



FACULTAD DE INGENIERÍA Y CIENCIAS AGROPECUARIAS

ESCUELA DE SONIDO Y ACÚSTICA.

ELABORACIÓN DE GUÍA METODOLÓGICA PARA LA APLICACIÓN DE LA
ACÚSTICA FORENSE MEDIANTE SOFTWARE DE RECONOCIMIENTO DE
LA VOZ HABLADA.

Trabajo de Titulación presentado en conformidad a los requisitos establecidos
para optar por el título de
INGENIERO EN SONIDO Y ACÚSTICA

Autores

Jonathan Paúl Calvache Fernández
Luis Francisco Sigcha Guachamin.

Año
2011



FACULTAD DE INGENIERÍA Y CIENCIAS AGROPECUARIAS

ESCUELA DE SONIDO Y ACÚSTICA.

ELABORACIÓN DE GUÍA METODOLÓGICA PARA LA APLICACIÓN DE LA
ACÚSTICA FORENSE MEDIANTE SOFTWARE DE RECONOCIMIENTO DE
LA VOZ HABLADA.

Trabajo de Titulación presentado en conformidad a los requisitos establecidos
para optar por el título de
INGENIERO EN SONIDO Y ACÚSTICA

Profesor Guía:
Marcelo Lazzati

Autores

Jonathan Paúl Calvache Fernández
Luis Francisco Sigcha Guachamin.

Año
2011

DECLARACIÓN DEL PROFESOR GUÍA

“Declaro haber dirigido este trabajo a través de reuniones periódicas con el estudiante, orientando sus conocimientos para un adecuado desarrollo del tema escogido, y dando cumplimiento a todas las disposiciones vigentes que regulan los Trabajos de Titulación.”

Marcelo Darío Lazzati Corellano
Ingeniero en Ejecución de Sonido
C. I.: 171163573-8

DECLARACIÓN DE AUTORÍA DE LOS ESTUDIANTES

“Declaramos que este trabajo es original, de nuestra autoría, que se han citado las fuentes correspondientes y que en su ejecución se respetaron las disposiciones legales que protegen los derechos de autor vigentes.”

Jonathan Paúl Calvache Fernández

180256133-0

Luis Francisco Sigcha Guachamin

171823546-6

AGRADECIMIENTO

Agradecemos profundamente y de todo corazón a las siguientes personas:

A nuestros familiares por sus grandes consejos, su gran apoyo moral y sobre todo por su aporte económico necesario para alcanzar nuestras metas como profesionales en el sonido y la acústica.

Marcelo Lazzati por sus sabios consejos y su excelente tutoría para entregar en buenos términos este trabajo de titulación.

Muchas gracias a todos nuestros profesores por ser una clave importante para el éxito de este trabajo de titulación.

DEDICATORIA

Jonathan Calvache dedica este trabajo a:

Mi madre Sonia Fernández por ser siempre ese apoyo sentimental y mi padre Luis Calvache que en paz descansa por ayudarme a formar la persona feliz que soy, a mis hermanos Ismael y Mishelle porque crecimos juntos y me ayudaron siempre cuando más lo necesité y siempre supieron aconsejarme en los peores momentos de mi vida.

A Carolina Sánchez por ser la mujer que amo y a la que llevo mucho dentro de mi corazón, a mis amigos y a todas aquellas personas que forman parte de mi vida.

Luis Sigcha dedica este trabajo a:

Mi querida madre Yolanda Guachamin, a mis abuelos maternos Francisco y Rosa Elena, a mi hermana Ximena y a mi tío Marco, gracias por estar ahí y apoyarme incondicionalmente en mi formación personal y académica.

Y a todos mis amigos y familiares que me han acompañado en transcurso de mi vida y que de uno u otro modo me han ayudado a ser la persona que soy.

RESUMEN

El presente trabajo es un aporte técnico hacia la labor judicial en el área del peritaje acústico, procurando solventar las deficiencias observadas en el uso de metodologías que muchas veces no son apropiadas para estos tiempos en los que la tecnología ha tenido grandes avances en diversas áreas de la ciencia gracias al análisis y procesamiento de señales por dispositivos digitales, mediante una exhaustiva investigación científica es posible crear una guía metodológica para ser utilizada en Ecuador por peritos en el análisis de voz.

Parte del desarrollo de esta tesis es crear un programa computacional utilizando el lenguaje de programación Max MSP perteneciente a la compañía "Cycling '74", el cual es un entorno gráfico basado en objetos que facilita la comprensión sobre conexiones y funcionamiento del nuevo programa denominado "Static Tone" desarrollada hasta su versión 0.9.

El software es capaz de extraer una serie características estadísticas de la frecuencia fundamental de la voz de una persona en tiempo real y establecer una identificación rápida del individuo frente a un grupo de personas analizadas, el programa computacional también permite establecer una puntuación para un cada uno de los individuos, dicha puntuación puede ser utilizada para validar la propiedad de la voz de un sospechoso.

ABSTRACT

The present work is a technical contribution to the judicial work in the area of acoustic expertise, trying to remedy the deficiencies in the use of methodologies that are often not appropriate for these times where technology has been progress in various biggest areas of science thanks the analysis and signal processing for digital devices, through an exhaustive scientific research is able to create a methodological guide for use in Ecuador by expert in voice analysis.

Part of the develop of this thesis is create a software using the programming language Max MSP belongs to the company "Cycling '74", is an object-based graphical environment that facilitates the understanding of connections and operation of the new program called "Static Tone" developed to the version 0.9.

The software is able to extract some statistical characteristics of the fundamental frequency of the voice from a person in real time and provide rapid identification of the individual within a group of people tested, the program also allows to establish a score for each of the individuals, this score can be used to validate the ownership of the voice of a suspect.

Índice

Capítulo 1. Introducción.....	1
1.1 Objetivos.	1
1.1.1 Objetivo general.....	1
1.1.2 Objetivos específicos.....	2
1.2 Antecedentes.	2
1.3 Justificación.....	3
Capítulo 2. Marco teórico.	4
2.1 Conceptos generales.	4
2.1.1 Peritaje acústico.	4
2.1.2 Fonología y fonética acústica.....	5
2.1.3 La inteligibilidad de la palabra.....	7
2.1.3.1 La inteligibilidad de la palabra en sistemas de sonido.....	7
2.1.4 Procedimiento penal, evidencia y su validez.	8
2.1.4.1 Procedimiento Penal.....	8
2.1.4.2 Evidencia.....	9
2.1.4.3 Validación de evidencia de registro sonoro.	9
2.1.5 Audio digital y analógico.	10
2.1.5.1 Audio analógico.	10
2.1.5.1.1 Transductor electroacústico de entrada:	10
2.1.5.1.2 Pre-amplificador	11
2.1.5.1.3 Especificaciones.....	13
2.1.5.1.4 Rango dinámico.....	13
2.1.5.1.4.1 Ruido de fondo.....	13
2.1.5.1.4.2 Headroom	13
2.1.5.1.4.3 Máximo nivel con distorsión tolerable	13
2.1.5.1.4.4 Relación señal/ruido	14

2.1.5.1.4.5 Nivel nominal.....	14
2.1.5.1.5 Respuesta de frecuencia:.....	14
2.1.5.2 Audio digital.....	14
2.2 Fundamentos del peritaje acústico.	19
2.2.1 Conversión de analógico a digital.....	19
2.2.1.1 Interface de audio.....	19
2.2.1.2 Tarjetas de audio.....	20
2.2.1.3 Ordenador.....	20
2.2.1.4 Software.....	21
2.2.1.5 Requerimientos mínimos y recomendados para un ordenador dedicado al análisis pericial.	22
2.2.2 Edición y procesamiento de audio digital.	22
2.2.2.1 Edición.....	22
2.2.2.2 Procesamiento.....	23
2.2.3 Parámetros de aceptación de audio para su análisis.....	23
2.2.3.1 Frecuencia de muestreo.....	23
2.2.3.2 Profundidad de bits.....	23
2.2.3.3 Cantidad de canales.....	24
2.2.3.4 Formato de archivos.....	24
2.2.4 Biometría.	24
2.2.5 Reconocimiento de locutor por medio de la voz.	25
2.2.5.1 Sistemas de reconocimiento automático de locutor.	25
2.2.6 Definición de parámetros básicos del comportamiento de la voz hablada.	26
2.2.6.1 Naturaleza de la voz.....	27
2.2.6.2 Extracción de características de la identidad del locutor.....	27
2.2.6.3 Técnicas de clasificación.	28

2.3 Herramientas comúnmente utilizadas en el peritaje acústico.....	30
2.3.1 Analizadores.....	30
2.3.1.1 Analizadores de propiedades del audio.....	30
2.3.1.2 Analizadores espectrales.	30
2.3.1.3 Medidores de nivel.	33
2.3.2 Procesadores.	35
2.3.2.1 Ruido.	35
2.3.2.1.1 Métodos de reducción y eliminación de ruido.	36
2.3.2.2 Distorsión.....	37
2.3.2.2.1 Métodos de reducción y eliminación de distorsión.....	38
2.3.3 Optimización y acondicionamiento del audio para analizar.	39
2.3.4 Herramientas de programación.....	40
2.3.4.1 Programación.	40
2.3.4.2 Max MSP.....	40
2.3.4.3 Nociones Básicas de Max MSP.....	40
2.3.4.4 Ventana del Patcher.	41
2.3.4.5 Objetos.....	41
2.3.4.6 Tipos de datos en Max MSP.	42
2.3.4.7 Conexiones.	43
2.3.4.8 Sub patcher.....	43
Capítulo 3. Análisis de la situación actual.	44
3.1 Descripción de la situación actual ecuatoriana en el peritaje acústico.....	44
3.1.1 Investigación y acerca del peritaje acústico.	44
3.1.2 Procesos realizados actualmente en el peritaje acústico.	44
3.1.3 Tiempos empleados actualmente en la realización de peritaje de voces.....	46

Capítulo 4. Descripción de la situación propuesta.....	47
4.1 Mejoras implementadas para una nueva metodología.....	47
4.2 Creación de una guía metodológica para realizar peritajes acústicos de reconocimiento de voz.....	48
4.3 Guía de procedimientos para la elaboración del informe de peritajes acústicos de reconocimiento de voz.....	52
Capítulo 5. Diseño del software de reconocimiento de voz hablada.....	56
5.1 Creación de un software de identificación de locutor.....	56
5.1.1 Diseño del algoritmo de identificación de locutor por la voz.	56
5.1.2 Determinación de la frecuencia fundamental.....	56
5.1.3 Distribución estadística de la frecuencia fundamental.	57
5.2 Programación de Static Tone en Max MSP.....	57
5.2.1 Etapa de la entrada de audio.	57
5.2.1.1 Conexiones de la etapa de entrada de audio.....	58
5.2.2 Etapa de análisis de la frecuencia fundamental:	60
5.2.2.1 Proceso de creación del sub patch de la etapa de análisis estadístico de la frecuencia fundamental.	61
5.2.3 Etapa de almacenamiento de información:.....	63
5.2.3.1 Creación de la etapa de almacenamiento de información.....	63
5.2.4 Comparación de similitud:.....	65
5.2.4.1 Conexiones de la etapa de comparación.	65
5.2.5 Interfaz de usuario.....	70
5.2.5.1 Proceso de creación de la interfaz de usuario.	70
5.2.6 Exportación del patcher como una aplicación independiente.....	71
5.2.6.1 Proceso de configuración y exportación:.....	72
5.3 Pruebas de funcionamiento.....	72
5.2.1 Presentación de datos para análisis.	72

5.2.1.1 Keele Pitch Database.....	72
5.2.1.2 Resultados del de la estimación del pitch y comparación con el programa PRAAT.....	73
5.2.2 Análisis de resultados.....	75
Capítulo 6. Análisis costo/beneficio.....	77
6.1 Elaboración del informe económico para el desarrollo, aplicación y ejecución del software propio con hardware adquirido.....	77
6.1.1 Cotizaciones.....	77
6.2 Presupuesto Referencial.....	86
6.3 Rentabilidad del proyecto aplicado a Ecuador.	90
Capítulo 7. Proyecciones.	94
Capítulo 8. Conclusiones y recomendaciones.	96
8.1 Conclusión general.	96
8.2 Conclusiones específicas.....	96
8.3 Recomendaciones.	98
Capítulo 9. Bibliografía.....	100
9.1 Libros.	100
9.2 Artículos y notas técnicas	100
9.3 Artículos de Internet.....	102
Capítulo 10. Anexos.	103

Capítulo 1. Introducción.

El presente trabajo investigativo del reconocimiento de la voz humana pretende mostrar la situación actual del Ecuador mediante un análisis detallado y exhaustivo de las actuales falencias en el peritaje acústico del audio, para posibles correcciones y mejoras para hacer más eficiente este proceso.

La importancia de este trabajo investigativo es la de comprender, analizar y mejorar en lo posible las metodologías aplicadas en países vecinos, para desarrollar y aplicar a nuestra realidad nacional.

La acústica forense es una subdivisión de la criminalística que se encarga de dar apoyo logístico a la labor policial en el esclarecimiento de hechos con la validación científica de evidencia para los tribunales.

Actualmente se poseen herramientas muy sofisticadas las cuales posibilitan un análisis eficiente del audio, ya que han habido avances científicos tanto en el procesamiento digital de señales, el desarrollo de software y la electroacústica, que facilitan enormemente la implementación de este nuevo tipo de servicio para el país.

Este trabajo abarca en su mayoría el comportamiento, ejecución, procesamiento y análisis de la voz humana hablada como tal, mencionando conceptos importantes como el estudio de las formantes, fonética, procesos penales vigentes en el Ecuador, así como su aplicación en una guía metodológica y principios acústicos de programación.

1.1 Objetivos.

1.1.1 Objetivo general.

Realizar una guía metodología para la realización de peritajes acústicos de reconocimiento de personas por medio de la comparación de audio muestreado, con la respectiva muestra de análisis para determinar un porcentaje de compatibilidad entre ellas, utilizando un software.

1.1.2 Objetivos específicos.

- ◆ Recopilar información relacionada a las distintas metodologías utilizadas para el reconocimiento de voz.
- ◆ Diseñar una metodología acorde a la realidad nacional para la realización de peritajes técnicos de reconocimiento de voz hablada, también esquematizar los procesos legales que deben realizar las personas, que requieran un trabajo de peritaje para usarlo como evidencia válida en un juicio.
- ◆ Desarrollar un software de reconocimiento de voz por comparación de las propiedades de audio anteriormente establecidas.
- ◆ Evaluar el software por medio de la comparación de muestras de audio con un banco de voces.

1.2 Antecedentes.

Desde los principios de la humanidad, el hombre ha necesitado del uso del DERECHO como tal, dado que todas las personas necesitan vivir con leyes para que así sea más justa la vida para los seres humanos.

Para inculpar a una persona que ha cometido un delito, es necesario el uso de evidencia en su contra para que este pueda ser declarado culpable. Pues como el mismo derecho lo dice “TODO HOMBRE ES INOCENTE HASTA QUE SE DEMUESTRE LO CONTRARIO”.

El campo de la acústica se encarga del estudio de la generación, propagación, y comportamiento de las ondas sonoras en un espacio físico.

La acústica forense se refiere a la aplicación práctica y científica dentro del proceso legal, convirtiéndose en una materia multidisciplinaria que cubre áreas como la acústica, la fonética, la lingüística, la ingeniería, electrónica, matemáticas, etc.

El peritaje aparece como un estudio realizado por un perito que es un experto que es nombrado por el Juez para la realización de exámenes o estudios, los que son utilizados como pruebas para el esclarecimiento de hechos o circunstancias, que requieran de conocimientos técnicos y/o científicos.

1.3 Justificación.

Su importancia es la creación de un sistema rápido y eficaz en peritaje de audio, puesto que es la primera vez en el país que se cuenta con profesionales y pioneros en el área acústica que son capaces de desarrollar técnicamente trabajos como éste, ya que se cuenta con los suficientes conocimientos y tecnología.

El aporte de este trabajo para la sociedad es que ayudará a la eficiencia del sistema judicial en país, mientras que por otro lado se podrían generar nuevas fuentes empleo para especialistas en el área de la acústica, pues la demanda de este tipo de servicios es alta en comparación de la oferta actual del Ecuador, debido al aumento de la inseguridad.

Capítulo 2. Marco teórico.

2.1 Conceptos generales.

2.1.1 Peritaje acústico.

Las palabras peritaje acústico hacen referencia a la ejecución de la acústica forense en el esclarecimiento de un delito.

Etimológicamente la palabra PERITO según la real academia de la lengua quiere decir:

“Persona que, poseyendo determinados conocimientos científicos, artísticos, técnicos o prácticos, informa, bajo juramento, al juzgador sobre puntos litigiosos en cuanto se relacionan con su especial saber o experiencia”¹.

En otras palabras el perito, es aquella persona que está en sus capacidades profesionales, académicas y artísticas de llevar a cabo un trabajo investigativo que es reconocido como fuente confiable de validación de hechos.

Si a esto se le suma la palabra ACÚSTICA que la real academia de la lengua dice:

“Parte de la física que trata de la producción, control, transmisión, recepción y audición de los sonidos.”²

Se obtiene como resultado que peritaje acústico es un trabajo profesional académico y artístico relacionado con la producción, control, transmisión, recepción de los fenómenos sonoros. Los cuales pueden ser utilizados en asuntos legales.

1 “Diccionario de la lengua española” - Vigésima segunda edición, 2008,
Web http://buscon.rae.es/draeI/SrvltConsulta?TIPO_BUS=3&LEMA=perito

2 “Diccionario de la lengua española” - Vigésima segunda edición,
Web <http://buscon.rae.es/draeI/SrvltGUIBusUsual?LEMA=acústico&origen=RAE>

2.1.2 Fonología y fonética acústica.

Se entiende por fonética al estudio de la ejecución física y acústica del lenguaje como tal, mientras que la fonología se encarga del estudio de la imagen mental que el ser humano entiende el lenguaje oral.

La fonología está compuesta por “*fonemas*” que son percepciones mentales del sonido, en lugar de las percepciones reales del sonido a la que se les denomina “*fonos*”.

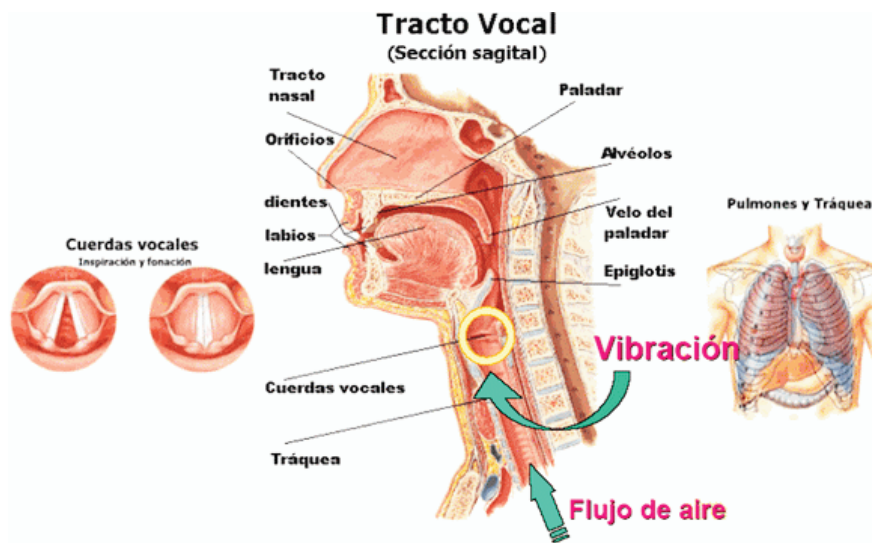
Los fonemas se clasifican en dos grandes grupos de estudio:

- ♦ El primero son las vocales debido a que su emisión no presenta ningún obstáculo a lo largo del aparato fonador.
- ♦ El segundo son las consonantes que si presentan algún obstáculo a lo largo del aparato fonador.

Los elementos articulatorios del aparato fonador son los siguientes:

- ♦ **Las cuerdas vocales:** Las cuerdas vocales humanas son directamente responsable de la producción de la voz. No tienen forma de cuerda, sino que se trata de una serie de repliegues o labios membranosos. Cuando estas vibran se obtiene el tono u sonido de la voz resultante.
- ♦ **La cavidad bucal:** Es la cavidad en donde se puede articular las palabras y se direcciona el sonido, es aquí en donde se encuentran los dientes, la lengua, el paladar, etc.
- ♦ **Los labios:** Son el límite exterior del aparato fonador, son muy importantes a la hora de articular palabras puesto que permiten o limitan el flujo de aire desde los pulmones hacia el exterior.
- ♦ **La cavidad nasal:** Es aquella encargada de limpiar el aire proveniente del exterior, también funciona como un resonador acústico y es aquí por donde se generan las letras (M, N, Ñ).

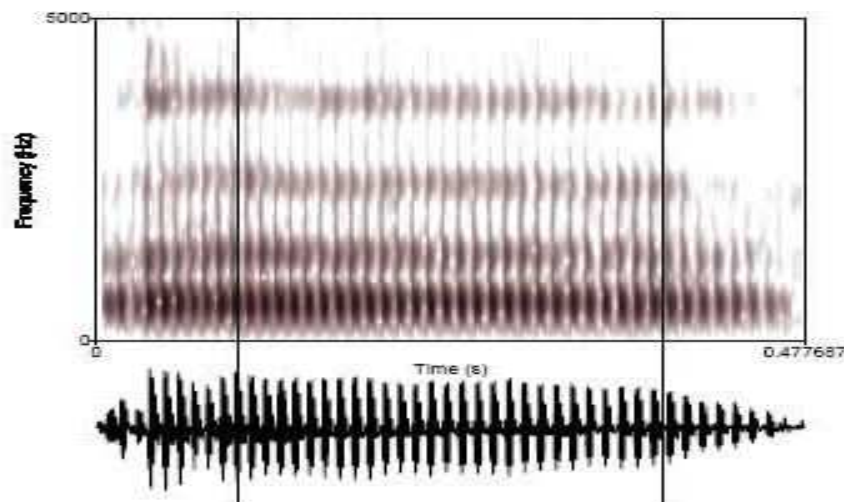
Figura 2.1: Tracto vocal.



Fuente: www.musicared.com/wp-content/uploads/2009/12/fonador.gif

Formantes: Un formante es el pico de intensidad en el espectro de un sonido; se trata de concentración de energía (amplitud de onda) que se da en determinada banda de frecuencia. Estos picos son el producto de las resonancias por filtrado que se producen en el tracto vocal por la configuración del aparato fonador.

Figura 2.2: Sonograma y forma de onda de una muestra de audio.



Fuente: <http://mingaonline.uach.cl/scielo.php?pid=S0718-025X>

2007000100004&script=sci_arttext

Los formantes permiten distinguir los diferentes tipos de sonidos que genera el habla humana, también sirven para los sistemas de reconocimiento de voz y las transposiciones de altura del audio digital.

Como se puede observar en el espectrograma de la figura 2.2 existen cuatro concentraciones de energía representadas en forma horizontal por un color más oscuro, gráficamente es aquí donde se encuentran representadas las formantes.

2.1.3 La inteligibilidad de la palabra.

Existen factores tales como el ruido, el contenido espectral y la duración pueden afectar la percepción del habla. Como la palabra es básicamente información codificada por el uso de vocales y consonantes, la disminución de la inteligibilidad se asocia a la pérdida parcial o total de dicha codificación.

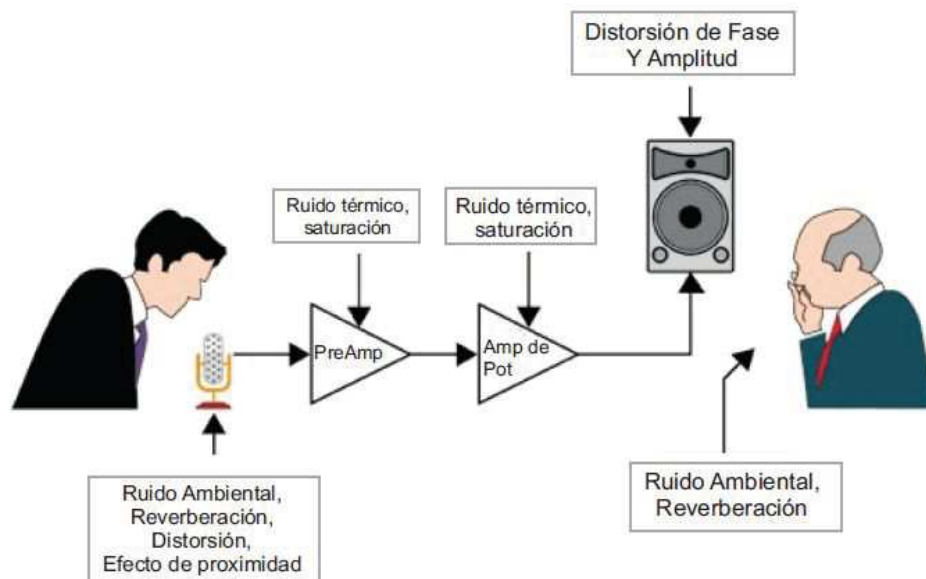
Cuando la presencia de ruidos de fondo es muy alta pueden enmascarar el habla, así como por la posición del oyente con respecto a la fuente, por la articulación del orador y la agudeza auditiva del oyente. Por lo tanto la inteligibilidad de la palabra puede verse afectada en cualquiera de las tres etapas de la comunicación ya sea en el orador, en el medio o en el receptor.

2.1.3.1 La inteligibilidad de la palabra en sistemas de sonido.

La inteligibilidad puede verse afectada por cualquiera de sus etapas en la cadena electro-acústica, dependiendo de sus componentes y su calidad.

Para el caso de micrófonos y altavoces es importante la consideración de su respuesta de frecuencia ya que puede comprometer la inteligibilidad al no respetar el ancho de banda del habla.

Figura 2.3: La inteligibilidad de la palabra en sistemas de sonido.



Fuente: www.cetear.com/InteligibilidaddelHablaParte1.pdf

2.1.4 Procedimiento penal, evidencia y su validez.

2.1.4.1 Procedimiento Penal.

Se conoce como procedimiento penal al protocolo de una sentencia establecida por el código penal, el mismo se encarga de investigar, identificar y sancionar a los acusados en caso de ser sentenciados como culpables.

El código vigente en el Ecuador "CÓDIGO DE PROCEDIMIENTO PENAL 2000" de 13 de Enero del 2000, publicado en el "Registro Oficial Suplemento 360", el cual está compuesto por 430 artículos.

Código penal: El código penal se encarga de estipular infracciones y definiciones legales para ser aplicadas durante un juicio.

El código vigente en el Ecuador es "EL CÓDIGO PENAL ECUADOR" publicado en el año 1999 consta de 631 artículos.

2.1.4.2 Evidencia.

Un sinónimo de la palabra evidencia es “Prueba” la cual es parte implícita de la escena del crimen, la cual puede ser valedera o rechazada según el Juez.

Una evidencia puede clasificarse en distintas etapas como:

- ♦ **Confesión de parte:** Es la evidencia que está dictaminada al declarar por una de las partes su incriminación directa o indirecta de un hecho.
- ♦ **Prueba testimonial:** Es aquella evidencia que implica el testimonio de terceras personas.
- ♦ **Documentos públicos o privados:** Es aquella evidencia que se encuentra registrada en documentos ya sean de carácter público como privado, por ejemplo publicaciones en periódicos, revistas, etc.
- ♦ **Informes de peritos:** Es la evidencia basada en el análisis científico de elementos circunstanciales hallados en escena.
- ♦ **Inspección personal del tribunal:** Es aquella evidencia que es proporcionada por el mismo tribunal, al designar personal calificado para la investigación de individuos sospechosos.
- ♦ **Presunciones establecidas por la ley o jurisprudencia:** Es aquella evidencia marcada por supuestos de una investigación, las cuales son definidas por el código penal.

2.1.4.3 Validación de evidencia de registro sonoro.

Para que una evidencia sea aceptada como válida se necesita de la participación de dos entes, el primero es el perito y el otro es el juez. Los aspectos más importantes para determinar que tan fidedigna es la procedencia y las cualidades de dicha evidencia, tales como: la fecha, hora, lugar de la grabación, persona que grabó el audio, personas a las que se grabó, el dispositivo de grabación, el formato de grabación, las condiciones del audio. Para poder determinar su validez se hace necesario de un estudio completo de

la evidencia, para poder iniciar un informe pericial, el mismo que una vez terminado es enviado al Juez, en donde él determinará su aceptación, según las conclusiones del informe pericial.

