empírico-analítico, debido a que los procedimientos de adquisición de muestras implementados por estas entidades no poseen la debida calidad, por esta razón se realizarán asesorías y mejoras en los procedimientos de captura, procesamiento, análisis de señales de audio, entre otras. Además puede ser enfocada como investigación crítico social por el impacto que este proyecto genera en la sociedad en lo pertinente a seguridad.

#### 3.2 CAMPO TEMÁTICO DEL PROGRAMA

La línea de investigación es Tecnologías actuales y sociedad, está relacionada directamente con este proyecto por la implementación de conocimientos técnicos y científicos de vanguardia que ayuden a la solución de problemas o faciliten los procesos de mejoramiento de la calidad de vida de las personas que pertenecen a un grupo social determinado en el país.

En cuanto a la sublínea corresponde el procesamiento de señales digitales y/o analógicas, ya que se estudiará la teoría y aplicación de detección, filtrado, codificación, reconocimiento, análisis de señales tanto en tiempo como su frecuencia.

El campo de investigación del programa en el que se encuentra el proyecto es: diseño de sistemas de sonido; ya que está enfocado hacia el análisis, procesamiento y control de señales.

#### 3.3 TÉCNICAS DE RECOLECCIÓN DE INFORMACIÓN

Los procesos propuestos para realizar la investigación se basan en el estudio teórico-práctico de propuestas con soluciones claras.

La siguiente encuesta se realizó con el fin de encontrar la viabilidad y aceptación del proyecto, conocer la situación actual del personal profesional frente al tema encuestando entre los empleados de las entidades estatales que velan por la seguridad del país. Esta es una encuesta es por sondeo y tiene preguntas de selección múltiple y preguntas cerradas. Las siguientes fueron las preguntas realizadas.

## ENCUESTA

1. ¿Considera usted que es necesario mejorar los procedimientos implementados en el reconocimiento del locutor?

- a. Si.
- b. No.

Respuesta:

2. ¿Sabe usted los procedimientos básicos para el análisis acústico de la voz?

- a. Si.
- b. No.

Respuesta:

3. ¿Le gustaría a usted tener acceso a esta información a través de asesorías técnicas sobre procesamiento de señales?

- a. Si.
- b. No.

Respuesta:

4. Cree que la mala calidad de un material fonográfico se debe a:

- a. Condiciones ambientales.
- b. Herramientas de grabación no adecuados.
- c. Falta de conocimiento sobre métodos de captura en muestras.

Respuesta:

5. ¿Conoce usted de algún tipo de dispositivo tanto análoga como digital especializada que facilite el análisis y comparación de voces?

- a. Si.
- b. No.

Determine el nombre del sistema.

Respuesta:

6. ¿Considera usted que es importante la implementación de un archivo de muestras de voz a nivel nacional facilitará la identificación del locutor?
- a. Si.
  - b. No.

Respuesta:

7. ¿Sabe usted que procesamiento de señal es utilizado en una muestra de voz para mejorar la inteligibilidad de la misma?
- a. Si.
  - b. No.

Determine cuales son.

Respuesta:

8. ¿Conoce los procedimientos de digitalización de señales?
- a. Si.
  - b. No.

Respuesta:

9. ¿Le gustaría tener un convenio que mantenga a la entidad en constante actualización en avances tecnológicos?
- c. Si.
  - d. No.

Respuesta:

### 3.4 POBLACIÓN Y MUESTRA

La encuesta fue aplicada a la población que labora en las entidades competentes en el área de acústica forense de la ciudad de Bogotá. Los profesionales encargados de esta área corresponden a fonoaudiólogos. Este personal pertenece a jefes de departamento de acústica forense a nivel central y 2 seccionales (a nivel nacional).

### 3.5 HIPÓTESIS

Este proyecto de grado pretende determinar que procedimientos y herramientas de alta tecnología son necesarios para atender de manera ágil y eficaz, los análisis solicitados por las diferentes unidades del DAS, policía y autoridades judiciales del país. Mediante la formulación de una técnica adecuada de registro y

análisis de señal, se pueden minimizar los errores en procesamiento además de mejoras del tiempo en emisión de resultados.

Estas recomendaciones tecnológicas serán soportadas por un protocolo de procedimiento, en consecuencia de las actividades realizadas en las asesorías sobre el reconocimiento del locutor.

### 3.6 VARIABLES

#### 3.6.1 Variables Independientes

- La acústica del recinto actual.
- El sistema de grabación digital multicanal.
- Respuestas tardías de informes periciales.
- Vencimiento de terminos en la parte legal.

#### 3.6.2 Variables dependientes

- La muestras no tienen una optima calidad.
- Material fonograbado con ruido.
- Ingreso de material fonograbado en gran cantidad que no puede ser analizado a tiempo.
- Desaprovechamiento de recusos fisicos.
- Falta de experiencia en el manejo de equipos.

#### **4. PRESENTACIÓN Y ANÁLISIS DE RESULTADOS**

Según las investigaciones realizadas durante los últimos 4 meses se han detectado ciertas falencias en el manejo de señales aplicadas en el área forense especialmente en el reconocimiento del locutor.

Teniendo en cuenta la importancia y la necesidad de obtener resultados con un mínimo porcentaje de error, se pretende brindar a los usuarios potenciales un protocolo en el cual se busca el mejoramiento de los procedimientos actualmente usados en el país sin desmeritar la metodología que desarrolla el personal profesional encargado de este tipo de funciones.

Durante este tiempo de investigación se observó que un gran porcentaje del material dubitado e indubitado fonograbado destinado para el análisis del locutor, presentaba factores que afectan directamente los procedimientos en el análisis de locutor, calidad y por consiguiente la obtención de resultados. Además, una gran cantidad de este material llegaba al personal de manera análoga (cassete, VHS, minicassete, etc.) siendo la digitalización un procedimiento básico para poder ser analizado por los sistemas biométricos, con los cuales cuentan estas entidades.

Las señales utilizadas para uso forense son sometidas a tres procesos:

- Captura de material dubitado e indubitado.
- Procesamiento de la señal.
- Análisis de señales.

##### **4.1 RESULTADOS DE INGRESO DE MATERIAL DUBITADO E INDUBITADO**

En las siguientes tablas se observa el material que ingresó durante el periodo de investigación. Este material fue clasificado y evaluado por parte del equipo de trabajo del área acústica forense, estableciendo el formato de ingreso y el estado de la calidad de la grabación.

La evaluación fue determinada mediante procedimientos audio-perceptivos en una escala de 1 a 5. Es de aclarar que los resultados solo pertenecen a una sola dependencia, es decir a nivel Bogotá.

El periodo de investigación comprende la recopilación de datos en un tiempo de 4 meses, los cuales fueron evaluados semanalmente.

### **Tabla de ingreso del material dubitado**

La tabla 5 muestra la cantidad total de unidades que ingresaron de material dubitado e indubitado en su respectivo mes.

<b>MES</b>	<b>DUBITADO</b>	<b>INDUBITADO</b>
<b>1</b>	12	15
<b>2</b>	12	19
<b>3</b>	10	24
<b>4</b>	9	20

**Tabla 5.** Tabla de ingreso total de material.

A continuación, podemos ver en la tabla si el material dubitado e indubitado que ingresa es análogo o digital.

### **Dubitado**

<b>Semana</b>	<b>análogo</b>	<b>digital</b>
<b>1</b>	3	1
<b>2</b>	3	2
<b>3</b>	3	0
<b>4</b>	3	0
<b>5</b>	2	3
<b>6</b>	1	2
<b>7</b>	4	0
<b>8</b>	5	2
<b>9</b>	5	0
<b>10</b>	2	2
<b>11</b>	5	2
<b>12</b>	7	1
<b>13</b>	1	2
<b>14</b>	5	0
<b>15</b>	1	2

<b>16</b>	7	2
<b>TOTAL</b>	57	21

**Tabla 6.** Tabla de ingreso material dubitado.

### Indubitado

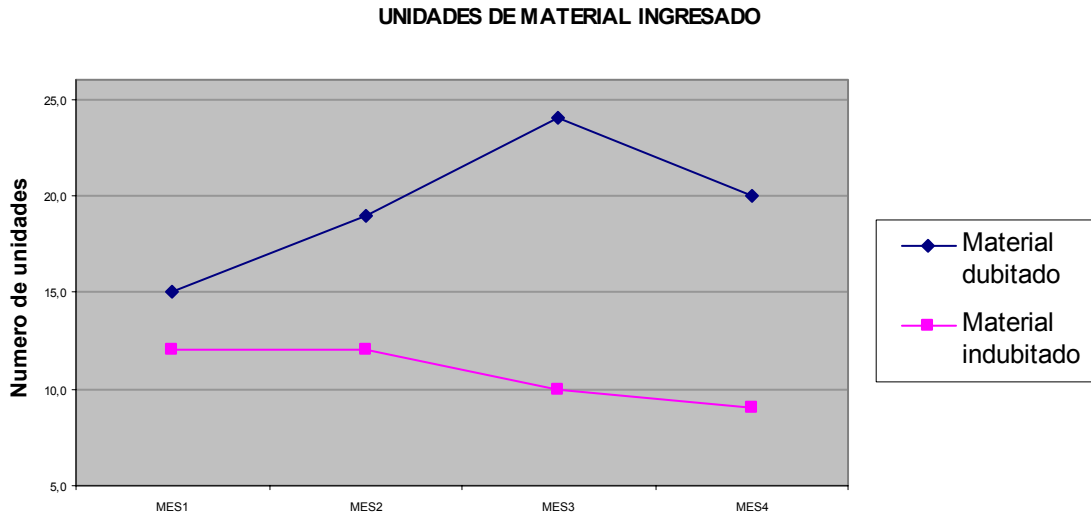
<b>Semana</b>	<b>análogo</b>	<b>digital</b>
<b>1</b>	4	0
<b>2</b>	2	0
<b>3</b>	4	0
<b>4</b>	2	0
<b>5</b>	3	0
<b>6</b>	4	0
<b>7</b>	2	0
<b>8</b>	2	0
<b>9</b>	3	1
<b>10</b>	1	1
<b>11</b>	3	1
<b>12</b>	3	0
<b>13</b>	1	0
<b>14</b>	3	0
<b>15</b>	2	0
<b>16</b>	3	0
<b>TOTAL</b>	42	3

**Tabla 7.** Tabla de ingreso análogo y digital.

Como se observa en los resultados de las tablas 5, 6 y 7, cada mes tiene un ingreso semanal de material análogo y digital dubitado e indubitado. En la grafica siguiente (Figura 18) se aprecia que el material que ingresa con mayor número de unidades es el dubitado.

Esto es debido a que al Departamento Administrativo de Seguridad (DAS) ingresan grabaciones de otros organismos de seguridad y también producto de la implementación de equipos que manejan los profesionales forenses.

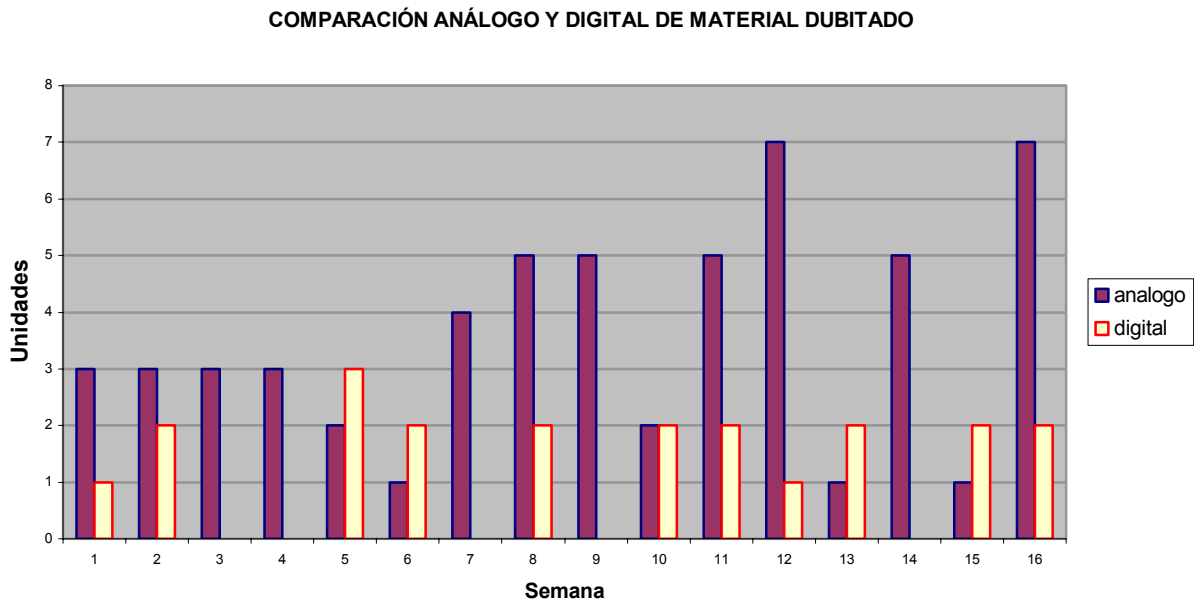




**Figura 18.** Comportamiento de material ingresado en los 4 meses.

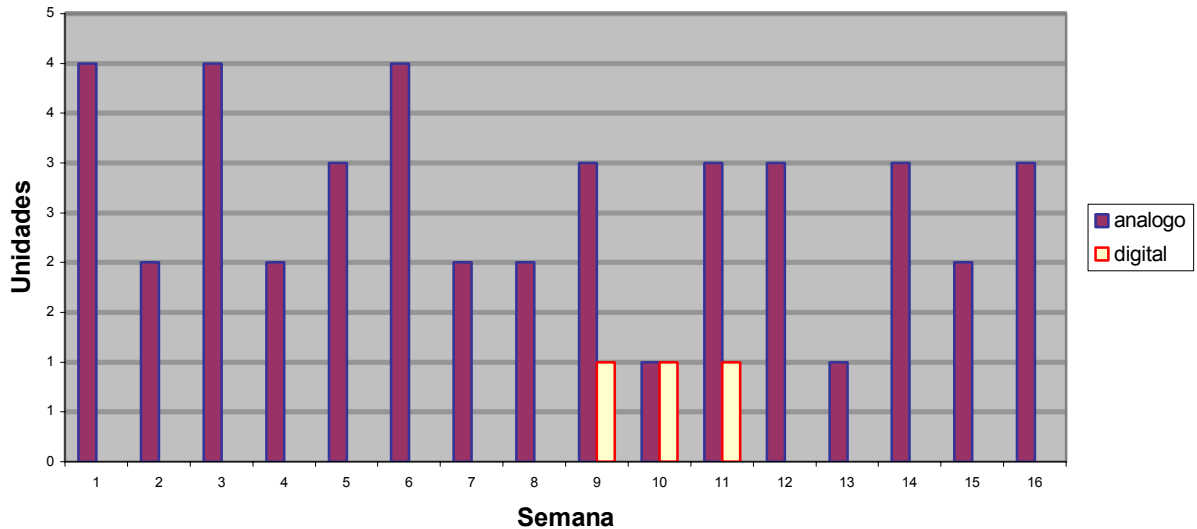
Existe un factor que afecta el número de unidades de material indubitado. Este depende del proceso investigativo que lleve el individuo en cuestión, es decir, cuando se requiera hacer la comparación de la muestra.

En las figuras 19 y 20 se observa la cantidad de grabaciones análogas y digitales que ingresa al DAS en material dubitado e indubitado.



**Figura 19.** Grabaciones de material dubitado (análogo Vs. Digital).

### COMPARACIÓN ANALOGO Y DIGITAL DE MATERIAL INDUBITADO



**Figura 20.** Grabaciones de material indubitado (análogo Vs. Digital).

Según los resultados, encontramos que tanto en el material dubitado e indubitado las grabaciones análogas sobrepasan a las digitales debido a los equipos que implementan los profesionales forenses como los ya mencionados (cassetes y microcassetes en su mayoría).

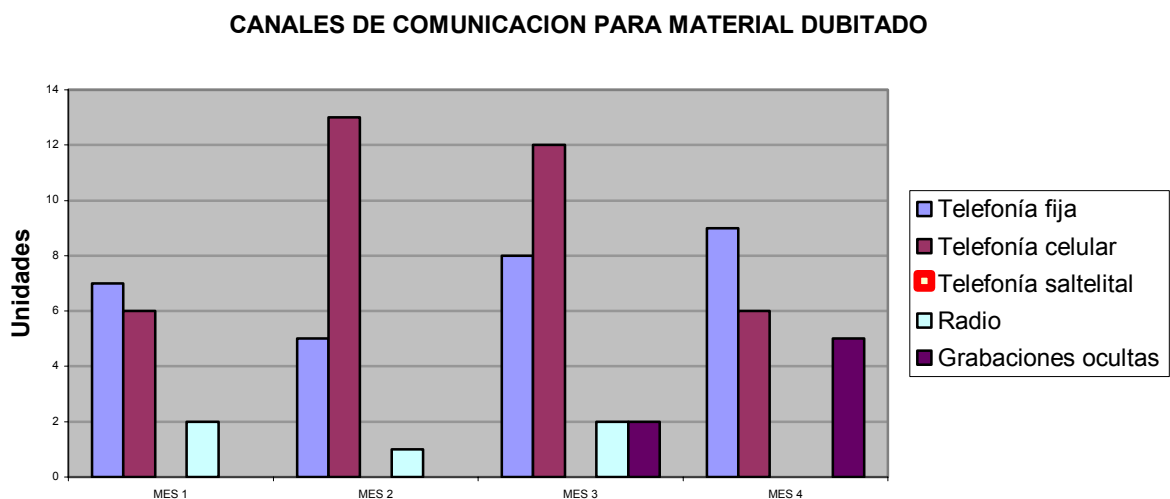
### TABLA DE CANALES DE COMUNICACIÓN

El material que ingresa en el DAS se encuentra definido por los canales de comunicación que usa la delincuencia, en la tabla 8 se muestra la información referente a los canales de comunicación utilizados para la adquisición de las muestras debitadas, es de notar que la información del tipo de canal denominado Telefonía satelital no fue posible tener acceso a ella.

tipo de canal	MES 1	MES 2	MES 3	MES 4	TOTAL
Telefonía Fija	7	5	8	9	29
Telefonía Celular	6	13	12	6	37
Telefonía Satelital	0	0	0	0	0
Radio	2	1	2	0	5
Grabaciones ocultas	0	0	2	5	7

**Tabla 8.** Canales de comunicación.

Los resultados muestran que los canales de comunicación usados en el DAS para obtener el material fonograbado durante el periodo de investigación, se encuentran en su mayoría entre la telefonía fija y al telefonía celular, debido a que gran parte de la delincuencia utiliza este tipo de canal de comunicación. En la figura 21 se muestran los resultados de tipo de canal de comunicación usado por mes.



**Figura 21.** Unidades de material dubitado según canal de comunicación.

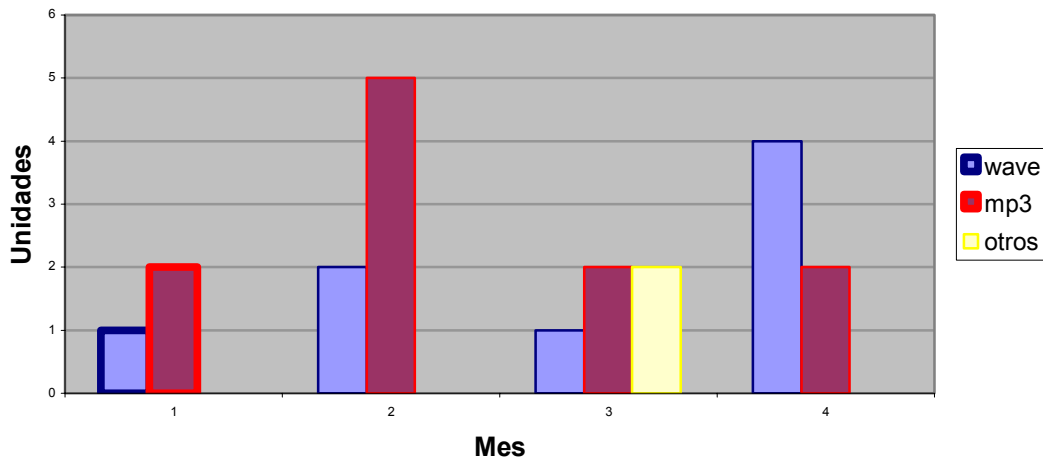
### TABLA DE TIPO DE FORMATO DE MATERIAL DUBITADO

Además de los canales de comunicación empleados, el material dubitado ingresa en ciertos formatos de audio que cambian la calidad de la señal. En la siguiente tabla 9 se observa los tipos de formato de material dubitado que llegan al DAS que por lo general son Wave, Mp3 y otros como por ejemplo Wma, CDA, etc.

Tipo de formato	MES 1	MES 2	MES 3	MES 4
<b>wave</b>	1	2	1	4
<b>mp3</b>	2	5	2	2
<b>otros</b>	0	0	2	0

**Tabla 9.** Unidades de material dubitado según tipo de formato.

## FORMATOS DE AUDIO DEL MATERIAL DUBITADO



**Figura 22.** Formatos de audio material dubitado.

Como se aprecia en la figura 22 el formato Mp3 resulta ser el de mayor uso, debido a lo generalizado que se encuentra y a su ahorro en el tamaño de almacenamiento. Por ejemplo una grabación de 1 hora continúa o más de duración, es más fácil almacenar en \*mp3 o \*wma que en \*wav.

