

**Universidad Autónoma Metropolitana Unidad Azcapotzalco  
División de Ciencias Básicas e Ingeniería  
Licenciatura en Ingeniería en Computación**

**Software inteligente capaz de diferenciar entre una situación normal y una de violencia por medio de la voz.**

**Guillermo Domínguez Maldonado**

**Asesora: Dra. Silvia Beatriz González Brambila  
Profesora Titular  
Departamento de Sistemas**

**Coasesor: Dr. Juan Gaspar Vargas Rubio  
Profesor Titular  
Departamento de Electrónica**

**Trimestre 2014 Otoño  
16 de Diciembre de 2014**

Yo, Silvia Beatriz González Brambila, declaro que aprobé el contenido del presente Reporte de Proyecto de Integración y doy mi autorización para su publicación en la Biblioteca Digital, así como en el Repositorio Institucional de UAM Azcapotzalco.

A handwritten signature in black ink, reading "Silvia Beatriz González Brambila", written over a horizontal line.

Yo, Juan Gaspar Vargas Rubio, declaro que aprobé el contenido del presente Reporte de Proyecto de Integración y doy mi autorización para su publicación en la Biblioteca Digital, así como en el Repositorio Institucional de UAM Azcapotzalco.

A handwritten signature in black ink, appearing to be "Juan Gaspar Vargas Rubio", written over a horizontal line.

Yo, Guillermo Domínguez Maldonado, doy mi autorización a la Coordinación de Servicios de Información de la Universidad Autónoma Metropolitana, Unidad Azcapotzalco, para publicar el presente documento en la Biblioteca Digital, así como en el Repositorio Institucional de UAM Azcapotzalco.

A handwritten signature in black ink, appearing to be "Guillermo Domínguez Maldonado", written over a horizontal line.

## Resumen

El presente proyecto de investigación muestra el diseño y la implementación de un software inteligente, capaz de diferenciar entre una situación normal y una de violencia por medio de la voz, a través de los parámetros: frecuencia, amplitud, tono e intensidad.

El proyecto surgió por el interés de generar un medio inteligente, capaz de identificar y dar a conocer cuándo se está manifestando una situación de violencia en espacios públicos y actuar para prevenir un riesgo mayor a las personas que está siendo víctimas de una situación de violencia verbal.

La primera actividad que se realizó fue la búsqueda de videos que presentaran situaciones normales por ejemplo: charlas, niños jugando, reuniones familiares, encuentros casuales, etc.

De igual forma, se buscaron videos donde ocurrían situaciones violentas, por ejemplo: personas discutiendo agredándose física y verbalmente, utilizando groserías, etc. La duración de los videos es distinta, de manera que el video más corto es de 30 segundos y el más largo tiene una duración de dos minutos. El número total de videos recopilados fue de 100, con una duración en tiempo promedio de un minuto.

De los 100 videos recopilados, se realizó la clasificación manual de ellos; es decir, separar los videos en los que ocurrían situaciones de violencia de los que presentaban situaciones normales. La clasificación arrojó 50 videos con situaciones normales y 50 con situaciones de violencia.

De cada video, mediante el uso del software denominado aTube Catcher se realizó la extracción del audio para obtener un archivo en formato .wav. De los archivos en formato .wav se obtuvo el parámetro de intensidad, tono, amplitud y frecuencia, para ello se hizo uso del software analizador de voz denominado Praat.

Los parámetros de intensidad, tono, amplitud y frecuencia propios de cada situación analizada (normal y de violencia) fueron almacenados en dos carpetas distintas para su análisis.

La siguiente parte del proyecto involucró la utilización del algoritmo DTW (*Dynamic Time Warping*), para realizar la comparación entre dos señales, la señal de referencia y la señal de

prueba. Estas señales son una representación gráfica del parámetro de intensidad, tono, amplitud y frecuencia.

Todas las señales de cada parámetro y cada situación fueron comparadas entre sí, teniendo una señal de referencia y otra de prueba para mostrar la región en la que ambas señales eran semejantes. El objetivo primordial de realizar estas comparaciones fue encontrar la señal de referencia de cada parámetro (intensidad, tono, amplitud y frecuencia) del conjunto de datos en estudio para situaciones normales y situaciones de violencia.

La señal de referencia fue aquella que al realizar las comparaciones mediante el algoritmo DTW, generó una mayor región de alineación; es decir, una mínima distancia entre las señales comparadas, indicando un alto grado de similitud.

Como resultado de estas comparaciones se obtuvo que la voz es un sonido tan complejo, que no es posible definir una situación de violencia con solo una de sus características o parámetros, razón por la cual se requiere de la combinación de más de una de ellas, por lo tanto y para nuestro caso de estudio se concluye que en el diseño e implementación del software inteligente se haga uso de la combinación de la intensidad, tono, y amplitud de la voz para identificar una situación de violencia. Se descarta el uso de la frecuencia de la voz como un indicativo de violencia porque la voz modula a diferentes frecuencias y depende de la edad y el sexo de la persona lo que dificultaría la interpretación del tipo de situación que se esté manifestando.

La proyección a futuro de esta propuesta puede estar encaminada a instalar el software en micrófonos ambientales ubicados en lugares públicos a fin de monitorear el comportamiento de la población e identificar de manera rápida una situación de violencia.

De acuerdo a los resultados obtenidos como parte de la investigación desarrollada, al medir el nivel de efectividad del software propuesto para detectar una situación de violencia a través de los parámetros de intensidad, tono, amplitud y frecuencia de la voz en una situación se identificó que tiene una eficiencia del 87%, determinado a través de una prueba experimental, por comparación entre el índice de concordancia de la señal de referencia para cada una de las características de la voz y el índice de concordancia que se obtuvo del archivo experimental.

# Índice

1. Introducción.....	8
2. Justificación.....	9
3. Objetivos.....	9
4. Objetivos específicos.....	9
5. Marco Teórico.....	10
5.1 Definición de violencia.....	10
5.2 Voz.....	12
5.2.1 Intensidad.....	12
5.2.2 Tono.....	13
5.2.3 Amplitud.....	13
5.2.4 Frecuencia.....	14
5.3 Algoritmo DTW.....	15
5.4 Software utilizado en el proyecto.....	16
5.4.1 aTube Catcher.....	16
5.4.2 Praat.....	17
5.4.3 Paquete R.....	18
5.4.4 Características de R.....	18
5.4.5 Extensiones y paquetes.....	18
5.4.6 Paquete usado en el proyecto.....	18
6. Desarrollo del proyecto.....	19
6.1 Búsqueda y clasificación de videos.....	19
6.2 Selección de de archivos para su análisis con Praat.....	20
6.2.1 Obtención del parámetro intensidad.....	22
6.2.2 Obtención del parámetro tono.....	23
6.2.3 Obtención del parámetro amplitud.....	23
6.2.4 Obtención del parámetro frecuencia.....	24
6.3 Aplicación del algoritmo DTW en el parámetro intensidad.....	24
6.3.1 Situación violenta.....	25
6.3.2 Situación normal.....	28
6.4 Aplicación del algoritmo DTW en el parámetro tono.....	31
6.4.1 Situación violenta.....	31
6.4.2 Situación normal.....	34
6.5 Aplicación del algoritmo DTW en el parámetro amplitud.....	36

<b>6.5.1</b>	Situación violenta.....	36
<b>6.5.2</b>	Situación normal.....	40
<b>6.6</b>	Aplicación del algoritmo DTW en el parámetro frecuencia.....	43
<b>6.6.1</b>	Situación violenta.....	43
<b>6.6.2</b>	Situación normal.....	46
<b>7.</b>	Resultados.....	49
<b>7.1</b>	Resultados obtenidos del parámetro intensidad, situación violenta.....	49
<b>7.2</b>	Resultados obtenidos del parámetro intensidad, situación normal.....	50
<b>7.3</b>	Resultados obtenidos del parámetro tono, situación violenta.....	50
<b>7.4</b>	Resultados obtenidos del parámetro tono, situación normal.....	51
<b>7.5</b>	Resultados obtenidos del parámetro amplitud, situación violenta.....	51
<b>7.6</b>	Resultados obtenidos del parámetro amplitud, situación normal.....	52
<b>7.7</b>	Resultados obtenidos del parámetro frecuencia, situación violenta.....	52
<b>7.8</b>	Resultados obtenidos del parámetro frecuencia, situación normal.....	53
<b>7.9</b>	Análisis de resultados.....	54
<b>8.</b>	Conclusiones.....	55
<b>9</b>	Referencias bibliográficas.....	56
<b>10.</b>	Apéndice A, código del proyecto de investigación.....	59

## Índice de Figuras.

<b>Figura 1.</b> Amplitud de una onda.....	14
<b>Figura 2.</b> Frecuencia de un sonido grave y uno agudo.....	14
<b>Figura 3.</b> Resultado de la comparación de dos funciones aplicando el Algoritmo DTW.....	15
<b>Figura 4.</b> Dos series de tiempo.....	15
<b>Figura 5.</b> Aplicación del algoritmo DTW a dos series similares.....	16
<b>Figura 6.</b> Similitud entre las dos series de tiempo.....	16
<b>Figura 7.</b> Pantalla de inicio de aTube Catcher.....	17
<b>Figura 8.</b> Pantalla de inicio de PRAAT.....	17
<b>Figura 9.</b> Conversión de videos a formato .WAV.....	20
<b>Figura 10.</b> Menú Read de Praat.....	20
<b>Figura 11.</b> Elección del archivo de audio de Praat.....	21
<b>Figura 12.</b> Oscilograma y Espectrograma del archivo de audio de Praat.....	21
<b>Figura 13.</b> Valores obtenidos con Praat.....	22
<b>Figura 14.</b> Obtención de los valores de intensidad del archivo de audio con Praat..	22
<b>Figura 15.</b> Obtención de los valores de tono del archivo de audio con Praat.....	23
<b>Figura 16.</b> Obtención de los valores de amplitud del archivo de audio con Praat...	23
<b>Figura 17.</b> Graficas del parámetro intensidad en situaciones violentas con Praat....	24
<b>Figura 18.</b> Gráfica y grado de concordancia del parámetro intensidad en situaciones violentas (Peor caso).....	26
<b>Figura 19.</b> Obtención de la señal de referencia ejemplo 1.....	27
<b>Figura 20.</b> Obtención de la señal de referencia ejemplo 2.....	27
<b>Figura 21.</b> Obtención de la señal de referencia ejemplo 3.....	28
<b>Figura 22.</b> Gráfica y grado de concordancia del parámetro intensidad en situaciones normales (Peor caso).....	29
<b>Figura 23</b> Señal de referencia situaciones normales ejemplo 1.....	29
<b>Figura 24.</b> Señal de referencia situaciones normales ejemplo 2.....	30
<b>Figura 25.</b> Señal de referencia situaciones normales ejemplo 3.....	30
<b>Figura 26.</b> Ejemplos de graficas del parámetro Tono (violencia).....	31
<b>Figura 27.</b> Gráfica y grado de concordancia del tono en situaciones violentas (Peor caso)..	32
<b>Figura 28.</b> Obtención de la señal de referencia del parámetro tono ejemplo 1.....	33
<b>Figura 29.</b> Obtención de la señal de referencia del parámetro tono ejemplo 2.....	33
<b>Figura 30.</b> Obtención de la señal de referencia del parámetro tono, ejemplo 3.....	34
<b>Figura 31.</b> Gráfica y grado de concordancia del tono en situaciones normales (Peor caso).	35

<b>Figura 32.</b> Ejemplo 1 de graficas del parámetro tono (situación normal).....	35
<b>Figura 33.</b> Ejemplo 2 de graficas del parámetro tono (situación normal).....	36
<b>Figura 34.</b> Ejemplos de graficas del parámetro Amplitud (violencia).....	36
<b>Figura 35.</b> Gráfica y grado de concordancia de la amplitud en situaciones violentas (Peor caso).....	37
<b>Figura 36.</b> Obtención de la señal de referencia del parámetro amplitud ejemplo 1.....	38
<b>Figura 37.</b> Obtención de la señal de referencia del parámetro amplitud ejemplo 2.....	39
<b>Figura 38.</b> Obtención de la señal de referencia del parámetro amplitud ejemplo 3.....	39
<b>Figura 39.</b> Ejemplos de graficas del parámetro amplitud (no violencia).....	40
<b>Figura 40.</b> Gráfica y grado de concordancia de la amplitud en situaciones normales (Peor caso).....	41
<b>Figura 41.</b> Obtención de la señal de referencia del parámetro amplitud mejor caso.....	42
<b>Figura 42.</b> Obtención de la señal de referencia del parámetro amplitud ejemplo 2.....	42
<b>Figura 43.</b> Obtención de la señal de referencia del parámetro amplitud ejemplo 3.....	43
<b>Figura 44.</b> Ejemplos de graficas del parámetro Frecuencia (violencia).....	43
<b>Figura 45.</b> Gráfica y grado de concordancia de la frecuencia en situaciones violentas (Peor caso).....	44
<b>Figura 46.</b> Obtención de la señal de referencia del parámetro frecuencia ejemplo 1.....	45
<b>Figura 47.</b> Obtención de la señal de referencia del parámetro frecuencia ejemplo 2.....	45
<b>Figura 48.</b> Obtención de la señal de referencia del parámetro frecuencia ejemplo 3.....	46
<b>Figura 49.</b> Gráfica y grado de concordancia de la frecuencia en situaciones normales (Peor caso).....	47
<b>Figura 50.</b> Señal de referencia del parámetro frecuencia (no violencia) ejemplo 1.....	47
<b>Figura 51.</b> Señal de referencia del parámetro frecuencia (no violencia) ejemplo 2.....	48
<b>Figura 52.</b> Señal de referencia del parámetro frecuencia (no violencia) ejemplo 3.....	48
<b>Figura 53.</b> Señal de referencia de una situación de violencia, parámetro intensidad.....	50
<b>Figura 54.</b> Señal de referencia de una situación normal, parámetro intensidad.....	50
<b>Figura 55.</b> Señal de referencia de una situación violenta, parámetro tono.....	51
<b>Figura 56.</b> Señal de referencia de una situación normal, parámetro tono.....	51
<b>Figura 57.</b> Señal de referencia de una situación violenta, parámetro amplitud.....	52
<b>Figura 58.</b> Señal de referencia de una situación normal, parámetro amplitud.....	52
<b>Figura 59.</b> Grafica de referencia de una situación violenta, parámetro frecuencia.....	53
<b>Figura 60.</b> Grafica de referencia de una situación normal, parámetro frecuencia.....	54

# 1. Introducción

El proyecto surgió por el interés de generar un medio inteligente, capaz de identificar y dar a conocer cuándo se está manifestando una situación de violencia en espacios públicos y actuar para prevenir un riesgo mayor a las personas que están siendo víctimas de una situación de violencia verbal.

Para dar atención a un acto violento debemos partir del hecho de entender que implica el término violencia.

La violencia se puede definir como “el uso intencional de la fuerza o el poder físico, de hecho o como amenaza, contra uno mismo, contra otra persona, un grupo o una comunidad, que cause o tenga muchas probabilidades de causar lesiones, muerte, daños psicológicos, trastornos del desarrollo o privaciones” [1].

Se reconocen varios tipos de violencia: la violencia física, verbal, psicológica, sexual, etc. Sus causas pueden variar en dependencia del contexto sociocultural que se tenga [1]. El proyecto muestra el diseño y la implementación de un software inteligente, capaz de diferenciar entre una situación normal y una de violencia por medio de la voz, a través de los parámetros de frecuencia, amplitud, tono e intensidad.

Para la obtención de los parámetros antes mencionados, se realizó una búsqueda y clasificación manual de videos que presentan situaciones violentas y situaciones normales. De este conjunto de videos se realiza la extracción del audio. Con ayuda de un software analizador de voz se obtuvieron las señales características o de referencia de los parámetros de frecuencia, amplitud, tono e intensidad de cada una de las situaciones analizadas.

Al definir los valores característicos que manifiesta la voz en los parámetros de frecuencia, amplitud, tono e intensidad, en una situación normal y en una situación violenta, inmediatamente al detectarse la combinación de estos valores se pudo identificar en qué momento un ser humano se encuentra expuesto ante una situación de violencia y poder actuar para evitar un riesgo mayor de exposición.

El presente proyecto de investigación se entrega en formato PDF en un CD, así como los archivos que contienen cada uno de los parámetros extraídos con el software Praat que representan situaciones violentas y no violentas.

## **2. Justificación**

Una de las necesidades más apremiantes en las instituciones de seguridad pública es resolver el problema de la violencia que vive la sociedad. Hoy día la violencia es común vivirla de distintas maneras (violencia física, verbal, psicológica, sexual, etc.), sin embargo, no existe un mecanismo capaz de identificar, medir e indicar en qué momento se está manifestando un suceso violento de cualquier tipo. Partiendo de esta problemática y en el entendido que la única violencia medible es la violencia física, surge el interés por diseñar e implementar un software con la capacidad de identificar el momento en el cual un ser humano está expuesto a una situación de violencia. La identificación, medición y alerta de una situación de violencia no física se realiza mediante comparaciones de la frecuencia, amplitud, tono e intensidad de la voz.

El proyecto realizado tiene la intención de identificar de manera rápida el momento en cual se manifiesta un hecho violento y su aplicación tendrá como beneficio identificar un acto violento que se presente en la sociedad para que en caso de ser necesario intervengan las fuerzas del orden en su atención.

## **3. Objetivo general**

Diseñar e implementar un Software inteligente con la capacidad de diferenciar entre una situación normal y una de violencia por medio de los parámetros frecuencia, amplitud, tono e intensidad de la voz.

## **4. Objetivos específicos**

- Obtener videos que contengan situaciones violentas y situaciones normales.
- Obtener archivos de audio en formato .wav con situaciones violentas y situaciones normales.
- Identificar la señal de referencia característica del parámetro frecuencia, amplitud, tono e intensidad de la voz en una situación de violencia y en una situación normal.

## 5. Marco Teórico

Este capítulo contiene los conceptos que permitieron fundamentar con bases teóricas el desarrollo del proyecto de investigación que se realizó.

Se cita una serie de definiciones de violencia, con el objetivo de encontrar las semejanzas y diferencias entre sí, y poder concluir con una definición particular de violencia que permita comprender en qué momento ocurre una situación violenta.

También se definen los conceptos de voz, intensidad, tono, amplitud y frecuencia, herramientas fundamentales para el desarrollo de la investigación, así como el algoritmo DTW.

### 5.1 Definiciones de violencia

Es muy común que escuchemos la palabra violencia pero, ¿cuál es el significado de esta palabra? A continuación se menciona una serie de definiciones de la palabra violencia con la finalidad de comprender ampliamente el significado de esta, y cuyas definiciones permitan establecer un panorama acerca del proyecto de investigación que involucra el reconocimiento de situaciones violentas y no violentas.

#### Conceptos de violencia

- “La violencia (del Lat. violentia), es un comportamiento deliberado, que provoca, o puede provocar, daños físicos o psíquicos a otros seres, y se asocia, aunque no necesariamente, con la agresión, ya que también puede ser psíquica o emocional, a través de amenazas u ofensas” [2].
- “La violencia es la cualidad o la acción y efecto de violentar o violentarse. Lo violento, por su parte, es aquello que está fuera de su natural estado, situación o modo; que se ejecuta con fuerza, ímpetu o brusquedad; o que se hace contra el gusto o la voluntad de uno mismo. La violencia, por lo tanto, es un comportamiento deliberado que puede provocar daños físicos o psíquicos al prójimo” [3].
- “La única violencia medible es la violencia física, la cual se define como el ataque directo, corporal contra las personas. Ella reviste un triple carácter: brutal, exterior y

doloroso. Lo que la define es el uso material de la fuerza, la rudeza, el uso de malas palabras, variaciones en parámetros de la voz que hacen distinguible este acto” [4].

- “Violencia, cualquier conducta física o verbal, activa o pasiva, realizada por una persona contra otra, cuya finalidad es producir en las víctimas intimidación, desvalorización, sentimientos de culpa o sufrimiento, para ganar poder o para mantener el abuso de poder, control y autoridad sobre esa persona” [5].
- “Se entiende por violencia a aquellos actos que tengan que ver con el ejercicio de una fuerza verbal o física sobre otra persona, que tenga por resultado la generación de un daño sobre esa persona de manera voluntaria o accidental. La violencia es uno de los actos más comunes del ser humano y es difícil determinar a ciencia cierta si el ser humano es capaz de vivir en sociedad sin ejercer ningún tipo de violencia” [6].

En un análisis comparativo de las definiciones antes citadas respecto al significado de violencia se concluye en las similitudes y diferencias existentes entre ellas.

Similitud	Diferencia
<ul style="list-style-type: none"> <li>✓ Comportamiento deliberado, mismo que se ejecuta con fuerza, ímpetu o brusquedad, aunque también existe la posibilidad que este sea de forma psicológica.</li> <li>✓ El objetivo primordial de la violencia es causar daño ya sea físico o psicológico a otra persona.</li> </ul>	<ul style="list-style-type: none"> <li>✓ La violencia que puede ser medible es la violencia física, por las características que la involucran como malas palabras y que implican un incremento de la voz.</li> </ul>

En conclusión, definimos a la violencia como todo acto que ocasiona daño físico o psicológico por parte de un individuo u hacia otros, donde las características más visibles que pueden asegurar la existencia de violencia son: golpes, insultos y aumento en el volumen de la voz y la velocidad de esta, que se pueden ver reflejados en los parámetros intensidad, tono, amplitud y frecuencia.

Asimismo la apreciación de violencia es subjetiva, es decir algo que es violento para una persona no necesariamente lo es para otra.