2.1.5 Audio digital y analógico.

2.1.5.1 Audio analógico.

Se define como audio a la representación eléctrica del sonido, en donde las variaciones de presión acústica son análogas a las variaciones de voltaje eléctrico en un conductor.

2.1.5.1.1 Transductor electroacústico de entrada: Es un dispositivo electro-mecano-acústico que tiene por objetivo el transformar las variaciones de presión sonora en variaciones de voltaje eléctrico en un conductor.

Electromecánicamente los transductores se clasifican en tres tipos entre:

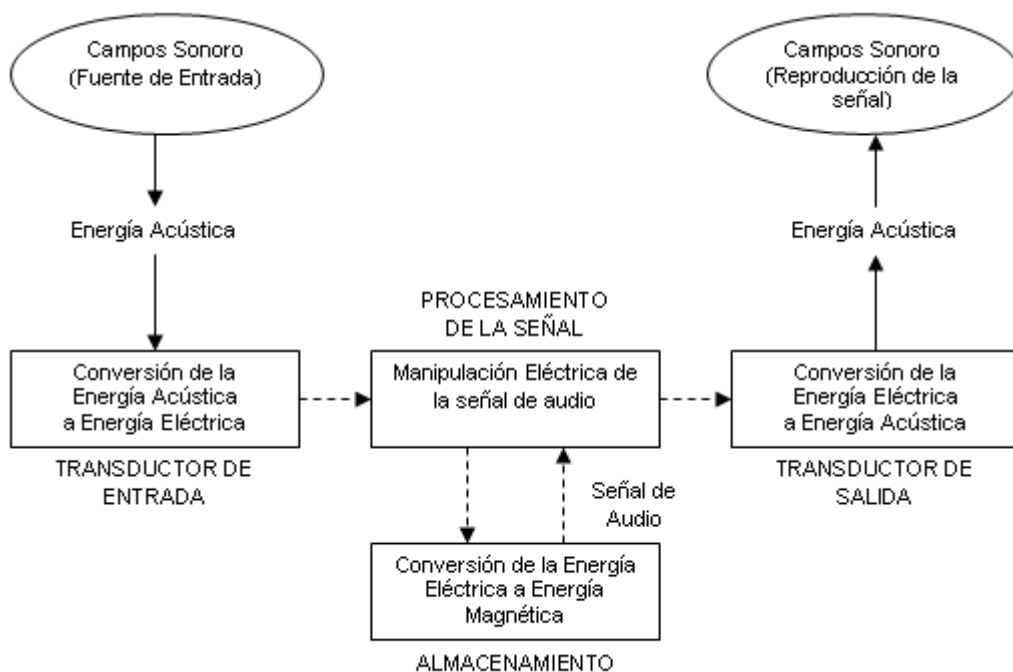
- ♦ **Micrófono electrostático:** Este micrófono se caracteriza por ser de principio activo eso quiere decir que necesitan de *alimentación fantasma* para poder funcionar, estos micrófonos generan cambios de voltaje cuando el diafragma se mueve por los cambios de presión liberando así proporcionalmente el voltaje acumulado en su capacitor, como ejemplo de estos micrófonos se tiene a los de condensador, a los de cinta, etc.
- ♦ **Micrófono electromagnético:** Son aquellos micrófonos que generan sus cambios de voltaje en una bobina móvil que interacciona con un entrehierro o imán permanente, como ejemplos de este tipo de micrófono se tiene a los de bobina móvil.
- ♦ **Micrófono electroresistivo:** Este micrófono se caracteriza por ser uno de los más antiguos, como ejemplo se tiene a los micrófonos de carbón, y estos funcionaban al tener una resistencia variable proveniente de la interacción de una placa como membrana oscilante y partículas de carbón como principio de *transducción*.

El micrófono al recibir un frente de onda, este excita la placa y las partículas de carbón generando cambios de resistencia, que al interactuar con una corriente continua se obtiene cambios de voltaje; cabe acotar que la respuesta de frecuencia de estos micrófonos oscila entre los 200 Hz y 3 kHz.

2.1.5.1.2 Pre-amplificador: Es necesario saber que la señal del transductor posee un nivel eléctrico bajo en términos de corriente y de voltaje, este dispositivo se encarga de elevar proporcionalmente el voltaje que entrega el micrófono el mismo que se encuentra en el rango de los 100 mili-volts para llevarlos a un nivel estandarizado denominado de línea que se comprende de 0,1 a 10 volts.

La señal pre-amplificada puede destinarse ya sea para dispositivos de grabación y/o procesamiento, o para su posterior etapa en reproducción por medio de amplificadores de potencia y altavoces, por ejemplo:

Figura 2.4: Diagrama de flujo de audio analógico.

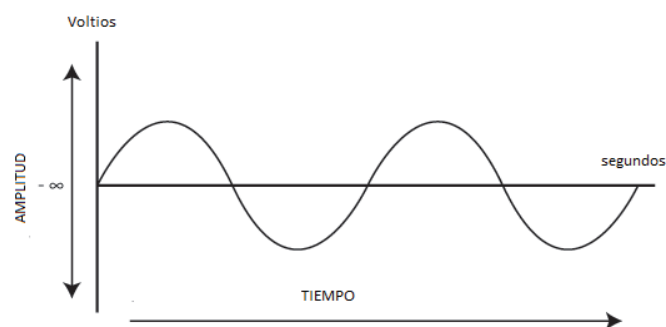


Fuente: Autores.

Se debe también considerar que una cadena electroacústica profesional puede considerar las etapas de captación tales como la microfónica, medusas, pre-amplificación, etc. Mientras que en su etapa de monitoreo hay amplificación de potencia, monitoreo de campo cercano y lejano, audífonos; posteriormente en la etapa de grabación, se considera edición, procesamiento, automatización, sincronismo, mezcla y masterización.

Al audio se lo puede diagramar de la siguiente manera:

Figura 2.5: Gráfico de voltaje vs tiempo.



Fuente: <http://www.indecentsmusic.com/blog/wp-content>

/uploads/2009/06/amplitude-and-frequency.png

Al audio analógico se lo define como a una señal continua de voltaje en función del tiempo, estas señales son de naturaleza senoidal y se comportan en base a la ecuación de onda simple descrita en la ecuación (2.1).

$$x = A \text{ sen } (\omega t + \varphi) \quad (2.1)$$

Donde: x es la ecuación de onda,

A es la amplitud de la onda,

ω es la velocidad angular [radianes],

t es el tiempo en [segundos],

φ es el ángulo de fase [radianes].

Los procesadores de audio basan su funcionamiento en la modificación de algunas de las componentes anteriormente mencionadas.

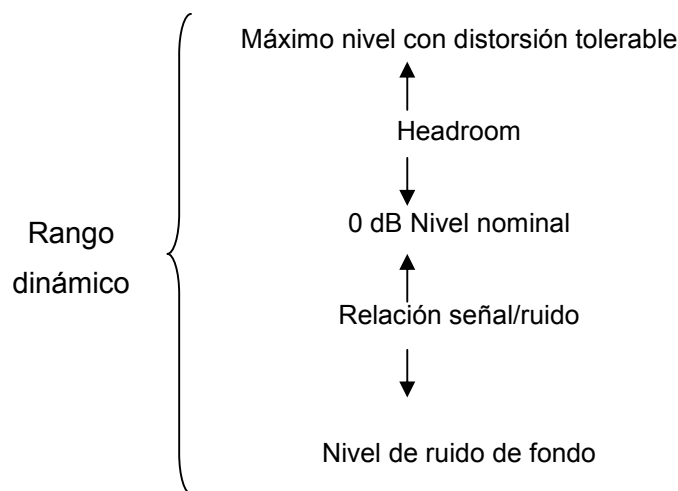
2.1.5.1.3 Especificaciones: Las grabaciones analógicas poseen un alto nivel de ruido, bajos niveles máximos sin distorsión y su respuesta de frecuencia es limitada tanto en los medios de almacenamiento como en los medios de transmisión.

2.1.5.1.4 Rango dinámico: Es una especificación asociada a una señal que representa la relación entre el máximo y el mínimo nivel de la señal, se expresa en dB definido en la ecuación (2.2).

$$RD = 20 \log_{10} (\text{Máximo nivel} / \text{Mínimo nivel}) \quad (2.2)$$

Donde: RD es el Rango dinámico [dB].

Figura 2.6: Niveles en el rango dinámico de una señal de audio.



Fuente: Autores

2.1.5.1.4.1 Ruido de fondo: Se conoce como ruido de carácter aleatorio de fondo aquel sonido presente componentes de circuitería de naturaleza reactiva.

2.1.5.1.4.2 Headroom: Es aquel rango comprendido entre el nivel nominal y el máximo nivel con distorsión tolerable.

2.1.5.1.4.3 Máximo nivel con distorsión tolerable: El rango dinámico como tal considera al máximo nivel con distorsión tolerable ejemplo: (+10dB con 0,01% THD o *distorsión total armónica* según componentes del fabricante), cuando la amplitud de la señal rebasa los límites de distorsión tolerables se genera una saturación por recorte de onda.

2.1.5.1.4.4 Relación señal/ruido: Es aquel rango comprendido entre el nivel nominal y el ruido de fondo, es en esta sección donde por defecto no se debe deberá registrar el audio analógico por su débil amplitud.

2.1.5.1.4.5 Nivel nominal: Se conoce como nivel nominal a aquella carga eléctrica ideal entre dispositivos electrónicos.

2.1.5.1.5 Respuesta de frecuencia: La respuesta de frecuencia da a conocer cuantitativamente el comportamiento energético en un determinado ancho de banda establecido por las *frecuencias de corte*.

La respuesta de frecuencia está establecida por estudios estadísticos y pruebas de campo, por ejemplo de respuesta de frecuencia de los micrófonos que están dictaminados por el fabricante en pruebas de laboratorio que en algunos casos pueden captar frecuencias ajenas al rango de la *audición humana*.

2.1.5.2 Audio digital.

Se define como audio digital a la discretización de una señal continua de audio analógico, mediante frecuencias de muestreo establecidas por el teorema de Nyquist que dice: que para poder reproducir de forma fiel una forma de onda, deberá muestrearse dicha señal al menos al doble de la máxima frecuencia que se desea reproducir según la ecuación (2.3).

$$F_s \geq 2 * F_n \quad (2.3)$$

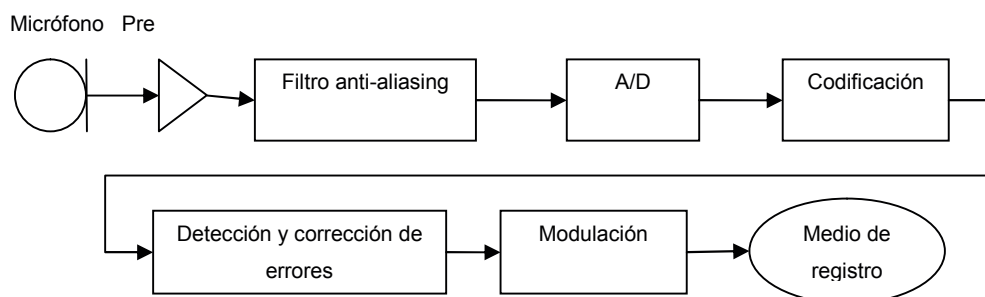
Donde: F_s es la frecuencia de muestreo [Hz],

F_n es la frecuencia de Nyquist [Hz].

Las etapas del audio digital están determinadas por los siguientes diagramas en bloques.

Etapa de grabación:

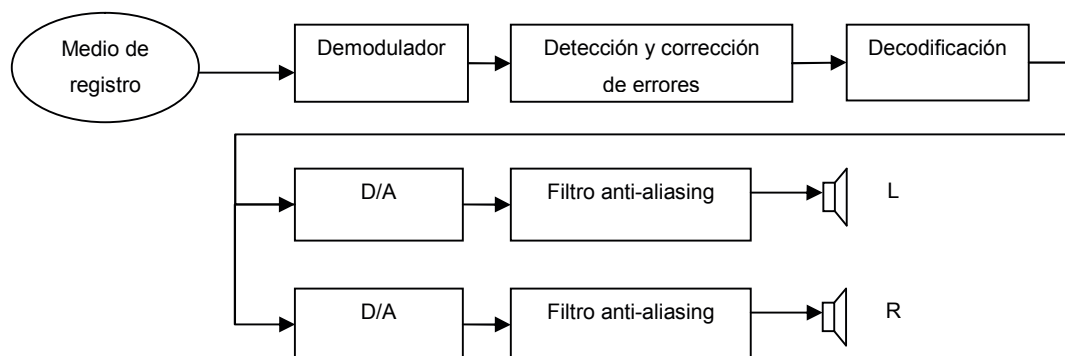
Figura 2.7: Cadena de componentes de grabación digital.



Fuente: Autores

Etapa de reproducción:

Figura 2.8: Cadena de componentes de reproducción digital.



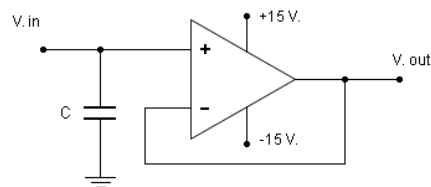
Fuente: Autores

- ♦ **Filtro anti-aliasing:** Es un tipo filtro pasa bajos de pendiente abrupta, que tiene como objetivo filtrar todas las frecuencias ajenas al rango de frecuencias de 0 Hz a 20 kHz y para evitar distorsiones por intermodulación de frecuencias, cuya distorsión aparece por la presencia de dos frecuencias de gran amplitud distantes en el espectro de frecuencia, que generan como efecto la presencia de nuevos armónicos

que alteran la forma de onda original al sumarse por defecto a la misma. Para grabación la frecuencia de corte es de 44.1 kHz mientras que para la reproducción es de 22.05 kHz.

- ♦ **Sample & Hold (muestra y retención):** Es un circuito destinado a retener el valor de voltaje en la entrada por un lapso de tiempo (periodo de muestreo) determinado por un reloj digital muy preciso llamado “*clock*”. Este circuito está compuesto por un amplificador operacional y un capacitor.

Figura 2.9: Circuito de muestra y retención.

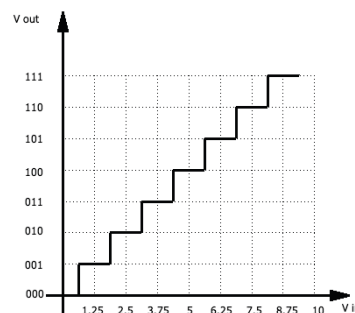


Fuente: Autores

- ♦ **Cuantización:** Es el proceso de aproximación de los voltajes instantáneos de cada una de las muestras al dominio de la amplitud; una mayor profundidad de bits da como resultado una mejor representación de la amplitud dentro del rango audible.

Como ejemplo para una cuantización a 3 bits de profundidad se tiene la siguiente curva de transferencia:

Figura 2.10: Curva de transferencia de 3 bits.



Fuente: Autores

La cuantización es codificada de manera serial como se muestra en el ejemplo.

Para el dominio digital el rango dinámico está determinado por la ecuación (2.4).

$$RD = 20 \log_{10} (2^n) \quad (2.4)$$

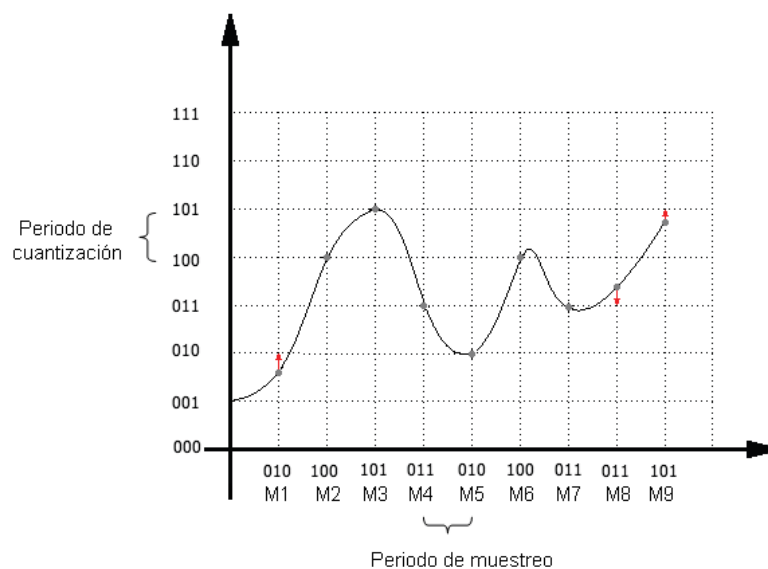
Donde: RD es el rango dinámico,

n es la cantidad de bits de la cuantización.

Los pasos de cuantización son directamente proporcionales al rango dinámico, es decir a mayor profundidad de bits mayor será el rango dinámico posible para la grabación.

Al momento de muestrear un voltaje instantáneo puede o no coincidir en un paso de cuantización, por lo tanto se debe aproximar dicho voltaje al paso de cuantización más cercano.

Figura 2.11: Proceso de muestreo, cuantización
y aproximación de voltajes.



Fuente: Autores

- ♦ **Error de cuantización:** Las señales de bajo nivel generan un error de cuantización, debido a la interpolación no exacta de los pasos de cuantización, es decir modifica la forma original de la señal, para obviar este problema se recomienda utilizar una mayor profundidad de bits y asegurarse de que la señal tenga idealmente un nivel nominal. Sin embargo existe un método que psicoacústicamente mejora la percepción del sonido llamado "dither".
- ♦ **Dithering:** Es el proceso por el cual se le suma a la señal de audio un ruido de naturaleza variable similar al ruido blanco, que posee una amplitud igual al bite más significativo y una frecuencia mayor al periodo de muestra seleccionado, en la práctica la forma de onda puede ser similar a otras formas de onda conocidas como la senoidal, cuadrada, triangular, etc.

Es importante mencionar que en la etapa de reproducción este proceso siempre es aplicado, también se utiliza en los procesos de re-cuantización al disminuir su profundidad de bits.

- ♦ **Codificación:** Es la etapa de organización de la información digital previamente antes del almacenamiento o transmisión de la señal. En esta etapa es posible anexar cualquier tipo de información de interés, dependiendo del sistema de codificación esta información adicional puede ser sincronismo, detección y corrección de errores, *address*, datos de propiedad intelectual, etc.
- ♦ **Detección y corrección de errores:** Es parte de la codificación se utiliza varios métodos dependiendo del sistema de codificación. Para la detección de errores se utiliza los métodos se basa en el método de redundancia de la información y el uso de bits de paridad. La corrección se da por el método de interpolación o competición automática de muestra y en el peor de los casos el muteo de la muestra de audio perdida.

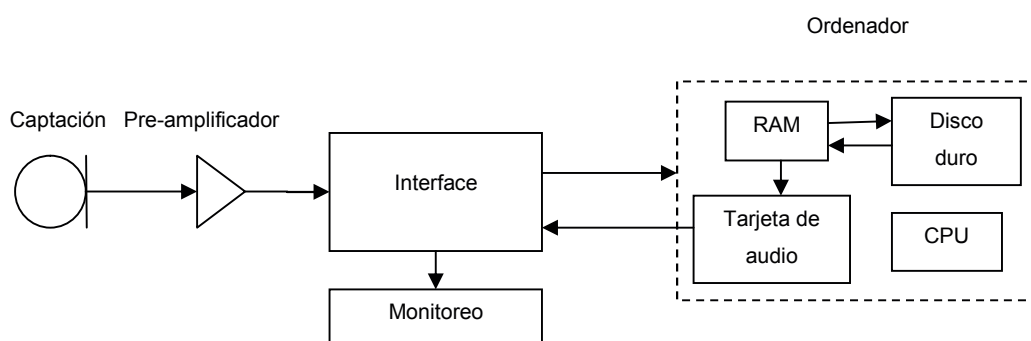
- ✦ **Características:** Las grabaciones digitales son menos susceptibles al ruido, poseen mayor rango dinámico según la profundidad de bits usada, mayor respuesta de frecuencia, el acceso a la información se la realiza de forma no lineal, la edición y el monitoreo son de tipo no destructivo; el procesamiento de la información puede ser realizado en tiempo real, es mucho más fiel y de menor costo que en el dominio analógico.

2.2 Fundamentos del peritaje acústico.

2.2.1 Conversión de analógico a digital.

En los apartados anteriores se mencionó sobre el audio analógico y digital y a continuación se presentará el encadenado físico básico necesario poder realizar procesamiento sobre señales de audio en un computador.

Figura 2.12: Cadena básica de componentes en un sistema de procesamiento de señales en peritaje forense.



Fuente: Autores

2.2.1.1 Interface de audio: Es un dispositivo donde se puede hallar todas las entradas y salidas de audio analógico y digital, para poder insertar una señal de audio preamplificada a nivel de línea para ser analizada o procesada dentro de un ordenador.

Las interfaces de gama baja poseen preamplificadores incorporados, sin embargo las interfaces de gama profesional y la interfaces internas no las poseen.

Existen distintas clases de interfaces:

- ♦ **Internas:** Son aquellas interfaces que se conectan directamente a la tarjeta madre de un computador. Para estas interfaces en especial hay que tener en cuenta que son más sensibles al ruido inducido por *campos electromagnéticos* sobretodo en sus convertidores A/D y D/A, de esta manera no son consideradas como interfaces profesionales.
- ♦ **Externas:** Son aquellas interfaces que se conectan por medio de conectores externos por algún puerto protocolario de transmisión de datos.

2.2.1.2 Tarjetas de audio: Es un conjunto de circuitos electrónico compuesto por procesadores especializados para el manejo de audio digital, *buffers*, *sintetizadores* en algunos casos. Para su funcionamiento requieren del control de programa informático denominado controlador.

La comunicación entre la tarjeta y el ordenador se la puede realizar por diferentes puertos entre los más importantes están: *PCI*, *PCI Express*, *USB*, *Firewire*, *PCMCIA*, etc.

El circuito procesador de audio es conocido como DSP (Digital Signal Processing) o en español "procesadores de señales digitales", y son especializados para procesar audio discreto evitando el uso de la CPU del ordenador.

2.2.1.3 Ordenador: Es una máquina electrónica compuesta de diversos dispositivos tales como: sistemas de control, monitoreo visual, almacenamiento y procesamiento del audio digital. Se utiliza para recibir instrucciones y procesar datos de manera digital para convertirlos en información útil.

Las partes necesarias para el funcionamiento de un ordenador son:

- ♦ **Teclado:** Es un periférico que en su interior posee interruptores tipo pulsador en cuyas teclas van escritos comandos, letras, números, y signos. Se utiliza para ingresar instrucciones al ordenador.

- ◆ **Ratón:** Es un dispositivo periférico de entrada cuya función es la de desplazar el cursor y poder seleccionar objetos en pantalla.
- ◆ **Pantalla:** Este dispositivo es utilizado como monitor de imagen, establece comunicación visual entre computador y el usuario. Es útil para el análisis y la edición gráfica de audio ya que permite visualizar formas de onda, espectrogramas, etc.
- ◆ **CPU:** Se conoce a la CPU, como unidad central de proceso o UCP en idioma español, se trata de un conjunto de circuitos microscópicos instalados sobre una placa que interpretan y ejecutan instrucciones del usuario. La CPU se ocupa del control y el proceso de datos en las computadoras. Una mayor velocidad de CPU permite ejecutar los programas más rápidamente.
- ◆ **Memoria RAM:** La memoria de acceso aleatorio o RAM, es una memoria de tipo no permanente esto quiere decir que todos los datos que han sido almacenados serán eliminados al quitar su alimentación eléctrica, está basada en semiconductores que puede ser leída y escrita por el microprocesador u otros dispositivos de hardware. Su acrónimo en inglés es *Random Access Memory*. La RAM está encargada de enviar instrucciones al CPU y guardar los resultados obtenidos de esta. Mayor cantidad de memoria RAM en un ordenador permite ejecutar una mayor cantidad de programas al mismo tiempo.
- ◆ **Disco duro:** Esta memoria se caracteriza por ser de tipo permanente esto quiere decir que todos los datos que han sido almacenados no serán eliminados después de quitar la alimentación eléctrica. Es aquí donde se registran todos los datos tales como programas computacionales, audio, sistema operativo, archivos de texto, etc. Existen de varios tipos y capacidades de almacenamiento, su unidad de medida es el Bit con sus respectivos múltiplos. Una mayor cantidad de disco duro permite registrar una mayor cantidad de datos permanentemente.

2.2.1.4 Software: Son programas computacionales encargados del soporte lógico necesarios para la realización de tareas en un sistema de computo.

Los softwares básicos necesarios para el funcionamiento de un ordenador son el *BIOS, sistema operativo y controladores*.

2.2.1.5 Requerimientos mínimos y recomendados para un ordenador dedicado al análisis pericial.

Se sugiere que para la instalación de programas computacionales aplicados al análisis de voz, se recomienda que posea características mínimas de hardware tales como: 1 GB de memoria RAM, 100 GB de espacio en disco duro, procesador de un núcleo con 1.5 GHz de velocidad, tarjeta gráfica, pantalla de 14", lector de CD, teclado, ratón, puertos de conexión externos tales como: USB 2.0, Firewire 400 o PCI.

Se recomienda que posea características mínimas de software son: sistema operativo Windows XP o Mac OSX, controladores de hardware.

Como requisitos recomendados se sugiere que equipo posea las siguiente características: 4 GB de memoria RAM, dos disco duros de 500 GB de espacio uno para el sistema operativo con sus programas y otro para biblioteca de audio, procesador de dos núcleos con 2 GHz de velocidad, tarjeta gráfica, pantalla de 17", lector de DVD, teclado, ratón, puertos de conexión externos tales como: USB 2.0, Firewire 400, PCI Express o aquel que incluya la interface de audio.

Estas características podrán variar en función de los programas computacionales que se desee usar, ya que estos son actualizados frecuentemente y necesitaran mayores recursos.

2.2.2 Edición y procesamiento de audio digital.

2.2.2.1 Edición: El proceso de edición se refiere a la captura o selección de las partes de interés de una toma audio, La edición en el dominio analógico se realizaba en sus principios de manera física, para ello se recortaba a la cinta

magnética para obtener la porción de la señal de interés. Actualmente para el dominio digital la edición es más precisa, rápida, libre de ruido, no lineal y no es destructiva.

La importancia de la edición es que mediante esta herramienta se puede simplificar el trabajo de procesamiento de reconocimiento de voz debido a que se analizará solo la señal de interés en este caso las que contienen voz hablada.

2.2.2.2 Procesamiento: El procesamiento de audio se refiere a la etapa de modificación matemática, algorítmica de una señal de interés en función de parámetros establecidos por el usuario.

Procesar una señal de audio implica en algunos casos la transformación parcial o total de una señal original, bajo algún criterio previamente establecido por el usuario y/o por el fabricante.

2.2.3 Parámetros de aceptación de audio para su análisis.

El método más conveniente para análisis y procesamiento de audio se lo realiza en el dominio digital ya que es mucho más eficiente, rápido y económico que el dominio analógico.

Los parámetros profesionales aceptados son los siguientes:

2.2.3.1 Frecuencia de muestreo: La frecuencia de muestreo aplicada únicamente para el audio digital, se puede decir que los estándares aceptables del mismo son de 32 kHz, 44.1 kHz y 48 kHz para la grabación, siendo el de 44.1 kHz el más utilizado actualmente por su respuesta de frecuencia de 20 Hz a 20 kHz garantizado un análisis más fiel que con frecuencias de muestreo menores.

Mientras que para el audio analógico es difícil obtener una respuesta de frecuencia de 20 Hz a 20 kHz, debido a que esta es dependiente de la corriente de polarización de cinta, por eso un mínimo aceptable de la respuesta de frecuencia es de hasta 15 kHz.

2.2.3.2 Profundidad de bits: La cuantización de 16 bits es la más comúnmente utilizada, cuyo rango dinámico es de 96 dB, mientras que para uso profesional es de 24 y 48 bits. Un análisis en 16 bits puede garantizar un buen procesamiento y análisis.

Cabe recalcar que para la grabación profesional moderna se la realiza a 24 bits, por sus 144 dB de rango dinámico.

