



INFORME PERICIAL

SOBRE LA MUESTRAS DE AUDIO

AUDIO 1 para identificación vocal.mp3

AUDIO 2 para identificación vocal.mp3

AUDIO 3 para identificación vocal.mp3

AUDIO 4 para identificación vocal.mp3

Graudio Forensics
Laboratorio de Acústica Forense
www.audioforensics.es
audioforensics@graudio.es
Avda de Brasil nº 6 1ºplanta.Madrid.España

INDICE

| | |
|-------------------------------------|---------|
| 1. Objeto..... | pág. 3 |
| 2. Fundamentos iniciales..... | pag. 4 |
| 3. Objeto de la pericia..... | pag .5 |
| 4. Análisis de autenticidad..... | pag.6 |
| 5. Análisis vocal | pag. 20 |
| 6. Conclusiones..... | pag 40 |
| 7. Declaración del perito..... | pag 41 |
| 8. Curriculum Vitae del perito..... | pag 42 |
| 9. Referencias bibliográficas..... | pag.43 |

1. OBJETO

A petición de D. Carlos Valverde (en adelante el Cliente), procedemos a redactar el presente Informe Pericial con el fin de emitir un **dictamen sobre ciertos aspectos concretos de varias grabaciones de audio**.

El informe técnico ha sido realizado por Miguel Ángel de la Torre Guijarro , técnico de **Graudio Forensics**, empresa española especializada en Ingeniería de audio y prueba electrónica para aplicaciones en Acústica forense, colabora con consultoras, despachos de abogados y empresas de investigación privada, para la realización de análisis, informes y dictámenes periciales . Entre sus filas, Graudio Forensics cuenta con una plantilla multidisciplinar formada por diferentes perfiles profesionales de la Acústica Forense (Linguistas, Foniatras, Ingenieros y técnicos especialistas de audio)

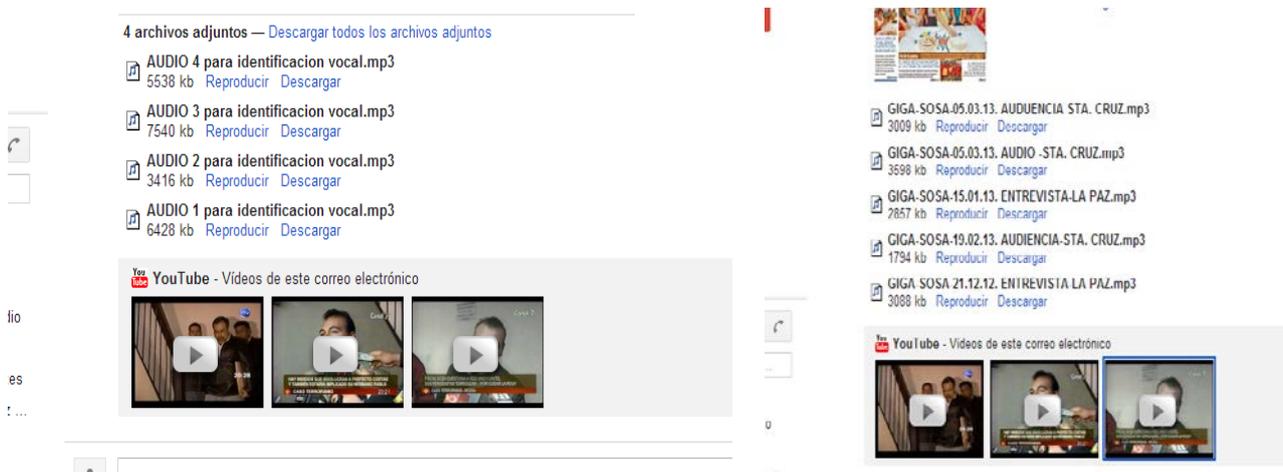
Todos los procedimientos y trabajos realizados por Graudio Forensics utilizan como referencia general los conceptos y procedimientos establecidos por **AES** (Audio Engineering Society), organización de la que es miembro.

**Como anexo se incluye un breve curriculum vitae del perito que firma el presente informe.*

2. FUNDAMENTOS INICIALES.

Para la realización de este informe, con fecha del 12 de Marzo del 2013 se facilita a D. Miguel Ángel de la Torre Guijarro, vía correo electrónico, cuatro muestras de audio dubitadas (cuestionadas) y varias muestras de audio indubitadas (fragmentos procedentes de la Televisión Boliviana), para realizar un proceso de autenticación e identificación vocal.

El analista trabaja en todo momento con **las copias** aportadas por el cliente mediante documento adjunto en un correo electrónico (muestras dubitadas e indubitadas). En ningún momento se facilita el material original.



2.1 Capturas de pantalla con los archivos adjuntos (muestras de voz dubitada y muestras de voz indubitada)

Tras finalizar la copia de cada uno de los archivos, extraemos la firma digital de las muestras dubitadas mediante algoritmo **hash MD5** dando como resultado la siguiente firma digital.

AUDIO 1 para identificación vocal.mp3 **d7c7b3a727756180c5e436be43bcd43f**

AUDIO 2 para identificación vocal.mp3 **dcoe2abd372f6691898361af756b7cca**

AUDIO 3 para identificación vocal.mp3 **05a08ab42fec22be023219f13a5572fe**

AUDIO 4 para identificación vocal.mp3 **c90e29b982a40425a61ef064259b11d1**

**La firma digital de cada archivo permite autenticar la copia y su correspondiente original*

Para la realización del presente informe se han empleado las siguientes herramientas de análisis:

- Herramienta de Análisis Forense ACUSTEK TD-Expert
- Software de edición Protools versión 8.
- Software Adobe Audition

- Sistema de reducción de ruido CEDAR DNS 1000.

- Herramientas de análisis vocal:

VoxMetria

BiometroSoft

SpeechAnalyzer

SIS II (STC-S521) v2.o.199

3. OBJETO DE LA PERICIA

El solicitante del presente Informe, requiere que el análisis se centre en los siguientes **aspectos de la grabación**:

1. Determinar si la voz cuestionada que aparece en los 4 archivos de audio (supuesta voz del abogado Sr. Marcelo Soza), se corresponde con la muestra indubitada (muestra GIGA.SOSA.21.12.12 ENTREVISTA.LA PAZ.mp3), procedente de la entrevista efectuada al Sr. Marcelo Soza por la Televisión de Bolivia.
2. Determinar si la grabación ha sido manipulada, editada o por el contrario mantiene una continuidad en el tiempo y el espacio.
3. Determinar si las muestras de audio analizadas corresponden a fragmentos de la misma grabación, o por el contrario no existen semejanzas o elementos de continuidad entre ellas.

4. ANÁLISIS DE AUTENTICIDAD

Los análisis de autenticación de grabaciones sonoras desarrollados por el laboratorio de Graudio forensics, utiliza como referencia general los conceptos establecidos en los estándares aprobados por el **AES Audio Engineering Society**

En nuestro entorno de trabajo, entendemos por proceso de autenticación, aquel estudio dirigido a determinar si una grabación sonora mantiene la integridad durante el intervalo temporal en que sucedió tal grabación. Dicho estudio también es extensivo a la verificación de si dicha grabación se trata de un registro original o de una copia.

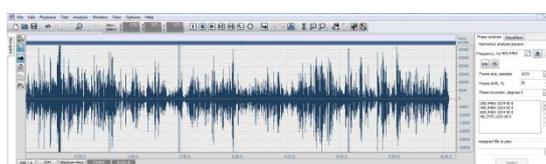
Los distintos análisis a los que se puede someter una grabación de audio, vendrá condicionada por el material aportado por nuestro cliente. En la mayoría de los casos, para realizar un análisis de autenticación exhaustivo, será necesario disponer tanto de la grabación original, como del conjunto de elementos de registro utilizados en el proceso de grabación (micrófono, dispositivo grabador, etc).

4.1 Características de las muestras analizadas

Para el presente informe hemos trabajado con las muestras de audio dubitadas e indubitadas, aportadas por nuestro cliente, todas ellas en formato de audio comprimido mp3, con 2 canales, frecuencias de muestreo de 44,1 khz y resolución de 16 bits.

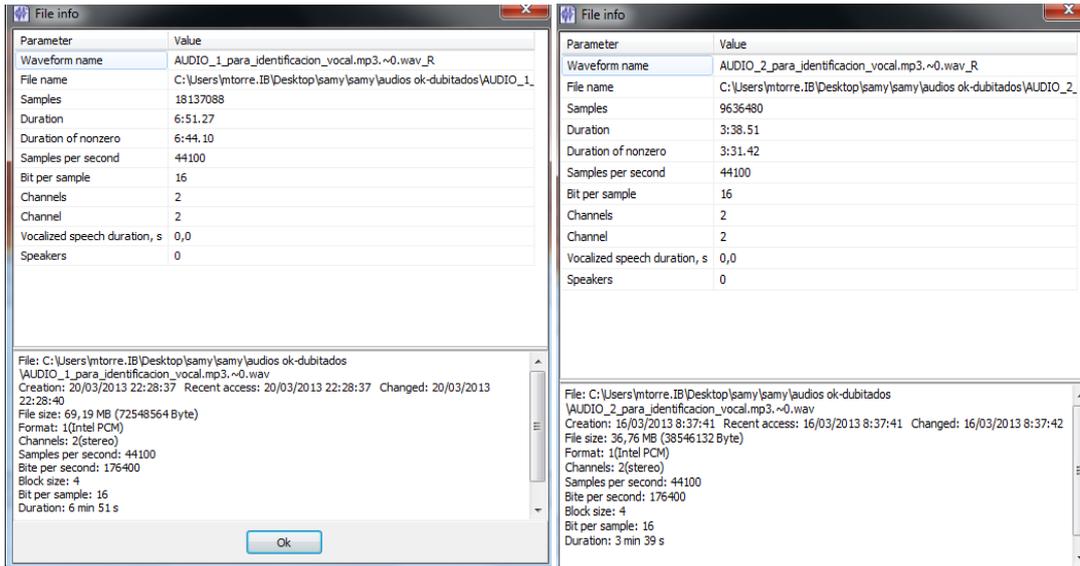
A continuación se adjunta la información obtenida de cada archivo de audio:

AUDIO 1 para identificación vocal.mp3 (6.51.27)

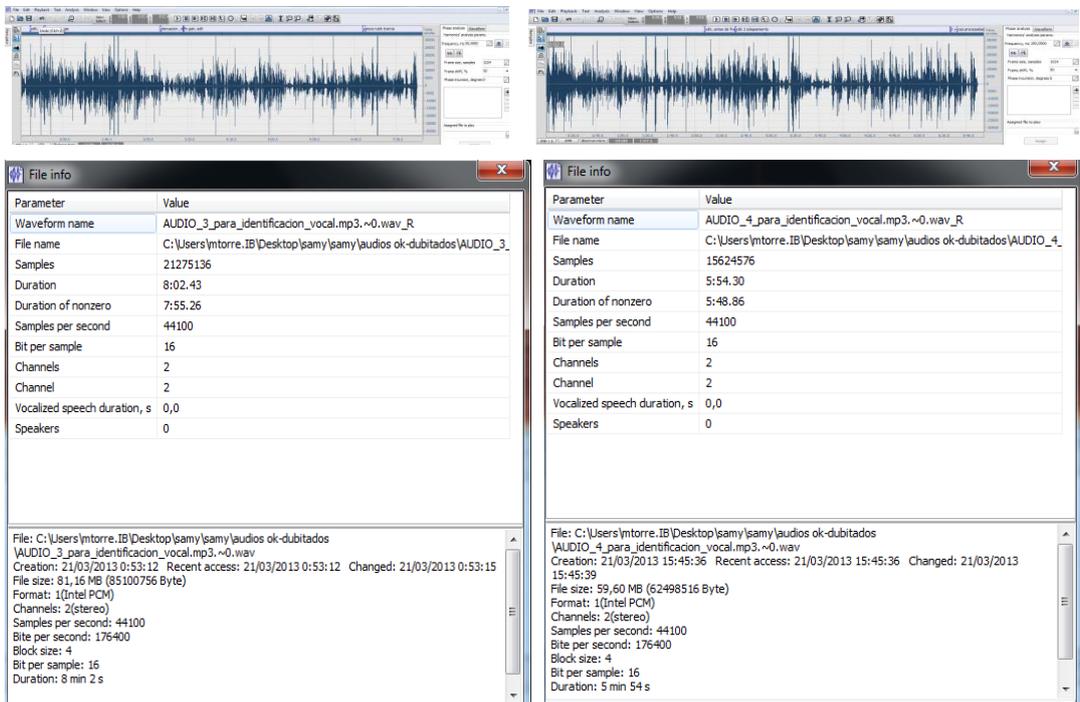


AUDIO 2 para identificación vocal.mp3 (3.38.51)