## 5.2 Voz.

La voz es el sonido que se produce por la vibración de las cuerdas vocales mediante el aire que es expulsado por los pulmones y sale a través de la laringe. Este término también se usa para hacer mención a la potencia, el timbre y otras propiedades de la voz [7].

Es importante resaltar que la voz es el instrumento de comunicación por excelencia del ser humano, aporta al lenguaje aspectos que permiten percibir sentimientos, emociones, por lo tanto, la voz puede considerarse como una herramienta que transmite información entre las personas.

### Características de la voz:

La voz tiene la característica de intensidad, tono, amplitud y frecuencia.

A continuación se definen los parámetros de la voz que se utilizan en el desarrollo de este proyecto, la elección de estos parámetros radica en la facilidad de su medición y análisis.

### 5.2.1 Intensidad.

La intensidad se define como el grado de fuerza con que un sonido llega a nuestro oído. La intensidad depende de la amplitud de la vibración. Es la potencia acústica (volumen) de un sonido, que se mide en decibeles (*dB*). De esta manera se tienen sonidos fuertes, medios o débiles, los cuales se pueden apreciar de forma subjetiva [8]. A continuación en la tabla 1 se presenta la gama de niveles de intensidad para el ser humano.

**Tabla 1. Gama de niveles de intensidad**

Emisión	Nivel Sonoro (dB)
Nivel mínimo de la voz humana	20
Mujer conversando en voz baja	25
Hombre conversando en voz baja	30
Mujer conversando en voz normal	50
Hombre conversando en voz normal	55

Mujer hablando en público	60
Emisión	Nivel Sonoro (dB)
Hombre hablando en público	65
Mujer hablando esforzándose	70
Hombre hablando esforzándose	75
Mujer cantando	80
Hombre cantando	85
Nivel máximo de la voz humana	90

### 5.2.2 Tono

El tono es producido por la frecuencia de vibración a la que se manifiesta una determinada onda sonora. En el caso de la voz, viene dada por la cantidad de movimiento que se produce en las cuerdas vocales al emitirla, es decir, por el número de vibraciones por unidad de tiempo que en ellas tienen lugar. Cuantas más vibraciones se produzcan por segundo, más aguda será la voz por lo que más alto será su tono. Por el contrario, cuantas menos vibraciones acontezcan por unidad de tiempo, más grave será el sonido resultante y más bajo será su tono.

La voz masculina tiene un tono fundamental de entre 100 y 200 Hz, mientras que la voz femenina es más aguda, típicamente está entre 150 y 300 Hz. Las voces infantiles son aún más agudas.

La unidad de medida del tono es el Hertz (*Hz*) [9].

### 5.2.3 Amplitud

La amplitud de una onda sonora es una medida de la variación máxima del desplazamiento que varía periódicamente en el tiempo, por lo que se puede decir también que es la distancia máxima entre el punto más alejado de una onda y el punto de equilibrio o medio [10], figura 1.

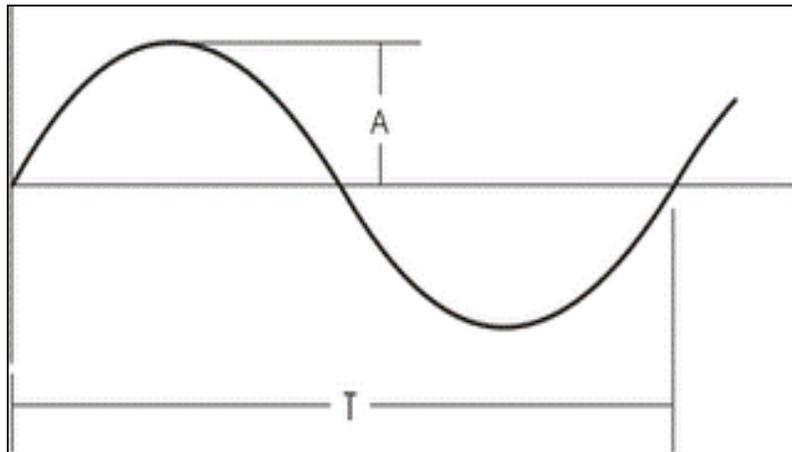


Figura 1. Amplitud de una onda.

### 5.2.4 Frecuencia

Es el número de ondas producidas por segundo. La frecuencia se indica con la letra  $f$  minúscula. Se mide en ciclos/segundo o hertz ( $Hz$ ). Coincide con el número de oscilaciones por segundo que realiza un punto al ser alcanzado por las ondas. A mayor frecuencia se tendrá un sonido más agudo en comparación, si se tiene menor frecuencia se tendrá un sonido más grave [11], ver figura 2.

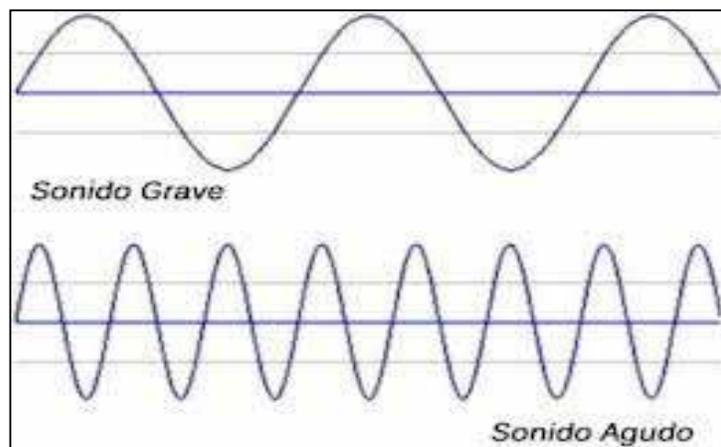


Figura 2. Frecuencia de un sonido grave y uno agudo.

### 5.3 Algoritmo DTW

El algoritmo DTW (*Dynamic Time Warping*), es un algoritmo utilizado para alinear dos series de tiempo similares, pero no sincronizadas y generalmente de diferente duración, de tal manera que la distancia entre las dos series sea mínima, esta diferencia es utilizada para caracterizar la disimilitud de dos series [12]. (Ver figura 3).

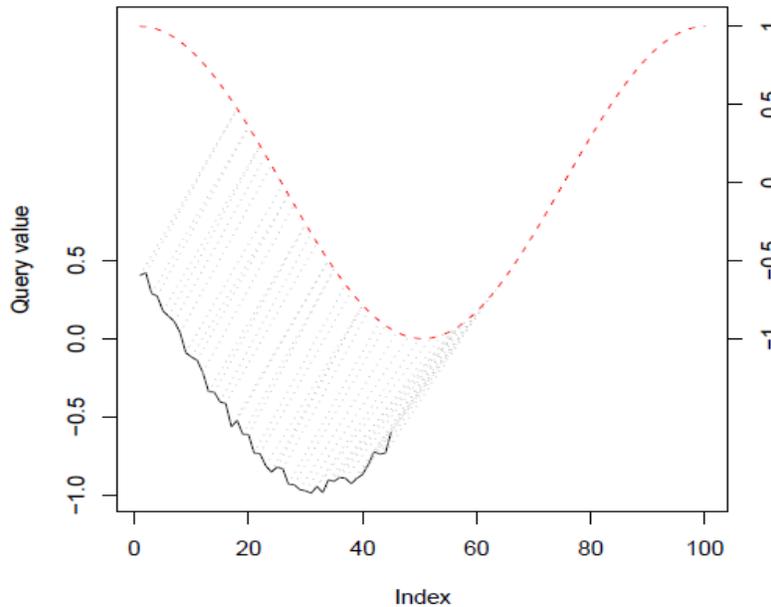


Figura 3. Resultado de la comparación de dos funciones aplicando el Algoritmo DTW.

Si la distancia es mayor que la de un umbral predefinido, las series no son reconocidas como similares. DTW ha sido utilizado durante muchos años en varios dominios de aplicación, incluyendo reconocimiento del habla, minería de datos, etc.

Para tener una mayor comprensión del funcionamiento del algoritmo DTW se presenta un ejemplo: Supongamos que tenemos las dos series de tiempo mostradas en la figura 4.

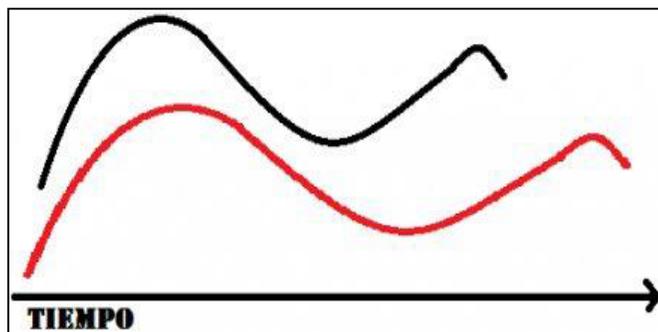


Figura 4. Dos series de tiempo.

La señal en color rojo tiene una mayor duración en el tiempo que la señal de color negro. Al aplicar DTW se puede estirar o encoger ambas señales para tratar de hacerlas coincidir lo mejor posible, de esta manera, si presentan el mismo patrón ambas señales, el algoritmo logrará reconocerlas con mayor exactitud e indicará en qué región existe similitud entre ambas señales, figura 5.

En la figura 6 vemos la similitud entre las dos series de tiempo que se compararon.

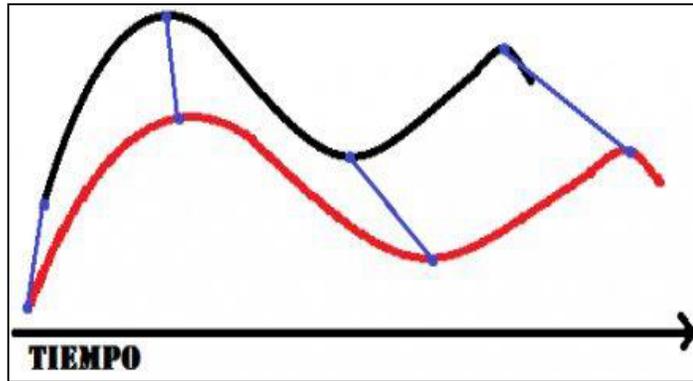


Figura 5. Aplicación del algoritmo DTW a las dos series mostradas en la figura 4.

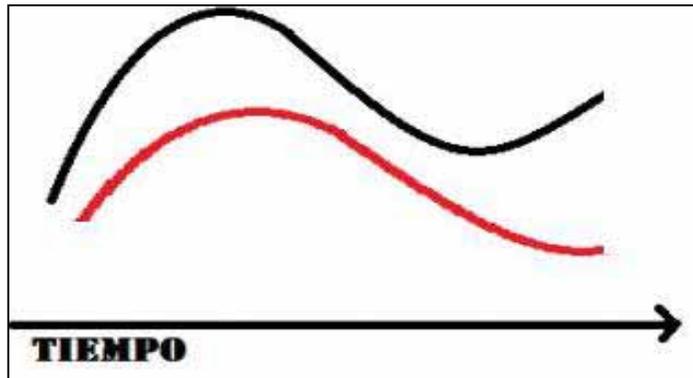


Figura 6. Similitud entre las dos series de tiempo.

#### 5.4 Software utilizado en el proyecto

Para el desarrollo del proyecto se utilizaron tres paquetes de software (**aTube Catcher**, **Praat**, **Paquete R**) descritos a continuación.

### 5.4.1 aTube Catcher

aTube Catcher [13], es un software libre que permite la descarga de videos multimedia de diversos sitios web y la conversión en diferentes formatos de reproducción, figura 7.



Figura 7. Pantalla de inicio de aTube Catcher.

### 5.4.2 Praat

Es un software científico, utilizado para el estudio del habla usado en lingüística. Permite analizar, sintetizar y manipular corpus de habla, así como crear imágenes de alta calidad para graficar dichos procesos.

Fue creado por Paul Boersma y David Weenink, del Departamento de Ciencias Fonéticas de la Universidad de Amsterdam en diciembre de 1995, actualmente es considerado como uno de los programas más utilizados para investigación en fonética acústica [14]. figura 8.

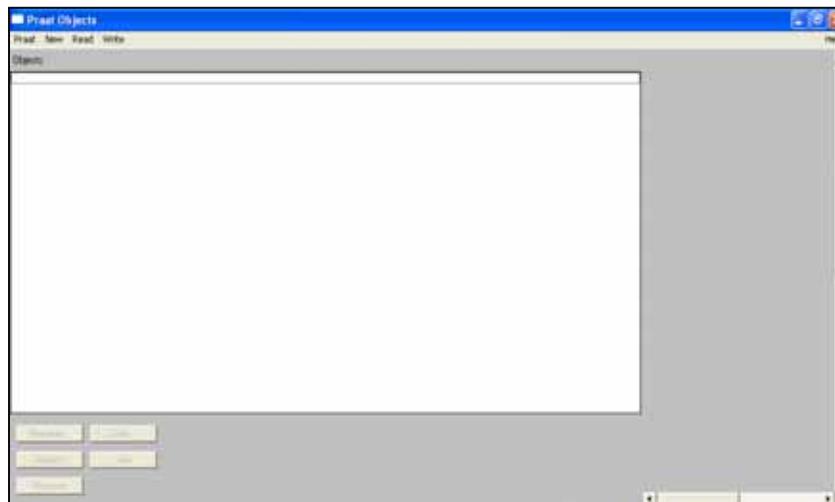


Figura 8. Pantalla de inicio de Praat.

### **5.4.3 Paquete R**

Fue desarrollado inicialmente por Robert Gentleman y Ross Ihaka del Departamento de Estadística de la Universidad de Auckland en 1993.

R es un paquete que permite la implementación de técnicas estadísticas en diversas disciplinas científicas aplicadas. Su carácter gratuito, la multitud de recursos disponibles para el programa y su elevada calidad tanto analítica como gráfica hacen que gradualmente se vaya convirtiendo en una herramienta poderosa para análisis estadístico entre otras funciones [15].

#### **5.4.3.1 Características de R**

Dentro de las principales características con las que cuenta R están:

- Almacenamiento y manipulación efectiva de datos
- Operadores para cálculo sobre variables indexadas (*Arrays*), en particular matrices.
- Una amplia, coherente e integrada colección de herramientas para análisis de datos.
- Posibilidades gráficas para análisis de datos, que funcionan directamente sobre pantalla o impresora, y
- Un lenguaje de programación bien desarrollado, simple y efectivo, que incluye condicionales, ciclos, funciones recursivas y posibilidad de entradas y salidas [16].

#### **5.4.3.2 Extensiones y paquetes**

Los paquetes son las extensiones que R necesita para poder hacer ciertas funciones. Sus usuarios pueden publicar paquetes que extienden su configuración básica. Existe un repositorio oficial de paquetes cuyo número supera los 4,000.

Dado el enorme número de nuevos paquetes, éstos se han organizado en vistas (o temas), que permiten agruparlos según su naturaleza y función. Por ejemplo, hay grupos de paquetes relacionados con estadística bayesiana, econometría, series temporales, etc.

Para facilitar el desarrollo de nuevos paquetes, se ha puesto a servicio de la comunidad una forja de desarrollo que facilita las tareas relativas a dicho proceso [17].

#### **5.4.3.3 Paquete usado en el Proyecto**

El paquete que se utilizó en el proyecto es el paquete *dtw*, permite calcular las alineaciones de series de tiempo libremente con una variedad de restricciones de continuidad, ventanas de

restricción, criterios de valoración, distancias locales, y así sucesivamente. El paquete también proporciona funciones para la visualización de las alineaciones y las limitaciones de uso de varios tipos de diagramas clásicos [15].

## **6. Desarrollo del proyecto**

El desarrollo del proyecto comprende cada una de las actividades realizadas, iniciando con la búsqueda aleatoria de videos con situaciones violentas y situaciones normales. Estos videos se descargaron para obtener de ellos archivos de audio en formato .wav usando el software aTube Catcher. Los videos descargados fueron clasificados de manera manual teniendo un grupo de videos con situaciones violentas y otro grupo de videos con situaciones normales.

De cada video recopilado, mediante la aplicación del software Praat se obtuvieron los parámetros de intensidad, tono, amplitud y frecuencia.

Cada parámetro se analizó por separado mediante la aplicación del algoritmo DTW con el objeto de encontrar la señal de referencia que representa una situación violenta y una situación normal.

### **6.1 Búsqueda y clasificación de videos**

La búsqueda de videos se realizó de manera aleatoria en el portal de internet Youtube para lo cual se utilizó como palabras de búsqueda “videos de violencia” y “videos con situaciones normales”. Los videos buscados contenían situaciones normales como charlas entre personas, niños jugando, etc. Y situaciones de violencia por ejemplo; discusiones, golpes, etc. El total de videos obtenidos fue de 100, con una duración variable. Es decir, el video más corto tiene una duración de 30 segundos y el video más largo tiene una duración de dos minutos. En promedio los videos en estudio tienen una duración de un minuto.

Se realizó la clasificación manual de ellos; es decir, separar los videos en los que ocurrían situaciones de violencia de los que presentaban situaciones normales. La clasificación arrojó 50 videos con situaciones normales y 50 con situaciones de violencia. Cada clasificación se

almaceno en dos carpetas distintas una para situaciones normales y otra para situaciones violentas.

A continuación se realizó la extracción del audio mediante el uso del software aTube Catcher obteniendo archivos en formato .wav, figura 9, los archivos almacenados se analizan a través del software Praat para obtener el parámetro de intensidad, tono, amplitud y frecuencia.

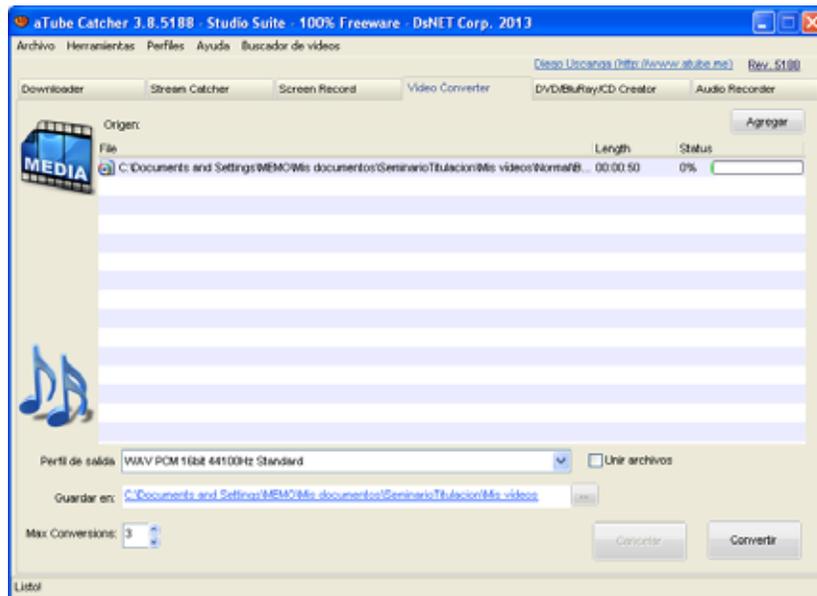


Figura 9. Conversión de videos a formato .wav

## 6.2 Selección de archivos para su análisis con Praat.

Para seleccionar un archivo de audio, se realizaron los siguientes pasos:

1. Abrir el software Praat, del menú **Read**, elegir la opción **Open long sound file**, figura 10.

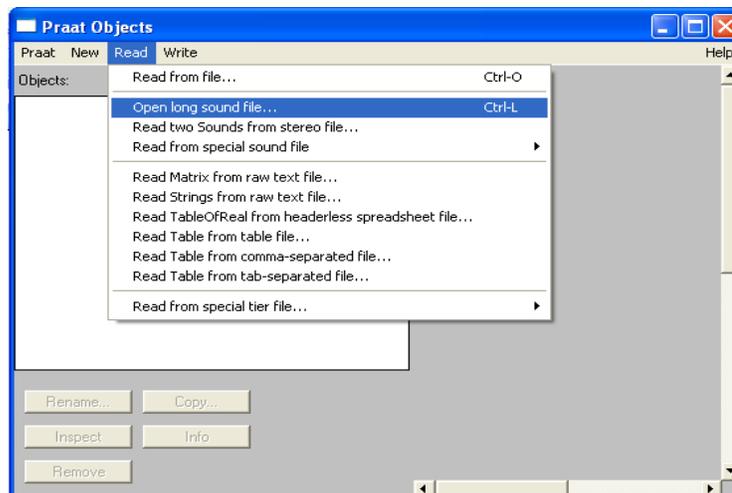


Figura 10. Menú Read de Praat.

- De la ventana que aparece, indicar la ubicación del archivo de audio que se desea analizar, figura 11.