2.2.3.3 Cantidad de canales: El análisis de audio de reconocimiento de voz no requiere de una gran cantidad de canales de entrada, debido a que la *grabación multipista* no se hace necesaria, sin embargo se recomienda que la interface de audio posea por lo menos dos canales de entrada debido a que la mayoría de grabadores portátiles usan dos canales, mientras que para grabar una voz por lo general se utiliza un solo canal.

Se necesita de dos canales de salida para el caso que se realice monitoreo por cajas acústicas.

2.2.3.4 Formato de archivos: Actualmente existen varias extensiones de formatos de audio digital, sin embargo los archivos por excelencia para ser analizarlos son los *WAV* y *AIFF*, debido a que son formatos de audio que no poseen ningún tipo de pérdida de información, no necesitan licencia para poder trabajar y son los más populares formatos de audio.

2.2.4 Biometría.

La biometría es parte de la biología y estudia los métodos de identificación de personas por medio de los rasgos biológicos anatómicos y psicológicos que son únicos en cada individuo.

Los métodos más importantes de reconocimiento biométrico usados actualmente son: el análisis de huella digital, las características del rostro, el iris del ojo, escritura, geometría de la mano y características de la voz.

Actualmente existen sistemas automáticos de reconocimiento biométrico debido a los avances tecnológicos en el procesamiento digital por medio de

computadoras.

Un sistema biométrico está integrado básicamente por:

- ♦ **Sensor:** Es usado para obtener la información de un individuo y transformarla a información digital.
- ♦ **Algoritmos de identificación:** Son usados para la extracción de parámetros biométricos de una persona a partir de la información obtenida por el sensor.
- ♦ **Almacenamiento de información:** Es una base de datos que se usa para guardar la información biométrica para una posterior comparación.
- ♦ **Comparación:** Es un sistema que busca similitudes entre la información biométrica de uno o varios individuos analizados previamente.
- ♦ **Decisión:** Este sistema usa la información obtenida en la etapa de comparación para establecer un determinado nivel en la decisión del sistema.

2.2.5 Reconocimiento de locutor por medio de la voz.

Este método utiliza las características físicas y de comportamiento de la voz de un individuo para crear modelos biométricos que serán utilizados para tareas posteriores de comparación y de reconocimiento. Los resultados obtenidos con este método de análisis se pueden ver afectado por factores internos del individuo como son: el estado emocional, patologías del aparato fonador, edad, cantidad de palabras habladas; y factores externos como: el ruido ambiental, reverberación, microfonía, ruido eléctrico, distorsión, codificación, etc.

Existen dos formas de reconocimiento del locutor el primero es dependiente del texto en el que un individuo debe pronunciar una frase previamente establecida dentro del sistema; y otro que es independiente del texto donde no existe ninguna frase específica para el reconocimiento este sistema resulta más flexible para su utilización pero su implementación resulta más compleja y poseen un menor porcentaje de acierto en el reconocimiento.

2.2.5.1 Sistemas de reconocimiento automático de locutor.

Después de obtener modelos matemáticos que representan las características de distintas señales de voz se procede a determinar si una o varias muestras de voz tienen cierta semejanza o pertenecen a algún modelo de voz previamente analizado.

A este conjunto de procesos se les denomina sistema de reconocimiento automático de locutor, en este sistema se puede encontrar dos tipos de análisis:

- ♦ **Identificación:** Un sistema de identificación de locutores tiene como finalidad reconocer si la voz de una persona se encuentra dentro de un conjunto de N locutores que han sido previamente modelados y registrados dentro del sistema. Este tipo de identificación se le denomina de conjunto cerrado.

En la salida de este sistema se puede encontrar N posibles resultados correspondientes a cada uno de los locutores.

En este tipo de sistemas también se puede dar el caso de que voz analizada no se encuentre dentro del conjunto de locutores, a este tipo de análisis se le denomina identificación de conjunto abierto, donde el sistema debe permitir la alternativa de que la voz analizada no pertenece al conjunto de N locutores.

- ♦ **Verificación:** En un sistema de verificación de locutores solo se tienen dos entradas: la señal de voz a verificar y una señal de evidencia.

En este tipo de sistemas se puede obtener dos posibles salidas las que son la aceptación o el rechazo de que las voces analizadas pertenecen a la misma persona. Esta respuesta dependerá de si las voces analizadas superan un determinado umbral de similitud.

El análisis de similitud se lo realiza por métodos estadísticos orientados a entornos judiciales, denominados relaciones de verosimilitud o "*Likelihood Ratios*", los que son usados ampliamente en otros tipos de

análisis forense como para la investigación del ADN.

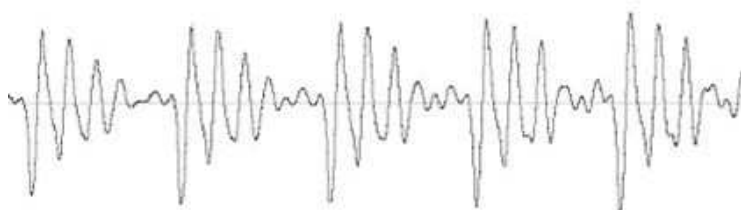
2.2.6 Definición de parámetros básicos del comportamiento de la voz hablada.

2.2.6.1 Naturaleza de la voz.

La voz humana es un tipo de señal que cambia en el transcurso del tiempo, esta posee a cambios fluctuantes en el transcurso del tiempo. Si se la analiza en el orden de los segundos la se puede considerar como no estacionaria,

Mientras que se si analizada en periodos cortos de tiempo en el orden de las decenas de milisegundos, se puede hallar segmentos que posee características cuasi estacionarias, mientras que otros segmentos son de apariencia ruidosa, a estos se le denomina sonidos sonoros y sordos respectivamente.

Figura 2.13: Gráfico de la forma de onda de muestra vocal durante 80 milisegundos.



Fuente: www.ub.edu/labfon/XVII-17.pdf

2.2.6.2 Extracción de características de la identidad del locutor.

Para realizar una investigación completa acerca de la identidad del locutor es necesario a analizar las características socio-psicológicas también llamadas de alto nivel y las características *fisiológicas* también conocidas como de bajo nivel.

Las características de alto nivel permiten distinguir la identidad del locutor, estas características se crean a partir del entorno o estado de ánimo en el que se desarrolló la persona analizada. En este grupo de características se puede

encontrar: el acento, el ritmo, la velocidad, la duración de las palabras, las pausas, el tipo de personalidad e influencia parental, la semántica, las pronunciaciones, el estatus socio económico, el nivel de educación, la modulación y la entonación.

Las características fisiológicas o de bajo nivel tienen una relación con las propiedades anatómicas del locutor, poniendo un especial interés en el análisis de los órganos de fonación los que guardan cierta relación con características físicas como son sexo, edad, altura, peso, etc.

En el análisis de señales de voz es común usar dos tipos de características: las características acústicas y las características de representación.

- ♦ **Características acústicas:** Estas características cuantifican las cualidades físicas de un sonido determinado, estos valores son obtenidos desde la representación digital de la voz.

El parámetro acústico se los puede dividir en parámetros cuasi periódicos y parámetros de perturbación:

- ♦ **Parámetros cuasi periódicos:** Frecuencia fundamental, formante, envolvente Espectral.
- ♦ **Parámetros de perturbación:** Relación armónico ruido, *Jitter* y *Shimmer*.
- ♦ **Características de representación:** Son conjuntos de valores calculados a partir de la representación digital de la voz. Generalmente estos valores no representan alguna cualidad física del la voz si no describen el comportamiento dinámico de una señal.

Los métodos usados actualmente para la extracción de características de representación son L.P.C. (Código de predicción lineal) o M.F.C.C. (Coeficientes de Frecuencia Mel Cesptrum).

2.2.6.3 Técnicas de clasificación.

Existen varios métodos de clasificar los resultados obtenidos después de

obtener las características, estos son adaptaciones de los métodos usados en el reconocimiento del habla. Entre estos los más relevantes tenemos:

- ♦ **Parámetros estadísticos (Long-Term-Statistics):** La serie de datos obtenidos de las características de la señal de voz es representada mediante medidas estadísticas como el promedio o la varianza. Son usados frecuentemente en reconocimiento del locutor independiente del texto.
- ♦ **Cuantización vectorial (VQ):** La cuantificación vectorial es una técnica muy utilizada en el procesamiento de señales, consiste en dividir un conjunto de vectores en grupos de vectores con valores parecidos los que serán representarlos por medio de su *centroide*. Este método es usado en el reconocimiento del locutor independiente del texto y posee un buen rendimiento cuando el número de muestras de voz es pequeño.
- ♦ **Alineamiento temporal dinámico (Dynamic Time Warping):** En este método se representa al locutor como conjunto de patrones que pueden ser palabras o frases clave. El reconocimiento del locutor se lo realiza utilizando medidas de distancia, aplicando el alineamiento temporal dinámico entre los patrones de voz del locutor y los patrones de referencia. Este método es utilizado en sistemas de reconocimiento del locutor dependiente del texto.
- ♦ **Modelos ocultos de Markov (HMM):** Es un modelo estadístico utilizado para el reconocimiento de patrones temporales como son: el reconocimiento del habla, el reconocimiento de gestos, etc. Tiene como objetivo determinar parámetros desconocidos de una cadena de Markov a partir de parámetros conocidos y observables. Los parámetros extraídos se pueden emplear para llevar a cabo sucesivos análisis. Este modelo se puede utilizar en el reconocimiento del locutor dependiente e independiente del texto.
- ♦ **Modelos de mezcla de gaussianas (GMM):** Este modelo caracteriza un conjunto de vectores de *coeficientes cepstrales* de un hablante por medio

de múltiples mezclas gaussianas para caracterizar la voz. Es el modelo estadístico de clasificación más exitoso en la actualidad y se puede utilizar tanto en sistemas de reconocimiento del locutor dependiente del texto como en los independientes del texto.

- ♦ **Redes neuronales artificiales (ANN):** Las redes neuronales son una simulación de las propiedades que se han observado en los sistemas nerviosos de los animales a través de modelos matemáticos que pueden ser representados por medio de mecanismos artificiales como circuitos integrados o conjuntos de válvulas. Estos mecanismos artificiales producen un estímulo de salida a partir de estímulos de entrada. Se utiliza en sistemas de reconocimiento del locutor independiente del texto.

2.3 Herramientas comúnmente utilizadas en el peritaje acústico.

2.3.1 Analizadores.

2.3.1.1 Analizadores de propiedades del audio.

El primer paso para la realización de un análisis de audio siempre será la determinación de las características del audio digital obtenido. Esto quiere decir que se hace necesario determinar la extensión del archivo, por ejemplo (.mp3, .ogg, .amr, etc.) conjuntamente con su frecuencia de muestreo y *bit rate* como por ejemplo. (44100 Hz a 128 kbps), el tamaño del archivo, su duración, su fecha de creación, y el nombre.

La manera más sencilla de poder determinar este tipo de parámetros, la realiza el mismo sistema operativo mediante la ventana de propiedades de los archivos.

Sin embargo existen aplicaciones que permiten ver con mayor detalle dichas propiedades como por ejemplo los editores de audio.

2.3.1.2 Analizadores espectrales.

El rango de frecuencias audible es el ancho de banda correspondiente es de 20 Hz a 20 kHz Sin embargo el audio eléctricamente se lo representa como al

voltaje versus el tiempo, para poder visualizar como es el comportamiento de una señal en el dominio de la amplitud versus frecuencia se utiliza a un analizador espectral.

El analizador espectral funciona mediante la FFT o transformada rápida de Fourier, que mediante algoritmos matemáticos descompone a la señal en sus componentes armónicas.

Existen tres tipos de analizadores espectrales entre ellos:

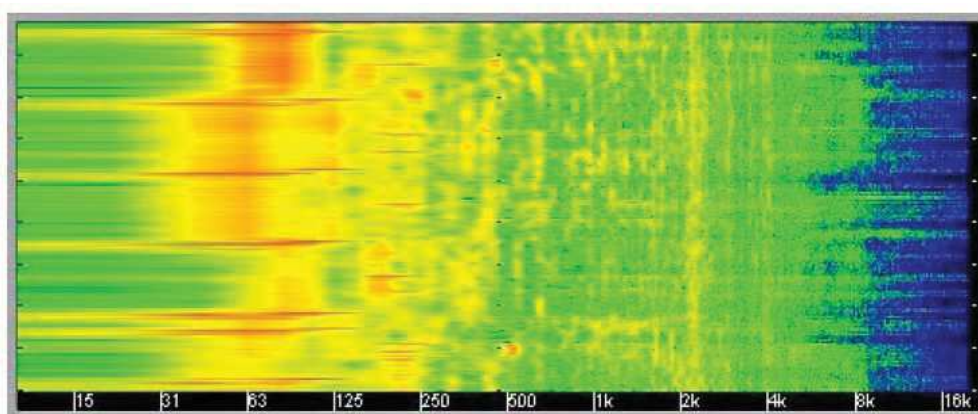
- ♦ **Espectrogramas:** Este gráfico es utilizado para mostrar la densidad del espectro de una señal utilizando tres dimensiones: tiempo, frecuencia y una codificación de colores para denotar su amplitud.

Figura 2.14: Gama de color en dB.



Fuente: www.astormastering.com.ar/Clase_2_Analisis_Espectral_y_Herramientas_Complementarias.pdf

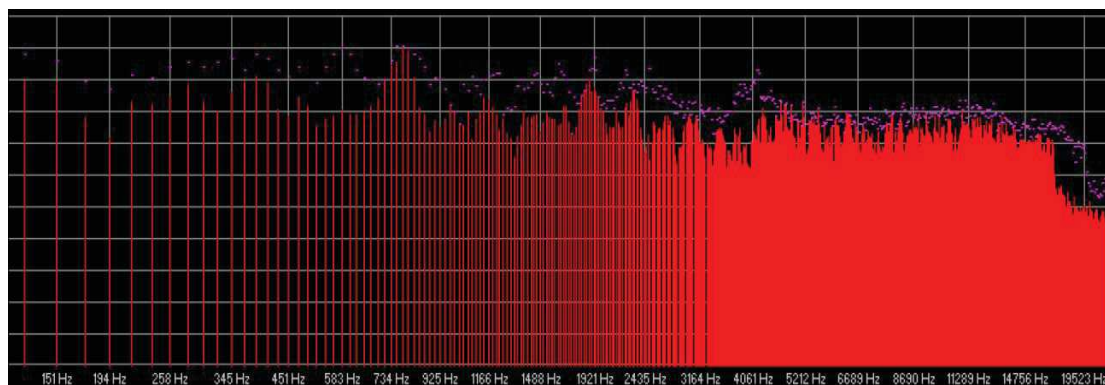
Figura 2.15: Espectrograma de un tema Rock cinco segundos.



Fuente: www.astormastering.com.ar/Clase_2_Analisis_Espectral_y_Herramientas_Complementarias.pdf

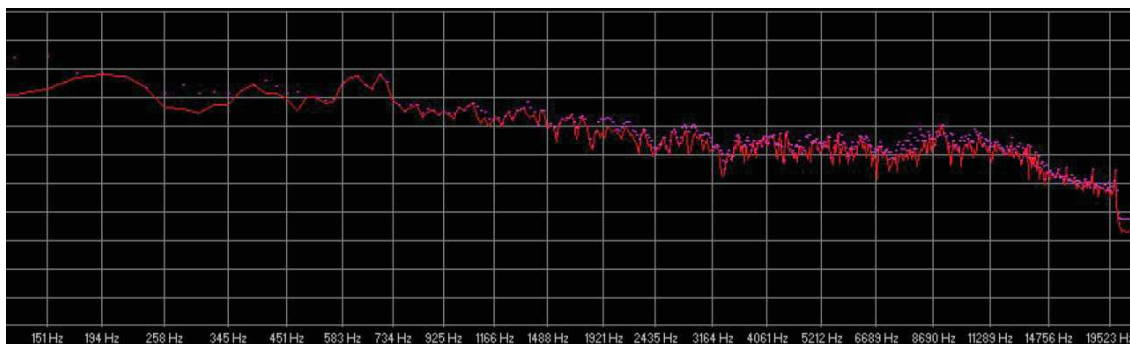
- ◆ **Analizadores FFT:** A diferencia del espectrograma este analizador lo realiza únicamente en dos dimensiones la amplitud y la frecuencia, cuya principal característica es la de hacer dicho proceso pero con anchos de banda equidistantes y en tiempos instantáneos, permitiéndose para altas frecuencias un análisis de hasta centésimo de octava, sin embargo el analizador es muy pobre para bajas frecuencias inferiores a 500 Hz, para lo cual se utiliza una interpolación para tratar de completar estos resultados.

Figura 2.16: Analizador FFT de Barras.



Fuente: www.astormastering.com.ar/Clase_2_Analisis_Espectral_y_Herramientas_Complementarias.pdf

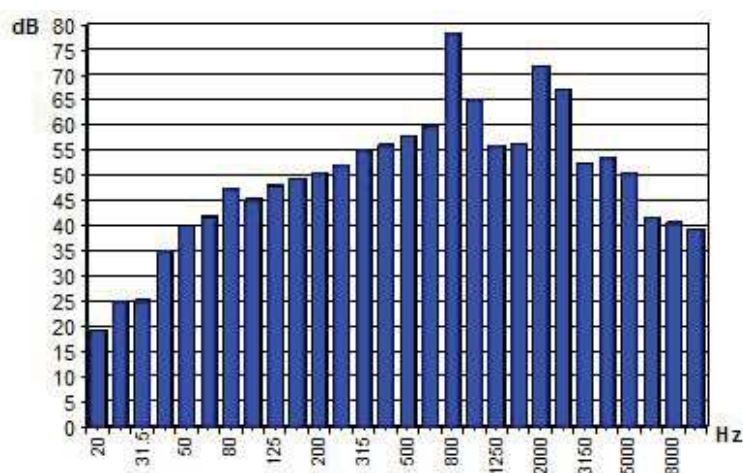
Figura 2.17: Analizador FFT de curvas.



Fuente: www.astormastering.com.ar/Clase_2_Analisis_Espectral_y_Herramientas_Complementarias.pdf

- ✦ **Analizadores por banda de octava o fracciones de octava:** Su principio de análisis es similar al del FFT, sin embargo este proceso se lo hace de manera logarítmica, así se optimizan recursos de procesamiento además se obtiene un mejor análisis para baja frecuencia debido a que no se realiza dicha interpolación para grafica lineal. Estos analizadores que utilizan anchos de banda variables dependiendo de la frecuencia central; mientras menor sea su frecuencia central, menor es su ancho de banda.

Figura 2.18: Analizador espectral por
banda de tercio de octava.



Fuente: www.bizkaia.net/Herri_Lanak/

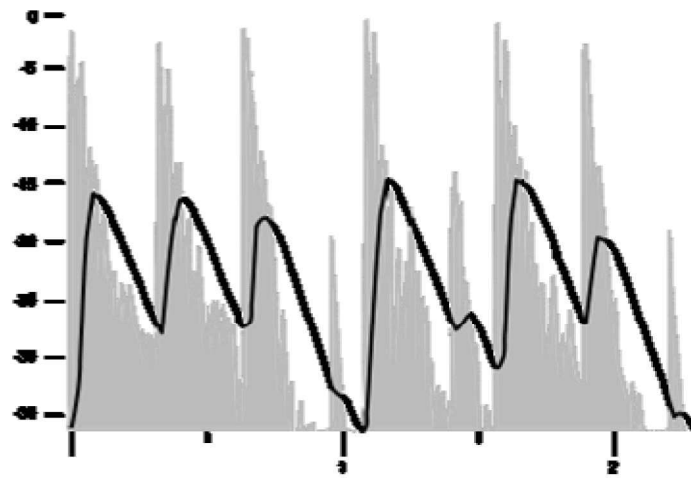
Mer/irudiak/grafico_sonido1.jpg

2.3.1.3 Medidores de nivel.

Se conoce como *meters* en audio a aquellos dispositivos analógicos o digitales capaces de medir el nivel de amplitud eléctrica de una señal en función de su tiempo de respuesta, el cual puede ser *peak* o *RMS*, siendo posible en cada uno de ellos configurar el tiempo de respuesta, como por ejemplo *fast* o *slow*.

Figura 2.19: Curva de respuesta de un meter.

RMS línea negra, *peak* línea gris.



Fuente: upload.wikimedia.org/wikipedia/commons/

f/f0/VU-meter-reponsegraph.svg

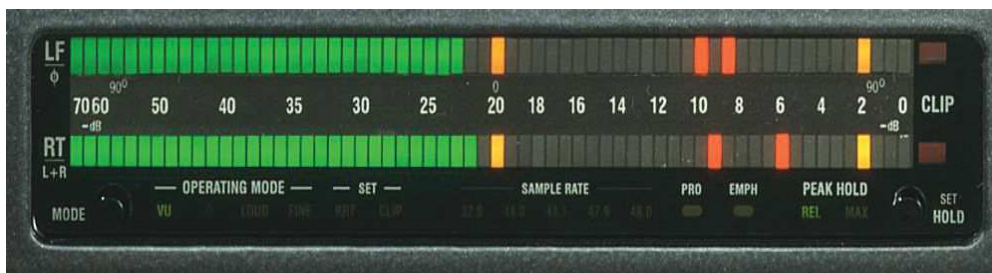
Entre los *meters* más conocidos se tienen a los *VU Meters* y a los *LED Meters*. Los primeros entregan obligatoriamente un tipo de respuesta RMS, mientras que los segundos pueden ser configurados para una respuesta RMS o *peak*.

Figura 2.20: VU Meter



Fuente: http://es.wikipedia.org/wiki/Archivo:VU_Meter.jpg

Figura 2.21: LED Meter.



Fuente: http://www.logitekaudio.com/assets/product_gfx/ultra_vu_large.jpg

2.3.2 Procesadores.

2.3.2.1 Ruido.

El ruido es toda señal que no se desea captar en la etapa de grabación y escuchar en la etapa de reproducción, cuya naturaleza es de carácter aleatorio.

En cada una de las etapas de una cadena electroacústica existe ruido de fondo proveniente de cada componente y sus efectos son acumulativos.

Existen de tres tipos de ruido:

- ♦ **El ruido de ambiente:** El cual se encuentra implícitamente en el medio ambiente como por ejemplo el ruido de tráfico, el viento, la respiración.
- ♦ **El ruido inherente:** Es un ruido producido por los dispositivos de naturaleza activa, entre ellos está el ruido por circuitos eléctricos, en el caso del audio analógico el ruido producido por cintas magnéticas y el ruido por interferencia electromagnética.
- ♦ **El ruido por inducción:** Es un ruido de característica eléctrica y magnética se puede presentar en cualquiera de las etapas de la grabación y reproducción de audio. El ruido por inducción puede ser de dos tipos:
 - ♦ **Campos magnéticos:** Este ruido es generado por la presencia de fuertes campos electromagnéticos que inducen una corriente sobre un

conductor, las partes más vulnerables para ser inducidas son los cables y las tarjetas de procesamiento.

- ✦ **Conducción:** Este ruido se presenta cuando se conectan dos o más dispositivos electrónicos entre sí, ya que cada uno adquiere una carga eléctrica por la circulación de corriente dentro de los mismos. Dichas cargas generan diferencias de tensión las que inducen una corriente eléctrica ajena a la señal original de audio.

2.3.2.1.1 Métodos de reducción y eliminación de ruido.

Cada ruido posee sonoridad y características distintas por lo tanto es necesario aplicar distintas técnicas para su eliminación.

- ✦ **Ruido de ambiente:** Este tipo de ruido es de naturaleza acústica por lo tanto su solución está basada en el uso de filtros acústicos, aislación acústica, sistemas anti vibratorios y técnicas de microfonía.
- ✦ **Ruido inherente:** Para tratar este ruido se utilizan métodos complementarios y no complementarios.
- ✦ **Sistemas complementarios:** Esta técnica de reducción de ruido se la usa previamente a la aparición del ruido, consiste en dos etapas: la primera es una etapa donde se comprime a la señal de audio conjuntamente con el ruido, y la segunda etapa consiste en expandir la señal de audio comprimida alejando así proporcionalmente el ruido de fondo de la señal útil. Esta técnica fue utilizada por los sistemas Dolby para eliminar ruido de cinta. Adicionalmente se puede encontrar en el mercado dispositivos de hardware especializados para el tratamiento de ruido fabricado por la empresa CEDAR.
- ✦ **Sistemas no complementarios:** Estas técnicas son utilizadas para una señal que ya presenta ruido, se utilizan procesadores sean de tipo hardware y software. Este tipo de técnicas son las más relevantes para tareas de peritaje acústico.
El método que utilizan estos sistemas está basado en la huella de ruido que consiste en analizar una porción de la señal que solo corresponde

al ruido y por medio de algoritmos matemáticos se las reduce de la señal original.

Actualmente se han desarrollado softwares profesionales tales como:

“iZotope RX, Waves Restoration Bundle, etc.” capaces de detectar y eliminar varias clases de ruido presentes tales como: clips, hum, ruido de fondo, ruidos ambientales, etc.

El uso de compresores de audio como sistemas no complementarios se hace útil a la hora de reducir la profundidad de bits en la etapa previa a la aparición del ruido de cuantización característico de este proceso.

- ◆ **Ruido por inducción:** Este ruido se lo trata dependido de su naturaleza.
 - ◆ **Campos magnéticos:** Para tratar este tipo de ruido se debe alejar las fuentes electromagnéticas lo más lejanamente posible de la interfaz de audio puesto que los convertidores A/D son los más sensibles a este tipo de ruido. Se recomienda también el aislamiento de los conductores mediante el uso de mallas y recubrimientos metálicos. Las fuentes electromagnéticas más comunes son: las fuentes de alimentación eléctrica, radio transmisores, cajas acústicas, etc.
 - ◆ **Conducción:** Para el tratamiento de este ruido es necesario verificar que los equipos tengan una buena conexión a *tierra* y que la mayoría de sus conexiones balanceadas.

2.3.2.2 Distorsión.

Se conoce como distorsión cuando se afecta a una señal de audio ya sea en amplitud frecuencia y fase, existen varias clases de distorsión entre ellas las más importantes.

- ◆ **Distorsión armónica:** se presenta cuando a una señal de entrada se ve afectada en el timbre en su etapa de salida, entre las distorsiones armónicas más comunes se tiene:

- ♦ **Distorsión por saturación:** Esta distorsión se presenta cuando se sobrealimenta el rango dinámico de un equipo, esta distorsión recorta los picos y valles de la forma de onda original, este tipo de saturación afecta a la señal cambiando su timbre.
- ♦ **Distorsión total armónica:** Conocido por sus siglas en ingles “THD”, y se produce por la naturaleza activa de la circuitería de un sistema, afecta a la forma de la onda resultante introduciendo frecuencias armónicas, que son frecuencias múltiplos de la señal de entrada. En otras palabras el THD es el porcentaje de la relación entre la energía aportada por las frecuencias armónicas no deseables con respecto a la energía de la señal original.

Se debe tener en cuenta que existen otros tipos de distorsión como por ejemplo la distorsión por fase y la producida por micrófono de condensador que se genera cuando las placas están a punto de contactarse produciendo recorte triangulares en los picos y valles de la onda.

- ♦ **Distorsión por cruce:** Esta distorsión se presenta por la no linealidad producida en los dispositivos electrónicos no se encuentren polarizados y afectan directamente a la curva de transferencia que altera la relación de los voltaje de entrada con los voltajes de salida de los dispositivos.

2.3.2.2.1 Métodos de reducción y eliminación de distorsión.

Cada tipo distorsión posee características distintas por lo tanto es necesario aplicar distintas técnicas para su eliminación.

- ♦ **Distorsión por saturación:** A esta distorsión se la puede solucionar en la etapa de grabación y conversión de la señal controlando la adaptación de la dinámica entre las conexiones de entrada y salida, sin embargo para señales ya grabadas con saturación se sugiere la utilización de programas computacionales especializados.

- ✦ **Distorsión total armónica:** Este tipo de distorsión es la más difícil de controlar, por lo tanto se sugiere utilizar equipos que tengan el THD más bajo posible en sus especificaciones.
- ✦ **Distorsión por cruce:** Para este tipo de distorsión se recomienda revisar el estado de los equipos de grabación y reproducción periódicamente, analizar si las corrientes de polarización se encuentran operando de manera normal y calibrarlos de ser posible; caso contrario se deberá reemplazar el equipo.