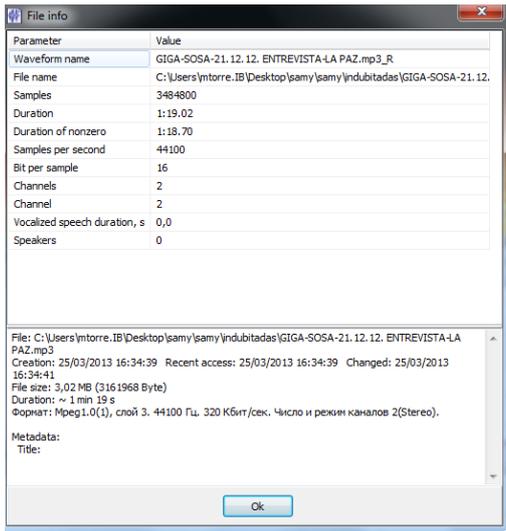




AUDIO 3 para identificación vocal.mp3(8.02.43) AUDIO 4 para identificación vocal.mp3 (5.54.30)

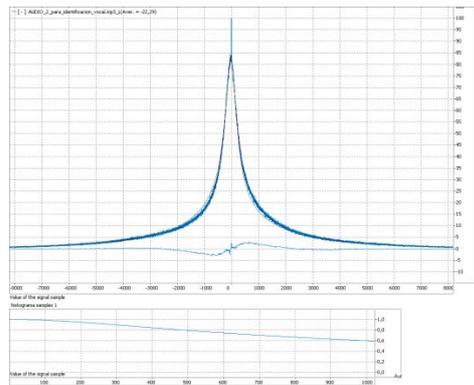
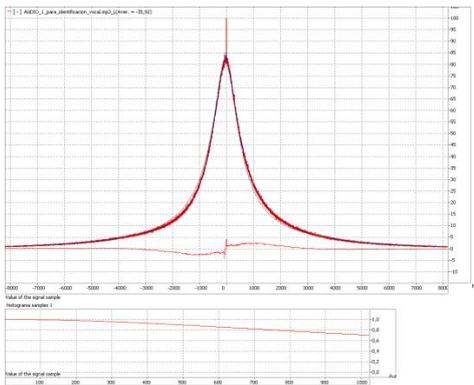


GIGA.SOSA.21.12.12 ENTREVISTA.LA PAZ.mp3 (muestra de voz indubitada)

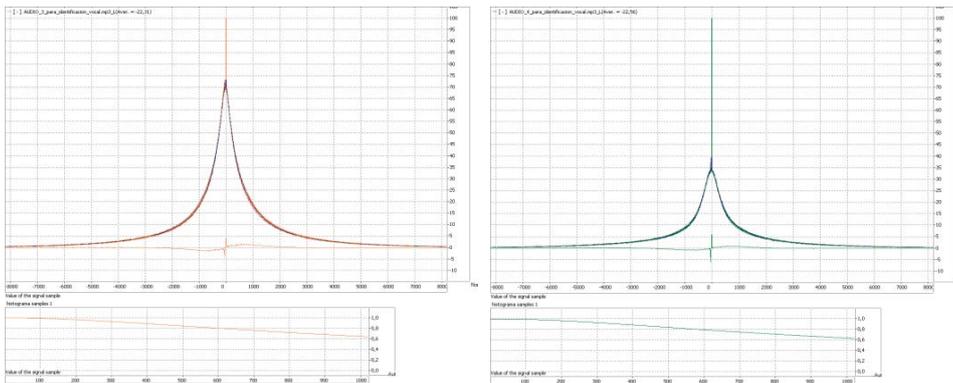


Histograma y distribución de samples de cada muestra de audio

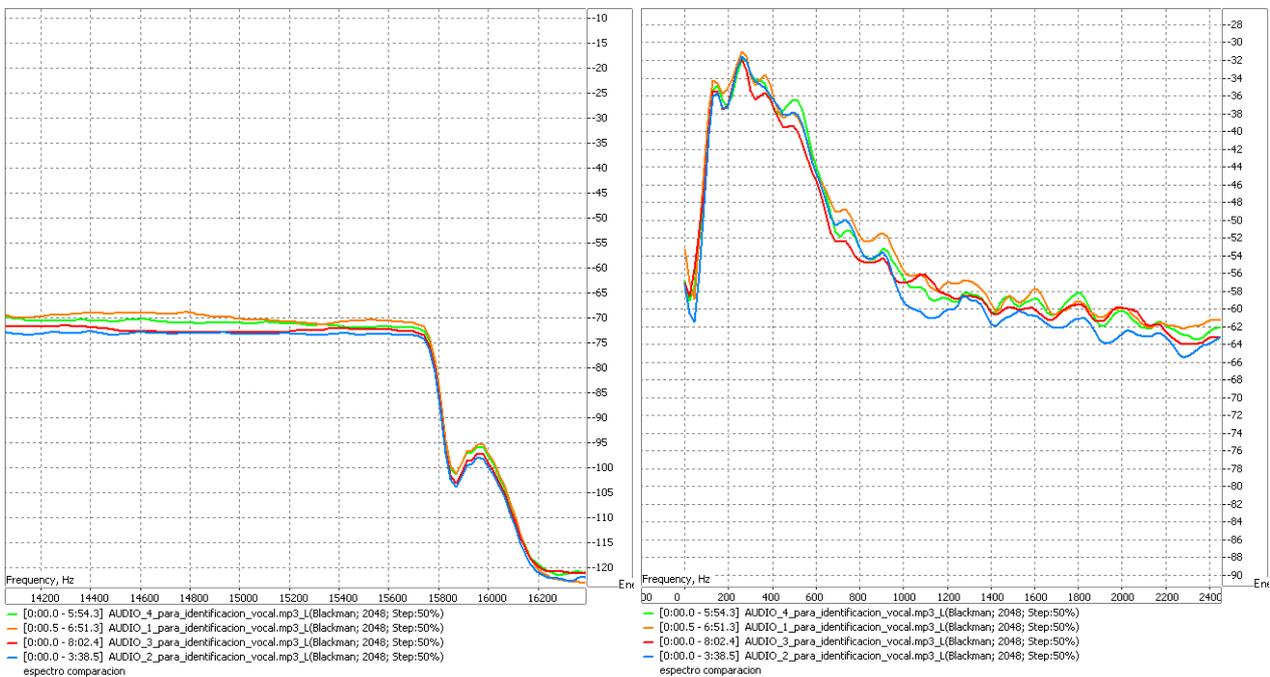
AUDIO 1 para identificación vocal.mp3 AUDIO 2 para identificación vocal.mp3



AUDIO 3 para identificación vocal.mp3 AUDIO 4 para identificación vocal.mp3



El ancho de banda de las muestras de audio no supera el margen de frecuencias superior a los 16Khz, con una atenuación drástica de la señal a partir de los 15,8Khz, común en las 4 muestras de audio analizadas. Lo mismo ocurre en baja frecuencia (entre 200-400hz), la señal de audio está sobremodulada, también característico en las 4 muestras de audio analizadas.



4.1 Fragmentos del espectro en las 4 muestras, características acústicas similares en los límites del ancho de banda, deficiencias en la electrónica del dispositivo grabador.

El análisis de espectro de cada una de las muestras analizadas indica que todas ellas fueron grabadas con el mismo dispositivo grabador.

En el proceso realizado para la identificación del locutor, todos los archivos de audio han de pasar por una serie de pre-procesados para adecuar las características el formato a las herramientas de análisis que se utilizarán a

continuación. El pre-procesado pasa primero por una descompresión convirtiendo cada archivo original MP3, en un archivo con formato WAV. A continuación se convierte el archivo de audio estéreo en uno mono con un solo canal, se elige la opción en este caso de la suma de ambos canales estéreo.

Finalmente se hace una copia del archivo haciendo un nuevo muestreo a 22,050kHz por medio de una diezmado con filtro antialiasing de máxima calidad, que no afectará a los registros de audio en el ancho de banda en que trabajaremos ni supondrá pérdida de cualidades de la voz que puedan afectar al análisis.

Todo este pre-procesado se realiza mediante la herramienta Adobe Audition

4.2 Análisis de las muestras de audio

AUDIO 1 para identificación vocal.mp3

AUDIO 2 para identificación vocal.mp3

AUDIO 3 para identificación vocal.mp3

AUDIO 4 para identificación vocal.mp3

Todas las archivos corresponden a una grabación de audio que recoge la conversación entre tres individuos, voces masculinas, (a partir de ahora: locutor 1, locutor 2 y locutor 3) realizada en un recinto cerrado y pequeño (muy posiblemente un vehículo), con baja influencia del ruido ambiente y utilizando un dispositivo grabador semiprofesional (tipo dictáfono, minigrabadora, etc), ubicado muy próximo a los locutores, principalmente locutor 1 (supuesta voz de Marcelo Soza) y locutor 2 (voz distorsionada intencionadamente). La posición de los locutores con respecto al elemento fonocaptor (micrófono), se mantiene constante en los 4 archivos analizados. No se aprecian variaciones sustanciales de amplitud y frecuencia en cada locutor. Las grabaciones aportan un sonido claro e inteligible.

La interacción de la acústica (proporción entre señal directa y señal reflejada) es mínima por 2 motivos:

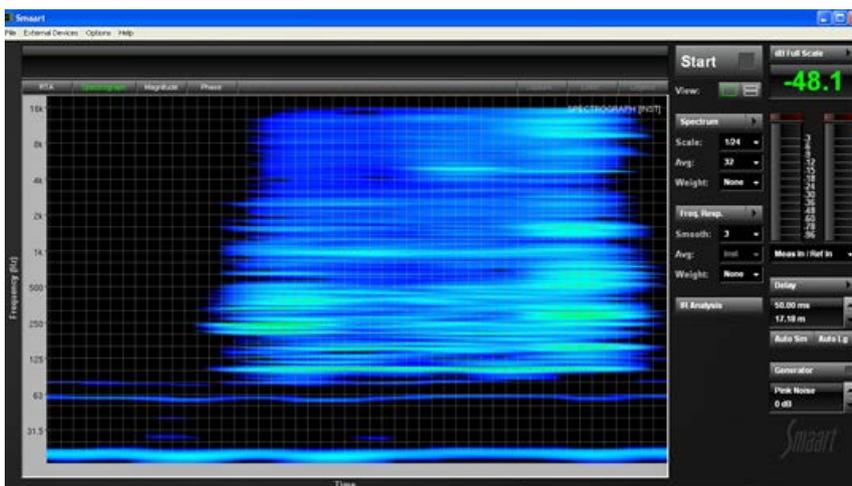
- Pequeña dimensión del recinto donde se desarrollan las conversaciones
- Proximidad de los locutores (locutor 1 y locutor 2) al micrófono o dispositivo grabador.

En los archivos analizados no se escuchan ruidos, que pudieran estar provocados por golpes o rozamiento del micrófono. La voz del locutor 1 (supuesto Marcelo Sosa) es clara e inteligible, con presencia de todo el espectro de voz, no parece una grabación oculta (en ropa u otro complemento). Son frecuentes los ruidos en el interior del vehículo por manipulación y movimiento de objetos.

La posición del micrófono con respecto a cada locutor se mantiene constante en todos los fragmentos analizados de cada una de las muestras, las voces presentan las mismas características acústicas y su relación con el entorno acústico y ruido ambiente es similar. Todos los archivos mantienen continuidad espacio temporal por lo que podemos afirmar que todas las muestras analizadas corresponden a fragmentos de la misma conversación.

La grabación se realiza en un recinto cerrado de pequeñas dimensiones (posiblemente un vehículo), con una constante presencia de ruido ambiente exterior (pasar de vehículos, transeúntes, actividad diurna) y similares características en todas las grabaciones analizadas.

La siguiente imagen corresponde a un análisis espectrográfico del ruido ambiente característico en las 4 muestras de audio (amplitud de la señal, considerando frecuencia y tiempo).



4.3 Análisis acústico y ubicación de los personajes

La grabación recoge la conversación entre tres individuos identificados en un único escenario, un recinto cerrado, posiblemente un vehículo, con características acústicas y ruido ambiente similar en las cuatro muestras analizadas.

El elemento fonocaptor (micrófono) puede aportar información relevante sobre las características acústicas de cada locutor, posición y ubicación con respecto al dispositivo grabador, en base a diferentes criterios acústicos:

- Criterios de nivel o intensidad
- Diferencias espectrales

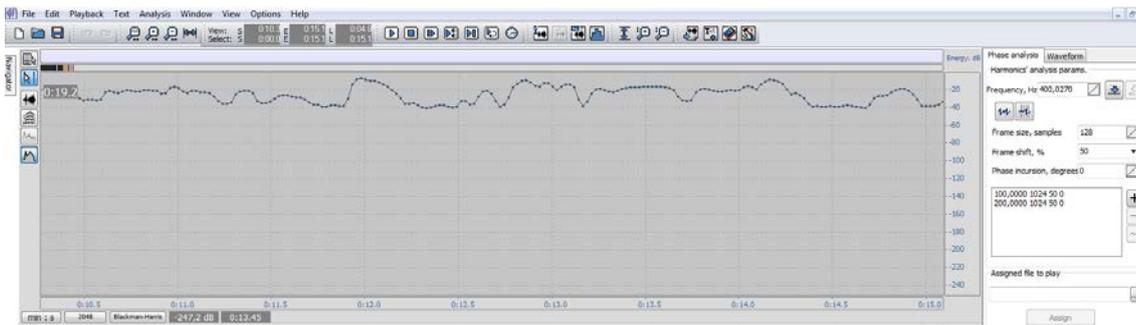
Tomando como referencia determinadas características acústicas de la voz, podemos analizar determinados parámetros (nivel sonoro y frecuencia) de cada locutor, y aproximarnos a conocer de una manera muy fiable:

- Nº de participantes
- Posición
- Distancia entre locutores

INTENSIDAD

El análisis de intensidad o energía, permite conocer el nivel sonoro o amplitud de la señal vocal en un determinado instante, aportando información relevante sobre la ubicación de cada individuo (con respecto al dispositivo grabador) y la relación de proximidad entre ellos, así como la influencia del entorno acústico en el que se desarrolla la acción.

