

TESIS DOCTORAL

DOCTORADO EN CIENCIAS DE LA TIERRA

**SISTEMA DE DETECCIÓN Y CLASIFICACIÓN DE
SEÑALES SÍSMICO-VOLCÁNICAS UTILIZANDO
MODELOS OCULTOS DE MARKOV (HMMs):
APLICACIÓN A VOLCANES ACTIVOS DE
NICARAGUA E ITALIA.**



UNIVERSIDAD DE GRANADA

Ligdamis Anaxis Gutiérrez Espinoza

Año 2013

Editor: Editorial de la Universidad de Granada
Autor: Ligdamis Anaxis Gutiérrez Espinoza
D.L.: GR 518-2014
ISBN: 978-84-9028-827-6

Foto de la Portada: Volcán Masaya. Nicaragua
Por: Ligdamis A. Gutiérrez E.

**INSTITUTO ANDALUZ DE GEOFÍSICA Y
PREVENCIÓN DE DESASTRES SÍSMICOS.
DEPARTAMENTOS DE FÍSICA TEÓRICA Y
DEL COSMOS. TEORÍA DE LA SEÑAL,
TELEMÁTICA Y COMUNICACIONES.
UNIVERSIDAD DE GRANADA**



M^a Del Carmen Benítez Ortúzar, Profesor Titular de la Universidad de Granada y **Jesús M. Ibáñez Godoy**, Catedrático de Universidad, de la Universidad de Granada.

Hacen Constar:

Que la presente memoria titulada “*Sistema de detección y clasificación de señales sísmico-volcánicas utilizando Modelos Ocultos de Markov (HMMs): Aplicación a volcanes activos de Nicaragua e Italia*”. Ha sido realizada bajo nuestra dirección por: **D. Ligdamis Anaxis Gutiérrez Espinoza** y cumple con las condiciones para que su autor pueda optar al grado de Doctor en Ciencias de la Tierra, con línea de investigación, Sismología y Geofísica.

En Granada, Septiembre de 2013

FDO. M^a Del Carmen Benítez Ortúzar

Jesús M. Ibáñez Godoy



Universidad de Granada, 2013

Compromiso de respeto de los derechos de autor

El doctorando **D. Ligdamis Anaxis Gutiérrez Espinoza** y los directores de la tesis: **D^a M^a Del Carmen Benítez Ortúzar** y **D. Jesús M. Ibáñez Godoy**. Garantizamos, al firmar esta tesis doctoral, que el trabajo ha sido realizado por el doctorando bajo la dirección de los directores de la tesis y hasta donde nuestro conocimiento alcanza, en la realización del trabajo, se han respetado los derechos de otros autores a ser citados, cuando se han utilizado sus resultados o publicaciones.

En Granada, Septiembre de 2013

Director/es de la Tesis

Doctorando

Fdo.: M^a Del Carmen Benítez Ortúzar

Fdo.: Ligdamis Anaxis Gutiérrez Espinoza

Fdo.: Jesús M. Ibáñez Godoy

*A Tututo e Itaita
A mi Familia*

*Mi Dios a su vez, proveerá a todas sus necesidades, según su inmensa riqueza en Cristo
Jesús.*

Filipenses 4:19

*Si me ofreciesen la sabiduría con la condición de guardarla para mí sin comunicarla a nadie,
no la querría.*

Séneca

ÍNDICE

Agradecimientos	1
Prólogo	3
Abstract Extendido	5
Extended Abstract	11
PRIMERA PARTE: INTRODUCCIÓN, ANTECEDENTES Y PRESENTACIÓN DEL CASO DE ESTUDIO	15
CAPÍTULO 1.- Introducción	17
1.1.- Impacto de terremotos y volcanes sobre la sociedad	19
1.2.- Las señales	22
1.2.1 Tipos de señales	22
1.2.2 El problema de la clasificación de las señales	23
1.2.3 Señales analógicas y digitales	23
1.2.4 Procesamiento de señales	25
1.3.- Reconocimiento y clasificación de patrones	26
1.4.- Ámbitos relacionados con el reconocimiento sísmico	28
1.5.- Los volcanes y sus manifestaciones	29
1.6.- Señales sísmicas producidas en volcanes	32
1.6.1.- Tipos de eventos sísmicos-volcánicos	33
1.6.2.- Problemas de reconocimiento de las señales sísmico-volcánicas	37
1.7.- Familia de eventos sísmico-volcánicos	39
CAPÍTULO 2.- Metodología y Herramientas	41
2.1 Introducción	43
2.2 La transformada de Fourier y el Espectrograma	43
2.2.1 La transformada de Fourier	43
2.2.2 El Espectrograma	45
2.3 Descripción de la señal en el dominio cepstral	46
2.4 Los Modelos Ocultos de Markov	48
Aplicación a un sistema de un sistema de reconocimiento	48
2.4.1 Definición de un HMM	48
2.4.2 HMM como generador de secuencias de observaciones	49
2.4.3 Los tres problemas básicos y soluciones propuestas	50
2.5 Esquema de un sistema de reconocimiento basado en HMM	51
2.6 Implementación con HTK	52
2.6.1.- Descripción de HTK	52
2.6.2.- Interfaz de usuario	53
2.6.3.- Preparación de los datos	54
2.6.4.- Entrenamiento	56
2.6.5.- Reconocimiento	58
2.6.6- Análisis (Evaluación de los resultados)	58

SEGUNDA PARTE: CAMPAÑAS, EMPLAZAMIENTOS Y DATOS	61
CAPÍTULO 3.- Zonas de estudio, Instrumentos y Campañas	63
3.1.- Introducción	65
3.2.- Marco Volcánico	67
3.2.1.- El volcán Etna	67
3.2.2.- El volcán Strómboli.....	69
3.2.3.- El volcán San Cristóbal	70
3.2.4.- El volcán Telica.....	72
3.3.- Estaciones sísmicas e instrumentos de medición en los volcanes de estudio	73
3.3.1.- Tipos de estaciones sísmicas	73
3.3.2.- Estaciones e instrumentos en el Etna	75
3.3.3.- Estaciones e instrumentos en el Strómboli.....	80
3.3.4.- Elección de los sitios de estudio en Nicaragua	83
3.3.5.- Estaciones e instrumentos en el San Cristóbal	85
3.3.6.- Estaciones e instrumentos en el Telica.....	89
3.4.- Transformación y unificación de formatos de los datos.	93
CAPÍTULO 4.- Análisis de los Datos	95
4.1.- Introducción	97
4.2.- Determinación de los tipos de eventos.....	97
4.2.1.- Tipos de eventos del volcán Etna.....	99
4.2.2.- Tipos de eventos del volcán Strómboli	100
4.2.3.- Tipos de eventos del volcán San Cristóbal	101
4.2.4.- Tipos de eventos del volcán Telica	105
4.3.- Análisis espectral de los eventos	109
4.3.1.- Análisis mediante la transformada rápida de Fourier.....	109
4.3.2.- Análisis de los espectrogramas de los eventos.....	114
4.4.- Estadísticas de la duración de los eventos.....	126
4.4.1.- Estadísticas en el volcán Etna	126
4.4.2.- Estadísticas en el volcán Strómboli.....	126
4.4.3.- Estadísticas en el volcán San Cristóbal	127
4.4.4.- Estadísticas en el volcán Telica.....	127
4.5.- Proceso de etiquetado manual de la base de datos	130
CAPÍTULO 5.- Aplicación del método y resultados	135
5.1 Introducción.....	137
5.2 Desarrollo de las pruebas en los volcanes	138
5.2.1.- Análisis de resultados en el Etna.....	138
5.2.2.- Análisis de resultados en el Strómboli	152
5.2.3.- Análisis de resultados en el San Cristóbal	164
5.2.4.- Análisis de resultados en el Telica	177
5.3.- Pruebas cruzadas con las bases de datos	188
5.3.1.- Pruebas cruzadas entre el Etna y el Strómboli	188
5.3.2.- Pruebas cruzadas entre el San Cristóbal y el Telica.....	191
5.3.3.- Resultado de las pruebas cruzadas con los cuatro volcanes	

(Etna, Strómboli, San Cristóbal, Telica).....	193
5.4.- Discusión de resultados.....	195
CAPÍTULO 6.- Conclusiones y trabajos a futuro	203
6.1. - Conclusiones y logros	205
6.1.1.- Conclusiones Metodológicas	205
6.1.2.- Conclusiones Experimentales	206
6.2. - Trabajos a futuro	209
6.3. - Conclusions and Successes	211
6.3.1 Methodological conclusions	211
6.3.2 Experimental conclusions.....	212
6.4. - Future Works.....	215
Bibliografía.....	217
Listado de Figuras	233
Listado de tablas	245
APÉNDICES	249
Apéndice A: Referencias a trabajos publicados	251
Apéndice B: Ámbito sísmico-volcánico de Centro América y Nicaragua.....	255
Apéndice C: Historia eruptiva y actividad sísmico-volcánica durante el período de recopilación de datos	273
Apéndice D: Herramientas complementarias en el análisis de señales	293
Apéndice E: Aplicaciones a las señales sísmicas de las herramientas complementarias de análisis de poder espectral y componentes Wavelet.....	299
Apéndice F: Procedimiento de conversión de Formatos.....	339
Apéndice G: Implementación Tecnológica	351
Apéndice H: Programas.....	369
Apéndice I: Formulario	383

AGRADECIMIENTOS

Esta tesis doctoral ha sido parcialmente financiada por los siguientes proyectos: Spanish project Epeostos, CGL2011-29499-C02-01, by the EU project EC-FP7 Mediterranean Supersite Volcanoes (MED-SUV), el proyecto regional “Grupo de Investigación en Geofísica y Sismología de la Junta de Andalucía, RNM104”, y el proyecto Algoritmos avanzados de procesamiento de señal para reconocimiento y caracterización de señales sísmo volcánicas (APASVO), TEC2012-315511 del Grupo de Investigación TIC-123, de la Junta de Andalucía.

A los proyectos: RNM-217 "Riesgo Sísmico y Microzonificación" del grupo de investigación de la Junta de Andalucía, cuyo investigador principal es el Dr. Carlos López Casado y El proyecto CGL2007-60535/BTE "Tectónica Activa en las Béticas. Peligrosidad Sísmica y Movimientos de Ladera cuyo investigador responsable es el Dr. Carlos López Casado. Con las diversas campañas de sismicidad que han sido subvencionadas con fondos del grupo de investigación y de los proyectos antes citados, en las que he participado con los conocimientos en instrumentación e informática y que me han permitido, financiar en parte mi estancia en Granada.

A la concesión de la beca MAE II de la Agencia Española de Cooperación Internacional (AECI), por haber hecho posible el llegar a Granada y financiar mi estancia y estudios de doctorado en la Universidad de Granada durante los períodos 2003 – 2005. A los proyectos NOVAC – Network for Observation of volcanic and atmospheric change (Monitoreo de Gases volcánicos) y TUCAN (Tucán Seismic Experiment), red sísmica temporal de banda ancha. De Suecia y USA respectivamente en colaboración con el INETER, por cuya participación fue posible la financiación de mi estancia durante 2006 y 2007 en el INETER, en las campañas de Nicaragua.

A mis directores de tesis, Dra. Carmen Benítez y Dr. Jesús Ibáñez, que me brindaron la oportunidad de realizar esta tesis, así como sus consejos, paciencia y apoyo. Al Dr. Carlos López Casado, Catedrático de Universidad, cuyo apoyo y ánimo fue fundamental para iniciar la etapa del Master en Geofísica y Meteorología, y al Dr. Fco. Javier Almendros González, quien fue mi tutor de Master durante el mismo período. Al Dr. Wilfried Strauch, la Lic. Virginia Tenorio, y el Lic. Emilio Talavera; de la Dir. General de Geofísica del INETER, quienes me brindaron su apoyo y ayuda, para que fuera posible el éxito de la recopilación de los datos en las campañas de Nicaragua. Al Dr. Javier Ramírez Pérez de Inestrosa, Catedrático de Universidad, quien supo orientarme y animarme en los inicios del doctorado y durante el Master en Tecnologías Multimedia.

Personalmente un párrafo bíblico, resume el camino que he recorrido hasta ahora, y el que aún falta por recorrer, un texto que manifiesta que realmente no hay valores totalmente absolutos en esta vida, pero que hay un tiempo para cada cosa.

*Hay un tiempo bajo el sol, un momento para todo,
Y un tiempo para hacer cada cosa:
Tiempo para nacer, y tiempo para morir;
Tiempo para plantar, y tiempo para arrancar lo plantado;
Tiempo para matar y tiempo para curar;
Tiempo para demoler, y tiempo para edificar;*

*Tiempo para llorar y tiempo para reír;
Tiempo para gemir, y tiempo para bailar;
Tiempo para lanzar piedras, y tiempo para recogerlas;
Tiempo para los abrazos, y tiempo para abstenerse de ellos;
Tiempo para buscar, y tiempo para perder;
Tiempo para conservar, y tiempo para tirar fuera;
Tiempo para rasgar, y tiempo para coser;
Tiempo para callarse, y tiempo para hablar;
Tiempo para amar, y tiempo para odiar;
Tiempo para la guerra, y tiempo para la paz;
Al final ¿qué provecho saca uno de sus afanes?*

Eclesiastés 3

Hay un tiempo para comenzar un sueño, y un tiempo para concretarlo, esta Tesis es el culmen de ese tiempo. La reflexión del texto anterior me hace decir que al final, el provecho que se saca es aquel, que se encuentra cimentado en la conjunción de los esfuerzos de muchos, de los que han podido influir o ser influidos por la vida de uno mismo, aquí la frase de **Bernardo de Chartres**; “*Somos enanos a hombros de gigantes*”, es más que verdadera. Ahora, este es el tiempo en que, al terminar la tesis, he de agradecer a todos aquellos, que directa o indirectamente, han contribuido a que llegara hasta este punto. La lista es bastante larga, y eso es muy bueno, porque así se está consciente que no se está solo en este mundo, tanto en lo bueno como en lo malo.

Agradecer sobre todo a Dios y a María Madre, que me han guiado y protegido mis pasos a lo largo de mi vida, para poder sacar lo mejor de mí para gloria de Él. A mi familia, presente y ausente, cercana y lejana, dispersa en varios ámbitos; Estados Unidos, México, Centro América, Nicaragua y España, especialmente a mi madre Auxilio y a mi hermana Xochilt, que han estado ahí siempre, a la memoria de mis abuelos maternos; Margarita y Gilberto, que desde el cielo, siguen guiando y fortaleciendo las convicciones, los valores y el ejemplo, que me supieron inculcar desde mi niñez. A mi esposa Inma, quien como compañera de camino, ha compartido parte de este recorrido y que ha tenido y sabido infundir el amor, la paciencia, el ánimo que necesitaba en los duros momentos, y el impulso necesario para seguir adelante. A todos los profesores, maestros y guías que con su sabiduría y apoyo, supieron suscitar el valor necesario para continuar luchando. A los asesores espirituales, mentores y amigos de diversos países y sitios, especialmente a la gran Familia de Aliados (FAMDAL), los que con sus oraciones, apoyo material, palabras, enseñanzas, ejemplo y su presencia física y espiritual, han sabido apoyarme y encausarme en el camino del saber y del bien.