Sin embargo, la utilización de estos formatos no es recomendable para el análisis de locutor, partiendo del punto que los sistemas semiautomáticos no permiten la entrada de estos formatos y se pierden componentes frecuenciales importantes para su estudio.

### **Estado de material Dubitado que ingresa al área de acústica forense**

Los resultados se evaluaron mediante procedimientos audio-perceptivos de la siguiente manera:

Se elaboró una tabla de valores en los cuales la escala de calidad era de 1 a 5 donde:

1. Baja.
2. Media baja.

3. Media.
4. Media alta.
5. Alta.

Para la obtención de los resultados, dentro del grupo de trabajo se escogieron 3 personas las cuales evaluaban la calidad de la grabación con respecto a su inteligibilidad. Después se realiza un promedio aritmético obteniendo los siguientes resultados:

<b>mes 1</b>		
<b>Estado de grabación escala 1-5</b>		
<b>Material dubitado</b>	<b>Análogo</b>	<b>Digital</b>
semana 1	2,5	4,0
semana 2	3,0	3,5
semana 3	2,3	X*
semana 4	2,8	X
<b>mes 2</b>		
<b>Estado de grabación escala 1-5</b>		
<b>Material dubitado</b>	<b>Análogo</b>	<b>Digital</b>
semana 1	2,3	4,0
semana 2	2,0	3,5
semana 3	3,3	X
semana 4	3,4	3,5
<b>mes 3</b>		
<b>Estado de grabación escala 1-5</b>		
<b>Material dubitado</b>	<b>Análogo</b>	<b>Digital</b>
semana 1	3,1	X
semana 2	2,9	4,0
semana 3	2,6	3,5
semana 4	3,4	4,5
<b>mes 4</b>		
<b>Estado de grabación escala 1-5</b>		
<b>Material dubitado</b>	<b>Análogo</b>	<b>Digital.</b>
semana 1	2,3	X
semana 2	3,2	X
semana 3	2,2	4,1
semana 4	2,3	3,5

**Tabla 10.** Estado de la grabación de material dubitado escala 1-5.

\* La X indica la ausencia de material grabado.

Los resultados obtenidos según las escalas demuestran una calidad media baja del material análogo como se puede apreciar en la figura 23, y percatándose que el volumen de material que ingresa para análisis es representativo, inevitablemente la revisión de los procedimientos y equipamientos utilizados para la grabación deben ser reconsiderados.

### Estado de material

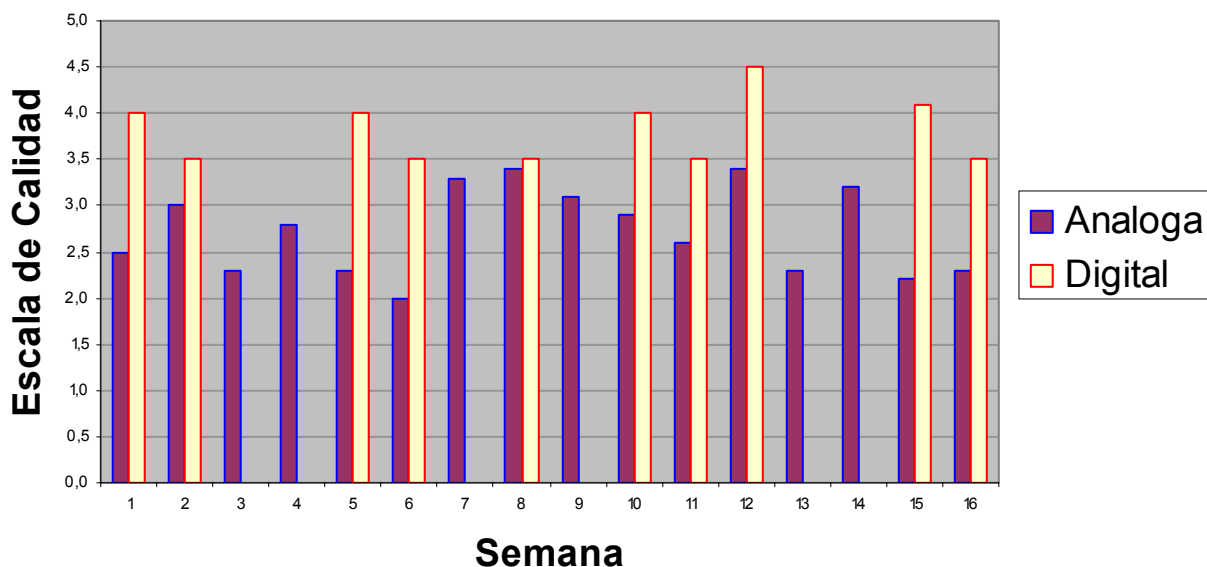


Figura 23. Escala de calidad de material dubitado.

En conclusión, según los resultados obtenidos, gran porcentaje del material sonoro que ingresa al área de acústica forense esta caracterizado por:

- Mayor número de unidades de grabaciones análogas (cassete y microcassete).
- Menor numero de unidades de grabaciones digitales.
- La calidad del material tiene un promedio medio bajo.
- Gran parte de las grabaciones digitales no cumplen con los requerimientos exigidos por los sistemas de análisis de voz semi-automáticos (formatos de audio).

En consecuencia a los problemas que caracterizan al material dubitado e indubitado, la digitalización se convierte en un proceso importante para la obtención de buenos resultados.

Por esta razón, se realizaron diferentes pruebas de grabaciones a un cierto número de locutores, para demostrar que un inadecuado equipamiento y un mal proceso de digitalización pueden sumar factores que deterioran aún más la calidad del material.

#### 4.2 RESULTADOS DE LA ENCUESTA

La encuesta fue respondida por el personal técnico criminalista que labora en el área de acústica forense de los cuales, 4 jefes de departamento del área de acústica forense respondieron la encuesta entregando los siguientes resultados:

PREGUNTA	RESPUESTA		
	A	B	C
1	X		
2	X		
3	X		
4		X	
5	X		
6	X		
7		X	
8		X	
9	X		

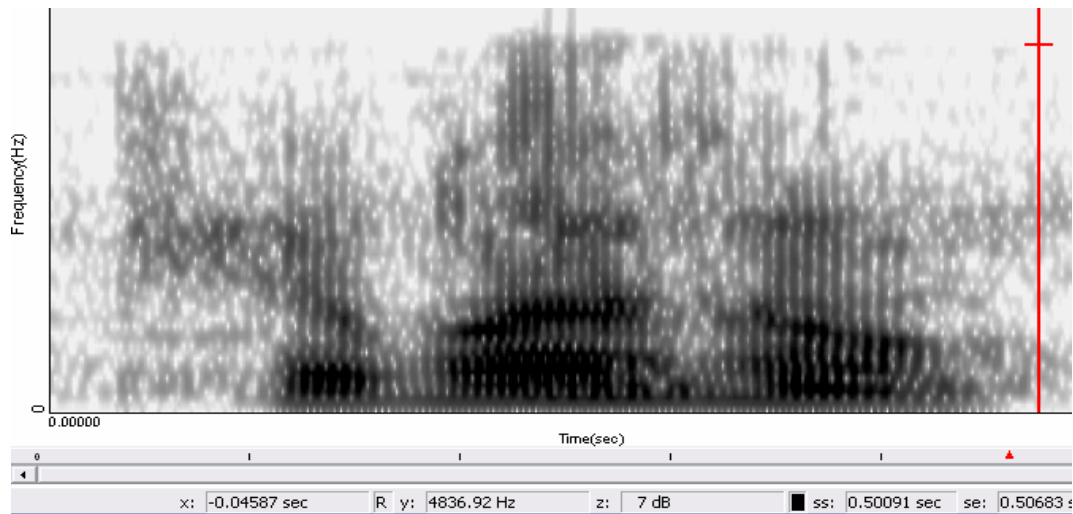
**Tabla 11.** Resultados de la encuesta.

Según los resultados de la encuesta, encontramos deficiencias en el manejo de digitalización y procesamiento de señales, como también el interés y la necesidad de suplir estas falencias por medio de capacitaciones o por medio de un protocolo para la optimización de resultados.





## Espectrograma

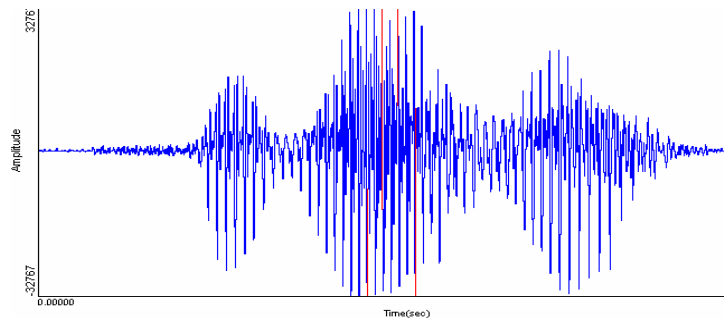


**Figura 25.** Espectrograma de la frase “que paso” con una frecuencia de muestreo correcta.

Como podemos ver (figura 25) en el espectrograma, cuando se cumple el teorema de Nyquist, la señal no sufre pérdidas de información relevante para el análisis de voz, como por ejemplo la apreciación de las formantes y las características originales de la señal.

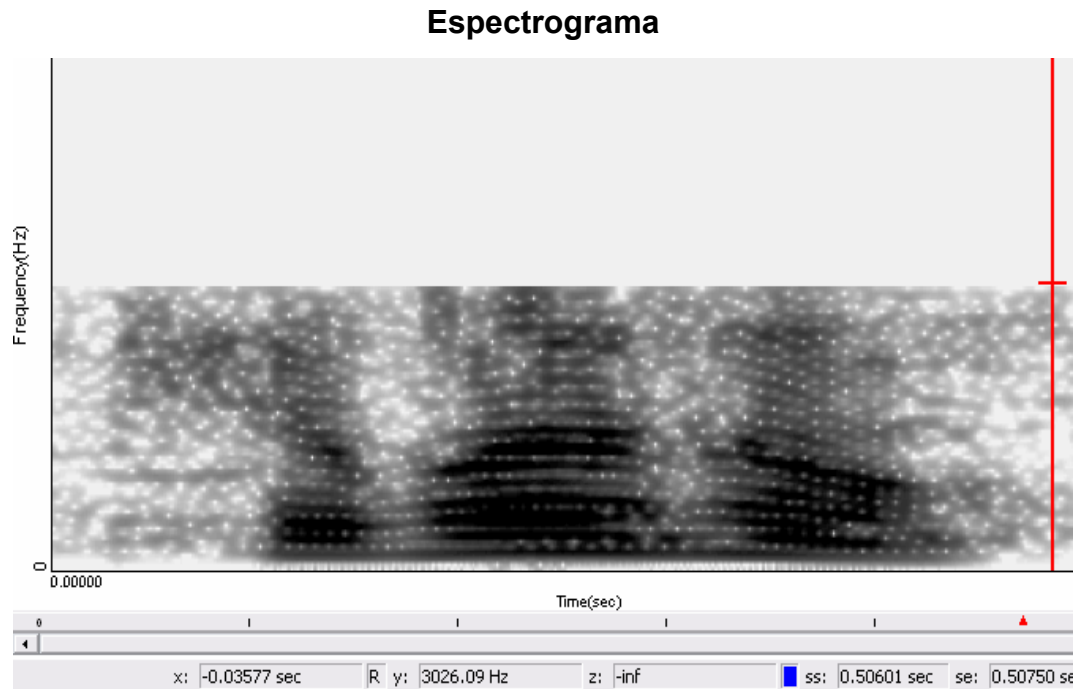
Posteriormente, la misma muestra original es digitalizada con los siguientes parámetros:

<b>Muestra de voz</b>	No Identificado
<b>Frase</b>	Que paso
<b>Formato</b>	Windows PCM
<b>Resolución</b>	16 Bits
<b>Frecuencia de muestreo</b>	6000Hz



**Figura 26.** Forma de onda en el oscilograma de la frase “que paso” grabada incorrectamente.

Es de considerar que los oscilogramas presentan muy pocas diferencias en la representación grafica de la forma de onda.



**Figura 27.** Espectrograma de la frase “que paso” con una frecuencia de muestreo incorrecta.

Ahora, podemos observar lo que ocurre en el espectrograma (figura 27) cuando no es aplicado correctamente el teorema de Nyquist, por lo cual existen modificaciones en comparación con la muestra original.

Teniendo en cuenta que la frecuencia de la muestra original es de 5000Hz, el espectro, obtenido en el segundo análisis, sólo alcanza una frecuencia máxima de 3000Hz eliminando un rango frecuencial de 2000Hz donde puede existir información vital.

## COMPARACIÓN DE ESPECTROGRAMAS

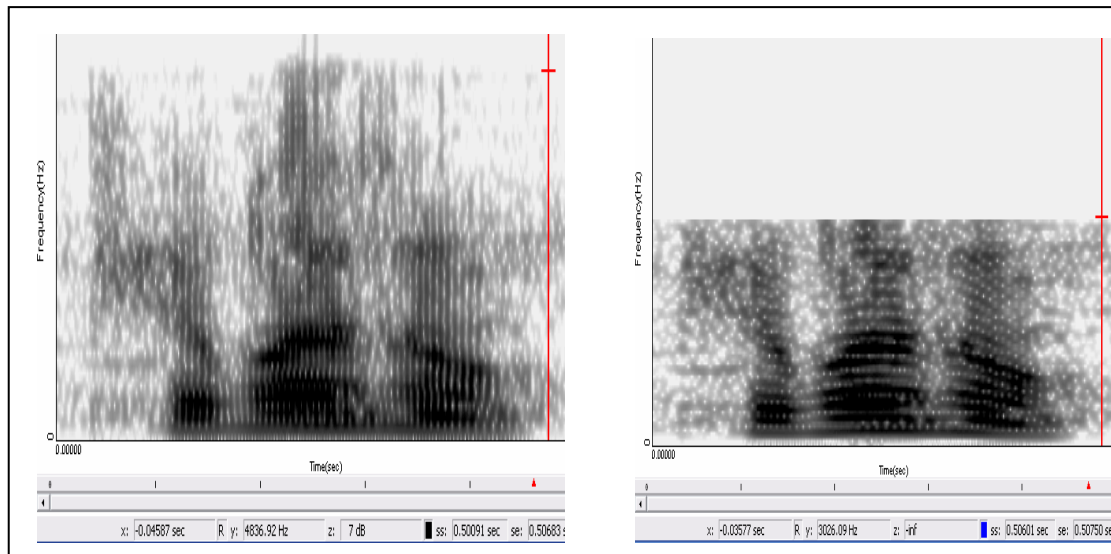
### Adecuado

Muestra de voz	Masculina
Frase	Que paso
Formato	Windows PCM
Resolución	16 Bits
Frecuencia de muestreo	11025Hz

Vs.

### Inadecuado

Muestra de voz	Masculina
Frase	Que paso
Formato	Windows PCM
Resolución	16 Bits
Frecuencia de muestreo	6000Hz



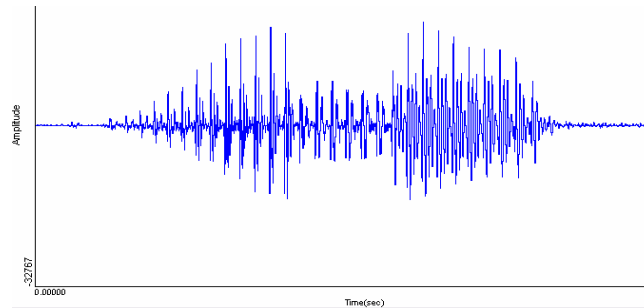
**Figura 28.** Comparación de espectrogramas de la frase “que paso” adecuado e inadecuado respectivamente.

El uso inadecuado de la frecuencia de muestreo (*Teorema de Nyquist*) a la hora de digitalizar trae como consecuencia la pérdida de información frecuencial reflejándose en el espectrograma.

### 4.3.2 ERRORES DE RESOLUCIÓN

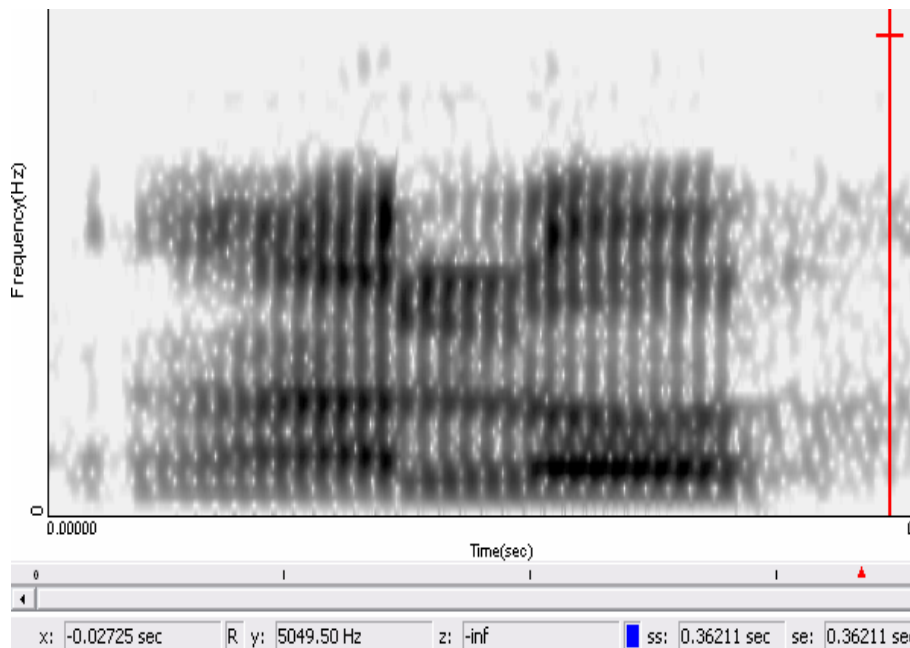
La siguiente prueba se basa en demostrar como afecta la resolución en la calidad de la señal. Se tomó una muestra de voz de un hombre en condiciones normales las cuales son:

<b>Muestra de voz</b>	Jorge Castro
<b>Frase</b>	Aló
<b>Formato</b>	Windows PCM
<b>Resolución</b>	16 Bits
<b>Frecuencia de muestreo</b>	11025Hz



**Figura 29.** Forma de onda en el oscilograma de la palabra "Aló" con una correcta resolución.

### Espectrograma

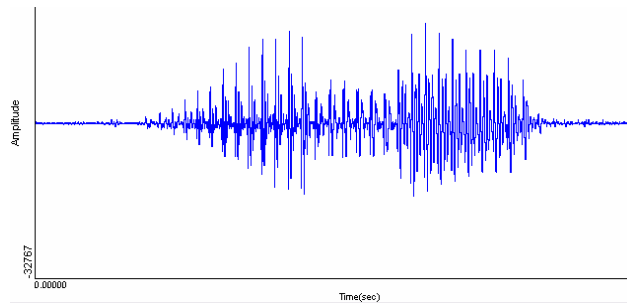


**Figura 30.** Espectrograma de la palabra "Aló" con una resolución correcta.

En el espectrograma (figura 30), vemos que la señal original al ser digitalizada con la resolución correcta no sufre modificaciones visuales, frecuenciales y a nivel de ruido, facilitando el análisis del locutor.

Usando una resolución menor con respecto al ejemplo anterior, la voz fue grabada con los siguientes aspectos:

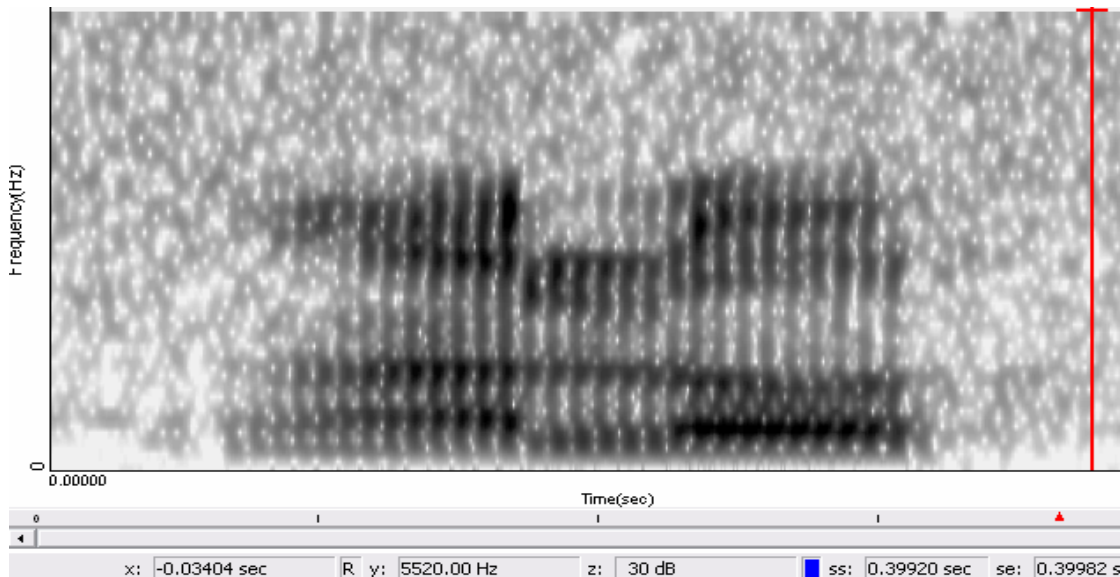
<b>Muestra de voz</b>	Jorge Castro
<b>Frase.</b>	Aló
<b>Formato</b>	Windows PCM
<b>Resolución</b>	8 Bits
<b>Frecuencia de muestreo</b>	11025Hz



**Figura 31.** Forma de onda en el oscilograma de la palabra “Aló” con una incorrecta resolución.

En este caso, se percibe una componente de ruido inexistente en la señal original reflejándose en el espectrograma.