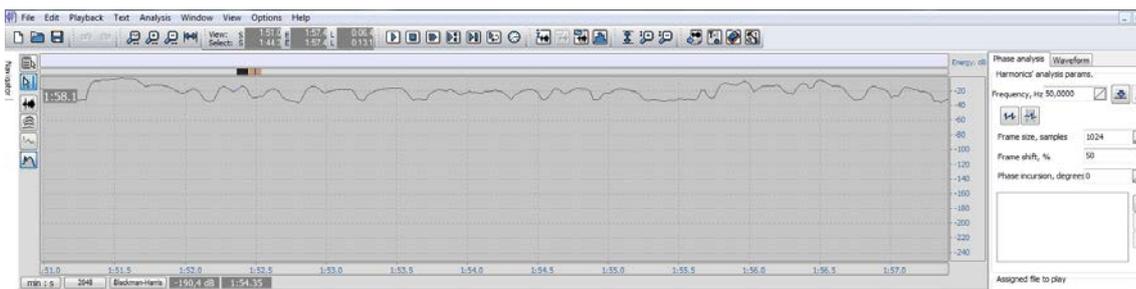
Las siguientes gráficas muestran los valores de intensidad sonora obtenidos tomando como muestra la voz del locutor principal (locutor 1 y supuestamente el abogado Marcelo Soza) en cada una de las muestras de audio aportadas en la prueba .



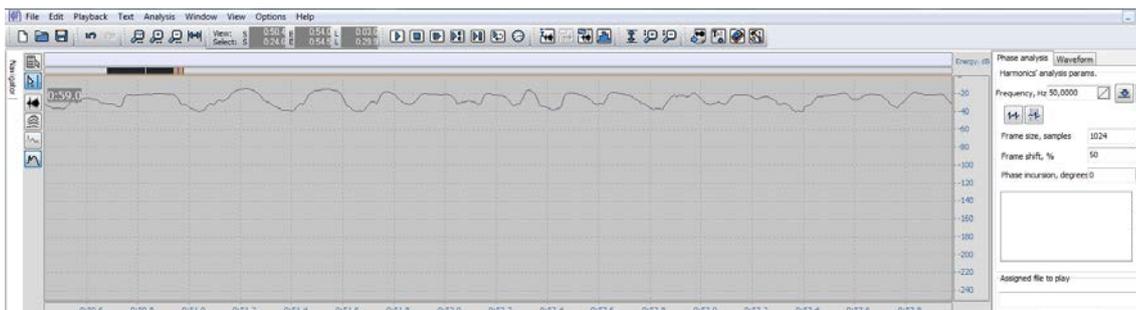
4.2 Niveles de Intensidad sonora correspondientes al locutor 1,(supuesta voz de Soza), en la muestra de audio nº 1



4.3 Niveles de Intensidad sonora correspondientes al locutor 1, (supuesta voz de Soza), en la muestra de audio n°2



4.4 Niveles de Intensidad sonora correspondientes al locutor B, (supuesta voz de Soza), en la muestra de audio n° 3

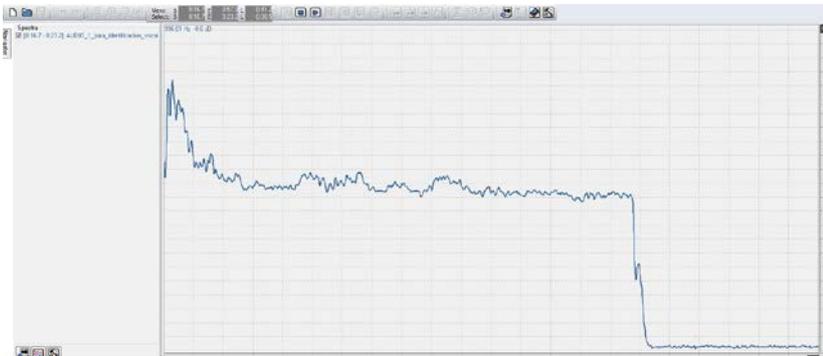


4.5 Niveles de Intensidad sonora correspondientes al locutor 1, (supuesta voz de Soza), en la muestra de audio n°4

Valores de intensidad similares en las cuatro muestras de audio analizadas, el locutor 1 (supuestamente Marcelo Soza) mantiene la misma posición y ubicación con respecto al dispositivo grabador.

FRECUENCIA

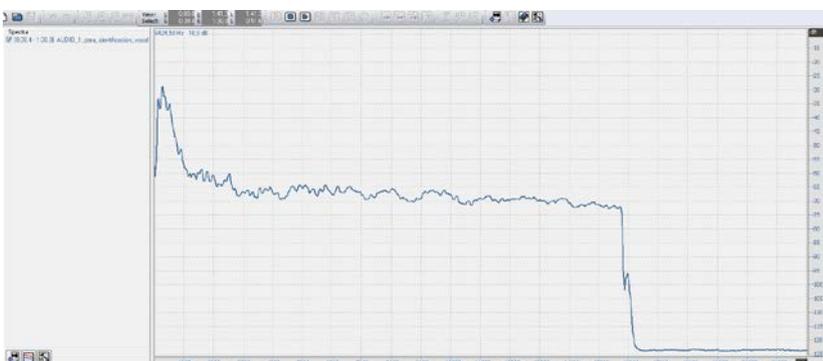
El análisis en frecuencia permite conocer las variaciones del espectro de voz y su amplitud, aportando información relevante sobre la ubicación de cada individuo y la relación de proximidad entre ellos, así como la influencia del entorno acústico en el que se desarrolla la acción.



4.6 Análisis espectral correspondientes al locutor 1 (supuesta voz de Soza), en la muestra de audio nº 1



4.7 Análisis espectral correspondientes al locutor 1 (supuesta voz de Soza), en la muestra de audio nº 2



4.8 Análisis espectral correspondientes al locutor 1 (supuesta voz de Soza), en la muestra de audio nº 3



4.9 Análisis espectral correspondientes al locutor 1 (supuesta voz de Soza), en la muestra de audio 4

En todas las grabaciones analizadas, el locutor 1 (supuesta voz de Marcelo Soza), presenta las mismas características espectrales y ancho de banda.

Las siguientes gráficas corresponden a las características espectrales de los 2 locutores que intervienen en la conversación.



4.10 Análisis espectral correspondientes al locutor 2, voz distorsionada y más próximo al locutor 1.



4.11 Análisis espectral correspondientes al locutor 3, voz más alejada del locutor 1 y del dispositivo grabador



4.12 Análisis espectral y comparación de los tres locutores. Valores muy similares en todas las muestras analizadas.

En base a estos criterios podemos afirmar que **todas las muestras analizadas corresponden a fragmentos de la misma conversación.**

4.4 Edición, cortes y manipulación del material sonoro

Todas las grabaciones analizadas presentan cortes o eventos electroacústicas de corta duración que no son lógicos en el entorno acústico en el que se desarrollan. Circunstancias que pueden asociarse a manipulación con posterioridad al proceso de grabación sonora. Sin embargo podemos afirmar que **la grabación mantiene una secuencia lógica, tanto en el plano semántico como expresivo.**

No se aprecian paradas, ni pausas en el dispositivo grabador.

La manipulación de audio se manifiesta de 2 maneras diferentes:

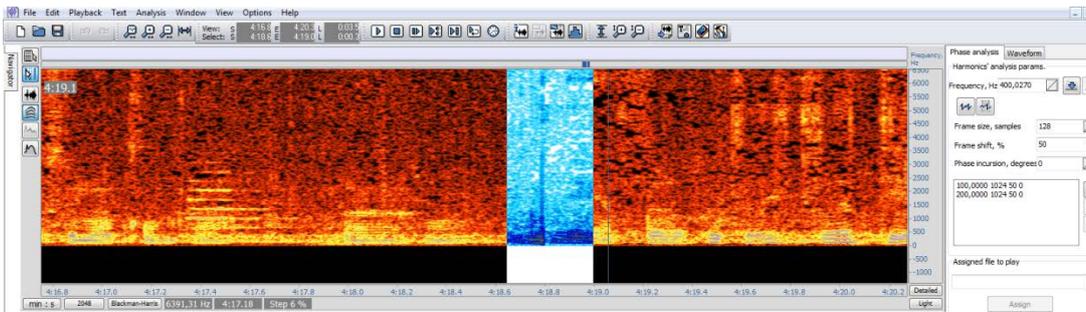
- Cortes y puntos de edición de audio en algunas partes de la conversación
- Las voces que corresponden a locutor 2 y locutor 3 han sido procesadas para evitar que sean reconocidas. El procedimiento empleado consiste en variaciones de la velocidad que afectan al tono e inteligibilidad de la voz.

A continuación identificamos el número de alteraciones encontradas en cada una de las muestras:

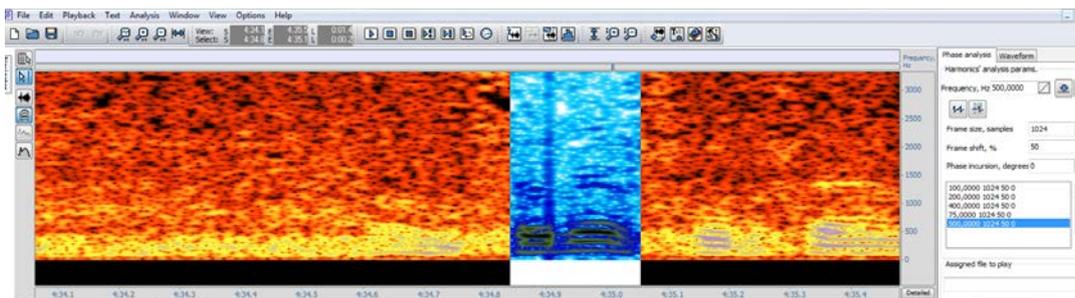
AUDIO 1 para identificación vocal.mp3

Cortes/Edición: 2 (minutos 4.18, 4.34)

Manipulación de las voces Loc 2, Loc 3: 10



4.13 Espectrograma que muestra punto de edición (minuto 4.18) con variación de la amplitud armónica de la señal, no hay ausencia de señal.

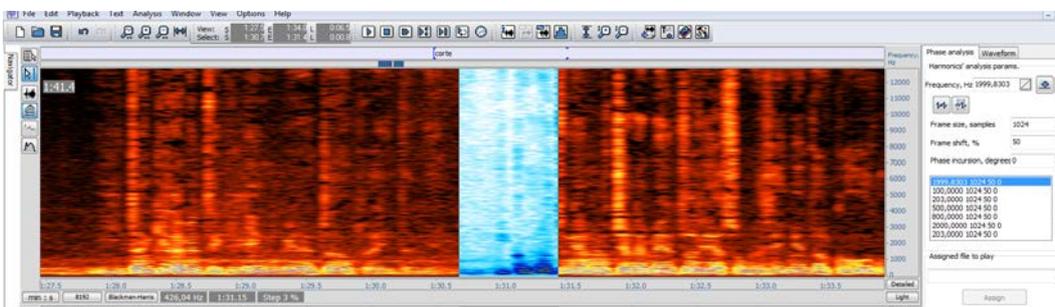


4.14 La zona marcada de color azul en el espectrograma representa el punto de edición donde se solapan varias frases procedentes del Locutor 1 (minuto 4.34)

AUDIO 2 para identificación vocal.mp3

Cortes/Edición: 3 (minutos 1.31, 2.48, 3.11,)