A todos aquellos, como mis alumnos, a los que lo bueno que han visto en mí, les ha sabido inspirar, para poder ser mejores profesionales, ciudadanos y personas. Al personal de INETER, los trabajadores, técnicos y administrativos de diversas áreas, sobre todo de la Dirección General de Geofísica, en especial a Martha Ibarra y Alan Morales entre otros, que sin su apoyo y ayuda no habría podido recopilar parte de la información que se plasma en este estudio. A las autoridades de las Universidades, Ministerios y Entidades de México, Nicaragua y España, que a través de las ayudas y becas otorgadas, pudieron ver en mí, lo suficiente para creer en lo que podía llegar a ser. A esa larga lista de personas hombres y mujeres que en este momento no puedo enumerar, pero que siempre llevo en mis pensamientos, oraciones, intenciones y acciones. A todos, por lo que al final, puedo decir que “el provecho que he sacado, bien ha valido la pena”, muchas Gracias.

PRÓLOGO

La presente tesis aplica técnicas de reconocimiento y clasificación de señales, mediante modelos matemáticos a una serie de datos sísmicos generados en continuo y que proceden de cuatro volcanes activos. Dos situados en Europa en Italia y dos en Centro América en Nicaragua. El objetivo principal es desarrollar un reconocedor de eventos sísmico-volcánicos que minimice el error de clasificación con un costo computacional aceptable, que permita llevar a cabo un análisis mucho más fiable de las señales sísmicas producidas en los volcanes del que actualmente se realiza de forma manual. Este proceso pasa por un reconocimiento de los tipos de eventos sísmicos que se producen en los volcanes estudiados, a través de un análisis de las características espectrales de la señales y de algoritmos de reconocimiento.

El desarrollo del estudio de los datos corresponde al registro de señales producidas por eventos acontecidos y recopilados en cuatro campañas; dos realizadas en dos volcanes de Italia; el Etna y Strómboli en un período de actividad de una semana durante 1997 y 1999 respectivamente, y otras dos en Nicaragua; con el volcán San Cristóbal y el volcán Telica, en un período de actividad de dos semanas del 16 de Febrero al 10 de Marzo del 2006 en ambos volcanes.

Para este fin se utiliza como herramienta para realizar el entrenamiento de los modelos y posterior reconocimiento de eventos, el razonamiento de los modelos ocultos de Markov (*Hidden Markov Models*: HMMs), a través de un proceso de parametrización de las señales en componentes cepstrales.

Uno de los logros del proyecto, ha sido el construir un prototipo de un sistema automático reconocedor y clasificador de eventos sísmicos en volcanes, que sea capaz en un futuro de funcionar en tiempo real, facilitando y ayudando en la labor de los especialistas y/o el operador humano en el análisis y futuras predicciones que suceden en el desarrollo de la actividad de un volcán. En esta misma línea, también se ha desarrollado una aplicación para ayudar en la conversión de formatos y el análisis de los eventos sísmico-volcánicos previo a la clasificación.

El presente trabajo, ha podido brindar un acercamiento a la definición particular de cada uno de los tipos de eventos que se producen en los volcanes estudiados, así como sus características, debido a que como ocurre en el caso particular de Nicaragua, aún no se maneja una clasificación en dichos volcanes. Por lo tanto, aportar información sobre los tipos de eventos que se producen en la cadena volcánica cuaternaria de Nicaragua es un éxito importante. Logrando de esta forma, al igual que en los de Italia, que se pueda identificar en el menor tiempo, el tipo de evento que ocurre en cada volcán. De esta manera, podrían despejarse dudas y conocer las causas y posibles consecuencias, llevando así a una toma de decisiones sobre un valor de la fiabilidad y no solamente del simple parecer humano. Por lo tanto, el método empleado en este trabajo no solo implicaría la existencia de eventos, también brindaría una medición técnica al reconocimiento visual de los mismos (lo que actualmente se realiza por los expertos en los observatorios), ya que estaría aportando el poder reconocer qué tipo de evento está ocurriendo en un determinado período de fuerte actividad sísmica, previa o posterior a una inminente erupción, lo que repercutirá en desarrollar sistemas de alerta temprana mucho más fiables, que minimicen las consecuencias de riesgo a la población en las acciones de vigilancia y prevención sísmico-volcánica.

Por último un logro particular a destacar, y que acompaña el desarrollo de esta tesis, ha sido la cooperación y el trabajo interdepartamental, ya que para alcanzar la presente tesis en Ciencias de la Tierra, se han unido dos departamentos de ámbitos de conocimientos muy distintos; los departamentos del Instituto Andaluz de Geofísica y Prevención de desastres sísmicos (IAG), con el departamento de Teoría de la Señal, Telemática y Comunicaciones (TSTC). Esto marca un hito en el futuro de la cooperación entre departamentos de la Universidad de Granada.

ABSTRACT EXTENDIDO

El objetivo de este trabajo es construir un prototipo de un sistema de detección y clasificación de eventos sísmico-volcánicos. Que sea capaz en un futuro de funcionar en tiempo real y facilitar así la labor de los especialistas, en el análisis y futuras predicciones que suceden en el desarrollo de la actividad de un volcán. En el estudio de los datos sísmicos, el problema de clasificación de las señales generadas por un volcán se resuelve de forma automática, utilizando técnicas similares a las implementadas en los sistemas de reconocimiento automático de voz. Dichos datos, corresponden a señales producidas por eventos acontecidos en cuatro volcanes activos; dos de ellos; los volcanes Etna e Strómboli de Italia, en un período de actividad de una semana. Los otros dos San Cristóbal y Telica ubicados en Nicaragua, Centro América se presentan en un período de actividad de dos semanas. Se ha determinado la selección de estos cuatro volcanes debido a que poseen características geológicas similares, aunque se encuentren en zonas geográficas diferentes. De esta forma, puede observarse si existe similitud en los eventos sísmicos que se producen en ellos. En el caso de Italia, se ha trabajado con eventos sísmicos ya catalogados, ampliamente conocidos y estudiados, que en el caso del Etna es el denominado “Tremor burst” (TB), y otro evento asociado a ruidos, producidos en el volcán por diferentes causas a los que se les denominó Noise Etna (NE). En el caso del Strómboli también se han utilizado eventos muy característicos, ampliamente conocidos, “las explosiones estrombolianas”, denominadas en el presente estudio como “explosiones Strómboli” (Exp), de manera similar que en el Etna, se ha considerado además, un evento relacionado a ruidos producidos en el volcán (asociados a diferentes causas), a los que se les denominó Silencio (Sil).

En el caso de los volcanes de Nicaragua no se dispone de ninguna clasificación, por lo que se está marcando pautas iniciales en eventos que no han sido analizados ni catalogados, llegando a ser pioneros en este campo, y nombrar de forma temporal durante el estudio, después de un exhaustivo análisis, los eventos encontrados de una forma genérica hasta lograr en un futuro un consenso con los sismólogos locales el tipo de evento al que pertenecen de acuerdo a una determinada escuela. De esta forma en Nicaragua se han podido catalogar siete tipos de eventos en los dos volcanes, en el San Cristóbal se analizaron tres tipos de eventos, denominados S1, S2 y S3 con su correspondiente ruido asociado, nombrado “Noise San Cristóbal (NS)”. Para el volcán Telica, fueron cuatro eventos llamados como T1, T2, T3 y T4, así como el ruido asociado o “Noise Telica (NT)”. Este procedimiento de estudio, incluye un detallado análisis espectral de los registros capturados por los sensores sísmicos, para obtener un conjunto de parámetros representativos de la señal y construir una fiable base de datos, sobre la que se aplique los algoritmos, tanto de entrenamiento de los modelos como de reconocimiento. Así, se utiliza como herramienta para realizar el entrenamiento de los modelos y posterior reconocimiento de eventos, el modelado oculto de Markov (*Hidden Markov Models, HMMs*). Obteniendo así, en las pruebas un porcentaje de aciertos en el reconocimiento de eventos en el Etna y Strómboli en pruebas cerradas superior al 93 % y un porcentaje de precisión de más de 94%. Además, se obtuvo una mejora en la matriz de confusión respecto al reconocimiento de los eventos, reduciendo el porcentaje de error en cada evento analizado. Para el caso de los volcanes de Nicaragua se realizaron pruebas tanto abiertas como cerradas y se obtuvieron resultados en el orden cercano al 95% de efectividad en pruebas cerradas.

Otro de los logros que manifiesta el sistema, ha sido el poder reconocer con éxito eventos que inicialmente no habían sido marcados por el observador. De esta forma, el número de eventos reconocidos se incrementó. Por otra parte, el sistema logra descartar con éxito eventos claramente definidos por el ruido y clasificándolos como tales. Por lo tanto, se afirma que se ha obtenido un sistema de entrenamiento de modelos fiable para el proceso de reconocimiento y clasificación de señales de eventos sísmicos en los volcanes, el cual puede en un futuro ser aplicado en tiempo real. Adicionalmente, en el curso de los experimentos, hemos establecido la premisa de que no es posible generalizar el modelo de entrenamiento. Esto se deduce al mezclar las bases de datos de los cuatro volcanes y no obtener resultados positivos. Por lo que se concluye que el modelo de entrenamiento no es posible emplearlo en la generalidad de los volcanes, sino más bien, debe tomarse cada estructura volcánica como un caso muy particular y único y realizar así un análisis y construcción de modelos de entrenamiento y clasificación de eventos de forma particular, pudiendo aplicar el método de forma general pero conformar la base de datos del entrenamiento de los modelos diferente en cada volcán.

Este estudio se inició de la experiencia recopilada en el desarrollo de un trabajo de investigación, cuyo eje central fue el estudio de eventos sísmicos en dos volcanes activos de Italia; el Etna y el Strómboli. Los datos que en esa ocasión se obtuvieron pertenecían a un equipo de investigación Ibero-Italiano en Septiembre de 1997. Puesto que no hay dos volcanes iguales aunque estén situados en una misma área geográfica y sean de similares características geológicas, el presente estudio ha recopilado datos de cuatro volcanes de características similares para establecer una comparación entre los eventos que se producen y poder establecer así un reconocimiento y clasificación de eventos de manera fiable.

El trabajo consta de dos grandes partes que a su vez se dividen en seis capítulos. La primera parte describe la presentación del problema y engloba la parte teórica y metodológica a utilizar. Abarca del capítulo uno al capítulo dos.

El capítulo uno tiene como objetivo describir el estado del arte de las señales y la importancia de la aplicación de técnicas de reconocimiento de patrones. De este modo, se presentan las características y la estructura general del procesamiento de señales. Adicionalmente, se describe el problema del reconocimiento de las señales sísmicas en los volcanes, los tipos de eventos producidos, sus relaciones con mecanismos volcánicos, la importancia de su caracterización y la necesidad de reconocer las señales en tiempo real. Por otro lado se describen las características generales de los sistemas de reconocimiento cuyo objetivo final es su aplicación a sistemas de alerta temprana en tiempo real de erupciones volcánicas.

El capítulo dos establece de forma teórica, cómo se realizará el análisis de las señales recopiladas en las distintas campañas de Italia y Nicaragua. Se presentará por lo tanto, el marco teórico matemático que permitirá el estudio de las señales sísmico-volcánicas, la caracterización de las mismas y el diseño y construcción de un sistema de reconocimiento automático de eventos sísmicos producido por los volcanes en estudio. En este capítulo se establecen las bases del análisis espectral de las señales, que va a permitir poder identificar los diferentes tipos de eventos que se utilizarán para la construcción de bases de datos etiquetada; dicho análisis se realiza en el dominio de la frecuencia, haciendo uso de la transformada de Fourier y el espectrograma de las

señales. También en este capítulo, se determinará la metodología a aplicar en el sistema de reconocimiento basado en los Modelos Ocultos de Markov. El análisis aquí propuesto, podría en un futuro complementarse con herramientas adicionales, que se detallan en los apéndices (ver apéndices D y E).

La segunda parte comprende la parte práctica, desde el trabajo de campo que involucra las campañas de recopilación de los datos, a la implementación del método y los resultados obtenidos. Aquí también se presenta la aplicación que se ha desarrollado a partir de los estudios implicados en este trabajo. Esta segunda parte engloba desde el capítulo tres hasta el capítulo seis.

El capítulo tres describe las zonas de estudio en los volcanes, las campañas a realizar y los instrumentos empleados. En este capítulo se recoge una descripción detallada en cuanto a historia y sismología del sitio de recopilación de datos de los cuatro volcanes de estudio: dos volcanes situados en Nicaragua y dos en Italia, así como las estaciones e instrumentos que permiten la recopilación de los datos. Esta información sirve de ayuda para el desarrollo del estudio dentro del marco geográfico y particular de la actividad sísmica que presenta cada volcán. En este mismo capítulo se desarrolla la adquisición de datos por parte del investigador. Aquí se recopilan los pasos realizados en la obtención de los registros sísmicos en los cuatro volcanes. En el apéndice F se describe el proceso de transformación de los datos, desde que son tomados en crudo en lenguaje binario en el sitio, hasta que son preparados mediante diversos formatos de conversión para ser capaces de ser procesados por los modelos de entrenamiento.

El capítulo cuatro presenta el análisis de los datos. También plantea una descripción de la obtención de los diferentes tipos de eventos y sus características. Con estos tipos de eventos que han sido definidos mediante la experiencia del observador, se han de desarrollar los análisis en el presente estudio. Aquí, hacemos hincapié en un análisis espectral mucho más detallado, con cada uno de los diferentes tipos de eventos sísmicos encontrados en los registros. Además de presentar las características espectrales de los eventos sísmicos de cada volcán, definiendo los tipos de eventos con los que se ha de trabajar en el proceso de reconocimiento que se verá en el siguiente capítulo. Se abordará la complejidad de los eventos sísmico-volcánicos al describir las estadísticas de duración de cada tipo de evento y se detallará el procedimiento de segmentación y etiquetación manual para la construcción de las bases de datos de entrenamiento, que como se indicará en este capítulo, es uno de los pasos más delicados, ya que una buena base de datos bien segmentada, asegura el éxito en las pruebas de entrenamiento y reconocimiento. Por último, en este capítulo se establecerá una clasificación genérica con los tipos de eventos determinados en los volcanes de estudio de Nicaragua, ya que en la actualidad no corresponden a ninguna clasificación en particular.

El capítulo cinco presenta la realización de la aplicación del método de estudio. Cada tipo de evento de los determinados en el capítulo anterior se ha de generar un conjunto de vectores parametrizados. En el apéndice H se detalla todo lo que involucra este proceso como son: la creación de los diccionarios de eventos y las gramáticas, la creación del modelo base o prototipo, y establecer los parámetros de configuración mediante scripts de entrenamiento y reconocimiento (Ver apéndice H).