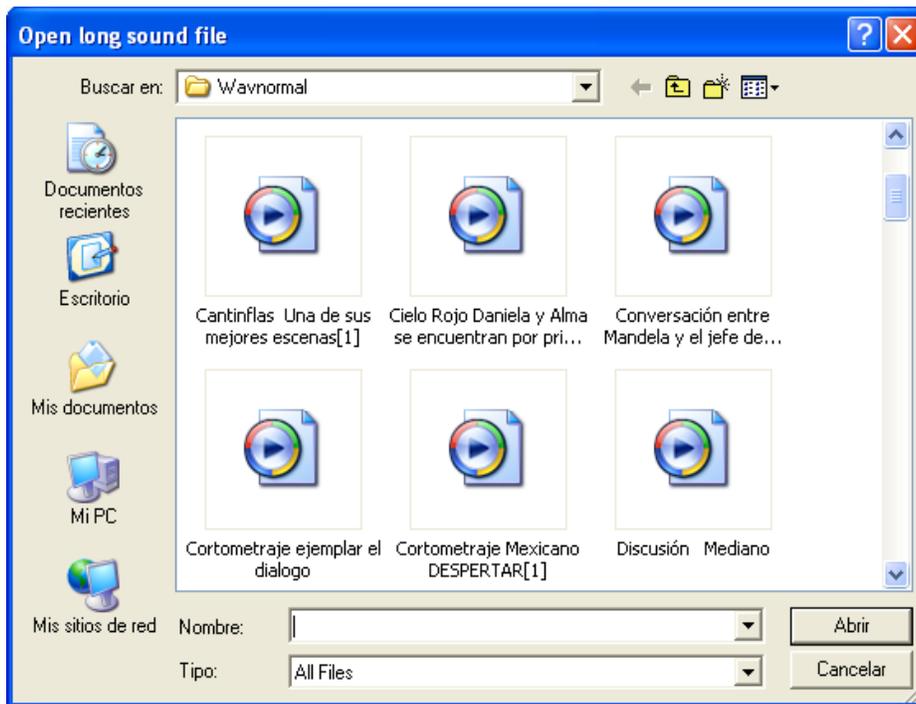


Figura 11. Elección del archivo de audio de Praat.

- De la nueva ventana que surge, elegir la opción **View**, con esto aparecerá el oscilograma y espectrograma del archivo de audio, figura 12.

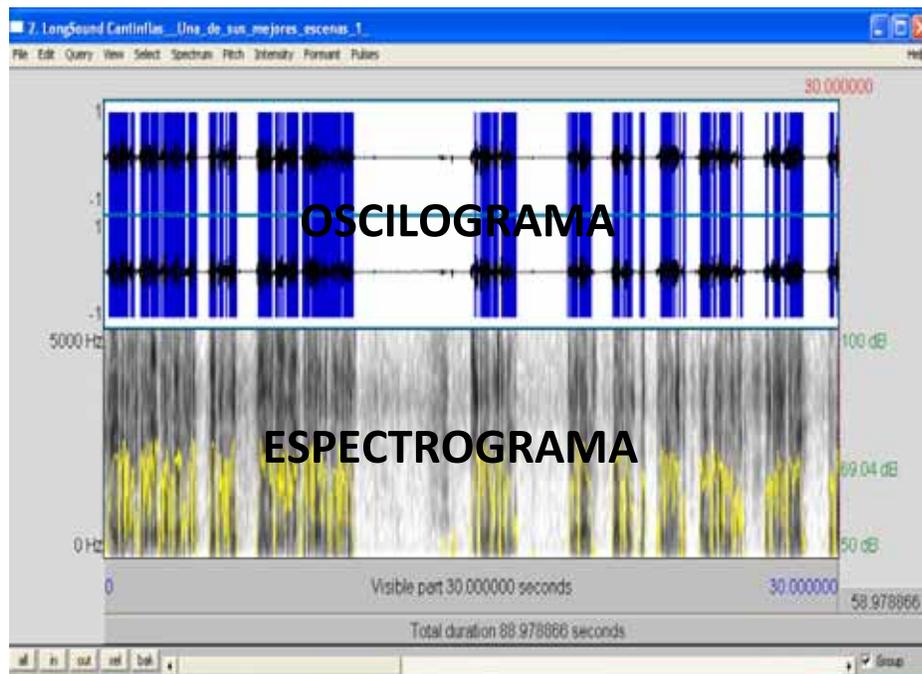
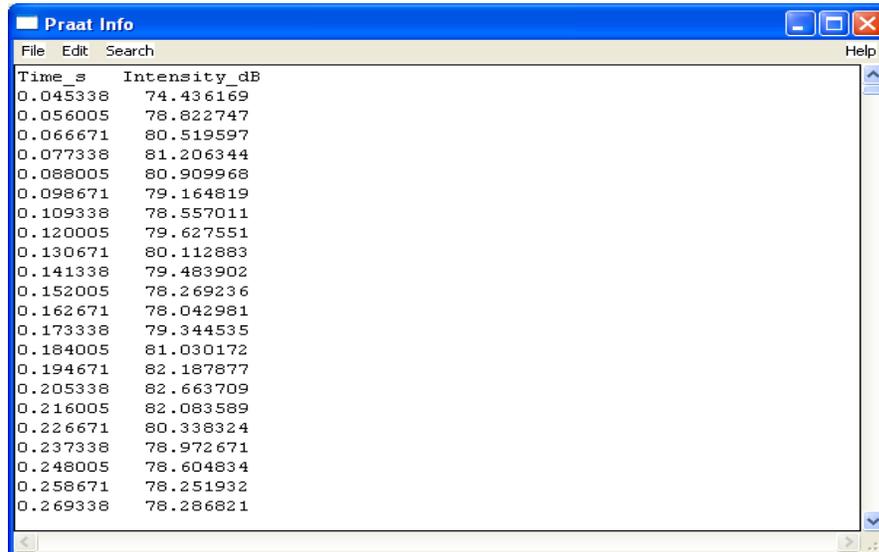


Figura 12. Oscilograma y Espectrograma del archivo de audio de Praat.

4. Seleccionar la parte del oscilograma y elegir el parámetro que se desea obtener, aparecerá un archivo con dos columnas, la primera representa el tiempo y la segunda los valores del parámetro indicado.



Time_s	Intensity_dB
0.045338	74.436169
0.056005	78.822747
0.066671	80.519597
0.077338	81.206344
0.088005	80.909968
0.098671	79.164819
0.109338	78.557011
0.120005	79.627551
0.130671	80.112883
0.141338	79.483902
0.152005	78.269236
0.162671	78.042981
0.173338	79.344535
0.184005	81.030172
0.194671	82.187877
0.205338	82.663709
0.216005	82.083589
0.226671	80.338324
0.237338	78.972671
0.248005	78.604834
0.258671	78.251932
0.269338	78.286821

Figura 13. Valores obtenidos con Praat.

### 6.2.1 Obtención del parámetro intensidad.

Realizados los pasos de la sección 5.2, seleccionar la parte correspondiente al oscilograma, ir al menú **Intensity** y elegir la opción **Intensity listing**, figura 14. Aparecerá un archivo con los valores de intensidad.

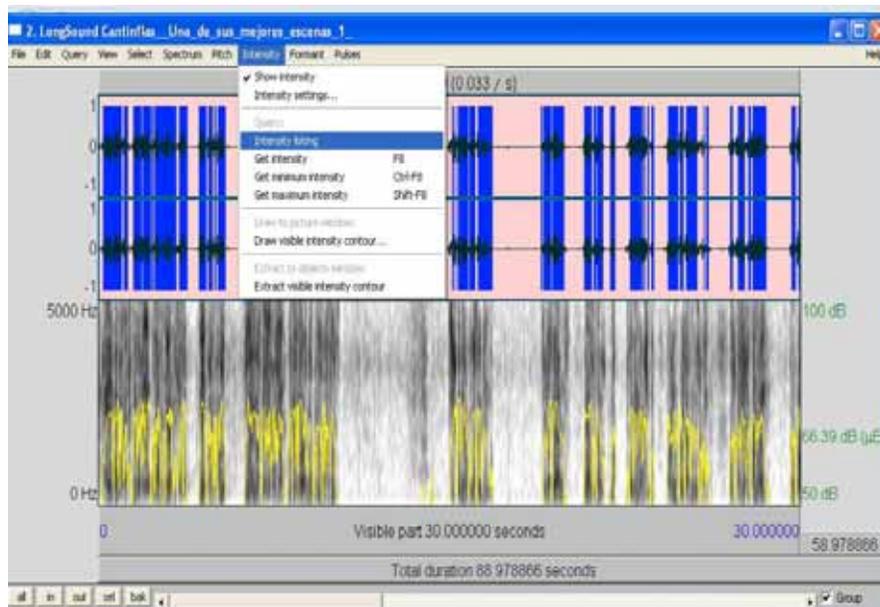


Figura 14. Obtención de los valores de intensidad del archivo de audio de Praat.

## 6.2.2 Obtención del parámetro tono.

Para obtener los valores del tono de un archivo de audio, de la parte del oscilograma, elegir la opción **pitch listing**, figura 15. Aparecerá un archivo con los valores del tono.

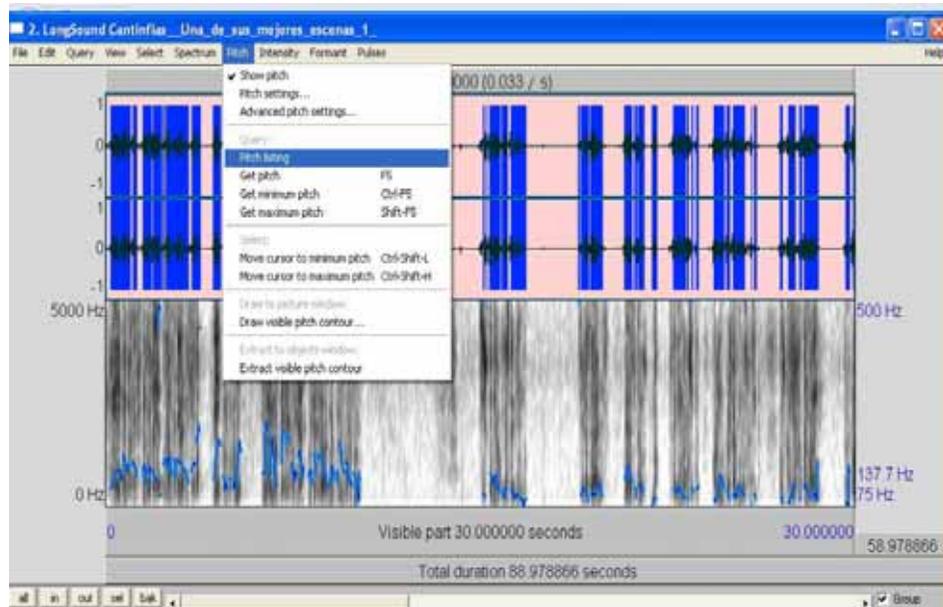


Figura 15. Obtención de los valores de tono del archivo de audio con Praat

## 6.2.3 Obtención del parámetro amplitud.

Para obtener los valores de la amplitud de un archivo de audio, se debe posicionar el cursor en donde se encuentra la señal de la voz y ver que amplitud se tiene, este proceso se debe repetir para obtener diversos valores de amplitud, figura 16.

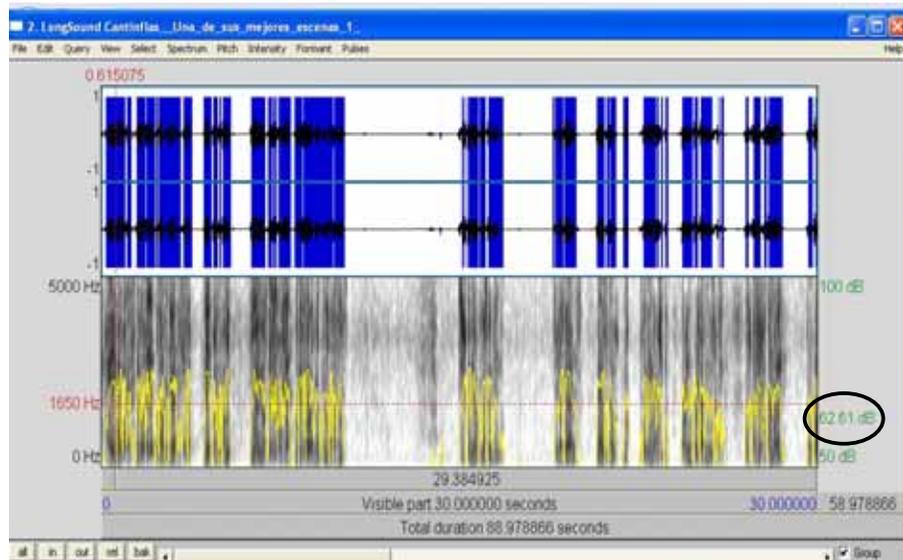


Figura 16. Obtención de los valores de amplitud del archivo de audio con Praat.

#### **6.2.4 Obtención del parámetro frecuencia.**

Para obtener los valores de frecuencia de un archivo de audio, seleccionar la parte del oscilograma, elegir la opción **formant listing**, con esto aparecerá un archivo con todos los valores de frecuencia.

### **6.3 Aplicación del algoritmo DTW en el parámetro intensidad, tono, amplitud y frecuencia.**

El algoritmo DTW realiza una comparación entre dos señales una contra otra (la señal de referencia y la señal de prueba), hasta obtener la señal de referencia. La señal de referencia fue aquella en donde presentó mayor semejanza en una misma región. Asimismo el algoritmo indica el grado de similitud entre las señales que se están comparando. Por esta razón, del conjunto de datos, el objetivo primordial fue encontrar la señal de referencia que proporcionara el mejor grado de similitud en ambas situaciones (violencia y no violencia), en el parámetro intensidad, tono, amplitud y frecuencia.

El algoritmo DTW permite calcular la distancia o grado de similitud que existe entre dos señales, dependiendo de esa distancia se decide si ambas señales son semejantes o no lo son, entre más pequeño sea el valor de la distancia, mayor semejanza de las señales, sin embargo si el valor obtenido es elevado, las señales tienen menor grado de similitud.

A continuación se presentan los resultados más relevantes de las comparaciones realizadas. Para ello se hace uso de una figura con dos secciones. El lado izquierdo de la figura indica la región de concordancia entre ambas señales, para todos los casos se presentan dos señales, una de color verde que representa la señal de referencia y otra de color rojo que es la señal a comparar. El lado derecho de la figura presenta el código aplicado para generar la gráfica correspondiente a cada caso y obtener el índice de concordancia (número en color azul) al comparar dos señales.

### 6.3.1 Situación violenta (intensidad)

Para obtener la señal de referencia del parámetro intensidad, se graficó cada uno de los archivos de audio con situaciones violentas. En la figura 17 se ven algunas de las señales obtenidas por medio del paquete R. El eje x, siempre representará el tiempo de duración del video, mientras que el eje y, muestra los valores de la intensidad.

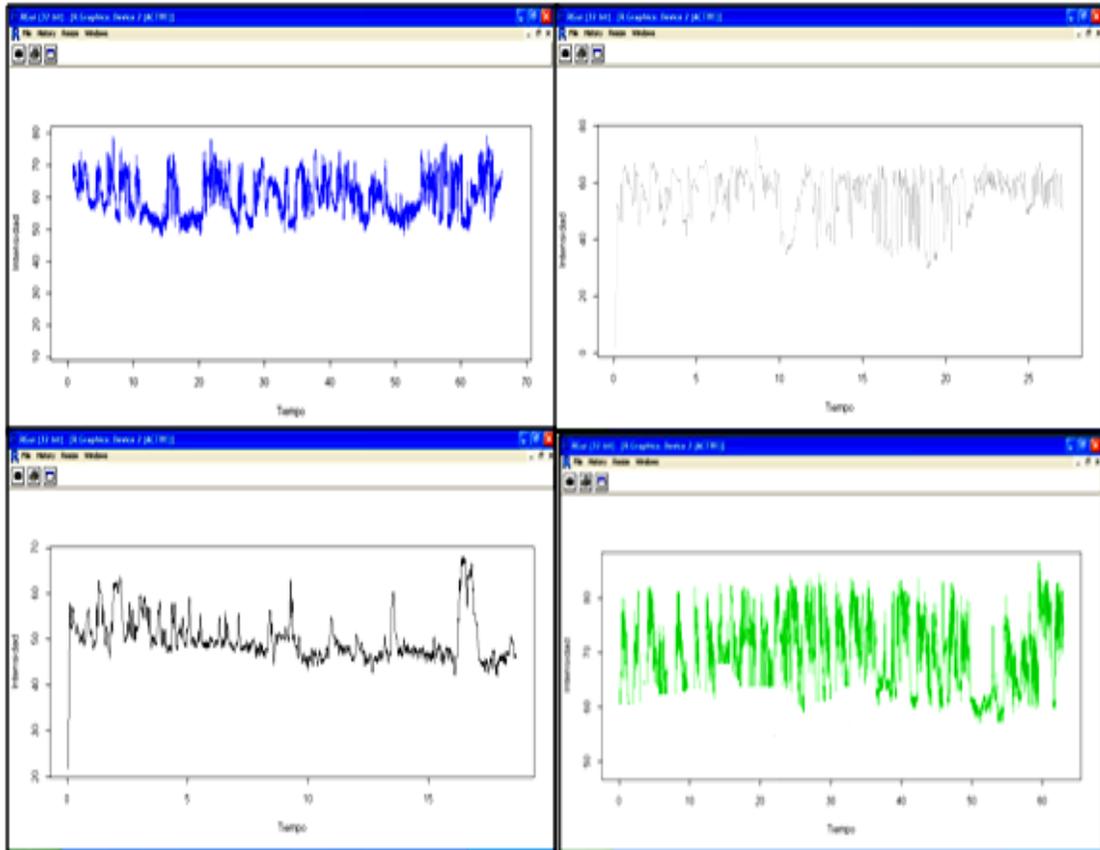


Figura17 .Gráficas del parámetro intensidad en situaciones violentas.

#### A) Peor de los casos.

La figura 18, representa el caso menos favorable, es decir de todas las comparaciones que se realizaron entre las distintas señales clasificadas manualmente como violentas fue la que menor grado de similitud presentó

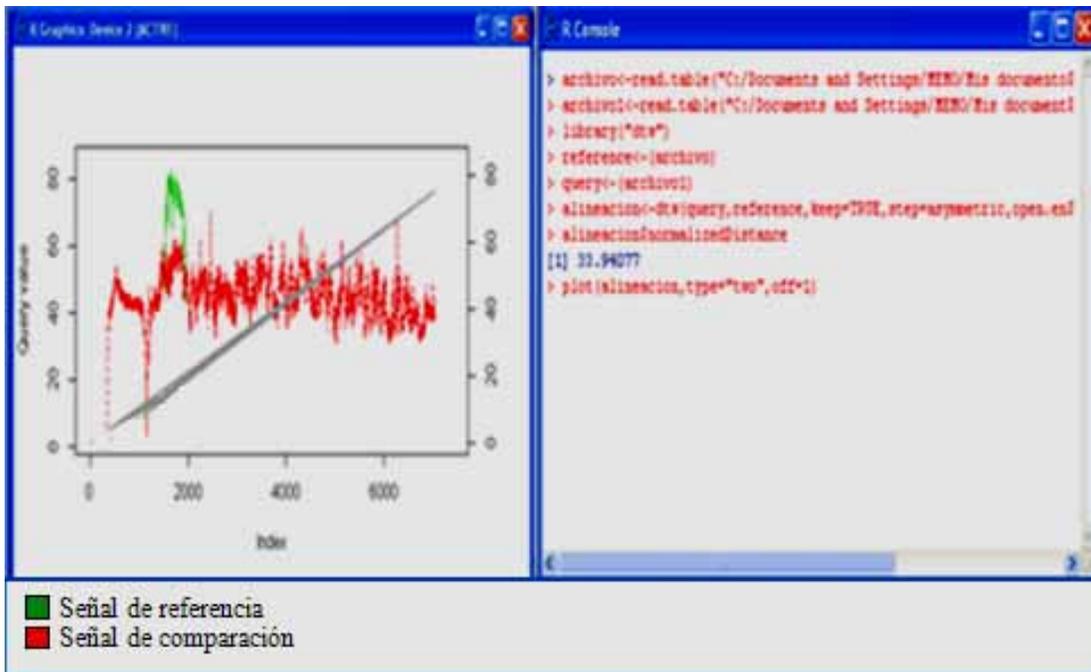


Figura18 .Gráfica y grado de concordancia del parámetro intensidad en situaciones violentas (Peor caso).

El lado izquierdo de la figura 18 indica la región de concordancia entre ambas señales, el lado derecho indica el valor del grado de similitud obtenido, como se ve en la figura hay una correspondencia entre el valor del grado de similitud el cual fue de 33.9407 y la región de concordancia, es decir al ser el valor más elevado que se obtuvo en las distintas comparaciones realizadas se tiene una pequeña región de concordancia entre las señales que se compararon, razón por la cual fue considerado el resultado menos favorable dentro de la clasificación de violencia con respecto al parámetro intensidad.

### **B) Mejor de los casos**

De las comparaciones realizadas se encontró una señal que al ser comparada con el resto proporcionó distancias pequeñas indicando el mejor índice de concordancia, razón por la cual fue elegida como señal de referencia, figura 19. Donde se muestra la región en la que ambas señales coinciden y el índice de concordancia fue de 6.06569. Este fue el valor más pequeño que se obtuvo en todas las comparaciones realizadas.