### Espectrograma

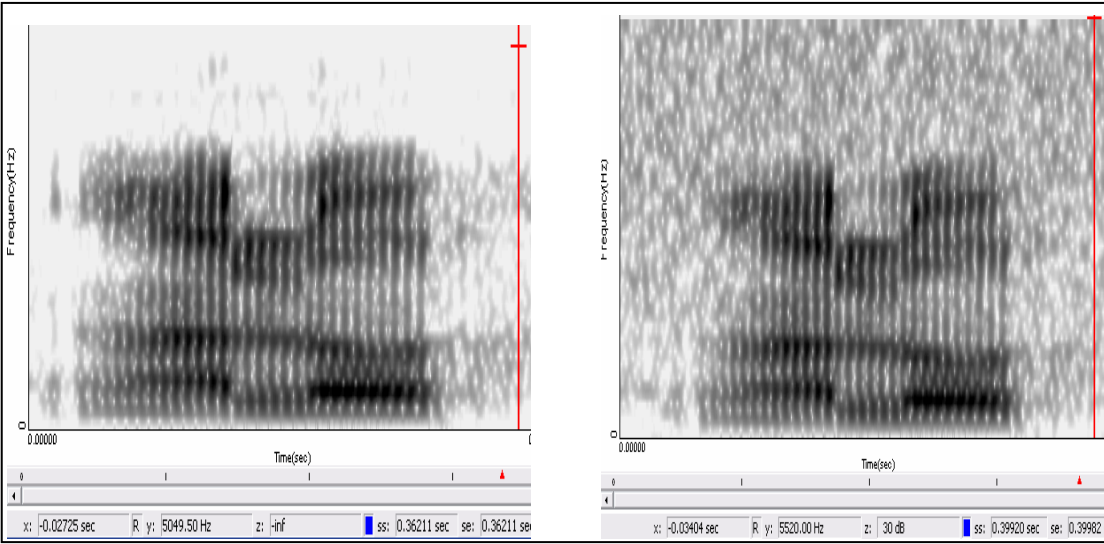


**Figura 32.** Forma de onda en el espectrograma de la palabra “Aló” con una incorrecta resolución.

Como se observa en la figura 32 como el nivel de ruido abarca todo el espectro frecuencial de la muestra representado visualmente como un oscurecimiento comparada con la señal original. Este oscurecimiento esta directamente relacionado con el incremento de ruido en todo el espectro frecuencial de la muestra.

**COMPARACIÓN DE ESPECTROGRAMAS**

Adecuado		Vs.	Inadecuado	
Muestra de voz	Jorge Castro		Muestra de voz	Jorge Castro
Frase.	Aló	Frase.	Aló	
Formato	Windows PCM	Formato	Windows PCM	
Resolución.	16 Bits	Resolución.	8 Bits	
Frecuencia de muestreo	11025Hz	Samplerate	11025Hz	



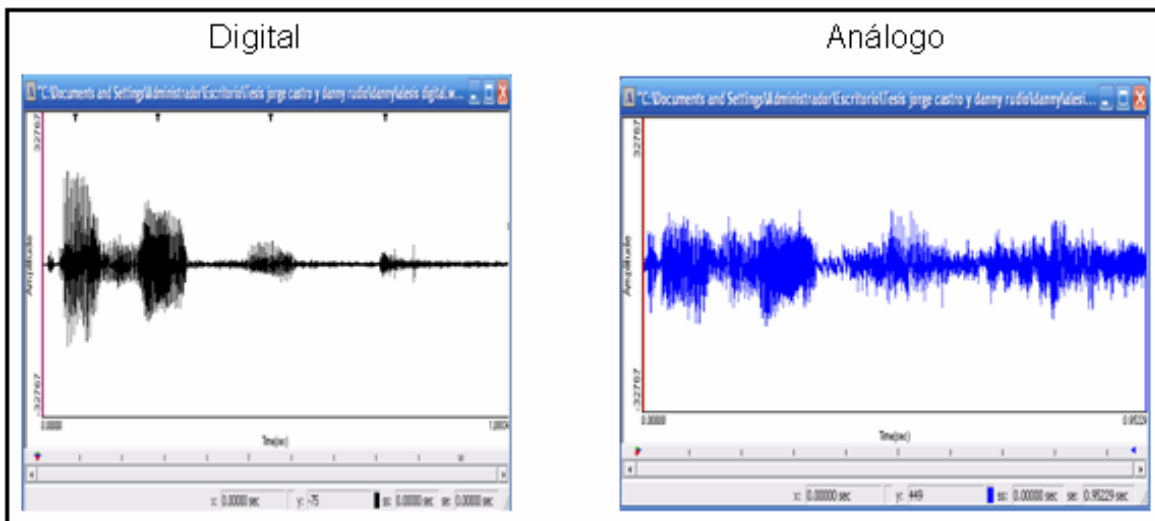
**Figura 33.** Comparación de espectrogramas de la palabra “Aló” con una resolución adecuada e inadecuada respectivamente.

Analizando los espectrogramas (figura 33), se demuestra que la calidad auditiva y visual es menor afectando la información que se encontraba en la muestra original. Además, este deterioro en la calidad del audio causa fatiga auditiva en los analistas por el incremento de ruido.

### 4.3.3 INFLUENCIA DEL RUIDO EN LA INTERPRETACIÓN DE RESULTADOS

La última prueba se realizó con los siguientes parámetros:

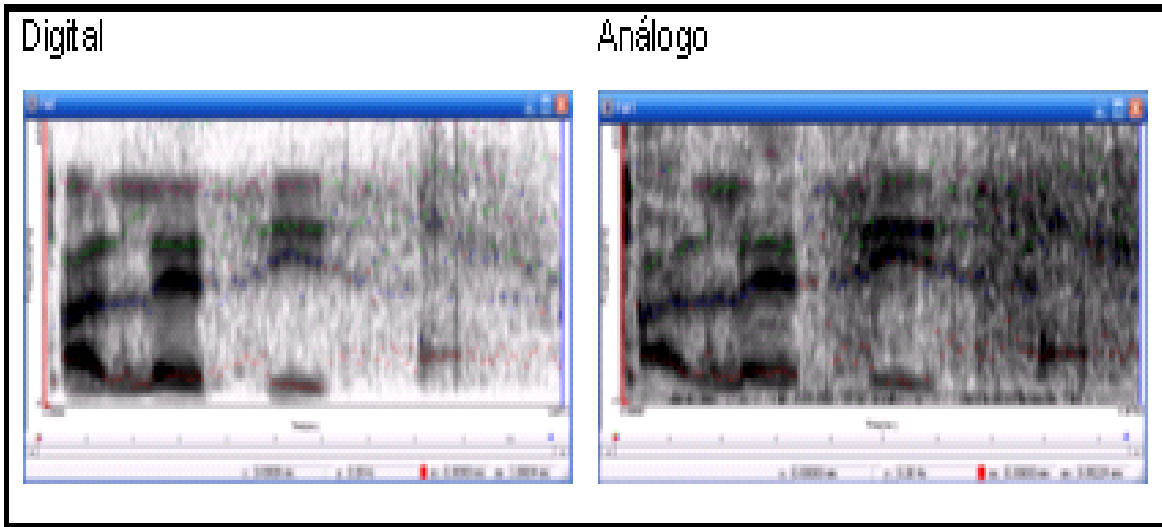
- Se grabó una señal de voz con el mismo texto.
- Una misma persona.
- La captura se llevo a cabo en un recinto acústicamente controlado.
- La grabación se realizó de manera simultáneamente en dos equipos distintos. El primer equipo corresponde a un sistema análogo de muy baja calidad. El segundo fue realizado en el software Smart Logger II.



**Figura 34.** Comparación de oscilograma (Digital vs. Análogo).

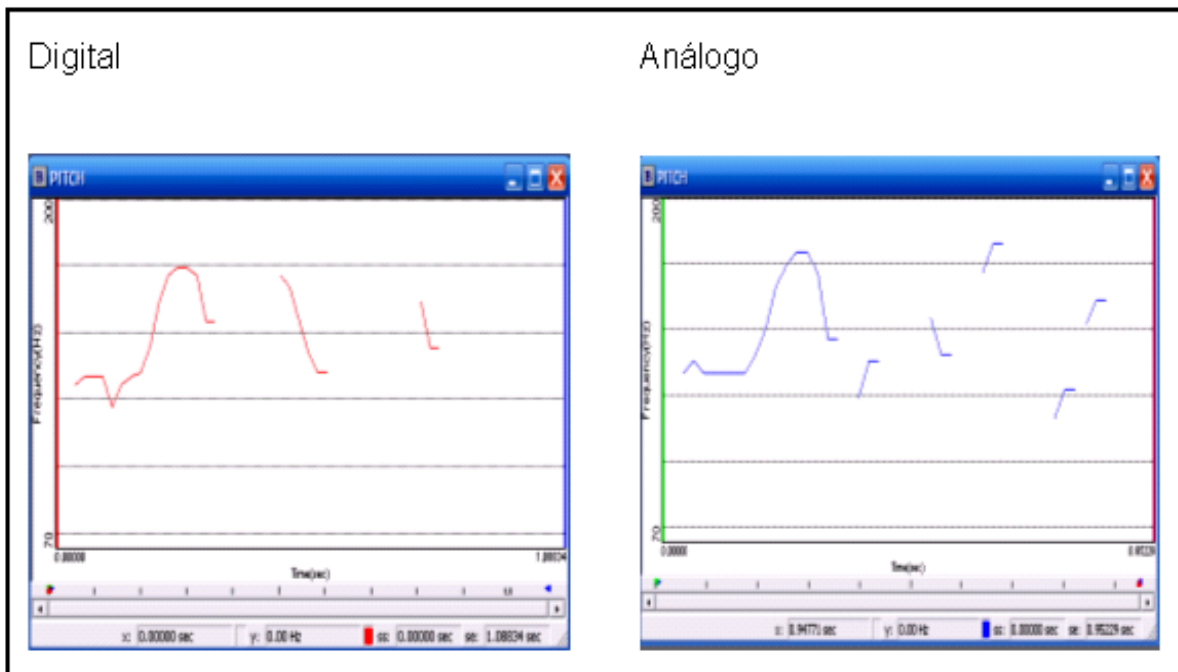
Los resultados demuestran la diferencia visual y sonora de la señal digital comparada con la señal análoga (figura 34), siendo la primera la de mejor calidad, puesto que cuenta con menor ruido, es decir, en este caso la relación señal-ruido es alta.

Esto se refleja en el análisis espectrográfico (figura 35), como se sabe que el ruido en el espectro se expresa como un oscurecimiento, el margen de error se incrementa en el dictamen de resultados.



**Figura 35.** Comparación de espectrogramas (Digital Vs Análogo).

Dentro del proceso de análisis, (figura 36) el tono posee una diferencia entre las dos muestras de más o menos 6Hz estando, 1Hz por encima del valor aceptado por expertos en la individualización del locutor en el ámbito forense.



**Figura 36.** Análisis del Rango del tono de la señal digital Vs. señal Análoga.

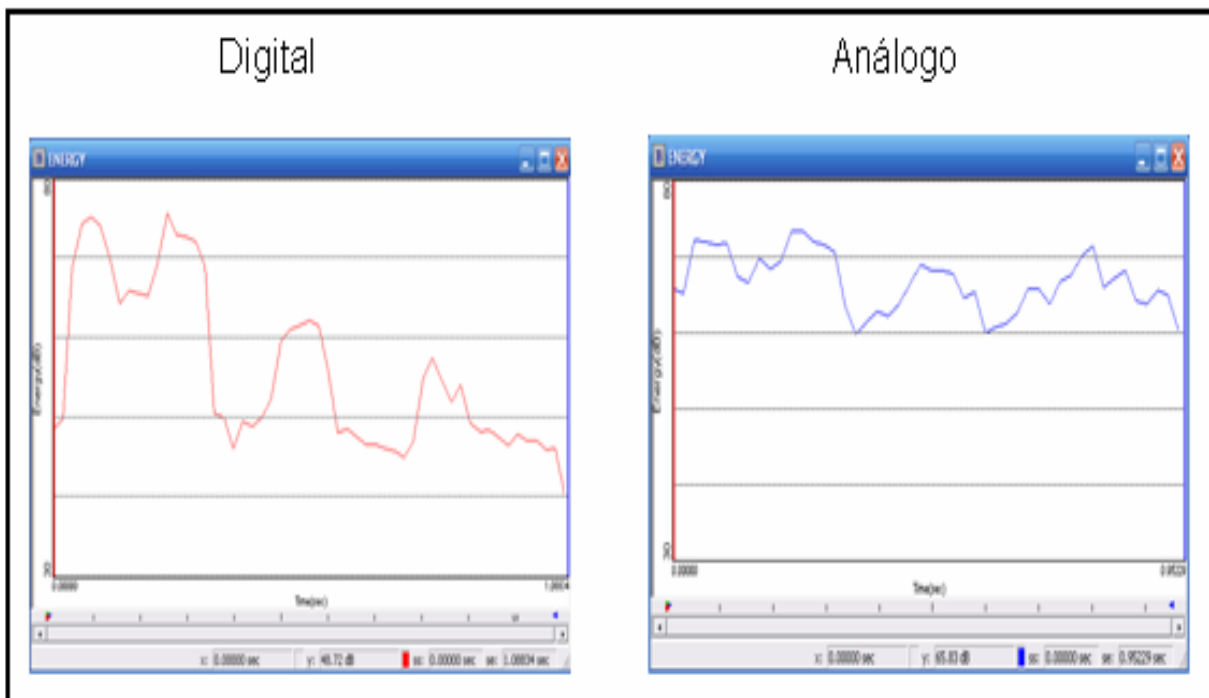


Los datos siguientes (Tabla 12) muestran los datos de tono del locutor apreciando poca variación entre las dos muestras

Digital		Análogo	
Minimum Pitch (Hz)	123.00	Minimum Pitch (Hz)	117.00
Maximum Pitch (Hz)	175.00	Maximum Pitch (Hz)	181.00

**Tabla 12.** Comparación de Tonos (Digital Vs Análogo).

Además, si queremos determinar el nivel de presión sonora del locutor a la hora de expresarse, los resultados son contradictorios (figura 37) al cotejar la información, ya que el A.D.S.R (attack, decay, sustain, release) no se aprecia con claridad.



**Figura 37.** Comparación de Energía (Digital Vs Análogo).

Si apreciamos los siguientes valores (tabla 13) comprobamos que el SPL aumenta significativamente en consecuencia del ruido que se encuentra en la grabación análoga.

Digital		Análogo	
Minimum Energy (dB)	40.12	Minimum Energy (dB)	59.98
Maximum Energy (dB)	75.81	Maximum Energy (dB)	73.51

**Tabla 13.** Comparación Energética (Digital Vs Análogo).

Este valor de ruido se puede obtener midiendo la relación señal ruido de la grabación, pero se debe recordar que la intención es mejorar los procedimientos y herramientas para la obtención de resultados ágiles y confiables.

Por esta razón, es necesario la utilización de sistemas digitales de grabación multi-canal, destinados a este propósito; sin embargo, esto no quiere decir que no seguirán llegando muestras de baja calidad, como muestras digitales en formato diferentes a los sugeridos.

Para concluir, dentro del reconocimiento de locutores, el análisis elaborado por los fonetistas y lingüistas dependerá de la calidad de la muestra. Por tanto, las técnicas, equipos y recurso humano deben cumplir con normas y estándares de calidad.

La ingeniería de sonido juega un papel muy importante en el reconocimiento de locutores y en la acústica forense, debido al aporte científico en la búsqueda de soluciones que brinden mejoramiento en la calidad de los resultados.

## **5. DESARROLLO INGENIERIL**

Después de observar y estudiar los resultados obtenidos durante el periodo de investigación, se determina que en consecuencia del mal estado del material dubitado e indubitado, es indispensable plantear soluciones ingenieriles que mejoren la obtención y calidad en los resultados.

Con este fin, se desarrolla un protocolo el cual plantea una metodología en técnicas de grabación, digitalización, procesamiento de la señal y análisis de la voz asistido por software especializado. .

Además, normalizar los procesos entre las diversas entidades competentes es vital para el desarrollo de la ciencia en el campo de la acústica forense. Por esta razón, este protocolo puede facilitar un trabajo en conjunto entre estas entidades.

### **5.1 PROTOCOLO**

#### **5.1.1 MÉTODOS DE GRABACIÓN PARA LA TOMA DE MUESTRA DE VOZ**

Las siguientes son las pautas para una adecuada captura de la señal de audio para material dubitado e indubitado:

##### **5.1.1.1 MATERIAL DUBITADO**

###### **GRABACIÓN MULTICANAL DE CANALES DE COMUNICACIÓN**

Para realizar grabaciones multicanales, se debe tener en cuenta el canal de comunicación el cual se desea grabar, debido a que cada uno de ellos cuenta con características propias en el manejo de la señal.

Estos canales pueden ser:

- Telefonía fija.
- Telefonía celular.
- Telefonía Satelital.
- Radio.
- Grabaciones ocultas (grabadoras de bolsillo)

Las áreas encargadas en el monitoreo de estas señales deben contar con equipos de grabación multicanal preferiblemente digital que cumpla con los siguientes requerimientos:

- El formato de grabación tiene que ser compatible con sistemas automático y semi-automáticos de análisis de voz
- El tiempo de grabación no debe ser limitado por la capacidad del PC.
- Es necesario contar con un sistema estable a pesar de la carga de trabajo, por ejemplo jornadas de grabación las 24 horas del día.
- Debe contar con opciones de grabación automática con parámetros preestablecidos.
- La información obtenida no puede ser modificada (encriptación de datos).

Además, estos equipos de grabación multicanal deben cumplir con ciertas especificaciones técnicas para su óptimo funcionamiento.

#### **5.1.1.1.2 EQUIPAMIENTO MATERIAL DUBITADO**

El software y hardware **Smart Logger II** es un sistema de monitoreo y registro de voz multicanal diseñado para grabación simultanea de datos de audio los cuales pueden ser procesados y archivados. Los datos de audio pueden ser obtenidos de líneas telefónicas, micrófonos y radios, dispositivos de reproducción de audio, etc.

Las especificaciones e información técnica se encuentran en el manual del usuario. A continuación se muestra dicha información de los requerimientos básicos para su funcionamiento.

## INFORMACIÓN TÉCNICA

### Formato de grabación

- Los datos de audio pueden ser grabados en cualquier formato, siempre que un codec de audio apropiado este instalado en su sistema.

### Formato de base de datos

- MS SQL Server 2000
- MSDE-Microsoft SQL Server Desktop Engine
- El número de grabaciones en una base de datos es solo limitado por el espacio dado para almacenar datos de audio.

### Medios extraíbles de archivo soportado

- CD-R, CD-RW
- DVD+/-R, DVD+/-RW

### Arquitectura de la red de trabajo

- Smart Logger II opera en ambientes TCP/IP y/o HTTP.
- El sistema de la red de trabajo soporta hasta 500 canales de grabación.

### Parámetros eléctricos

- Los parámetros eléctricos de los circuitos de entrada y salida están definidos por los parámetros eléctricos de las tarjetas I/E de sonido y dispositivos utilizados en el sistema. Estos parámetros están listados en los documentos técnicos respectivos para tarjetas I/E de sonido. El número de tarjetas I/E de sonido y dispositivos instalados en un computador sencillo es limitado por el número de puertos PCI y USB disponibles.

## Conexiones de línea telefónica

- Smart Logger trabaja con señales típicas en las extensiones de teléfono conectadas en modo paralelo. Mientras se graban los datos de audio de una línea telefónica, el programa genera señales de tono típicas de 0.4 segundos de largo a 425 KHz cada 15 segundos para advertir a los locutores que su conversación esta siendo grabada.

## REQUERIMIENTOS DEL SISTEMA

### Configuración de hardware requerida

- Computador IBM compatible con Intel Pentium-II 800MHz o CPU similar 10 GB HDD.
- 256 MB RAM.
- Unidad DVD-RW o CD-RW.
- Tarjeta de sonido.

### Configuración de hardware recomendada

- Computador IBM compatible con Intel Pentium-IV HT 3.6 GHz CPU 10 GB HDD.
- 1 GB RAM.
- Unidad DVD-RW o CD-RW.
- Tarjeta de sonido.

### Configuración de hardware mínima para grabación de estación de 30 hasta 60 canales.

- Intel Pentium-III 600M CPU.
- 256 GB RAM.

### Configuración de hardware recomendada para grabación de estación de 30 hasta 60 canales.

- Intel Pentium-IV 2 GHz o Athlon XP 2000 + CPU.

- 512 GB RAM.

Configuración de hardware recomendada para grabación de estación mayor a 60 canales.

- Intel Pentium-IV 3 GHz o Athlon XP 3000 + CPU.
- 1 GB RAM.