2.3.3 Optimización y acondicionamiento del audio para analizar.

Se le conoce como optimización al proceso que se le aplica a una muestra audio para obtener su mejor desempeño en las etapas de reproducción y análisis.

Cuando se elimina ruido del audio, puede decir que se lo está “optimizando”, sin embargo este proceso no es suficiente debido a que el audio puede presentar distintos tipos de problemas ajenos al ruido como grabaciones de bajo nivel, poca profundidad de bits y bajas frecuencias de muestreo.

Es importante recalcar que los formatos de almacenamiento de audio sin compresión más utilizados son los denominados “*WAV* y *AIFF*” siendo así los más óptimos para su procesamiento.

Para el análisis óptimo de audio, se debe reprocesar el audio de la siguiente manera una vez que sea detectado el respectivo problema:

- ✦ **Para señales de bajo nivel:** Se debe adaptar el rango dinámico de la señal a un nivel nominal, este proceso se lo realiza mediante un aumento de ganancia, sin embargo existe otro método denominado normalización que mediante algoritmos matemáticos incrementa proporcionalmente el nivel de toda la señal.
- ✦ **Para poca profundidad de bits:** En este caso se deberá recuantizar el audio a profundidades iguales o superiores a 16 bits. Sin embargo aunque

se aumente la profundidad de bits el audio no mejorará en su calidad pero tendrá mejor desempeño en la etapa de procesamiento. En caso de reducción de la profundidad de bit se hace necesario la utilización del dithering en los programas editores de audio.

- ♦ **Para baja frecuencia de muestreo:** Se deberá resamplear a frecuencias de muestreo iguales o superiores a 44100 Hz. El audio original no mejorará en su calidad aunque se aumente la frecuencia de muestreo, pero tendrá mejor desempeño en la etapa de procesamiento.

2.3.4 Herramientas de programación.

2.3.4.1 Programación.

La programación ha existido como la manera de esquematizar procesos que permiten organizarlos y resolverlos, el modo utilizado para la programación es por medio de la informática.

- ♦ **Lenguajes:** Se lo realiza como un idioma artificial diseñado para expresar cálculos matemáticos y lógicos que pueden ser llevadas a cabo por máquinas como las computadoras. En la década de los 60's aparecieron los lenguajes de programación más conocidos como: Basic, Pascal, C, C++, Perl, Python, Java. Actualmente la programación es utilizada varias áreas de la ciencia.

2.3.4.2 Max MSP

Es un entorno de programación gráfico basado en objetos, inicialmente fue desarrollado por Miller Puckette, Max fue diseñado como un sistema de control para el protocolo MIDI. Sin embargo se han implementando nuevas herramientas como el MSP que son utilizadas para el procesamiento de audio en tiempo real.

2.3.4.3 Nociones Básicas de Max MSP.

En el entorno de programación Max MSP el documento donde se procede a interconectar las distintas herramientas es denominado patcher.

2.3.4.4 Ventana del Patcher.

Como primer paso en la creación de un patcher es abrir una ventana de patcher desde el menú archivo y seleccionando Nuevo Patcher. Se abre una pantalla en blanco donde se procede a seleccionar, modificar y arrastrar distintos tipos de herramientas conocidas como objetos los que son conectados unos a otros por medio de cables virtuales.

Las características de la ventana del patcher así como de los objetos pueden ser personalizadas utilizando el Inspector de patcher que se encuentra en el menú Ver.

2.3.4.5 Objetos.

Los objetos son bloques de construcción que ejecutan determinadas órdenes dentro de un patcher, los objetos que pueden ser interconectados para la creación de programas más grandes. Los objetos acompañados del símbolo ~ son especializados en el manejo de señales de audio.

Los distintos objetos se los puede encontrar en la paleta de Objetos de Max MSP, desde donde se los debe arrastrar hacia la ventana de patcher. Para crear un objeto específico se debe crear un cuadro de objeto y escribir el nombre de la tarea que debe realizar, adicionalmente se puede escribir uno o varios argumentos que determinan su forma de funcionamiento.

Los argumentos de algunos objetos pueden ser modificados durante la ejecución del programa utilizando mensajes de texto que se envían a sus entradas.

Existen varias categorías de objetos de los que se puede citar a los objetos más importantes para la elaboración del programa de reconocimiento de locutor:

- ◆ Básicos: cuadro de objeto, mensaje, comentario, número entero, número decimal, inlet y outlet.
- ◆ Deslizadores de ganancia: Gain~, multi-deslizador.

- ◆ Botones: Button, LED y conmutador.
- ◆ Interface: Editor de texto y umenu.
- ◆ Audio: Medidor de nivel y espectroscopio.

Dentro de Max MSP existen una gran cantidad de objetos listos para ser usados de acuerdo a las necesidades del programador, también es posible utilizar objetos creados por terceras personas los que permiten expandir las posibilidades de procesamiento tanto en datos numéricos como en señales de audio.

2.3.4.6 Tipos de datos en Max MSP.

- ◆ **Mensaje:** Entre los tipos de datos que puede manejar Max MSP están los mensajes que pueden ser conjuntos de palabras o símbolos. Se utilizan para establecer el funcionamiento de algunos objetos o para informar los resultados de forma textual. Uno de los mensajes más importantes en Max MSP es el denominado bang que puede iniciar el funcionamiento de objetos.
- ◆ **Número:** Los datos numéricos expresan una determinada cantidad, en el entorno Max MSP considera dos tipos de mensajes numéricos: los números enteros y los números con decimales. Los datos numéricos pueden ser utilizados para la realización de operaciones matemáticas o la visualización de resultados.
- ◆ **Lista:** Al combinar dos o más mensajes de texto o numéricos se crea un nuevo tipo de datos denominados lista. Las listas resultan útiles para la creación de estructuras de datos para el análisis y para reducir número de conexiones hacia un mismo lugar.
- ◆ **Audio:** Una de las principales características del programa Max MSP es que permite manejar señales de audio en tiempo real. Dentro del programa las señales pueden ser sintetizadas, analizarlas o procesadas.

2.3.4.7 Conexiones.

Para conectar los distintos objetos en Max MSP se utilizan cables virtuales, existen varios tipos de cables dependiendo de la información que esta lleve.

- ◆ **Cable de Max:** Es un cable delgado de color negro se utiliza para transmitir información numérica, caracteres, mensajes o listas.
- ◆ **Cable de MSP:** Es un cable de color amarillo se utiliza para transporta información de audio de un solo canal, la velocidad de transmisión del audio dependerá de la frecuencia de muestreo y profundidad de bits escogida en el menú DSP Status del menú Opciones.

2.3.4.8 Sub patcher.

Este es un objeto que permite crear un patcher dentro de un patcher, es útil para organizar la venta de patcher, permitiendo agrupar objetos en distintos módulos con tantas entradas y salidas como sean requeridas utilizando los objetos inlet y outlet.

Capítulo 3. Análisis de la situación actual.

3.1 Descripción de la situación actual ecuatoriana en el peritaje acústico.

3.1.1 Investigación y acerca del peritaje acústico.

Para la realización de este capítulo se acudió al departamento criminalística de la Policía Nacional sede Pichincha, sección de “Audio y Video”, ubicado en la Av. Occidental y Mariana de Jesús cuyo horario de atención es de lunes a viernes de 08h00 a 16h30.

El departamento posee un equipo de siete oficiales activos, donde se entrevistó al Cabo Primero Patricio Vascones, encargado de la unidad de audio y cotejamiento de voces, cuyo cargo lo desempeña durante un lapso de dos años y medio. El mismo dio a conocer que en el Ecuador este tipo de trabajo logístico se lo realiza únicamente en las ciudades de Quito y Guayaquil, siendo el primero el más actualizado en equipamiento. Estos departamentos acogen las grabaciones magnetofónicas a nivel nacional las mismas que son financiadas por el estado, razón por la cual solo se realizan peritajes cuando son requeridos en un proceso legal, designado por la fiscalía y los juzgados.

Este tipo de evidencia es validada y acogida por el “Código de Procedimiento Penal”.

3.1.2 Procesos realizados actualmente en el peritaje acústico.

La evidencia recolectada normalmente por el departamento de audio y video son diálogos grabados en dispositivos magnetofónicos.

La sede de Quito realiza transcripciones a texto y cotejamientos de voz por medio del sistema “Ikar Lab” propiedad del “Centro de tecnologías de la voz de Rusia”, el sistema consta de estos componentes:

- ◆ Sistema interactivo de análisis y procesamiento de señales acústicas “SIS 6.0”.
- ◆ Programa para filtrar ruidos y mejorar la calidad de señales acústicas “Sound Cleaner Premium”.

- ◆ Programa para diagnosticar la autenticidad de grabaciones “EdiTracker”.
- ◆ Programa Asistente para la transcripción “Caesar”.
- ◆ Interface de medición y entrada/salida de señales acústicas “STC-H246”.
- ◆ Interface de entrada/salida de señales acústicas “STC-H216”.
- ◆ Micrófono profesional y audífonos.

Los detalles de IKAR LAB se encuentran en el (Anexo 2).

Las transcripciones a de voz a texto se las realiza auditivamente por un perito a un informe oficial utilizando la siguiente simbología:

- ◆ P1: para redacción de las palabras del hablante uno.
- ◆ P2: para redacción de las palabras del hablante número dos.

Y así sucesivamente según el número de participantes en el diálogo magnetofónico y el orden en el que aparecen. Por la tanto nunca se incluyen nombres de implicados para evitar posibles acusaciones o malos entendidos.

Mientras que para el cotejamiento de voz el departamento policial toma como primer paso que el sospechoso haya sido registrado en una grabación con la pronunciación de por lo menos 15 palabras, una vez aprobado el siguiente paso a realizarse es la citación por fiscalía de los sospechosos para obtener nuevas muestras de voz, las que son obtenidas al grabar la lectura de un texto igual o parecido al dialogo de la grabación.

La policía realiza el análisis solo superposición de formantes en un espectrograma considerando tres factores: el dispositivo con el que se grabó, la distancia entre el emisor y el dispositivo de grabación, y el ambiente acústico en el que se realizó la grabación.

En el informe pericial se incluye una impresión de pantalla con los resultados comparativos de por lo menos 15 silabas similares o iguales, siempre y cuando

posean más de tres formantes o “manchas” en el espectrograma para poder identificar a la persona.

3.1.3 Tiempos empleados actualmente en la realización de peritaje de voces.

Los informes periciales de la policía nacional tienen un plazo máximo de entrega de 15 días dependiendo la demanda y la urgencia que estos requieran, que se hasta se han entregado como tiempo record en cinco días.

Actualmente el área de peritaje acústico más demandada es la transcripción textual de una grabación, la frecuencia de petición de este tipo de informe es de alrededor de dos oficios o más al día.

Entonces se puede decir que a la Policía Nacional le toma un tiempo de alrededor de 15 minutos por cada palabra analizada.

La Policía mencionó que existen peritos civiles y se los puede encontrar en listados de fiscalía, sin embargo la función judicial destina la mayoría de los casos a este departamento debido a la falta de oferta de este tipo de servicios, también comentó que para mejorar la eficiencia es necesario de un mayor número de unidades de análisis y distribuir las en más regiones del país.

Capítulo 4. Descripción de la situación propuesta.

4.1 Mejoras implementadas para una nueva metodología.

Para evitar el manejo empírico de esta rama del peritaje, se analizaron los procedimientos técnicos y se detectaron los siguientes inconvenientes:

Por humanística:

- ◆ Falta de especialistas en el área.
- ◆ Exclusión previa de evidencia por parte de jueces y tribunales debido a su relevancia.
- ◆ Capacitación periódico al personal.
- ◆ Exámenes de la salud auditiva de personal.
- ◆ Desconocimiento popular de este tipo de servicio.

Por técnicas empleadas:

- ◆ Desconocimiento del comportamiento del ruido y distorsión.
- ◆ Falta de unificación de procedimientos de análisis.
- ◆ Falta de normativa de calidad por organismos externos.
- ◆ Desconocimiento de fundamentos de audio digital y analógico.

Por herramientas utilizadas:

- ◆ Dificil acceso a tecnología forense por falta de recursos.
- ◆ Falta de bibliografía referente al audio.
- ◆ Análisis realizados únicamente por laboratorios de Quito y Guayaquil.

Considerando todos estos aspectos más aporte investigativo y científico se realizó una guía metodológica aplicada al Ecuador.

4.2 Creación de una guía metodológica para realizar peritajes acústicos de reconocimiento de voz.

Para poder realizar una guía metodológica fue necesaria una investigación de campo conjunto con material bibliográfico proveniente de los sitios web de profesionales en el campo laboral.

Durante el desarrollo de la presente investigación se pudo conocer cómo se realizan las tareas de peritaje acústico en el Ecuador, proceso que se lo realiza normalmente pero con falencias debido al poco conocimiento popular de este tipo de servicio entre otros. Razón por la cual se utilizó como referencia al texto denominado “Guías Metodológicas de las Especialidades Periciales” proveniente de la Procuraduría General de la República de México editada en el año 2003, bibliografía detallada en el (Anexo 5).

GUÍA METODOLÓGICA PARA LA APLICACIÓN DE LA ACÚSTICA FORENSE EN EL RECONOCIMIENTO DE VOZ HABLADA APLICADA A ECUADOR.

Concepto: Se hace necesario un exhaustivo análisis de voz, en vista de la popularidad de los sistemas magnetofónicos de grabación de audio, por lo tanto se convierten como evidencia circunstancial implícita en tiempo real, para el esclarecimiento de hechos.

Objetivo: Identificar a un sospechoso mediante comparaciones auditivas y extracción de las características fonéticas propias de la emisión de la voz, utilizando un software analizador de voz.

Realidad Nacional: Actualmente en el Ecuador se analiza la relevancia de una grabación magnetofónica, para que así una vez aprobado fiscalía o el juez puedan ser enviadas al departamento de audio y video pertenecientes a Criminalística de La Policía Nacional, por medio de investigación de campo se identificó que se descarta mucha evidencia debido masificación de grabadores portátiles, la poca participación de peritos civiles y la irrelevancia de

información en algunas grabaciones puesto que algunos juzgados con criterio personal lo desechan.

Casos comunes para intervención de análisis:

- ♦ **Grabación de audio implícito en video:** Este tipo de grabaciones se las obtiene al extraer el audio proveniente de una grabadora de video, siendo de tipo analógico o digital, para lo cual se hace necesario un editor que permita separar estas dos componentes.
- ♦ **Grabación de audio en dispositivos portátiles:** Estas grabaciones siendo las más comunes son las que se las realiza por medio de dispositivos como grabadoras de tipo periodista, grabaciones por micrófono de teléfono celular, es decir las grabaciones realizadas con equipamiento de tamaño reducido.
- ♦ **Registro de audio de grabadoras profesionales:** Es aquel material obtenido de grabadoras profesionales de audio, es el menos común que se lo obtiene cuando se realiza grabaciones en conciertos, ruedas de prensa, estudios de grabación, es decir utilizar un equipamiento de alto desempeño.
- ♦ **Grabación de audio intervenido en llamada telefónica:** Este tipo de grabación se obtiene cuando se interviene a una llamada telefónica como es muy común para el caso de secuestros, casos de corrupción, charlatanería etc. Se la puede realizar por medio de mismo dispositivo telefónico, por equipamiento especializado para grabación como el policial y el realizado por las mismas operadoras de telefonía.

Requerimientos: Actualmente se puede realizar análisis sobre cualquier registro magnetofónico, sin embargo la efectividad de un análisis es muy dependiente de la calidad del audio obtenido, para lo cual se dictaminan diferentes sugerencias y parámetros para evitar la pérdida de tiempo que esto implica para casos complicados.

A continuación se enlistarán los requerimientos por áreas:

Requerimientos del perito:

- ◆ El perito debe ser especialista en acústica forense y tener conocimientos técnicos de audio.
- ◆ El perito debe encontrarse en perfecto estado de salud con respecto a su *aparato auditivo* y debe realizarse periódicamente exámenes de *audiometría* para evitar falencias inconscientes en análisis auditivos.
- ◆ El perito debe tener sano juicio y valores de moral y ética.

Requerimientos del laboratorio:

- ◆ El laboratorio debe contar con tecnología de análisis de voz competente a época.
- ◆ Se debe contar la mayor cantidad posible de reproductores de audio digital y analógico.
- ◆ El laboratorio debe contar con condiciones acústicas favorables y aislamiento contra ruidos y vibraciones externas.
- ◆ El laboratorio debe poseer una infraestructura confortable para evitar la fatiga del perito.
- ◆ El tiempo de entrega de un informe pericial dependerá de la cantidad del contenido a analizarse.

Requerimientos de la evidencia y muestras de voz.

- ◆ Se debe entregar al perito la grabación original de la evidencia recolectada.
- ◆ La voz de él o los implicados deberían ser lo más legibles, para determinar un mejor resultado.
- ◆ Las grabaciones se deben analizar para determinar si cumplen con condiciones mínimas de calidad, tanto como en profundidad de bits si

como de frecuencia de muestreo, de lo contrario se deberá realizarse un proceso de optimización del audio.

- ◆ Factores acústicos externos muy enfatizados en una grabación como por ejemplo: mucha distancia emisor-dispositivo y la influencia del medio ambiente afectan directamente a la inteligibilidad de la palabra, por lo tanto se debe descartar esta grabación como evidencia.
- ◆ Debería existir un mínimo de palabras mencionadas por el sospechoso y que estén implícitas en la grabación para hacerlo participe en un informe.
- ◆ En caso de existir responsables previos al análisis se deberá obtener muestras de audio por orden judicial, y dichas muestras deberán ser tomadas por el perito a cargo del análisis.
- ◆ En caso de no existir responsables previos a análisis es necesario que el perito a cargo de la grabación, realice un perfil acústico en caso de futuras investigaciones.
- ◆ Para el caso de la obtención de muestras de voz para comparación se hace muy importante que el individuo se encuentre en perfecto estado de salud con respecto a su aparato fonador y en lo posible que las muestras sean tomadas sin ningún tipo de manipulación intencional al momento de hablar, por ejemplo la utilización de voz dispersa, voz nasal, falsetes y voz gutural.

Procedimiento del análisis: Una vez que se ha cumplido con los requerimientos previamente establecidos, el perito debería cumplir con los siguientes pasos.

- ◆ Digitalizar la evidencia con la mejor calidad posible en el caso que provenga de una fuente analógica y para el caso de audio digital se lo resampla y recuantiza a la mejor calidad posibles para su análisis por medio de software.

- ◆ En caso de ser necesario se optimizará la grabación digitalizada por procesamiento de software especializado.
- ◆ En el cotejamiento de voz el perito debe analizar la frecuencia fundamental en intervalos de tiempo y calcular los sus valores estadísticos: mínimo, máximo, promedio y desviación estándar.
- ◆ Para el análisis de formante el perito debe tomar en cuenta por lo menos de 15 a 20 fonemas por análisis en el peor de los casos.
- ◆ Cada fonema analizado deberá contener un mínimo de tres formantes para poder acreditarlo.
- ◆ Para la verificación de la identidad del locutor se debe yuxtaponer fonemas correspondientes a una misma vocal y comparar las similitudes en frecuencia entre las primeras tres formantes. Además se debe compara los valores estadísticos de la frecuencia fundamental entre los distintos sospechosos.
- ◆ La transcripción a texto se realizará como manera de análisis auditivo y vía software en el mejor de los casos, utilizando simbología legible para cualquier ciudadano.
- ◆ Una vez que el perito haya obtenido resultados se deberá redactar su posterior informe, detallando particularidades y redactar todo de la manera más sencilla posible para que pueda ser entendida por jueces, fiscales, ciudadanía, etc.

4.3 Guía de procedimientos para la elaboración del informe de peritajes acústicos de reconocimiento de voz.

Realizada una previa investigación, a fines de comparar similitudes en informes periciales españoles como ecuatorianos y basados en la guía metodológica mexicana, se puede decir que todo informe pericial debe constar de mínimo seis puntos claves para poder hacer válido dicho informe.

A continuación se detallarán los distintos pasos en el proceso de redacción:

- ◆ **Detalle del recurso humano:** En esta sección del informe por ser la inicial se la representará con una carátula donde se incluyan el logotipo, el nombre, lugar, fecha, teléfonos y dirección del laboratorio técnico en donde se realiza el informe. A continuación se procede a enumerar la cantidad de personal científico indispensable en donde se detallará sus cargos y su currículum académico conjunto con sus años de experiencia en la elaboración de informes periciales.
- ◆ **Detalle del equipamiento técnico:** Se hace necesario enumerar el equipamiento utilizado con sus características técnicas tales como su año de fabricación, su procedencia, su marca, número de unidades que posee el departamento y de ser posible la frecuencia de utilización y mantenimiento.
- ◆ **Detalle de percepción auditiva de parte del perito:** Para ello se hace necesario que el perito posea una buena salud auditiva, por lo tanto este sujeto deberá poseer experiencia, ética y buen criterio, ya que esta es la primera etapa de análisis sobre la muestra de audio, es aquí en donde se realiza las transcripciones a texto, es decir; se utiliza la simbología para redactar a manera de dictado, lo que el sospechoso intenta comunicar en mensaje oral.

Esto se utiliza en la mayoría de los casos para entender su intención, su energía, su estado de ánimo, su sexo, muletillas, dialecto. Puesto que estos son parámetros para determinar posibles actos de grabación forzada o manipulación, de ser así para el caso de confesiones ya no sería válido como evidencia.

- ◆ **El estudio fono articulatorio y acústico:** Esta es la etapa más importante de todo el informe pericial debido a que aquí se elimina cualquier parámetro de subjetividad con el referente a sospechosos con voces iguales o parecidas.

Para esto se hace necesario el estudio fono articulatorio partiendo con la edición previa del audio digitalizado, dividiéndolo en fonemas vocálicos

consonantes y mixtos de mínimo tres formantes. Detallando la frecuencia central a la que se encuentra cada una en la posición del espectrograma, los datos de comparación con las nuevas muestras obtenidas en el laboratorio, mientras mayor sea el número de fonemas a analizar mayor será la probabilidad de determinar la identidad del sospechoso, sin embargo será prudente realizar este tipo de análisis en número superior a los diez fonemas.

Como sugerencia si el sospechoso presenta algún tipo de muletilla al hablar repetitiva e inconscientemente una palabra a la largo de su expresión, se recomienda hacer el análisis sobre esa palabra en específico debido a que es la muestra más natural.

- ♦ **Definición de parámetros de comparación:** Normalmente toda rama de la ciencia se encarga de elaborar tablas comparativas o tablas de referencia con el fin de entender la naturaleza de algún fenómeno, para este caso se pueden hacer análisis como la frecuencia fundamental a la que el individuo normalmente en promedio ejecuta su voz, también se puede encerrar un rango de frecuencia al que el individuo ejecuto su voz; es decir se puede realizar todo tipo de procesamiento de audio para entender parámetros que para este caso son subjetivos pero no descartables a la hora de dar una conclusión.

Como sugerencia este es el lugar en donde se pueden adjuntar anexos, imágenes y extras que se quieran agregar al informe, siempre y cuando contenga información relevante que aporte a la conclusión del mismo.

- ♦ **Redacción de conclusiones:** Esta etapa es la más importante para la aplicación judicial debido a que aquí el perito determinará si la voz es o no perteneciente a un sospechoso.

Para ello se debe establecer niveles de similitud en la diversas etapas del informe para lo cual la mayor ponderación la tendrá el examen fono articulatorio y acústico, por lo tanto será o no así un sospechoso identificado.

En caso de existir cualquier duda o inconformidad con el examen realizado, es en esta etapa del informe en donde se detallará los distintos tipos de inconvenientes que imposibilitaron el análisis.

Capítulo 5. Diseño del software de reconocimiento de voz hablada.

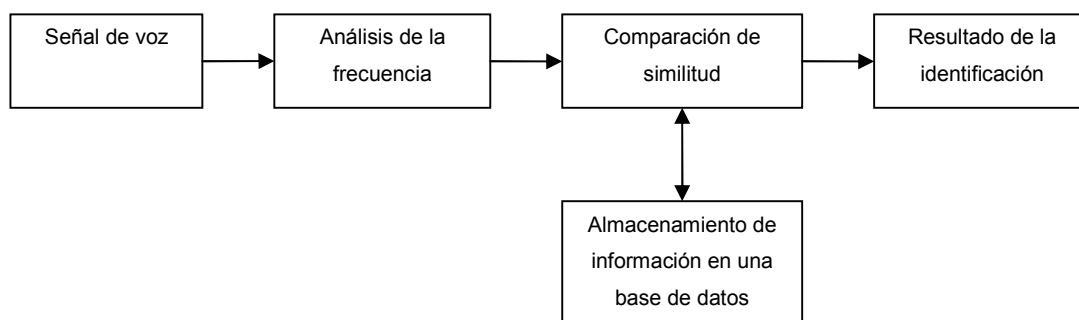
5.1 Creación de un software de identificación de locutor.

La idea planteada es la de crear un programa de identificación de locutor en un conjunto cerrado de individuos basado en la estadística de la frecuencia fundamental.

5.1.1 Diseño del algoritmo de identificación de locutor por la voz.

El programa debe realizar el análisis en tiempo real de la frecuencia fundamental así como calcular sus datos estadísticos.

Figura 5.1: Esquema del algoritmo de reconocimiento de locutor.



Fuente: Autores.

5.1.2 Determinación de la frecuencia fundamental.

La frecuencia fundamental F_0 o pitch, es una de las principales características para la identificación de locutor en ámbito forense, codificación de voz y sistemas de control acceso, puede ser determinada midiendo la velocidad con que se abren y cierran las cuerdas vocales durante la fonación de sonidos sonoros, también se puede estimar la frecuencia fundamental mediante el análisis digital usando un algoritmo computacional.

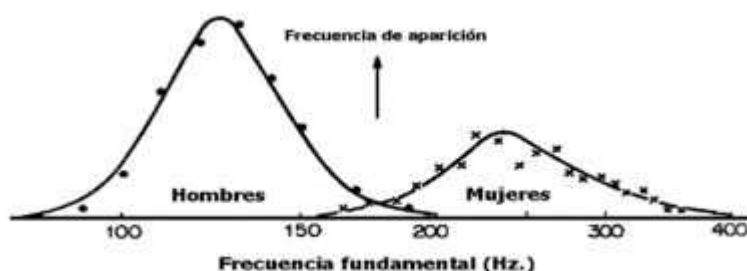
La frecuencia fundamental varía en el tiempo, por lo tanto para su análisis se hace necesario dividirla en intervalos finitos de tiempo a este proceso de análisis localizado se lo denomina *eventanado*.

5.1.3 Distribución estadística de la frecuencia fundamental.

La distribución en el espectro de la frecuencia fundamental entre hombres y mujeres es distinta debido a que ambos poseen distintas características fisiológicas. En promedio la frecuencia fundamental en hombres se halla por los 125 Hz mientras que para mujeres tiende a ser más alto entorno los 250 Hz.

Figura 5.2: Distribución de la frecuencia

fundamental en hombres y mujeres.



Fuente: www.ub.edu/labfon/XVII-17.pdf

La dispersión de los valores en mujeres es mayor que en los hombres, pudiéndose encontrar valores entre 80 y 200 Hz para hombres y entre 150 y 400 para mujeres.

Debido a que la frecuencia fundamental posee un alto margen dinámico, su estimación se debe restringir a valores que van desde 75 a 500 Hz.

5.2 Programación de Static Tone en Max MSP.

Los objetos utilizados para la programación de Static Tone se encuentran detallados en el Anexo 3.

5.2.1 Etapa de la entrada de audio.

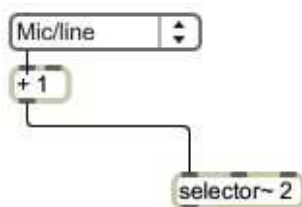
En esta sección del programa se crean las conexiones necesarias para poder analizar una señal proveniente desde un micrófono conectado a una interfaz de audio o analizar un archivo de audio desde un medio de almacenamiento de información permanente como un disco duro.

5.2.1.1 Conexiones de la etapa de entrada de audio.

La primera etapa del programa permite elegir entre la entrada de micrófono y la lectura de un archivo de audio utilizando el objeto selector~ 2 el que se controla por medio de un cuadro de menú al que previamente se le ha cargado las opciones: 0 Mic/line in y 1 Archivo de sonido.

Debido a que el objeto selector necesita en su entrada los números 1 y 2 para su funcionamiento se conecta el objeto +1 entre los objetos umenu y selector~ 2.