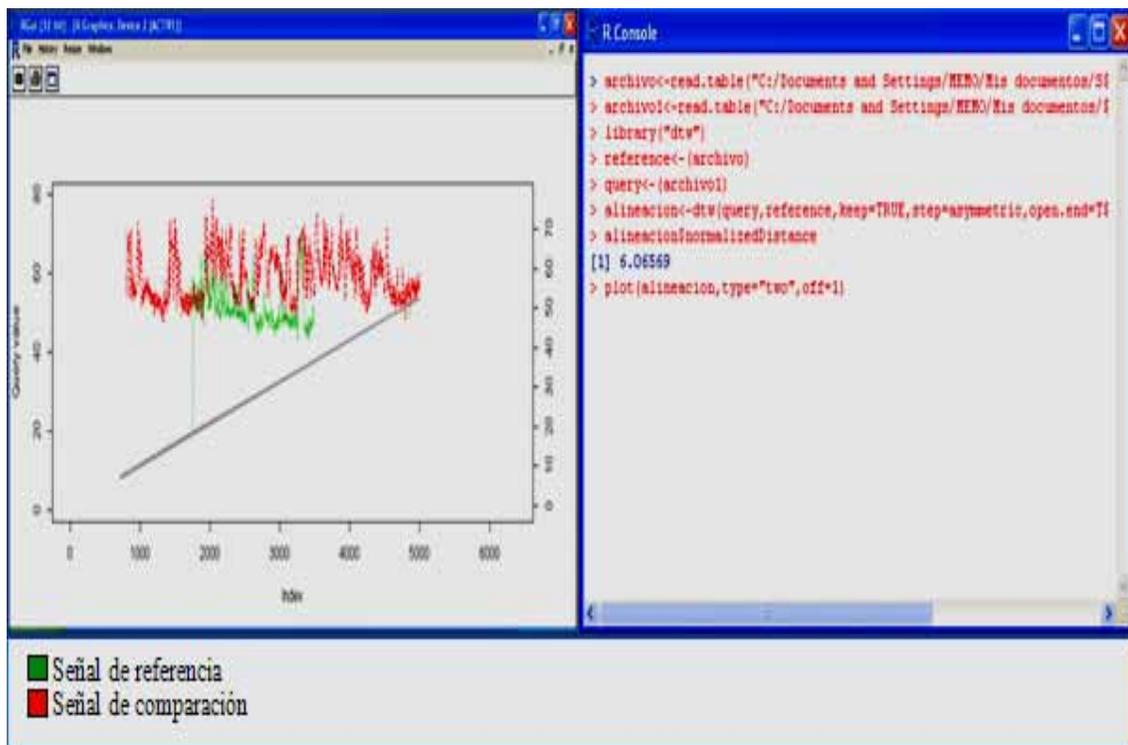


Figura 19. Obtención de la señal de referencia, ejemplo 1.

De los experimentos realizados se presentan dos casos más, con resultados muy parecidos al anterior, ver figura 20 y figura 21, donde el índice de similitud es de 8.7056 y 6.2462, respectivamente.

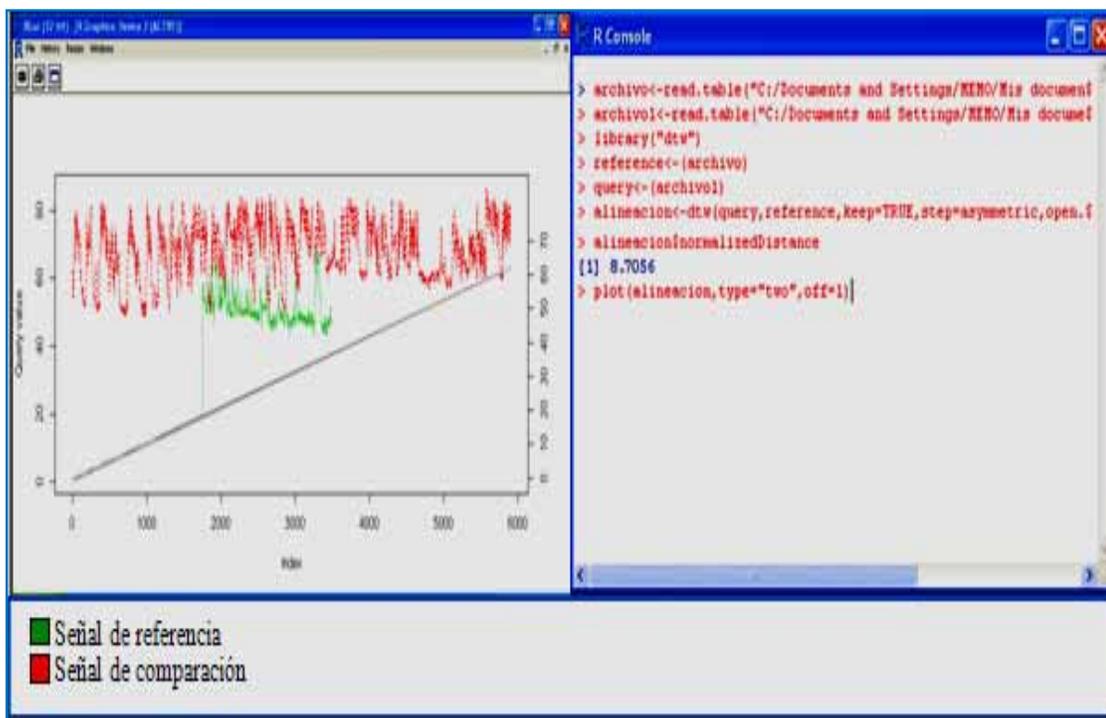


Figura 20. Obtención de la señal de referencia, ejemplo 2.

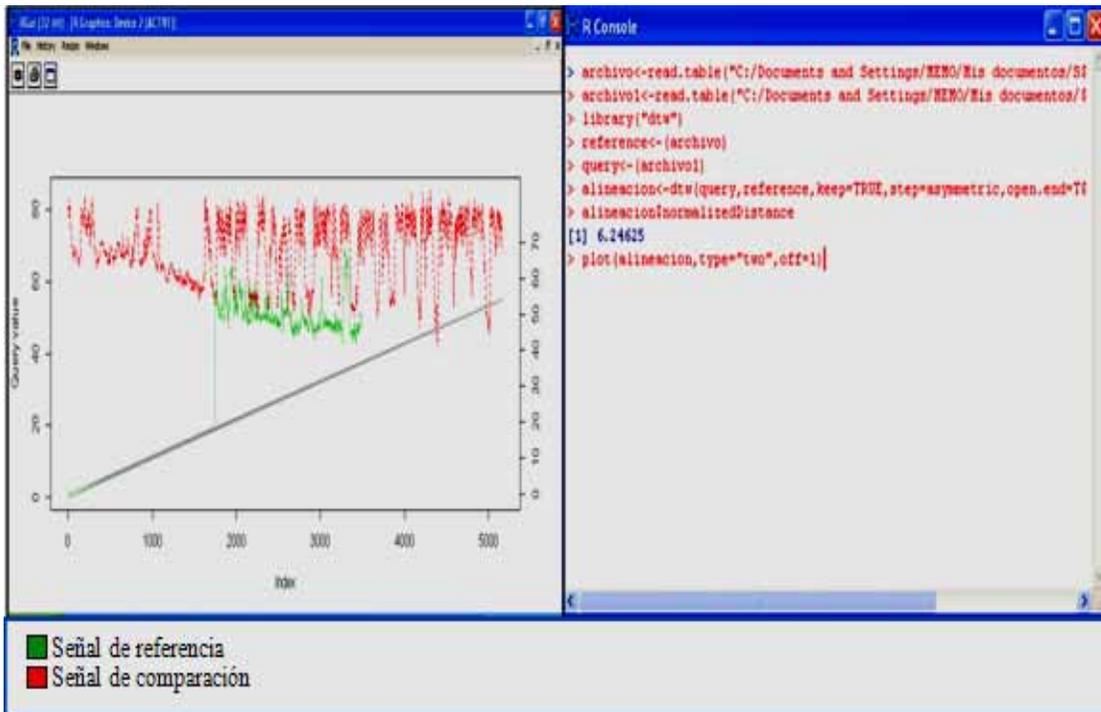


Figura 21. Obtención de la señal de referencia, ejemplo 3.

### 6.3.2 Situación normal (intensidad)

Para situaciones normales fueron comparados cada uno de los archivos que se tenían clasificados manualmente como no violentos unos contra otros por medio del algoritmo DTW con el objetivo de encontrar la señal de referencia para el parámetro de intensidad.

De los resultados obtenidos en las comparaciones se presenta el peor y el mejor de los casos, a continuación se describen cada uno de ellos.

#### A) Peor de los casos

De las comparaciones realizadas entre señales se obtuvo la que menor grado de similitud presentó, esta es la correspondiente a la figura 22, el valor del grado de similitud que se obtuvo para este caso fue de 24.9733.

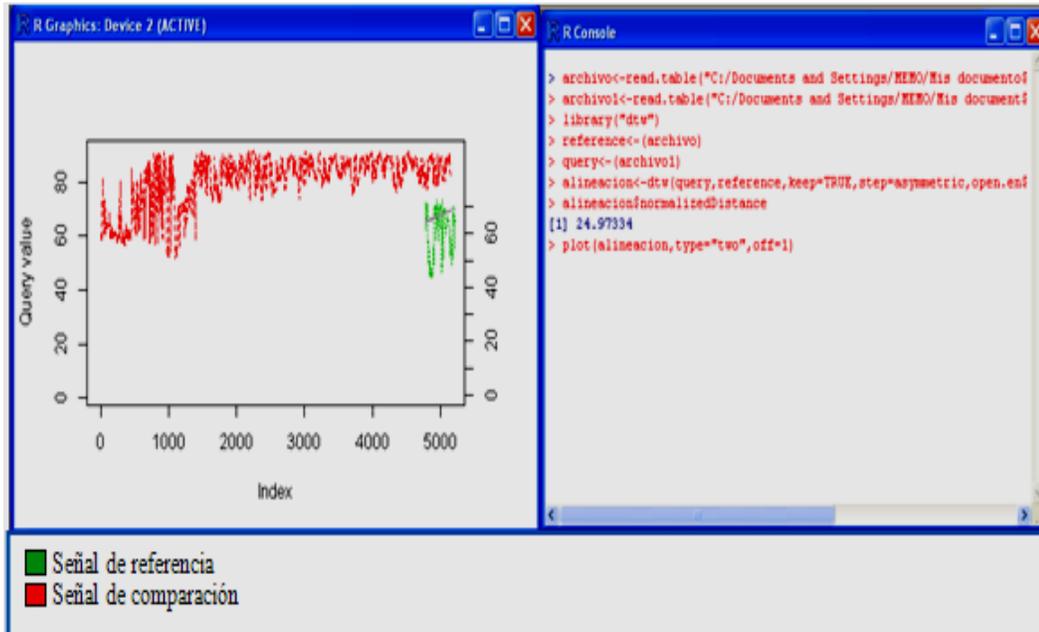


Figura 22 .Gráfica y grado de concordancia del parámetro intensidad en situaciones normales (Peor caso).

En la figura 22, la región de concordancia es pequeña debido a que no hay gran similitud entre las señales.

### A) Mejor de los casos

Las figuras 23 a 25 representan la señal que mayor grado de concordancia tuvo con el resto.

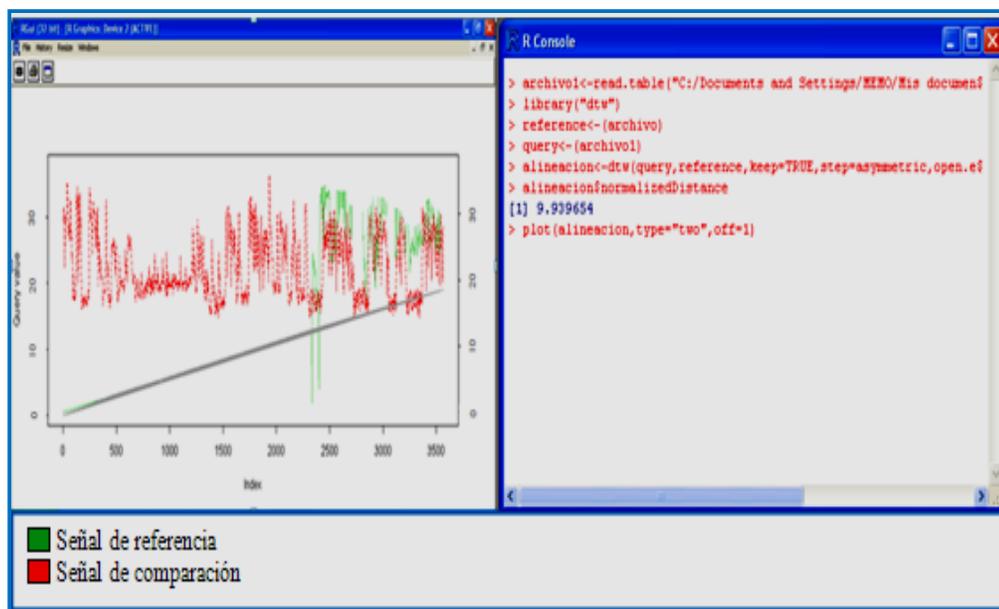


Figura 23. Señal de referencia en situaciones normales, ejemplo 1.

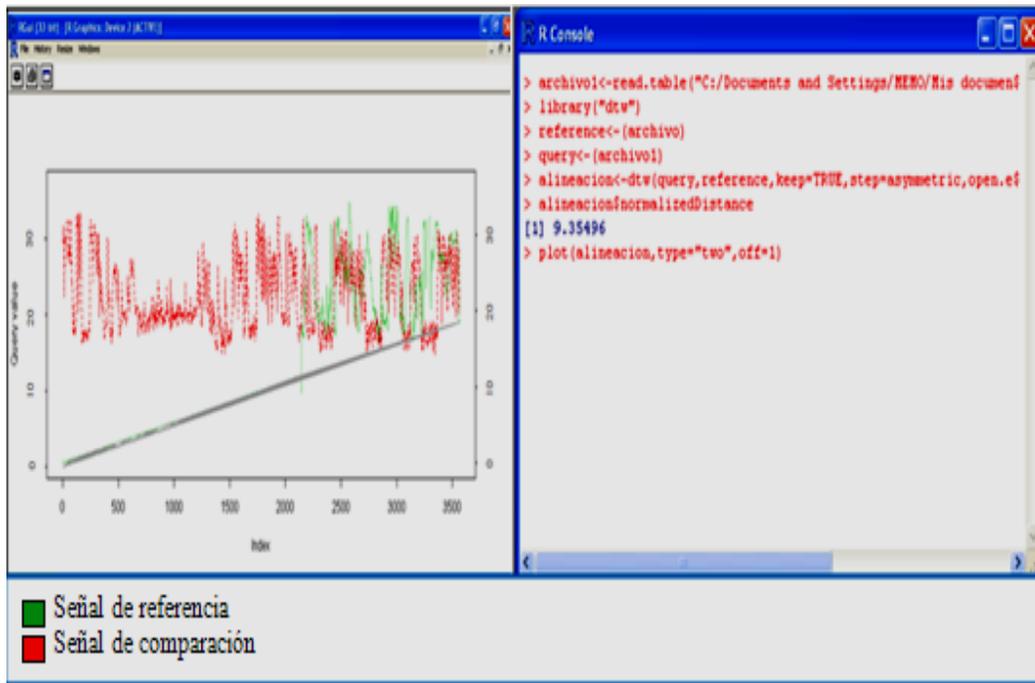


Figura 24. Señal de referencia en situaciones normales, ejemplo 2.

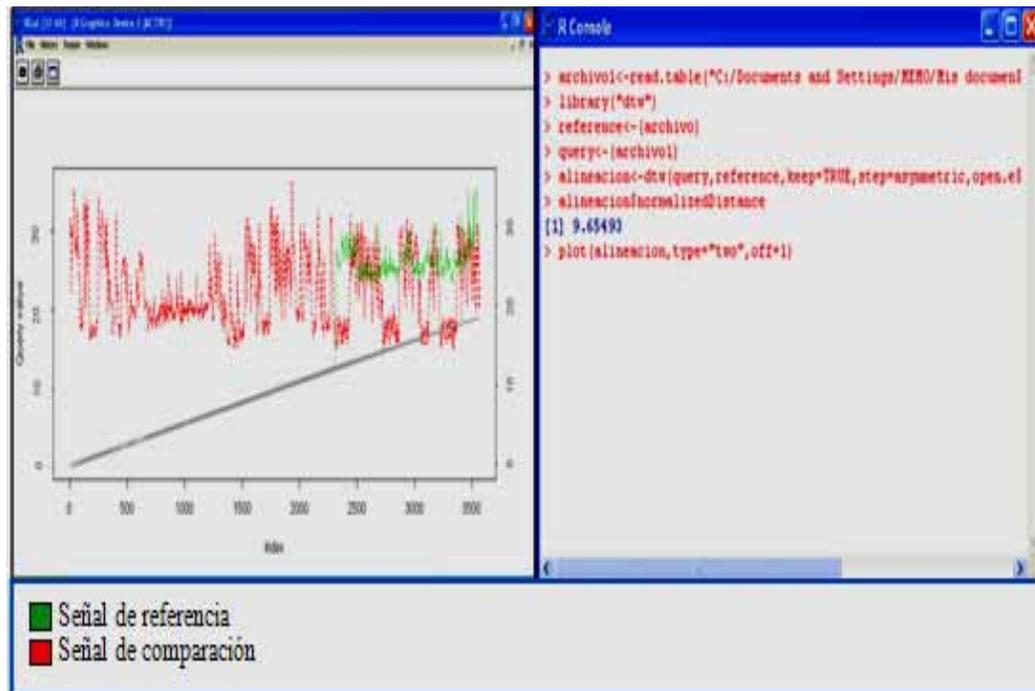


Figura 25. Señal de referencia en situaciones normales, ejemplo 3.

Recordemos que el eje x, corresponde al tiempo y el eje y, a los valores de intensidad. En los tres ejemplos mostrados, la señal en color verde representa la señal de referencia para el

parámetro intensidad dentro de las situaciones normales, la señal en color rojo es la que se compara con la de referencia.

Existe un comportamiento parecido entre estos y otros experimentos, cuyos valores de grado de similitud son de 9.93965, 9.3549 y 9.6549 respectivamente.

## 6.4 Aplicación del algoritmo DTW en el parámetro tono

Para la obtención de la señal de referencia representativa del parámetro tono de una situación violenta o normal, se procedió de la misma forma en que se obtuvieron las señales de referencia del parámetro intensidad.

### 6.4.1 Situación violenta.

Primero, se obtuvieron las señales de cada uno de los archivos con los valores de tono que se tenían clasificados como situaciones violentas. En la figura 26, se presentan algunas de las gráficas obtenidas por medio del paquete R y que sirvieron para saber cómo serían las gráficas que el parámetro tono tendría bajo esta clasificación.

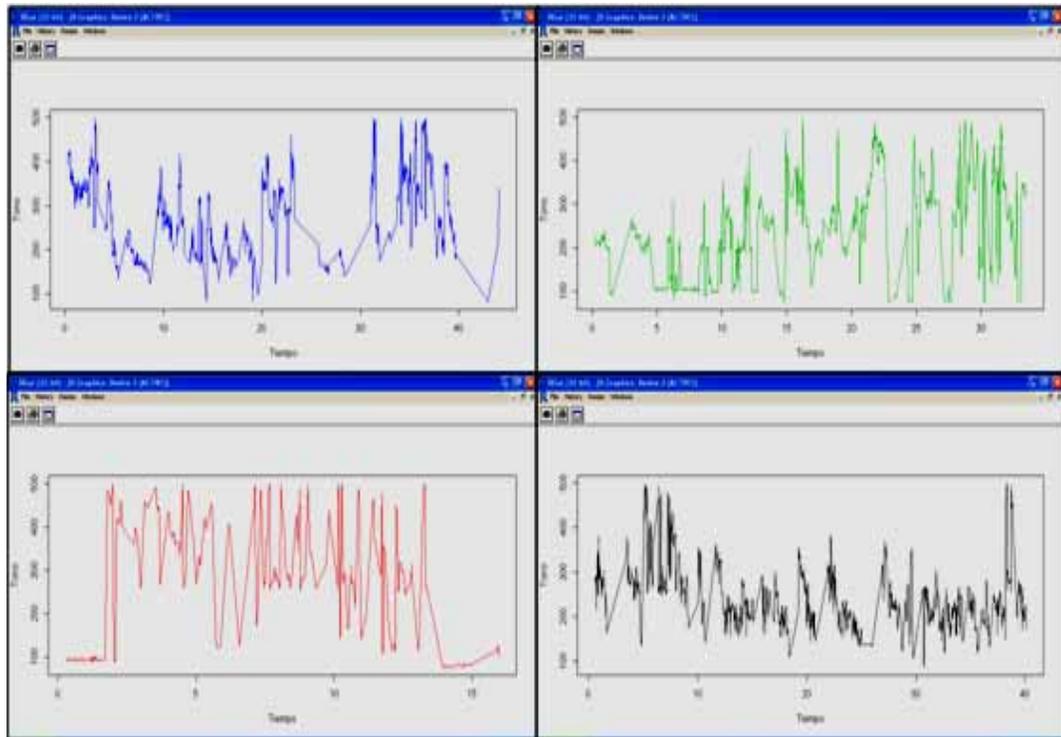


Figura 26. Ejemplos de gráficas del parámetro tono (violencia).

Cada uno de los archivos que contenían los valores numéricos de tono bajo la clasificación de situaciones violentas fueron comparados unos contra otros, para esto fue utilizado el algoritmo DTW.

El grado de similitud es el criterio que sirvió para determinar cuál de las señales fue elegida como señal de referencia, de las comparaciones realizadas se verán los siguientes casos.

### A) Peor de los casos

Después de realizar cada una de las comparaciones se obtuvo el valor 61.2456 que representó la distancia entre dos señales con el menor grado de similitud entre ellas, por ello fue considerado este caso como el peor, figura 27.