Manipulación de las voces Loc 2, Loc 3: 9

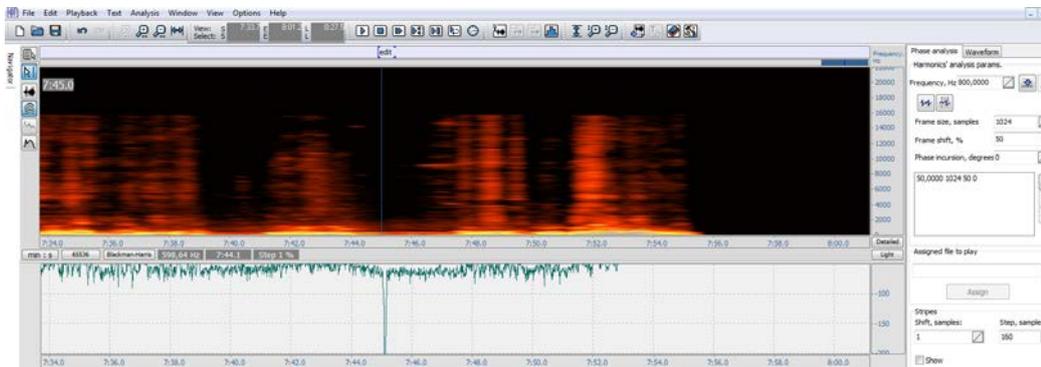


4.15 Espectrograma que muestra punto de edición correspondiente al minuto 1.31

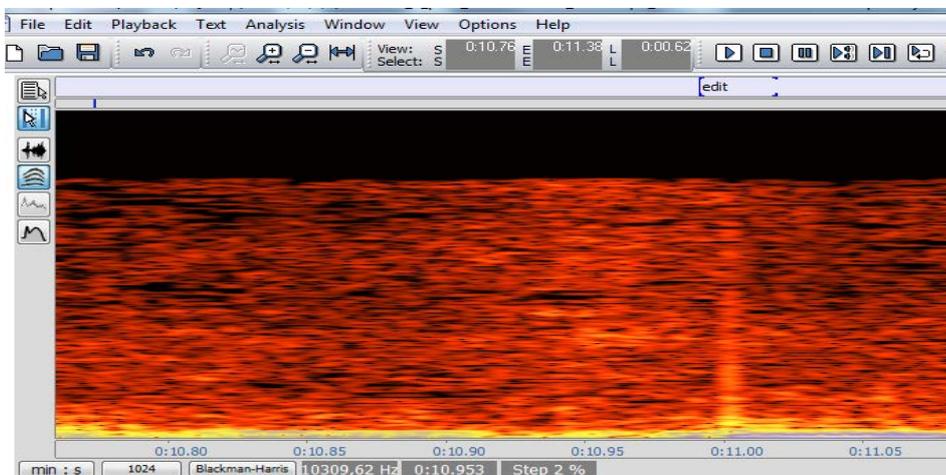
AUDIO 3 para identificación vocal.mp3

Cortes/Edición: 7 (minuto 0.10, 2.45, 3.13, 3.21, 5.48, 5.57, 7.39)

Manipulación de las voces Loc 2, Loc 3: 12



4.16 Espectrograma que muestra punto de edición entre 2 frases (minuto 7.39), fuerte atenuación de la amplitud armónica de la señal

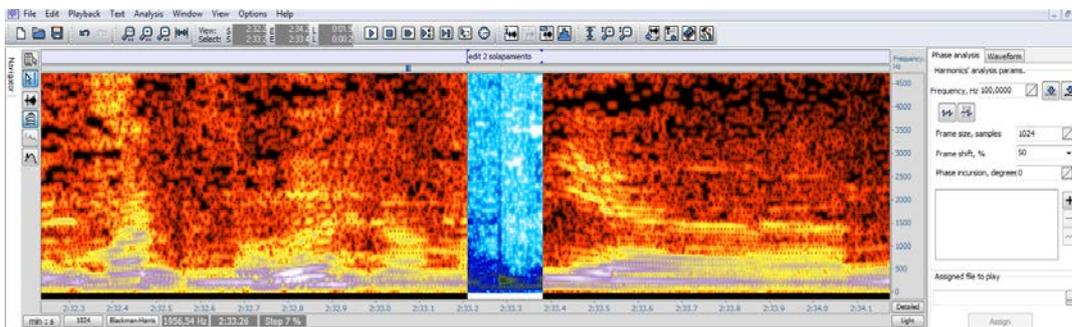


4.17 Espectrograma que muestra punto de edición en el minuto 0.10.58

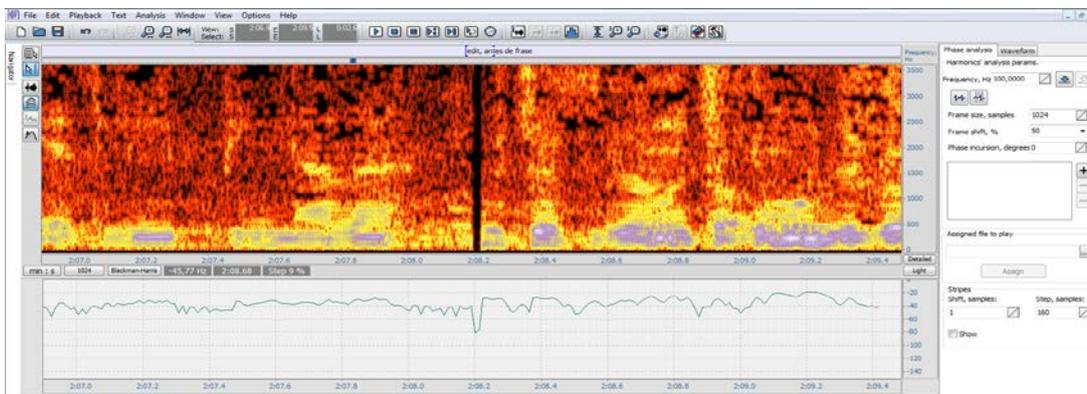
AUDIO 4 para identificación vocal.mp3

Cortes/Edición: 3 (minuto 2.08, 2.33, 5.31)

Manipulación de las voces Loc 2, Loc 3: 5



4.18 Espectrograma corresponde al minuto 2.33, solapamiento entre frases.



4.19 Espectrograma correspondiente al punto de edición, minuto 2.08

5. ANÁLISIS VOCAL

5.1 Materiales y métodos

El material aportado para el análisis consta de 5 pistas estéreo en formato mp3, una de ellas etiquetada como voz indubitada (GIGA.SOSA.21.12.12 ENTREVISTA.LA PAZ.mp3) y por lo tanto la atribuiremos al L-A (locutor A: abogado Marcelo Soza) y las otras 4 atribuidas al mismo locutor (supuestamente el abogado Marcelo Soza) pero que serán etiquetadas en un principio como L-B/1, L-B/2, L-B/3 y L-B/4. Esto se debe a que recibimos los archivos en 4 fragmentos y aunque en partes anteriores del documento se sostuvo que los cuatro fragmentos son de la misma conversación, haremos un análisis de diferentes fragmentos intralocutor antes del cotejo de ambos locutores.

5.1.1. Pre-procesado

En primer lugar los archivos de audio han de pasar por una serie de pre-procesados para adecuar las características el formato a las herramientas de análisis que se utilizarán a continuación. El pre-procesado pasa primero por una descompresión convirtiendo cada archivo original MP3, en un archivo con formato WAV. A continuación se convierte el archivo de audio estéreo en uno mono con un solo canal, se elige la opción en este caso de la suma de ambos canales estéreo. Finalmente se hace una copia del archivo haciendo un nuevo muestreo a 22,050kHz por medio de una diezmado con filtro antialiasing de máxima calidad, que no afectará a los registros de audio en el ancho de banda en que trabajaremos ni supondrá pérdida de cualidades de la voz que puedan afectar al análisis.

Todo este pre-procesado se realiza mediante la herramienta Adobe Audition.

5.1.2. Limpieza de fragmentos de voz

Para el análisis de identificación automático y herramientas estadísticas se decide utilizar un tiempo de unos 10 minutos de voz dubitada para poder ser comparada y verificada. En los archivos recibidos se aprecian otras voces además de las del locutor LA y LB las cuales se eliminan de las pistas para su procesado, de la manera que se explica a continuación:

- Fragmento L-A_22kHz

Se trata de una entrevista en la que el L-A contesta a preguntas formuladas por una segunda persona. Todas sus intervenciones se realizan respetando el orden de la conversación siendo su voz no interferente, por lo que es eliminada de la pista en los fragmentos del minuto 0:26.333 a 0:29.000 y de 0:49.000 a 0:51.144.

- Fragmento L-B/4_22kHz

Se elige el fragmento 4 para este tipo de análisis. En este fragmento existe otro locutor que asiente de vez en cuando, respetando el turno de la conversación por lo que simplemente eliminamos su voz mediante el algoritmo "limpieza automática" de Adobe Audition. En el minuto 00:02:08 aparece otra voz muy grave que parece distorsionada al mismo tiempo que habla L-B. Por esta razón silenciamos el fragmento entre los minutos 2:08.214 y 2:12.447 y

nuevamente entre 2:30.388 y 2:36.369 y más tarde entre 3:53.028 y 3:54.737 y entre 4:50.348 y 4:51.785 y entre 5:27.025 y 5:31.568. Una nueva voz se encuentra entre 2:26.000 2:27.088 que también puede perturbar los resultados con lo que la silenciamos también. Entre los minutos 3:08.679 y 3:15.475 el locutor L-B es registrado a menor nivel, pero decidimos no manipular el sonido.

- Fragmento L-B/3_22kHz

Este fragmento contiene gran cantidad de interferencias tipo voz, tipo ruidos de lo que parecen ser cuerdas o rodillos. Los primeros segundos se nota ruido de movimiento del fonocaptor.

En la se Tabla 5-1 se detallan los procesos de limpieza de la voz dubitada.

Tabla 5-1: Procesado de limpieza y adecuación de las pistas de voz dubitada para el análisis estadístico y automático

| | Inicio | Fin | Audio | Decisión |
|-----------|--|--|---|---|
| LB_03_22k | 0:17.127 0:26.000 1:05.533 2:09.749 4:11.322 | 0:21.409 0:27.242 1:14.177 2:14.517 4:16.000 | Interferencia y voz de otro locutor que interfiere con LB | Eliminado |
| LB_03_22k | 1:26.218 4:50.634 | 1:28.516 5:00.788 | Voces interferentes | Eliminado |
| LB_03_22k | No se toman todos los tiempos 2:32.595 | No se toman todos los tiempos 2:33.007 | Voces no interferentes tipo asentimientos | Limpieza automática |
| LB_03_22k | 2:54.756 3:03.366 3:19.386 | 3:00.279 3:05.493 3:23.468 | Voces no interferentes | Eliminado |
| LB_03_22k | 0:21.391 | 0:46.220 | Interferencia | Elimina de la pista pero se procesará por separado como archivo LB_03_22k_INC01 |
| LB_04_22k | 2:08.214 2:30.388 3:53.028 4:50.348 5:27.025 | 2:12.447 2:36.369 3:54.737 4:51.785 5:31.568 | Voz grave hablando al mismo tiempo que L-B | Silenciado |
| LB_04_22k | 2:26.000 | 2:27.088 | Otra voz interviene a la vez que LB | Silenciado |

5.1.3. Herramientas de análisis

Las principales herramientas de análisis que serán utilizadas durante el presente trabajo serán:

- VoxMetria: Análisis estadístico, histograma y triángulo vocálico.

- BiometroSoft: Caracterización biométrica de la onda glótica.
- SpeechAnalyzer: Espectrograma, LPC y cepstral.
- SIS II (STC-S521) v2.o.199: Identificación automática.

5.2. Marco lingüístico de los locutores

Para conocer algunas características relativas al español boliviano y a sus rasgos dialécticos, encontramos el documento (CORDÓN, 1980) del Instituto Cervantes, en él concluye que la existencia de dos áreas dialectales en Bolivia. El altiplano y los llanos se separan tajantemente en la articulación de /r/, en la realización fonética de /s/ en posición implosiva, y, aunque en menor grado, en la articulación de las vocales átonas. En el altiplano, el empleo de [r̥] es más frecuente que el de [r], pero entre los llanos hay marcado predominio de [r]. En posición implosiva, [s] suele realizarse mediante [s̥] en el altiplano, mientras que alternan [s], [h] y [θ], con predominio de los dos últimos, en los llanos. Entre los llanos, las vocales átonas a veces se articulan de manera imprecisa, fenómeno que no se da en las tierras bajas.

Algunas otras consideraciones son, que Bolivia es un país *lleista* ya que usa [λ] de forma predominante, La pronunciación de /r/ tras /t/ reveló otra diferencia más entre la fonética del altiplano y la de los llanos. En las tierras altas, el 18% de los informantes emplearon [t̥r] exclusivamente y otro 59% alternaron [t̥r] y [tr].

También se considera distintivo entre clases populares del llano, la confusión entre [l] y [r] a principios de palabra.

Por todo esto consideraremos el análisis de las realizaciones del fonema /r/ como uno de los rasgos que pueda diferenciar a un locutor de otro debido a su lugar de origen.

5.3. Identificación del locutor por medio de análisis vocal

5.3.1. Biometría de la voz

La biometría se puede definir en pocas palabras como el estudio de métodos automáticos para el reconocimiento único de humanos basados en uno o más rasgos conductuales o rasgos físicos intrínsecos. El caso más conocido por todos de rasgo biométrico sería el de las huellas dactilares. En todo sistema biométrico se trata de extraer las características del rasgo biométrico y crear un modelo con estas características que pueda ser único y propio del individuo.