Tipos de sistemas operativos soportados

- Windows 2000 Professional
- Windows 2000 Server
- Windows XP Home Edition
- Windows XP Professional
- Windows Server 2003 Standard Edition

## DESCRIPCIÓN DE HARDWARE Y SOFTWARE

### *Smart Logger II Hardware*

El programa puede incluir las siguientes tarjetas I/E de sonido y dispositivos:

- STC-H197, PCI proporciona conexión a 4 extensiones de teléfonos análogos, micrófonos o dispositivos de reproducción con salidas lineales.
- STC-H197D, PCI proporciona conexión a 4 micrófonos o dispositivos de reproducción con salidas lineales.
- STC-H205, PCI proporciona conexión a 8 extensiones de teléfonos análogos, micrófonos o dispositivos de reproducción con salidas lineales.
- STC-H205/4, PCI proporciona conexión a 4 extensiones de teléfonos análogos, micrófonos o dispositivos de reproducción con salidas lineales.
- STC-H199, PCI proporciona conexión a extensiones de teléfonos digitales (ambas 2 y 4 líneas alámbricas) y cajuelas E1.
- STC-H219, USB proporciona conexión a 4 extensiones de teléfonos análogos, micrófonos o dispositivos de reproducción con salidas lineales.

- STC-H219D, USB proporciona conexión a 4 extensiones de teléfonos análogos, micrófonos o dispositivos de reproducción con salidas lineales.

### *Smart Logger II Software*

El software incluye los siguientes componentes:

- Modulo de grabación (sb.exe).
- Archivador (ArchiverService.exe).
- Interfase de usuario (GUI) para operaciones de base de datos (SmartLogger2.exe).
- Reproductor (Saplayer.exe).
- Transcriptor (Trans.exe).
- Controladores de dispositivos I/E de sonido.
- Software del servidor de base de datos.

## **GRABACIONES OCULTAS**

A la hora de realizar una grabación oculta es necesario contar con los siguientes equipos que cumplen con los requerimientos de alta fidelidad.

### **GNOME 2M**

La micrograbadora digital profesional *Gnome 2M* desarrollada por Speech Technology Center es una grabadora ganadora del diploma y medalla del primer lugar en el foro Internacional *Seguridad y tecnología segura, 2003* por su excelente calidad y características.

Entre estas características encontramos su modo mono o estéreo en la grabación de voz, diseñada para grabaciones de voz en-línea (espontánea, no-preparada) en condiciones acústicas pobres, en sitios interiores o externos o de máximo ruido. Además, emplea micrófonos externos e internos incorporados, almacena las grabaciones registradas en la tarjeta compacta de memoria Flash incluida.



Además posee características como:

- Empieza o detiene la grabación de forma manual o automática de acuerdo al temporizador o nivel de ganancia (Modo de Activación de Voz).
- Emplea una compresión de datos para aumentar la duración total de la grabación.
- El dispositivo se controla por medio del teclado, switch externos, la pantalla LCD y los auriculares.
- El reloj incorporado determina la fecha y la hora para empezar y acabar automáticamente cualquier grabación.
- Todos los datos son protegidos simultáneamente. Esta función asegura la autenticidad de los datos grabados y ofrece una protección para el acceso no autorizado.
- Reproduce cualquier grabación protegida directamente usando un PIN (código de restricción de 8 dígitos), o con el software especial incluido.
- Estructura de metal sólida y delgada con 2mm de espesor asegura la durabilidad del dispositivo, así como la protección contra detectores o bloqueadores de sistemas de grabación, permitiendo ocultar la grabadora y disminuir la posibilidad de ser descubierto por equipo especializado.

## ESPECIFICACIONES TÉCNICAS DE LA GNOME 2M

Rango de grabación		6, 8, 10, 16,32, 48 Khz.
Formato de grabación		PCM 16 bit
Tipo de Compresión		μ-law, A-law
THD (En canal directo a 1Khz)		No inferior a 0.04%
SNR (En canal directo a 1Khz)		No inferior a 72 dB
El rango dinámico (con respecto al control de ganancia)		90 dB
Micrófono	Tipo	Electret contra radiaciones HF
	Sensibilidad	-55 dB
	Ruido bajo	25 dB
Tipo de tarjeta de memoria intercambiable		Flash Compacta I(CF) FAT 32

Fuente de Alimentación	Autónomo	2xLR03 (1,5V AAA Alkaline)
	Externo	3V (al conductor central del enchufe)
	Consumo de energía (En modo grabación)	No mayor a 0.4W
Duración máxima de la grabación (sin cambiar la fuente de alimentación)		Aprox. 5 horas (usando Duracell Plus o batería Ultra Alcalinas)
Tamaño		115×55×15 mm
Peso		130g

**Tabla 14.** Manual de usuario Gnome 2M elaborado por Speech Technology Center.

## **GNOME 2P**

La *Gnome 2P* es una grabadora de voz digital profesional portátil que permite grabaciones de alta calidad mono y estereo de señales de voz usando micrófono externos o internos. Esta grabadora es de menor tamaño que la *Gnome 2M* por lo que facilita su camuflaje, pasando más desapercibida.

Otras características:

- Los sonidos al ser grabados son almacenados en una memoria Flash.
- Además, a la hora de grabar en modo mono esta grabadora permite la compresión de los datos de audio para aumentar el tiempo de grabación.
- Las grabaciones pueden empezar o detenerse de forma manual o automática de acuerdo al temporizador o nivel de ganancia (Modo de Activación de Voz).
- La información grabada puede ser transferida a un PC por medio de puerto USB. Esta información puede ser reproducida y modificada por medio de un software especializado.

## ESPECIFICACIONES TÉCNICAS

- Numero de canales 1 o 2.
- Memoria de 1 GB, expandible hasta 2 GB.
- Formato de grabación mono/estereo PCM 16 bit o 2x comprimido.
- Sampling rate 8 ò 16 kHz
- Pasa banda flatness 3 dB o menos.
- SNR para entrada de micrófono menor que 72 dB
- SNR para entradas auxiliares menor que 78 dB

Distorsión de armónicos para entrada de micrófonos (1 kHz at 16 kHz FR).

- Sin compression 0, 1% o menos.
- Con Compression 1, 0% o menos.

Distorsión para entrada auxiliar (1 kHz at 16 kHz FR).

- Sin compresión 0, 01% o menos.
- Con compresión 1, 0% o menos.
  
- AGC depth - up to 12 dB.
- Nivel de modo de activación de voz (VOX mode threshold level) 1...60 dB.

Tipo de micrófono:

- Electret, HF resistant.
- Sensibilidad 55 dB.
- Nivel de sonido interno 28 dB o menos.

Seguridad:

- Código de seguridad de 8 dígitos contra borrado, edición o modificación de la data grabada.
- Etiqueta de seguridad con firma digital contra borrado, edición o modificación de la data grabada.

#### Ajustes de parámetros:

- Etiqueta de fecha y tiempo en cada archivo de grabación.
- Ajuste parámetros y submenús de grabación, fecha / tiempo, vox, temporizadores y loop directamente desde la PC
- Modos grabación manual Star / Stop, por VOX y por 5 temporizadores programables

#### Tiempo de grabación con operación continua sin alimentación externa:

- Modo grabación 24 horas.
- Modo grabación en VOX 250 horas.
- Modo Stand by 1500 horas.
- Máxima duración en grabación (mono, con sampling rate de 8 kHz y compresión), con 1 GB de memoria 37 horas.
- Duración de grabación garantizada (estereo, FR=16kHz, descomprimida) a 1 GB memoria - 9 horas.

#### Alimentación:

- Batería recargable de Li-Ion, 900 mAh para trabajo 24 horas continuas.
- Alimentación externa 4.5V con adaptador o por puerto PC USB.
- Tiempo de carga cerca de 4 horas.

#### Transmisión de datos:

- PC interfase USB 2.0 Hi.
- Tamaño: 73x47x12.5 mm, caja metálica (B95).
- Peso: 90 gramos.

## **5.1.1.2 MATERIAL INDUBITADO**

### **GRABACIÓN TOMA DE MUESTRA DE VOZ**

Debido que el material indubitado se adquiere directamente del individuo es necesaria que esta grabación sea realizada con el equipamiento necesario y con las condiciones acústicas ideales para la sala.

#### **5.1.1.2.1 CONDICIONES ACÚSTICAS PARA LA SALA**

Tratar acústicamente la sala es de vital importancia para no afectar la muestra con factores externos, tales como tiempo de reverberación, flutters, modos de resonancia, ruido de fondo, etc.

Para el tratamiento acústico de la sala es necesario considerar:

- Insonorización y aislamiento acústico con el exterior
- Acondicionamiento acústico interior

Las salas de grabación de voz generalmente son de dimensiones estructurales pequeñas y el comportamiento del sonido dentro de ellas esta definido por las mismas.

Para el tratamiento acústico es necesario tener en cuenta:

- Prevenir ondas estacionarios e interferencias acústicas que afecten la respuesta en frecuencia del cuarto de la sala de grabación.
- Reducir modos de resonancia y disminuir el tiempo de reverberación de la sala.
- La absorción o difusión del sonido en el cuarto para evitar ecos y reflexiones mejorando la proyección de imagen estereo.
- Pérdidas por transmisión.

Debido a que la información más significativa de la voz son las consonantes, que a su vez son más débiles y cortas que las vocales, por tanto con un tiempo de reverberación alto las vocales se prolongarían demasiado, enmascarando a las consonantes que le siguen, lo que produce una disminución en la inteligibilidad de la palabra.

## **ESPECIFICACIONES ACÚSTICAS DE LA SALA**

- Las superficies internas de la sala no deben ser paralelas entre si.
- La respuesta en frecuencia del recinto debe ser plana.
- Ruido de ambiente menor a 50dB.
- Tiempo de reverberación entre 0.3 y 0.5 seg.
- La claridad  $C_{50}$  para la palabra comprende valores mayores a + 2 dB.

### **5.1.1.2.2 EQUIPAMIENTO PARA SALA**

La sala debe contar con un software para la captura y análisis de la señal de voz que cumplan con los requerimientos necesarios para uso forense.

## **SOFTWARE Y HARDWARE**

### **SIS versión 6.X**

#### **(Speech Interactive System) Sistema Interactivo De La Voz**

El *SIS* es un software para edición, análisis, reducción de ruido e identificación del locutor desarrollado por Speech Technology Center que permite mejorar la inteligibilidad de la voz, refuerza la calidad de señal y reduce el ruido en la voz grabada. El programa proporciona gran manejo en entrada / salida de señal, análisis, edición, presentación y procesamiento de voz y otras señales de frecuencia baja.

El sistema contiene una hardware (interfase STC-H216 ò STC-H246) de alta precisión para entrada y salida señales de voz, permitiendo la digitación de señales análogas y su copiado al disco duro de un computador sin ninguna distorsión y con todas las propiedades de la señal relevantes para el examen forense. Otras de las capacidades proporcionadas por el hardware incluyen conversión de la señal digital a análoga, reproducción de señales procesadas con ajustes de parámetros ADC y DAC deseados.

Este software esta diseñado para expertos en procesamiento de señal de baja frecuencia, lingüística y el audio forense.

#### UTILIDADES DEL SISTEMA

El software *S/S* permite:

- Conversión de señal análoga-a-digital, Entrada / salida al computador vía tarjeta de sonido de 16-bit ò 24-bit.
- Grabación en tiempo real de entrada de señal de voz a un disco duro o RAM. El espacio de dirección virtual es de 2000 Mb.
- Control de calidad de grabación de sonido y detección de tipos de distorsión e interferencia en la señal.

Edición de señal de voz digitalizada:

- Ajuste de marcas delimitando los componentes de señal.
- Modificar la amplitud y duración de los componentes de señal.
- Normalización de señal.
- Compilar muestras de prueba de varios fragmentos de entrada de señales de voz.
- Remover componentes de ruido.
- Ajuste de pausas a cero.
- Reproducción de muestras de voz de prueba generadas para establecer los

contenidos de una grabación de sonido

Reducción de ruido mediante:

- Filtración adaptable de ruido armónico de banda-angosta cuasi-estacionaria en amplitud y fase.
- Filtración adaptable de ruido armónico de banda-angosta, cuasi-estacionaria solamente en su amplitud.
- Filtración adaptable de ruido de banda-ancha.
- Supresión de impulso de ruido.

Análisis de fragmentos de señal seleccionados con estimación y presentación de los siguientes parámetros:

- Duración de componentes de señal.
- Amplitud de componentes de señal.
- Amplitud instantánea de espectro Fourier de un fragmento seleccionado.
- Amplitud promedio de espectro Fourier de un fragmento seleccionado.
- Función fundamental de frecuencia para segmento / fragmentos seleccionados.

Características dinámicas de un componente de señal con representación cuasi-tridimensional de tipo “voz visible” en los ejes de tiempo/número del coeficiente / amplitud parametrizado para:

- Coeficientes de amplitud de espectro de poder.
- Cepstral real.
- Función de autocorrelación.
- Coeficiente de predicción lineal.
- Coeficiente de correlación parcial.
- Respuesta de frecuencia a LPC.
- Promedio dinámico de características de la señal.



- Características estadísticas de curvas de tono.

Comparación de características paramétricas medidas de señales de voz y adquisición de medidas similares para diversos ejemplos, comparando:

- Promedio espectral de muestras de señal.
- Características estadísticas de curvas de tono.
- Espectro instantáneo de fragmentos de señal de referencia seleccionados.
- Promedio de características de Cepstral, coeficientes de autocorrelación, predicción lineal, correlación parcial y de estructuras de formato a puntos de referencia en grabaciones de voz.

Presentación de señales con adquisición subsiguiente de imágenes en pantalla o copia dura de:

- Formas de onda.
- Espectrogramas visibles (voz visible).
- Cepstrograma dinámico (funciones de periodicidad de señal).
- Autocorrelogramas dinámicos.
- Contornos de tono.
- Trayectorias de formato.

Las representaciones obtenidas pueden ser subsecuentemente usadas para análisis visual subjetivo tan bien como para:

- Tomar decisiones acertadas / no acertadas a cerca de muestras de voz comparadas.
- Determinar propiedades de voz específicas.
- Determinar características individuales del hablante.
- Verificación instrumental de fenómenos fonéticos y prosódicos revelados en la etapa de auditoria de examinación lingüística.

## REQUERIMIENTOS DEL COMPUTADOR

Este hardware y software es 100% compatible con IBM PC y que cumplan los siguientes requerimientos mínimos:

- CPU Intel Pentium 300MHz o mayor.
- RAM 512 MB o mayor.
- Adaptador de video y monitor SVGA con resolución no inferior a 1024\*768.
- DOS MS Windows 2000/XP.
- Tarjeta de sonido I/O STC-H216 ó STC-H246.
- Espacio en disco duro 10 GB recomendado.
- Impresora.

### TARJETA DE SONIDO I/O STC-H216 Ó STC-H246

Esta tarjeta de sonido profesional permite la entrada y salida de voz diseñado para la medición y conversión análoga-digital y digital-análoga de corriente eléctrica, medición de amplitud y frecuencia de transformación de las señales eléctricas, así como para la formación de cambios en señales análogas con tensiones dadas, frecuencia, fases e intervalos de tiempo.

El equipo permite la introducción salida precisa bi-canal de la PC de información de sonido en forma análoga o digital, así como permite escuchar al operador a través de audífonos las señales entrantes y salientes.

Los siguientes elementos (tabla 15) hacen parte del conjunto de la tarjeta.

N	Nombre	Especificación	Marca	Cant.
1	Aparato para la medida de las características y de la formación de señales eléctricas en el diapasón de frecuencia de sonido "Kamerton"			1
2	Drivers de Interfase	USB 2.0		1
3	Cable USB	USB (A-B)	5	1
4	Cable de entrada-salida de señal Análoga XLR (M) – XLR (F)	Cable 950/1,5 Negro.	1	2

5	Cable de alimentación eléctrica	SCZ – 1R		1
6	Cable Óptico	TBD-OPT-612	2	1
7	Cable de entrada/salida Digital XLR (M) – XLR (F)	Cable 952/1,5 Negro.	3	1
8	Cable de entrada/ salida digital RCA (M) – RCA (M)	Cable 902/1,5 Negro.	4	1
9	Cable de entrada análoga RCA (M)-XLR(M)		6	2
10	Cable de salida análoga XLR(F)-RCA(M)		7	2
11	Adaptador de entrada análoga 2RCA(M)-Stereo Jack 3.5	AC-010		1
12	Maletín de transporte			1
13	Manual del usuario			1

**Tabla 15.** STC-H246 Tarjeta Profesional I / O De Sonido Manual de usuario (Speech Technology Center).

## CARACTERÍSTICAS TÉCNICAS

La tarjeta se conecta al computador a través del cable USB 2.0. El PC se utiliza como medio de visualización de los resultados del trabajo del equipo y para la conservación o grabación de los datos de entrada o formados.

La tarjeta soporta MS Windows con interfase multimedia para aparatos de audio. Esta permitido OC- Windows XP (Service Pack 2).

La asistencia de energía eléctrica para este aparato es de la corriente eléctrica de 110 / 220V a una frecuencia entre 50/60 Hz.

Condiciones ambientales de trabajo del equipo:

- Diapasón de temperatura del aire alrededor de: 10 - 35 °C.
- Humedad relativa del aire en la temperatura de 25°C: 80.
- Diapasón de presión atmosférica 80-107.

La tabla 16 muestra la ficha técnica de la tarjeta de sonido I/O STC-H216 Ó STC-H246.

## ESPECIFICACIONES TÉCNICAS

N	Parámetros comprobados	Significación en Condiciones técnicas
1	Banda libre en el nivel –2dB para la frecuencia de muestro (sampling rate): 200 kHz: 96 kHz: 48 kHz:	0.01-88 kHz 0.01-43 kHz 0.01-21 kHz
2	Correlación señal-ruido en el canal de referencia en la frecuencia de muestreo de 48-192 kHz con el nivel nominal de testeo de señal de 1kHz, en las bandas de frecuencia 20 Hz - 20 kHz, dB, No menor a :  Considerando el paso a dB, no menor:	105  110  112
3	Distorsión armónica total en el canal de referencia con la frecuencia de muestreo de 48 kHz, bajo el nivel nominal de la señal de testeo en la frecuencia de 1kHz, %, no mayor : Tipo de sentido, %	0.001 0.0004
4	Extinción íter canal de cambio en el canal de referencia con la entrada nominal en la frecuencia de 1000 Hz en la frecuencia de muestreo de 48 - 192 kHz, dB, no menor a:	105
5	Diapasón dinámico en el canal de referencia en la frecuencia de muestreo de 48 – 192 kHz no menor a: Tipo de sentidos, dB:	105 112
6	Diapasón de regulación de volumen de la entrada, dB:	0 a + 64 con paso de 0,5
7	Desigualdad de la amplitud de respuesta de frecuencia en los canales de entrada en la frecuencia de muestreo de 48 kHz en las bandas 20 a 200.000 Hz, no mayor a (en dB)	± 0.02
8	Resistencia en la entrada en k-Ohm no menor a	10
9	Resistencia mínima de tensión: - en los audífonos, en Ohm: - En la salida lineal, en Ohm:	16 600

10	Coeficiente de supresión de las interferencias en fase en las salidas en la frecuencia de 1000 Hz, (en dB), no menor a:	100
11	Coeficiente de distorsiones por íntermodulación, (en %), no mayor a: Tipo de sentidos, %:	0.005 0.003
12	Diapasón de regulación del volumen de salida, (en dB):	0 a 127,5 (en pasos de 0.5)
13	Amplitud máxima de señal en la salida de los audífonos con carga de 16 Ohm, de tensión efectiva, no menor a:	1
14	Amplitud nominal de señal para las entradas y salidas análogas, tensión efectiva:	1,5
15	Amplitud máxima de señal para las entradas análogas, tensión efectiva, no menor a:	6.0
16	Amplitud máxima de señal para las salidas análogas, tensión efectiva, no menor a:	3.0
17	Descargas de la conversión análoga-digital:	24 bits
18	Descargas de la conversión digital- análoga:	24 bits
19	Frecuencia de muestreo (sampling rate) permitida en los canales digitales de entrada y salida, kHz:	22.05 a 100
20	Frecuencias de muestreo (sampling rate) básicas para los canales análogos de entrada-salida, kHz:	1, 8, 10, 11.025, 11, 167, 16, 22.05, 32, 44.1, 48, 96, 192, 200.
21	Error máximo en la selección de las frecuencias de muestreo para los canales análogos de entrada y salida en los diapasones de 4.0 - 200 kHz, no mayor a:	± 5
22	Potencia de consumo VA, no mayor	20 Watios
23	Alimentación eléctrica, Voltios	110-240(50 – 60 Hz)

**Tabla 16.** Especificaciones técnicas de la tarjeta de sonido STC-H246 Manual de usuario (Speech Technology Center).

## **MULTI-SPEECH AND CSL 4500 SOFTWARE (Computerized Speech Lab)**

El CSL es un sistema de hardware y software para la adquisición, análisis acústico, muestra y reproducción de señales de voz. CSL Y Multi-Speech es usado para grabar y estudiar los parámetros acústicos del habla y señales de voz para lingüistas, forenses y aplicaciones científicas de comunicación.