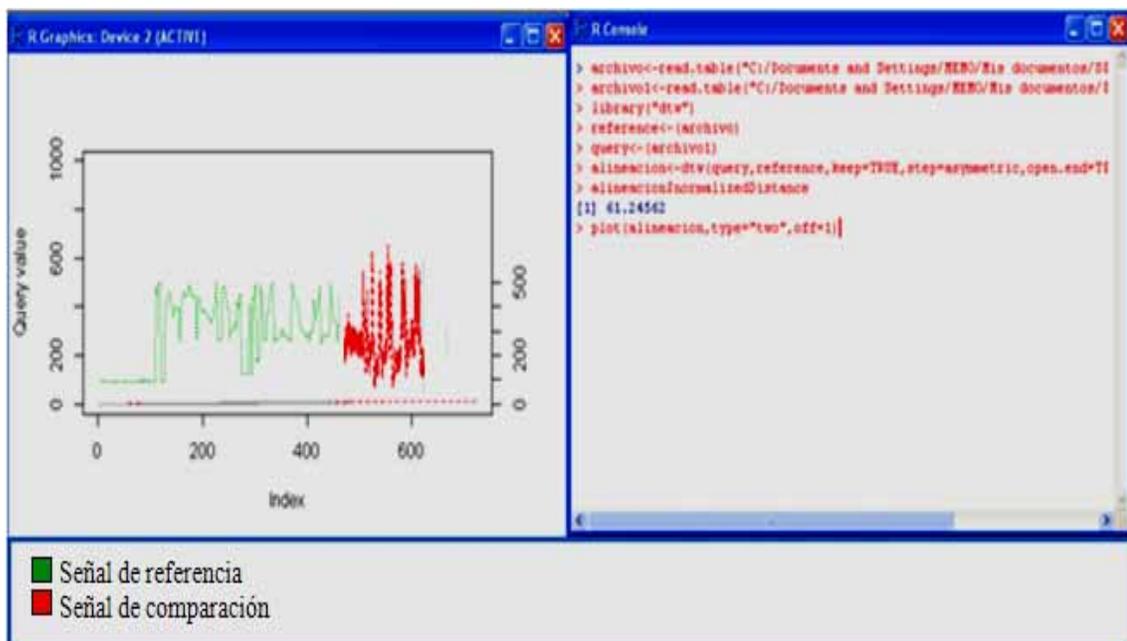


Figura 27 .Gráfica y grado de concordancia del tono en situaciones violentas (Peor caso).

En la figura 27, prácticamente no existe región de concordancia entre la señal de referencia y la señal de comparación, debido a la poca similitud existente entre ambas señales.

### B) Mejor de los casos

La señal presentada en la figura 28, se considera como la señal de referencia representativa del parámetro tono en una situación violenta presentando grado de similitud con un valor de 10.4217.

En las figuras 29 y 30 se ven dos comparaciones realizadas con esta misma señal elegida como referencia y cuyos grados de similitud fueron de 12.6472 y 14.0172 respectivamente.

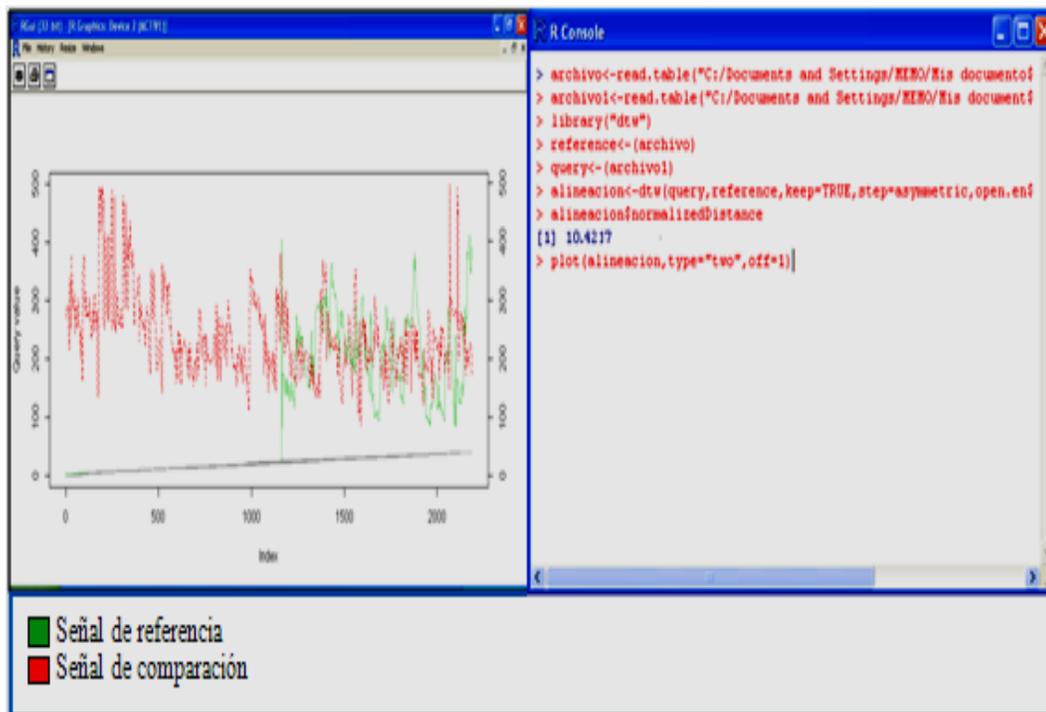


Figura 28. Obtención de la señal de referencia del parámetro tono, ejemplo 1.

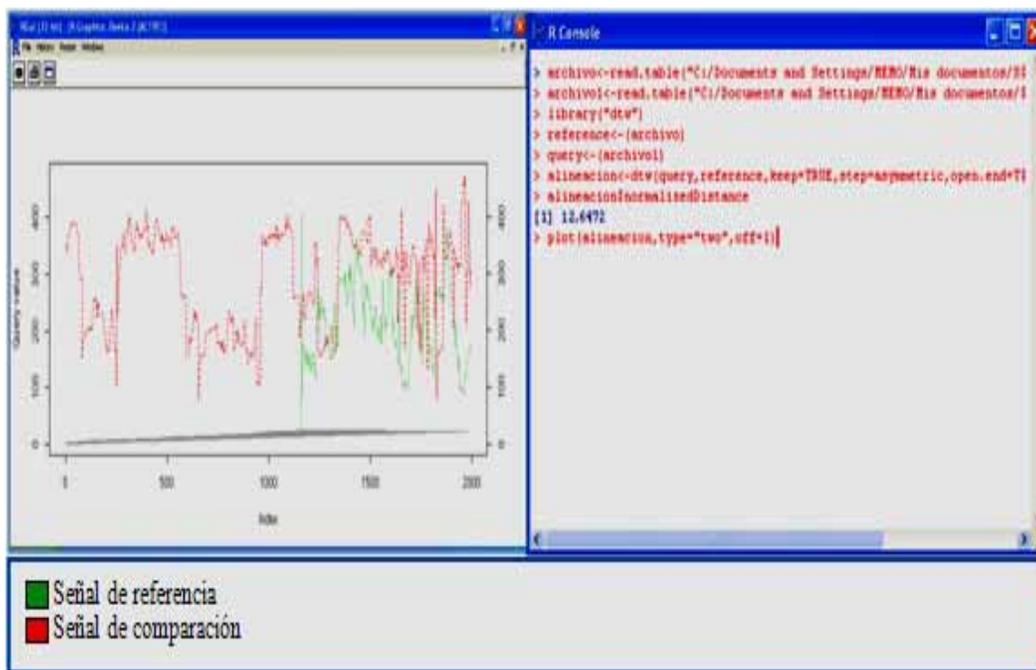


Figura 29. Obtención de la señal de referencia del parámetro tono, ejemplo 2.

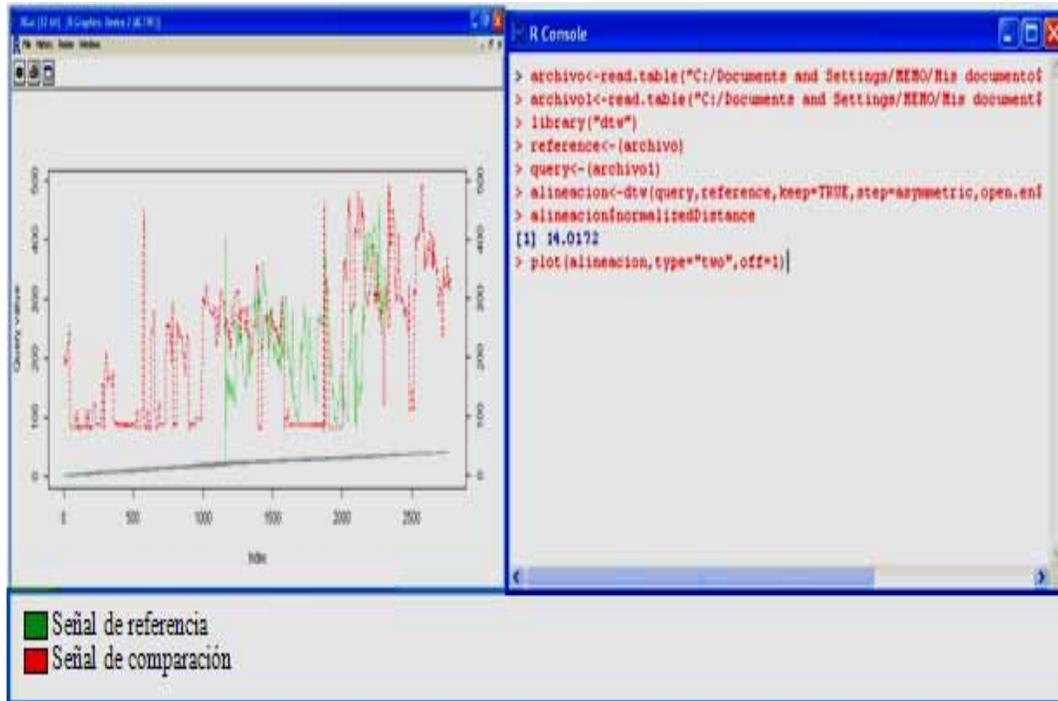


Figura 30. Obtención de la señal de referencia del parámetro tono, ejemplo 3.

#### 6.4.2 Situación normal

Aquí analizaremos los dos casos más relevantes, es decir la peor y la mejor de las comparaciones realizadas con la finalidad de saber cuál de todas las comparaciones realizadas fue la que menor grado de similitud presentó y cuál fue la que mayor grado de similitud tuvo y por ende fue elegida como señal de referencia para el parámetro tono en situaciones normales.

##### A) Peor de los casos

La figura 31 es el caso donde las señales comparadas presentaron mayor distancia entre si y por lo tanto poco grado de similitud. Se aprecia la escasa región de concordancia entre las señales que se están comparando, el grado de concordancia obtenido fue de 84.2113, siendo este el valor más elevado en todas las comparaciones realizadas.

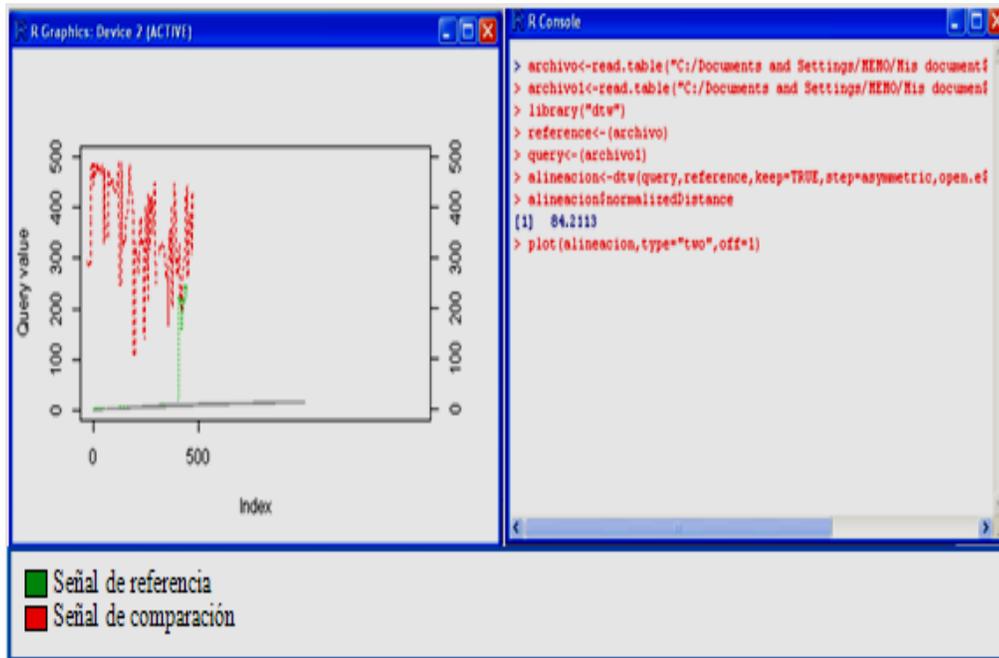


Figura 31 .Gráfica y grado de concordancia del tono en situaciones normales (Peor caso).

### B) Mejor de los casos

Las figuras 32 y 33, muestran la señal que fue elegida como referencia, presentando mayor grado de concordancia.

La señal en color verde representa la señal de referencia, la de color rojo, son las señales con las que se comparó. Las distancias obtenidas fueron de 25.6011 y 26.0115.

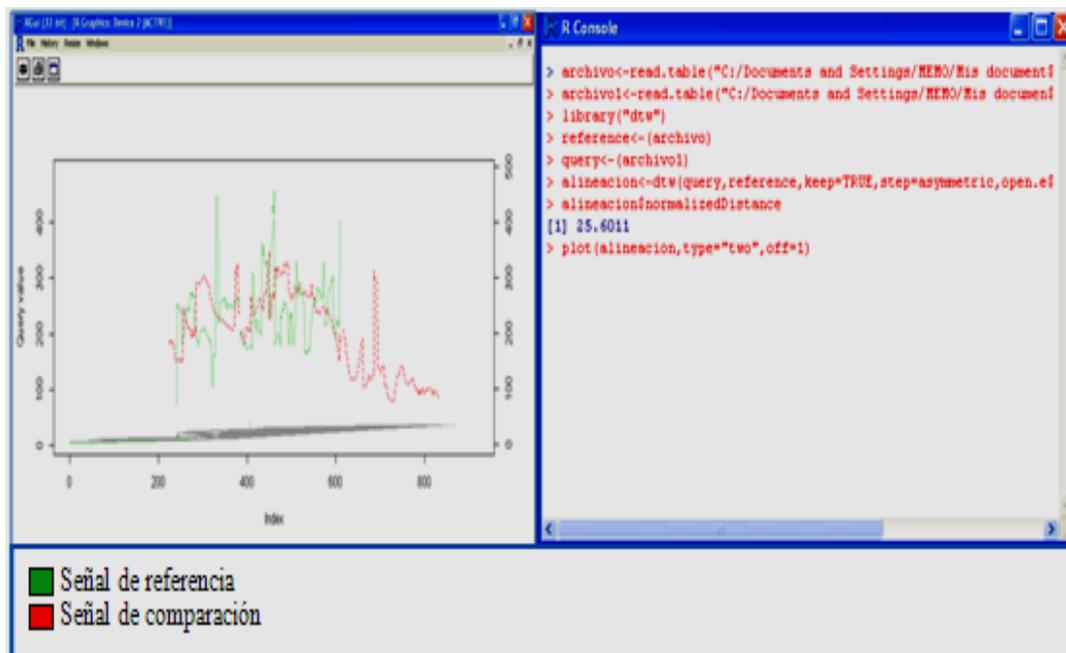


Figura 32. Ejemplo 1 de gráficas del parámetro tono (situación normal).

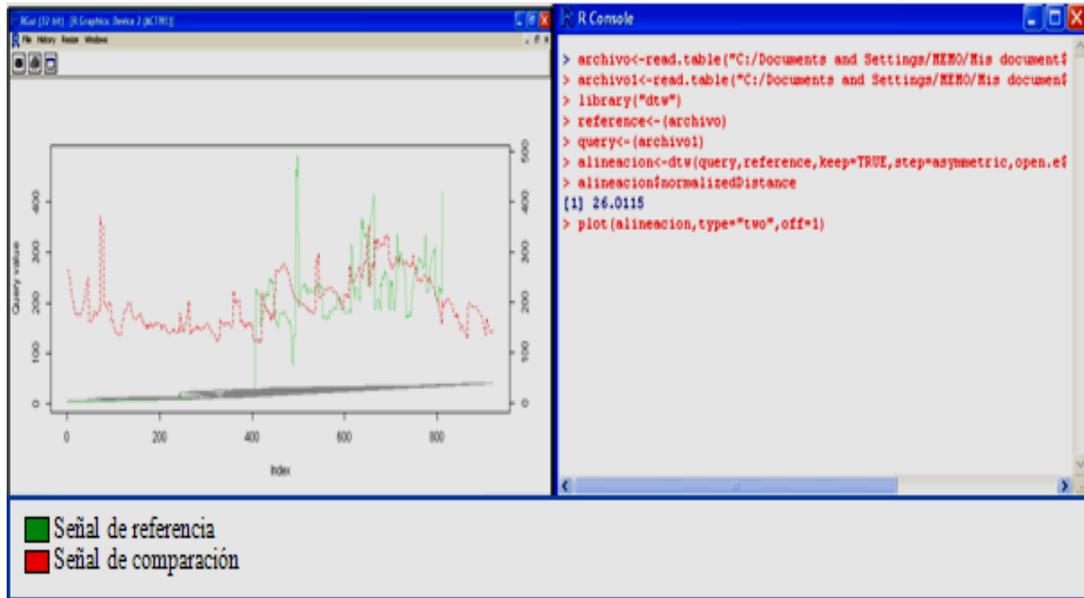


Figura 33. Ejemplo 2 de gráficas del parámetro tono (situación normal).

## 6.5 Aplicación del algoritmo DTW en el parámetro amplitud.

### 6.5.1 Situación violenta

Para obtener la señal de referencia del parámetro amplitud de una situación de violencia, lo primero que se hizo fue obtener la señal característica de cada uno de los archivos que habían sido clasificados de forma manual como violentos. Algunas de las graficas obtenidas se presentan en la figura 34.

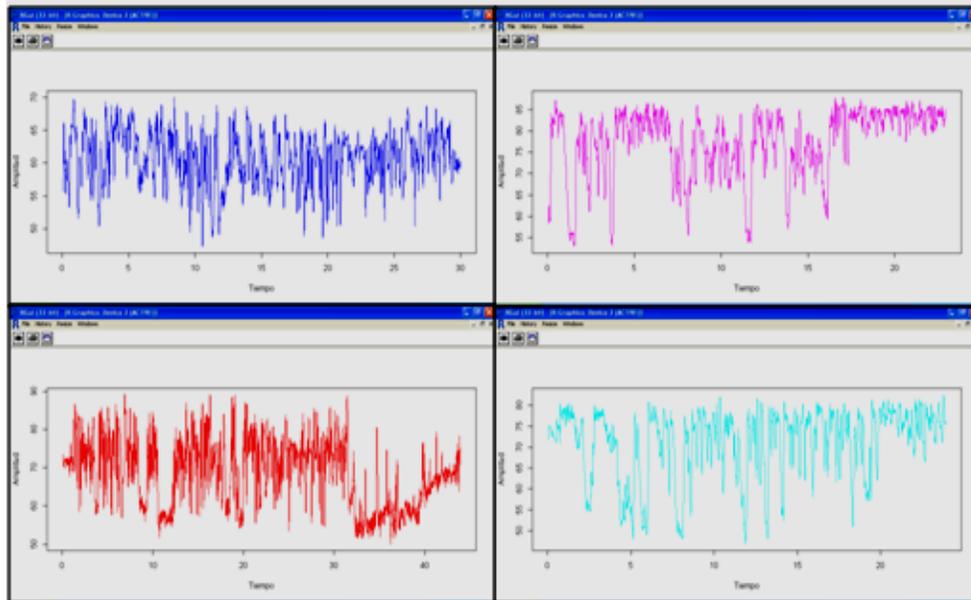


Figura 34. Ejemplos de gráficas del parámetro amplitud (violencia).

Recordemos que el eje x, representa el tiempo y el eje y, los valores de amplitud.

A continuación se presenta el peor y el mejor de los resultados encontrados tras realizar las comparaciones.

### A) Peor de los casos

La figura 35, representa la señal con menor similitud en el parámetro de amplitud en una situación de violencia. El valor del grado de similitud que se encontró por medio del algoritmo DTW al realizar la comparación de estas dos señales fue de 58.2591.

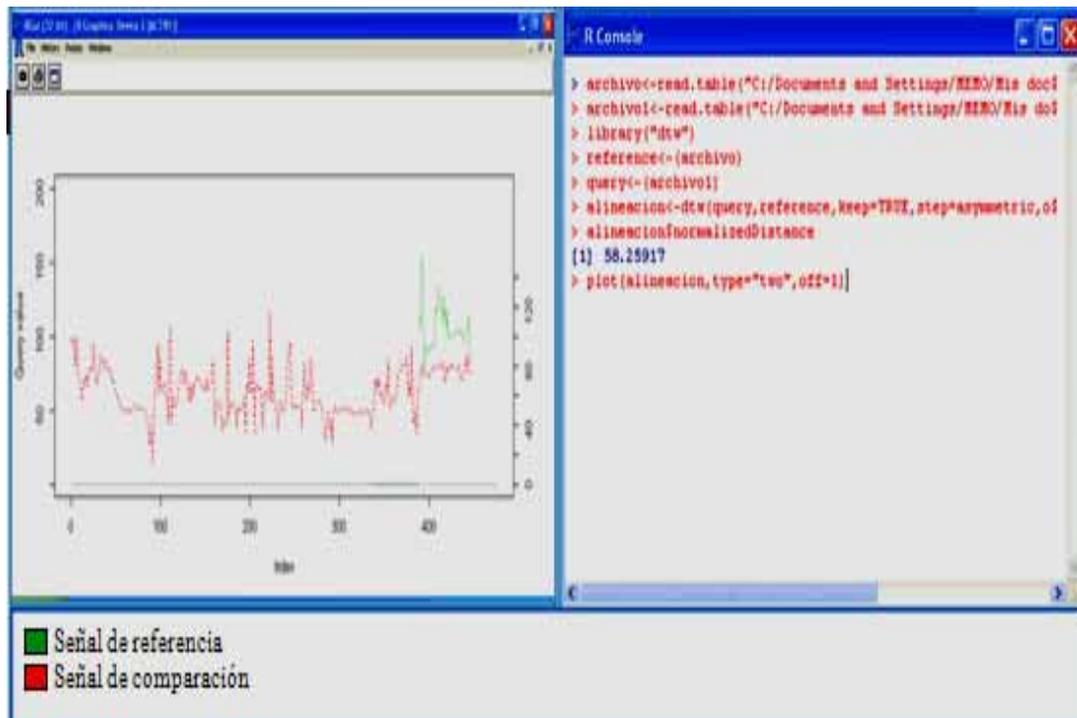


Figura 35 .Gráfica y grado de concordancia de la amplitud en situaciones violentas (Peor caso).