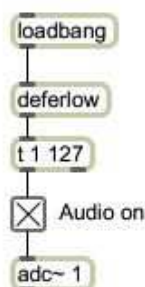
Figura 5.3: Etapa de selección de la entrada de audio o carga directa de archivo en el ordenador.



Fuente: Autores.

Lo siguiente es configurar la entrada de audio desde la tarjeta de sonido. Esta entrada se enciende automáticamente al abrir la aplicación con la ayuda del objeto loadbang que envía un mensaje bang a la entrada del objeto deferlow para que el mensaje tenga menor prioridad del que llega al selector~ 2. Luego se conecta al objeto trigger 1 127 cuya primera salida enciende primero al objeto toggle y luego al objeto adc~ 1. Con estos pasos se crea una entrada de audio que se conectada a la segunda entrada de selector~ 2.

Figura 5.4: Etapa de entrada de audio desde una señal de micrófono o línea.



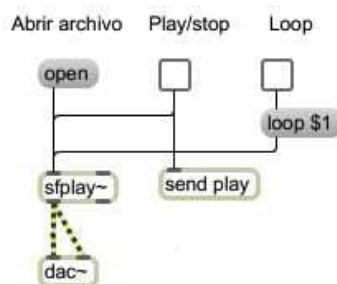
Fuente: Autores.

La siguiente etapa consiste en crear la opción para cargar archivos de audio desde un medio de almacenamiento, esto se lo realiza por medio del objeto `sfplay~` al que se le ha conectado a su entrada los controles de `play/stop`, `loop` y el mensaje `open` que abre una ventana de navegación para encontrar y abrir el archivo que se desee analizar.

La salida de audio del objeto `sfplay~` se conecta a la tercera entrada del `selector~ 2`.

A continuación se agrega un objeto `dac~` para el monitoreo auditivo de los archivos de audio. Aunque este objeto no tiene relación con la entrada de audio se lo incluye en esta etapa del desarrollo ya que se conecta después del objeto `sfplay~`.

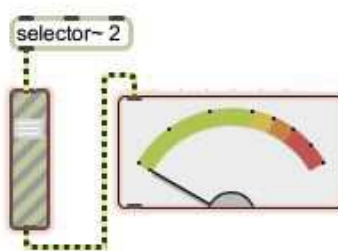
Figura 5.5: Etapa de carga de archivos de audio y control de reproducción.



Fuente: Autores.

Como último paso se agrega un `gain~` a la salida del `selector~ 2` para controlar el nivel con que se analiza a la señal de voz con la ayuda visual de el `levelmeter` que se conecta a continuación.

Figura 5.6: Control de nivel de la señal previo al análisis.

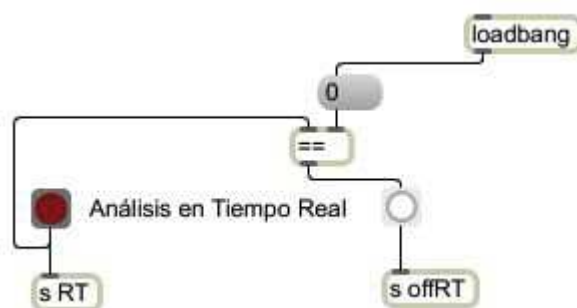


Fuente: Autores.

En algunas de etapas de creación del programa se utilizan los objetos send y receive para operaciones de encendido, apagado y reiniciación automático, estos funciona enviando y recibiendo mensajes bang en distintas etapas del programa.

Como caso especial de encendido y pagado se tiene al control de análisis en tiempo real, que utiliza un objeto led y el objeto == cuyo propósito es enviar un bang cuando se detiene el análisis en tiempo real, estos objetos controlan el flujo de la señal que es enviada a la etapa de análisis.

Figura 5.7: Control de análisis en tiempo real.



Fuente: Autores.

5.2.2 Etapa de análisis de la frecuencia fundamental:

En esta etapa se puede utilizar varias formas de caracterizar a la voz ya sean acústicas o de representación. En este caso se utiliza un sub patcher que analiza la frecuencia fundamental en tiempo real y obtiene los siguientes datos estadísticos:

- ♦ **Frecuencia promedio:** Este dato estadístico muestra la frecuencia promedio de altura tonal a lo largo del tiempo.
- ♦ **Frecuencia mínima:** Este dato estadístico indica la altura tonal mínima hallada en el análisis a lo largo del tiempo.
- ♦ **Frecuencia máxima:** Este dato estadístico indicar la altura tonal máxima hallada en el análisis a lo largo del tiempo.

- ♦ **Desviación estándar:** Este dato estadístico indica en cuanto ha variado los máximos y mínimos de la frecuencia promedio a lo largo del tiempo.
- ♦ **Oblicuidad (*skewnees*):** Este dato estadístico cuyo coeficiente indica el grado de oblicuidad en el que se encuentra una gráfica de distribución de probabilidad.
- ♦ **Curtosis:** Esta dato estadístico determina el grado de concentración que presentan los valores en la región central de la distribución en un análisis gráfico.

5.2.2.1 Proceso de creación del sub patch de la etapa de análisis estadístico de la frecuencia fundamental.

Este sub patch se compone de dos partes principales una es la de estimación de la frecuencia fundamenta por medio del objeto analyzer~ y del análisis estadístico utilizando el objeto lp.stacey.

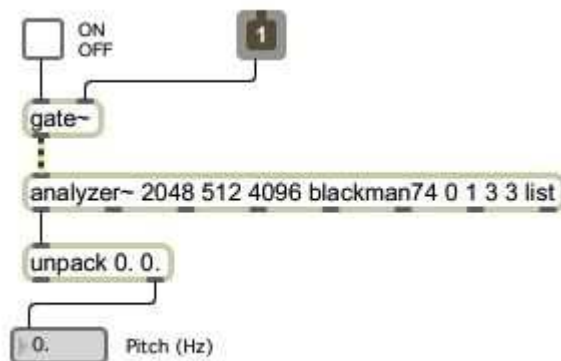
El primer paso en la creación del sub patcher es crear el objeto patcher analysis donde se insertan los objetos utilizados en este sección. Al crear el objeto p analysis se abre una nueva ventana de patcher donde se coloca un objeto inlet para crear una entrada desde el patcher principal. La salida del inlet se conecta a la segunda entrada del objeto gate~ que controla el flujo de la señal hacia la entrada del objeto analyzer.

El objeto analyzer entrega en su primera salida una serie de datos en forma de lista correspondientes a la estimación del la frecuencia fundamental en formatos MIDI y Hz.

Los argumentos utilizados en el objeto analyzer se establecieron buscando similitud en la estimación del pitch con programa gratuito de análisis *PRAAT* utilizando varias muestras de voz de prueba.

La estimación de la frecuencia en Hz se obtiene después de usar el objeto unpack 0. 0. para separar los datos MIDI de los de frecuencia en Hz, los resultados obtenidos son presentados en un cuadro de número.

Figura 5.8: Etapa de estimación de la frecuencia fundamental en Hz.



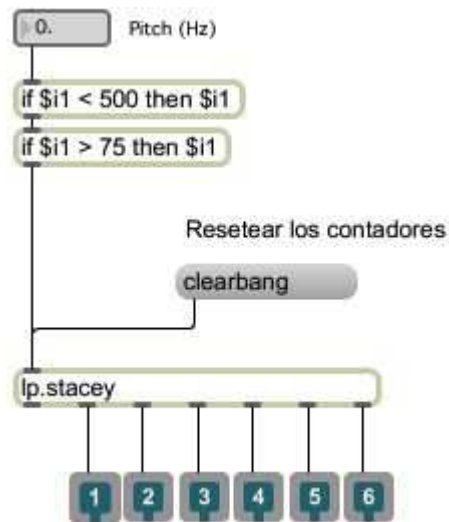
Fuente: Autores.

Después de obtener una serie de datos de la frecuencia fundamental es necesario restringirlos a valores entre 75 y 500 Hz con el objeto `if`. Estas frecuencias corresponden al margen dinámico que posee la frecuencia fundamental en las personas adultas y niños, las frecuencias superiores e inferiores pueden ser consideradas como falsas estimaciones correspondientes a armónicos superiores o patologías en el sistemas fonador.

Con la serie de valores restringidos se los envía a la entrada del objeto `lp.stacey` para calcular un conjunto de valores estadísticos que caracterizaran a la señal de voz de una persona. Se utiliza el mensaje `clearbang` para reiniciar los contadores y los cálculos que realiza el objeto.

En las salidas de `lp.stacey` se conecta seis objetos `outlet` para transmitir la información hacia el patcher principal.

Figura 5.9: Etapa de cálculo estadístico de datos de la frecuencia fundamental



Fuente: Autores.

5.2.3 Etapa de almacenamiento de información:

En esta etapa se recopila en una base de datos la información obtenida después de calcular los datos estadísticos de la frecuencia fundamental de una señal de voz.

5.2.3.1 Creación de la etapa de almacenamiento de información.

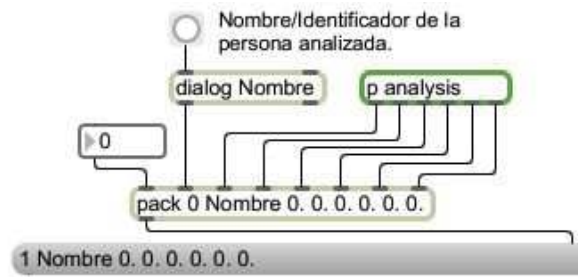
En esta etapa se debe guardar las características de la frecuencia fundamental de cada persona con un nombre en una base de datos para poder identificarlas y compararlas.

Después de analizar una señal de voz con el sub patcher análisis se debe anteponer un número y un nombre único a los datos estadísticos para combinarlos en una sola lista con el objeto pack.

El formato establecido para la base de datos es el siguiente:

- ◆ # Nombre mínimo promedio máximo desviación estándar skew kurtosis.

Figura 5.10: Etiquetado y empaquetado en una lista de los datos analizados.

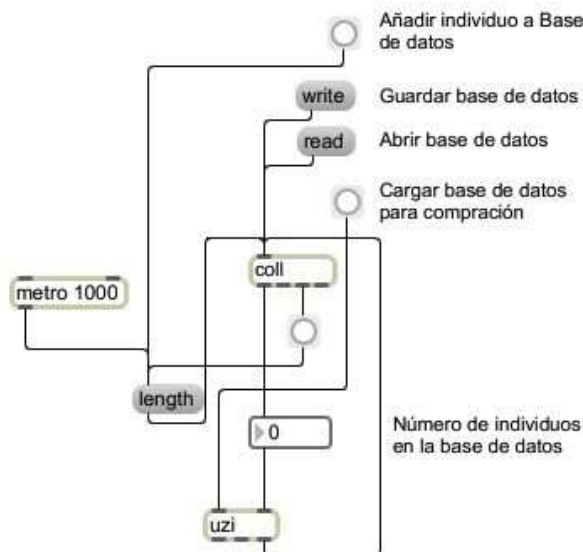


Fuente: Autores.

Para agregar un nombre de identificación a cada lista se utiliza el objeto dialog que es accionado por un bang, esto abre una ventana donde se escribe el nombre que se desee agregar a la voz que se va analizar y luego se oprime el botón OK.

La base de datos se crea recopilando varios datos de análisis de distintas personas y enviándolos a la entrada del objeto coll al que se le ha agregado mensajes de control que permiten agregar una nueva persona a base de datos, guardar y abrir bases de datos y un mensaje length que indica cuantas listas (personas) están almacenadas.

Figura 5.11: Etapa de creación y edición de la base datos.



Fuente: Autores.

Para recuperar y cargar la base de datos para comparación se debe accionar con un bang el objeto uzi que envía una cuenta ascendente números a la entrada de coll para recuperar las listas de la base de datos en su salida.

Finalmente se agrega un objeto metro 1000 que envía un bang al mensaje lenght cuya finalidad es actualizar cada 1000 milisegundos el número de individuos que están almacenados en la lista.

5.2.4 Comparación de similitud:

En esta etapa se busca similitudes entre las personas analizadas y registradas en la base de datos mostrando en su salida el nombre o identificación utilizados cuando se registro en la base de datos. Para el cómputo de la similitud se calcula la distancia aritmética de cada persona analizada.

5.2.4.1 Conexiones de la etapa de comparación.

En esta etapa se divide en dos partes: en la primera se hace una identificación rápida del nombre de una persona desconocida dentro de un conjunto de varias personas que fueron previamente analizadas y registradas dentro de la base de datos y en la segunda parte se calcula una puntuación después de evaluar de las características obtenidas para cada una de las personas con la persona desconocida.

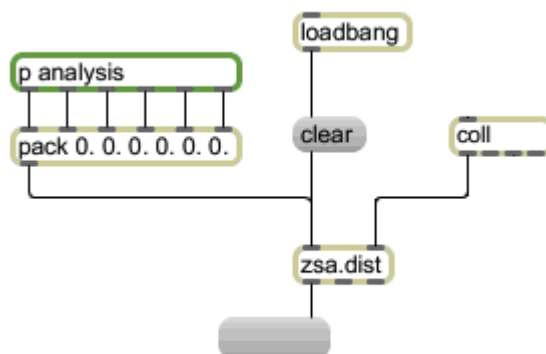
El objeto que realiza la comparación es el zsa.dist. En su primera entrada se envía los datos de analizados y empaquetados de la persona de la que se desconoce su identidad y en su segunda entrada la base de datos proveniente del objeto coll que son accionados con el button conectada al objeto uzi creados en la etapa anterior.

Es necesario enviar el mensaje clear para borrar bases de datos de análisis previos esto se lo realiza automáticamente al abrir el patcher por medio de un loadbang.

La identificación rápida se la obtienen en salida del objeto zsa.dist el que se

conecta un cuadro de mensaje que indica el nombre de la persona identificada.

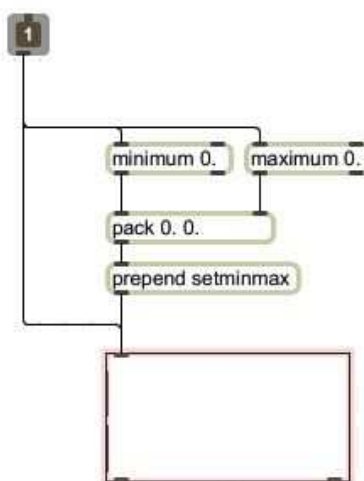
Figura 5.12: Conexiones de la etapa de identificación rápida.



Fuente: Autores.

Para la realización de una tabla de puntuación se crea un sub patcher denominado score con dos entradas: en la primera se conecta la segunda salida del objeto zsa.dist donde se tiene una lista con las distancias aritméticas del análisis de cada una de las personas en la base de datos con la voz desconocida y en la segunda entrada se conecta la primera salida del objeto coll con el propósito de obtener los nombres de las personas.

Figura 5.13: Conexiones para la creación del gráfico de barras de la puntuación.

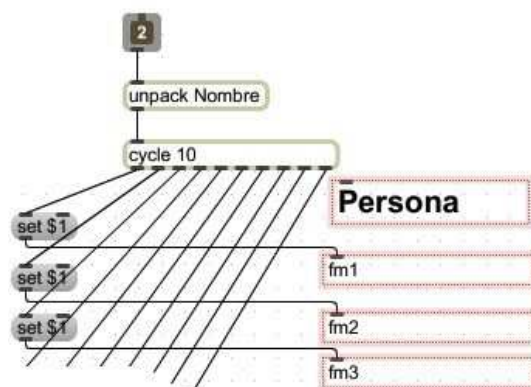


Fuente: Autores.

Dentro del sub patcher se crea un grafico de barras de las puntuaciones obtenidas del primer inlet con el objeto multislider. La información que se grafique debe ser normalizada de modo que la distancia con menor valor sea igual a cero y la de mayor valor sea igual a uno, esta normalización se calcula automáticamente en el objeto multislider estableciendo el valor mínimo y el máximo con los objetos minimum y maximum respectivamente los que a su vez son unidos con el objeto pack para después agregar el mensaje setminmax con el objeto prepend para dar el formato que puede entender el objeto multislider.

Para poder listar los nombres de las personas registradas que ingresan al sub patch por el segundo inlet se utilizo el objeto unpack que separa la información adicional y dejar solo los nombres como una serie de datos, a continuación se utiliza el objeto cycle para separar a cada uno de los nombres en salidas independientes las que son conectadas a un cuadro de mensaje que le agrega el mensaje set que permite visualizar la información mediante un cuadro de comment.

Figura 5.14: Conexiones para la visualización de nombres de las personas.

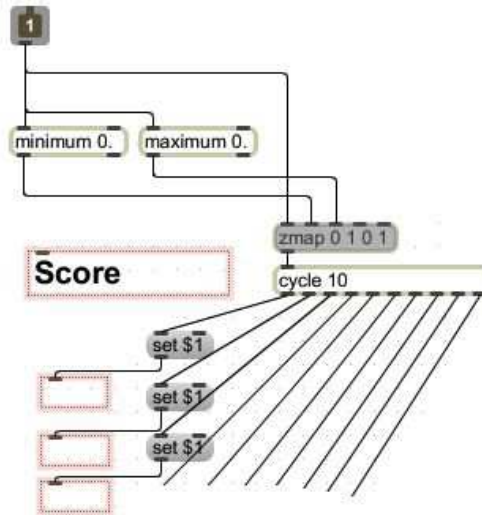


Fuente: Autores.

Finalmente para obtener la información numérica de la puntuación de cada persona se utiliza el objeto zmap que calcula la normalización cero a uno de las distancias obtenidas en el primer inlet utilizando los objetos minimum y maximum que se crearon previamente y conectándolos a la segunda y tercera entrada respectivamente, en la salida del zmap se obtiene una serie de datos numéricos que pueden ser visualizados del mismo modo que en la etapa del

listado de nombres mediante el objeto cycle, varios mensajes y comments.

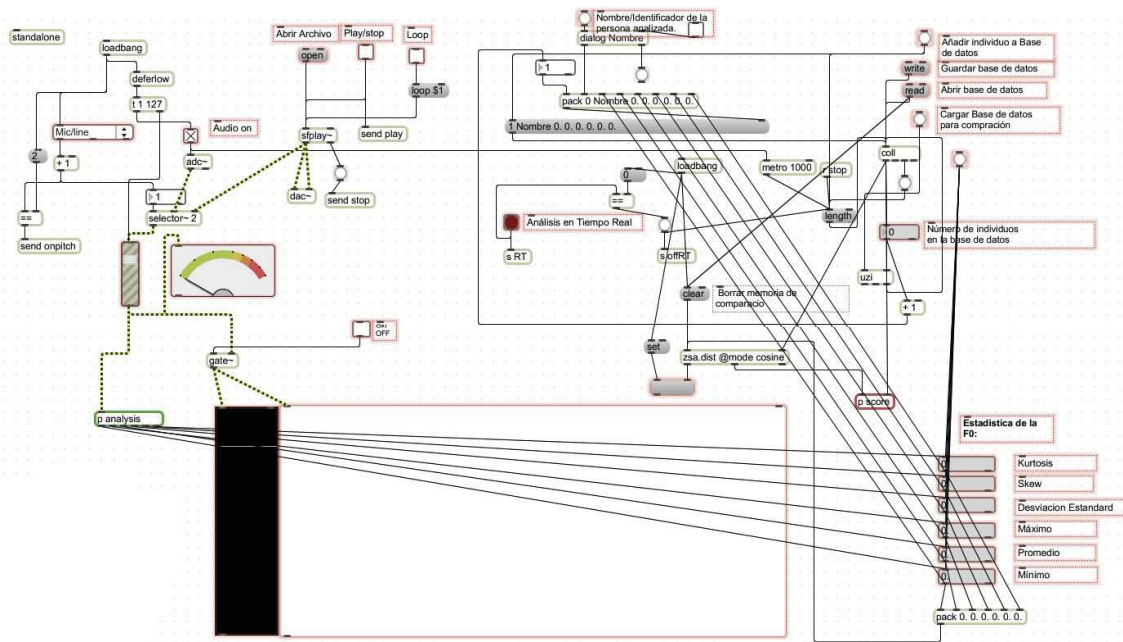
Figura 5.15: Conexiones para el calculo y visualización de las puntuaciones de cada persona.



Fuente: Autores.

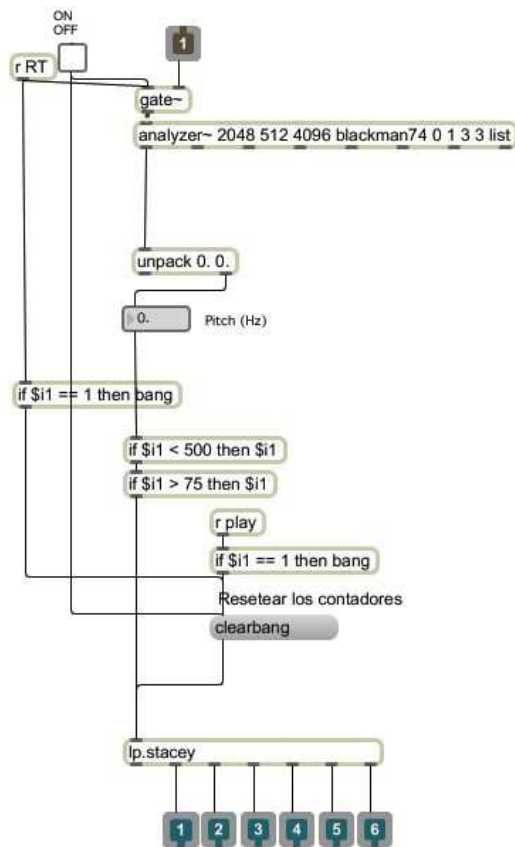
Con las conexiones del programa terminadas se procede a crear una interfaz de usuario a partir de las ventanas del patcher principal y del sub patcher score para que puedan ser utilizados como una aplicación independiente.

Figura 5.16: Vista general del patcher principal en modo edición.



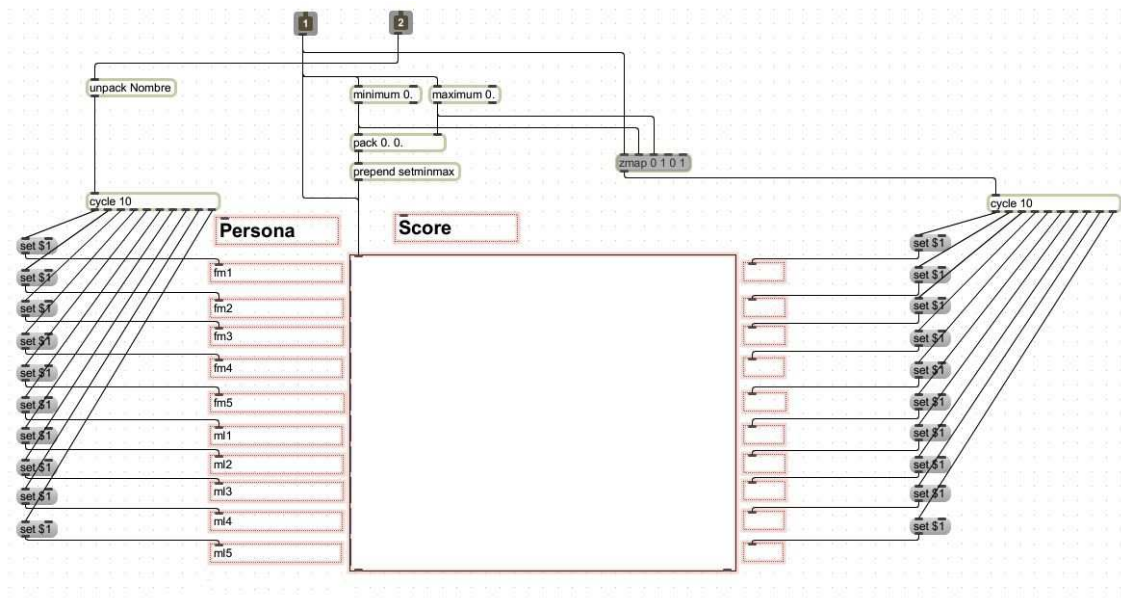
Fuente: Autores.

Figura 5.17: Vista general del sub patcher analysis en modo edición.



Fuente: Autores.

Figura 5.18: Vista general del sub patcher score en modo edición.



Fuente: Autores.

5.2.5 Interfaz de usuario.

En esta etapa se selecciona solo los elementos de control y visualización de mayor importancia para la crear una interfaz de usuario amigable para el usuario.

Es posible crear esta interfaz de usuario gracias a la función que posee Max MSP denominada Presentación que permite seleccionar, mover y redimensionar los objetos y cables que el usuario requiera en el momento de utilizar el programa, para evitar la distracción visual que producen los diferentes objetos y cables usados en el patcher.

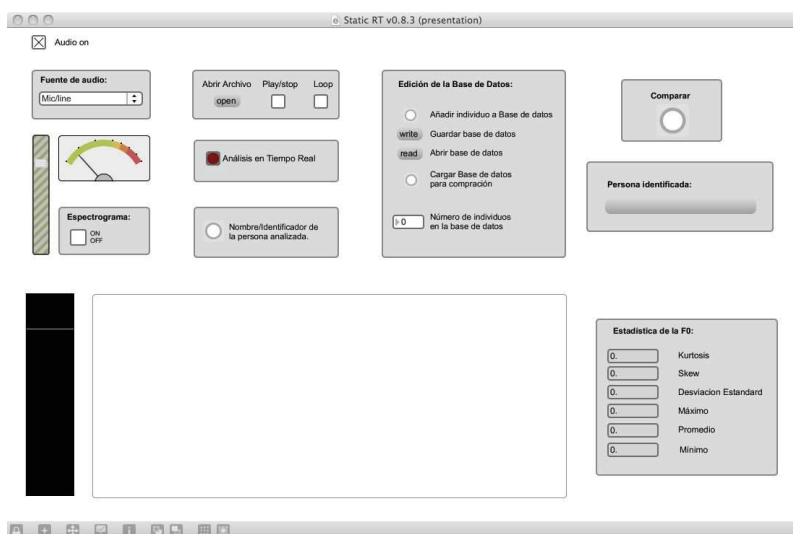
5.2.5.1 Proceso de creación de la interfaz de usuario.

Para la creación de la interfaz principal se secciona los objetos que se visualiza en el modo de presentación con la opción Agregar a Presentación en el menú Objeto, entre estos se selecciona botones, mensajes, meters, analizadores visuales, cuadros de número, comentarios, paneles y el control de nivel.

También se añade dos objetos espectroscope que pueden ser encendidos por medio del toggle conectado al objeto gate, estos analizadores sirven como representación visual del espectro, la frecuencia fundamental las formantes de la señal que está siendo analizada.

La interfaz utilizada sigue un orden lógico de procesamiento de la información que va de izquierda a derecha.

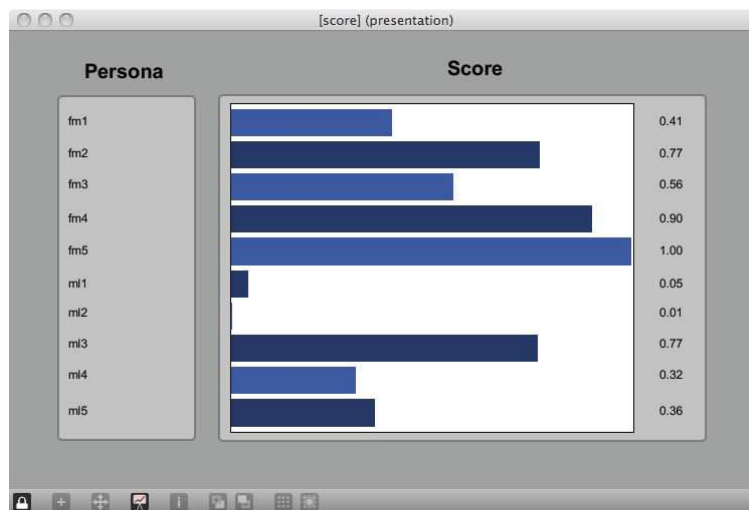
Figura 5.19: Interfaz de usuario principal del programa Static Tone.



Fuente: Autores.

Para la creación de la ventana de puntuación se selecciona el objeto multislider, las etiquetas de comentario y paneles.

Figura 5.20: Ventana de puntuación del programa Static Tone.



Fuente: Autores.