En el caso de la voz, la parametrización directa de la señal de voz íntegra contiene información propia de la biometría del locutor (estructura y dimensiones del tracto vocal, biomecánica de las cuerdas vocales, etc.), pero está sujeta a variaciones fuertemente condicionadas por el proceso articulatorio. Existe la creencia de que la parametrización por separado de los rasgos biométricos de tracto vocal y de pulso glótico pueden describir mejor las características de un locutor dado, es por esto que a lo largo del documento trataremos por una parte la señal vocal como suma del efecto de las cuerdas vocales, el tracto vocal y la articulación y por otro evaluaremos la onda glótica.

5.3.2. Características y visualización de la señal de voz

5.3.2.1. Espectrograma

El espectrograma muestra la señal utilizando tres variables. En el eje horizontal se encuentra el tiempo, en el vertical la frecuencia, y en la intensidad o coloración de la representación de la señal tenemos la intensidad de la emisión.

El espectrograma nos muestra las características de la señal vocal en estas tres dimensiones, con lo que nos da información en un vistazo de las formantes de la voz, duración y reparto de energía en los diferentes sonidos del habla.

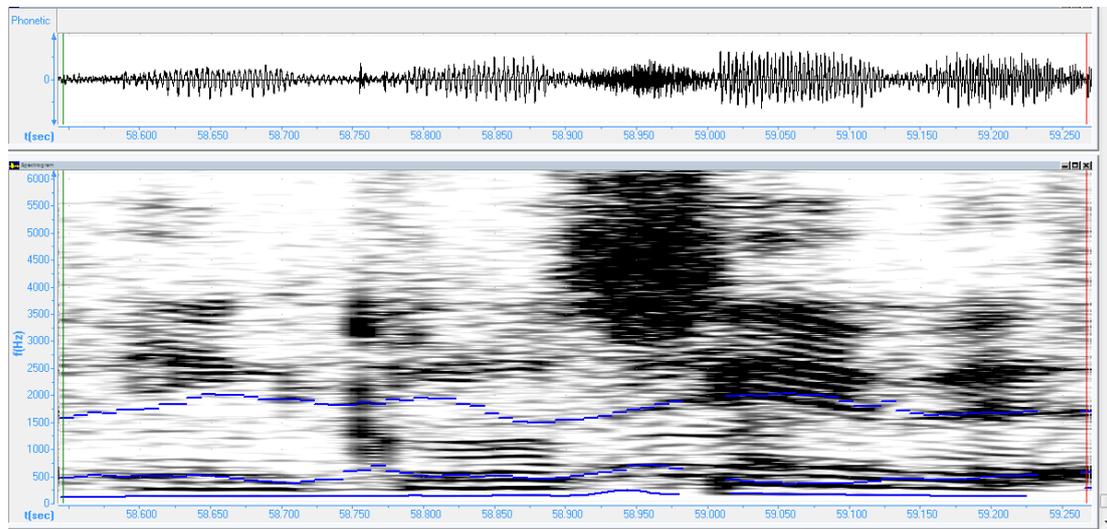


Figura 5-1: Espectrograma de Banda Estrecha

El espectrograma se realiza mediante Transformada Rápida de Fourier con lo que los resultados dependerán del tipo de enventanado y características de la FFT que se impongan. Generalmente se utilizan dos tipos de espectrograma en análisis vocal, el de banda estrecha y el de banda ancha. El de banda ancha nos permite obtener de forma visual una mayor claridad a la hora de estudiar los formantes y la energía en los ruidos no periódicos, mientras que el de banda estrecha ofrece más información armónica y de la estructura fina del habla.

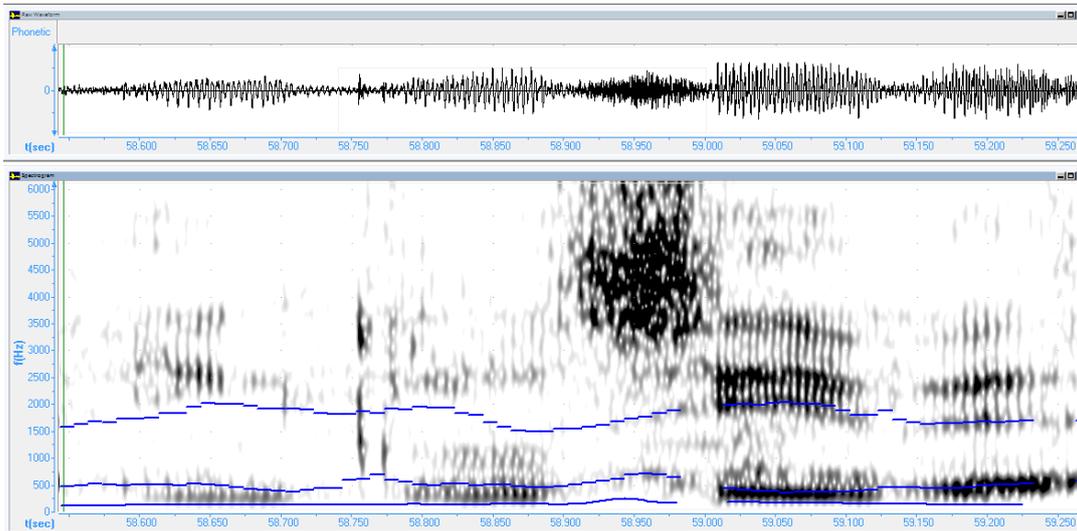


Figura 5-2: Espectrograma de Banda Ancha

5.3.2.2. Espectro, LPC y dominio Cepstral

El espectro de la señal de voz obtenido des la trasformada de Fourier de la señal acústica, nos muestra una señal compleja y con cierto rizado. Este espectro cuenta en los sonidos vocálicos con una serie de formantes o frecuencias de acumulación de energía, y puede entenderse como un filtro todo-polos con 3 frecuencias enfatizadas (F_1 , F_2 y F_3). En síntesis de voz se ha modelado la señal vocálica como la suma del efecto de varios filtros en cascada, uno modela la onda glótica, el otro modela la el tracto vocal y un último filtro modela la radiación vocal.

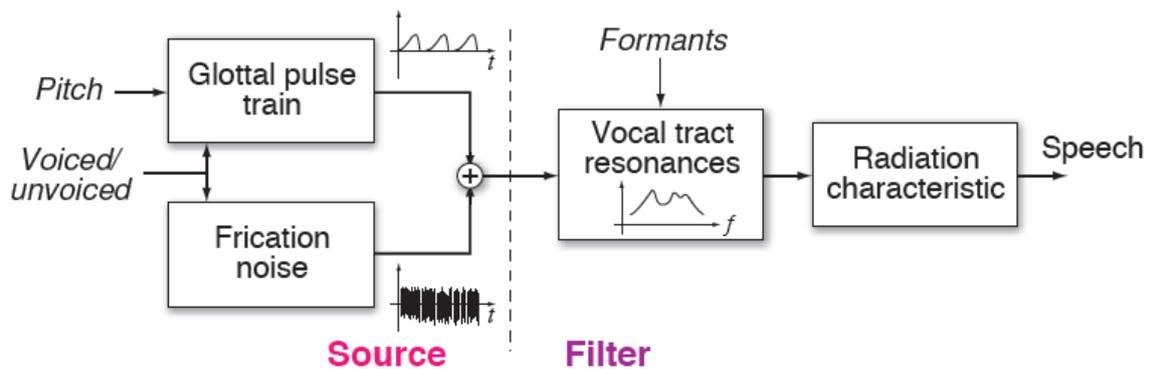


Figura 5-3: Modelo de síntesis de voz.

El filtro que se ocupa del tracto vocal ha sido modelado en síntesis de voz como un filtro de predicción lineal (LPC), el cual dependiendo del fonema a reproducir tendrá unos coeficientes diferentes modelando las formantes F_1 , F_2 y F_3 (Chen, Bilmes, & Ellis, 2005).

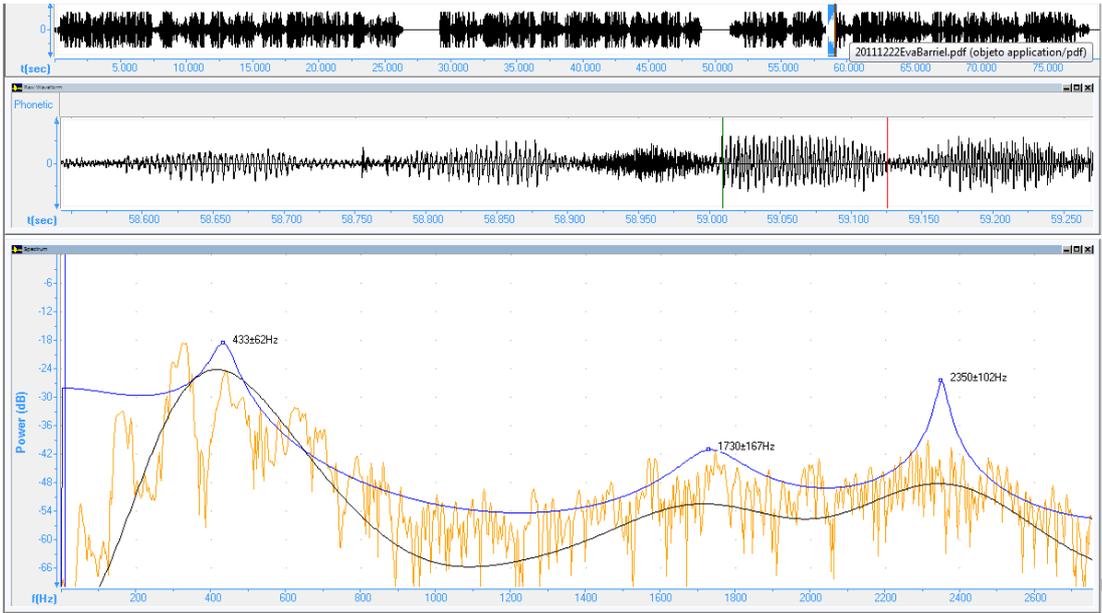


Figura 5-4: Ejemplo de representación del espectro mediante FFT (amarillo), LPC (en azul) y de *cepstrum* (negro).

Otro método que se ha utilizado para un mejor análisis vocal es el dominio *cepstral*, método que permite obtener descomponer la señal de voz en dos componente, la señal periódica modelada como un tren de deltas en la Figura 5-1 y la función de transferencia de los filtros.

5.3.2.3. Triángulo vocálico

Cada realización vocálica se produce por la acción de las cuerdas vocales en ausencia de obstáculos, pero modificando el volumen del tracto vocal. La posición de la lengua juega el papel más importante ya que su posición en la cavidad bucal hace que fonética y acústicamente cada vocal sea diferente.

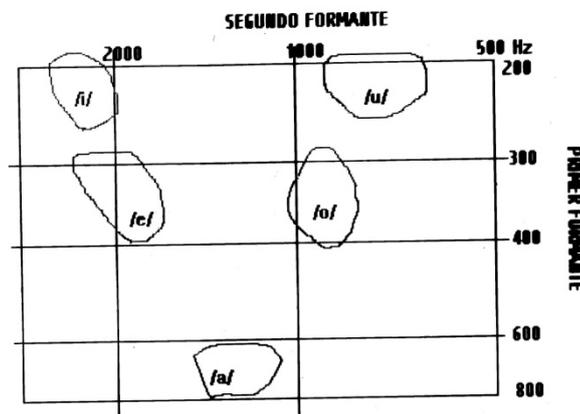


Figura 5-5: Triángulo vocálico del español, representado mediante gráfico F1 vs F2

Las vocales se caracterizan acústicamente por ser sonidos periódicos y tener 3 formantes que las caracteriza y las segmentan en abiertas y cerradas, y en altas, media, y bajas.

a: $F_1=660$, $F_2=1130$ i: $F_1=250$, $F_2=2600$
 e: $F_1=480$, $F_2=2120$ o: $F_1=500$, $F_2= 930$
 u: $F_1=270$, $F_2= 650$

5.3.2.4. Histograma F_0

El pitch o frecuencia fundamental, se debe a la acción de las cuerdas vocales. Se ha observado que esta vibración no es constante a lo largo del discurso, detectándose variaciones a lo largo de la frase y también dentro mismo de una palabra. Estas variaciones se deben tanto a la entonación de la frase, como a la acentuación de los fonemas así como al estado emocional del orador.

Mediante procesos estadísticos puede estudiarse a lo largo de un discurso como el locutor varía esta frecuencia fundamental en función a la entonación y la prosodia, y comprobar cuales con los valores más probables en dicho locutor.

5.3.2.5. Onda Glótica

Recordando el esquema anterior sobre producción de voz, podemos sintetizarlo en la realización de vocales de la siguiente manera.

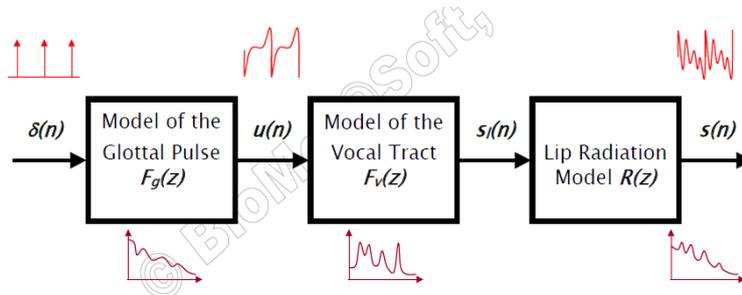


Figura 5-6: Explicación de las diferentes funciones de transferencia en la fonación humana.
 Fuente:BioMetroSoft.