### B) Mejor de los casos

La figura 36, presenta la señal con mayor región de concordancia con un grado de concordancia de 13.1827.

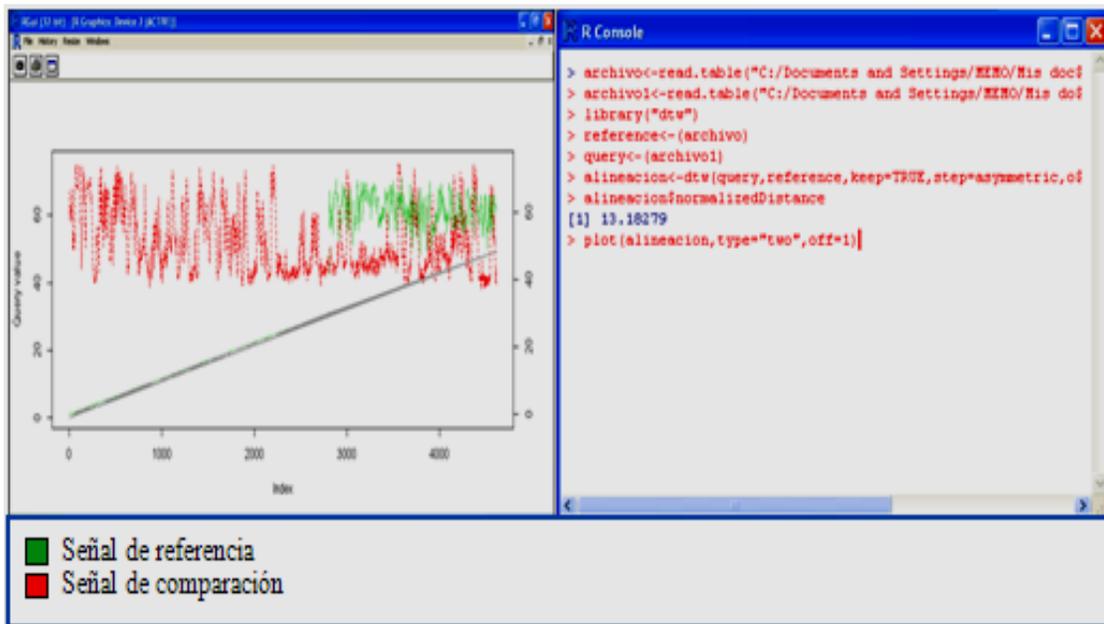


Figura 36. Obtención de la señal de referencia del parámetro amplitud, ejemplo 1.

Las figuras 37 y 38, son comparaciones que se realizaron entre la señal elegida como referencia y otras señales, el resultado obtenido es muy parecido al anterior con respecto al grado de concordancia al obtenerse valores de 15.5829 y 17.5825.

La señal en color verde representa la señal de referencia del parámetro amplitud en una situación violenta.

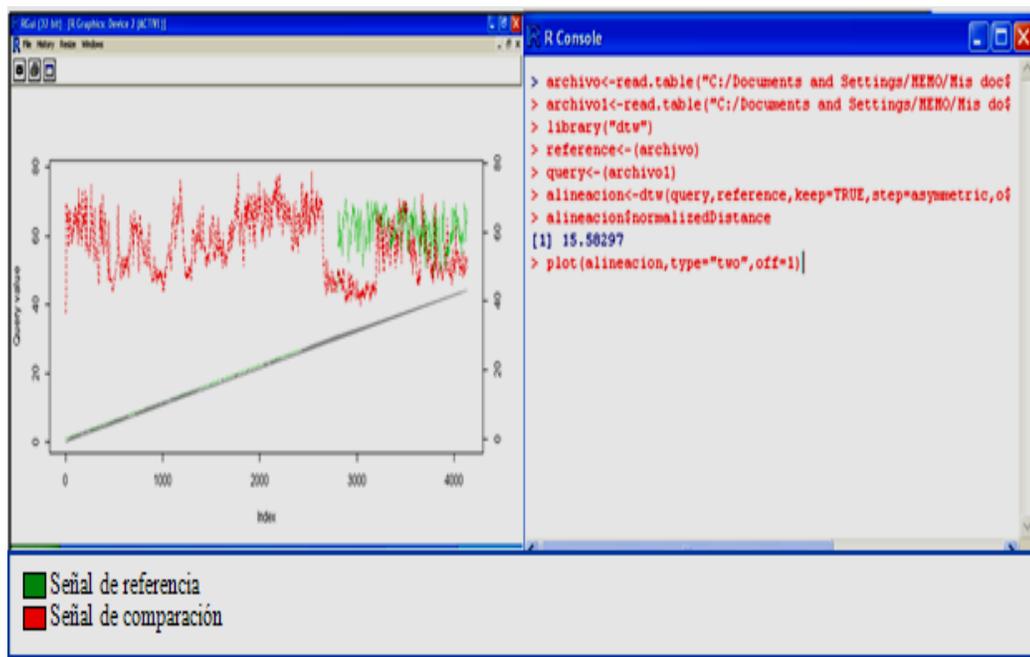


Figura 37. Obtención de la señal de referencia del parámetro amplitud, ejemplo 2.

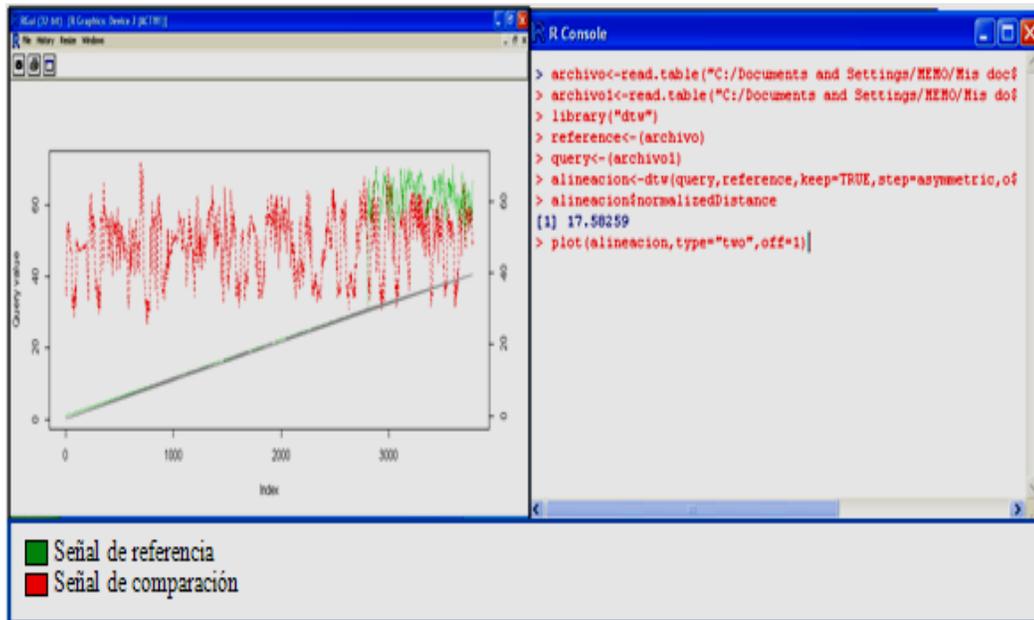


Figura 38. Obtención de la señal de referencia del parámetro amplitud, ejemplo 3.

### 6.5.2 Situación normal

De cada uno de los archivos que se tenían clasificados de forma manual como situaciones normales se obtuvo por medio del paquete R sus respectivas señales, figura 39. Cada señal fue comparada, una contra otra, hasta encontrar aquella que tuviese la mayor similitud posible con el resto. Después de hacer estas comparaciones se obtuvieron los siguientes resultados.

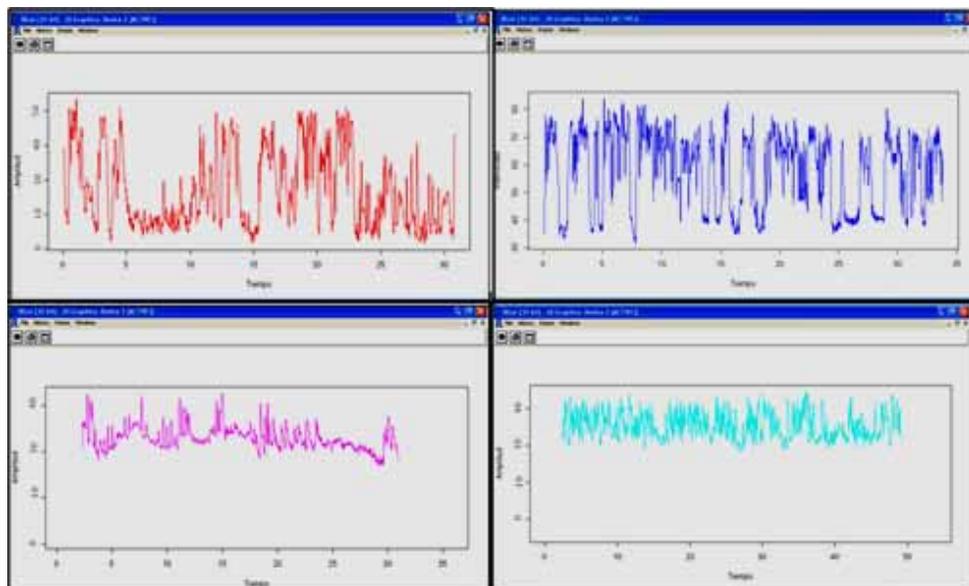


Figura 39. Ejemplos de gráficas del parámetro amplitud (no violencia).

### A) Peor de los casos

Las comparaciones que se realizaron con el algoritmo DTW mostraron que la comparación que menor grado de similitud presentó es la que se muestra en la figura 40. El grado de concordancia fue de 82.3114.

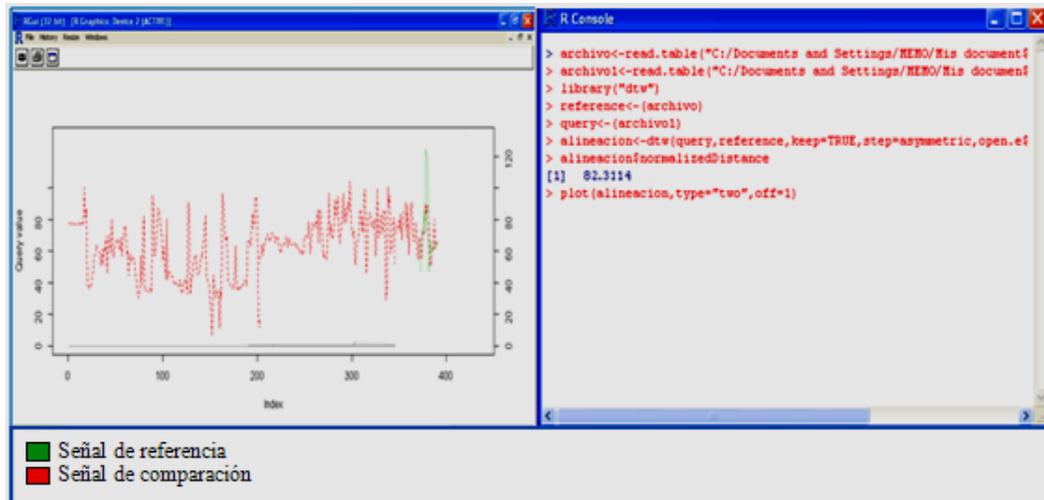


Figura 40. Gráfica y grado de concordancia de la amplitud en situaciones normales.

### B) Mejor de los casos

El valor de la distancia más pequeño que se encontró después de realizar una serie de comparaciones entre los archivos que fueron clasificados manualmente como no violentos con respecto al parámetro amplitud fue de 23.9734, figura 41, este valor representa el mejor grado de similitud.

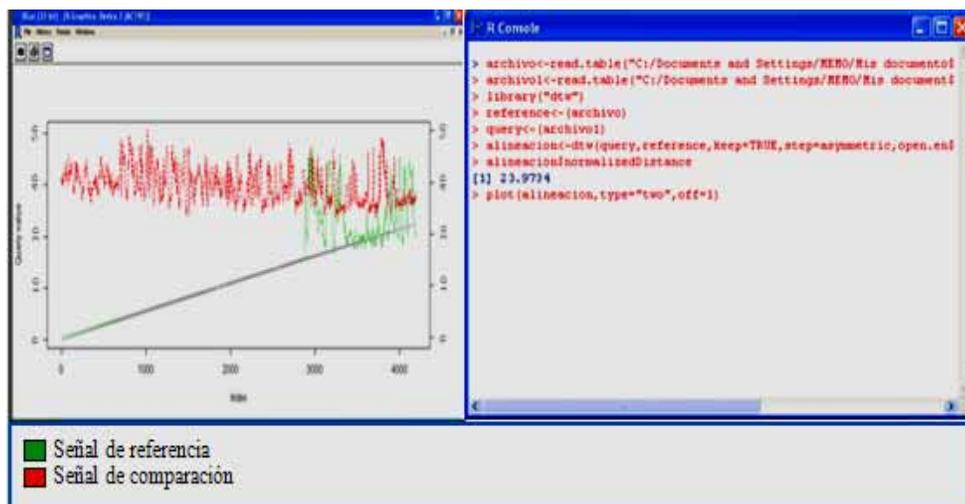


Figura 41. Obtención de la señal de referencia del parámetro amplitud, (Mejor caso).

La figura 41 presenta la región de concordancia que existe entre la señal elegida como referencia (color verde) y la señal con la que se compara (color rojo). A continuación se presentan dos comparaciones más que se realizaron con la señal elegida como referencia, el primero, figura 42, con un valor en el grado de similitud de 30.9437 y el segundo, figura 43, con un valor en el grado de similitud de 36.8948.

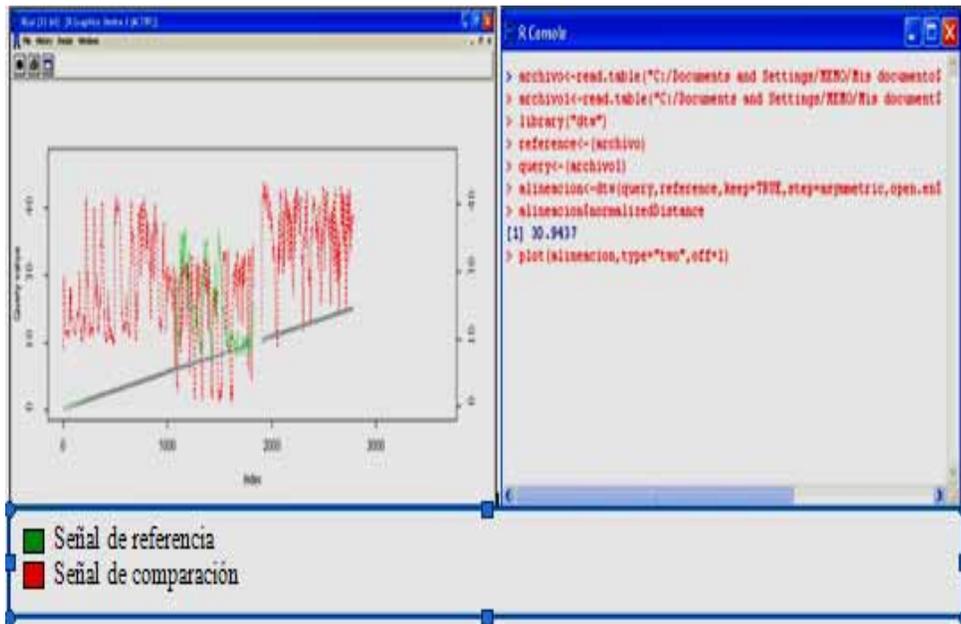


Figura 42. Obtención de la señal de referencia del parámetro amplitud, ejemplo 2

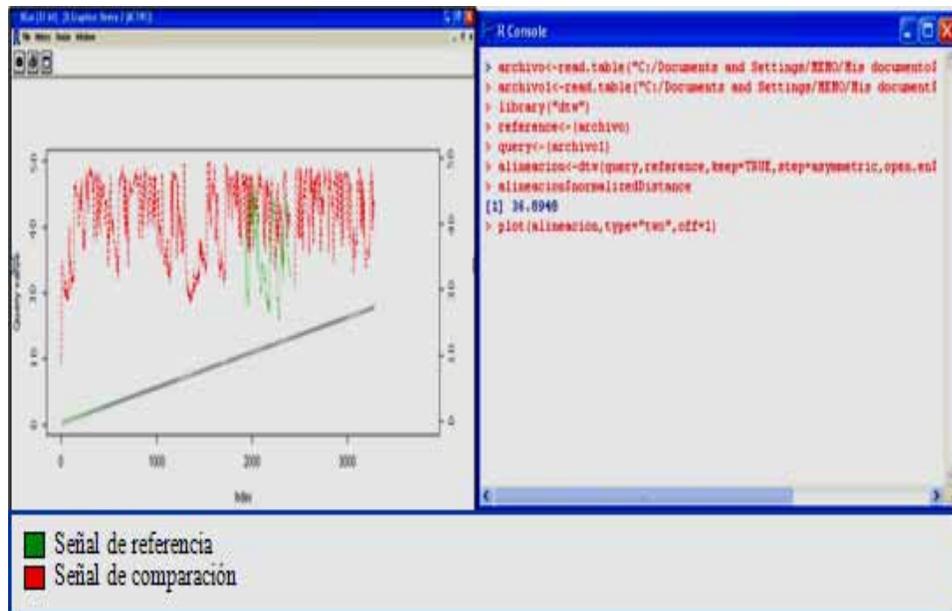


Figura 43. Obtención de la señal de referencia del parámetro amplitud, ejemplo 3.

## 6.6 Aplicación del algoritmo DTW en el parámetro frecuencia

### 6.6.1 Situación violenta

Con ayuda del paquete R, se obtuvieron cada una de las señales de los archivos con los datos de frecuencia que estaban clasificados como violentos, figura 44.

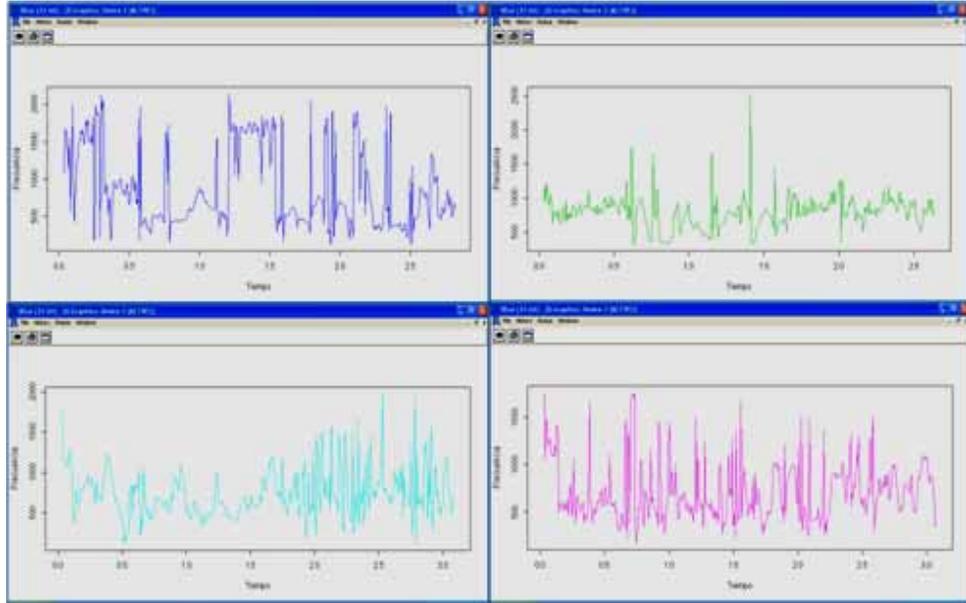


Figura 44. Ejemplos de gráficas del parámetro frecuencia (violencia).

De las gráficas obtenidas, se aplicó el algoritmo DTW con la finalidad de encontrar una señal que pueda ser utilizada como patrón para representar una situación violenta.

#### A) Peor de los casos

De las comparaciones realizadas el caso con menor similitud entre dos señales comparadas fue el que se presenta en la figura 45, donde la región de concordancia fue prácticamente nula, el valor del grado de similitud encontrado fue de 94.86983.