5.2.6 Exportación del patcher como una aplicación independiente

En esta etapa se exporta el patcher como una aplicación independiente que puede ser utilizada en ordenadores Mac con sistema operativo OSX sin la necesidad de tener instalado el entorno de programación Max MSP.

5.2.6.1 Proceso de configuración y exportación:

Antes de exportar el patcher a una aplicación independiente se debe colocar el objeto standalone en el patcher con el nivel más alto, este objeto no requiere ninguna conexión. Su configuración se la realiza en el inspector en donde es importante seleccionar el soporte de audio.

Para que la aplicación solo funcione en modo presentación se debe abrir el inspector del patcher dando clic derecho en un área vacía de la ventana del patcher y seleccionar Abrir en presentación.

Finalmente se exporta el archivo ejecutable desde el menú Archivo y se secciona Build Collective/ Application, se abre una ventana donde se da clic en Build. Aparece una venta de exploración donde se selecciona la carpeta en la que se guarda la aplicación y los archivos necesarios para su funcionamiento, a continuación se selecciona Crear Aplicación en el menú formato y se oprime en el botón Save.

5.3 Pruebas de funcionamiento.

5.2.1 Presentación de datos para análisis.

5.2.1.1 Keele Pitch Database.

Es una base de datos de voz usada para evaluar algoritmos de reconocimiento del pitch fue creada por Departamento de Comunicación y Neurociencia y el Departamento de Ciencias Computacionales de La Universidad Keele del Reino Unido.

Las señales fueron grabadas utilizando un texto fonéticamente balanceado correspondiente a la lectura "The North Wind Story". Las que fueron grabadas un en una habitación insonorizada utilizando un micrófono y grabador DAT.

La base de datos obtenida está compuesta de 10 locutores con habla nativa inglesa, cinco de ellos son hombres con edades entre 21 y 60 años y cinco son mujeres con edades entre 20 y 37 años, la lectura del texto tiene una duración de 27 a 40 segundos para hombres y de 30 a 38 segundos para mujeres.

Las grabaciones de voz fueron guardadas utilizando la extensión .pes, y tienen una frecuencia de muestreo de 20 kHz, una profundidad de 16 bits y utilizan un solo canal. Paralelamente se grabó la señal de un laringógrafo con la misma frecuencia de muestreo y profundidad de bits con la extensión .pel que puede ser usada como una referencia de la estimación del pitch.

Además se adjunta una serie de archivos con la extensión .pev con una estimación del pitch con ventanas de 25.6 milisegundos y con intervalos de tiempo de 10 milisegundos entre cada estimación.

Para conseguir la base de datos se contactó vía correo electrónico con el Dr. Georg Meyer para obtener la información de acceso.

La base de datos puede ser descargada directamente desde un servidor ftp al que se puede acceder con la dirección y claves facilitadas por George Meyer.

La información detallada sobre la metodología utilizada en la creación de la base de datos se la puede encontrar en el (Anexo 4).

Para proceder con el análisis de prueba en el programa se debe convertir los archivos .pes al formato aiff con la utilización del programa editor de audio Cool Edit Pro según la recomendación del Dr. Georg Meyer.

5.2.1.2 Resultados del de la estimación del pitch y comparación con el programa PRAAT.

Para comprobar la funcionalidad del programa en la estimación del pitch se analizó los datos estadísticos de la frecuencia mínima, máxima y promedio de las 10 muestras de voz de la base de datos Keele y se comparó los resultados con los que se obtuvo al analizar las mismas voces con el programa *PRAAT*.

Tabla 5.1: Comparación de datos estadísticos de la frecuencia fundamental entre los programas PRAAT y Static Tone.

CANDIDATOS	PRAAT			STATIC TONE			
	MUJER	PROMEDIO	MIN	MAX	PROMEDIO	MIN	MAX
1		191,2	76,4	495	177,59	76	484
2		228,9	76	485	209,26	76	484
3		191,2	76,3	488	181,97	77	467
4		235,5	76,89	484,7	228,97	76	493
5		231,7	76,8	470	210,2	76	444
	HOMBRE	PROMEDIO	MIN	MAX	PROMEDIO	MIN	MAX
1		106,1	76	479,9	113,05	76	378
2		138,7	76,2	462	142,44	76	463
3		135,8	85,7	184	141,53	76	272
4		95,13	76	431	108,02	76	336
5		108	76,4	181	122,44	76	364

Fuente: Autores.

Tabla 5.2: Resultados de análisis en función del tiempo de muestra utilizado.

TIEMPO DE MUESTRA				
MUJER	COMPLETA	30 seg.	20 seg.	10 seg.
1	√	√	√	√
2	√	√	√	√
3	√	√	√	√
4	√	√	x	x
5	√	√	√	√
HOMBRE	COMPLETA	30 seg.	20 seg.	10 seg.
1	√	√	√	√
2	√	√	√	x
3	√	x	x	x
4	√	x	x	x
5	√	√	√	√

Fuente: Autores.

5.2.2 Análisis de resultados.

Por lo observado en la tabla 5.1 de la comparación entre el programa PRAAT y el programa elaborado Static Tone de determino que la variación máxima entre resultados de la frecuencia fundamental es de 20 Hz.

Se encontró que no existe mucha variación de resultados en la frecuencia mínima y máxima debido a que se limito de igual manera frecuencias

estimadas de 75 a 500 Hz.

Por lo observado en la Tabla 5.2 los factores de éxito del programa Static Tone son dependientes del tiempo que se analizó la muestra, de la inteligibilidad del audio registrado como la mayoría de softwares comerciales. Al utilizar la base de datos Keele Pitch para el análisis completo de la muestra, el programa Static Tone fue capaz de identificar con un 100% de efectividad la identidad de los locutores, mientras que utilizar los primeros 30 segundos de la muestra se determinó que hay un 80% de efectividad en la identificación, para 20 segundos hay un 70% de efectividad en la identificación y para los primeros 10 segundos hay una efectividad del 60%.

Tras analizar todos los resultados se concluye que Static Tone es un programa efectivo para la identificación de locutores basado en el análisis de la frecuencia fundamental y comparación automatizada con una base de datos de locutores previamente analizados y registrados cuya principal característica es de realizarlo en tiempo real.

En función a las distintas alternativas que existen para el análisis de voz deberían realizarse de manera simultánea la mayor cantidad de procesos puesto que esto aseguraría la identidad de un locutor.

Capítulo 6. Análisis costo/beneficio.

6.1 Elaboración del informe económico para el desarrollo, aplicación y ejecución del software propio con hardware adquirido.

Para poder aplicar un sistema de reconocimiento de voz eficiente es necesario tener una inversión inicial y un referente para saber del costo que esto implica.

Existen algunas alternativas para alcanzar este objetivo, existen actualmente softwares creados por empresas cuyos derechos son reservados; sin embargo se tiene como propósito poder desarrollar un programa de reconocimiento de voz, el cual por el desarrollo ingenieril reduciría en gran cantidad el presupuesto requerido.

A continuación se detallarán los costos aproximados para una inversión adecuada en la implementación de nuestro propio software con hardware adquirido y compararlo con productos comerciales.

6.1.1 Cotizaciones.

- ◆ **Ordenador:** A continuación se presenta el ordenador de buenas prestaciones aplicables para el peritaje forense.

Se puede utilizar una computadora marca Apple, modelo MacBook pro de 15 pulgadas, debido a que los computadores de esta empresa poseen mejor desempeño en aplicaciones multimedia y la proliferación de virus de computadora es menos frecuente para ordenadores de este tipo. Una ventaja de este ordenador en especial es que es portable.

Figura 6.1: Ordenador MacBook Pro.



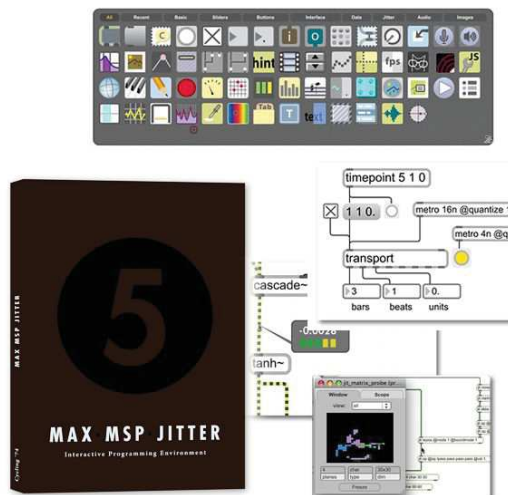
Fuente: http://store.apple.com/us/browse/home/shop_mac

[/family/macbook_pro?mco=MTM3NDc2NDA](http://store.apple.com/us/browse/home/shop_mac/family/macbook_pro?mco=MTM3NDc2NDA)

Este ordenador se lo puede encontrar en varias versiones, sin embargo se selecciono la que posee un procesador "Intel Core i5" con cuatro GB de memoria RAM lector de tarjetas SD, y tarjeta de video "Nvidia GeForce GT 330M" con 256 MB de memoria. Este ordenador esta avaluado por 1999.00 dólares americanos, cotizados directamente desde el sitio web oficial de Apple (<http://store.apple.com>).

- ◆ **Software:** Se selecciona como software de programación a "Max MSP" desarrollado por "Cycling 74" debido a su interfaz gráfica y su facilidad de programación aplicada al audio, se encuentra actualmente en la versión 5.1.5 cuya fecha de actualización fue el 26 de Agosto del 2010.

Figura 6.2 Max MSP.



Fuente: <http://www.taringa.net/posts/mac>

[/5542569/Max-MSP-5_06-Jitter.html](http://www.taringa.net/posts/mac/5542569/Max-MSP-5_06-Jitter.html)

En sus versiones iniciales este programa trabajaba únicamente con datos MIDI y en versiones únicamente para sistemas operativos Macintosh, hasta sus versiones para audio en el año 1997.

Este programa está valorado en 495.00 dólares americanos, cotizados en sitio web (<http://cycling74.com/shop/max>) propio de su empresa.

El modo de adquirirlos puede ser con la petición del CD impreso o por descarga directa por enlace de internet.

- ♦ **Monitoreo:** Es importante para la sección del monitoreo se lo haga con audífonos de tipo cerrado, debido a cualquier ruido filtrado desde el exterior puede encarecer el análisis auditivo.

Se seleccionó a los audífonos K44 de la empresa australiana “AKG Acoustics” encargada de fabricar transductores de gama profesional.

Figura 6.3: Audífonos AKG K44.



Fuente: [http://www.akg.com/site/products/powerslave,
id,328,nodeid,2,_language,EN.html](http://www.akg.com/site/products/powerslave,id,328,nodeid,2,_language,EN.html)

Su principio de transducción es de tipo dinámico, con un rango de frecuencias de 18 a 20,000 Hz, posee una impedancia de 32 ohms, conector mini jack convertible a jack plug de 6.3mm. Estos audífonos están valorados en 49.99 dólares americanos cotizados en la página web de la tienda Estado Unidense “Guitar Center” (www.guitarcenter.com) distribuidor directo de varias marcas.

Este audífono posee un buena relación costo beneficio, debido a que brinda un gran desempeño a un bajo costo.

- ♦ **Micrófono:** Para la toma de una muestra de voz se hace necesario trabajar con micrófonos de gama profesional y especializados a la transducción de voces humanas, debido a su relación costo benéfico se han seleccionado micrófonos de tipo dinámicos, para lo cual se seleccionó el siguiente:

Se escogió al micrófono profesional vocal “D5” de la empresa Australiana “AKG”.

Figura 6.4: Micrófono AKG D5.

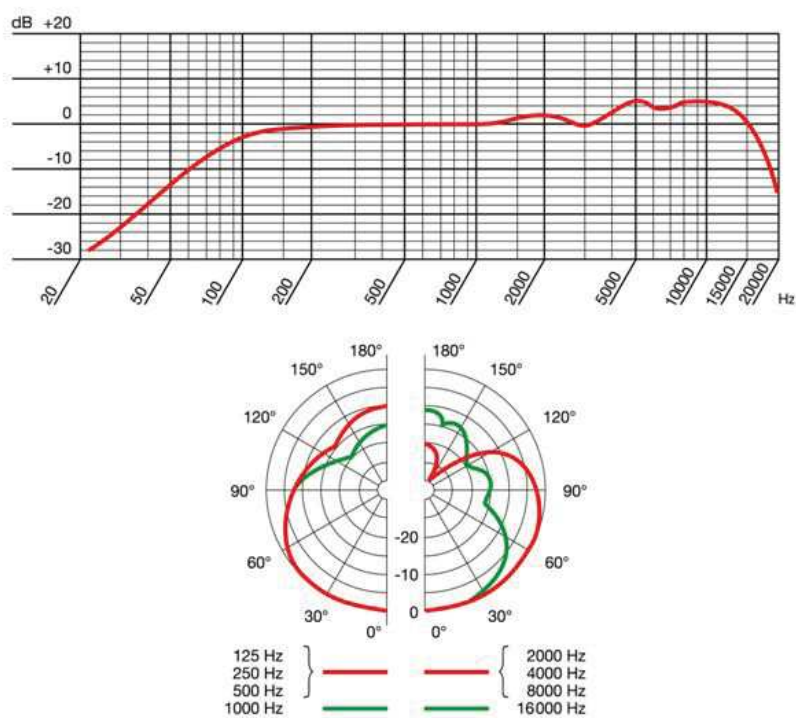


Fuente: www.ake.com/site/products/powerslave

,id,980,pid,980,nodeid,2,_language,EN.html

Este micrófono posee altas prestaciones como la de ser de patrón polar súper cardiode, con una respuesta de frecuencia de 70 a 20,000 Hz.

Figura 6.5: Respuesta de frecuencia y patrón polar del micrófono AKG D5.



Fuente: http://www.ake.com/site/products/powerslave,id,980,pid,980,nodeid,2,_language,EN,view,diagram.html

Además posee una sensibilidad de 2.6 mV/Pa (-52dBV), una impedancia mayor o igual a 600 ohms, y un conector de tres pines XLR.

Estos micrófonos están valorados en 99.99 dólares americanos, cotizados en la página web de la tienda Estadounidense “Guitar Center” (www.guitarcenter.com) distribuidor directo de varias marcas.

- ♦ **Interface de audio:** Se selecciona la interface “UltraLite MK3” de la empresa estadounidense “Motu” especializada en la fabricación de hardware y software para audio y video profesional.

Figura 6.6: Interface Motu UltraLite MK3.



Fuente: <http://www.motu.com/products/motuaudio/ultralite-mk3/hybrid.html>

Esta interface posee una característica en especial que la hace diferente a sus competidoras y la que presenta una conexión híbrida, eso quiere decir que se la puede vincular a una computadora ya sea por conexión Firewire 400 y la USB 2.0.

Su frecuencia de muestreo puede llegar hasta los 192 kHz y alcanzar una profundidad de 24 bits, un su etapa de entrada posee dos conectores de señal de micrófono con sus respectivos preamplificadores, además posee seis conectores para nivel de línea.

En su etapa de salida posee diez conectores analógicos de nivel de línea.

Cabe mencionar que además de tener un diseño moderno, puede operar como un mezclador digital de 36 bits, sin la necesidad de estar conectada

a una computadora. Posee una pantalla LCD de 16x2 caracteres como indicador de entradas y salidas habilitadas así como el tipo de proceso interno (efecto por DSP) a realizarse.

Esta interface está valorada en 549.00 dólares americanos, cotizados en la página web de la tienda Estadounidense “Guitar Center” (www.guitarcenter.com) distribuidor directo de varias marcas.

Se seleccionó este interface por su relación costo beneficio y su función en el ámbito de la acústica forense.

- ♦ **Regulador de Voltaje:** Se seleccionó este regulador de voltaje Tripp-lite VR1208R debido a si buena marca y su popularidad en el mercado, se encuentra valorado en 24,99 dólares americanos cotizados en ebay.

Figura 6.7: Regulador de voltaje Tripp-lite VR1208R.



Fuente: <http://www.intelcompras.com/tripp-lite-regulador-tension->

tripplite-600w-1200va-supresion-picos-ruidos-contactos-prot-linea-p-45342.html

- ♦ **Cable:** Se seleccionó al cable Rapco SJ 10 K debido a si buena marca y por su calidad de componentes, se encuentra valorado en 19,99 dólares americanos cotizados en guitar center.

Figura 6.8: Cable Rapco SJ 10K.



Fuente: http://www.clavedesol.com.ar/Rapco/rapco_cables.HTM

- ♦ **Pedestal:** Se seleccionó al pedestal Hercules MS100B debido a su buena marca y por su practicidad para uso en laboratorio, se encuentra valorado en 49,99 dólares americanos cotizados en guitar center.

Figura 6.9: Pedestal Hercules MS100B.



Fuente: <http://www.musiccenter.com/i-2862>

[/hercules-ms100b-stalak-za-mikrofon](#)

- ♦ **Reproductor y grabador de mini casete:** Aunque actualmente se encuentran discontinuados este tipo de dispositivos, es importante que un laboratorio pericial de audio posea todo tipo de reproductores para sus

análisis, se cotizó a este reproductor y grabador Sony M-450 valorado en 149.99 dólares americanos en: www.amazon.com.

Figura 6.10: Grabador reproductor Sony M-450.



Fuente: www.amazon.com/Sony-M-450-Microcassette-Recorder-battery/dp/B0000615TG

- ♦ **Radio grabadora CD y Casette:** Es importante que un laboratorio pericial de audio posea todo tipo de reproductores para sus análisis, se cotizó a este reproductor y grabador Sony CFD-S07 valorado en 104,22 dólares americanos en: www.mercadolibre.com.

Figura 6.11: Reproductor y grabador Sony CFD-S07.



Fuente: <http://articulo.mercadolibre.com.ec/MEC-7446867->

radiograbadora-sony-cfd-s07-kissu-con-garantia-_JMFuente

- ♦ **Lector de tarjetas de memoria externo:** Actualmente en el mundo digital es muy común encontrar todo tipo de tarjetas, por lo tanto el laboratorio debe contar con este dispositivo marca Usa Net, modelo UC SD 33-61-4

con capacidad de reproducir tarjetas tipo C-Flash I y II, T-Flash, SD, Mini SD, XD, MMC, MS, MS PRO, MSDuo y MSPRO Duo; y conectividad USB, el cual está valorado en 15 dólares americanos en ebay.

Figura 6.12: Lector multi tarjeta de memoria externa.



Fuente: <http://articulo.mercadolibre.com.ec/MEC-7446867>

-radiograbadora-sony-cfd-s07-kissu-con-garantia-_JMFuente:

6.2 Presupuesto Referencial.

Los equipos seleccionados para realizar un óptimo encadenado electro acústico son listados en la siguiente tabla.

Tabla 6.1: Inventario de equipos necesarios.

DISPOSITIVO	MARCA	REFERENCIA	COSTO (USD)
Ordenador	Apple	Mac Book Pro	1.999,00
Interface	Motu	UltraLite MK3	549,00
Audífonos	A.K.G.	K 44	49,99
Micrófono	A.K.G.	D 5	99,99
Cable	Rapco	SJ 10 K	19.99
Pedestal	Hercules	MS100B	49,99
Regulador de voltaje	Tripp-lite	VR1208R	24.99
Reproductor mini casete	Sony	M-450	149.99
Radio grabadora	Sony	CFD-S07	104,22
Lector multitarjeta	Usa net	UC SD 33-61-4	14,99
Software	Cycling 76	Max/MSP	495,00

Fuente: Autores.

Sin embargo algunos de estos costos no incluyen impuestos y costos de envío propios del país para la compra de productos, por lo tanto se investigó sus tributos correspondientes dando como resultado el siguiente total cotizados en la ciudad de Miami en el Estado de Florida, debido a que el pasaje aéreo Quito-Miami es el de menor costo para ingresar a Estados Unidos, a demás el precio promedio del pasaje es de aproximadamente 494 dólares.

Tabla 6.2: Costo total de inventario de equipos.

REFERENCIA	COSTO	ENVIÓ	IMPUESTO (TAX 7%)	TOTAL
Mac Book Pro	1.999,00	0,00	139,93	2.138,93
Max/MSP	495,00	0,00	34,65	529,65
K 44	49,99	0,00	3,50	53,49
D 5	99,99	0,00	7,00	106,99
UltraLite MK3	549,00	7,03	38,43	594,46
Tripp-lite	30	0,00	2,10	32,10
Rapco	20	0,00	1,40	21,40
Hercules	25	0,00	1,75	26,75
M-450	150,00	0,00	10,50	160,50
CFD-S07	104,22	0,00	7,30	111,52
UC SD 33-61-4	15,00	0,00	1,05	16,05
TOTAL				3.749,04

Fuente: Autores.

Cabe recalcar que según la página web oficial de Aduanas del Ecuador en uno de sus literales dice que se pueden transportar:

“Bienes de uso profesional, herramientas y equipos necesarios para el desempeño de funciones o actividades laborales propias del viajero (únicamente los que sean portátiles y de fácil transportación por parte del viajero), situación que deberá justificarse”.³

6.3 Rentabilidad del proyecto aplicado a Ecuador.

Para determinar una tasa de rentabilidad de este tipo de servicio se hace necesario investigar sobre los costos alternos aproximados para la implementación y funcionamiento de un laboratorio sencillo, además de la inversión inicial en equipamiento, esto es basado en costos aproximados en nuestro país actualmente, para lo cual se genero la siguiente tabla.

Tabla 6.3: Consumo eléctrico aproximado mensual.

UNID.	BIEN ELÉCTRICO	CONSUMO ELÉCTRICO	TIEMPO HORAS AL MES	CONSUMO kW/h
1	Mac Book Pro	85 W/h	160	13,6
5	Toma Corrientes	70 W/h	100	7
3	Focos	40 W/h	160	6,4
COSTO kW/h (dólares) = 0,078			TOTAL CONSUMO kW/h	27
Valor Facturable				2,106
Otros Valores a pagar				11,69
TOTAL A PAGAR				13,796

Fuente: Autores.

³ “Aduanas del Ecuador” Web <http://www.aduana.gov.ec/contenido/procViajeros.html>

Una vez obtenido el consumo eléctrico aproximado del laboratorio, se puede sacar un total de operación mensual que puede poseer un laboratorio, sin incluir su inversión inicial y el costo ingenieril del desarrollo del software.

Tabla 6.4: Costo de operación mensual sin inversión inicial.

UNID.	TIPO	COSTO MENSUAL (USD)
INFRAESTRUCTURA		
1	Arriendo	250
INSUMOS		
1	Hojas de papel	20
1	Material de oficina	30
SERVICIOS BÁSICOS		
1	Agua	5
1	Teléfono	20
1	Total consumo eléctrico	13,79
1	Internet	20
1	Servicios contables	50
RECURSO HUMANO		
1	Secretaria	240
2	Peritos	1000
TOTAL OFICINA		1648,79

Fuente: Autores.

A esta tabla se la ha separado de la inversión inicial y el costo ingenieril, debido a que este costo se puede financiar al tiempo que se desee recuperar la inversión, estos gastos son correspondientes a los activos fijos que posee el laboratorio.

Tabla 6.5: Tabla activos fijos del laboratorio.

UNID.	INVERSIÓN INICIAL	COSTO (USD)
1	Equipos de laboratorio	3.749,04
BIENES Y ENCERES		
2	Escritorios	500
6	Sillas	360
EQUIPOS DE COMPUTO		
1	Impresora	75
TOTAL		4684,04

Fuente: Autores.

Tabla 6.6: Tabla del precio final del software a cinco USD la hora.

PROCESO	TIEMPO (HORAS)	COSTO (USD)
Análisis de proyecto	448	2240
Diseño	400	2000
Programación	200	1000
Pruebas de funcionamiento	148	740
PRECIO FINAL SOFTWARE		5980,00

Fuente: Autores.

Una vez obtenidos los costos del software, conjuntamente con los activos fijos, estos se los debe diferir al tiempo deseado de recuperación de la inversión el cual se decidió que será de cinco años, para lo cual se explica en la siguiente tabla:

Tabla 6.7: Tabla de financiamiento del equipamiento mas software.

EQUIPAMIENTO	COSTO (USD)
Precio final software	5.980,00
Costo Activo Fijo	4684,04
TOTAL	10.664,04
Tiempo de financiamiento	60 meses
TOTAL MENSUAL	177,74

Fuente: Autores.

Tabla 6.8: Tabla de facturación mínima mensual del laboratorio.

FACTURACIÓN MÍNIMA MENSUAL	
COSTO OPERACIÓN DE OFICINA	1648,79
COSTO FINANCIADO	177,74
TOTAL	1826,53

Fuente: Autores.

Al obtener la facturación mínima mensual, se puede calcular el costo unitario del producto, en este caso será la unidad de análisis forense para un registro de audio; por lo tanto para sacar dicho costo se procedió a dividir el mensual de facturación con la frecuencia mínima del producto que se aspira que sea a un peritaje diario por día laboral del laboratorio que es de 20 análisis por mes.

Tabla 6.9: Tabla del costo unitario del análisis.

COSTO PROMEDIO POR ANÁLISIS	
Facturación mínima mensual	1826.53
Frecuencia mínima de análisis	20
TOTAL	91,33

Fuente: Autores.

Los valores podrían ser más altos o menores dependiendo de la frecuencia de casos en el laboratorio o también en el cobro de sueldo del recurso humano, ya sea en peritos o extras.

Capítulo 7. Proyecciones.

- ◆ Aplicación de guías metodológicas periciales que sean utilizadas como futuras normativas una vez aprobadas por los organismos competentes.
- ◆ Creación de bases de datos de voces estandarizada como un nuevo proyecto ecuatoriano que facilite en gran medida al desarrollo investigativo en varias de áreas de la ciencia como: la verificación de locutor con fines forenses, las telecomunicaciones e interacción del habla con computadoras.
- ◆ Desarrollo de mejores softwares para identificación y verificación de locutor utilizando de forma complementaria técnicas innovadoras de caracterización y clasificación de las muestras de voz o mejorar las técnicas existentes.
- ◆ Realización de departamentos de restauración de audio donde ciudadano común pueda llevar sus registros sonoros y transfórmalos a un formato de mejor calidad para su conservación.
- ◆ Creación de un museo ecuatoriano de formatos de audio que posean la mayor cantidad de grabadores y reproductores con fines didácticos e investigativos.
- ◆ Creación en el país de auditecas que recopilen registros sonoros que motiven la investigación y desarrollo científico tales como: efectos de sonido, grabaciones históricas, bancos de voces y musicología étnica.
- ◆ Creación de una base de datos estandarizada de voces a nivel nacional, la cuál facilitaría el desarrollo investigativo en varias de áreas de la ciencia como: la verificación de locutor con fines forenses, las telecomunicaciones y la interacción del habla con computadoras.
- ◆ Creación de nuevos métodos de análisis de voz con muestras de audio que posean distintas calidades de registro o distintos formatos tales como

grabadoras de audio en teléfonos celulares, grabadoras de video, archivos mp3 o aac,

- ♦ Análisis el rendimiento del software de una manera más estandarizada utilizando métodos tal como EER (*Equal Error Rate*).

Capítulo 8. Conclusiones y recomendaciones.

8.1 Conclusión general.

Se concluye que sí es posible la realización de una guía metodológica que logra unificar los conceptos técnicos referentes al audio, fonética y procedimiento judicial recopilado de distintas bibliografías para la creación de un nuevo documento que puede ser utilizado como referencia estandarizada para futuros peritos que se dediquen al análisis de voz en el Ecuador. Esta guía metodológica está enfocada al análisis comparativo utilizando los métodos de análisis espectral de formantes y frecuencia fundamental como técnica de análisis de verificación de la identidad del locutor basada en software, en la que se demuestra que si se puede realizar efectivamente con presupuesto accesible.

8.2 Conclusiones específicas.

- ◆ Metodologías de aplicación técnica estandarizadas son muy difíciles de encontrar o simplemente no existen por lo tanto se hace necesario complementar con conocimientos de algunas ramas de la ciencia como jurisprudencia, tratamiento de audio, procesamiento digital de señales, investigación de campo.
- ◆ Es posible el diseño de una guía metodológica basada en la operación actual de organismos competentes en el Ecuador añadiendo información relevante proveniente de una investigación técnica, además se determinan ausencias en el procedimiento técnico que pueden ser mejoradas y aplicadas. Proponiendo a este como uno de los primeros documentos de este tipo en el campo forense ecuatoriano.
- ◆ Se obtuvo un porcentaje del 100% en la correcta identificación de los locutores dependiente del texto, debido a la utilización de muestras de voz de excelente calidad, por lo tanto se concluye que el software creado es de impecable calidad técnica en análisis estadístico de la frecuencia fundamental.