El desarrollo de la aplicación de método implica el desarrollo de cuatro tipos de pruebas: las primeras son las denominadas “**test cerrado o entrenamiento**”, realizadas sobre la base de datos previamente etiquetada. Un resultado exitoso en las pruebas, indicaría un porcentaje de precisión superior al 80%. A partir de los resultados iniciales, se introducirá un proceso de mejoras en las pruebas cerradas, que consistirá en realizar una serie de modificaciones a algunos de los parámetros de configuración iniciales, para observar cuál de ellos incrementa los porcentajes de acierto y precisión, y a la vez determinar si se pueden disminuir los valores de las inserciones, borrados y sustituciones en los eventos. Dicho proceso de mejoras constará de cinco pasos; los primeros cuatro, comprenderán la modificación de cuatro de los parámetros de configuración iniciales, el quinto no versará sobre el fichero de configuración, sino que significará realizar un proceso de depuración en el último resultado de los cuatro pasos previos, para eliminar de la base de datos, aquellos registros que obtengan menores porcentajes de aciertos y que estuvieran perjudicando tanto los porcentajes de resultado como el reconocimiento de los eventos. Los cinco puntos son los siguientes:

- a) Cambio en el número de iteraciones
- b) Cambio en el número de Gaussianas
- c) Cambio en el factor de penalización (-p) de la herramienta de reconocimiento Hvite
- d) Cambio en el número de estados *
- e) Proceso de depurado selectivo de los registros, con menores porcentajes de acierto

En el paso “d”, se determina marcar un hito (*), en donde se repetirán las pruebas con el número de gaussianas del paso “b”, y se ha determinado establecer este punto, para tomar un valor de referencia de las mejoras que hasta ahora se han obtenido con las modificaciones. Una vez finalizados estos cinco pasos del proceso de mejoras del test cerrado, el siguiente paso a realizar es el segundo tipo de pruebas, denominadas “**test abierto o ciego**”, la que esencialmente trata en dividir en dos partes la base de datos. Una primera parte la formarán los modelos entrenados, para pasarlos después a la parte restante y observar los porcentajes de reconocimiento del modelo entrenado sobre la parte restante. El tercer tipo de pruebas a realizar son las denominadas “**ciego - ciego**”, en donde la base de datos de entrenamiento reconocerá registros adicionales, que no han sido previamente parametrizados y que se encuentran sin segmentar (no etiquetados), donde la mayoría de ellos posee un fuerte contenido de ruido de fondo. Con esta prueba, podrá constatar si el modelo de entrenamiento puede reconocer eventos en registros que no han sido manipulados por el observador. Los resultados de estos tres grupos de pruebas podrán medir la eficacia del sistema, para poder reconocer las señales sísmico-volcánicas en cada volcán. El cuarto y último grupo de pruebas, consistirán en “**mezclar las bases de datos de entrenamiento**”, para así poder ver la generalización o no, de los resultados obtenidos. Es decir, si los modelos de entrenamiento de un volcán pueden reconocer eventos en los registros de otro volcán. Para ello se mezclarán las bases de datos y los modelos en dos fases; la primera mezclará las del Etna con las del Strómboli, y las del San Cristóbal con las del Telica, en la segunda fase se realizarán la mezcla de las bases de datos y los modelos de entrenamiento de los cuatro volcanes. Al final del capítulo se atenderá a la discusión de resultados.

El capítulo seis, presenta las conclusiones, metodológicas y experimentales obtenidas durante el desarrollo de este trabajo, así como las futuras investigaciones que pueden resultar de las experiencias y conclusiones a las que se han llegado mediante los resultados del presente trabajo.

Por último, se han incluido varios apéndices que ayudan y complementan toda la información que se maneja en la presente tesis. El apéndice A presenta las publicaciones científicas, y las ponencias que se han derivado fruto del desarrollo de la presente tesis. El apéndice B ayuda a situar a la región Centro Americana y concretamente a Nicaragua en el marco geodinámico, presentando el ámbito sísmico-volcánico. Este apéndice ha sido incluido con el fin de aportar a dicha recopilación, un marco teórico sobre la importancia y la historia de los volcanes que conforman la cadena cuaternaria del Pacífico de Nicaragua. Debido a que la información sobre este tema no es particularmente abundante, se ha creído conveniente incluirlo aquí, para poder brindar una breve información sobre los procesos sísmicos, la historia y la influencia sobre la población actual y futura de Nicaragua. El apéndice C describe la historia eruptiva y actividad sísmico-volcánica de los volcanes de estudio durante el período de recopilación de datos. Los apéndices D y E incluyen las herramientas complementarias en el análisis de la caracterización de las señales sísmico volcánicas que se tratan en el capítulo segundo, así como su aplicación a los diferentes tipos de eventos aquí estudiados. El apéndice H realiza una detallada descripción de los principales programas utilizados en el presente estudio.

En esta misma línea, en el apéndice G se ha agregado una implementación tecnológica tendiente a la ayuda en los análisis de las señales sísmico-volcánicas. A partir de los análisis de los eventos sísmico-volcánicos, se ha logrado desarrollar un software que provee de herramientas para desarrollar un análisis espectral de las señales, así como la conversión, para ser utilizado como herramienta de ayuda en el posterior reconocimiento y clasificación de dichas señales mediante los modelos ocultos de Markov (HMMs). El sistema ha sido desarrollado utilizando Matlab V.9, mediante la programación de interfaces gráficas de usuario (GUI).

Key words: Tremor Burst, Noise Etna, HMMs, parametrización, componentes cepstrales



EXTENDED ABSTRACT

The main objective of this work is to build a prototype of a system of detection and classification of seismic volcanic events. To be able in the future to operate in real time and facilitate the work of specialists in the analysis and future predictions that occurs in the development of the activity of a volcano. In the study of the seismic data, the problem of classification of the signals generated by a volcano is solved automatically, using techniques similar to those implemented in systems for automatic speech recognition. These data correspond to signals produced by events happening in four active volcanoes, two of them Etna and Stromboli volcanoes in Italy, in a period of activity a week. The other two San Cristóbal and Telica located in Nicaragua, Central America over a period of two weeks of volcanic activity. We have determined the selection of these four volcanoes because they have similar geological characteristics, although they are in different geographical areas. Thus, it can be seen whether there is similarity in the seismic events that occur in them. In Italy, we have worked with already cataloged seismic events, widely known and studied, in the Etna is the "Tremor burst" (TB), and other event associated with noise, the volcano produced by different causes which were called Noise Etna (NE). In Strómboli, have also been used very characteristic events, widely known "estrombolianas explosions" referred to in this study as " Strómboli Explosions " (Exp), similarly as in Etna has been considered also an "noises events" related to the volcano (associated with different causes), which were named Silence (Sil).

For Nicaragua, there is no any classification, so it is being marked initial guidelines on events that have not been analyzed or cataloged, becoming pioneers in this field and a temporary name for the study, after an exhaustive analysis, the events found in a generic way, to achieve consensus on a future with local seismologists, the event type they belong according to a particular " seismic school ". Thus, in Nicaragua have been able to catalog seven kinds of events in the two volcanoes in the San Cristóbal analyzed three types of events, called S1, S2 and S3 with a corresponding noise associated, named "San Cristobal Noise (NS)" . For Telica volcano were four events called as T1, T2, T3 and T4, and the noise associated or "Noise Telica (NT)". This study procedure, including a detailed spectral analysis records captured by the seismic sensors, to obtain a set of parameters representative of the signal and reliably construct a database on which to apply the algorithms both for training as recognition models. Thus, the hidden Markov modeling (Hidden Markov Models, HMMs) it is used as a tool for training the models and recognition seismic events, and we have obtained a percentage of correct recognition greater than 98%, and an accuracy of over 97% in the closed tests in the Etna and Stromboli. In addition, we obtained an improvement in the confusion matrix in the recognition of events, reducing the error rate in each event analyzed. In the volcanoes of Nicaragua, were tested both open and closed and the results were close to 89% effective in open test.

Another achievement that shows the system has been successfully able to recognize events that initially had not been marked by the observer. Thus, the number of events recognized increased. Moreover, the system can successfully rule clearly defined events and classifying them as such noise. Therefore, it is stated that it has obtained a training system reliable model for the process of recognition and classification of seismic signals at volcanoes events, which may in the future be applied in real time. Additionally, in the course of experiments, we have been established the premise that it is not possible to

generalize the training model. What can be deduced, by mixing the databases of the four volcanoes and achieve positive results. It is concluded that the model of training is not possible to use it in most of the volcanoes, but rather, should be taken every volcanic structure as a very special and unique event and perform well analysis and modeling training and classification of events in a particular way and to apply the method in general but form the basis of training data of different models in each volcano.

This study was initiated on the experience gathered in developing a research project, whose centerpiece was the study of seismic events on two active volcanoes in Italy, Etna and Stromboli. The data were obtained at that time belonged to a research team Ibero-Italian in September 1997. Since no two volcanoes same though situated in the same geographical area and are of similar geological characteristics, the present study has collected data from four volcanoes similar characteristics to establish a comparison between the events that occur and to establish and recognition event classification and reliably.

This work consists of two parts divided into six chapters. The first part describes the presentation of the problem and covers the theory and methodology used. It covers Chapter One to Chapter Two.

Chapter one describes the state of the art of the signs and the importance of the application of pattern recognition techniques. Thereby, displaying the features and the general structure of signal processing, additionally, we describe the problem of recognition of seismic signals at volcanoes, types of events produced, its relations with volcanic mechanisms, the importance of characterization and the need to recognize the signals in real time. On the other hand describes the general characteristics of the recognition system whose ultimate goal is to apply to systems of real-time early warning of volcanic eruptions.

Chapter two provides a theoretical way, how will be the analysis of the collected signals in the various campaigns in Italy and Nicaragua. Therefore be presented, the mathematical framework that allows the study of volcanic seismic signals, the characterization of the Dwellings and the design and construction of a system of automatic recognition of seismic events produced by volcanoes under study. This chapter establishes the basis of spectral analysis of signals, which will allow identifying different types of events, which will be used to construct labeled databases; the analysis is performed in the frequency domain, using the Fourier transform and the spectrogram of the signals. Also in this chapter, will determine the methodology to be applied in the recognition system based on Hidden Markov Models analysis proposed here, could in the future be complemented by additional tools, which are detailed in the appendices (see D and E appendices).

The second part comprises the practical, from the fieldwork campaigns involving data collection, implementation of the method and the results obtained. Here also presents the application has been developed from the studies involved in this work. This second part covers from chapter three to chapter six.

Chapter three describes the study areas in volcanoes, campaigns and tools to make employees. This chapter contains a detailed description about the site history and seismology data collection study the four volcanoes, two volcanoes located in

Nicaragua and two in Italy, as well as stations and instruments for data collection. This information helps the development of the study within the geographic and seismic activity especially presenting each volcano. In this chapter develops the acquisition of data by the researcher. Are collected here the steps performed in obtaining seismic records in the four volcanoes. Appendix F discloses the transformation of the data, which are taken from the raw binary language on site until they are prepared by converting various formats to be able to be processed by training models.

Chapter four presents the analysis of the data. It also presents a description of the production of different types of events and their characteristics. With these types of events that have been defined by the observer's experience, have to develop the analysis in this study. Here, we focus on a much more detailed spectral analysis, with each of the different types of seismic events found in the records. Besides presenting the spectral characteristics of seismic events each volcano, defining the types of events that you have to work on the recognition process is discussed in the next chapter. We discuss the complexity of the seismic-volcanic events describing the statistics of duration of each type of event and will detail the segmentation and labeling procedure manual for the construction of the training database, which, as indicated in this chapter, is one of the most delicate steps, since a good well segmented database, ensures success in training and recognition tests. Finally, this chapter will establish a generic classification of event types identified in the study volcanoes Nicaragua, since now not correspond to any particular classification.

Chapter five shows the embodiment of the application of the method of study. Each type of event to those determined in the previous chapter to generate a set of vectors parameterized in Appendix H, outlined everything involved in this process such as: the creation of dictionaries and grammars event, the creation base model or prototype, and set the parameters configuration parameters using training and recognition scripts (see Appendix H). The application of this method involves the development of four types of tests, the first is called "closed test or training", made on the basis of pre-labeled data. A successful outcome in trials, indicate accuracy above 80%. From the initial results, we introduce a process of improvement in closed testing, consisting of a series of modifications to some of the initial settings, to see which of them increases success rates and accuracy, and also to determine if you can decrease the value of the insertions, deletions and substitutions in the events. This improvement process will consist of five steps, the first four, include the modification of four of the initial settings, the fifth will focus not on the configuration file, it will mean making a purge in the last issue of the four previous steps to eliminate the database records that obtain lower percentages of correct and that they were both percentages impairing the recognition result of the events. The five points are:

- a) Change in the iterations number
- b) Change in the Gaussian number
- c) Change in the penalty factor (-p) recognition tool Hvite
- d) Change in the number of states *
- e) selective debugging process records with lower success rates

In step "d", a milestone is determined (*), where the tests are repeated with the Gaussian number (step b), and determined to establish this point, to make a reference value of the improvements which so far have been achieved with modifications. After

completing the five steps of the process to improve the closed test, the next step to do, is the second type of tests called "open or blind test", which is essentially divided into two parts in the database. The first part, the trained models form, to pass then the remaining part and observe the pattern recognition percentages trained on the remaining part. The third type of test to be performed, is called "blind - blind", in which the training data base will recognize additional records that have not previously been parameterized and are unsegmented (not labeled), and in which most of them has a strong containing background noise. With this test, may be established if the training model can recognize events in logs that have not been manipulated by the observer. The results of these three groups of tests may measure the effectiveness of the system, to detect seismic-volcanic signals at each volcano. The fourth and last group of tests consist of "mixed basis of training data", so, to see whether or not the generalization of the results. That is, if the training models of a volcano can recognize events in the records of another volcano. This will mix the databases and models in two phases: the first will blend with those of Etna Stromboli, and the San Cristobal with the Telica, in the second phase will take place the mix of databases and training models of four volcanoes. At the end of the chapter shall address the discussion of results.

Chapter six presents the conclusions, methodological and experimental obtained during the development of this work, and future research that may result from the experiences and conclusions that have been reached in this work.

Finally, we have included several appendices that help and complement all the information used in this thesis. Appendix A presents the scientific publications, and presentations that have been derived result of the development of this thesis. Appendix B helps situate the Central American region and specifically to Nicaragua in the geodynamic framework, presenting the seismic-volcanic field. This appendix has been included in order to contribute to this collection, a theoretical framework about the importance and history of volcanoes comprising quaternary Pacific chain of Nicaragua, because the information on this topic is not particularly abundant, has seen fit to include it here in order to provide brief information about seismic processes, history and the influence on the current and future population of Nicaragua. Appendix C describes the eruptive history and seismic-volcanic activity of the volcanoes of study during the data collection period. Appendices D and E include complementary tools in the characterization analysis of volcanic seismic signals discussed in the second chapter and their application to different types of events studied here. Appendix H performs a detailed description of the main programs used in this study.

Similarly we have developed in Appendix G, a technological implementation that it will help in the analysis of the seismic-volcanic signals. From the analysis of volcanic seismic events, we have a software developed that provides tools for developing a spectral analysis of signals and conversion, to be used as an aid in the subsequent recognition and classification of these signals by hidden Markov models (HMMs). The system was developed using Matlab V.9; by programming a graphical user interfaces (GUI).