En las ciencias de la salud el análisis acústico es usado para visualizar y valorar el comportamiento de las cuerdas vocales y el rendimiento del habla y de la voz. El sistema completo de hardware y software tiene características como:

- Excelente fidelidad en la adquisición de la señal y adecuación play back.
- 4 canales A/D con un sampling arriba de 200kHz por canal y una resolución de 24-bit.
- Alta resolución y definición en colores de los espectrogramas en VGA graphics.
- Entrada y salida digital con filtros anti-aliasing para todos los rangos.
- Interfase PCI.
- Una series de opciones de software disponibles para pitch, análisis espectrográfico, pedagogía alfabética fonética, análisis de silabas y vocales, etc.

### UTILIDADES DEL SOFTAWRE modelo 4500

El software CSL permite:

- Conversión de señal analoga-a-digital, Entrada / salida al computador vía tarjeta de sonido.
- Resolución de 8-bit a 32-bit.
- Sampling rate hasta 200KHz.
- Grabación en tiempo real de entrada de señal de voz.
- Control de calidad de grabación de sonido con filtros.
- Permite la visualización de varias ventanas simultáneas.

- Fácil manejo implementando short cuts y comandos de acceso.

Edición de señal de voz digitalizada:

- Importa grabaciones realizadas en DAT sin pérdidas de resolución.
- Ajuste de marcas delimitando los componentes de señal.
- Modificar la amplitud y duración de los componentes de señal.
- Normalización de señal.
- Compilar muestras de prueba de varios fragmentos de entrada de señales de voz.
- Reproducción de muestras de voz de prueba generadas para establecer los contenidos de una grabación de sonido.

Análisis de fragmentos de señal seleccionados:

- Amplitud de componentes de la señal.
- Duración de componentes de señal.
- Función fundamental de frecuencia para segmento / fragmentos seleccionados.

El software permite los siguientes análisis:

- Espectrográfico.
- LPC Frequency Response.
- LPC WaterFall.
- FFT Power spectrum.
- FFT WaterFall.
- LTA (Long Term Average)
- Cepstrum analysis.
- Formant History.
- Pitch Contour.
- Energy Contour.

## REQUERIMIENTOS DEL COMPUTADOR

- Pentium III mayor a 800 MHz.
- MB video RAM.
- 64 MB de RAM.
- Una ranura libre PCI.
- Tarjeta de sonido.
- Windows 2000 a Windows XP

## CSL 4500 HARDWARE

El hardware como dispositivo de grabación de entrada-salida para un PC responde a las especificaciones y características necesarias para la captura de la señal.

Incluye una interfase PCI, usando la configuración ASIO para mantener una baja latencia entre el módulo externo y el computador. Posee abundantes características para el análisis de la voz.

El CSL hardware ofrece una relación señal ruido de 20-30dB superior a las a las tarjetas de sonido genéricas.

## ESPECIFICACIONES TÉCNICAS

Entradas análogas:

- 4 canales: 2 XLR y 2 RCA con 5mV-10.5V de pico a pico.
- Switch acoplador para CA o CC.
- Entrada calibrada, gama ajustable >38dB con 24-bit A/D.
- Sampling rates: 8000-200,000Hz.
- THD+N: <-90dB FS.
- Respuesta de frecuencia (CA acoplada): 20 a 22kHz +.05dB en 44.1kHz



Interfase de Digital:

- Formato de AES/EBU o de S/P DIF, transformador acoplado.

Interfase de software:

- ASIO y MME

Interfaz de la computadora:

- PCI (versión 2.2).
- tarjeta del PCI 5.0" H x 7.4" W x 0.75" D (tarjeta PCI medio clasificada)

Salida análoga:

- canales, líneas y altavoces,
- salida del auricular, canales 1 y 2 proporcionan la línea salidas del &speaker

Dimensiones físicas:

- 4" W x 8.25" H x 12.5" D.
- 4 libras. 12 onzas.
- 45 vatios de potencia.

## **MICRÓFONO**

Para la grabación de las muestras, se pueden emplear dos tipos de micrófonos; dinámicos y de condensador. Los micrófonos dinámicos no son recomendados para grabaciones de voz debido a que no registran muy bien las variaciones mínimas de presión producidas por el habla. Sin embargo, Kaypentax Elemetrics permite el uso de este micrófono siempre y cuando se use con la tarjeta de sonido CSL 4500 Hardware. Uno de estos micrófonos por su facilidad de adquisición es el

Shure SM-48 que se encuentra dentro de estos requerimientos y recomendado por Kaypentax para el uso con el CSL.

## SHURE SM-48

El micrófono dinámico unidireccional SM-48 mantiene un verdadero patrón cardiode a través de su rango de frecuencia, asegurando una alta ganancia.

La respuesta de frecuencia es ideal para voz, aplicaciones de palabras habladas contando con filtros anti-pop para disminuir ruidos explosivos o ruidos producidos por la respiración y controla efectos de proximidad.

Además, incluye un soporte para reducir ruidos por vibraciones y rozamientos.

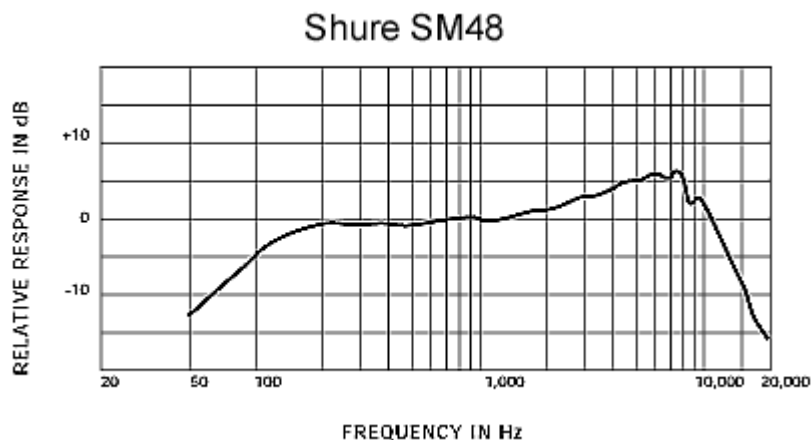
### ESPECIFICACIONES

Tipo:

- Cardiode.

Respuesta en frecuencia:

- 55 Hz a 14000 Hz.



**Figura 38.** Respuesta en frecuencia de SM-48 obtenido de [www.shure.com/ProAudio/Products](http://www.shure.com/ProAudio/Products)

Patrón polar:

- Cardiode (unidireccional) rotacionalmente simétrico con respecto al eje del micrófono uniforme con frecuencia.

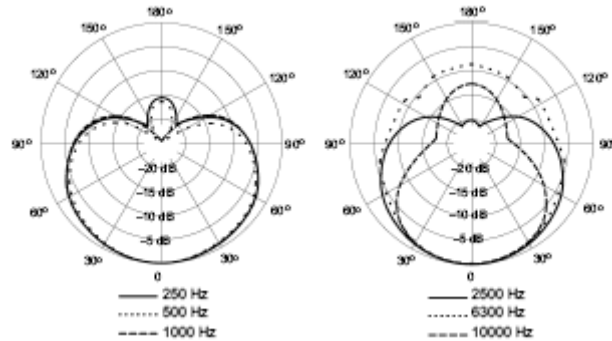


Figura 39. directividad de SM-48 obtenido de [www.shure.com/ProAudio/Products](http://www.shure.com/ProAudio/Products)

Impedancia:

La impedancia especificada es de 150  $\Omega$  (270  $\Omega$  efectivos) para su conexión a entradas de micrófono especificadas entre 19 y 300  $\Omega$ .

Nivel de salida (a 1KHz):

- Voltaje de circuito abierto : -57.5 dBV/Pa (1.3 mV)  
1 Pa = 94 dB SPL

Conector:

- Audio profesional XLR.

Polaridad:

- La presión positiva produce un voltaje positivo en el pin 2 respecto al 3.

Por otro lado, tenemos los micrófonos de condensador los cuales tiene una respuesta más exacta en frecuencia y rango dinámico de la señal, su directividad puede ser omnidireccional, cardiode o hipercardiode, sin embargo su desventaja es la dependencia de una potencia externa (phantom power).

Para este micrófono se recomienda usar del tipo Head-mounted (con soporte para la cabeza), para evitar pérdidas de intensidad producidas por variaciones de desplazamiento de la boca al micrófono.

Este tipo de micrófono es recomendado por las asociaciones de investigación más importantes sobre grabación del habla (NCVS, Department of Speech Music and Hearing Royal Institute of Technology, Scientific Department of Speech Technology Center, etc).

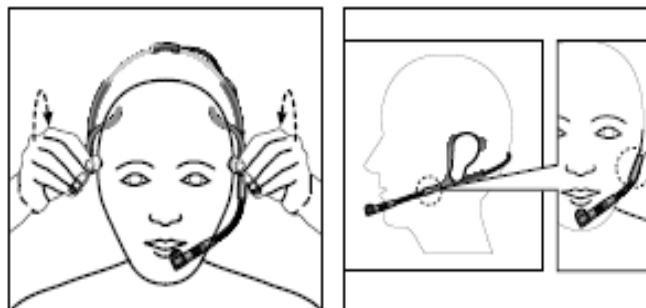
La universidad de Michigan, Bartek Plichta, recomienda el AKG C420 para grabación de la voz.

### **El AKG C420**

El C 420 es un micrófono miniatura de condensador, que se lleva en la cabeza con un cabezal de nuca. El micrófono tiene característica direccional cardiode y el sonido lo capta preferentemente adelante de la boca.

La caja del micrófono con el transductor de condensador está unida al brazo del micrófono por medio de una suspensión elástica para reprimir ruidos de cable y de vibraciones.

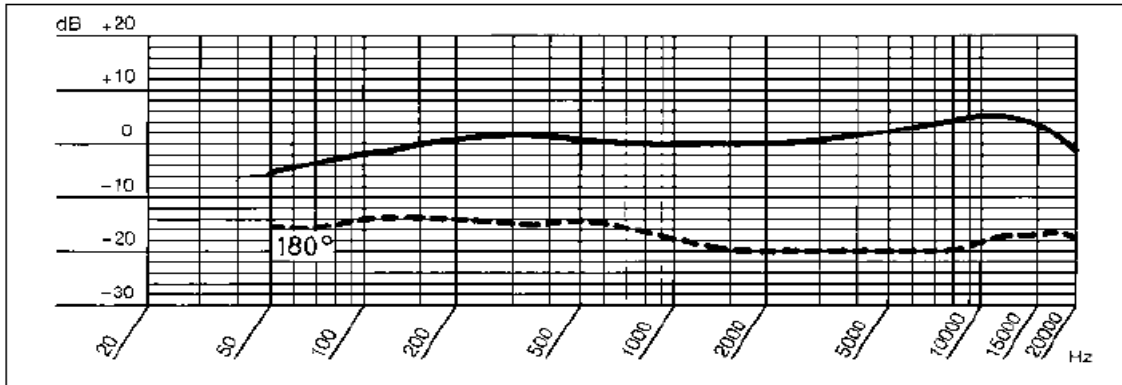
El brazo del micrófono (figura 40) posiciona el micrófono al lado y por delante de la boca. Con ello se evitan en gran medida los sonidos pop y el micrófono queda protegido contra suciedad. La pantalla antiviento suministrada reprime además sonidos de viento y pop.



**Figura 40.** Configuración de montura AKG C 420. Tomado de <http://www.akg.com>

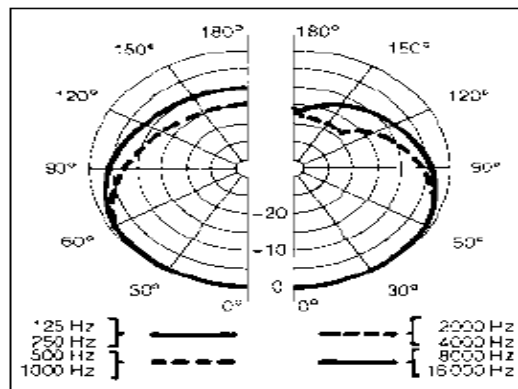
## ESPECIFICACIONES

Respuesta en frecuencia:



**Figura 41.** Respuesta en frecuencia AKG C 420. Tomado de <http://www.ake.com>

Diagrama polar:



**Figura 42.** Diagrama polar AKG C 420. Tomado de <http://www.ake.com>

Modo de funcionamiento:

- Micrófono de condensador con carga permanente

Característica direccional:

- Cardioide

Gama de frecuencia:

- 20 - 20000 Hz

Sensibilidad:

- mV/Pa (-46 dB referido a 1 V/Pa)

Impedancia eléctrica a 1000 Hz:

- 200  $\Omega$

Impedancia de carga recomendada:

- >2000  $\Omega$

Nivel de ruido equivalente:

- 33 dB (según DIN 45412)

Tensión de alimentación:

- C 420: Alimentación fantasma universal 9 - 52 V

Toma de corriente:

- aprox. 2 mA

Tipo de conector:

- XLR de 3 polos

Presión sonora límite para factor de distorsión no lineal de 1% / 3%:

- 126 / 130 dB SPL

## AUDÍFONOS

Para el monitoreo de la grabación es necesario disponer de audífonos de optima calidad con respuesta plana y aislamiento del exterior.

Los **HD 250 II** Linear de Sennheiser son audífonos dinámicos estéreo de alta fidelidad, cerrados, con la sensación de un diseño abierto. Son una selección ideal

no solo para monitoreo profesional y grabación de sonido en vivo, sino también para uso con sistemas de alta fidelidad.

## CARACTERÍSTICAS

- Reproducción precisa del sonido con dinámicas amplias.
- Bobinas de voz de aluminio extremadamente ligeras aseguran una excelente respuesta transitoria.
- Sistemas de magneto de neodimio ferroso para óptima sensibilidad.
- Campo difuso ecualizado para audición no-adulterada.
- Cable de cobre OFC desmontable.

## ESPECIFICACIONES TÉCNICAS

Impedancia nominal:

- 300 Ohm.

Valor nominal de carga:

- 200 mW.

Presión de contacto:

- 3 N.

Peso sin cable:

- 215 g

Clavija hembra/receptáculo hembra

- 3,5/6,3 mm estereo.

Principio del transductor:

- Dinámico, cerrado.

Acoplamiento al oído:

- Circumaural

Longitud del cable:

- 3 metros.

THD, distorsión armónica total:

- Menor a 0.2 %.

Nivel de presión sonora (SPL):

- 94 dB(SPL)

Respuesta en frecuencia:

- 5 a 40000Hz (-10dB) y de 10 a 25000Hz (-3dB).

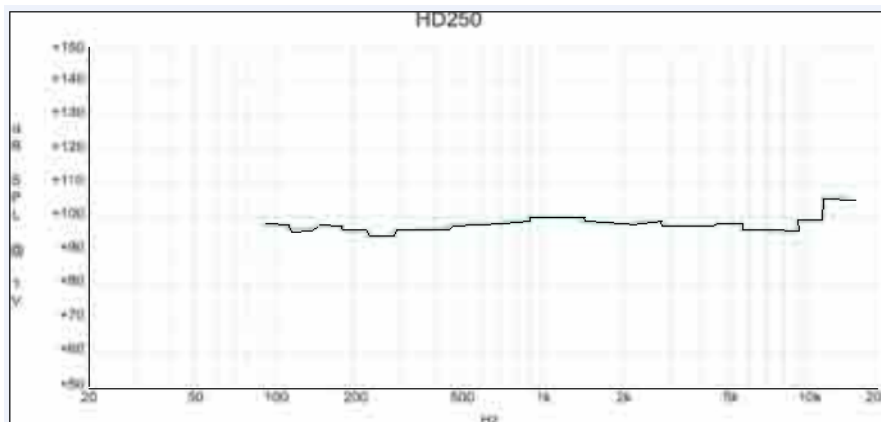


Figura 43. Respuesta en frecuencia de los HD 250 II extraído de <http://.sennheiser.com>.

## MONITORES

### HS 50M POWERED MONITOR SPEAKER

Los *HS 50M* son monitores de alto rendimiento para sistemas de monitoreo profesional con un diseño especial que evita vibraciones espurias y resonancias permitiendo al parlante una completa capacidad de potencia sonora. Además, este



sistema puede ser complementado con subwoofer para realzar bajas frecuencias, el tweeter esta diseñado para minimizar perdidas en altas frecuencias.

El monitor auto-potenciado al ser bi-amplificado separa los amplificadores de potencia que son usados para el manejo de las unidades de frecuencia low-mid y high eliminando la interacción entre las unidades así como anomalías en respuesta y fase introducidas por redes crossover pasivas. Además, permite una confiabilidad de monitoreo de campo medio y cercano.

## ESPECIFICACIONES TECNICAS

Tipo:

- Bi-amplificado 2 - vías parlante auto potenciado.

Crossover Frequency:

- 2 kHz.

Respuesta en frecuencia:

- 42 Hz—20 kHz (-10 dB).

Dimensiones

- (W x D x H) 165 x 222 x 268 mm 250 x 332 x 390 mm 300 x 386 x 350 mm.
- Weight 5.8 kg 11.3 kg 12.5 kg.

Componentes del parlante:

- LF: 8" cone (Magnetic shielding Type) 8" cone.
- HF: 1" Dome (Magnetic shielding Type).
- Enclosure Tipo Bass-reflex.
- Material MFD

Potencia de salida:

- Salida de potencia total 120 W (dynamic power) 150 W 4 ohms (dynamic power).

- (LF: 75 W, 4 ohms)
- (HF: 45 W, 8 ohms)

Sensibilidad de entrada (impedancia):

- 10 dBu/10 k ohms.

Conectores de entrada:

- 1: XLR-3-31 tipo (balanceado).
- 2: TRS PHONE (balanceado).

Controles:

- Control de nivel (+4 dB/center click)
- LOW CUT switch.
- (FLAT/80/100 Hz, 12 dB/octava.
- Equalizador MID (+/- 2 dB at 2 kHz).  
HIGH (+/- 2 dB at HF).
- ROOM CONTROL.
- (0/-2/-4 dB under 500 Hz).
- Indicador de encendido ON: LED blanco.
- Consumo 60 W.

### **5.1.2 TÉCNICAS DE GRABACIÓN PARA MATERIAL INDUBITADO**

Cuando se tiene todo un sistema de grabación y monitoreo establecido, éste debe ser calibrado para revisar posibles errores de conexión y evaluar el funcionamiento de los equipos.

Para empezar la grabación de la muestra en primera medida, el micrófono (debe cumplir con las especificaciones mencionadas anteriormente) debe estar montado

en una base fija para evitar su manipulación y desplazamiento dentro del recinto o tener la configuración de head-mounted.

La *National Center for Voice and Speech* en la publicación “Workshop on acoustic voice analysis” realizada por INGO R. TITZE, PH.D (1993), establece algunas recomendaciones técnicas para grabación de elocuciones de vocales y consonantes. Las siguientes pautas para la grabación de las muestras están soportadas por la mayoría de las asociaciones y organizaciones que estudian la voz en aplicación forense.

- La distancia máxima de la boca al micrófono es de 30cm cuando se emplean micrófonos soportados en bases fijas.
- Es indispensable mantener una distancia constante entre la boca y el micrófono para evitar pérdidas de intensidad en la voz o efectos de proximidad.
- Cuando la distancia del micrófono y la boca es cercana es necesario posicionar (45° a 90° del eje de la boca) el micrófono fuera de eje para reducir el ruido aerodinámico de la boca al hablar.
- La distancia es preferiblemente de 3 a 4cm siempre y cuando se emplee un micrófono de configuración head-mounted para evitar un wow artificial (fluctuación periódica de la amplitud o modulación de la misma) y mantener una alta relación señal-ruido.
- Es importante determinar que patrón polar se debe usar a la hora de grabar; por ejemplo si el cuarto no está tratado acústicamente en su totalidad, se debe usar un patrón cardiode o hipercardiode para evitar la captura de posibles efectos producidos por el mismo (aleteos, sensación

ambiente de cuarto, ruido, problemas de fase, etc). De lo contrario, es posible usar un patrón omnidireccional.

- Los parámetros de reverberación, ruido, efecto de proximidad y reflexiones de las superficies de la sala inciden en la grabación de la señal, por esta razón deben ser controlados.
- La frecuencia de muestreo debe comprender valores desde 20-100Khz dependiendo del grado de interpolación entre las muestras y las que provee el software de análisis (Titze, Horii & Scherer, 1987; Milenkovic, 1987; Deem et al., 1989).
- Evitar en la menor medida mantener encendido durante la grabación equipos que no se estén utilizando, con el fin de suprimir el hum de 60Hz producido por la mayoría de dispositivos durante su funcionamiento o por ruidos externos como puertas o procesadores de computadores. Debido a que la permanencia de ruido de bajas frecuencias puede interferir con el análisis acústico de la frecuencia fundamental y el primer formante.

Las grabaciones se deben hacer en un cuarto tratado acústicamente (ruido de ambiente menor a 50dB); dado que 120Hz están muy cerca de la frecuencia fundamental del habla masculina normal promedio, se debe tener un cuidado especial para evitar fuentes de ruido en el cuarto que crean frecuencias de 60Hz de hum y armónicos asociados. Uno debe especificar generalmente la carga espectral del ruido permisible en el cuarto tratado acústicamente. Esto es particularmente importante si una filtración inversa se realiza desde de la señal del micrófono.

Ahora si no cuenta con una tarjeta profesional mencionada con anterioridad puede reemplazar por un dispositivo de grabación (DAT).