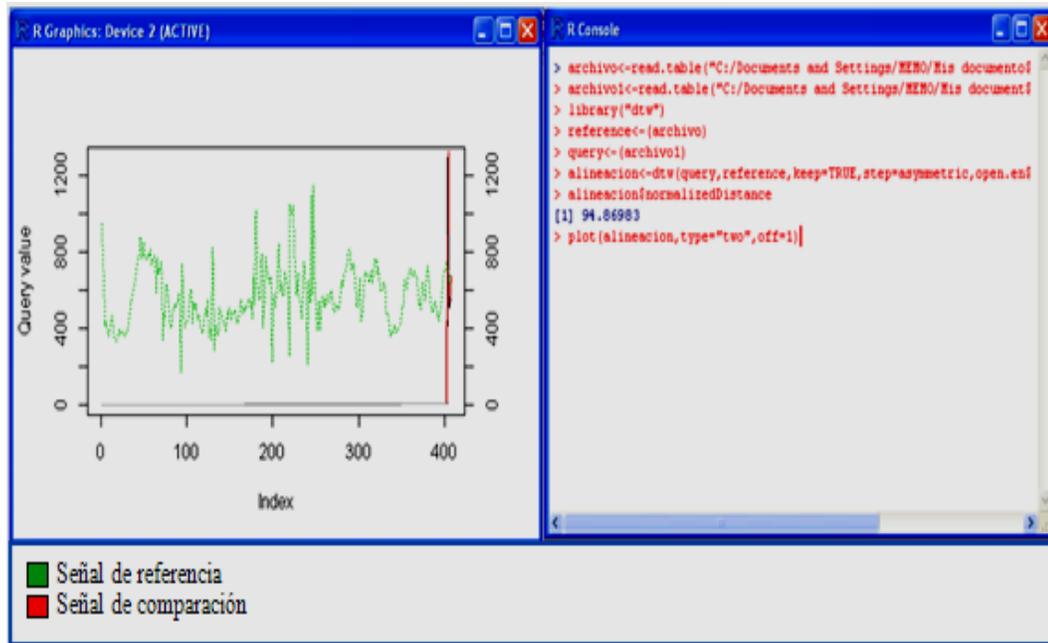
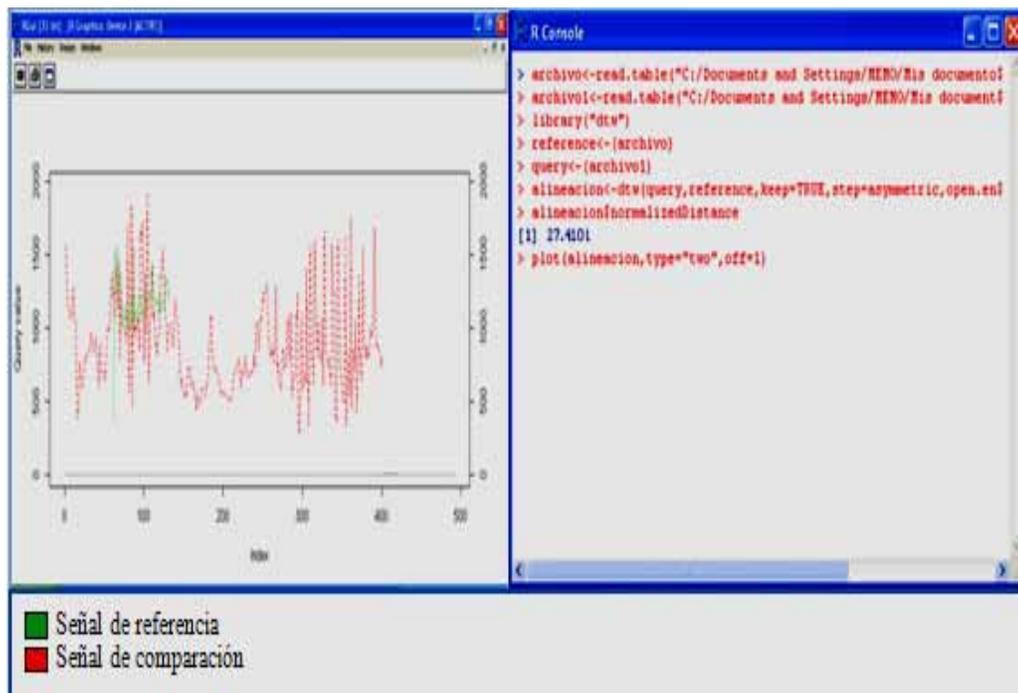


Figura 45 .Gráfica y grado de concordancia de la frecuencia en situaciones violentas (Peor caso).

### B) Mejor de los casos

Algunos de los resultados obtenidos, se presentan en las figuras 46 a 48.



reporteFigura 46. Obtención de la señal de referencia del parámetro frecuencia, ejemplo 1.

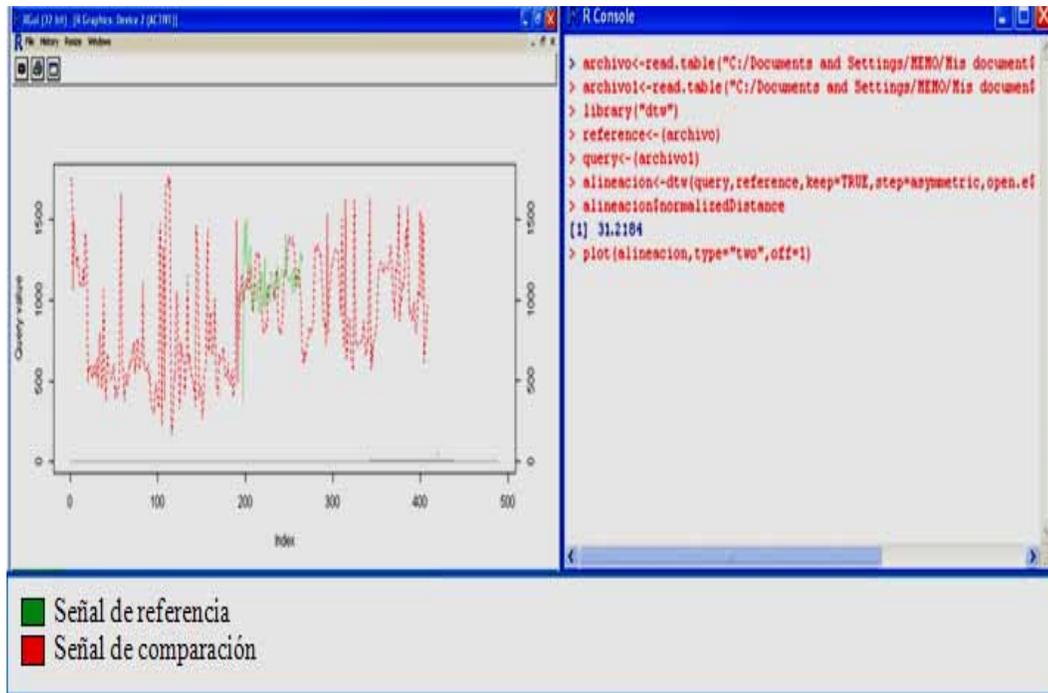


Figura 47. Obtención de la señal de referencia del parámetro frecuencia, ejemplo 2.

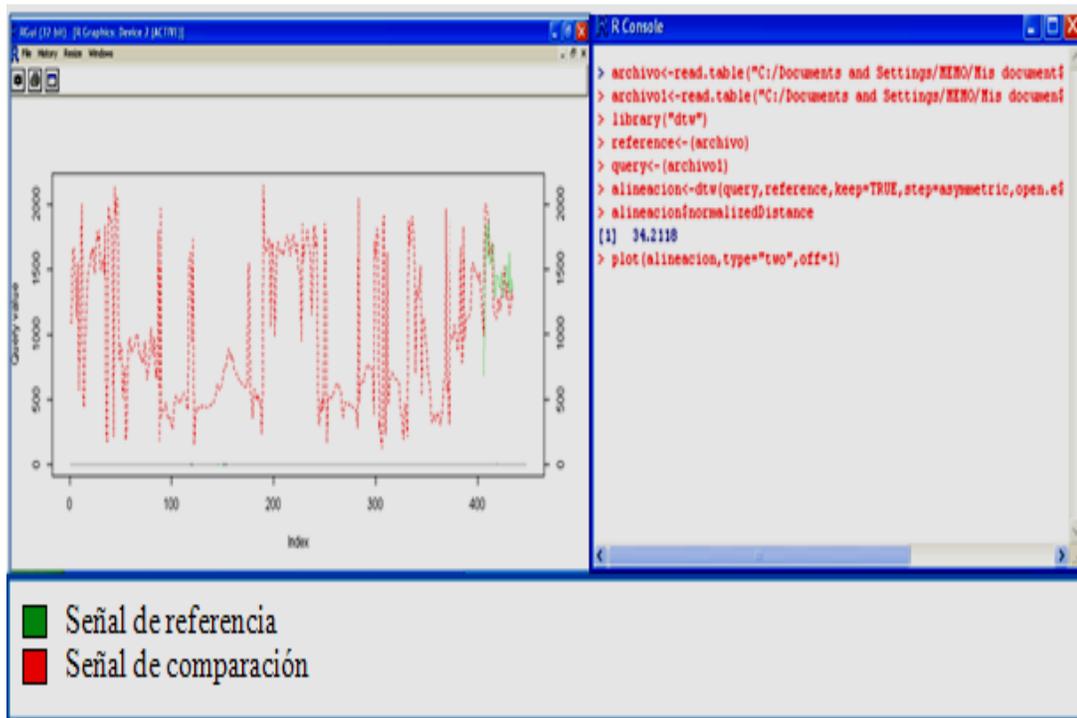


Figura 48. Obtención de la señal de referencia del parámetro frecuencia, ejemplo 3.

En la figura 46 el valor obtenido del grado de similitud fue de 27.4101, la segunda de 31.2184 y la última de 34.2118. Debido a estos valores obtenidos se decidió elegir a la señal en color

verde con un grado de similitud de 27.4101 como la señal de referencia para la frecuencia ante una situación violenta.

### 6.6.2 Situación normal

El siguiente paso, fue encontrar la señal que representara al parámetro frecuencia en una situación no violenta. El proceso fue el mismo, primero se obtuvieron las señales de cada uno de los archivos clasificados como no violentos, posteriormente se aplicó el algoritmo DTW para encontrar la señal de referencia, a continuación veremos los dos siguientes casos.

#### A) Peor de los casos

Los resultados obtenidos se tienen en la figura 49, con un grado de similitud de 93.8694 el cual representa el valor más alto encontrado y por ende el peor de los casos.

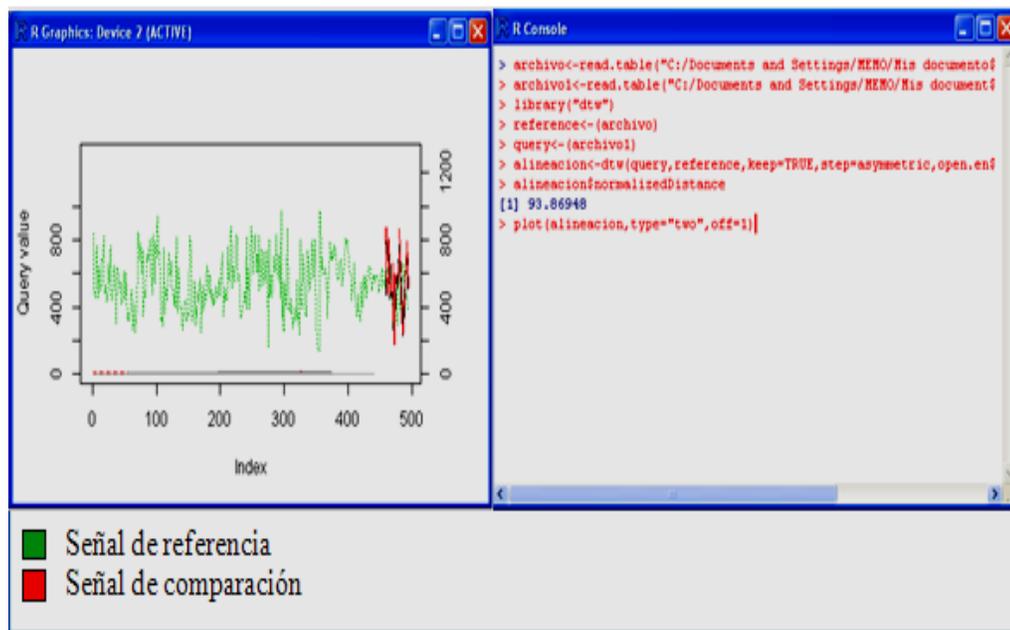


Figura 49 .Gráfica y grado de concordancia de la frecuencia en situaciones normales (Peor caso).

#### B) Mejor de los casos

Los mejores resultados obtenidos son los que se ven en las figuras 50 a 52, donde se ve la región de concordancia que presentan, la señal en color verde representa la señal elegida como referencia y la de color rojo son señales con las que fue comparada.

Los valores en el grado de concordancia entre estas señales son de 31.2184, 32.8141 y 33.2118 respectivamente, estos resultados fueron el fundamento para elegir a esta señal como referencia

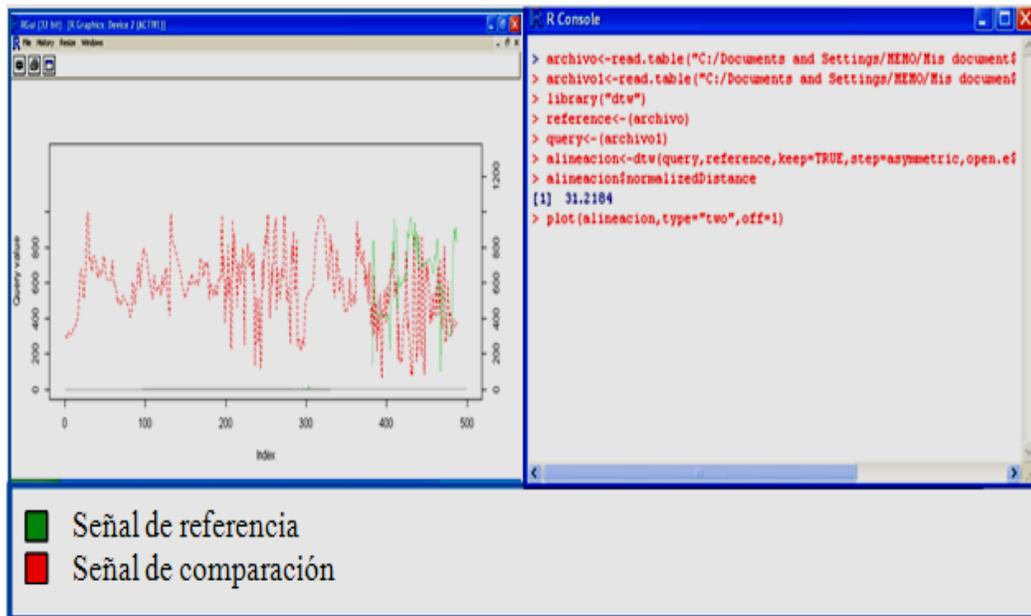


Figura 50. Obtención de la señal de referencia del parámetro frecuencia, ejemplo 1.

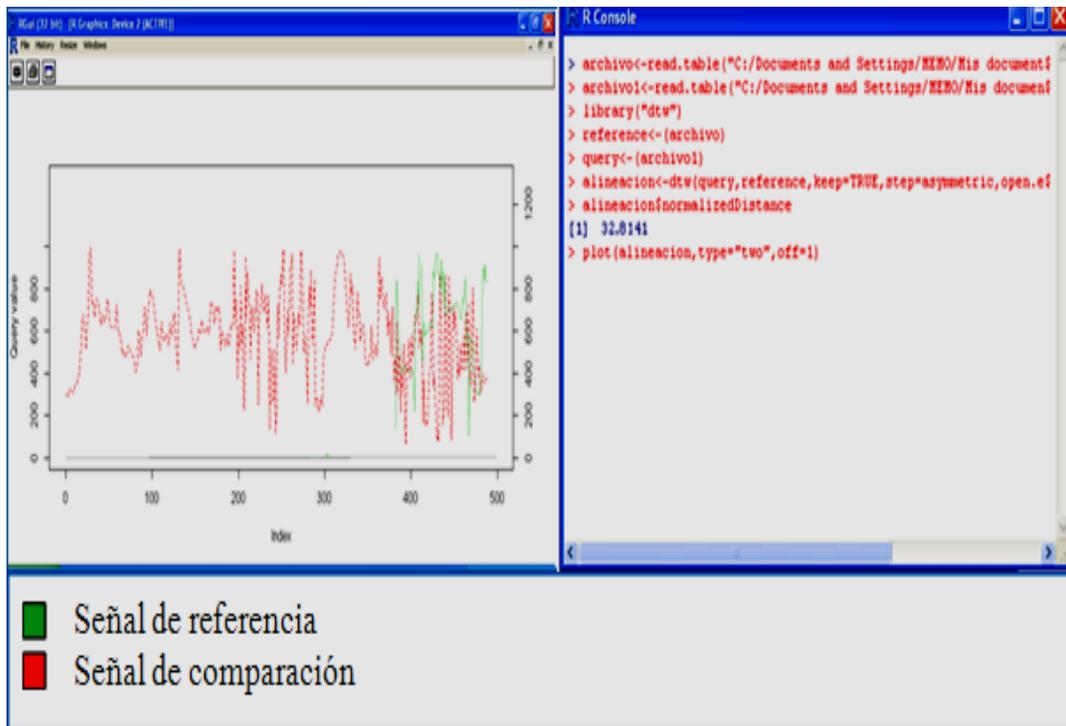


Figura 51. Obtención de la señal de referencia del parámetro frecuencia, ejemplo 2.

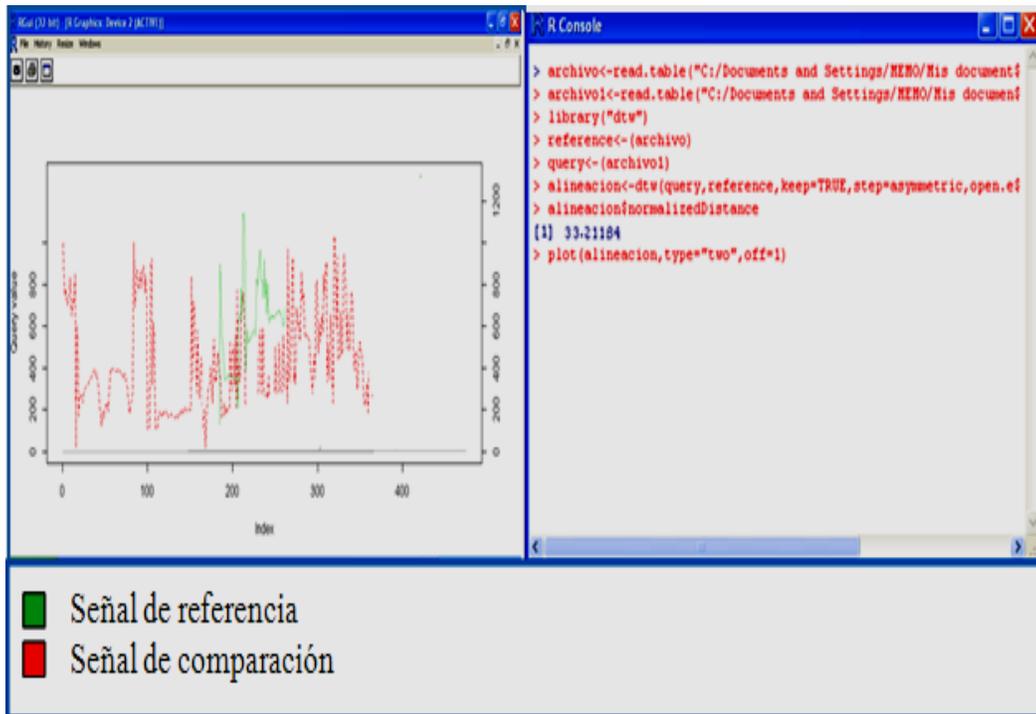


Figura 52. Obtención de la señal de referencia del parámetro frecuencia, ejemplo 3.

## 7 Resultados (presentación y análisis)

Con el fin de lograr el objetivo general planteado al inicio de este proyecto de investigación, e identificar qué parámetro(s) de la voz (intensidad, tono, amplitud y frecuencia) era crucial para detectar una situación de violencia, se realizó el análisis de cada archivo de audio recopilado en formato .wav, comparando todas las señales generadas por parámetro, de manera que al final del proceso se obtuvo una señal representativa con el mejor índice de concordancia y una señal con el peor índice de concordancia, considerado a partir del conjunto de archivos en estudio.

Esta sección presenta un resumen de los resultados obtenidos durante el desarrollo del proyecto de investigación, dando a conocer las señales de referencia características de cada parámetro de la voz (intensidad, tono, amplitud y frecuencia) en una situación de violencia y en una situación normal. La señal de referencia propia de cada parámetro de la voz se esquematiza en una gráfica de color verde; así mismo, se da a conocer el índice de concordancia correspondiente a la señal de referencia obtenida y el intervalo en cual se estuvo manifestando cada parámetro de la voz.

Definimos al índice de concordancia como la distancia existente entre dos señales al ser comparadas entre sí. A menor distancia cuantificada, mayor fue la similitud de las señales comparadas, es decir tienden a ser idénticas, sobreponiéndose entre sí mismas. El valor ideal del índice de concordancia es cero.

El criterio utilizado para definir si una señal podía considerarse como referencia fue el índice de concordancia. De todos los casos evaluados se consideró como señal de referencia aquella que diera el índice de concordancia más pequeño (el más cercano a cero), considerado este a partir de la población de videos evaluados.

A continuación se presentan los resultados obtenidos para cada uno de los parámetros evaluados en situaciones de violencia y en situaciones normales.