Key words: Tremor Burst, Noise Etna, HMMs, parametrización, componentes cepstrales



PRIMERA PARTE

INTRODUCCIÓN, ANTECEDENTES Y PRESENTACIÓN DEL CASO DE ESTUDIO



*Volcanes Momotombo y Momotombito, Nicaragua.
Por: Ligdamis A. Gutiérrez E.*





CAPÍTULO I

INTRODUCCIÓN



*Cráter Santiago, Volcán Masaya, Nicaragua.
Por: Ligdamis A. Gutiérrez E.*



1.1- IMPACTO DE TERREMOTOS Y VOLCANES SOBRE LA SOCIEDAD

Una gran parte de la población del planeta se concentra, alrededor o sobre zonas sísmicas muy activas y volcanes, la humanidad vive y se desarrolla es un planeta en constante movimiento, donde los volcanes juegan un papel crucial, en la vida y el entorno del ser humano. Por las mismas causas de desarrollo y supervivencia, la humanidad, desde sus inicios ha establecido sus núcleos urbanos muy cerca de las zonas volcánicas, debido principalmente a los minerales que generan las emanaciones volcánicas y que hacen muy productivas a las tierras que las rodean. Por lo mismo, los volcanes representan simultáneamente efectos tanto de riqueza como de peligro para quienes viven en sus alrededores. Sin embargo, el hecho de que la economía de las primeras civilizaciones, estuviera basada primordialmente en la agricultura, justificaba el peligro de convivir con tales amenazas. Es así que, aunque las erupciones volcánicas y los terremotos constituyen una serie de sucesos y fenómenos naturales que representan un riesgo potencial para la seguridad humana, incluido el medio ambiente, también han desempeñado un papel crucial en el mismo desarrollo de las sociedades. Los fenómenos volcánicos y sísmicos, en todas sus variadas manifestaciones, han despertado el interés a los estudiosos y curiosos a través de todas las épocas de la historia. Desde la antigüedad hasta los tiempos modernos, han sido descritos grandes desastres: terremotos que han destruido poblaciones enteras y causado miles de víctimas, erupciones volcánicas que han sepultado pueblos y destrozado todo el ecosistema de los alrededores. Una de las primeras crónicas escritas sobre la destrucción ocasionada por la erupción de un volcán data del año 79 D.C. donde Plinio el joven, sobrino de Plinio el viejo describe la erupción del volcán Vesubio que destruyó las ciudades de Pompeya y Herculano y donde murió su tío [Plinio el joven 2005]. Actualmente la ciudad de Nápoles se encuentra en los alrededores del volcán, con el riesgo de sufrir el mismo final. Por otra parte, existe un problema de recopilación de información, ya que existen muy pocos registros históricos de terremotos anteriores al siglo XVIII o en el mayor de los casos no son muy verídicos. Es hasta entrado el siglo XX y lo que se lleva del siglo XXI, en que las nuevas tecnologías de la información y de la comunicación, permiten conocer a nivel global y al instante, los sucesos y fenómenos naturales que ocurren a miles de kilómetros de distancia. Aunque se ha avanzado mucho, las características de eventos sísmico-tectónicos como los terremotos producidos en diversidad de fallas geológicas representan aun muchos retos. Países como Japón, lideran la prevención a través de la construcción de estructuras y viviendas que soporten las sacudidas de grandes sismos sin colapsarse, pero aún falta mucho para poder predecir realmente con tiempo cuando y en que sitio ocurrirá un terremoto. Sin embargo, se ha avanzado mucho más en el estudio de los fenómenos que se producen en los volcanes. En este campo, los científicos pueden recurrir a una serie de factores en la búsqueda de eminentes erupciones volcánicas, como son los factores geoquímicos, de deformación de termometría y de sismología entre otros. El estudio del presente trabajo se enfocará precisamente a los eventos sísmico-volcánicos que se generan en ellos.

Se han logrado muchos éxitos en la predicción de alertas tempranas basándose en la sismología. En Junio del 2003, de acuerdo a un informe final de la consulta hemisférica sobre alerta temprana, realizado en Antigua Guatemala, con la participación de 19 países del continente americano se definió que *“Los sistemas de alerta temprana son concebidos como estructuras multi-sectoriales, multi-institucionales que usan pronósticos para responder de manera eficiente en caso de que ocurra un evento de gran intensidad”*. [Declaración la Antigua 2003] En el caso de la sismología, estos

pronósticos pasan por la aparición de eventos predictivos como terremotos, tremores o eventos de largo período (LP) y representan al día de hoy, un factor fiable a la hora de poder afirmar que se acerca una eminente erupción en un determinado volcán. Ante esta realidad, uno de los objetivos primordiales es poder predecir la ocurrencia de los fenómenos eruptivos, con el fin de poder mitigar y minimizar los daños. Es vital poder realizar tareas de evacuación de la población de las zonas de peligro a zonas seguras antes de que una erupción o un terremoto se produzcan. Ante cualquier posible catástrofe, el bien humano sigue y seguirá siendo la prioridad de todo plan de alerta o de mitigación y prevención de riesgo, Es por ello que la clasificación de eventos sísmicos en tiempo real representa un aspecto importante, que requiere un mejor estudio y análisis de los procesos Vulcano-tectónicos. Por consiguiente, en un futuro, una mejor predicción de las erupciones volcánicas que puedan poner en riesgo tanto propiedades, medio ambiente y sobre todo vidas humanas pasa por identificar de manera fiable y en tiempo real las señales sísmicas que se producen en el interior de un edificio volcánico.

Sin embargo, existe el inconveniente de que no todos los volcanes activos del mundo están siendo instrumentados y vigilados, de hecho solo unos pocos cuentan con algo más que no sea vigilancia sísmica [Ortiz 2000]. Además, analizar miles de eventos producidos a diario por los volcanes, supone un aspecto demasiado costoso tanto en tiempo como en recursos. Una consecuencia es que actualmente, en donde existe un adecuado estudio de un volcán, la clasificación y reconocimiento de eventos sísmicos que se producen es una tarea manual y en muchos casos empírica. Por otro lado, la actividad sísmica presente en un volcán activo¹ es difícil de clasificar. En muchos casos, depende de cada escuela [Minakami 1969] [Minakami 1974], [Ibáñez y Carmona. 1997], además, dicha clasificación varía de acuerdo al criterio que manejan los sismólogos en cada observatorio. Es necesario, por lo tanto, el empleo de técnicas avanzadas que permitan una mejor manipulación y visualización de los eventos en los períodos de actividad sísmica. En este sentido, éste es el principal objetivo del presente estudio: “realizar mediante un análisis de las propiedades de las señales, una herramienta que contribuya a facilitar el trabajo del investigador y de esta forma, poder garantizar la correcta interpretación de los datos, mejorando en tiempo y forma los resultados, para obtener mejores conclusiones de vigilancia y predicción sísmica”.

En el estudio de los volcanes, se considera que una erupción volcánica, “*es el proceso en que se libera la energía contenida en el magma a la superficie de la tierra, que puede darse en la atmósfera o en el mar*”. Una gran mayoría de dicha energía presente en el magma es térmica y la liberación se realiza mediante la expansión del gas contenido en el, además, de por una disipación térmica al ambiente [Ortiz 2000]. Un volcán activo, aunque esté en etapa de reposo, genera múltiples procesos que están interrelacionados con el movimiento del magma en su interior, ya sean productos del mismo sistema magmático o por un reajuste del edificio volcánico. De todos estos procesos muchos están formados niveles de energía sumamente pequeños, y por lo mismo son muy difíciles de detectar. De forma similar, gran parte de la actividad en un volcán también obedece a la circulación de gases a través del sistema de fracturas que posee el edificio volcánico, cuando estos gases se combinan con aguas meteóricas, originan lo que se denomina acuíferos geotérmicos, además de las llamadas fumarolas, fuentes termales y los suelos calientes. Es todo este proceso de circulación de los fluidos

¹ Volcanes que han demostrado algún signo de actividad en los últimos 10,000 años o que hayan tenido o no erupciones “recientes” y que en el momento actual presentan alguna manifestación relacionada con la actividad eruptiva (fumarolas, anomalías térmicas, microsismicidad).

y especialmente a los anteriormente mencionados que se producen en la evaporación a través del sistema de fracturas del volcán, lo que genera la aparición de eventos sísmicos, que normalmente se conocen como temblor o temblor volcánico [Ortiz 2000]. En este sentido, puede mencionarse que las señales sísmicas del temblor volcánico se definen como el resultado de la resonancia de estructuras rellenas de fluidos, en estado estático o en movimiento, que se ubica en el interior de los volcanes activos. Aristizábal y otros mencionan que *los tipos de resonadores más comúnmente asumidos como modelos de fuente de temblor son conductos (cilíndricos o rectangulares) o grietas planares*. [Aristizábal et. al 2004]. Por lo tanto, los cambios de presión o el movimiento de magma, provocan series de eventos sísmicos que rodean al sistema magmático. Similarmente, los cambios en el interior del volcán, como la inyección de nuevo magma, el aumento de temperatura de los acuíferos o la entrada de agua en el sistema, provocan deformaciones del edificio volcánico y también el incremento de la actividad sísmica. Aquí de acuerdo a Ortiz la dificultad radica en poder detectar estos fenómenos, identificar sus causas, sus características, establecer los pronósticos de evolución y con toda esta información poder manejar adecuadamente parámetros de alertas predictivas [Ortiz 2000].

Por lo tanto, el estudio de la sismicidad volcánica se basa sobre todo en establecer de acuerdo a una serie de observaciones el estado de actividad de un volcán en un determinado momento, para así poder realizar una alerta predictiva. Aquí es muy importante la identificación en tiempo real de los eventos sísmicos que se producen. De acuerdo al tipo de eventos se puede conocer la fuente que lo origina determinando su posición en el tiempo y espacio. Así, es posible saber las causas o consecuencias que podrán sobrevenir. De ahí, que el primer punto sea establecer un sistema de detección y clasificación de eventos, fiable y objetivo. Entrenar a un clasificador de forma fiable, implica disponer de suficientes repeticiones de los eventos a detectar. Al reconocer, se puede clasificar y ello lleva a atribuir o asignar una etiqueta a cada evento detectado. Una vez detectados y clasificados los distintos eventos producidos por la actividad del volcán, se puede realizar un estudio de los mismos y de este modo, caracterizar a la fuente que los produce. Por lo tanto, poder conocer la dinámica que se está produciendo y establecer las consecuencias que se deriven de ella. Analizando los eventos sísmico-volcánicos que acontecen antes y durante la fase de actividad se puede comprender los procesos eruptivos y los terremotos producidos por el volcán; en este sentido, existen muchas manifestaciones de eventos sísmicos producidos por los volcanes. Sin embargo, autores como Chouet mencionan que dos de ellas son particularmente importantes en el estudio de la sismología volcánica y son el Tremor y los eventos de largo período (LP) [Chouet 1992], [Chouet 2003].

Ahora bien, como se ha mencionado líneas atrás, dicha comprensión parte del poder reconocer qué está ocurriendo en el interior de un volcán, qué señales se generan en la dinámica ya antes mencionada y para tal caso, es necesario precisar de un sistema que de forma automática, pueda manejar la complejidad de dichas señales y producir resultados efectivos y fiables. Esto no es un proceso sencillo ya que las señales sísmicas de origen volcánico registradas por los sensores pueden verse afectadas por agentes externos al volcán; estas interferencias pueden ser producidas tanto por la acción natural (aire, cercanía de fuentes de agua como ríos, mares, etc.), como por la acción humana o animal (personas o animales caminando, medios de transporte, maquinaria, etc. en las cercanías), es deseable minimizar dicho efecto y obtener así una señal más limpia con la cual poder llevar a cabo un proceso de detección y clasificación más efectivo.

Por otro lado, en un volcán activo, de acuerdo al estado y materia, el número de eventos que se producen puede ser muy elevado. Se precisa entonces, reconocer en el menor tiempo posible el tipo de eventos que están ocurriendo. Esto implica, una vigilancia para así establecer una alerta adecuada, una predicción oportuna y establecer el nivel de riesgo en un determinado momento de actividad. Esto conduce a un contexto social, en que se deben de movilizar mecanismos y organizaciones en salvaguarda del bien humano. Hay que recalcar que hasta ahora, esto se está haciendo mediante operadores humanos, los cuales establecen un reconocimiento manual. La decisión de establecer alertas, predicciones y riesgos se basa en la experiencia humana, y por lo tanto no se puede cuantificar. En este sentido, el presente estudio trata de establecer un sistema automático basado en la experiencia humana, que de solución al problema, de reconocer, en el menor tiempo posible, el tipo de evento que se está generando en un momento determinado, que pueda ayudar al operador humano a despejar dudas sobre qué tipo de evento se está desarrollando, conocer sus causas y posibles consecuencias. En este estudio se propone hacer uso de las técnicas ya bien establecidas de reconocimiento de voz (en el proceso de emisión y clasificación de sonidos) [Lee et. al. 1996], [Kaufman et. al. 1992], [De Bruijin et. al. 1995], [Ren et. al. 2008], [Dhir et. al. 2003], debido a que se podría asumir una semejanza entre las señales generadas en una estructura volcánica y el tracto vocal humano. Es decir, si se es capaz de reconocer y clasificar emisiones de señales de voz, la misma premisa se puede aplicar a los volcanes y a cualquier fuente emisora de impulsos. Esto facilitaría establecer un reconocimiento fiable y objetivo.

1.2.- LAS SEÑALES.

1.2.1.- Tipos de Señales:

En la actualidad, en el mundo moderno se puede mencionar que los conceptos de señal y sistema han permitido encarar el estudio de cualquier problema del mundo físico mediante la representación de un “*modelo*” adecuado de la realidad, lo que permite describir dicho problema e inmediatamente poder tener a disposición toda la teoría para intentar resolverlo. Este tipo de modelos se han difundido de variadas formas en los tiempos actuales, debido primordialmente a las bases matemáticas de la teoría de la comunicación y a los avances cada vez más rápidos generados en el área informática. Estos hechos han permitido llevar las soluciones al campo digital, invadiendo casi todas las actividades de la sociedad moderna. Así surgen aplicaciones tecnológicas en campos tan diversos como por ejemplo: la investigación, el comercio, la medicina, la educación y la política, entre muchos otros.

Al tratar el tema de las señales se puede afirmar que éstas transportan información acerca del sistema que las ha producido, contenido o codificado en un patrón de variaciones de alguna magnitud física. Desde el punto de vista matemático las señales son descritas por medio de funciones, y los sistemas en términos de transformaciones. Estas transformaciones modifican a las denominadas señales de entrada para dar lugar a otras señales de salida del sistema.

De acuerdo al concepto de señal mencionado por A.V. Oppenheim en el prólogo de su libro “Señales y Sistemas” *Las señales son funciones de una o más variables independientes que contienen información acerca de la naturaleza o comportamiento de algún fenómeno, mientras que los sistemas responden a señales particulares*

produciendo otras señales". Y agrega "Aunque las señales se pueden representar de muchas formas, en todos los casos la información en una señal está contenida en un patrón de variaciones que presenta alguna forma determinada" [Oppenheim et al. 1998]. Etimológicamente, la palabra "Señal" proviene del latín "Signale" lo que significa "marca" que se coloca en algo ya sea para darlo a conocer o distinguirlo de otras formas similares. Así una Señal se puede representar matemáticamente como una función que contiene una o más variables independientes [Oppenheim et al. 1998], si a esta función se le puede describir por medio de una función matemática, se denomina señal determinista, en caso contrario se tratará de una señal de naturaleza aleatoria. Ejemplos de señales son: la actividad eléctrica del corazón, las señales de voz producidas por el aparato fonador humano o las señales sísmicas producidas por la actividad volcánica que es lo que compete al presente estudio. En el desarrollo de este capítulo, se tratará en las diferentes secciones sobre la problemática de la clasificación de las señales en general y la consecuente y agregada complejidad de las señales sísmico-volcánicas tratadas en el presente trabajo.

1.2.2.- El problema de la clasificación de las señales

Una adecuada clasificación permite el estudio sistemático de las señales. Dicho de esta forma, las señales se pueden clasificar de acuerdo a varios criterios:

1. Dimensional: Esta clasificación se basa en el número de variables independientes del modelo de la señal.
2. Energético: Dicha clasificación se basa en que si poseen o no energía finita.
3. Espectral: Está basada en la distribución de frecuencias del espectro de la señal.
4. Fenomenológico: Esta clasificación se basa en el tipo de evolución de la señal si es predefinido o aleatorio.
5. Morfológico: Esta clasificación se basa en el carácter continuo o discreto de la amplitud de la señal o de la variable independiente.