- Consecuentemente, se debe realizar una prueba de señal de entrada para establecer el nivel entre los preamplificadores de la interfase y el nivel de presión sonora de la fuente, para obtener una señal con un rango dinámico aceptable.
- Es necesario ajustar los parámetros de grabación del software (frecuencia de muestreo de 11025Hz y 16 bit de resolución), revisar el funcionamiento del sistema de monitoreo para controlar la grabación durante su proceso y mantener la comunicación con el perito.
- Iniciar la grabación verificando la calidad de la muestra de voz para almacenarla en formato de audio Wave PCM o aiff.

### **5.1.3 PROCESAMIENTO DEL MATERIAL DUBITADO**

Debido a las condiciones en que son grabados estas muestras es necesario el filtrado para reducción de ruido, mejoramiento de la inteligibilidad, rescate de información, edición, separación de locutores (búsqueda de palabras claves) y normalización de muestras.

En el momento de captura de la señal de sonido, la baja calidad en grabaciones o en los canales de comunicación deben ser controlados, es posible procesarla por los medios de incremento de calidad de voz. En principio, la calidad de voz e inteligibilidad debe alcanzar un nivel cómodo para la percepción y la comprensión del personal.

Como sabemos los canales de comunicación introducen interferencias y distorsiones adicionales. Por esta razón la aplicación de herramientas y técnicas de cancelación de ruido y normalización de señal permite obtener mejores resultados de mejoramiento de señal.

Un equipo de procesamiento de señal debe contar con las siguientes características:

- Ecuador.
- Filtros adaptativos para ruidos periódicos y de banda ancha.
- Filtro de suavizado espectral adaptable para suprimir ruidos periódicos de sistemas eléctricos o vibraciones mecánicas.
- Filtro para interferencias de banda angosta y regulares.
- Filtración de impulsos.
- Atenuación de señal de entrada y salida.
- Debe contar con procesos dinámicos.
- Sistema para monitoreo y procesamiento en tiempo real.

Es importante resaltar que las muestras de voz son pruebas dentro de un proceso legal, por tal razón su contenido original no puede ser manipulado.

Si dado el caso es inevitable la manipulación, justificar con un informe científico detallado los procesos que se le aplicaron para su mejoramiento.

Para uso forense las muestras obtenidas, tanto dubitadas e indubitadas cumplen diferentes funciones, ya que algunas son destinadas al peritaje y forman parte de la evidencia de un proceso legal, otras están destinadas a recolectar información que conlleve al esclarecimiento de un delito.

En el caso que la muestra recolectada sea parte de un proceso legal contra algún individuo, su procesamiento es restringido ya que se estaría modificando de alguna manera u otro la evidencia original.

Sin embargo, existen casos especiales donde la muestra presenta baja inteligibilidad, para este caso es recomendado realizar una copia de la señal original para someterla de modificación requeridos, sustentándolo con un soporte científico detallado de los procesos realizados junto con un informe de la persona encargada.

Para muestras las cuales son destinadas a recolección de información, como nombres, lugares, ubicación geográfica, direcciones o cualquier otro tipo de información vital que aporte a una investigación, la muestra puede ser editada, procesada, amplificada para su respectivo mejoramiento.

Según los factores que afecten la muestra, existen diferentes procesos que permiten el mejoramiento de la señal.

Estos Procesos utilizados en el mejoramiento de la señal pueden lograrse con una calidad óptima con la implementación del software SoundCleaner.

#### ECUALIZADOR

El ecualizador debe contar con filtro inverso automático-manual, donde la señal es analizada. Después de este proceso el software determina los picos y valles que se debe atenuar para que la señal sea lo mas plana posible en el espectro.

El ecualizador se utiliza para eliminar ruidos estacionarios como ruidos de línea de poder, mecánicos, ruido de motor, etc.

#### FILTRACIÓN DE BANDA ANCHA ADAPTABLE

Se basa en un algoritmo de frecuencia adaptable diseñada para suprimir ruidos periódicos y de banda ancha producidos por levantamientos eléctricos, interferencias de equipo, vibraciones mecánicas, ruido de salón o de calle, canal de comunicación o interferencias en los equipos de grabación.

#### FILTRACIÓN INVERSA ADAPTABLE

Se basa en un algoritmo de corrección espectral adaptable o también llamado suavizado espectral adaptable. La adaptación inversa adaptable efectivamente suprime los ruidos periódicos fuertes de levantamientos electrónicos o vibraciones mecánicas recuperando la señal de la voz.

Este filtro amplifica los componentes de señal más débil y suprime los más fuertes al mismo tiempo, aprovechando que el espectro plano desenmascara la señal de voz y mejora su inteligibilidad.

Por ejemplo, cuando tenemos una comunicación telefónica donde participan dos locutores y uno de ellos tiene un nivel más bajo de ganancia que el otro es necesario establecer un punto de equilibrio entre la reducción de ruido y la percepción de la voz.

### COMPENSACIÓN DE FRECUENCIA

Este filtro se basa en el algoritmo de filtración adaptable para compensación adaptable de un canal. Es más efectivo para interferencias estacionarias de banda angosta regulares con un ajuste automático, manteniendo estable la calidad de la voz en la señal útil.

La compensación adaptable permite remover, dependiendo su configuración, ruidos de dispositivos eléctricos, música estable, ruido de habitación, ruido vehicular, reverberaciones entre otros.

La principal ventaja de la compensación adaptable es la capacidad de preservar la señal de la voz mucho mejor que otros filtros convencionales.

### DISMINUCIÓN

Se utiliza para incrementar o disminuir la velocidad de reproducción de la señal sin alterar el tono de la voz (slowing). El ajuste de velocidad usualmente se emplea ya al final del sistema de filtrado justo antes de ser procesado el sonido.

### CLIPPING

Es un limitador de amplitud que remueve fuertes rompimientos y suaviza el nivel de la señal afectando poco la inteligibilidad de la señal. Es muy aplicado en la limpieza de interferencias armónicas.



## TRASFORMACIÓN-MU

Este proceso es aplicado cuando los fragmentos de señal difieren en gran medida en nivel, por ejemplo si un locutor esta ubicado cerca del micrófono y el otro lejos de este, los fragmentos débiles son altamente amplificados y los fuertes en menor medida durante la transmisión.

## FILTRACIÓN DE IMPULSOS

La filtración de impulso adaptable automáticamente restaura los fragmentos de voz y música distorsionados, producto de los enmascaramientos producidos por interferencias de pulsos como los clics, ruidos de radio, ruido impulsivo, golpes, etc.

## PROCESO DINÁMICO

Este proceso ayuda a mejorar la inteligibilidad si las muestras de la señal difieren en gran medida el nivel, como por ejemplo golpes resonantes, impulsos largos, ruidos de habitación. Los algoritmos de proceso dinámico mejoran y desenmascaran la señal suprimiendo dichos impulsos y clics poderosos, reduciendo la fatiga del analista. Los filtros anteriormente mencionados pueden ser empleados para grabaciones estereo y monofónicas.

### **5.1.4 MÉTODOS DE ANÁLISIS ACÚSTICOS DEL HABLA**

La etapa posterior a la captura o grabación de señal, bien sea análoga o digital, y procesamiento de la misma es la del análisis. Recordando que para ser ingresadas a un sistema automático (hardware) o semi-automático (software), el material atraviesa el proceso de digitalización mencionado con anterioridad.

A través de este análisis es posible determinar parámetros acústicos del habla y de señales de voz, para lingüistas, forenses, áreas de la salud, inteligencia militar, entrenamiento en lenguaje y ciencias aplicadas a la comunicación.

El enfoque de una investigación de uso forense abarca muchas aplicaciones dentro de las cuales encontramos como:

- Identificación de locutor
- Determinación de acento y dialecto
- Identificación de huella vocal
- Autenticación de grabaciones
- Detección de estrés
- Análisis de cajas negras en accidentes aéreos
- Mejoramiento y restauración de material grabado

En la actualidad existen sistemas automáticos y semi -automáticos que facilitan el proceso de análisis de las muestras. Los programas a los cuales se tuvo acceso limitado comprenden:

- Multi-speech 3700 y C.S.L Software (semi-automático)
- S.I.S. Version 6.x (semi-automático)
- Trawl (automático)
- Batvox (automático)

El principio de análisis que utilizan estos programas es el mismo, la diferencia radica en que los semi-automático son aquellos sistemas donde el factor humano (analista) interviene en la obtención de resultados.

En contraparte, en los sistemas automáticos el hombre solo interviene para el funcionamiento del sistema, más no en el análisis y en la obtención de resultados.

A continuación se realizará por medio de un ejemplo la descripción del proceso en el análisis de una muestra.

La información previamente digitalizada es ingresada al sistema semi-automático (Multi-speech and CSL software) para posteriormente someterla a los siguientes análisis:

#### **5.1.4.1 TIPOS DE ANÁLISIS**

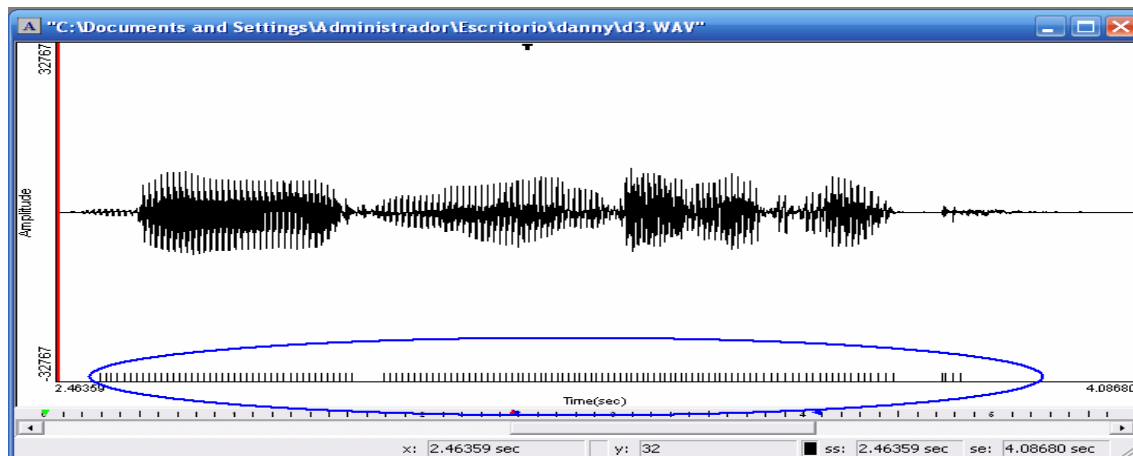
- Spectrogram.
- LPC Frequency Response.
- LPC WaterFall.
- FFT Power spectrum.
- FFT WaterFall.
- Cepstrum analysis.
- Formant History.
- Pitch Contour.
- Energy Contour.

#### **ANÁLISIS DE VOZ**

Con la siguiente muestra de voz se realiza la metodología para identificar las características de una persona y por la cual se puede individualizar.

Frase: ***deee una banda de rock***

## Oscilograma



**Figura 44.** Forma de onda en el oscilograma de la frase “de una banda de rock”.

La señal se grabó directamente con el software Multi-Speech and CSL aplicando los procedimientos de digitalización anteriormente mencionados, en la figura 44 se resaltan las marcas periódicas de la voz, cada línea representa la periodicidad de la señal.

Información de la señal digitalizada

### 5.4.1.2 Espectrograma

El análisis en el espectrograma muestra la representación de las variaciones de la frecuencia en eje vertical y el tiempo en eje horizontal. La energía se observa en escala de grises (gradiente de color) en valores de nivel de presión sonora. Estos espectros son computarizados por medio de una serie de secciones de información muestreada en un rango específico.

Los espectrogramas son la herramienta base para el análisis de formantes, respuesta en frecuencia, periodo glótico, turbulencias, etc.



**Figura 45.** Espectrograma de la frase “de una banda de rock”.

Por la calidad de la señal, (figura 45) resulta un espectro fácil de interpretar por el analista, es muy útil sincronizar el oscilograma con el espectro para relacionar la información proveniente de ambas ventanas.

#### 5.4.1.3 LPC Frequency Response

El análisis LPC (*Linear Predictive Coding*) opera bajo el modelo físico del tracto vocal representándolo mediante un filtro digital. La señal es analizada por medio de un número de patrones que varían en el tiempo con los cuales es posible reconstruirla adecuadamente.

Esta reconstrucción se produce por la generación de un tren de pulsos periódicos o ruido aleatorio que viaja a través de los valores que tomaron los patrones. La información relevante que pasa por el filtro se transformará en la señal original.

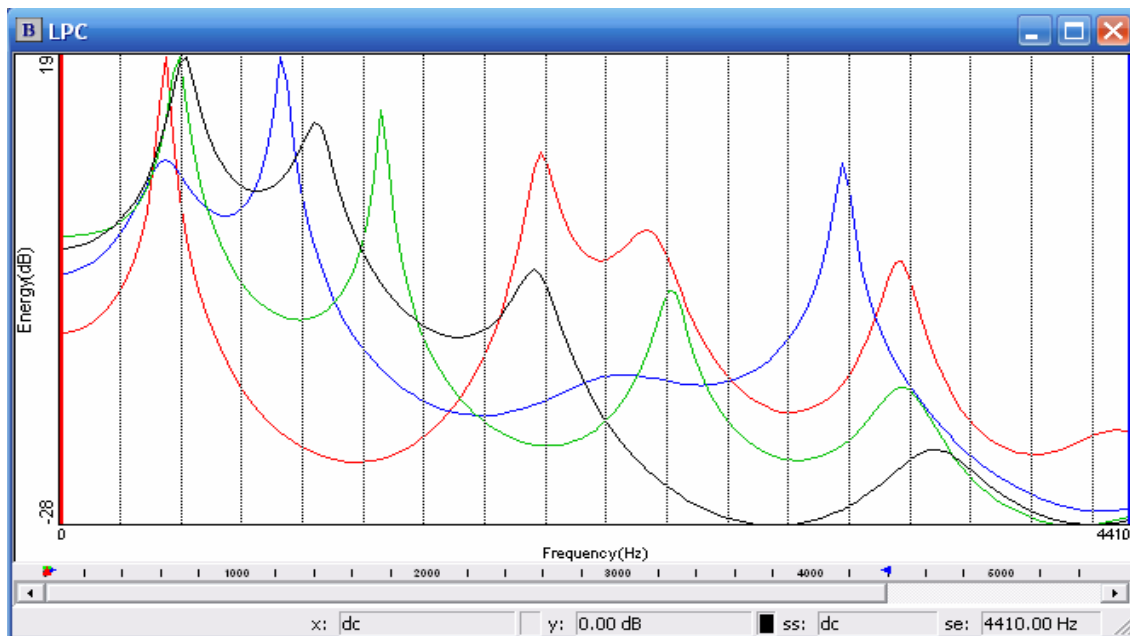
El LPC analiza la señal del habla para estimar las formantes, intensidad y frecuencias de resonancia. El *LPC WaterFall* realiza el mismo análisis pero en tres dimensiones (figura 38) lo que se denomina caída de agua, permitiendo ver el

comportamiento energético y frecuencial de la voz en función de la variable de tiempo.

Análisis LPC Response.

Se divide el oscilograma por etiquetas la frase **Deeee una banda de rock** en nuestro ejemplo.

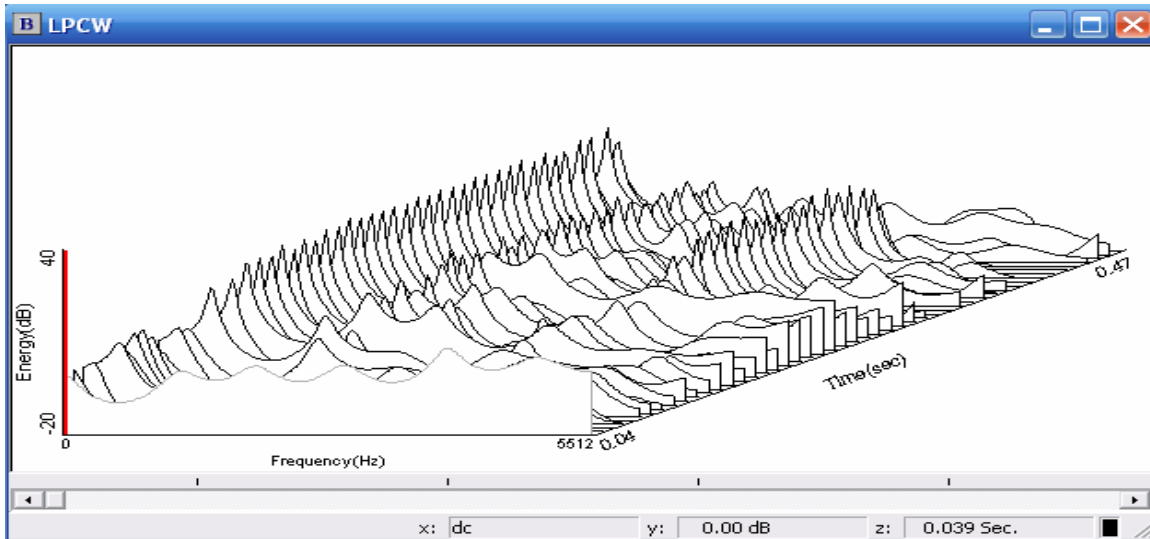
- **Deeee**
- **Una**
- **Banda de**
- **Rock**



**Figura 46.** Análisis LPC frequency Response de la frase “de una banda de rock”.

La figura 46 muestra un análisis mas detallado desde el primer formante hasta el quinto, separadas por palabras representativas dentro de la frase, para casos de cotejo se compara la frase dubitada con la indubitada. Además se aprecian los armónicos de la frecuencia fundamental de cada palabra que compone la frase.

## LPC Waterfall



**Figura 47.** Análisis LPC WaterFall de la silaba “deeee”.

### DATOS NUMERICOS DEL ANALISIS LPC DE FORMANTES

El programa CSL permite la visualización de los resultados en forma numérica para facilitar su estudio. Los siguientes datos muestran las primeras 5 formantes del locutor de cada palabra que conforma la frase.

- El primer formante F1 determina la apertura de la boca del locutor, entre más abierta esté, la frecuencia de dicho formante es mas elevada. Por el contrario cuando la lengua se acerca al paladar la apertura vocálica decrece y la frecuencia del formante disminuyes.
- El segundo formante F2 es determinado por la posición de la lengua dentro del resonador bucal, por acción de los labios y posición de la lengua. Al ser mas pequeño el resonador bucal la frecuencia resultante es mas aguda. Por el contrario entre mas grande sea el resonador bucal la frecuencia son mas bajas. En los labios, cuanto más redondeados y abocinados estén, mas baja es la frecuencia.

- El tercer formante F3 esta determinado por el descenso del velo del paladar y por la relación directa entre la elevación de la punta de la lengua hacia una posición retrofleja, obteniendo un descenso frecuencial en el formante.

### Deeee

#### Formant/Bandwidth Results

Formant (Hz) Bandwidth (Hz)

434.51	35.16
1971.81	95.23
2426.70	212.79
3445.24	130.58
4366.06	435.98

### Una

#### Formant/Bandwidth Results

Formant (Hz) Bandwidth (Hz)

427.73	181.21
914.51	36.38
2268.41	591.76
3211.70	54.16
3372.18	1355.33
4884.91	282.04

### Banda de

#### Formant/Bandwidth Results

Formant (Hz) Bandwidth (Hz)

486.45	48.56
1322.04	31.25
2507.39	106.94
3461.62	230.66
4755.71	814.35



## Rock

### Formant/Bandwidth Results

Formant (Hz)    Bandwidth (Hz)

510.36	77.90
1055.01	116.16
1958.30	148.32
3588.11	365.80
5142.31	416.3

#### 5.4.1.4 FFT Power spectrum

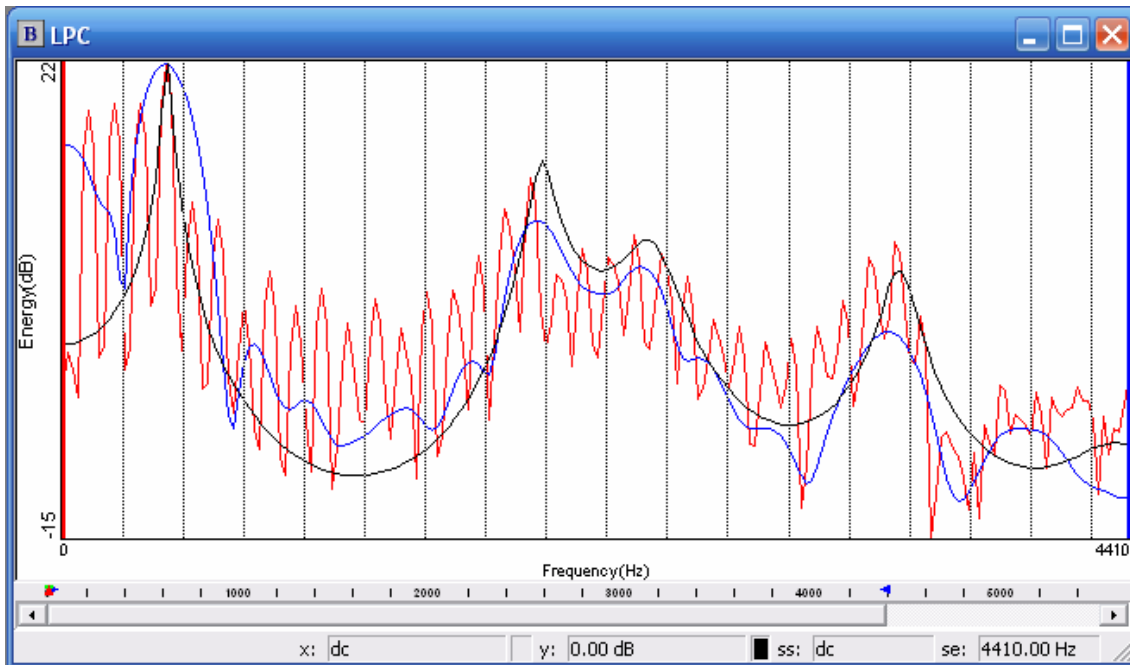
Para el análisis de la FFT se emplea un algoritmo de la transformada rápida de Fourier que calcula el espectro de la voz en un determinado tamaño de datos. Por medio de la FFT se puede calcular la transformada discreta de Fourier, que a su vez es una aproximación discreta y finita de la integral de Fourier.