- ◆ Se afirma que el celo empresarial por la comercialización de sus productos, dificulta en gran medida la realización de nuevos softwares puesto que existe poco detalle de su funcionamiento y su metodología es patentada, por lo tanto se concluye se debe investigar en mayor escala para poder alcanzar competitividad en su mercado.
- ◆ La forma de programación basada en objetos facilita la creación de programas computacionales debido a que es más fácil de visualizar y desarrollar sus líneas de flujo y principio de funcionamiento que si se programara mediante métodos tradicionales de líneas de código.
- ◆ Actualmente es posible la creación de softwares de identificación de locutor mediante otros lenguajes de programación, pero se puede concluir que la programación mediante Max MSP se la realiza en menor tiempo.
- ◆ Económicamente se demostró que para la realización de softwares es viable sin una gran inversión, sin embargo con el apoyo de inversionistas y una mayor cantidad de tiempo "Static Tone version 0.9" puede alcanzar altos índices de utilidad.
- ◆ La creación del software Static Tone se facilitó en gran medida por la ayuda de objetos gratuitos para Max MSP, por lo tanto se afirma que el desarrollo de nuevos softwares en varias áreas del audio sean factibles.
- ◆ Los objetos externos de Max MSP están en constante desarrollo, y es posible un mejor desempeño tanto del programa como su funcionalidad siempre y cuando se los actualice.
- ◆ Se dificulta en gran medida la evaluación del software Static Tone con un banco de voces puesto adquisición estos bancos son costosas y para su realización requieren de una tecnificación que rebasa los límites de este proyecto.
- ◆ Actualmente en Internet se encuentran programas computacionales aplicables a la acústica forense que se pueden adquirir de manera gratuita como el programa PRAAT citado anteriormente, por lo tanto el

desarrollo de nuevos softwares deberá tener en cuenta la creación de nuevas prestaciones que las de estos programas.

- ◆ Se establece que la comercialización del software Static Tone en la versión 0.9 no puede ser rentable actualmente, debido a sus limitaciones y la presencia de programas gratuitos, por lo tanto su futuro desarrollo podrá permitir la competitividad en su comercialización.
- ◆ Debido a la diversidad de equipos, herramientas y formatos. El llegar a ser especialistas en el dominio de todos ellos requiere de revisión de manuales de usuario y especialización técnica haciendo de esta una tarea muy ambiciosa de realizar, por lo tanto se deberá investigar más a fondo los que tengan mayor relevancia.
- ◆ Los servicios periciales estatales son gratuitos pero consideran solo casos muy relevantes, sin embargo la masificación de peritos civiles acompañado del conocimiento público de estos servicios darán nuevas fuentes de empleo fortalecerá el sistema judicial, la fe y seguridad ciudadana.
- ◆ Se afirma que existe un descuido por los organismos competentes en esta área puesto que la poca confianza en la justicia ecuatoriana conjuntamente con el desconocimiento popular y esto desmotiva a la ciudadanía y perjudican al desarrollo de la acústica forense.
- ◆ Se conoce que los altos aranceles dificultan la adquisición de equipamiento técnico en Ecuador por lo tanto se concluye que para hacer viable de mejor manera a este tipo de proyectos se deberá adquirir equipos de preferencia en su lugar de fabricación.

8.3 Recomendaciones.

- ◆ Se recomienda el uso de interfaces de audio que posean excelente convertidores A/D y D/A y que posean frecuencias de muestreo superiores a 96 kHz y profundidades de bits superiores a 24.

- ◆ Implementar carreras ecuatorianas de especialización dedicadas exclusivamente al desarrollo en el área de la acústica forense.
- ◆ Sería de gran utilidad la creación de departamentos de peritaje acústico, de ser posible en todas las provincias del país puesto que existen una centralización de los análisis en las ciudades de Quito y Guayaquil.
- ◆ Resultaría útil la libre importación de equipamiento técnico en el área de sonido y acústica que impulse el desarrollo tecnológico y científico para aumentar la competitividad de Ecuador a nivel mundial.
- ◆ Se recomienda al departamento de justicia de poseer mayor confianza en el peritaje ecuatoriano ya que en la actualidad es una ciencia muy relevante.
- ◆ Se recomienda a futuros peritos acústicos adquirir la mejor tecnología y buscar capacitación periódicamente ya que es una ciencia que esta a vanguardia de la tecnología y tiene una gran responsabilidad a su cargo.
- ◆ Se recomienda el uso de softwares actualizados que hayan sido probados y entreguen resultados fiables para evitar cualquier posible manipulación de las conclusiones.

Capítulo 9. Bibliografía.

9.1 Libros.

[1] MIYARA, Federico, Acústica y Sistemas de Sonido, Tercera Edición, Editora de la Universidad Nacional de Rosario, 2003.

[2] HARRIS, Cyril, Handbook of Acoustical Measurements and Noise Control, Tercera Edición, Acoustical Society of America (A.S.A.).

[3] FONÉTICA ACÚSTICA DE LA LENGUA ESPAÑOLA. Dr. D. Antonio Quilis Morales. Biblioteca Románica Hispánica. Editorial Gredos, Madrid. 1981.

[4] *Forensic Phonetics*, John Balwin & Peter French, Pinter Publishers, 1990 Londres y *Forensic Speaker Identification*, Philip Rose, Taylor and Francis Forensic Science Series, 2002.

[5] “*Statistics and the Evaluation of Evidence for Forensic Scientists*”, C.G.G. Aitken y F. Taroni, John Wiley & Sons, 2ª edición, 2004.

[6] MANUAL DE MEDIDAS ACÚSTICAS Y CONTROL DE RUIDO, Cyril M. Harris, Editorial McGraw-Hill, 3ª edición, Diciembre 1995.

9.2 Artículos y notas técnicas

[7] REDVET, Vol. VIII, Nº 4, Metodología de trabajo del perito forense veterinario (Methodology of work of veterinary expert forense), Abril 2007.

[8] FORENSICLAB, informe pericial de análisis de voz, Jordi Cicres Bosch, 29 de mayo de 2006.

[9] INGENIERÍA ACÚSTICA APLICADA A LA CRIMINALÍSTICA “ACÚSTICA FORENSE”, Eduardo Andrés Pérez Bello, Universidad Austral de Chile, Valdivia, Chile 2008.

[10] LA INTELIGIBILIDAD DEL HABLA (PARTE 1), Ralph Jones, Meyer SOUND, 2009.

[11] ANÁLISIS DISCRIMINANTE DE UN CONJUNTO DE PARÁMETROS FONÉTICO ACÚSTICOS DE LAS PAUSAS LLENAS PARA IDENTIFICAR HABLANTES, J. Cicres Síntesis Tecnológica. V.3 N° 2 (2007) 87-98.

[12] RECONOCIMIENTO DE LOCUTOR DEPENDIENTE DE TEXTO MEDIANTE ADAPTACIÓN DE MODELOS OCULTOS DE MARKOV FONÉTICOS, Cristina Esteve Elizalde, Universidad autónoma de Madrid, 2007.

[13] MEZCLA Y MASTERIZACIÓN II, secretaría de extensión, facultad de bellas Artes.

[14] FUNDAMENTOS BÁSICOS DEL RECONOCIMIENTO DE VOZ, Cristian Martínez Bernaldo de Quirós.

[15] LA ACÚSTICA FORENSE, JOSÉ JUAN LUCENA MOLINA Comandante de la Guardia Civil. Jefe del Departº de Acústica e Imagen del Servicio de Criminalística de la Guardia Civil, Instituto Universitario De Investigación Sobre Seguridad Interior.

[16] ALGORITMOS Y MÉTODOS PARA EL RECONOCIMIENTO DE VOZ EN ESPAÑOL MEDIANTE SILABAS, José Luis Oropesa Rodríguez, Segui Suarez Guerra, México, 2006.

[17] BATVOX: SISTEMA AUTOMÁTICO DE RECONOCIMIENTO DE LOCUTOR, Beatriz González Sigüenza, EFE, ISSN 1575-5533, XVII, 2008, pp. 301-316.

[18] A PITCH EXTRACTION REFERENCE DATABASE, F. Plante, G.F. Meyer, W.A. Ainsworth Fabrice, Dept of Communication and Neuroscience Keele University.

[19] FUSIÓN DE SISTEMAS DE RECONOCIMIENTO BASADOS EN CARACTERÍSTICAS DE ALTO Y BAJO NIVEL, Mireia Farrús, Jan Anguita, Javier Hernando, Ramón Cerda, Universidad Politécnica de Cataluña, Universidad de Barcelona.

9.3 Artículos de Internet

[20] INFORMES PERICIALES, <http://www.iula.upf.edu/forensiclab>

[21] ACÚSTICA ARQUITECTÓNICA: EL TIEMPO DE REVERBERACIÓN,

Ing. Alejandro Bidondo, inteligibilidad de la palabra,

<http://www.ingenieriadesonido.com>

[22] WIKIPEDIA, Max (programa), [http://es.wikipedia.org/wiki/Max_\(programa\)](http://es.wikipedia.org/wiki/Max_(programa)), 2009, 16/06/2009.

[23] FUNDAMENTOS PARA EL RECONOCIMIENTO DE VOZ Y ANUNCIOS VARIOS PARA LA WEB,

<http://www.adictosaltrabajo.com/tutoriales/tutoriales.php?pagina=complementor>

[24] WIKIPEDIA, Espectro de frecuencias,

http://es.wikipedia.org/wiki/Espectro_de_frecuencias, 2009, 24/04/2009.

[25] GUÍAS METODOLÓGICAS PERICIALES,

<http://www.scribd.com/doc/26009318/PGR-Guias-metodologicas-periciales>

[26] DIRECCIÓN GENERAL DE LA POLICÍA, España, 2004, <http://www.cita.es>

[27] CUERDAS VOCALES http://es.wikipedia.org/wiki/Cuerdas_vocales

Capítulo 10. Anexos.

Anexo 1. Glosario.

Address: Son direcciones dentro de un archivo que sirven para poder acceder a la información de una manera no lineal

AIFF: Conocido por sus siglas en inglés por “Audio Interchange File Format” es un formato de audio estándar usado para almacenar datos de sonido. El formato fue desarrollado por Apple Inc. en 1988 basado en el IFF¹ (Interchange File Format) de Electronic Arts, es internacionalmente utilizada por varios sistemas operativos pero en mayor escala por computadoras Apple Macintosh.

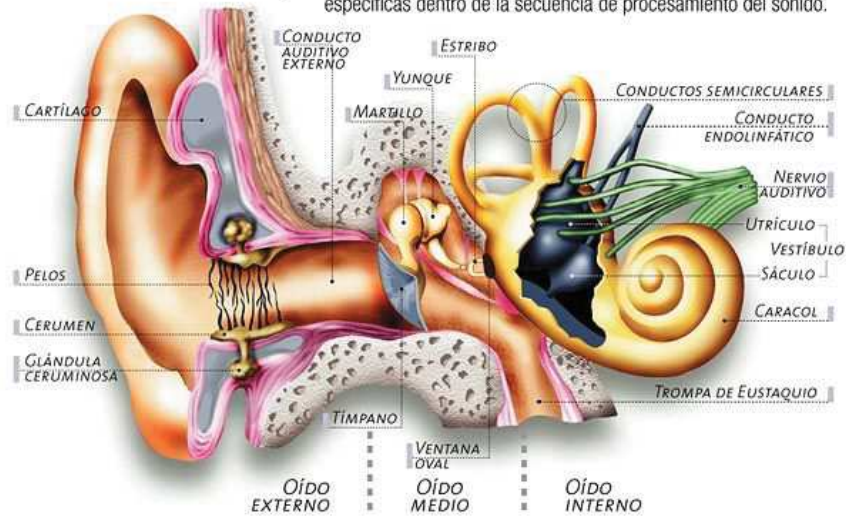
Los datos de audio en el estándar AIFF no están comprimidos y emplea una modulación por impulsos codificados (PCM).

Alimentación Fantasma: Conocida en inglés como “Phantom power” es la forma de proporcionar alimentación eléctrica (corriente continua) a los dispositivos de audio que lo necesiten. El estándar define tres tipos de alimentación (P48, P24 y P9) que suministran respectivamente 48V a través de resistencias de 6.8k ohmios, 24V a través de resistencias de 1.2k ohmios, o 12V a través de resistencias de 680 ohmios y puede suministrar una corriente de 10 a 15 mA.

Aparato auditivo: Estructura del cuerpo humano responsable de la audición y el equilibrio. Se divide en tres zonas: externa, media e interna.

El oído

Una de las funciones principales del oído es la de convertir las ondas sonoras en vibraciones que estimulen las células nerviosas, para ello el oído tiene tres partes claramente identificadas. Estas secciones están interconectadas y son el oído externo, el medio y el interno. Cada parte tiene funciones específicas dentro de la secuencia de procesamiento del sonido.



Fuente:

http://usuarios.discapnet.es/ojo_oido/esquemas_cuerpo_humano/oido.jpg

Audiometría: Es un examen auditivo que cifra las pérdidas auditivas mediante un análisis con equipos especializados y determina la magnitud de la altura tonal y la sonoridad percibida.

BIOS: Por su sigla en inglés de “basic input/output system” es un software básico instalado en la placa base que se ejecuta al encender la computadora, y que localiza y reconoce todos los dispositivos necesarios para cargar el sistema operativo en la memoria RAM.

Bit: Se conoce como bit a la unidad mínima de información utilizado comúnmente en sistemas informáticos, corresponde a un digital del sistema binario que puede ser 0 o 1.

Bit rate: Se lo conoce como a la velocidad de tasa de transferencia de bits, y está determinado como el número del flujo de bits por unidad de tiempo.

Buffers: Es una memoria que se utiliza para almacenar datos, con el objetivo de tener un respaldo temporal de información, su propósito es de dar una continuidad al procesamiento de datos.

Campo electromagnético: Es un campo en el plano físico sus características son del tipo eléctrico y magnético, y al interactuar con conductores afecta a sus partículas con una carga eléctrica.

Centroide: El centroide o baricentro de un objeto X que pertenece a un espacio de n dimensiones puede obtenerse mediante el promedio de todos los puntos de X .

Coefficientes cepstrales: Son coeficientes para la representación del habla basados en la percepción auditiva humana. Se derivan de la Transformada de Fourier (FT) o de la Transformada de coseno discreta (DCT).

Conexión a tierra: La conexión a tierra es un camino con poca resistencia a cualquier corriente, consiste en una pieza metálica, enterrada en una mezcla especial de sales y conectada a la instalación eléctrica a través de un cable. su principal función es la de descargar la tensión y corriente de redes eléctricas.

Controladores: En inglés conocidos como "Drivers" son aplicaciones encargadas de la coordinación entre el sistema operativo y los distintos dispositivos periféricos, además son complementarias del sistema operativo computacional y permiten expandir su manejo y procesamiento.

Enventanado: el enventanado es una función matemática utilizada en el análisis y el procesamiento de señales su usa para prevenir las discontinuidades que se crean al principio y al final de un bloque de análisis.

Firewire: Es un protocolo de comunicación muy masificado como el USB, el puerto firewire también se lo llama IEEE 1394 existen dos versiones:

Tabla 10.1: Velocidades de comunicación de las conexiones Firewire.

VERSIÓN	VELOCIDAD
400	50 MB/s
800	100 MB/s

Frecuencia de corte: Es aquella frecuencia en la cual la respuesta de nivel a decaído 3 dB con respecto a la ganancia unitaria (ni atenuación ni amplificación).

Jitter: En audio se conoce como jitter a la distorsión de fase percibida como ruido, debida a la falta de precisión en la regularidad del periodo de muestreo en la grabación o reproducción de audio digital. Fisiológicamente es la anomalía en la duración de los pulsos glóticos, lo que produce fluctuaciones en la frecuencia fundamental (F0).

LED: Se trata de un diodo emisor de luz, cuando se polariza de forma directa la unión de semiconductores PN, este dispositivo emite luz de espectro reducido.

MIDI: Sus siglas quieren decir “Interfaz digital de instrumentos musicales”. Consiste en un protocolo de comunicación serial estándar entre los computadores, sintetizadores, secuenciadores, controladores y otros dispositivos musicales electrónicos.

Normalización: Normalización audio es el proceso del aumento o de disminución de la amplitud de audio señal. La normalización aumenta la amplitud de la forma de onda al nivel máximo nivel posible sin introducir distorsión armónica.

PCI: Por sus siglas en inglés "Peripheral component interconnect" o Interconexión de Componentes Periféricos, es un bus de ordenador estándar cuya función es la de conectar dispositivos periféricos directamente a su placa madre.

Algunos de estos dispositivos pueden ser circuitos integrados o tarjetas de expansión que se ajustan en conectores.

PCI Express: PCI Express es un desarrollo del bus PCI, se basa en un sistema de comunicación serial mucho más rápido que su predecesor. Este sistema fue desarrollado principalmente por la empresa Intel.

PCMCIA: Por sus siglas en inglés "Personal Computer Memory Card International Association", es una asociación Internacional encargada del desarrollo un formato de tarjetas de memoria para ordenadores personales, que permiten una expansión de sus funciones. Este formato de tarjeta PCMCIA es usado frecuentemente en módems, tarjeta de sonido y tarjeta de red.

Polarización: Es la suma de una corriente eléctrica para la etapa de grabación con el fin de obtener linealidad en la representación eléctrica-magnética en la cinta.

Peak: Respuesta o medición de tipo instantánea, considerada así por una ventana de análisis de tiempo muy corta.

PRAAT: Es una herramienta computacional para el análisis fonético del habla desarrollada por Paul Boersma y David Weenink en el "Instituto de Ciencias Fonéticas de la Universidad de Ámsterdam".

RMS: Siglas de Root Mean Square. Respuesta o medición de tipo promedio, considerada así por una ventana de tiempo de mayor duración que en la respuesta Peak.

Shimmer: Fisiológicamente son los cambios en la intensidad de los pulsos de la glotis, los que producen una perturbación de la amplitud de la frecuencia fundamental.

Sintetizador: Un sintetizador es conocido como un instrumento musical electrónico diseñado para generar sonido de una manera artificial, entre sus técnicas de síntesis se tiene a la síntesis aditiva, substractiva, de modulación de frecuencia, de modelado físico o modulación de fase, tablas de ondas y mixtas de esta manera se crean distintos sonidos.

Sistema operativo: Es un software cuya función es la de comunicación digital entre los dispositivos de hardware , los programas computacionales y el usuario.

Entre los sistemas operativos más populares del mercado se tiene a Windows, Macintosh y a Linux.

Transducción: Se le conoce como transducción a la transformación de un tipo de señal o energía en otra de distinta naturaleza mediante dispositivos llamados Transductores. Entre los transductores más conocidos se tiene al transductor electroacústico, transductor electromagnético, transductor electromecánico, transductor electroquímico, transductor electrostático, transductor fotoeléctrico, transductor piezoeléctrico y transductor electroresisivo.

USB: El protocolo de comunicación USB o Universal Serial Bus es un puerto universal de interconexión de dispositivos actualmente disponible en 3 versiones la 1.0, 1.1, 2.0 y 3.0 cuyas diferencias son su velocidad de transmisión puesto que utilizan el mismo tipo de conector. La gran ventaja de estas interfaces es que al apartarse de la CPU, se evita el ruido por inducción electromagnética.

Tabla 9.2: Velocidades de comunicación de las conexiones USB.

VERSIÓN	VELOCIDAD
1.0	192 KB/s
1.1	1,5 MB/s

2.0	60 MB/s
3.0	600 MB/s

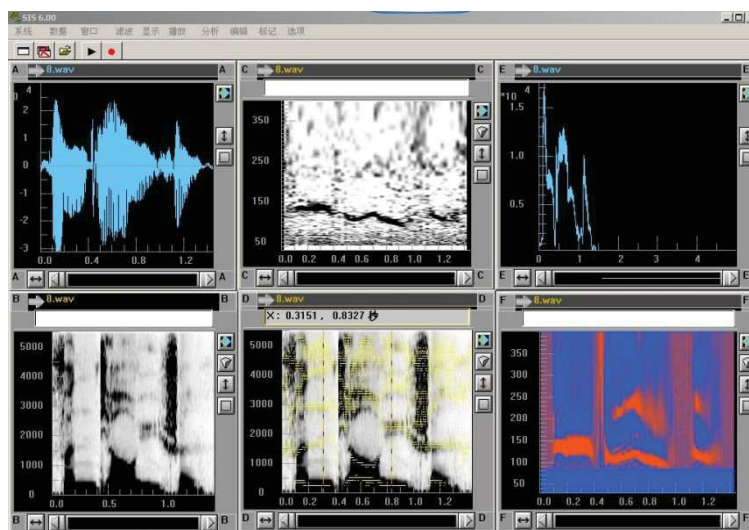
WAV: El formato WAV o WAVE conocido por sus siglas en ingles “*Waveform audio file format*”, es un formato de audio digital codificado sin compresión de datos más masificado a nivel mundial, fue desarrollado en conjunto por las empresas Microsoft e IBM, su extensión es .wav.

Anexo 2. IKAR LAB.

Es un software de reconocimiento de voz el programa S.I.S. versión 6 de la compañía "Speech Technology Center" de procedencia rusa, que es actualmente usado por la Policía Nacional de Ecuador.



Fuente: http://www.brokkesystem.eu/audio_analys.html?num=5



Fuente: <http://www.rebeck.cn/stc.files/lkar%20lab.jpg>

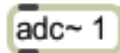
La adquisición del software S.I.S. se la puede realizar con la compra del equipo completo denominado "IKAR Lab" que incluye tanto hardware como varios software de análisis y procesamiento de señales acústicas. Se contactó directamente con la empresa fabricante el 31 de Agosto del 2010, donde se contactó el Sr. Miguel Valarezo Popov, donde por vía telefónica comunicó que este producto se encuentra valorado en 25,000.00 dólares americanos, haciendo de este un sistema costoso para una inversión inicial de un perito civil.

Anexo 3. Objetos utilizados de Max MSP.

Etapa de la entrada de audio:

adc~: Es un objeto que permite crear una entrada de audio en un patcher de Max MSP desde una entrada física de la tarjeta de sonido del ordenador para un posterior análisis o procesamiento.

Dentro de Max MSP este objeto puede tener entre una y 512 entradas dependiendo de las características de la tarjeta de sonido, el orden interno de las entradas puede ser modificado desde la ventana DSP status que se puede encontrar en el menú Opciones. El objeto tiene una entrada para su encendido o apagado por medio de mensajes 1 y 0 respectivamente, además posee un determinado número de salidas correspondientes al número de canales de entrada de la tarjeta de audio.

The image shows a rectangular object in a Max/MSP patcher window. The object is labeled 'adc~ 1' in a black font. The object has a light green border and a small black square on the left side, indicating it is selected.

sfplay~: Este objeto permite reproducir un archivo de audio desde un disco los formatos que se puede reproducir son aiff, au, wav, de uno a 32 canales.

El número de salidas del objeto dependerá del número que se haya escrito en el argumento, además crea una salida adicional que envía un mensaje bang al terminar de reproducir un archivo de audio.

Los mensajes más importantes para controlar a este objeto son:

- ◆ Open: que abre una ventana para buscar el archivo a reproducir.
- ◆ 1: Inicia la reproducción del archivo cargado.
- ◆ 0: Detiene la reproducción del archivo cargado.
- ◆ loop 1: Enciende la reproducción en bucle.

The image shows a rectangular object in a Max/MSP patcher window. The object is labeled 'sfplay~' in a black font. The object has a light green border and a small black square on the left side, indicating it is selected.

selector~: Es un objeto tiene como función seleccionar desde una o varias

entras a una sola salida.

Este objeto posee una entrada de control y tantas entradas como se haya escrito en el argumento, además posee una sola salida para la señal seleccionada.

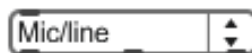
Entre los mensajes de control más importantes están:

- ◆ 0: Apaga el objeto y no permite pasar ninguna señal hacia su salida.
- ◆ 1, 2, 3, etc.: Este mensaje envía selectivamente cualquiera de las señales conectadas a las entradas hacia la salida.



umenu: Es un cuadro menú que muestra varias opciones establecidas por el usuario, a cada opción ingresada en el objeto se puede asociar un mensaje numérico que se muestra en la salida del objeto. El método más fácil de agregar nuevas opciones es mediante el inspector.

Este objeto posee dos salidas útiles la primera es el mensaje numérico asociado a la opción seleccionada, y la segunda es la opción seleccionada como un mensaje de textual.



toggle: Es un tipo de botón que puede funcionar como un conmutador de encendido y apagado, el objeto puede ser controlado por mensajes o directamente con un clic sobre él.

Los mensajes de control de este objeto son:

- ◆ 0: apaga el objeto obteniendo en su salida el mensaje numérico 0
- ◆ bang: con cada mensaje bang el objeto se enciende o se apaga.
- ◆ 1, 2, 3, etc: con cualquier número entero positivo y negativo distinto al 0 el objeto se enciende.

En su salida se obtiene el mensaje 1 cuando esta encendido y 0 cuando esta apagado.



comment: Este objeto es una etiqueta donde se puede escribir algún texto, resulta útil para escribir explicaciones o nombres.



loadbang: Es un objeto que envía automáticamente un mensaje bang cuando se abre una ventana de un patcher o cuando se da doble clic sobre el objeto.

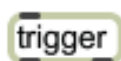
Resulta útil para encender automáticamente ciertos objetos, utilizando sus dos salidas por donde se envía el mensaje bang.



trigger: Este objeto envía los mensajes escritos en el argumento hacia sus salidas cuando es activado por un bang. El objeto puede ser abreviado por la letra t, posee una sola entrada y tantas salidas como argumentos se escriba en el objeto.

Los argumentos y tipos de datos que puede manejar este objeto son:

- ◆ l para lista
- ◆ b para bang
- ◆ int para número entero
- ◆ symbol para símbolos
- ◆ float para números decimales.



deferlow: Es un objeto que reduce la prioridad con que se envía un mensaje.

Este objeto es útil para que un mensaje responda con cierto retardo con relación a otro de mayor prioridad.

Posee una entrada y una salida por donde ingresa el mensaje y sale el mensaje.



gain~: Es un objeto de audio similar a un potenciómetro deslizable, posee una escala logarítmica y puede ser usado como control de nivel de una señal de entrada.

Posee dos entradas: una para ingresar la señal de audio y manejar el control del nivel y una segunda entrada para establecer el tiempo de rampa en milisegundos, y posee dos salidas una para la señal de audio escalda y otra que representa el valor numérico del deslizador.

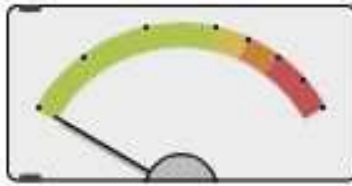
El nivel puede ser configurado moviendo manualmente el deslizador virtual o por medio de un mensaje de tipo numérico que puede ir desde el cero equivalente a silencio, 128 como valor nominal de 0 dB y un máximo de 157 equivalente a +18 dB.



levelmeter~: Es un objeto similar a un medidor de tipo *VU meter*.

Se utiliza para indicar visualmente el nivel RMS de una señal.

Es posible configurar todos los rangos de visualización por medio de un mensaje, así como los tiempos ataque y relajación en milisegundos.



LED: Es un interruptor luminoso que cuando está encendido envía un mensaje de 1 u cuando está apagado envía el mensaje 0.



dac~: Es un objeto que permite crear una o varias salidas de audio en un patcher de Max MSP hacia una salida física de la tarjeta de sonido del ordenador para señales que pueden haber sido analizadas, procesadas o sintetizadas.

Este objeto puede tener entre una y 512 salidas dependiendo de las características de la tarjeta de sonido, el orden de las salidas puede ser modificado desde la ventana DSP status.

El objeto posee una entrada para su encendido o apagado por medio de mensajes numéricos 1 o 0 y para la salida del canal número uno, también posee tantas entradas como argumentos y salidas tenga la tarjeta de audio.



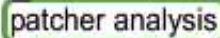
Etapas de análisis de la frecuencia fundamental:

patcher: Este objeto crea un sub patcher dentro de otro patcher, puede tener un determinado número de entradas y salidas .