### 7.1 Resultados obtenidos del parámetro intensidad, situación violenta.

En términos de intensidad de la voz para una situación de violencia referida a la población en estudio, el algoritmo DTW identificó que la señal de referencia característica es la que se presenta en la figura 53, donde se presenta un índice de concordancia de 6.06569 siendo este el valor más pequeño obtenido de todos los videos analizados.

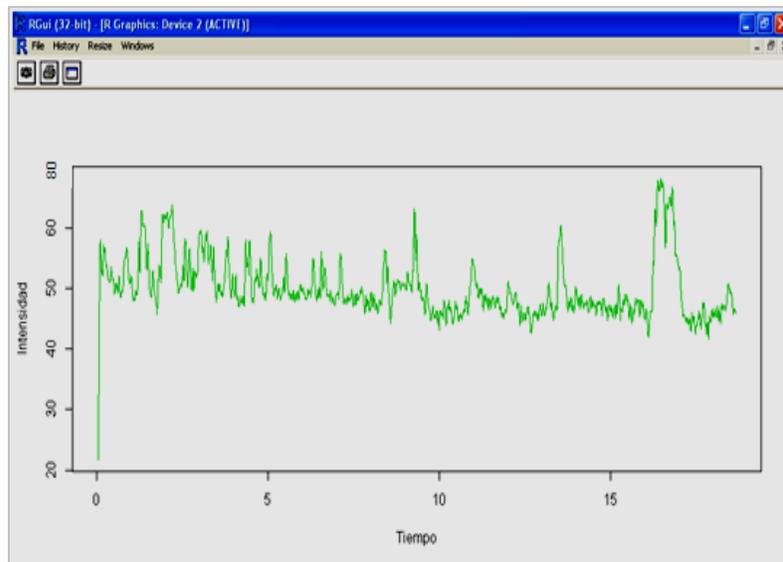


Figura 53. Grafica de referencia de una situación de violencia, parámetro intensidad.

Analizando todos los casos, y en función a la señal de referencia encontrada, la intensidad de la voz en situaciones violentas, está de los 40-70 dB.

## 7.2 Resultados obtenidos del parámetro intensidad, situación normal

La señal de referencia para el parámetro intensidad en situaciones normales se presenta en la figura 54.

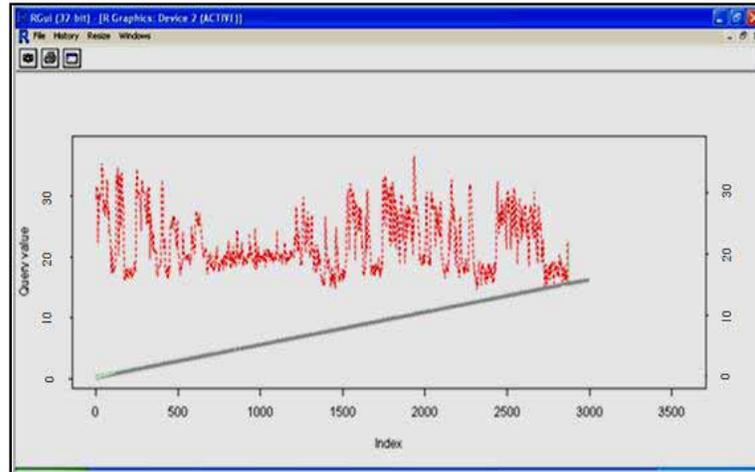


Figura 54. Señal de referencia de una situación normal, parámetro intensidad.

La señal de referencia tiene un índice de concordancia de 9.35496 para situaciones normales y en función a la figura 54, la voz para una situación normal tiene una intensidad de 15 – 35 dB.

## 7.3 Resultados obtenidos del parámetro tono, situación violenta.

El resultado obtenido después de realizar las comparaciones por medio del algoritmo DTW, muestran que la gráfica que representa la señal de referencia del parámetro tono en una situación de violencia, es la de la figura 55.

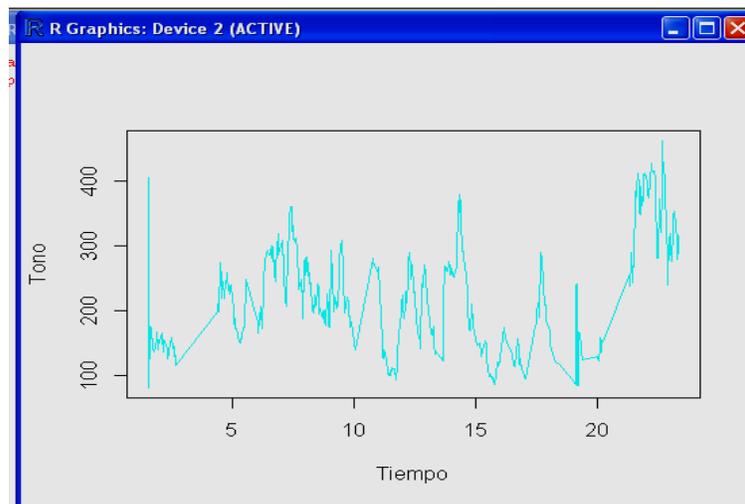


Figura 55. Señal de referencia de una situación de violencia, parámetro tono.

El índice de concordancia obtenido fue de 10.4217

El intervalo en el que está definida la señal de referencia, va de los 90-400 Hz; sin embargo, la mayor concentración de la señal se encuentra, entre los 90-350 Hz.

Tomando como base esta observación, se concluye que una situación de violencia en base al tono de la voz, se presenta de los 90 – 350 Hz.

#### 7.4 Resultados obtenidos del parámetro tono situación normal

La señal de referencia del tono de la voz que corresponde a una situación normal se representa en la figura 56. El índice de concordancia obtenido para esta señal fue de 25.601.

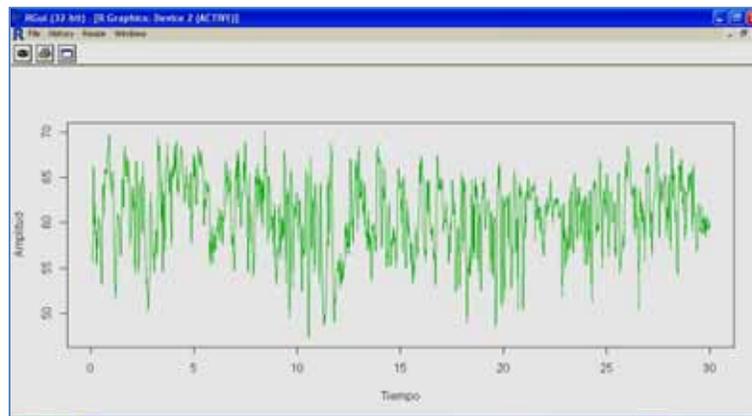


Figura 56. Señal de referencia de una situación normal, parámetro tono.

De acuerdo a la figura 56, el tono de la voz en una situación normal va de los 47-69 Hz.

#### 7.5 Resultados obtenidos del parámetro amplitud, situación violenta.

Como resultado del análisis realizado a los videos clasificados como situaciones violentas, mediante el algoritmo DTW se obtuvo que la señal que servirá como referencia para la amplitud de la voz ante una situación violenta es la que se observa en la figura 57.

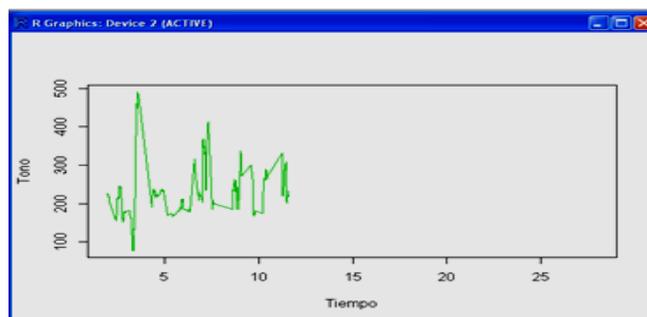


Figura 57. Señal de referencia de una situación violenta, parámetro amplitud.

La señal de la figura 57 fue elegida como referencia porque al realizar la comparación de las señales tuvo una mayor región de similitud, presentando un índice de concordancia de 13.18279. El intervalo dentro del cual está definida la amplitud de la voz es de los 90-480 dB, aproximadamente.

### **7.6 Resultados obtenidos del parámetro amplitud, situación normal.**

Tras una serie de comparaciones entre señales que estaban clasificadas manualmente como situaciones normales por medio del algoritmo DTW, el resultado obtenido fue la señal que representa la amplitud de la voz en una situación normal, figura 58.

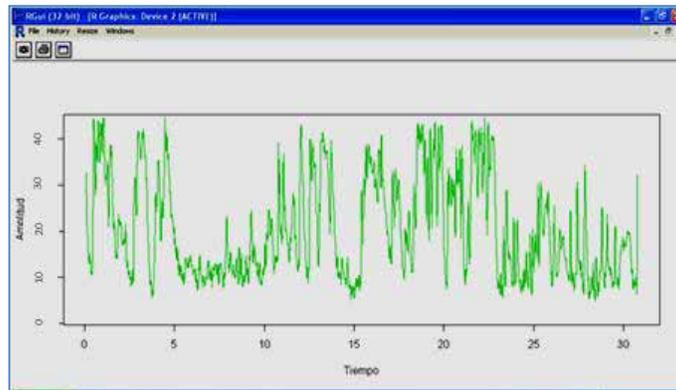


Figura 58. Señal de referencia de una situación normal del parámetro amplitud.

La señal de referencia de la amplitud de la voz en una situación normal tuvo un índice de concordancia de 23.9734.

La figura 58 indica que para la población de videos analizados, la amplitud de la voz en una situación normal va de los 5-45dB, aproximadamente.

### **7.7 Resultados obtenidos del parámetro frecuencia, situación violenta.**

La señal representativa de la frecuencia de la voz en una situación violenta se esquematiza en la figura 59 donde se observa un pico que está por arriba de los 2000 Hz, sin embargo se descarta por considerarse que la señal se encuentra mejor definida en el intervalo de los 800 - 1700 Hz, siendo éste el intervalo característico de la frecuencia de la voz en situaciones violentas.

La señal de referencia para la frecuencia de la voz en situaciones violentas tiene un índice de concordancia de 27.4101

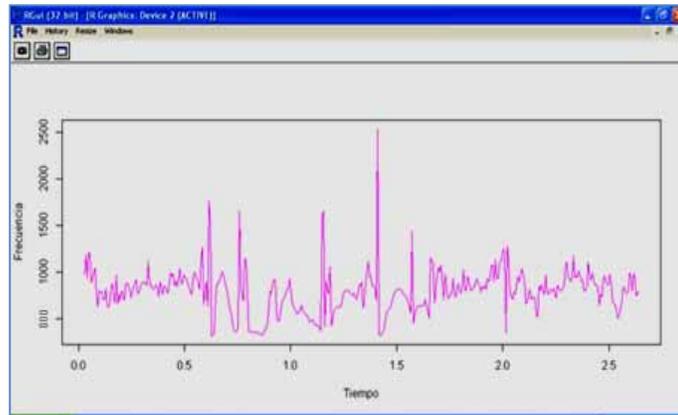


Figura 59. Señal de referencia de una situación violenta del parámetro frecuencia.

### 7.8 Resultados obtenidos del parámetro frecuencia situación normal.

La figura 60, muestra la señal representativa que tiene la frecuencia de la voz en una situación normal donde se muestra que la frecuencia de la voz en una situación normal, se encuentra en el intervalo de los 300-1300 Hz, aproximadamente.

La señal de referencia representativa de la frecuencia de la voz en una situación normal tiene un índice de concordancia de 31.2184

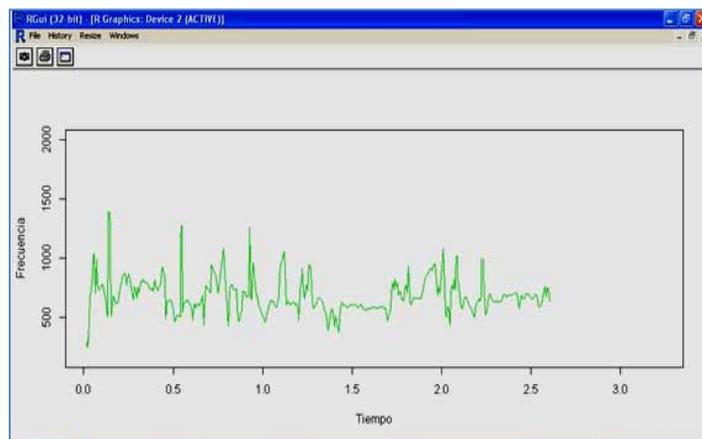


Figura 60. Señal de referencia de una situación normal del parámetro frecuencia.

Las tablas 2 y 3 son un resumen de los resultados obtenidos para cada parámetro de la voz evaluado.

Tabla 2. Resultados obtenidos en una situación violenta.

Situación evaluada	Parámetro	Índice de concordancia		Intervalo de identificación
		Peor índice de concordancia	Óptimo índice de concordancia (Señal de referencia)	
Violencia	Intensidad	33.9407	6.06569	40-70 dB
	Tono	61.2456	10.4217	90-350 Hz
	amplitud	58.2591	13.18279	90-480 dB
	Frecuencia	94.86983	27.4101	800 -1700 Hz

Tabla 3. Resultados obtenidos en una situación normal.

Situación evaluada	Parámetro	Índice de concordancia		Intervalo de identificación
		Peor índice de concordancia	Óptimo índice de concordancia (Señal de referencia)	
Normal	Intensidad	24.9733	9.35496	15-35 dB
	Tono	84.2113	25.601	47-69 Hz
	amplitud	82.3114	23.9734	5-45 dB
	Frecuencia	93.8694	31.2184	300-1300Hz

De todos los parámetros evaluados de la voz, se presenta una situación no esperada en el caso de la frecuencia, porque se da una superposición de los valores registrados en la situación de violencia y en la situación normal, la explicación a esta eventualidad se debe a que la voz modula en un amplio espectro de frecuencias y al haber hecho una selección aleatoria de los videos; es decir, en los videos intervenían voces femeninas, masculinas y de niños ocasionando este desequilibrio en el análisis.

## 8. Conclusiones

La violencia es un término subjetivo, difícil de medir, porque está en función de rasgos culturales, y la apreciación de los hechos varía de un ser humano a otro.

La cultura y educación que recibe cada individuo marca la diferencia para tipificar a una situación violenta, de ahí la importancia de identificar los rasgos característicos que pudieran ser indicativos de una situación violenta.

En la actualidad, la única violencia medible, es la violencia física, encontrar y proponer un medio para identificar la violencia no física fue una tarea ardua y parte de este proyecto de investigación.

La propuesta del diseño de un software inteligente, capaz de identificar y diferenciar entre una situación violenta y una de no violencia a través de la voz condujo a decir que la voz, es un sonido tan complejo, que no es posible definir una situación de violencia con solo una de sus características o parámetros (intensidad, tono, amplitud, frecuencia), razón por la cual se requiere de la combinación de más de una de ellas, por lo tanto y para nuestro caso en estudio se concluye que en el diseño e implementación del software inteligente se haga uso de la combinación de la intensidad, tono, y amplitud de la voz para identificar una situación de violencia. Se descarta el uso de la frecuencia de la voz como un indicativo de violencia porque la voz modula a diferentes frecuencias y depende de la edad y el sexo de la persona lo que dificultaría la interpretación del tipo de situación que se esté manifestando.

De acuerdo a la investigación desarrollada, la voz, en sus características de intensidad, tono, amplitud y frecuencia, para una situación de violencia presenta el siguiente comportamiento, tabla 4.

Tabla 4: Características de la voz en una situación violenta

Característica de la voz	Intervalo de identificación
Intensidad	40-70 dB
Tono	90-350 Hz
Amplitud	90-480 dB
Frecuencia	800 -1700 Hz

De acuerdo a los resultados obtenidos como parte de la investigación desarrollada, al medir el nivel de efectividad del software propuesto para detectar una situación de violencia a través de los parámetros intensidad, tono, amplitud y frecuencia de la voz se identificó que tiene una

eficiencia del 87%, determinado a través de una prueba experimental, por comparación entre el índice de concordancia de la señal de referencia para cada una de las características de la voz y el índice de concordancia que se obtuvo del archivo experimental. El porcentaje obtenido se considera aceptable, tomando en consideración que el término violencia es subjetivo.

La proyección a futuro de esta propuesta puede estar encaminada a instalar el software en micrófonos ambientales ubicados en lugares públicos a fin de monitorear el comportamiento de la población e identificar de manera rápida una situación de violencia.

## 9. Referencias bibliográficas.

[1] Definición de violencia.

[http://apps.who.int/iris/bitstream/10665/67411/1/a77102\\_spa.pdf](http://apps.who.int/iris/bitstream/10665/67411/1/a77102_spa.pdf)

Visitado el 14 de agosto del 2014

[2] Definiciones de violencia y características.

<http://urbanres.blogspot.mx/2009/02/violencia-definicion-y-caracteristicas.html>

Visitado el 14 de agosto del 2014

[3] Definiciones de violencia.

<http://definicion.de/violencia/>

Visitado el 16 de agosto del 2014

[4] Definición de violencia

[http://www.scielo.org.mx/scielo.php?script=sci\\_arttext&pid=S0188-77422009000200002#notas](http://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S0188-77422009000200002#notas)

Visitado el 16 de agosto del 2014

[5] Definición de violencia

[http://www.congresoestudioviolencia.com/2010/image/articulos\\_cientifico.pdf](http://www.congresoestudioviolencia.com/2010/image/articulos_cientifico.pdf)

Visitado el 16 de agosto del 2014

[6] Definición de violencia

<http://www.definicionabc.com/social/violencia.php>

Visitado el 17 de agosto del 2014

[7] Definición de voz

[http://otorrinopamplona.com/?page\\_id=1548](http://otorrinopamplona.com/?page_id=1548)

Visitado el 17 de agosto del 2014

[8] Definición de intensidad.

<http://www.educarchile.cl/ech/pro/app/detalle?id=223728>

Visitado 23 de agosto del 2014

[9] Definición de tono.

<http://riunet.upv.es/bitstream/handle/10251/32683/Memoria.pdf?sequence=1>

Visitado 23 de agosto del 2014

[10] Definición de amplitud.

<http://blogmazuera.wordpress.com/marco-teorico-2/>

Visitado 23 de agosto del 2014

[11] Definición de frecuencia.

<http://www.profesorenlinea.com.mx/fisica/SonidoOndas.htm>

Visitado 23 de agosto del 2014

[12] Artículo sobre el algoritmo DTW

Toni Giorgino, Computing and Visualizing Dynamic Time Warping Alignments in R: The dtw Package.

[13] Definición del software Atube Catcher.

<http://atube-catcher.softonic.com/>

Visitado el 24 de agosto del 2014

[14] Sitio oficial de Praat

<http://www.praat.org>

Visitado el 24 de agosto del 2014

[15] Sitio oficial de r-project

<http://www.r-project.org>

Visitado el 24 de agosto del 2014

[16] Características de R

<http://docencia.udea.edu.co/cci/linux/dia9/r.htm>

Consultado 30 de agosto

[17]. Extensiones y paquetes de R

<http://www.r-bloggers.com/lang/uncategorized/309>

Consultado 30 de agosto

## Apéndice A

### Código fuente del proyecto de investigación.

```
archivo1<-read.table("C:/Documents and Settings/MEMO/Mis
documentos/SeminarioTitulacion/comisariadefamilia.txt")

plot(archivo1,type="l",xlab="Tiempo",ylab="Intensidad",col=4)

archivo2<-read.table("C:/Documents and Settings/MEMO/Mis
documentos/SeminarioTitulacion/asidelprecipicio.txt")

plot(archivo2,type="l",xlab="Tiempo",ylab="Intensidad",col=6)

archivo3<-read.table("C:/Documents and Settings/MEMO/Mis
documentos/SeminarioTitulacion/bala.txt")

plot(archivo3,type="l",xlab="Tiempo",ylab="Intensidad",col=7)

archivo4<-read.table("C:/Documents and Settings/MEMO/Mis
documentos/SeminarioTitulacion/cacheteabarbara.txt")

plot(archivo4,type="l",xlab="Tiempo",ylab="Intensidad",col=8)

archivo5<-read.table("C:/Documents and Settings/MEMO/Mis
documentos/SeminarioTitulacion/calamidades.txt")

plot(archivo5,type="l",xlab="Tiempo",ylab="Intensidad",col=9)

archivo6<-read.table("C:/Documents and Settings/MEMO/Mis
documentos/SeminarioTitulacion/campañan_violencia.txt")

plot(archivo6,type="l",xlab="Tiempo",ylab="Intensidad",col=10)

archivo7<-read.table("C:/Documents and Settings/MEMO/Mis
documentos/SeminarioTitulacion/cielorojo.txt")

plot(archivo7,type="l",xlab="Tiempo",ylab="Intensidad",col=11)

archivo8<-read.table("C:/Documents and Settings/MEMO/Mis
documentos/SeminarioTitulacion/cielorojomentiraspatricia.txt")

plot(archivo8,type="l",xlab="Tiempo",ylab="Intensidad",col=12)

archivo<-read.table("C:/Documents and Settings/MEMO/Mis
documentos/SeminarioTitulacion/dtw.txt")

archivo1<-read.table("C:/Documents and Settings/MEMO/Mis
documentos/SeminarioTitulacion/aguaoleche.txt")
```

```
library("dtw")  
ref<-(archivo1)  
test<-(archivo)  
alineacion<-dtw(ref,test)  
alineacion$normalizedDistance  
plot(alineacion,type="two",off=1)
```