A su vez cada uno de estos criterios se clasifica en uno o más elementos. Por ejemplo, las señales Fenomenológicas se pueden clasificar en determinísticas y aleatorias o estocásticas y las señales Morfológicas clasificarse a su vez en continuas y discretas. De estas dos últimas clasificaciones se determina que en el caso de que la amplitud y la variable independiente sean continuas, entonces se indica que la señal es analógica. En el otro caso de que la variable independiente y la amplitud de la señal sean discretas, entonces se determina que la señal es digital.

1.2.3.- Señales analógicas y digitales

El mundo natural o real en el que vive el ser humano está compuesto en su mayoría por señales de tipo analógico (la voz, los sonidos, los sismos, etc.). Sin embargo, los ordenadores y los avances tecnológicos en el mundo actual, determinan que el procesamiento en tiempo discreto, plantee muchas más ventajas que el procesamiento en tiempo continuo. De ahí que sea de vital importancia el poder analizar las señales digitales. De acuerdo a la definición general, una señal es analógica cuando varía de forma continua en tiempo y amplitud, entre un límite inferior y uno superior pudiendo tomar infinitos valores, ejemplo de ello son la voz y el movimiento cardíaco. Por otro lado, las señales digitales, respecto a las señales analógicas, no varían en forma

continua, sino que los valores en tiempo y amplitud se modifican o alternan mediante aumentos discretos, así la señal digital toma valores discretos.

Por lo tanto, una señal analógica es continua tanto en dominio (tiempo) como en rango (amplitud), por lo que al realizar la conversión se deben realizar dos procesos de discretización: discretización en tiempo (muestreo) y discretización en amplitud (cuantización). Este proceso se lleva a cabo mediante el uso de sistemas de conversión analógica a digital (A/D), los cuales muestrean (sampling) o sea, retienen cada muestra por un instante de tiempo. La velocidad con que se toma esta muestra, es decir, el número de muestras por segundo, (es lo que se conoce como frecuencia de muestreo). Dicho de otra forma, al ser los instantes equidistantes el tiempo que transcurre entre uno y el siguiente se denomina **periodo de muestreo** (suele denotarse como T_s). En consecuencia, se define la **frecuencia de muestreo (fs)** como la inversa de este periodo:

$$f_s = 1 / T_s \quad (1.1)$$

El teorema de Nyquist dice que la frecuencia mínima de muestreo necesaria para poder reconstruir la señal a partir de sus muestras es la frecuencia de Nyquist que es el doble de la frecuencia máxima de la señal. Previo al muestreo es necesario emplear siempre un filtro activo analógico pasa bajo sobre la señal a muestrear con objeto de evitar lo que se conoce como aliasing. El aliasing podría hacer que componentes de frecuencia fuera de la banda de interés quedaran registrados como componentes falsos de frecuencia dentro de la banda de interés. Asimismo, durante la reconstrucción de la señal en la posterior conversión D/A, se hace también necesario aplicar un filtro activo analógico del mismo tipo (pasa bajo) conocido como filtro de reconstrucción. En la figura 1.1 se observa el proceso de muestreo de una señal “X(t)” y de su reconstrucción “Y(t)” a partir de las muestras. El proceso (figura 1.1) parte de una señal continua X(t), que se muestrea usando un convertidor A/D para producir una sucesión de valores discretos $x[n] = x(nT_s)$, donde n es un valor entero que se utiliza como índice de muestreo y T_s es el período de muestreo. La frecuencia de muestreo es el valor $f_s = 1/T_s$. El convertidor D/A ideal permite recuperar de nuevo la señal a partir de los valores discretos.

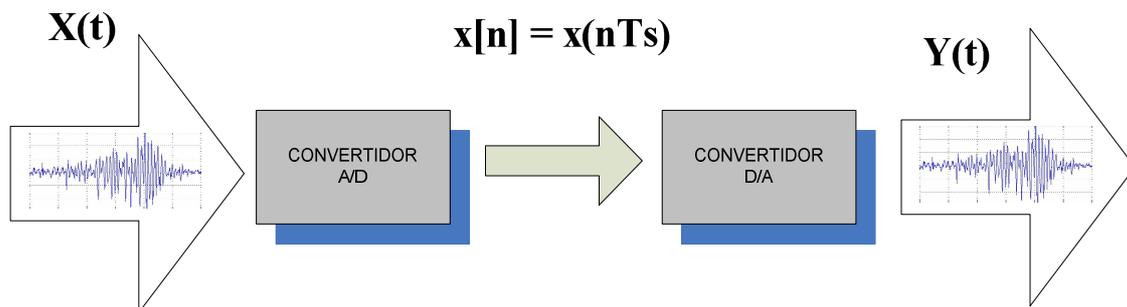


Fig. 1.1. Muestreo y reconstrucción de una señal continua en el tiempo

Posteriormente, la señal se cuantiza en valores discretos. La cuantización consiste en asignar un margen de valor de una señal analizada a un único nivel de salida. Incluso en su versión ideal, añade, como resultado, una señal indeseada a la señal de entrada: el ruido de cuantificación. Por último se procede a realizar la codificación del resultado de la cuantización, en código binario.

Las ventajas de las señales digitales pueden enunciarse de la siguiente forma:

1. Cuando una señal digital es atenuada o experimenta perturbaciones leves, puede ser reconstruida y amplificada mediante sistemas de regeneración de señales.
2. Cuenta con sistemas de detección y corrección de errores, que se utilizan cuando la señal llega al receptor; entonces comprueban (uso de redundancia) la señal, primero para detectar algún error y algunos sistemas, pueden luego corregir alguno o todos los errores detectados previamente.
3. Facilidad para el procesamiento de la señal. Cualquier operación es fácilmente realizable a través de cualquier software de edición o procesamiento de señal.
4. La señal digital permite la multigeneración infinita sin pérdidas de calidad.
5. Es posible aplicar técnicas de compresión de datos sin pérdidas o técnicas de compresión con pérdidas basados en la codificación perceptual mucho más eficientes que con señales analógicas.

1.2.4.- Procesamiento de señales

El procesado de señal tiene como objetivo fundamental la descripción y manipulación matemática de las señales. Proporciona el modo de enfatizar las características fundamentales de una señal, como pueden ser su distribución espectral de energía o su distribución estadística de amplitudes. Uno de los principales instrumentos de la teoría de la señal es la expansión en términos de funciones ortogonales, siendo el análisis de Fourier el caso más considerablemente difundido, y cuya forma generalizada es la Transformada de Fourier (vista con detalle en el siguiente capítulo). Uno de los objetivos principales en el estudio del procesamiento de señales es el poder extraer de una determinada señal la información útil que se encuentra en ella, lo que se realiza mediante diversas técnicas como son: análisis, filtrado, regeneración, medición, detección, e identificación, para posteriormente mostrar los resultados correspondientes en una forma adecuada, ya sea para un observador humano o un ordenador. En la figura 1.2 se observan los principales elementos que intervienen en este proceso.

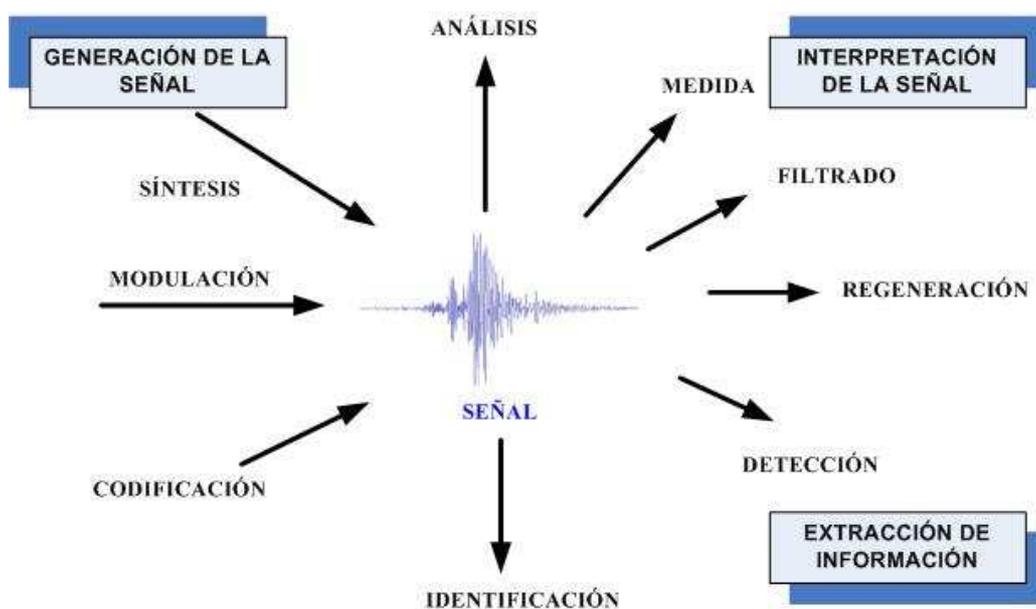


Fig. 1.2. Elementos que intervienen en el procesamiento de la señal.

1.3.- RECONOCIMIENTO Y CLASIFICACIÓN DE PATRONES.

La extracción de la información de una señal para establecer sus características y poder clasificarlas, pasa por el reconocimiento y la clasificación de patrones, por lo que la clasificación es el punto neurálgico de un sistema de reconocimiento de patrones. En este sentido se requiere clasificar una señal reconociendo las características que esta presenta. Atendiendo a la definición más general, propuesta por Devijver y Kittler en 1982, un patrón de características “ p ”, es aquel que describe a una entidad o un objeto, a través de un conjunto de características, las cuales toman la forma de variables aleatorias, de acuerdo a la forma: $p = \{x_1, x_2, x_3, \dots, x_n\}$ [Devijver and Kittler 1982]. De esta forma, los diferentes tipos de patrones se diferencian de acuerdo con la cantidad, significado y naturaleza de dichas características. En un sistema de reconocimiento de patrones una de las principales tareas es la de asignarle a cada patrón de entrada una etiqueta. Por lo tanto, dos patrones diferentes estarían asignados a una misma clase en caso de ser similares y a distintas clases en caso contrario. De acuerdo a Devijver y Kittler, todo sistema de reconocimiento de patrones, básicamente se encuentra compuesto por tres etapas bien definidas: una primera etapa de adquisición de datos, una segunda de extracción de las características o parametrización de la señal y una tercera de clasificación de los patrones. En la figura 1.3, puede observarse un esquema general de las partes de este proceso.

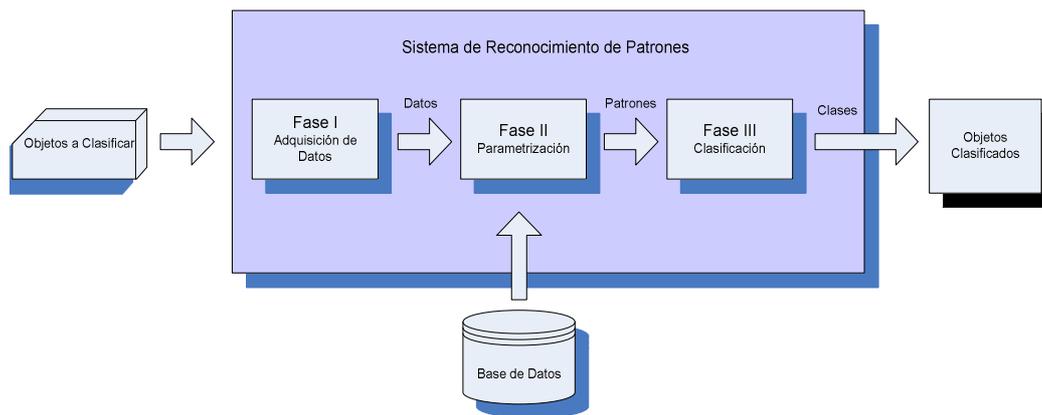


Fig. 1.3. Esquema general de un Sistema de reconocimiento de patrones.

En este sentido, el proceso de reconocimiento y clasificación de patrones pasa por el diseño de un sistema clasificador. En el libro “*Pattern Classification and Scene Analysis*”, Duda y Hart, mencionan que no puede hablarse que hasta hoy exista una teoría unificada en el reconocimiento y clasificación de patrones o de las herramientas utilizadas para tal fin [Duda y Hart 1973].

Diversos autores han abordado desde diferentes perspectivas los algoritmos de clasificación supervisados y no supervisados. De acuerdo a autores como Godoy Calderón, determinan que: “*un problema de clasificación se designa como supervisado cuando previamente se dispone de una muestra de patrones previamente clasificados en cada una de las categorías de que se compone*” [Godoy Calderón et. al. 2006]. A esta muestra de patrones comúnmente se le denomina “muestra de entrenamiento”, ejemplo de clasificación supervisada son los modelos estocásticos. Análogamente, un problema de clasificación no supervisado es aquel en el que: “*no existe un conocimiento anterior,*

así que el problema parte de un universo de patrones, sin una determinada estructura pero que tanto la naturaleza como el número de las clases a determinar forman parte de las definiciones iniciales que son necesarias para poder resolver dicho problema". [Godoy Calderón et. al. 2006]. Ejemplos de estos algoritmos se encuentran utilizando la lógica difusa, propuesta en trabajos por Bezdek. [Bezdek 1981]. Como se ha mencionado, dentro de los algoritmos de clasificación supervisados se encuentran los modelos estocásticos, que se producen cuando por lo menos una variable es tomada como un dato al azar. Las relaciones entre variables se toman por medio de funciones probabilísticas y el aprendizaje de los modelos se basa en la estimación de las distribuciones de probabilidad.

Los *Modelos Estocásticos* proporcionan métodos, técnicas y herramientas de análisis y diseño de sistemas en los que intervienen factores que no se pueden pronosticar con precisión. Para el estudio de estos sistemas, es conveniente medir la incertidumbre a través del uso de la probabilidad. Las cadenas y los modelos ocultos de Markov son un ejemplo de ellos. En el presente trabajo se realiza el estudio de la detección y la clasificación de patrones de señales sísmico-volcánicas mediante el uso de Modelos ocultos de Markov (HMMs).

En el modelado oculto de Markov cada uno de los eventos a clasificar está caracterizado por un modelo estadístico, la señal a clasificar se representa por un conjunto de vectores o símbolos. Si se considera que W es una secuencia de eventos y X es una secuencia de símbolos que representan a la señal a clasificar, el problema a resolver consiste en: dada una secuencia de vectores o símbolos representativos de la señal, determinar cuál es la secuencia más probable de eventos que se ha producido, esto es $P(W|X)$. Para una secuencia de eventos dada y un conjunto de modelos dado, se pueden evaluar la probabilidad de la secuencia de símbolos representativos de la misma, esto es $P(X|W)$; así si $P(W)$ es la probabilidad de la secuencia de eventos y $P(X)$ la probabilidad de la secuencia de símbolos, aplicando la regla de Bayes a cada modelo obtenemos una expresión para calcular la secuencia de eventos más probable que representa una señal dada (fórmula 1.2):

$$P(W|X) = P(X|W)P(W)/P(X) \quad (1.2)$$

El capítulo siguiente, tratará sobre la construcción de dichos modelos, en referencia a la aplicación de los HMMs en el reconocimiento y entrenamiento de eventos sísmico-volcánicos.