En cuanto a la naturaleza discreta de los cálculos, aunque la transformada es continua solo podremos obtener muestras de la misma, que pueden constituir una buena aproximación si se toman suficientes puntos. La función FFT calcula la transformada de Fourier de una señal finita en el número de puntos equiespaciados especificado en la función.

Cuando se aplica, además de la transformada discreta de Fourier un análisis LPC, se puede observar la coincidencia en las resonancias que vienen a ser las que caracterizan el contenido frecuencial de la señal vocal, pero el espectro de la señal tratada por el sistema LPC se caracterizaría por tener contornos más suaves de lo normal suavizando el espectro.

La FFT WaterFall realiza el mismo análisis de la FFT power spectrum con la diferencia que muestra la señal en tres dimensiones (figura 48) que permiten ver el comportamiento de energético y frecuencial de la voz en la variable de tiempo.

## Análisis FFT Power Spectrum de la palabra **Deee**



**Figura 48.** Análisis FFT Power Spectrum de la silaba “deeee”.

- Match FFT size
- Pitch Synchronous
- LPC

En la figura 48 muestra el análisis frecuencial de la FFT que facilita la determinación de puntos de aumento y decrecimiento en la señal, el Match FFT size especifica la longitud de la sección de información que fue analizada. Esto puede ser igual al numero de coeficientes de la FFT (tamaño de análisis).

Otra forma de ajuste es delineado por las marcas periódicas de la voz (Pitch Synchronous).

## FFT Waterfall de palabra deee

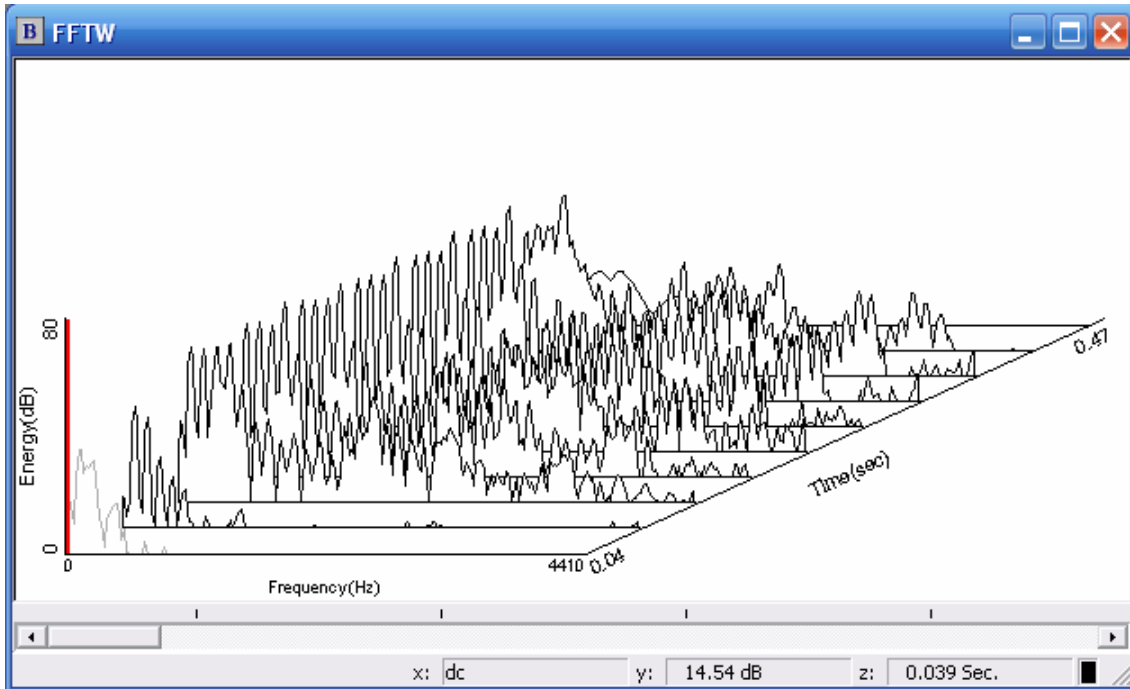


Figura 49. Análisis FFT WaterFall de la palabra “deeee”.

### Información de parámetros

Esta información numérica muestra la frecuencia de muestreo el tiempo de análisis, longitud de cada sección, el nivel máximo y mínimo de SPL. Además, el programa realiza un análisis estadístico del comportamiento de SPL en el ancho de banda escogido.

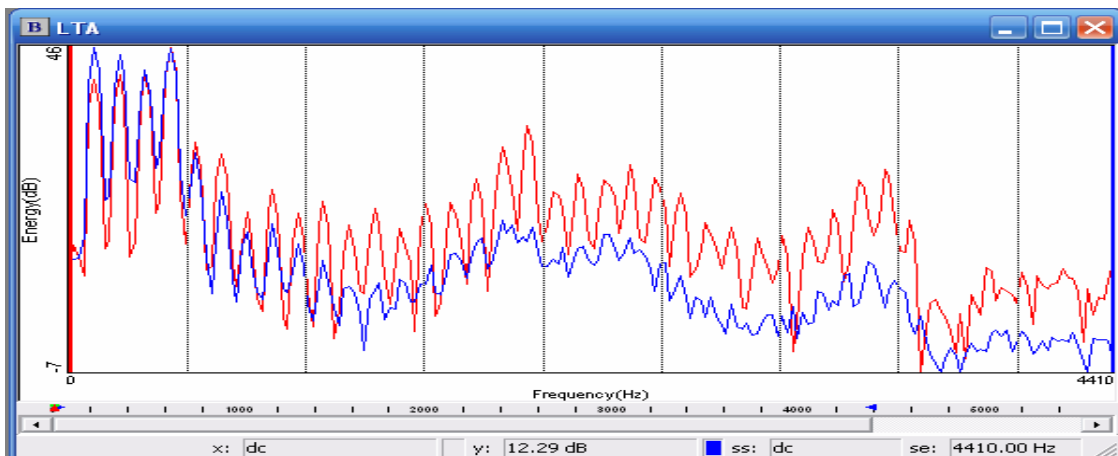
Source	A
Channel	1
Frecuencia de muestreo	11025
Start of Analysis (sec)	0.27283
End of Analysis (sec)	0.28209
Frame Length (points) From Impulses	
FFT Size (points)	512

Minimum Power (dB)	-7.40
Maximum Power (dB)	43.75
Pre-emphasis Level	0.000
Window Weighting	Hanning
Smoothing Level	None

#### ANÁLISIS ESTADÍSTICO

Samples	205
Minimum Frequency (Hz)	0.0000
Maximum Frequency (Hz)	4392.7734
Minimum Power (dB)	-3.26
Maximum Power (dB)	43.75
Mean Power (dB)	12.77
Standard Deviation (dB)	11.70
Median Power (dB)	10.08
Root Mean Squared (dB)	17.30
Spectral Mean (Hz)	462.30
Spectral Std Dev (Hz)	347.97

Análisis FFT Power Spectrum Vs. análisis LTA Power Spectrum de palabra [Deee](#).



**Figura 50.** Análisis FFT Power Spectrum Vs. análisis LTA Power Spectrum de la silaba “deeee”.

### 5.4.1.5 Cepstrum

Se define como la transformada inversa de Fourier del módulo del espectro en escala logarítmica de esa señal. Las bajas componentes cepstrales corresponden a variaciones lentas de las componentes espectrales donde contienen información de la envolvente del espectro, la cual se relaciona con la respuesta en frecuencia del filtro que modela el tracto vocal.

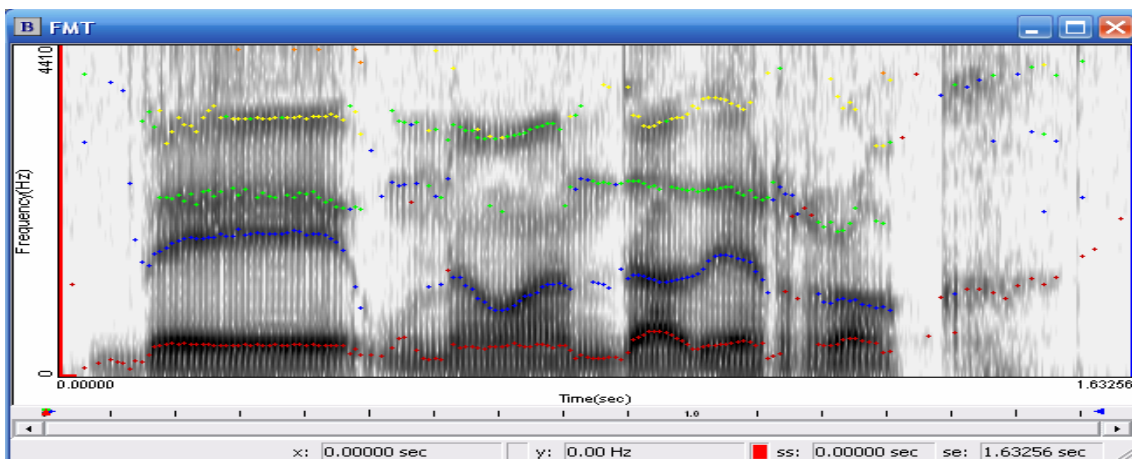
El propósito del Cepstrum es separar la energía armónica que esta representada en el espectro de la voz.

En el caso de reconocimiento de patrones de voz, la información relevante se encuentra en las características del tracto vocal, por lo que se usan bajas componentes cepstrales para reconocer voz más no locutores.

### 5.4.1.6 Formant History

Este análisis muestra información de las formantes y su comportamiento en el tiempo. Es representada por una sucesión de puntos los cuales trazan la formante para transferir los datos gráficos a valores numéricos facilitando su análisis.

En el historial de formantes se consideran picos espectrales a los puntos de mayor energía en el espectro y formantes a los agrupamientos claros de picos espectrales entorna a una banda definida de frecuencias.



**Figura 51.** Análisis histórico de formantes de la frase “deeee una banda de rock”.

En la figura 51 se puede observar los diferentes tipos de formantes descritos por una secuencia de puntos con colores específicos para cada formante, los cuales recrean la información. Esta información puede ser almacenada y mostrada en datos numéricos para análisis un más detallado.

La tabla 17 muestra los datos de F1, F2, F3, F4 y F5 especificando el rango de frecuencia conformado por la historia del formante, es decir la cantidad secuencial de puntos que se encuentre dentro de dicho rango. Por ejemplo la primer formante tiene un rango de 109 Hz hasta 4664 Hz.

### Formant/Bandwidth Statistics

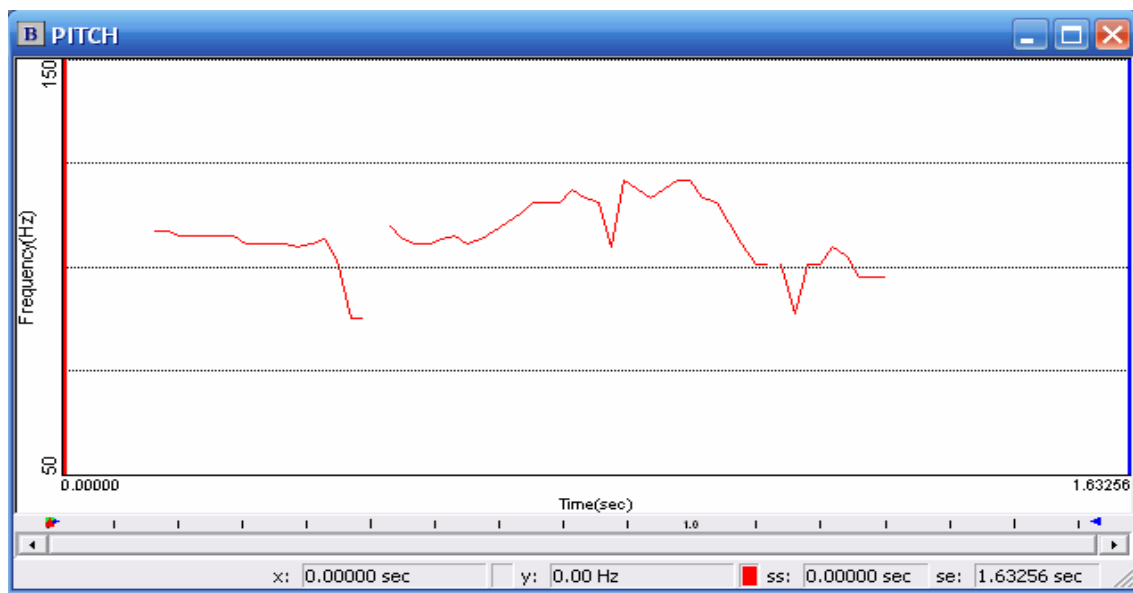
Frames Processed	151					
Start of Analysis (sec)	0.0000					
End of Analysis (sec)	16.326					
	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>
<b>Samples</b>	151	142	131	101	42	0
Minimum	109.80	895.01	1946.18	3089.97	4058.81	
	22.31	24.21	27.56	24.20	152.82	
Maximum	4664.72	4925.62	5098.57	5023.27	5357.28	
	495.53	493.37	493.12	498.15	494.75	
Mean	652.57	1878.59	2979.89	3910.26	4785.85	
	146.79	152.03	211.70	233.31	380.43	
Median	437.03	1656.63	2559.01	3616.81	4782.59	
	95.80	92.80	193.81	233.24	408.55	
Standard Deviation	714.78	911.69	797.35	623.87	301.45	
	128.66	128.47	131.99	142.56	98.21	
Root Mean Squared	966.11	2086.73	3083.94	3959.23	4795.11	
	194.91	198.75	249.21	273.05	392.61	
Geometric Mean	497.08	1707.21	2887.60	3863.85	4776.44	
	100.47	107.88	165.52	175.69	364.49	

**Tabla 17.** Información numérica de formantes del análisis histórico de la frase “deeee una banda de rock”.

### 5.4.1.7 Pitch Contour

Este tipo de análisis trabaja con tonos que cambian durante su realización debido a sus ascendentes, descendentes o cualquier combinación de elevación y descenso. Los tonos de registro son tonos que se perciben como constantes durante su realización, pueden ser altos, bajos o medios.

Este análisis extrae los valores de la frecuencia fundamental para computarlos entre cada marca periódica de la voz y la FFT inversa, dibujando los valores como tonos de contorno.



**Figura 52.** Análisis de los tonos de contorno de la frase “deeee una banda de rock”.

Según la figura 52 podemos identificar el rango de tono del locutor que se encuentra entre valores de 88 y 121Hz permitiendo al analista una variable más de individualización del individuo. Igualmente es posible ver los datos estadísticos de los tonos de contorno.

#### Parameter Information

Source	A
Channel	1
Frecuencia de muestreo	11025

Start of Analysis (sec)	0.00000
End of Analysis (sec)	1.63256
Minimum Pitch (Hz)	88.00
Maximum Pitch (Hz)	121.00
Frame Length (msec)	20
Frame Advance (msec)	20

### Analysis Statistics

Samples	82
Start of Analysis (sec)	0.0000
End of Analysis (sec)	1.6326
Minimum Pitch (Hz)	88.20
Maximum Pitch (Hz)	121.15
Mean Frequency (Hz)	108.67
Mean Fo (Hz)	108.17
Mean Period (msec)	9.24
Standard Deviation (Hz)	7.32
Median Pitch (Hz)	108.09
Root Mean Squared (Hz)	108.91

Cuando el profesional forense tiene basta experiencia en el análisis de este tipo, puede sugerir perfiles de la persona tales como:

- grado de educación
- nivel cultural
- melodía de la voz

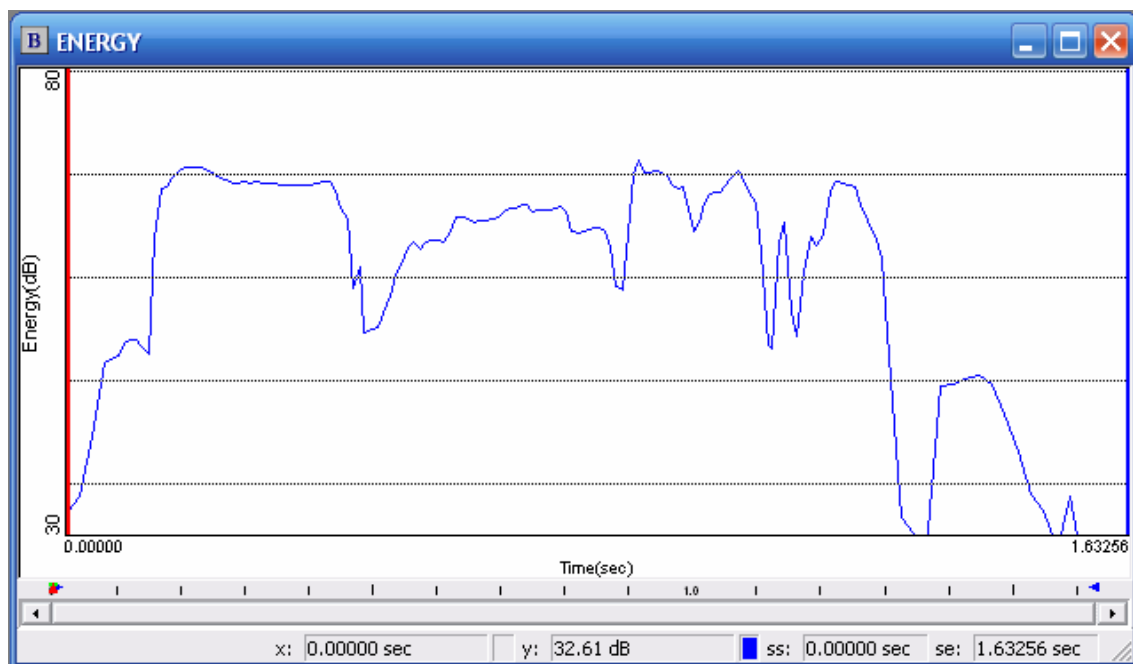
El grado de articulación para reproducir el habla también es reflejado en el contorno de tono, por ejemplo los locutores de radio o televisión.



### 5.4.1.8 Energy Contour

La energía es la suma de los valores de amplitud absolutos en una sección de información dividida por el número de puntos dentro de esta sección, cuando la duración de la sección de datos es un valor en milisegundos.

La energía es computada y después convertida a valores en nivel de presión sonora. Tanto la energía como la magnitud son útiles para distinguir segmentos sordos y sonoros en la señal de voz.



**Figura 53.** Análisis energético de la frase “deeee una banda de rock”.

La figura 53 muestra el ataque, el sostenimiento, la caída y la zona de relajación de la energía del locutor durante su expresión, cayendo a nivel muy bajos en los segmentos donde el emisor realiza pausas durante la oración.

El locutor al inicio de la oración presenta un ataque rápido de la oración, realizando un sostenimiento de la silaba “deee”, posteriormente se mantiene en un rango relativamente constante de energía en las silabas “una ban da de” pero la disminución de energía el la silaba “roc” y la consonante “k” es mas significativa.

## Parameter Information

Source A  
Channel 1  
Frecuencia de muestreo 11025

Start of Analysis (sec) 0.00000  
End of Analysis (sec) 1.63256  
Minimum Energy (dB) 25.19  
Maximum Energy (dB) 70.33

Frame Length (msec) From Impulses  
Smoothing Level None

## Analysis Statistics

Samples 151  
Start of Analysis (sec) 0.0000  
End of Analysis (sec) 1.6326

Minimum Energy (dB) 25.19  
Maximum Energy (dB) 70.33  
Mean Energy (dB) 59.74  
Standard Deviation (dB) 11.32  
Median Energy (dB) 63.99  
Root Mean Squared (dB) 60.80

## 6. CONCLUSIONES

A través de esta investigación se ha demostrado la importancia y el aporte que brinda el ingeniero de sonido dentro del análisis e identificación de locutores.

Podemos concluir que la voz es un patrón que permite la identificación e individualización de las personas siempre y cuando los recursos técnicos, científicos y humanos sean los adecuados.

Los resultados obtenidos por la encuesta, el personal carece de herramientas y técnicas adecuadas para la captura, procesamiento y análisis de muestras de voz. El perfil de los profesionales que se desempeñan en estas áreas solo puede desenvolverse en el análisis audio-perceptivo.

Las sustentaciones de peritajes carecen de soporte científico para ser justificadas. No existen normas que regulen y controlen los procesos involucrados en estas áreas de identificación.