Esta herramienta permite crear módulos específicos de procesamiento o análisis que pueden ser visualizados en varias ventanas de patcher, evitando la fatiga visual que producen una gran cantidad de objetos y conexiones en una sola pantalla.

Para el funcionamiento de este objeto es necesario utilizar los objetos inlet y

outlet que crean las entradas y salidas del sub patcher.



inlet: Es un objeto usado para recibir mensajes y señales desde un patcher con un nivel superior. Posee una sola salida correspondiente a los mensajes y señales recibidos.



gate~: Es un objeto que envía una señal de entrada hacia una de varias salidas. Posee dos entradas: una entrada de control y una entrada para la señal de audio, y tantas salidas como se haya especificado en el argumento.

Entre los mensajes de control más importantes están:

- ♦ 0: Apaga el objeto y no permite pasar ninguna señal hacia cualquiera de las salidas.
- ♦ 1, 2, 3, etc.: Este mensaje envía selectivamente la señal de entrada conectadas hacia la salida especificada por el mensaje numérico.



analyzer~: Es un objeto externo para Max/MSP creado por Tristan Jehan, está basado en el objeto fiddle~ creado por Miller Puckette. Es un objeto diseñado para analizar una señal de audio por métodos preceptuales basados en la transformada rápida de Fourier (F.F.T.) de hasta 4096 bandas.

Argumentos:

- ♦ Tamaño de buffer (Número de muestras).
- ♦ Tamaño del salto entre ventanas. (Número de muestras).
- ♦ Tamaño de muestras por F.F.T. (número de muestras).

- ◆ Tipo de ventana: Rectangular (blackman62), hamming (blackman72), hanning (blackman74), hamming 2 (blackman92).
- ◆ Retraso inicial.
- ◆ Número de pitches para estimar (número entero).
- ◆ Número de frecuencias senoidales para buscar (número entero).
- ◆ Número de salidas para las frecuencias senoidales.

Este objeto posee una sola entrada para la señal de audio y para establecer ciertos parámetros de funcionamiento por medio de mensajes, en sus salidas se puede encontrar:

- ◆ Pitch: en esta salida se obtiene la estimación de la frecuencia fundamental en Hz y en codificación MIDI presentados como una lista.
- ◆ Loudness: Entrega una estimación del nivel de la señal en dB.
- ◆ Brightness: Entrega la claridad de la señal.
- ◆ Noisiness: Entrega información de cuán ruidosa es la señal.
- ◆ Attack: Esta salida entrega un mensaje bang cuando detecta un pico.
- ◆ Pitch polifónicos: Esta salida entrega la frecuencia y la amplitud de pitch polifónicos estimados por el objeto en forma de una lista, se puede obtener un máximo de 3 pitch.
- ◆ Componentes senoidales: Esta salida entrega la frecuencia y la amplitud de pitch de los componentes senoidales estimados por el objeto en forma de una lista. Se puede obtener un máximo de 100 componentes senoidales dependiendo del argumento utilizado.

```
analyzer~ 2048 512 4096 blackman74 0 1 3 3 list
```

lp.stacey: Es un objeto externo de la empresa Litter Power su función es calcular algunos valores estadísticos de un conjunto de valores ingresados en

tiempo real. Posee una entrada para mensajes de tipo numérico con o sin decimal y seis salidas correspondientes a los cálculos de: número de valores ingresados, valor mínimo, valor máximo, valor promedio, desviación estándar, skew y kurtosis.

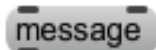
Se puede enviar un mensaje clearbang a la entrada para reiniciar todos los contadores a cero.

A screenshot of a text input field with a light green border and a dashed line at the bottom. The text 'lp.stacey' is entered in the field.

message: Es un objeto en forma de botón que puede enviar datos de tipo de mensaje, número, número con decimal, o lista. Este objeto tiene la habilidad de cambiar argumentos específicos en ciertos objetos.

En el objeto se puede encontrar dos entradas una para enviar el mensaje a la salida después de recibir un bang y otra que carga información dentro del objeto.

El mensaje dentro del objeto también puede ser enviado dando un clic sobre si mismo.

A screenshot of a button with a grey background and rounded corners. The text 'message' is centered on the button.

unpack: Este es un objeto que separa los elementos de una lista y los envía a salidas independientes.

El número de argumentos y su formato determina el número de salidas, los tipos de datos que puede manejar son número, número con decimal y símbolos.

A screenshot of a button with a light green border and rounded corners. The text 'unpack 0. 0.' is centered on the button.

number : Es un cuadro con el que se puede visualizar datos de tipo numérico enviarlos por su primera salida.

A screenshot of a button with a light green border and rounded corners. The text '0' is centered on the button.

send: Es un objeto que puede enviar mensajes sin la necesidad de una conexión directa por cables virtuales. Es un objeto útil para enviar mensajes a sub patchers sin la necesidad del uso de inlets. Puede ser abreviado con la letra r.

En el argumento se debe especificar un nombre que será común entre el objeto send y receive.

send argumento

receive: Es un objeto que puede recibir mensajes sin la necesidad de una conexión directa por cables virtuales, es un objeto complementario al send.

receive argumento

outlet: Es un objeto usado para enviar mensajes y señales hacia un patcher con un nivel superior.

Este objeto posee una entrada correspondiente a la salida de los mensajes y señales.

Etapas de almacenamiento de información:

coll: Es un objeto que permite guardar, organizar y editar en una lista un conjunto de mensajes. En la entrada de este objeto se puede ingresar listas de datos anteponiendo un número único para posteriores tareas de identificación.

Las listas de mensajes guardados en el objeto pueden ser recuperados en la primera salida enviando el número de identificación de la lista hacia la entrada.

Una función útil del objeto es la conocer el número de elementos que se obtiene enviando el mensaje length a la entrada.

coll

dialog: Es una ventana de diálogo que se utiliza para escribir texto dentro de Max MSP. Este objeto posee dos entradas: en la primera entrada activa la

ventana de diálogo al recibir el mensaje bang, la segunda define el formato que tendrá el texto escrito. En su primera salida envía el mensaje escrito en la ventana de diálogo y en la segunda envía un bang cuando el usuario cancela la ventana de diálogo.

`dialog Nombre`

pack: Este es un objeto que combina varios números y símbolos en una sola lista. El número de argumentos y su formato determina el número de entradas, los tipos de datos que puede manejar son número, número con decimal y símbolos.

`pack 0. 0. 0. 0. 0. 0.`

uzi: Este objeto envía un número específico de mensajes bang a su salida. Posee una entrada que acciona el inicio de los bang y una segunda entrada que establece el número de bangs que se obtendrá en la salida. Posee tres salidas: la primera entrega los bang consecutivos, la segunda entrega un bang cuando todos los bang haya sido detonados y la tercera entrega una serie de números enteros en lugar de bangs.

`uzi`

button: Es un objeto que destella cuando recibe un mensaje o un número y en su salida se obtiene un mensaje bang. Se utiliza para detonar otros mensajes o para iniciar procesos en determinados objetos.



+: Es un objeto que suma dos números y entrega el resultado en su salida. Se puede usar un argumento para sumar un número constante con los números en la primera entrada.

`+ 1`

metro: Este objeto envía un mensaje bang en un intervalo de tiempo regular establecido en milisegundos.

Posee una entrada para encender o apagar el objeto y una salida por donde se obtienen los mensajes bang.

Etapa de la comparación de similitud:

zsa.dist: Es un objeto externo que forma parte del paquete de descriptores zsa de la empresa Ircam.

Este objeto compara una lista de números en la primera entrada con un conjunto de listas etiquetadas en la segunda entrada y en su salida identifica la etiqueta de la lista más parecida, para encontrar la similitud utiliza medidas de distancia vectorial.



= =: Este objeto compara dos números en sus entradas y presenta 1 en su salida si estos son iguales caso contrario 0.

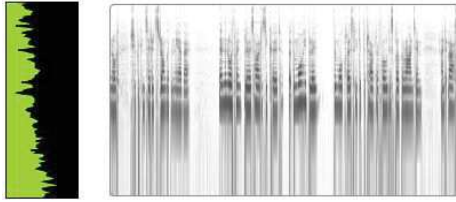


5 Interfaz de usuario:

panel: Este objeto permite crear una superficie rectangulares de fondo dentro de un patcher para crear interfaces de usuario. Es posible configurar el color de fondo esquinas redondeadas y efectos de sombra.



spectroscope~: Es un objeto que permite analizar gráficamente una señal puede ser configurado para funcionar como espectrograma o como sonograma. Este utiliza la transformada rápida de Fourier (FFT) para su análisis.



minimum: Es un objeto que entrega en su salida el menor numero de una serie de valores que ingresan en su entrada.

```
minimum 0.
```

maximum: Es un objeto que entrega en su salida el numero de mayor valor de una serie de valores que ingresan en su entrada.

```
maximum 0.
```

prepend: Es un objeto que agrega un mensaje al principio de un mensaje o una lista de mensajes. El argumento es el mensaje que se desee agregar al principio.

```
prepend setminmax
```

cycle: Este objeto separa una serie de de datos hacia varias salidas independientes. Se pueden utilizar un número argumento que define el número de salidas.

```
cycle 10
```

zmap: Es un objeto que adapta rangos de valores de su entrada hacia rangos mas grande o mas pequeños. El objeto posee cuatro entradas correspondientes a los siguientes argumentos:

- ◆ Argumento 1: Define el valor mínimo de entrada.
- ◆ Argumento 2: Define el valor máximo de entrada.
- ◆ Argumento 3: Define el valor mínimo de salida.
- ◆ Argumento 4: Define el valor máximo de salida.

zmap 0 1 0 1

Exportación de cómo una aplicación independiente:

standalone: Este objeto permite configurar una serie parámetros generales del programa por medio del inspector entre los más importantes están: Soporte de audio, ventanas activas, subcarpetas personales, nombre del programa, etc. Estos parámetros deben ser configurados previa la exportación como de aplicación independiente.

standalone

Anexo 4. Metodología utilizada en la creación de la base de datos Keele Pitch.

A PITCH EXTRACTION REFERENCE DATABASE

F. Plante(+), GF Meyer(*), W.A. Ainsworth(+,*)

fabrice@cs.keele.ac.uk, georg@cs.keele.ac.uk, w.a.ainsworth@keele.ac.uk
(+) Dept of Communication and Neuroscience, (*) Dept of Computer Science
Keele University, Keele, Staffs ST5 5BG, UK

ABSTRACT

Many pitch extraction algorithms have been proposed in the past. The comparison of these algorithms is difficult because each study tends to be carried out on a unique data set. The purpose of this project is to develop a database for the comparison of these algorithms. This database is based on a core speech module and several additional modules. The core module contains speech and laryngograph data for 15 speakers reading a phonetically balanced text. A voiced/unvoiced reference file is provided with the speech data. Currently a psychophysics module is available to test the performance of pitch extraction stages on commonly used pitch perception stimuli. The database is intended to be open: contributions and remarks can be send to georg@cs.keele.ac.uk.

1. INTRODUCTION

Pitch extraction as a problem is as old as speech processing. Over the past 30 years a number of fundamentally different approaches have been proposed to extract the pitch of speech. All areas of speech research (analysis, recognition, coding, synthesis, perception, pathology) need more or less information about pitch such as voicing decision, fundamental frequency estimation, glottal closure instant detection or glottal wave reconstruction.

The complexity of the task is expressed in the number of methods and algorithms currently available. It is obvious that "none of them work perfectly for every voice, application and environmental condition" (Hess, 1983 [1]). Nevertheless it is important to evaluate the different algorithms for specific applications. Currently, such an evaluation is difficult because evaluations are typically restricted to limited sets of algorithms on proprietary data sets [2, 3, 4, 5].

One of the reasons for this is that research on pitch

extraction tends to be 'algorithm-driven' rather than data driven. Details of the algorithms are made public so that they can be re-implemented but little data is shared.

Recent developments in speech recognition research shows how much benefit a common data set can have (DARPA evaluations). The evaluations also show that an approach based on common data has its problems, for instance algorithms can be optimised to perform well on a very restrictive task.

2. AIMS

We propose to make available a common database of speech and psychophysical stimuli. This would allow a systematic evaluation of the strengths and weaknesses of the wide range of pitch extraction algorithms.

Pitch extraction is applied to problems ranging from perceptual models to synthesis applications. It is clear that the requirements are problem specific [1].

A number of core requirements were identified the database should be open, easily obtainable and practical.

Openness

The most important requirement of the database is that it is open. The aim is to provide a small general purpose core database to which more modules, for instance databases used in previous evaluations, or more problem specific data sets can be added (pathological speech, telephone speech, noise, etc). The range of voices should be large enough to allow useful performance measures.

Availability

The database is stored on an ftp server <ftp://cs.keele.ac.uk> and available by anonymous ftp from <pub/pitch>.

Users are able to choose which of the modules will be used for evaluation.

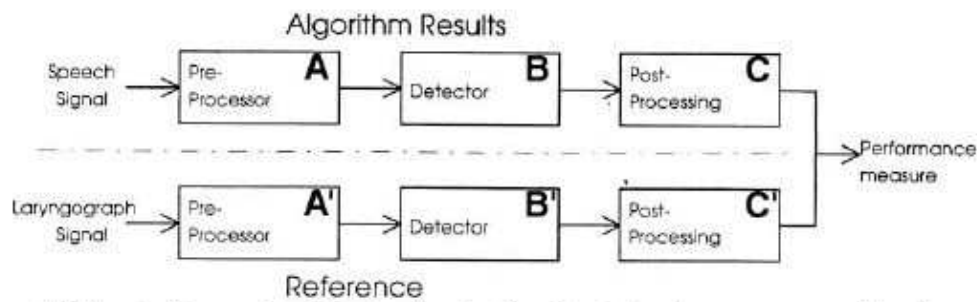


Figure 1: Schematic diagram of a pitch extraction algorithm. Speech signals are pre-processed to enhance FO related information (A), this information is fed into a detector module (B) to obtain F0 values. Many systems also contain a post-processing stage (C) to smooth out eventual wrong estimates. Reference figures, obtained from a laryngograph trace, also depend on the preprocessor used (A'). Ideally the reference and the pitch estimate should be obtained with the same technique.

Practicality

The database is intended to be used for evaluation rather than optimisation. This means that the size of the core data set and the different modules can (and should) be kept relatively small. New modules will be developed according to user requirements.

3. SPEECH CORE MODULE

3.1. Signals

The core data consists of a phonetically balanced text, 'The North Wind Story', read by 15 native English speakers: 5 mature males, 5 mature females and 5 children (3 males, 2 females). The age and duration of the text are summarised in table 1.

Group	Age	duration (s)
Males	21-60	27-40
Females	20-37	30-38
Children	8-12	30-50

Table 1: Age and duration range for each group of speakers.

The adult readers were recorded in a soundproof room, the children were recorded in a quiet office environment to minimise stress.

The speech data was recorded simultaneously with a laryngograph [6] using a head mounted microphone and a DAT recorder. Both traces were digitised at 20KHz with 16 bit resolution.

The raw data corresponding to the signals were saved in the SAM format: new file extensions are *.pes for the speech signal and *.pel for the laryngograph signal.

A coarsely labelled file (*.pet) containing the beginning and end of each phrase with its orthographic transcription is provided.

3.2. Reference data

Reference files based on the laryngograph signal are provided as the primary reference. It is clear that a laryngograph reference only has application in a subset of algorithms: those concerned with fundamental frequency (FO) estimation and glottal closure extraction.

There are two main problems with the reference:

- 1) The term 'pitch' implies a perceptual quality, which is known to diverge from the measured F0 for certain stimuli.
- 2) Whichever algorithm is used to extract an objective pitch estimate from the laryngograph signal will compromise the relative performance of possible pitch extraction algorithms, fig 2. This is due to, for instance, different window sized used in the algorithms and the very different fundamental assumptions made about the signal processed.

For speech data the laryngograph trace, nevertheless is the only feasible hard reference. Users are encouraged to use the laryngograph data to build their own reference data sets.

For convenience a set of more abstract reference files are supplied. These reference files (*.pev) contain a voiced/unvoiced decision and pitch estimates for each 10 ms block of speech in the database. This reference is computed from the laryngograph trace using a floating autocorrelation of 25.6ms duration.

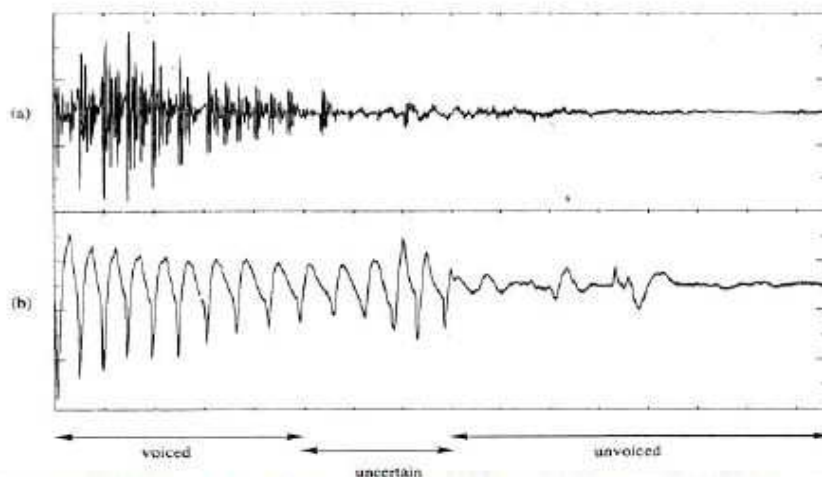


Figure 2: Example of asynchrony between the speech signal (a) and the laryngograph signal (b).

To allow a comparison of algorithms based on different models (production/perception) it is necessary initially to obtain a consistent voicing decision for both speech and laryngograph signals. In theory this should be trivial, in practice there are serious problems. Real recordings contain both sections where the speech waveform looks and sounds voiced while the laryngograph picked up no activity. At the same time laryngeal activity does not always cause a speech waveform. It was decided that segments where no consistent and obvious decision could be made by visual inspection (periodicity in one signal and not in another) are labelled 'uncertain' using a negative flag. Figure 1 shows an example of these uncertain frames. In this case the periodicity recorded by the laryngograph does not reflect a periodicity on the signal.

For each frame, the reference value is one of the following:

value a pitch estimate for voiced sections
 0 for unvoiced sections
 -value uncertain voiced section (lx data)
 -1 uncertain voiced section (voice data)

When periodicity is observed in the laryngograph, but no clear periodic activity can be seen in the speech trace, the corresponding frame is labelled with the negative of the pitch estimate obtained from the laryngograph.

When periodicity is observed in the speech data, but

not in the laryngograph the frame is set to -1 to indicate that it may well be perceived as voiced but that no independent reference exists.

Most of the frames occur at voicing onsets and around plosive bursts. Uncertain frames represent between 2.5 and 5% of the number of frames, depending of the subject.

The philosophy underlying this decision is that, whichever decisions are taken for a uncertain frame will influence the performance figures for one subset of extraction algorithms, but not others.

If only one source (speech or laryngograph) exists objective evaluation is not possible. The data, consequently, should not be included in comparisons between algorithms.

The data is nevertheless supplied as far as possible to allow maximum flexibility in the algorithm evaluation.

While experimenters using the database are encouraged to use the laryngograph trace to calculate a reference matching the processing performed on the speech data, the labelled uncertain frames should be ignored as far as possible in performance comparisons.

A comparison based on this data is likely to produce good results because many of the critical frames are removed. It is nevertheless far from clear what, for instance, the perceived pitch of the questionable frames would be.

4. ADDITIONAL MODULES

4.1. Psychophysics modules

Pitch is defined as a perceptual quality. It could therefore be argued that a production based reference, such as laryngograph traces, is not appropriate and indeed is likely to bias any performance comparisons towards production based algorithms, such as inverse filtering.

Currently no pitch perception data is available for continuous speech. A number of psychophysical stimuli, however, have been proposed and extensively used in the past to explore human pitch perception mechanisms. These are typically very short signals with one or more perceived pitches. The psychophysics module contains the following signals

- 1) Pure tones
- 2) Missing fundamental stimulus,
- 3) Pitch shift sequence,
- 4) Musical chords,
- 5) Ambiguous pitch,
- 6) Amplitude modulated noise
- 7) Comb filtered noise

The stimuli and models of pitch perception based on them are explained in detail in Meddis and Hewitt [8] or Meyer and Dewar [9].

4.2. Digital Signal Processing Toolbox

A number of DSP tools are available on the same ftp site under pub/DSP. This toolbox contains classical DSP routines (fft, filtering, data generation, etc) and also some pitch extractor algorithms developed at Keele [7]. The reference was computed using this toolbox. The package runs under UNIX and can be used directly as a set of command line commands or can be used with a very simple Xli graphical user interface. The system has been implemented on SUN workstations and tested on a range of other workstations (Silicon Graphics, HP, DEC, LINUX).

5. CONCLUSION

The database proposed here is the first step towards a public database to aid evaluation of pitch extraction algorithms. The database is open and external contributions and remarks are welcome. To keep the size of the database to practical limits, contributions should either cover areas of interest not currently included or have been used previously in algorithm evaluations.

Regular updates will be made. Researchers interested in this database are encouraged to contact the authors by email: georg@cs.keele.ac.uk.

6. ACKNOWLEDGEMENTS

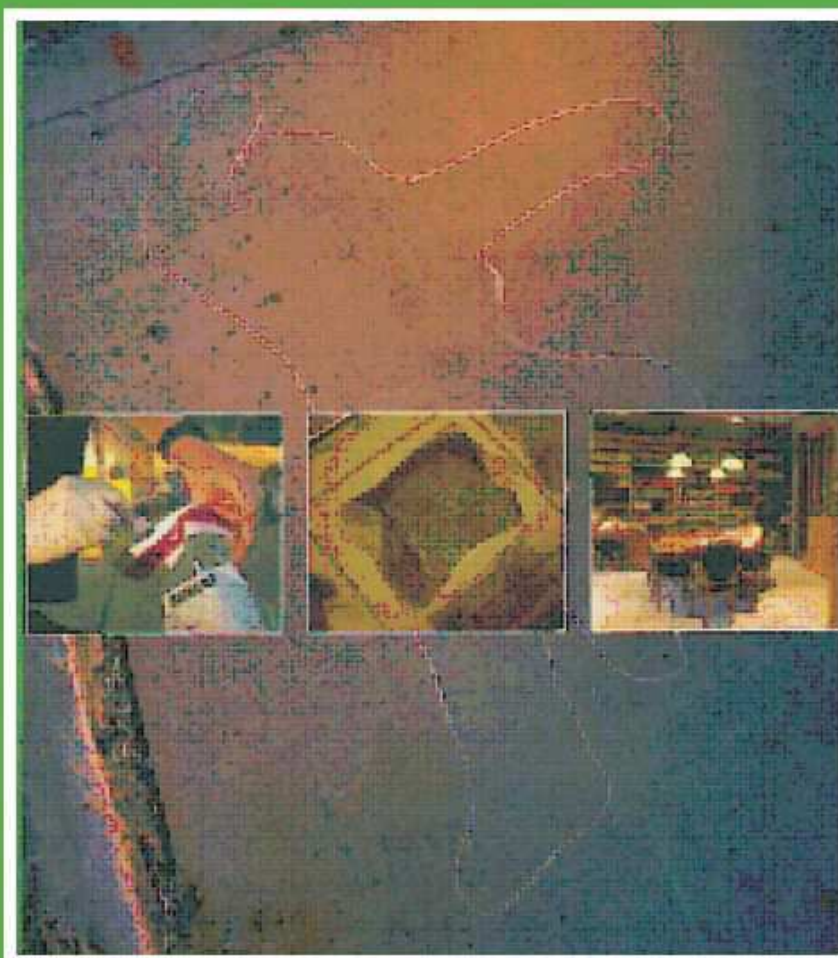
The work was supported in part by contract SCI-CT92-0786 of the EC Science Programme.

7. REFERENCES

- [1] W. Hess "Pitch determination of speech signals", Springer-Verlag, Berlin, 1983.
- [2] P.C. Bagshaw, S.M. Hiller, M.A. Jack, "Enhanced pitch tracking and the processing of F0 contours for computer aided intonation teaching." Proc Eurospeech93, pp. 1003-1006, Berlin 1993.
- [3] D.J. Hermes "Pitch analysis" in Visual representations of Speech, Cooke, Beet and Crawford Eds, pp. 1-25, 1993.
- [4] L.M. Rabiner, M.J. Cheng, A.E. Rosenberg, C.A. McGonegal, "A comparative performance study of several pitch detection algorithms". IEEE ASSP, Vol.24, pp. 399-418, 1976.
- [5] L. Van Immerseel, J.P. Martens, "Pitch and voiced/unvoiced determination with an auditory model". JASA, Vol.91, pp. 3511-3526, 1992.
- [6] A.J. Fourcin, E. Abberton, "First application of a new laryngograph". Medical and biological Illustration, Vol.21, pp. 172-182, 1971.
- [7] F. Plante, G. Meyer, W.A. Ainsworth, "Pitch detection: Auditory model versus inverse filtering" Procs IOA, Vol.16, pp. 81-88, Windermere 1994.
- [8] R. Meddis and M.J. Hewitt "Virtual Pitch and Phase sensitivity of a Computer Model of the Auditory Periphery. 1: Pitch Identification" J Acoust Soc AM. 89 pp. 2866-2882, 1991.
- [9] G.F. Meyer and I.D. Dewar, "Comparing Pitch Extraction in the Cochlear Nerve and Cochlear Nucleus". Proc IOA, 16(5), pp. 263-271, 1994.

Anexo 5. Guías metodológicas de las especialidades periciales (sección análisis de voz).

Guías Metodológicas de las Especialidades Periciales



Procuraduría General de la República

México 2003

DEFINICIÓN

Desde el punto de vista criminalístico es el estudio de las características fonéticas y espectrales sobre las cuales se busca relacionar los elementos semejantes para identificar o descartar su origen.

OBJETIVO

Lograr la identificación de una persona mediante el análisis comparativo de dos o más emisiones de voz.

TIPOS DE INTERVENCIÓN

I.- EN CASOS DE SECUESTRO Y ESTUDIO DE VOCES EN DIVERSOS SISTEMAS DE AUDIO Y VIDEO

En el Laboratorio de Análisis de Voz se realizan estudios en hechos en los cuales intervienen una o varias personas, examinando la grabación de voz en audiotape y/o videotape, de agresores, víctimas o personas que se encuentren involucradas en presuntos hechos delictuosos.

Requerimientos

- La petición debe ser clara sobre el motivo de la intervención del perito, especificándose cuál o cuáles son las voces que se desean analizar.

- Remitir los audiocasetes o videocasetes objeto de estudio al laboratorio.
- En caso de existir probables responsables es necesario contar con los audiocasetes o videocasetes problema, antes de tomar las muestras de voz, para conocer el contenido de la conversación, a fin de que la muestra testigo cuente con elementos similares a la cuestionada.
- Las muestras de voz deberán ser tomadas por especialistas en la materia.
- Los dictámenes serán emitidos siempre y cuando se disponga de elementos testigo y problema.
- Que el audiocasete o el videocasete problema tengan buena calidad de grabación.
- Que se envíen las grabaciones originales.

II.- TRANSCRIPCIONES

Se realizan reproducciones estenográficas del contenido de la grabación de un audiocasete o un videocasete.

Requerimientos

- La petición debe ser clara y específica sobre el motivo de la intervención del experto.
- El material sujeto a estudio se remitirá al Laboratorio de Análisis de Voz para su examen.
- Es conveniente contar con el original de la grabación, tanto de audio como de video.
- El tiempo en que se emita el dictamen dependerá del contenido de las grabaciones.

Forma y requisitos que debe cumplir la solicitud de intervención

La autoridad ministerial envía solicitud dirigida al Director General de Coordinación de Servicios Periciales, la cual deberá indicar:

- a) Número de averiguación previa.
- b) Número de oficio.
- c) Especialidad(es) que pide.
- d) Nombre y cargo de la autoridad solicitante.
- e) Anexado el material de estudio.

Cuando la solicitud se haga por vía telefónica o fax es indispensable enviar la petición en original.

Esta especialidad brinda servicio en el Distrito Federal a través de la Dirección General de Coordinación de Servicios Periciales, con domicilio en: calle Jaime Nunó No. 25, Colonia Ampliación Morelos, Delegación Cuauhtémoc, México, D. F., C. P. 06200, teléfono 53-46-19-20, fax 53-46-19-80.