La onda glótica es la onda producida por el movimiento de las cuerdas vocales, y produce una función de transferencia análoga a dicho movimiento. Las cuerdas vocales no vibran todo el tiempo de forma estática durante los ciclos de fonación sino que ocurre un movimiento entre las cuerdas como se muestra en la Figura 5-4.

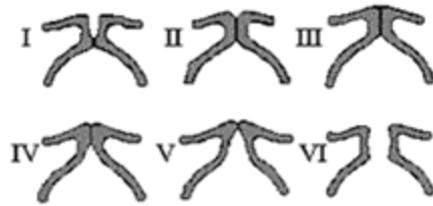


Figura 5-7: Movimiento de las cuerdas vocales en 6 pasos

Para caracterizar la onda glótica como rasgo biométrico se le han atribuido a su forma unos 5 parámetros al igual que se hace con los PQRST de los electrocardiogramas.



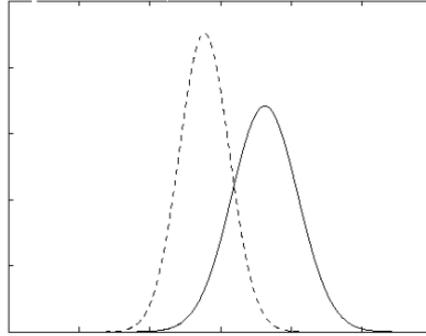
Figura 5-8: parámetros de la onda glótica

La herramienta que utilizaremos para este análisis mostrará los valores de CROA de las ondas glóticas estudiadas y las cotejará con los parámetros del modelo biofísico correspondiente, dándonos información del locutor más profunda. De esta manera obtendremos los parámetros más distintivos como pueden ser la rigidez de las cuerdas vocales o la masa de éstas (Gómez Vilda et al., 2008).

5.3.3. Identificación automática

La identificación automática se realiza en sistemas de seguridad, de reconocimiento de voz, de huella dactilar, etc. En (Nolan, 2001) se hace referencia a las limitaciones y formas de la identificación de locutores ampliamente.

La identificación se basa en la creación de un modelo gaussiano que caracterice la probabilidad de obtener un falso aceptado (FA) y un falso impostor (FR). La creación de estos modelos depende de la base de datos y de las características que compongan el modelo. El resultado de esta identificación automática nos dirá en valores de % la probabilidad de que el locutor L-B sea aceptado como L-A.



La curva DET resume la discriminación del conjunto experimental de valores de LR en una única curva. En ella se representa la probabilidad de FA frente a la probabilidad de FR para todos los puntos de funcionamiento del sistema.

5.4. Análisis locutor muestra L-A (voz indubitada)

5.4.1. Análisis subjetivo

El locutor A (L-A) identificado como Marcelo Soza se encuentra en la grabación indubitada en un contexto concreto, una entrevista en un medio de comunicación. La percepción subjetiva de la grabación es la de seguridad, firmeza y casi de discurso ensayado. En algunas ocasiones rellena las pausas utilizando la muletilla “eh” como recordando el discurso. Otra característica que no puede pasar por alto es el seseo, realizando siempre [s] en las realizaciones de /θ/ (Quilis, 1997).

- Muletillas: A parte de ese “eh” dubitativo entre algunas frases, cuando concluye algunas partes del discurso añade la muletilla “no?”
- Ritmo: El ritmo no es constante, acelerando y decelerando el discurso de forma intencionada y enfática.
- La entonación es variante, no realizando entonación acorde con el acento del grupo fónico, siendo una entonación propia de la zona demográfica pero incluyendo acentos forzados enfatizando algunas ideas.

5.4.2. Análisis articulatorio

- Neutralización. Los archifonemas que encontramos en la locución son tratados de forma poco clara, por ejemplo en “...averiguación de la verdad” /aBERigwaθiÓN de la veRdáD/ realiza claramente [n], [r] y en la realización de /D/ hay cierto debilitamiento. También las dos veces que dice /malkobíG/ lo hace de forma distinta, en la primera realiza el archifonema de forma glotal, y en la segunda ver lo vuelve fricativo.
- Oclusivas: la realización [k] tiene una doble explosión de duración entre ellas unos 20ms y unos 30ms hasta la realización de la vocal. El resto de oclusivas sordas tienen el intervalo de explosión (*Voice onset time*) en unos 20ms.
- Fricativas: en la [s] la energía se concentra a partir de los 3-3,7kHz

- Pitch: La media de Fo se encuentra en 151Hz y tiene una variación de 262Hz.

5.4.3. Análisis de la fonación

Para el análisis de la fonación capturamos el análisis de onda glótica de la vocal /a/.

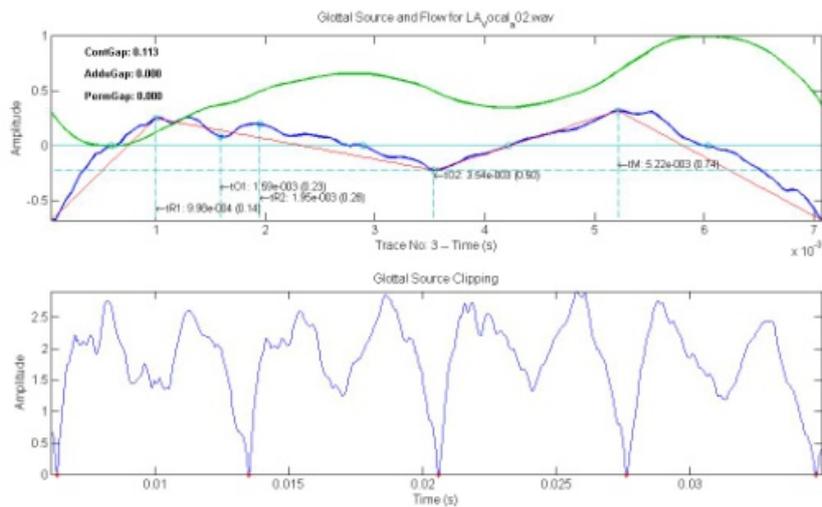


Figura 5-9: Onda Glótica vocal L-A /a/

La curva azul muestra el movimiento de las cuerdas vocales y la curva verde el paso de aire a través de la laringe. La onda glótica masculina estándar tiene una forma más lineal ambas curvas. En la mitad de cada ciclo aparece una inflexión.

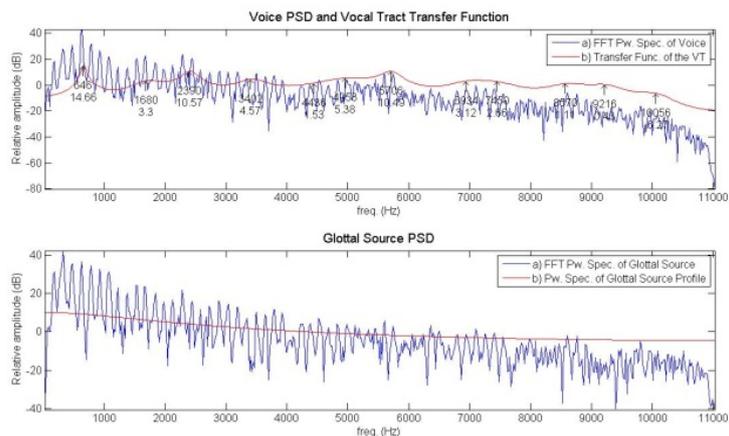


Figura 5-10: Funciones de transferencia de tracto vocal y glótica de L-A.

La función de transferencia muestra que existen formantes hasta los 10kHz aproximadamente

5.5. Análisis locutor L-A frente a L-B

5.5.1. Análisis subjetivo

La voz de locutor L-A y la de L-B (supuestamente el abogado Marcelo Soza) pueden ser segmentadas de igual manera: voz masculina, habla propia la misma zona geográfica, rasgo de seseo, edad similar, habla con ritmo variante, con diferencias en entonación de la voz y en intensidad a lo largo de la locución. Voz en ocasiones apresurada. En cuanto al timbre es suficientemente similar para un oyente externo.

Las principales diferencias a nivel de características subjetivas las encontramos en que L-A es a veces dubitativo y aunque fluido parece forzado, además de hablar de manera afirmativa y segura, casi monològística, mientras que L-B parece más fluido, utilizando una manera más expresiva incluso descuidada. En la locución de L-B (voz dubitada) no se aprecian pausas rellenas con el sonido “eh” mientras que sí ocurre en la locución que tenemos de voz indubitada L-A.

5.5.2. Análisis articulatorio

En la articulación de vocales podemos observar en la Figura 5-8 cómo se distribuyen las realizaciones en ambos locutores.

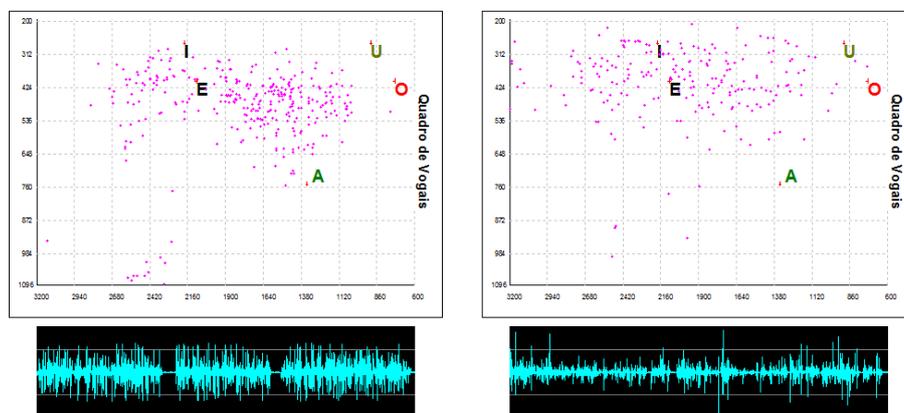


Figura 5-11: Análisis de formantes vocálicos: L-A vs L-B

Las realizaciones en L-A se distribuyen de forma más o menos clara dentro del triángulo vocálico propuesto para el español, aunque los valores de F2 parece que se encuentran por encima de la estadística. En el caso de L-B (voz dubitada) encontramos esta misma tendencia hacia la parte del espectro más aguda en F2. Cabe destacar también que la dispersión en L-B también parece algo desplazada hacia frecuencias graves en cuanto a F1. No obstante este hallazgo puede ser compatible con la manipulación y procesado de voz, pérdida de velocidad en un grabador de cinta magnética, etc.

En cuanto al histograma de Fo, obtuvimos una clara diferencia si utilizamos los fragmentos completos que tenemos de L-B (más de 1 minuto) en la comparación con L-A (1 minuto), diferencia en cuanto a la Fo más probable y a la función de distribución. La frecuencia Fo para el locutor B (voz dubitada) era de 131,77Hz, y la forma era mucho más

picuda, lo que no era coherente con la percepción subjetiva ya que en ambos casos L-A y L-B varían su entonación en varias ocasiones a lo largo de la grabación. Por esto decidimos utilizar un fragmento de igual duración para obtener valores más ajustados.

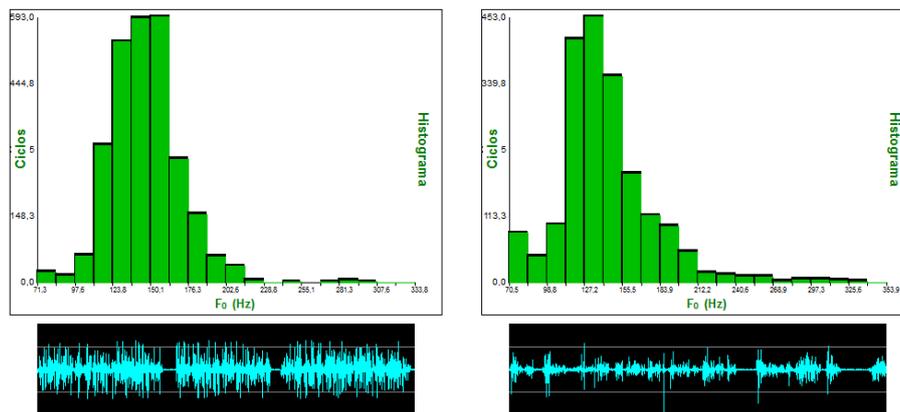


Figura 5-12: Histograma Fo L-A vs L-B

Sin embargo siguen apreciándose diferencias estando claro en ambos casos que la Fo media está próxima a los 150Hz pero habiendo diferencias que si se tratara del mismo hablante podrían achacarse a la diferente entonación en ambas grabaciones.