Se demostró un mejoramiento en el material dubitado e indubitado por medio del protocolo planteado en este proyecto, el cual consiste en capturar la señal de forma digital (los resultados obtenidos muestran un alto grado de confiabilidad), implementar técnicas equipos, dispositivos y software de grabación propuestos, procesar la señal para eliminación de ruido y finalmente ser analizada en el software requerido.

## **7. RECOMENDACIONES**

En este proyecto se realizan ciertas recomendaciones para elaborar un equipo de trabajo para las muestras de voz. En primer lugar se deben determinar los perfiles de recurso humano que intervienen en el proceso de identificación del locutor.

### **7.1 EQUIPO INTERDISCIPLINARIO**

#### **INGENIERO DE SONIDO**

- Manipular modernas tecnologías de uso para el procesamiento de la señal sonora en salas de supervisión Identificando factores que afectan la variabilidad intra locutor.
- Implementa técnicas y herramientas para el procesamiento señales de audio (grabación, edición, certificación, análisis espectro gráficos entre otros)
- Diseño de recintos acústicos destinados para procesamiento de señales sonoras (laboratorios).

#### **FONOAUDIOLOGO**

- Identifica factores achacables al locutor ya sean voluntarios e involuntarios denominados factores intrínsecos: edad, estado emocional, estado físico, estar sometido a estrés, velocidad de articulación o tipo de habla – leída, susurrada, conversacional, etc.

#### **LINGUISTA FORENSE**

- Determina el origen geográfico del hablante, edad, ámbito social en el que se desenvuelve.

- Descubre rasgos de personalidad o hábitos de procedencia patológica u aprendida que lo caractericen de alguna manera.
- Utiliza sus conocimientos de sintaxis, semántica, morfología, lexicología, dialectología, sociolingüística o psicolingüística para subrayar características que individualicen al locutor.

#### SICOLOGO FORENSE

- analiza los resultados de los diferentes departamentos y crea un perfil sobre la persona.

### **7.2 TIPOS DE ANÁLISIS QUE INTERVIENEN EN LA IDENTIFICACIÓN DEL LOCUTOR**

Las siguientes son las recomendaciones para mejoramiento de procesos que intervienen en la identificación de locutor.

1. Análisis auditivo preliminar
2. Análisis audio perceptivo
3. Análisis acústico del habla

#### ANÁLISIS AUDITIVO PRELIMINAR

Consiste en determinar la calidad y contenido del material (análogo o digital).

#### ANÁLISIS AUDIO PERCEPTIVO

Identifica patrones de la voz del individuo de manera perceptiva, identificando patrones de velocidad, timbre, entonación, intensidad, resonancia, tono, modulación. Además de patrones de articulación, fluidez y respiración.

Estos patrones son los relevantes y al ser identificados determinan cuales son las oraciones, frases, palabras, silabas, fonemas en los cuales se debe concentrar el análisis.

La siguiente tabla identificará los patrones que se disponen a analizar.

<b>Análisis audio perceptivo</b>	
Oraciones	
Frases	
Palabras	
Silabas	
Fonemas	
Observaciones:	

## ANÁLISIS ACÚSTICO DEL HABLA

Ejecución de técnicas y herramientas para el procesamiento señales de audio (grabación, edición, autenticación, análisis espectro gráficos entre otros) determinando las características acústicas cuantitativas de las muestras como: el tono, intensidad, variaciones de amplitud, timbre preseleccionadas en el análisis audio perceptivo

El siguiente formato pretende estandarizar y controlar los procesos elaborados por el ingeniero en el análisis de locutor. Este es un ejemplo de cómo registrar la información en el informe.

### Ejemplo.

<b>TIPO</b>	DIGITAL
<b>FORMATO</b>	MP3
<b>CALIDAD DE LA GRABACION</b>	BUENA
<b>CANAL DE COMUNICACIÓN</b>	TELEFONIA CELULAR
<b>PROCESOS APLICADOS AL MATERIAL</b>	FILTRO PASABANDA 300 A 4000Hz
<b>TIPOS DE ANALISIS</b>	TONOS DE CONTORNO Y ENERGETICO

<b>INFORME DE RESULTADOS</b>	TONO ENTRE 80 A 150 Hz, VOZ HABLADA
<b>OBSERVACIONES</b>	RUIDO VEHICULAR, INTERFERENCIA
<b>INFORME DE PROCEDIMIENTO PARA SOPORTE CIENTIFICO</b>	SE DETERMINÓ NIVEL DE RUIDO, CON RELACION SEÑAL/RUIDO DE 15Db.....
<b>RESPONSABLE</b>	INGENIERO ENCARGADO

### 7.3 BANCO DE VOZ

Se propone un sistema automático que permita la introducción muestras de diferentes canales de comunicación (BATVOX, TRAWL)

Se recomienda que las muestras cumplan con las siguientes recomendaciones:

- La muestra debe tener mínimo de duración de 20 segundos.
- La grabación debe tener la mejor calidad posible.
- La muestra debe estar etiquetada con la información de las características físicas del individuo describiendo antecedentes.

## ANEXOS A

Las siguientes tablas describen detalladamente el movimiento semanal de las unidades de material dubitado e indubitado, análogo digital, tipo de canal de información y tipo de formato.

### A.1 TABLA DE INGRESO DE MATERIAL DUBITADO E INDUBITADO

<b>Mes 1</b>					
<b>Material</b>	<b>semana 1</b>	<b>análogo</b>	<b>Digital (CD)</b>	<b>análogo %</b>	<b>Digital (CD) %</b>
dubitado	4	3	1	75,0	25,0
indubitado	4	4	0	100,0	0,0
	<b>semana 2</b>	<b>análogo</b>	<b>Digital (CD)</b>	<b>análogo %</b>	<b>Digital (CD) %</b>
dubitado	5	3	2	60,0	40,0
indubitado	2	2	0	100,0	0,0
	<b>semana 3</b>	<b>análogo</b>	<b>Digital (CD)</b>	<b>análogo %</b>	<b>Digital (CD) %</b>
dubitado	3	3	0	100,0	0,0
indubitado	4	4	0	100,0	0,0

	semana 4	análogo	Digital (CD)	análogo %	Digital (CD) %
dubitado	3	3	0	100,0	0,0
indubitado	2	2	0	100,0	0,0
<b>Total dubitado al mes</b>	<b>15</b>	<b>12</b>	<b>3</b>	<b>80,0</b>	<b>20,0</b>
<b>Total indubitado al mes</b>	<b>12</b>	<b>12</b>	<b>0</b>	<b>100,0</b>	<b>0,0</b>

<b>Mes 2</b>					
<b>Material</b>	<b>semana 1</b>	<b>análogo</b>	<b>Digital (CD)</b>	<b>análogo %</b>	<b>Digital (CD) %</b>
dubitado	5	2	3	40,0	60,0
indubitado	3	3	0	100,0	0,0
<b>Material</b>	<b>semana 2</b>	<b>análogo</b>	<b>Digital (CD)</b>	<b>análogo %</b>	<b>Digital (CD) %</b>
dubitado	3	1	2	33,3	66,7
indubitado	4	4	0	100,0	0,0
<b>Material</b>	<b>semana 3</b>	<b>análogo</b>	<b>Digital (CD)</b>	<b>análogo %</b>	<b>Digital (CD) %</b>
dubitado	4	4	0	100,0	0,0
indubitado	2	2	0	100,0	0,0
<b>Material</b>	<b>semana 4</b>	<b>análogo</b>	<b>Digital (CD)</b>	<b>análogo %</b>	<b>Digital (CD) %</b>
dubitado	7	5	2	71,4	28,6
indubitado	3	2	1	66,7	33,3
<b>Total dubitado al mes</b>	<b>19</b>	<b>12</b>	<b>7</b>	<b>63,2</b>	<b>36,8</b>
<b>Total indubitado al mes</b>	<b>12</b>	<b>11</b>	<b>1</b>	<b>91,7</b>	<b>8,3</b>

<b>Mes 3</b>					
<b>Material</b>	<b>semana 1</b>	<b>análogo</b>	<b>Digital (CD)</b>	<b>análogo %</b>	<b>Digital (CD) %</b>
dubitado	5	5	0	100,0	0,0
indubitado	3	3	1	100,0	33,3
<b>Material</b>	<b>semana 2</b>	<b>análogo</b>	<b>Digital (CD)</b>	<b>análogo %</b>	<b>Digital (CD) %</b>
dubitado	4	2	2	50,0	50,0
indubitado	1	1	0	100,0	0,0
<b>Material</b>	<b>semana 3</b>	<b>análogo</b>	<b>Digital (CD)</b>	<b>análogo %</b>	<b>Digital (CD) %</b>
dubitado	7	5	2	71,4	28,6
indubitado	3	3	0	100,0	0,0
<b>Material</b>	<b>semana 4</b>	<b>análogo</b>	<b>Digital (CD)</b>	<b>análogo %</b>	<b>Digital (CD) %</b>
dubitado	8	7	1	87,5	12,5
indubitado	3	3	0	100,0	0,0
<b>Total dubitado al mes</b>	<b>24</b>	<b>19</b>	<b>5</b>	<b>79,2</b>	<b>20,8</b>
<b>Total indubitado al mes</b>	<b>10</b>	<b>10</b>	<b>1</b>	<b>100,0</b>	<b>10,0</b>

<b>Mes 4</b>					
<b>Material</b>	<b>semana 1</b>	<b>análogo</b>	<b>Digital (CD)</b>	<b>análogo %</b>	<b>Digital (CD) %</b>
dubitado	3	1	2	33,3	66,7
indubitado	1	1	0	100,0	0,0
<b>Material</b>	<b>semana 2</b>	<b>análogo</b>	<b>Digital (CD)</b>	<b>análogo %</b>	<b>Digital (CD) %</b>
dubitado	5	5	0	100,0	0,0
indubitado	3	3	0	100,0	0,0
<b>Material</b>	<b>semana 3</b>	<b>análogo</b>	<b>Digital (CD)</b>	<b>análogo %</b>	<b>Digital (CD) %</b>



dubitado	3	1	2	33,3	66,7
indubitado	2	2	0	100,0	0,0
	<b>semana 4</b>	<b>análogo</b>	<b>Digital (CD)</b>	<b>análogo %</b>	<b>Digital (CD) %</b>
dubitado	9	7	2	77,8	22,2
indubitado	3	3	0	100,0	0,0
<b>Total dubitado al mes</b>	<b>20</b>	<b>14</b>	<b>6</b>	<b>70,0</b>	<b>30,0</b>
<b>Total indubitado al mes</b>	<b>9</b>	<b>9</b>	<b>0</b>	<b>100,0</b>	<b>0,0</b>

## A.2 TABLA DE CANALES DE COMUNICACIÓN

<b>Mes 1</b>					
<b>Material grabado en Formato Análogo y digital</b>	<b>Tipo de Canal</b>	<b>Tipo de canal</b>	<b>Tipo de canal</b>	<b>Tipo de canal</b>	<b>Tipo de canal</b>
Dubitado	Telefonía Fija	Telefonía Celular	Telefonía Satelital	Radio UHF	Grabaciones ocultas
<b>15</b>	7	6	<b>0</b>	2	0
<b>Mes 2</b>					
<b>Material grabado en Formato Análogo y digital</b>	<b>Tipo de Canal</b>	<b>Tipo de canal</b>	<b>Tipo de canal</b>	<b>Tipo de canal</b>	<b>Tipo de canal</b>
Dubitado	Telefonía Fija	Telefonía Celular	Telefonía Satelital	Radio UHF	Grabaciones ocultas
<b>19</b>	5	13	<b>0</b>	1	0
<b>Mes 3</b>					
<b>Material grabado en Formato Análogo y digital</b>	<b>Tipo de Canal</b>	<b>Tipo de canal</b>	<b>Tipo de canal</b>	<b>Tipo de canal</b>	<b>Tipo de canal</b>
Dubitado	Telefonía Fija	Telefonía Celular	Telefonía Satelital	Radio UHF	Grabaciones ocultas
<b>24</b>	8	12	<b>0</b>	2	2
<b>Mes 4</b>					
<b>Material grabado en Formato Análogo y digital</b>	<b>Tipo de Canal</b>				
Dubitado	Telefonía Fija	Telefonía Celular	Telefonía Satelital	Radio UHF	Grabaciones ocultas
<b>20</b>	9	6	<b>0</b>	0	5

## A.3 TABLA DE TIPO DE FORMATO DIGITAL DUBITADO

<b>Mes 1</b>			
<b>Material grabado en formato digital</b>	<b>Tipo de formato</b>	<b>Tipo de formato</b>	<b>Tipo de formato</b>
Dubitado	Wave	mp3	otros
<b>3</b>	<b>1</b>	<b>2</b>	<b>0</b>
<b>Mes 2</b>			
<b>Material grabado en formato digital</b>	<b>Tipo de formato</b>	<b>Tipo de formato</b>	<b>Tipo de formato</b>
Dubitado	Wave	mp3	otros

<b>7</b>	<b>2</b>	<b>5</b>	<b>0</b>
<b>Mes 3</b>			
<b>Material grabado en formato digital</b>	<b>Tipo de formato</b>	<b>Tipo de formato</b>	<b>Tipo de formato</b>
Dubitado	Wave	mp3	otros
<b>5</b>	<b>1</b>	<b>2</b>	<b>2</b>
<b>Mes 4</b>			
<b>Material grabado en formato digital</b>	<b>Tipo de formato</b>	<b>Tipo de formato</b>	<b>Tipo de formato</b>
Dubitado	Wave	mp3	otros
<b>6</b>	<b>4</b>	<b>2</b>	<b>0</b>

## GLOSARIO

*Acústica*: Parte de la física, que trata de la producción, transmisión y recepción de las ondas sonoras.

*Alta Fidelidad (Hi-Fi)*: Término usado para describir la capacidad de un sistema de reproducir en sonido con un alto grado de realismo.

*Alvéolo dental*: Cavidad de los maxilares en que están engastados los dientes en los vertebrados.

*Alvéolo pulmonar*: En los pulmones de los vertebrados, cada uno de los sacos en que terminan las últimas ramificaciones de los bronquiolos.

*Armónicos*: Las ondas que son múltiplos de una determinada frecuencia fundamental y cuyas intensidades suelen ser menores que la frecuencia fundamental.

*Aspiración*: Introducción del aire exterior en los pulmones.

*Efecto peine*: Efecto producido por la combinación de dos señales no sincronizadas y que cuya respuesta en frecuencia se caracteriza por valles y picos de frecuencia que se asemejan a un peine, particularmente visto en un eje de frecuencias lineal.

*DAT*: Digital Audio Tape, cinta digital de audio de alta calidad.

*Difusión*: Proceso de dispersión de la energía radiada de forma uniforme en todas direcciones.

*Direccional*: Que capta o emite más sonido en una dirección que en el resto. Sinónimo de uni-direccional. Opuesto a omnidireccional.

*Distorsión*: Deformación de la onda de imágenes, sonidos o señales durante su propagación

*Distorsión armónica*: Adición de nuevas frecuencias a una señal que están relacionadas a la frecuencia de entrada en forma de múltiplos. Así, una señal frecuencia 400 Hz puede distorsionarse con armónicos de 800 Hz, 1200 Hz, etc.

*Fase*: La relación de tiempo para cada frecuencia entre dos señales. Dos señales idénticas están "en fase" si los picos y valles de la onda ocurren simultáneamente. Están "fuera de fase" si los picos de una onda coinciden con los valles de la otra.

*Fonemas*: Cada uno de los sonidos simples del lenguaje hablado.

*Frecuencia de Resonancia*: Frecuencia en la que un sistema determinado vibra, o entra en resonancia.

*Ganancia*: Cambio en el volumen de una señal. Se habla de ganancia positiva si la señal se hace más grande y negativa si ésta se hace más pequeña. Se expresa habitualmente en dB (por ejemplo +20 dB) o como razón (ganancia unidad, ganancia mitad).

*Glottis*: Abertura superior de la laringe que controla la entrada de aire en la tráquea.

*Grabación*: Proceso de almacenaje de información sonora en soportes específicos.

*Identidad*: Conjunto de rasgos o informaciones que individualizan o distinguen algo y confirman que es realmente lo que se dice.

*Impedancia:* Resistencia que ofrece un elemento al paso de la corriente. Los altavoces suelen ser de 4 a 8 ohmios (aunque es variable con la frecuencia y podría ir de 4 a 60 Ohm).

*Jack:* Conector estándar para auriculares, micrófonos o instrumentos musicales.

*Locutor:* Persona que habla ante el micrófono.

*Mono:* Opuesto a estereo, información de audio a través de un único canal.

*Phantom power:* Alimentación de corriente continua para un micrófono de condensador. Se le dice fantasma porque utiliza el mismo cable de tres conductores que transporta la señal del micrófono, y es neutralizada antes de llegar a la entrada del mezclador o pre-amplificador. Normalmente es de 48 voltios en las aplicaciones más profesionales, aunque a veces también se usan 12 voltios.

*Psicoacústica:* Ciencia que estudia la percepción de los sonidos.

*Rango de frecuencias:* Margen de frecuencias que es capaz de reproducir con respuesta lineal plana un sistema.

*RCA:* En audio, se suelen denominar a un tipo estándar de conectores. Son por ejemplo los que van de un lector CD al amplificador, etc.

*Respuesta en frecuencia:* Parámetro de un dispositivo que indica su capacidad de transmitir frecuencias, se suele usar en fuentes, amplificadores, altavoces.

*Sensibilidad:* Capacidad de un sistema para recibir señales de muy bajo nivel

*Verificación:* Comprobación de la verdad o autenticidad de algo

*XLR*: Conectores de entrada y salida balanceadas de audio tipo profesionales. También conocidos como Canon. Otros tipos de conectores muy utilizados son: BNC, S-Vídeo, RCA, Plug, multi-pin entre otros.

## BIBLIOGRAFÍA

- Carlos Delgado Romero, IDENTIFICACIÓN DE LOCUTORES EN EL ÁMBITO FORENSE.
- Speech Technology Center, CANCELACION DE RUIDO Y DECODIFICACIÓN DE TEXTO DE VOZ DE BAJA CALIDAD.
- J. Gonzáles, T. Cervera, J. Miralles, ANÁLISIS ACÚSTICO DE LA VOZ.
- Néstor Becerra Yoma, Miguel Villar Fernández, VERIFICACIÓN DE IDENTIDAD DE INDIVIDUOS MEDIANTE LA VOZ.
- Federico Miyara, LA VOZ HUMANA.
- Cristian Martínez Bernardo de Quirós, FUNDAMENTOS BÁSICOS DEL RECONOCIMIENTO DE VOZ.
- [www.biometrics.org](http://www.biometrics.org)
- Harris Cyril, MANUAL DE MEDIDAS ACUTICAS Y CONTROL DEL RUIDO.
- Manuel Delgado Cañizares, SISTEMAS DE TELECOMUNICACION E INFORMATICA.
- Jose M Huidobro, MANUAL DE TELECOMUNICACIONES.
- Agustin Alvarez Marquina, FUNDAMENTOS DE LA PRODUCCION Y PRECEPCION DE LA VOZ.
- Federico Miyara, INTRODUCCION A LA PSICOACUSTICA.
- Alberto lopez, INGENIERIA DE ONDAS FORMATOS DE AUDIO DIGITAL.
- Jose Juan Molina, LA ACUSTICA FORENSE.
- Matias Zañartu Salas, APLICACIONES DEL ANALISIS ACUSTICO DE LA VOZ HUMANA.
- Federio Miyara, LA VOZ HUMANA.
- Enrique Vargas, BIOMETRIA LA CLAVE SECRETA.
- Antonio M. Pinado, FUNDAMENTOS DE RECONOCIMIENTO DE VOZ.
- Javier Alejandro Bustos, ESTUDIO DE SISTEMAS DE COMPRESION DE VOZ DIGITAL ORIENTADOS A LA TELEFONIA CELULAR.

- Joseph Picon, FUNDAMENTALS OF SPEECH RECOGNITION
- David Vernet, Xavier Canaleta, LA BIOMETRIA Y SU LEGALIDAD.
- Kittsteiner, Escobar, Algoritmos de compresión  
[http://electricidad.uvipro.cl/teleco/guia\\_teleco/comp-sen.ppt](http://electricidad.uvipro.cl/teleco/guia_teleco/comp-sen.ppt)
- LAIP (Laboratoire d`analyse informatique de la parole), IAFP (Internacional Association for Forensic Phonethics).
- Department of Speech, Music and Hearing (TMH), Royal Institute of Technology (KTH), Stockholm, Sweden
- Groningen Voice Research Lab, Department of Biomedical Engineering, University of Groningen, The Netherlands
- Joaquím Llsterri. Phonetics analysis acoustic (2006).
- Ángel Vega De la Torre. Procesamiento de señales de voz (2002)  
<http://ceres.ugr.es/~atv/Documents/Docencia/voz.ppt>
- Collado, Concejero, Tapias, Castellanos, Hernández. Medida de la inteligibilidad del habla: comparación entre humanos y sistemas automáticos de reconocimiento.  
[http://www.tid.es/documentos/boletin/numero21\\_1.pdf](http://www.tid.es/documentos/boletin/numero21_1.pdf)
- Federico Miyara. El ruido y la inteligibilidad de la palabra (2004).  
<http://www.eie.fceia.unr.edu.ar/~acustica/biblio/inteligibilidad.pdf>
- Juan Carlos Pérez. Reconocimiento de formas en biometría (2005).
- José Juan Molina. La acústica forense (2005).
- NCVS. Department of Speech Music and Hearing Royal Institute of Technology.