Otro de los algoritmos utilizados en los reconocimientos de patrones son las redes neuronales. El estudio de las redes neuronales artificiales (ANN, por su término en inglés *Artificial Neural Networks*), comenzó a partir de los trabajos del matemático John Von Neumann (1903-1957), en donde se toma como base el funcionamiento y estructura de un sistema biológico, como el sistema nervioso conformado por elementos como las neuronas. En la actualidad, existen varios sistemas clasificadores de reconocimiento de patrones de señales sísmico-volcánicas basados en redes neuronales, entre los cuales se pueden nombrar: El sistema de clasificación automática de señales sísmicas propuesto por Scapetta y otros del volcán Monte Vesubio en Italia [Scapetta et. al. 2005], El sistema de clasificación automática de eventos sísmicos propuesto por Langer y otros sobre el volcán Colinas Sufrière, en Monserrat [Langer et. al. 2006]. En Venezuela se diseñó un sistema automático de adquisición de Datos Sísmicos en tiempo

real, para ser implementado en la red sismológica de los Andes venezolanos operada por el laboratorio de Geofísica de la Universidad de los Andes. Este sistema utiliza redes neuronales para el reconocimiento de patrones sísmicos en el dominio de la frecuencia [Guada Barráez y Silva Trujillo 2002]. Otro país que ha implementado sistemas de reconocimiento a partir de redes neuronales es Colombia. En este caso Restrepo y otros utilizaron un sistema de reconocimiento de eventos sísmicos registrados en el volcán Nevado del Ruiz. Dicho sistema fue implementado por la Universidad Tecnológica de Pereira [Restrepo et. al. 2007]. Otro sistema desarrollado en Colombia fue una aplicación de redes neuronales para el pronóstico de sismos, desarrollada por la facultad de ingeniería de la Universidad de la Sabana, utilizando datos del año 2006 del observatorio sismológico del suroccidente. [Manríquez et. al. 2007]. Incluso diversos autores como Bridle [Bridle 1990], Hwang [Hwang et. al. 1989], Niles [Niles y Silverman 1990] y Verdejo [Verdejo et. al. 1993] entre otros, proponen sistemas híbridos argumentando que es posible diseñar una red neuronal con una cierta topología, que evalúe probabilidades que se desarrollan en un HMM. Futuros trabajos podrían incluir este tipo de topología en los análisis desarrollados en el presente estudio.

1.4.- ÁMBITOS RELACIONADOS CON EL RECONOCIMIENTO SÍSMICO

Desde el Neolítico, cuando la humanidad dejó de ser nómada como cazador recolector, y pasó a crear cultivos, a asentarse y formar sociedades complejas, un factor determinante para dichos asentamientos fue el de la fertilidad del terreno que deberían cultivar, y por esta razón dichas sociedades se concentraron alrededor de los volcanes o sobre fallas geológicas. La tierra en constante movimiento y los volcanes juegan un papel crucial en la vida y el entorno del ser humano, las erupciones volcánicas y los terremotos han impactado en todas las sociedades a lo largo de los tiempos, constituyen fenómenos naturales que representan un riesgo para la seguridad personal y colectiva, así como el medio ambiente. En este contexto, los volcanes representan una dualidad para el ser humano, ya que brindan tanto riqueza como destrucción. La humanidad ha tenido que aceptar estas dos realidades, que no han de cambiar con el tiempo, por lo que el hecho de saber convivir con la naturaleza, sus posibles amenazas y peligros se convierte en una necesidad para poder minimizar los daños que puedan presentarse.

Ante dichos peligros, el hombre desde tiempos remotos ha observado los volcanes con los que convive [Plinio el joven 2005], ya en tiempos modernos ha creado los observatorios que se encargan de vigilar y analizar el comportamiento de los volcanes. La vigilancia volcánica se establece a través de diversos métodos, entre los principales se encuentran: los geofísicos, los geodésicos, los geoquímicos, los geológicos y los visuales [Ortiz 2000], [Pérez et. al 2008], [Sparks 2003]. Entre los métodos geofísicos, la vigilancia sísmica es una de las principales. Ahora bien, entre todos los volcanes que existen en el mundo, hay entre 500 a 600 activos [Siebert y Simkin 2002] [Simkin y Siebert 1994]. Sin embargo, no todos los volcanes activos están siendo instrumentados y vigilados por los diversos métodos antes expuestos, de hecho, muchos de ellos solo poseen algún tipo de vigilancia sísmica [Ortiz 2000]. Por lo que en los observatorios, una de las principales necesidades y retos con los que el personal encargado se enfrenta, es poder realizar una clasificación automática del gran número de señales sísmicas que a diario se producen en los volcanes, debido a que dicho análisis se realiza de forma manual por expertos. Así el tener que analizar miles de eventos producidos a diario supone un elevado coste, tanto en tiempo como en recursos, es un proceso muy exhaustivo y que en muchos casos se realiza de forma

empírica. Por otro lado, hay que agregar que la actividad sísmica presente en un volcán activo es difícil de clasificar y en muchos casos depende de cada escuela [Minakami 1960]. Esto último es un ingrediente determinante, ya que cada observatorio tiene en cuenta diferentes clases de eventos sísmicos, tanto tectónicos como volcánicos.

Por lo tanto, la idea principal en el desarrollo de este trabajo en relación al reconocimiento sísmico en los volcanes es implementar un sistema de clasificación, utilizando técnicas de análisis espectral de señales, métodos estadísticos y teniendo en cuenta los sistemas existentes de adquisición en los observatorios, que sobre todo pueda dar un valor numérico fiable a la hora de poder tomar decisiones, que minimice el error humano con un costo computacional aceptable. Ahora bien, al crear un sistema de reconocimiento, este puede determinarse para que trabaje en exclusiva con señales aisladas, que es lo que generalmente realizan los sistemas de reconocimiento automático de la voz [Borrero, et. al. 2011], [Furui 1986], [Lamel, et. al. 1981], o para procesar señales continuas tal y como se generan las señales sísmicas en los volcanes, lo que repercute en un aumento en la complejidad del sistema. Otro aspecto a tomar en cuenta al tratar con los sistemas de reconocimiento general, es el conjunto de eventos que los reconocedores son capaces de tratar; así un sistema sólo puede reconocer el conjunto de eventos sísmico-volcánicos para los cuales ha sido entrenado. Idealmente, un sistema experto de reconocimiento sísmico-volcánico debería de reconocer cualquier tipo de evento sísmico, incluso debería ser capaz de detectar aquellos eventos para los cuales no ha sido entrenado e incluir un procedimiento que permita incrementar la base de datos de entrenamiento con nuevos eventos. Sin embargo, no todos los volcanes son iguales, ni se comportan de la misma manera en cada erupción. Las circunstancias sobre las cuales se generan las señales sísmicas varían de una erupción a otra. Por lo que entre los principales problemas que surgen a la hora de implementar un sistema clasificador se encuentran: la variabilidad y complejidad tanto de las señales, así como las diversas opiniones de los observadores humanos expertos, sobre el tipo de evento que están tratando, ya que como se ha mencionado en la actualidad, los expertos aún no se han puesto de acuerdo sobre estandarizar los diferentes eventos que ocurren en los volcanes, debido a que cada observatorio maneja un criterio diferente para clasificar los eventos que aparecen más frecuentemente en los volcanes que monitorizan. Más adelante en este capítulo se presentarán las principales clasificaciones de los diferentes tipos de eventos sísmico-volcánicos existentes.

1.5.- LOS VOLCANES Y SUS MANIFESTACIONES

En los sistemas de reconocimiento de señales sísmico-volcánicas, el punto de partida de la generación de dichas señales es un volcán. Se ha mencionado que un volcán (figura 1.4) se considera un punto de la superficie terrestre el cual puede situarse ya sea en los continentes o en el fondo oceánico por donde son expulsados al exterior diversos materiales, ya sea en estado sólido (piroclastos), líquidos (magmas) o gaseosos (CO_2 , SO_2 , BrO_2 , sulfuro de hidrógeno, entre otros). Los materiales que produce un volcán provienen del magma que asciende desde el interior de la tierra, son estos movimientos de magma y de otros fluidos volcánicos que al ascender producen fracturas y generan las señales sísmico-volcánicas o terremotos volcánicos. En la figura 1.4 se aprecia un esquema típico con las principales partes de un volcán, así como los productos más importantes que de él emanan.

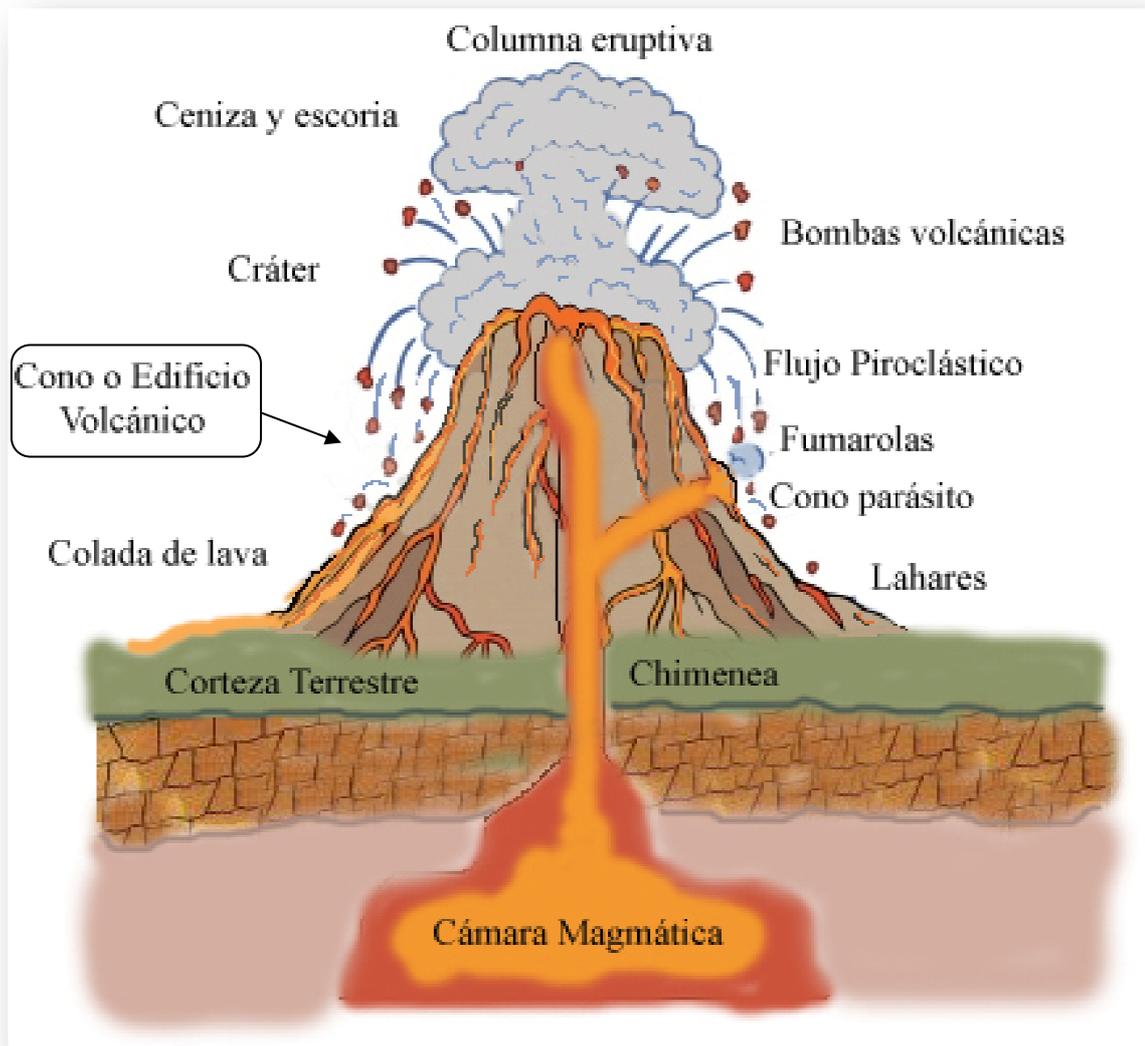


Fig. 1.4. Esquema típico de un volcán y sus principales productos.

En la figura 1.5 puede observarse la distribución global de las placas litosféricas y la relación con la distribución de volcanes y terremotos. Se observa como la mayoría de los volcanes y los terremotos se presentan en la conjunción de las placas, sobre todo en el llamado “Cinturón de Fuego del Pacífico”. En este sentido, debido a que es de particular interés para el autor, así como hacer un compendio de la bibliografía existente, se ha visto interesante, incluir en los apéndices una reseña sobre el ámbito sísmico-volcánico de Centro América y Nicaragua (Ver apéndice B).

Como se observa en la figura 1.4, los volcanes en el proceso de una erupción emiten diferentes productos, lo que implica un determinado riesgo tanto directo como indirecto. La figura 1.6 presenta a los eventos eruptivos que caracterizan lo que se conoce como amenaza volcánica. La cercanía de un volcán implica tener un adecuado método de prevención ante las posibles erupciones volcánicas que impliquen riesgo a la población y a las estructuras de la zona, es necesario por lo tanto, conocer la historia, la frecuencia, y la intensidad de las erupciones de un volcán y ver qué es lo que sucede dentro y en su entorno, para así ante una inminente erupción poder emitir una alerta.

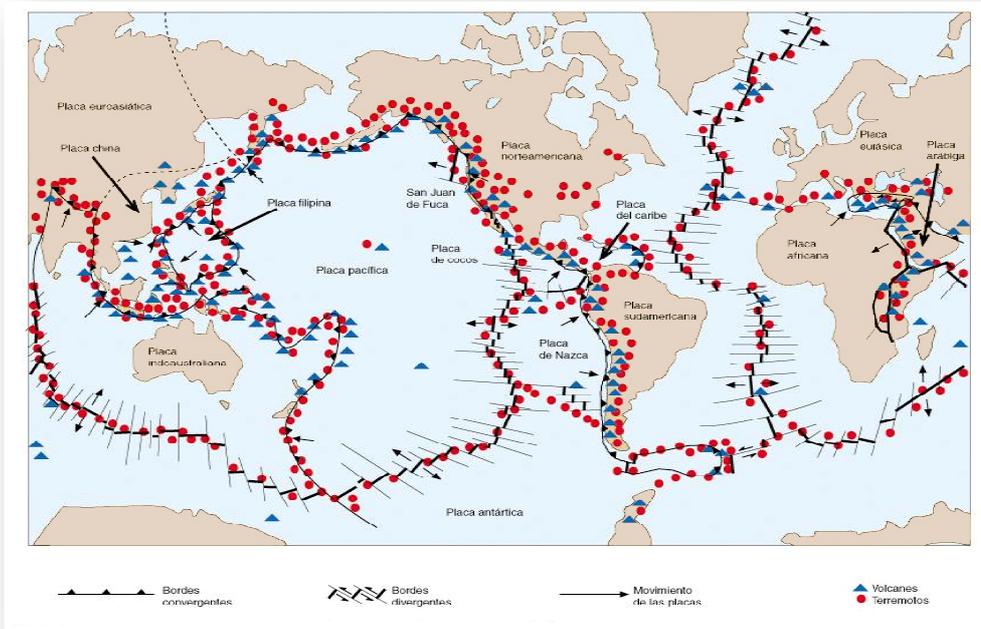


Fig. 1.5 Placas litosféricas y la relación con la distribución de volcanes y terremotos.
Fuente: Kramer S., Geotechnical Earthquake Engineering. Ed. Prentice Hall., 1996

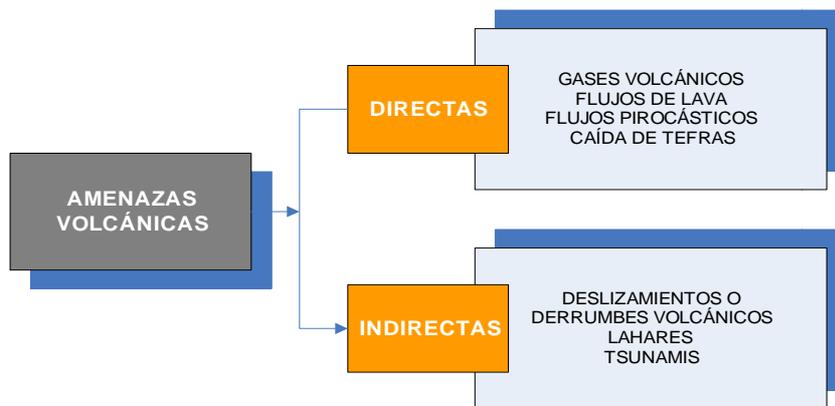


Fig. 1.6 Eventos eruptivos que caracterizan la amenaza volcánica

En la sección 1.4 se mencionaban diferentes métodos de vigilancia volcánica, varios son los factores que se manejan con estos métodos y que preceden a los eventos eruptivos y las amenazas volcánicas. Estos factores son llamados “precursores volcánicos”, entre los más generales se encuentran los siguientes:

- Gases emitidos.
- Pequeños sismos y/o ruidos emitidos en el interior del volcán.
- Cambios producidos en la forma o en las laderas del volcán.
- Variaciones en las propiedades eléctricas, magnéticas, gravimétricas, etc.
- Cambios detectados en imágenes sacadas por satélite.