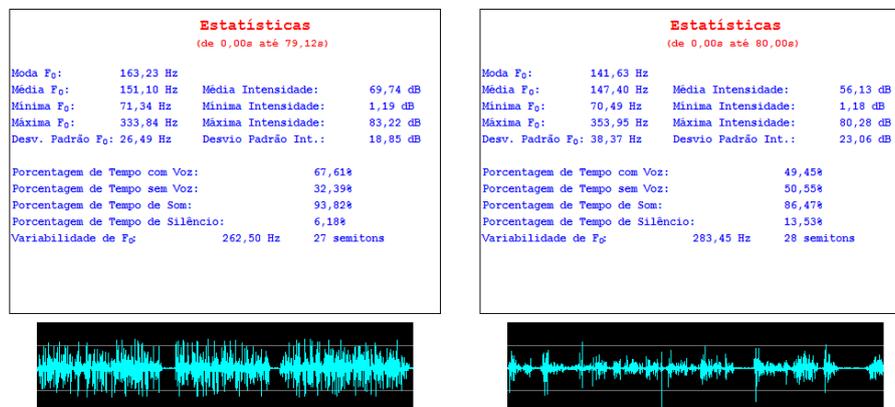


Figura 5-13

Para el análisis de la función de transferencia del tracto vocal elegimos una palabra repetida en ambos casos /bonko malkobík/. El análisis se realiza mediante espectro FFT, LPC de 5 coeficientes y análisis *cepstral*. La caída del filtro LPC es muy similar y las frecuencias centrales de sus formantes varían en unos 50Hz en ambos casos. La representación *cepstral* es también bastante semejante en frecuencias inferiores a 1000Hz, pero en el caso de L-A aparece una nueva formante alrededor de los 2400Hz, la cual es compatible con una equalización de la voz con énfasis en esas frecuencias.

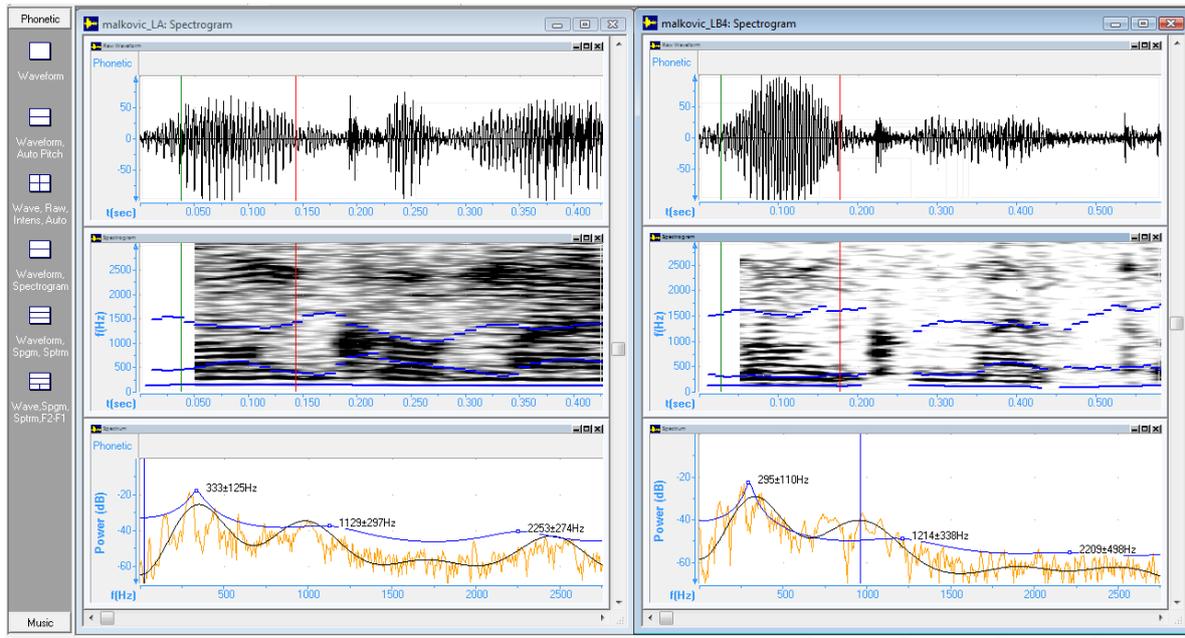
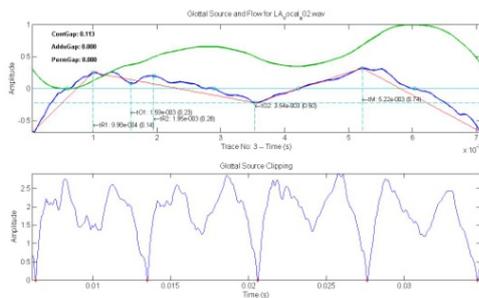


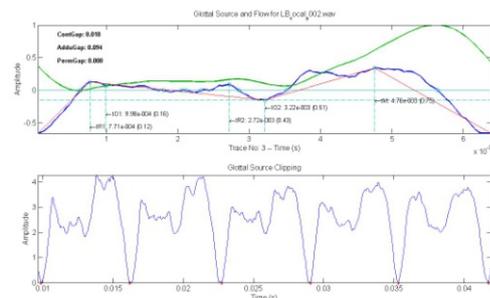
Figura 5-14: Análisis espectrográfico y LPC de L-A frente a L-B ante la palabra

5.5.3. Análisis de la fonación

Los resultados de la comparación del análisis de la fonación de L-A (abogado Marcelo Soza) y L-B (voz sospechosa), fue una tarea difícil principalmente por el ritmo del habla. Obtuvimos varias muestras de vocales que no mostraban una correcta onda glótica en la que no aparecían claramente las características que queremos comparar.

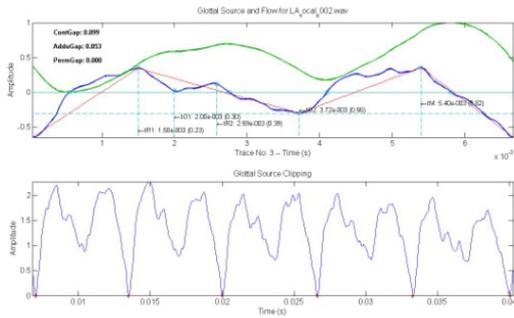


Onda Glótica de L-A vocal /a/

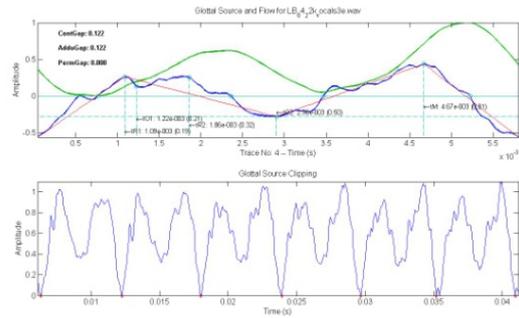


Onda Glótica de L-B vocal /a/

En el caso de la vocal /a/ de L-B (supuestamente el abogado Soza) observamos que la onda sigue teniendo una inflexión en la mitad del ciclo pero que es menor que la que vemos en L-A.



Onda Glótica de L-A vocal /o/



Onda Glótica de L-B vocal /e/

Probando con otras vocales abiertas para hacer la comparación obtenemos nuevas muestras las cuales coinciden y demuestran la existencia de esa inflexión en ambos locutores.

Las características sustraídas del análisis muestran como rasgos distintivos de la fonación de L-A algunos parámetros biomecánicos, parámetros temporales y de ciclo. En el caso de la voz indubitada encontramos los mismos parámetros como distintivos aunque existe una variedad en los datos numéricos de los parámetros que deberán ser estudiados con detenimiento.

| | Nombre del parámetro | | |
|-----------------------|-----------------------------|-----|----|
| Parámetro Biomecánico | Body Stiffness | 3,4 | 6 |
| Parámetro Biomecánico | Body Stiffness Unbalance | <7 | <8 |
| Parámetro temporal | Rel Open 2 Ampl | 3 | -5 |
| Parámetro cíclico | 2nd Order Cycl Coefficients | 3,8 | 3 |

5.6. Análisis automático

La herramienta utilizada para el análisis automático es una parte dedicada a la identificación de locutores que está incluida en el software SIS II.

A continuación se detalla la información extraída del manual del software de STC

5.6.1. Método de estadísticas del tono

Este método usa dieciséis diferentes características del tono fundamental de voz, tales como: el valor medio del tono fundamental, sus valores máximo y mínimo, la mediana, el porcentaje de los intervalos con tono creciente, la dispersión de logaritmo del tono, la asimetría de logaritmo del tono, el exceso de logaritmo del tono y otros parámetros. El valor de la equi-probable tasa de error igual (en inglés (Equal Error Rate – EER) para el método de estadísticas del tono depende de la duración de los fragmentos comparados y puede llegar al valor ~ 18–19%.

No obstante, la dependencia de confiabilidad de este método del estado emocional y psicológico del locutor en el momento de pronunciación permite utilizarlo como un auxiliar solamente.

Este método sigue la misma filosofía que la utilizada en el punto 5.5.2.

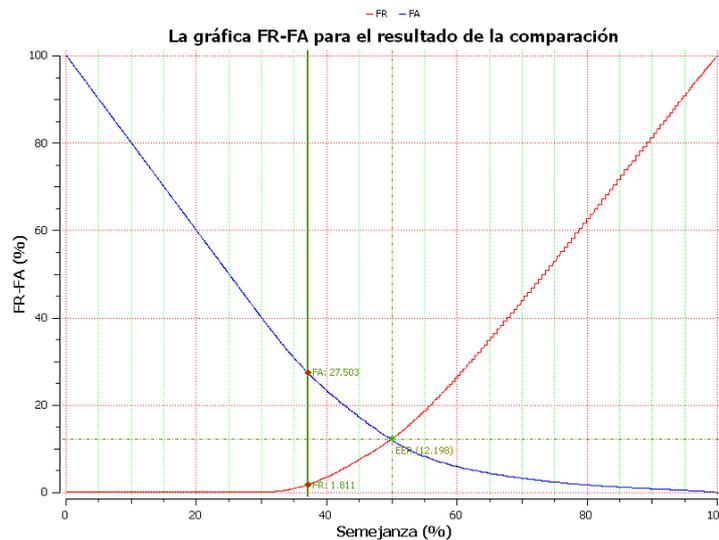


Figura 5-15: Comparación de Tono

5.6.2. Método de espectros-formantes

Este método se basa en la extracción y comparación de posiciones y de dinámica de tres o más formantes (el método está protegido por un patente de invención de Rusia).

Para hacer la comparación se construye el espectro de grabación de habla que se usa para crear la tarjeta de identificación de un locutor. Luego se calcula la función de normalización para esta grabación la que se utilizará más tarde para calcular los espectros instantáneos de la señal de voz. En calidad de los indicios de identificación se usan las posiciones de tres máximos espectrales los más correspondientes a los formantes de la señal de voz en cada corte espectral donde se hayan determinados bastante confiables. Estos indicios definen en grado superlativo algunas particularidades individuales del tracto vocal. La densidad de distribución de los indicios de identificación se modela usando la MDN.

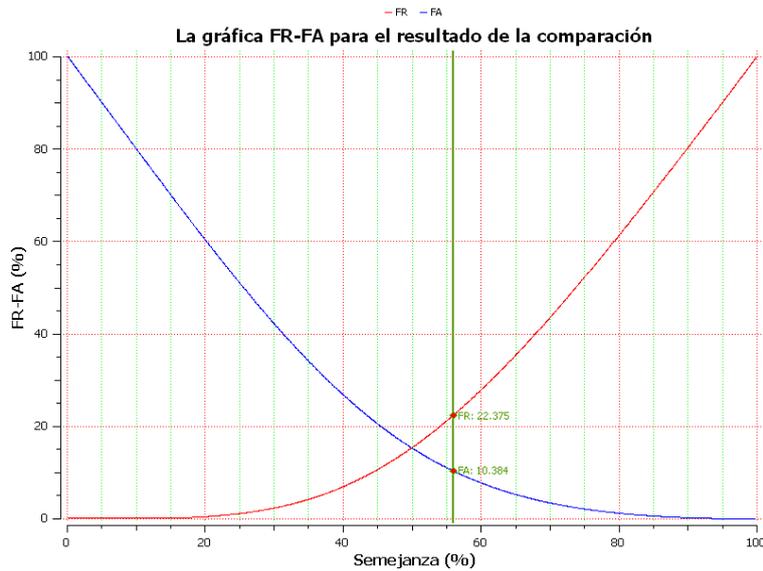


Figura 5-16: Comparación automática de EF

5.6.3. Método de variabilidad total

El método de variabilidad total (o bien TotV del inglés *total variability*) es el método contemporáneo de identificación por voz más rápido.

Se calculan los indicios de habla de MFCC (del inglés Mel Fourier Cepstrum Coefficients – coeficientes de cepstrum por la escala Mel); ellos son las características que describen representación espectral de una señal de voz en momentos de tiempo particulares. La densidad de distribución de los indicios de identificación se modela utilizando la mezcla de distribuciones normales (MDN).