De esta forma, y siguiendo la línea mencionada en las anteriores secciones, uno de los principales **precursores** de los eventos eruptivos en un volcán son los sismos o terremotos que se producen en su interior [Aki y Ferrazzini 2000], [Chastin y Main 2003], [Chouet 1994]. Este es el eje central del presente trabajo, por lo que es fundamental hacer un buen análisis de los eventos sísmicos que el volcán produce. Hay que conocer y clasificar el tipo de señal sísmica que se está manifestando en determinado período de actividad del volcán, para establecer si dicho tipo de señal sísmica es un precursor de una posible erupción volcánica. Así se podrán establecer mejores y más fiables sistemas de alerta temprana, que establecerán acciones encaminadas a poner a salvo las vidas humanas. Sin embargo, como ya se verá en el desarrollo de este trabajo, dicha clasificación es un tema complejo.

1.6.- SEÑALES SÍSMICAS PRODUCIDAS EN VOLCANES

En la sección 1.1 se contemplaba que el origen de la mayoría de los eventos sísmicos en los volcanes activos, se produce cuando las rocas que se encuentran dentro del edificio volcánico se fracturan, de la misma forma, cuando un volcán se encuentra inactivo o en reposo, la mayor actividad sísmica se debe a la circulación de gases a través de las fracturas en el interior del volcán. Cuando estos gases se mezclan con aguas procedentes de la lluvia, suelen originar acuíferos geotérmicos; ya sea fumarolas, fuentes termales o los denominados suelos calientes. Por lo tanto, todos o la mayor parte de los eventos sísmicos en un volcán se generan debido a la circulación de fluidos (magma y agua), especialmente debido a los cambios que se producen en el agua de los mantos acuíferos al pasar de líquido a vapor, y que provoca la aparición de los sismos, que normalmente se denominan como temblor o temblor volcánico [Ortiz 2000]. Además hay que recalcar que de acuerdo con Arittizábal las señales sísmicas de un temblor volcánico, podrían ser el producto de una resonancia en estructuras al interior del volcán que se encuentran llenas de fluidos, ya sea estáticos o en movimiento. De acuerdo con Arittizábal, los tipos más comunes de los resonadores que son la fuente de temblor son: conductos (cilíndricos o rectangulares), o grietas planares [Arittizábal et. al. 2004]. Por otra parte, también se producen eventos sísmicos por cambios en el interior del volcán, como pueden ser: la acumulación y/o traslado del magma desde el manto a la cámara magmática, el movimiento de gases, los cambios de presión, el aumento de la temperatura en los acuíferos producto de la actividad volcánica o la entrada de agua en el sistema. Todos estos fenómenos provocan alteraciones en el edificio volcánico y por consiguiente el incremento de la actividad sísmica. Estos eventos sísmicos son registrados por los sismógrafos que se instalan alrededor de un volcán, las características que poseen dichas señales como son el tamaño y la forma, dependen básicamente de la intensidad de los diferentes procesos geofísicos que se desarrollan en el interior del volcán. Así pues, los cambios físicos y químicos del sistema magmático, reflejan las condiciones de inestabilidad del sistema y la producción de los eventos sísmicos, que dependiendo de su magnitud, algunos pueden ser sentidos por los habitantes que viven en las cercanías del volcán, mientras que otros, únicamente pueden ser detectados mediante el uso de los sensores o sismógrafos. De esta forma, la vigilancia o monitoreo sísmico en los volcanes, presupone el instalar diversos sensores en las proximidades, en este sentido se recomienda para un mejor seguimiento de la actividad sísmica, la instalación de antenas o array sísmicos [Almendros et. al 1999], [Almendros et. al 2007], [Saccorotti et. al. 2001]. Abril e Ibáñez definen que una antena o array sísmico se compone de un grupo de sismómetros, cuyo instrumental posee características conocidas y similares entre sí. Dichos sismómetros se encuentran

distribuidos en un área determinada, de acuerdo a una configuración que puede estar planificada o ser de forma aleatoria. En dicho estudio mencionan la dificultad de la localización de la fuente de algunos eventos como el temblor o los eventos de largo período con redes sísmicas tradicionales, y hacen una comparación con la técnica del array. Afirman que el problema de localización puede solventarse por medio del uso de esta técnica, ya que los estudios derivados de ellos “*permiten la determinación y el seguimiento de la evolución espacio-temporal de la zona fuente de las señales asociadas a movimientos de fluidos*” [Abril e Ibáñez 2000]. Sin embargo, en muchos de los volcanes en donde existe vigilancia sísmica a través de sensores, solamente se da la existencia de una sola estación, telemétrica con una sola componente, la componente vertical. Aun así, el proceso de recopilación de datos y posterior análisis es el mismo, ya sea mediante el uso de un array o una sola estación. Por regla general, en las estaciones la forma en que se recopilan las señales sísmicas que se producen (ya sea por los array o por una sola), se realiza mediante el uso de instrumentos conocidos como “Convertidor Analógico Digital”, o por sus siglas CAD. Un CAD es capaz de generar un registro digital a partir de una señal eléctrica analógica generada por un sensor. Dicho de otra forma, el proceso de recopilar las señales generadas por la actividad sísmica de un volcán, comienza por establecer estaciones sísmicas en zonas o áreas delimitadas y para poder almacenar y posteriormente tratar las señales sísmicas, que normalmente se hace por medio de un CAD. Los registros pueden ahora ser analizados y almacenados de forma digital. De acuerdo a Tárraga [Tárraga et. al. 2000], el trabajo del conversor es cuantificar la señal analógica en distintas partes de un potencial de referencia utilizado. Un CAD posee 2^N códigos, siendo N el número de bits del conversor. Actualmente, existen conversores de hasta 24 bits. Cuanto mayor es el número de bits, menor es el bit menos significativo (*Less Significant Bit LSB*) y se consigue una mayor resolución en el conversor. Sin embargo, aunque en la teoría los CAD de 24 bits posean mayores resoluciones, en la práctica se trabaja con una resolución de 16 bits, ya que solo los eventos sísmicos muy grandes saturan la señal, y además ya que el ruido térmico generado por los componentes del equipo es mayor a 2^{-16} , con un sobre muestreo y un procesamiento digital posterior, se pueden conseguir mayores resoluciones. Así que con conversores a 16 bits, se puede obtener un equilibrio entre la calidad de los datos obtenidos y la extensión de los ficheros que los contienen. De esta forma, la resolución que posee el CAD está determinada por la mínima cantidad que puede llegar a detectar, así el *LSB* o una cuenta viene determinada por la fórmula 1.3:

$$LSB = \frac{x \text{ voltios}}{2^N} \quad (1.3)$$

Siendo X la cantidad de variación mínima a máxima del voltaje de entrada. Por ejemplo, si el voltaje de entrada varía de -3 a $+3$ voltios su variación sería de $X = 6$ voltios y un CAD de 16 bit dará como resultado una cuenta de 0.36×10^{-14} V. Además del CAD, existen más equipos que son necesarios para el estudio de las señales sísmicas que se producen en un volcán. Algunos autores en sus trabajos presentan descripciones detalladas de la configuración y características de dichos equipos [Havskov y Alguacil 2004]; [Ortiz 2000]. En el presente estudio, el capítulo tres tratará con mayor detalle las zonas de estudio, las campañas y los instrumentos sísmicos utilizados.

1.6.1.- Tipos de eventos sísmicos-volcánicos

La clasificación propuesta por Minakami en 1974, fue hasta hace poco considerada como la referencia para la evaluación de las señales sísmicas. Sin embargo, en la actualidad dicha clasificación de acuerdo con los estudios que se han realizado a lo

largo de los años ha variado. Minakami clasifica las señales sísmicas en 6 tipos: señales Tipo-A, Tipo-B, Tipo-C, Período Largo (LP), Explosión y Tremor [Minakami et. al. 1974]. Sin embargo, algunos observatorios vulcanológicos han extendido esta clasificación de acuerdo a las características observadas en el proceso eruptivo de sus volcanes, ya sea de manera descriptiva o a partir del análisis del contenido de frecuencias típicas en cada señal sísmica. De hecho, la clasificación del tipo de eventos sísmicos es una materia pendiente en la regularización y normalización por parte de los diversos organismos, países y observatorios ya que cada uno interpreta los sismos locales de una forma que no necesariamente coincide con los de otra entidad, organismo o escuela. Así para un observador en algún instituto vulcanológico lo que puede ser un evento LP para otro el mismo puede ser un terremoto, una explosión u otro tipo de señal. Pero volviendo a la clasificación general, Minakami propone cuatro tipos de eventos sísmico-volcánicos:

- **Eventos tipo "A"**, Son eventos que de acuerdo a Minakami se componen de altas frecuencias y cuyas frecuencias dominantes se encuentran entre los 5 y 15 Hz.

- **Eventos tipos "B"** Eventos que se caracterizan por estar compuestos por bajas frecuencias. Las frecuencias dominantes de estos eventos se encuentran entre los 1 y 5 Hz.

Eventos sismos volcánicos explosivos, son aquellos que llegan a presentarse durante erupciones explosivas, tienen magnitudes generalmente pequeñas y pueden ser sentidos a corta distancia del volcán.

- **Eventos tremor volcánico**, Este tipo de evento se caracteriza por una vibración continua, que puede prolongarse desde varios minutos, varias horas, o incluso días, con amplitudes regulares y que muestran un contenido de frecuencia más o menos estable. Este tipo de evento es similar al de tipo B en cuanto a que sus frecuencias dominantes se encuentran a baja frecuencia entre 1 y 5 Hz. En cuanto al origen de este tipo de evento los científicos aún no se ponen de acuerdo, aunque generalmente se piensa que se producen por una oscilación continua de los elementos del edificio volcánico, o posiblemente al desplazamiento de magma en el interior del volcán.

En una extensión de esta clasificación se encuentran los denominados sismos Vulcano-Tectónico (VT). Los eventos Vulcano-Tectónicos (VT) son señales sísmicas que se asemejan claramente a los eventos de tipo tectónicos regionales, tienen frecuencias típicas de 5 Hz. o mas y las fases de P y S son claramente reconocibles. Estos eventos VT se producen en la roca sólida originando fracturamientos o fallamientos de corte muy similares a los producidos por los eventos tectónicos como los de subducción. Los eventos híbridos son eventos con componentes en las altas y bajas frecuencias, por ello se les denominan híbridos. En las altas frecuencias tienen fases impulsivas de P y S. Generalmente presentan amplitudes de P y S menores a las que presentan los eventos Vulcano-Tectónicos. Minakami indica que los sismos de tipo-A poseen fases de las ondas "P" y "S" y frecuencias mayores a 10 Hz. y el tipo-B con no claras fases de "P" y "S" y frecuencias bajas 1 a 5 Hz., [Minakami 1960]. Adicionalmente, trabajos publicados en esa misma línea han indicado que algunos sismos tectónicos (tipo-A) son de altas frecuencias a 5 Hz. En cambio los tipo-B son aquellos eventos que tienen frecuencias más bajas entre 1.0 y 1.6 Hz., también se ha caracterizado al tipo-AB, el cual tiene una componente de alta frecuencia seguida de baja frecuencia de 1.25 a 0.9 Hz. Algunos eventos los clasifican en un tipo-A con frecuencia dominante entre 6-10 Hz., y un tipo-B con frecuencias en su parte inicial de 0.5 y 2 - 4 Hz. y en su parte final entre 2 - 4 Hz. [Valdez González y González 1994],

[Lermo et. al. 1995]. Una clasificación más actual es la clasificación de los tipos de eventos sísmico-volcánicos de acuerdo a Ibáñez [Ibáñez et. al. 1997], y utilizados en trabajos como el de Wassermann para su nuevo manual de sismología [Wassermann 2002]. Ibáñez en su clasificación ha definido cinco eventos principales a seguir:

- a) Terremotos volcano-tectónicos (VT) (Figura 1.7). También llamados eventos de alta frecuencia (en inglés, High Frequency events (HF)) Ibáñez lo define como un terremoto que ocurre en un ambiente volcánico, siendo una señal que posee una duración variable, desde pocos segundos hasta varios minutos, con un contenido espectral amplio, llegando incluso a los 30 Hz. Además, posee un comienzo (llegada ondas P), algo impulsivo, en donde se identifica con facilidad la llegada de la onda S (sobre todo si se tiene el registro de las tres componentes). [Ibáñez et. al. 1997]. Wasserman indica que este tipo de evento es producido por fracturas de cizalla en grietas y que se da aproximadamente en profundidades por debajo de los 2 Km., con arribos en frecuencias superiores a 5 Hz. [Wassermann 2002].

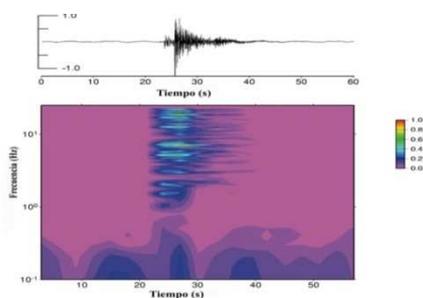


Fig. 1.7 Evento Volcano-Tectónico (VT) en el Monte Merapi, Indonesia.
Fuente: [Wassermann 2002]

- b) Eventos híbridos (Figura 1.8). Ibáñez indica que este tipo de evento posee un comienzo caracterizado por señales que se encuentran en altas frecuencias, con una amplia banda espectral (incluso superior a los 10 Hz.), al igual que el anterior, en este tipo de evento se pueden identificar las fases P y S, por lo que a veces se le confunde como un pequeño terremoto. Este tipo de evento suele asociarse con períodos pre-eruptivos [Ibáñez et. al. 1997]. Por su parte Wasserman indica que Algunas señales sísmico-volcánicas comparten las características en frecuencia de los eventos LP y VT. Este tipo de señales son generalmente llamadas eventos híbridos. Son procesos mixtos por superposición de terremotos volcano-tectónico y eventos de largo período [Wassermann 2002].