Luego los parámetros del modelo de MDN con ayuda del análisis adaptado de factores se representan en forma del i-vector de dimensión baja en el así llamado espacio de variabilidad total, el que contiene los subespacios de canales propios y de voces propias, los cuales se usan en el JFA (del inglés Joint Factor Analysis – análisis factorial conjunto). La particularidad distintiva de esta representación del modelo de MDN es su claridad alta y tamaño de datos pequeño.

En la etapa de comparación de los i-vectores el clasificador de SVM el cual se considera bueno respecto a la velocidad y calidad de decisión de identificación.

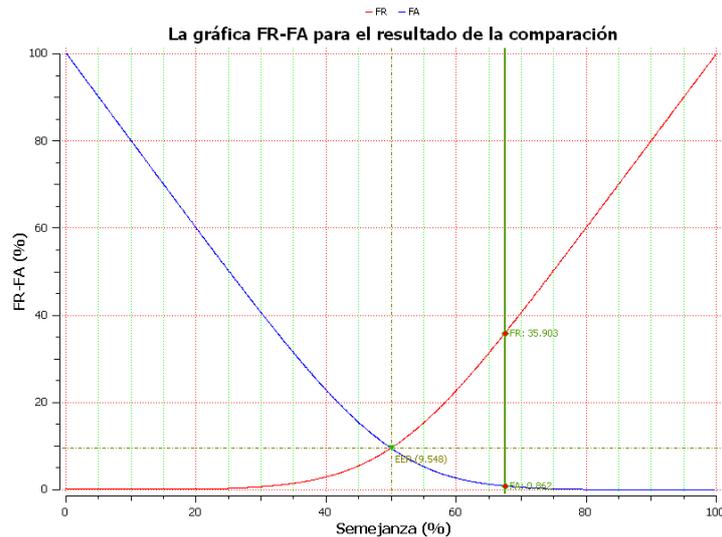


Figura 5-17: comparación mediante método de variabilidad total

5.6.4. Fundamentos teóricos del método de la solución total

El método de la solución total de la identificación de locutores se realiza con uso de la solución total la que se basa en los resultados de identificación realizada mediante un método o varios métodos de identificación independientemente de su cantidad.

El resultado es la pseudoprobabilidad P de pertenencia de las dos grabaciones comparadas a un locutor; este valor está dentro del diapasón desde 0 hasta 100 %; 0 % significa la similitud de los locutores de algunas dos grabaciones más mínima posible y 100 % significa su similitud completa.

El algoritmo de cálculo de la solución total se basa en uso del método de la votación ponderada:

$$P = \sum_{i=1}^N w_i * (FR_i FA_i)$$

donde N es la cantidad de los métodos de identificación usados para obtener la solución; w_i es el coeficiente de ponderación de un i método; su valor se cambia automáticamente dependiendo de calidad de la señal; FR_i/FA_i son los errores de tipos primero y segundo de un i método; los valores de FR_i/FA_i se calculan en porcentaje y están dentro del diapasón desde 0 hasta 100 %.

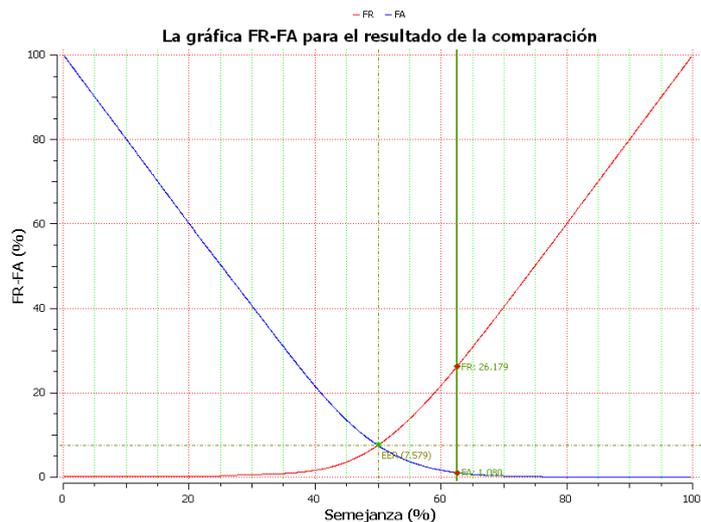


Figura 5-18 Comparación Solución total.

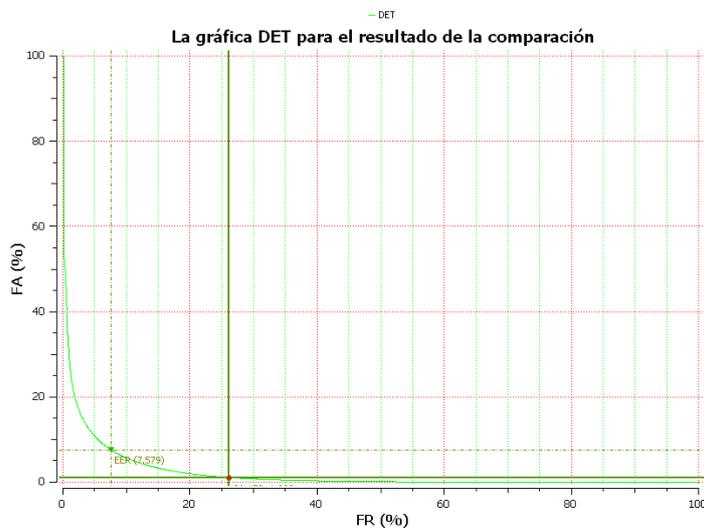


Figura 5-19: Curva DET de la Solución total.

Resultado de la comparación

Archivo 1: LA_22k_preproc.wav

Formato: 16 bit; mono; 22050 Hz; 79.12 seg.; el habla no se segmentó. habla pura: 49.22 seg.;

Archivo 2: LB_04_22k_preproc1min.wav

Formato: 16 bit; mono; 22050 Hz; 80.00 seg.; el habla no se segmentó. habla pura: 35.98 seg.;

| Métodos | FR [min,max], % | FA [min,max], % | LR [min,max] | P [min,max], % |
|--|----------------------|---------------------|--------------------|--------------------|
| <input checked="" type="checkbox"/> EF | 22.38 [18.4, 26.3] | 10.38 [10.16, 10.6] | 2.15 [1.77, 2.5] | 56.0 [54.0, 57.98] |
| <input checked="" type="checkbox"/> Tono | 1.8 [0.5, 3.08] | 27.5 [27.17, 27.8] | 0.07 [0.02, 0.1] | 37.15 [36.5, 37.8] |
| <input checked="" type="checkbox"/> MMG | 35.9 [31.3, 40.47] | 0.86 [0.79, 0.9] | 41.67 [35.5, 47.8] | 67.5 [65.2, 69.8] |
| <input checked="" type="checkbox"/> Solución total | 26.18 [21.99, 30.37] | 1.08 [1.0, 1.16] | 24.2 [20.05, 28.4] | 62.5 [60.46, 64.6] |

Los rasgos de identificación de los locutores coinciden con el nivel de confianza más de 99%.

La probabilidad de coincidencia es más de 60.46%.

Queda también para futuros análisis la posibilidad de realizar una rueda de reconocimiento mediante el software BioMetroSoft@Fore el cuál no hemos podido realizar a fecha de hoy, y que es complementario a la identificación automática pero desde el punto de vista de la ónda glótica del locutor.

6.CONCLUSIONES DEL DICTAMEN

1. Determinar si la voz cuestionada que aparece en los cuatro archivos de audio aportados (supuesta voz del abogado Sr. Marcelo Soza), se corresponde con la muestra indubitada (muestra GIGA.SOSA.21.12.12 ENTREVISTA.LA PAZ.mp3),procedente de la entrevista realizada al Sr. Marcelo Soza en la Televisión de Bolivia.

A la vista de los resultados obtenidos en todos los análisis efectuados, podemos decir que **hay suficientes indicios como para considerar que L-A (voz indubitada del abogado Marcelo Soza) y L-B (voz cuestionada, atribuida a Marcelo Soza), corresponden al mismo locutor.**

2. Determinar si las grabaciones han sido manipuladas, editadas o si por el contrario mantienen una continuidad en el tiempo y el espacio.

La cuatro muestras de audio analizadas (AUDIO 1 para identificación vocal.mp3 , AUDIO 2 para identificación vocal.mp3, AUDIO 3 para identificación vocal.mp3, AUDIO 4 para identificación vocal.mp3), con una duración aproximada de 24 minutos, presentan un total de 51 eventos electroacústicos de corta duración,que no son lógicos en el entorno en el que se desarrollan. Circunstancias que **no deben asociarse** a parada o mal funcionamiento del dispositivo grabador, sino más bien **a la manipulación intencionada con posterioridad al proceso de grabación sonora.**

Sin embargo todos los archivos **mantienen continuidad espacio temporal, y una secuencia lógica tanto en el plano semántico como expresivo.**

En todas las grabaciones analizadas, la voz del locutor 1 (correspondientes al supuesto Marcelo Soza) **mantiene las mismas características acústicas, misma posición y ubicación con respecto al dispositivo grabador.**

3. Determinar si las muestras de audio analizadas corresponden a fragmentos de una misma grabación, o por el contrario no existen semejanzas o elementos de continuidad entre ellas.

Todos los archivos de audio analizados **comparten características acústicas y aportan información con rasgos de continuidad entre sí, por lo que podemos afirmar que las cuatro muestras analizadas corresponden a fragmentos de la misma conversación.**

También podemos afirmar que **todas las muestras de audio analizadas fueron registradas con el mismo dispositivo grabador.**

Firmado en Madrid, a 29 de Marzo del 2013



Fdo.: Miguel Ángel de la Torre Guijarro



DECLARACIÓN DEL PERITO

D. Miguel Ángel de la Torre Guijarro, con DNI: 01833260E, especialista de audio y Director Gerente de la empresa española Graudio Servicios integrales de sonido y formación (a partir de ahora **Graudio Forensics**).

Declaro:

- Que juro actuar con la mayor objetividad posible, tomando en consideración tanto lo que pueda favorecer como lo que sea susceptible de causar perjuicio a cualquiera de las partes.
- Que estoy informado del contenido de los artículos 458 y 459 del Código Penal Español .
- Que he cumplido los siguientes deberes específicos de toda actuación pericial:
 - . Actuar con la debida diligencia para no demorar ni entorpecer la actuación de la Administración Pública.
 - . Examinar personal y directamente el objeto de la pericia.
 - . Actuar con veracidad,objetividad,imparcialidad e independencia de criterio basando el dictamen en hechos comprobables y criterios explícitos y limitando el contenido del dictamen a lo que se ha requerido.
 - . Guardar la discreción exigible con arreglo al deber del secreto profesional respecto a las cuestiones objeto de la pericia y a los hechos que conozca con motivo del desempeño de mi trabajo.

Firmado en Madrid, a 29 de Marzo del 2013



Fdo: Miguel Ángel de la Torre Guijarro



Breve Curriculum Vitae del Perito

Miguel Ángel de la Torre Guijarro

Formación Académica

- . Licenciado en Comunicación Audiovisual, rama imagen y sonido. Universidad Complutense de Madrid
- . Master en Dirección de Seguridad ICADE
- . Programa de Dirección General (PDG) en la Escuela de Negocios IESE de Madrid
- . Miembro activo de la *Audio Engineering Society* (AES)

Trayectoria Profesional

- . Técnico de sonido en el departamento de Audio frecuencia de Antena 3 TV (1995_2.010)
- . Director gerente de la empresa Graudio Forensics, primer laboratorio de Audio Forense en España
- . Actualmente forma parte del equipo de *Grupo pericial*, la primera y mayor firma de Peritos Judiciales de toda España
- . Profesor de Audio Forense en la escuela de Derecho y Criminología de la Universidad CEU San Pablo de Madrid
- . Organización de innumerables cursos y programas de capacitación de audio profesional para empresas y profesionales.
Colaboración con la Universidad Politécnica y Universidad CEU San Pablo, de Madrid .España

Bibliografía

Koenig B.E, 1990 . Authentication of Forensic Audio Recording , Journal of AES, vol.38,nº 12

Owen T. 1989. An introduction to Forensic Examination of audio and video tapes, Rodgers and Hammerstein Archives of Record Sound Public Library

Chen, C., Bilmes, J., & Ellis, D. P. (2005). Speech feature smoothing for robust ASR. 2005 IEEE International Conference on Acoustics, Speech, and Signal Processing: Proceedings: March 18-23, 2005, Pennsylvania Convention Center/Marriott Hotel, Philadelphia, Pennsylvania, USA, pp. 525-528.

Cordón, A. M. (1980). Notas sobre la fonética del castellano en bolivia. Instituto Cervantes.

Gómez Vilda, P., Álvarez Marquina, A., Mazaira Fernández, L. M., Fernández-Baillo Gallego De La Sacristana, Roberto, Nieto Lluís, V., Martínez Olalla, R., et al. (2008). Decoupling vocal tract from glottal source estimates in speaker's identification. Language Design, (Special Issue), 111-118.

Nolan, F. (2001). Speaker identification evidence: Its forms, limitations, and roles. Proceedings of the conference 'Law and Language: Prospect and Retrospect', Levi Finland,

Quilis, A. (1997). Principios de fonología y fonética española Arco libros.