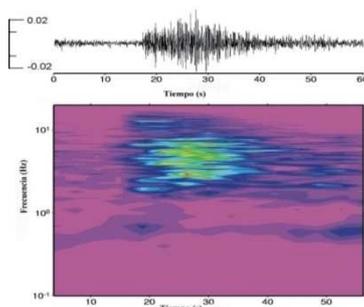


Fig. 1.8 Evento Híbrido en el Monte Merapi, Indonesia.
Fuente: [Wassermann 2002]

- c) Eventos de largo período (LP) (Figura 1.9). También llamados eventos de baja frecuencia (en inglés Low-frequency (LF) events). Ibáñez lo determina como una señal típica de ambientes volcánicos, que está caracterizada por tener una duración desde pocos segundos hasta un poco más de un minuto, además, posee un contenido espectral muy limitado en bandas de frecuencias bastante estrechas desde 0.5 hasta 5 Hz. A diferencia de los anteriores, en este tipo de evento no se pueden definir claramente las llegadas de las ondas P y S, siendo difícil determinar el inicio de la señal [Ibáñez et. al. 1997]. Wassermann por su parte, menciona respecto a este evento que se produce por fluidos a presión, y generalmente no presentan claramente una llegada de las ondas S. Este tipo de evento se da a profundidades menores a los 2 Km. Son de baja frecuencia y mayor duración (largo período), comparados con los terremotos volcano-tectónicos, aunque están restringidos en una banda de frecuencia entre 2 y 10 Hz. El espectro de este tipo de evento está formado por una sucesión de picos estrechos con un espaciado irregular, lo que le da un aspecto monocromático [Wassermann 2002].

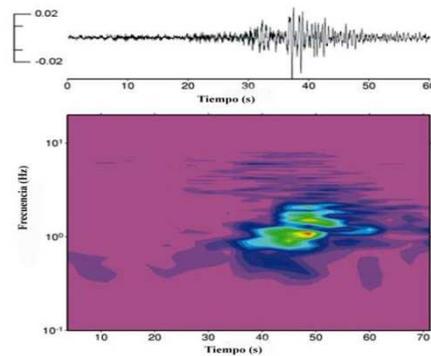


Fig. 1.9 Evento LP en el Monte Merapi, Indonesia.
Fuente: [Wassermann 2002]

- d) Tremor volcánico (Figura 1.10). Evento producido por fluidos a presión continua. Este tipo de evento se cataloga como una superposición de muchos eventos de largo período. En muchos casos se asocia a la aparición de constantes eventos de este tipo en un corto período de tiempo con la eminente actividad eruptiva en un volcán, por lo que se le considera un precursor muy significativo a la hora de poder definir y/o establecer los estados de una alerta volcánica. Ibáñez (1997), define el tremor volcánico como una señal sísmica caracterizada por mantener la amplitud constante durante un largo periodo que puede oscilar entre varios minutos y horas, incluso hasta meses, con contenido espectral centrado en bandas de frecuencia relativamente estrechas. Así, el tremor es considerado una superposición de eventos mucho más simples [Ibáñez et. al. 1997].

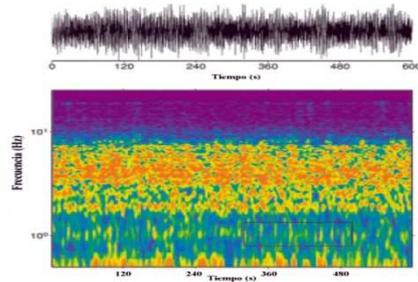


Fig. 1.10 Evento Tremor en el Monte Merapi, Indonesia.
Fuente: [Wassermann 2002]

- e) Evento explosión (Figura 1.11). Ibáñez indica que tanto este tipo de evento como el tremor, corresponden a las típicas señales sísmicas producidas durante los episodios eruptivos, y generalmente suele ocurrir debido a procesos de desgasificación, tal es el caso por ejemplo, de las explosiones del Strómboli que aparecen superpuestas sobre un tremor de fondo. Este tipo de evento se identifica por el aumento brusco de amplitud y el incremento en las frecuencias de la señal [Ibáñez et. al. 1997].

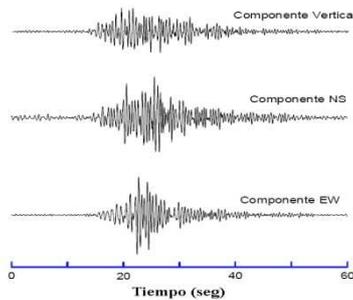


Fig. 1.11 Evento Explosión en el Strómboli, Italia.
Fuente: [Ibáñez et. al. 1997]

1.6.2.- Problemas de reconocimiento de las señales sísmico-volcánicas

Las señales sísmicas que producen los volcanes están básicamente definidas e influenciadas tanto por el medio que atraviesan y por la fuente que las origina. De hecho, tanto el propio edificio volcánico como sus alrededores presentan o están compuestos de múltiples fallas y fisuras y diversas capas de estratos y suelos, por lo que las señales, al ser producidas y al atravesar medios diversos también necesariamente van a ser igualmente muy diversas. Así se tiene que cada tipo de señal sísmica podría equivaler a un mecanismo de fuente muy diferente, como podrían ser: la fractura de medios sólidos, o los movimientos de algún líquido como la lava. Por lo tanto, hay que tener en cuenta no solo que los volcanes son diferentes, sino que las señales sísmicas que se generan en ellos también van a ser diferentes, además de que poseen una duración muy variable, incluso en un mismo período de actividad. Por otra parte, en su trabajo de análisis, los sismólogos se enfrentan a diversas dificultades, como es el poder determinar si antes de una eventual erupción o crisis volcánica se están produciendo algún o algunos tipos de eventos precursoros de dicha crisis [McNutt 1996]. Su labor se basa en poder identificar y reconocer un determinado evento y tratar de asociarle posibles características que puedan ser empleadas en una predicción sísmico-volcánica. En la actualidad, algunos de dichos eventos precursoros han podido ser identificados.

Sin embargo, la comunidad científica no ha podido llegar a un consenso definitivo, ya que ninguno de dichos eventos ha podido ser completamente confirmado como un verdadero evento precursor. Gran parte de esta dificultad radica en que la morfología de un volcán está en un continuo cambio, y lo que en una determinada crisis podría ser denominado un evento precursor, para otro período de crisis, al cambiar la situación del volcán, podría no serlo, ya que las circunstancias que han generado dicha crisis han cambiado de una erupción a otra.

Así que entre las diversas dificultades que afronta un sismólogo en un reconocimiento, se encuentra el que tiene que distinguir entre eventos que se presentan de manera periódica y eventos que se presentan como anomalías. Pero además existe una gran variedad de posibilidades entre dichas anomalías, que pueden ser de naturaleza mecánica, física y química. Por lo tanto, un evento podría catalogarse como precursor, solo cuando fuera posible estimar la probabilidad de la ocurrencia de un determinado evento; entonces se le podría llamar, “*un potencial evento precursor*”. Otro aspecto que complica el estudio y la observación de los eventos sísmico-volcánicos son las diferencias entre los métodos para la determinación de los probables eventos precursores cuando estos se producen a corto, mediano y largo plazo. A largo plazo, la mayoría de las predicciones que actualmente se realizan, generalmente se refieren a las tendencias de una determinada actividad sísmica, de una región específica, y de acuerdo con un determinado tratamiento probabilístico. Esto no es más que el lapso en el que se presentan a largo plazo; ya sea los grandes eventos sísmicos, las brechas sísmicas, o las acumulación de deformaciones, entre otros factores. Cuando esto ocurre a mediano plazo (meses o pocos años), los métodos estadísticos antes aceptados, ahora son generalmente rechazados. En estos casos se utilizan diversos factores como fenómenos precursores, entre los que se encuentran;

- Una rápida liberación de energía,
- Valores irregulares en las velocidades de propagación de ondas,
- La deformación en la corteza, lo que se denomina “*domo volcánico*”,
- Cambios drásticos de resistividad del suelo,
- Cambios drásticos del campo geomagnético,
- Cambios drásticos del nivel de agua subterránea,
- Cambios drásticos de los contenidos químicos de los acuíferos en el volcán.

Al hablar de un corto plazo, o en tiempo real, generalmente se utilizan precursores característicos de los que se han mencionado a mediano plazo, pero llevados a un nivel mucho más intenso en cuanto a los fenómenos observados. Aquí es donde precisamente no se debe de perder de vista, el poder reconocer, que tipo de eventos se están generando, para así poder determinar si dichos eventos pueden tratarse de potenciales precursores. Otro punto de vista de acuerdo a autores como Blong y Tilling, presentan que la dificultad radica en poder detectar los eventos sísmico-volcánicos, identificar sus causas, sus características, establecer los pronósticos de evolución y con toda esta información poder manejar adecuadamente parámetros de alertas predictivas [Blong 1984]; [Tilling 1989]. Esta complejidad puede reflejarse al referirse a señales sísmicas, por ejemplo las de tipo tremor volcánico (mencionado al inicio de la sección 1.6), las cuales pueden deducirse como un producto de la resonancia de estructuras rellenas de fluidos, ya sea estáticos o en movimiento, que están ubicados en el interior de los volcanes activos. Por lo que un mismo tipo de señal como en este caso el tremor, varía dependiendo de cómo ha sido generado, y esta complejidad se

extiende a los otros tipos de señales sísmico-volcánicas. En este sentido, autores como Garcés y Hanses en un análisis sobre el estudio de la erupción ocurrida en el volcán Pavlof en Alaska describen al Tremor como señales de baja frecuencia limitadas en banda, con un pico de energía situado en torno a los 2Hz [Garcés y Hansen 1997]. Otros autores como Konstantinou y Lin en 2004 realizan un análisis comparativo del Tremor en diferentes volcanes y reafirman la baja concentración de energía en la que se sitúa este evento [Konstantinou y Lin 2004]. Chouet en 1992 menciona que el tremor se identifica por una señal continua de amplitud sostenida en frecuencia con ancho de banda limitado [Chouet 1992]. En muchas ocasiones el tremor se enmascara conjuntamente con las diferentes manifestaciones de ruido lo que hace que la identificación y análisis de los eventos sean bastante complicados. Saccorotti en 2004, ha indicado que el Tremor es uno de los temas de estudio de mayor dificultad para un sismólogo [Saccorotti et al. 2004]. Todo lo anterior solo en cuanto al tremor, y hay que tomar en cuenta que en volcán genera múltiples tipos de señales en diferentes circunstancias. Por lo tanto, en la vigilancia o monitoreo volcánico, el poder reconocer y determinar con precisión los diferentes tipos de eventos sísmico-volcánicos que ocurren en una determinada crisis, corresponde a una tarea bastante delicada y que requiere de mucho análisis.

1.7.- FAMILIA DE EVENTOS SÍSMICO-VOLCÁNICOS

Las familias de eventos sísmicos o también conocidas como familias de terremotos son un conjunto de eventos con los mismos hipocentros o relativamente cercanos. Además presentan “*formas de onda similares*”, estas características sugieren un origen común. Se han observado familias de eventos sísmicos tanto en la sismicidad tectónica como la volcánica. Trabajos de Hamaguchi y Hasegawa en 1975 localizaron las replicas del terremoto de 1968 en Tokachi Oki Japón y notaron la similitud de las formas de ondas de eventos que ocurren a la misma profundidad [Hamaguchi y Hasegawa 1975]. Por otra parte, para el caso de los sismos de origen volcánico Espíndola y asociados identificaron 3 familias de eventos sísmicos en el volcán Popocatepetl en México [Espíndola J. et. al. 2005]. Asimismo, Mizukoshi y Moriya en 1980 y Okada en 1981, reconocieron grupos de eventos similares, asociados al período de actividad de 1977-1978 en el volcán Usu en Japón [Mizukoshi y Moriya 1980], [Okada et. al. 1981]. Fremont y Malone en 1987 también reportaron familias de eventos en la sismicidad del monte Santa Elena [Fremont and Malone 1987]. En otros volcanes como el Tarumi de Japón Nishimura en 1992 observa familias de eventos durante el período de actividad de Enero a Febrero de 1988 [Nishimura et. al. 1992]. El análisis de las familias de eventos sísmicos presentes en los volcanes activos puede resultar de suma importancia a la hora de poder definir eventos precursores de eminentes erupciones en los volcanes. Tal es el caso de eventos precursores o de disparo observados por Neuberg en 2006 en familias de enjambres sísmicos de baja intensidad en volcanes como el Soufrière Hills en la isla de Montserrat [Neuberg et. al. 2006]. El hecho de que existan familias de terremotos en ambientes volcánicos, aún no ha podido ser del todo evaluado, en particular los eventos del tipo Tectónico-volcánicos, sin embargo, su existencia podría dar pistas sobre lo que ocurre en el interior de un volcán. Para el caso de los volcanes de estudio en el presente trabajo, los análisis que se han realizado, podrían indicar la posibilidad de existencia de familias de sismos en los volcanes de Nicaragua, lamentablemente no se cuenta con la cantidad de datos necesarios para afirmar tal suposición, de forma que este hecho podría ser causa de futuras investigaciones.





CAPÍTULO II

METODOLOGÍA Y HERRAMIENTAS

*Interior del Cráter del Volcán
Cerro Negro, Nicaragua.
Por: Ligdamis A. Gutiérrez E.*





2.1.- INTRODUCCIÓN

En este capítulo, se establece de forma teórica, como se realizará el análisis de las señales recopiladas en las distintas campañas de Italia y Nicaragua. Se presentará por lo tanto, el marco teórico matemático que permitirá el estudio de las señales sísmico-volcánicas, la caracterización de las mismas y el diseño y construcción de un sistema de reconocimiento automático de eventos sísmicos producido por los volcanes en estudio. En este capítulo se establecen las bases del análisis espectral de las señales, que va a permitir el poder identificar los diferentes tipos de eventos, que se utilizarán para el etiquetado y construcción de las bases de datos de entrenamiento. Dicho análisis se realiza en el dominio de la frecuencia, haciendo uso de la transformada de Fourier y el espectrograma de las señales. También en este capítulo, se determinará la metodología a aplicar en el sistema de reconocimiento basado en los Modelos Ocultos de Markov. El análisis aquí propuesto, podría en un futuro complementarse con herramientas adicionales, mismas que se detallan en los apéndices (ver apéndices D y E).

2.2.- LA TRANSFORMADA DE FOURIER Y EL ESPECTROGRAMA

En el estudio de las señales sísmico-volcánicas, una herramienta muy útil y de uso muy extendido es el análisis del contenido en frecuencias de la señal. Un ejemplo que lo puede demostrar es el de la senoide infinita, que ve concentrada toda su energía en una sola frecuencia, la frecuencia de la senoide. Al considerar una señal sísmico-volcánica en el dominio de tiempo, esta puede ser expresada como la suma de n señales sinusoidales, cada una conteniendo una amplitud, frecuencia y fase determinada, por lo que es posible llegar a diferenciar claramente los registros de señales sísmico-volcánicas por su contenido espectral. La “*Transformada de Fourier*” permite transformar una señal expresada en el dominio de tiempo, en una expresada en el dominio de la frecuencia. Es decir, que al aplicar la transformada de Fourier a las señales sísmico-volcánicas, será posible ver el contenido espectral que se busca.

2.2.1.- La transformada de Fourier

En el ámbito analógico, la serie y la transformada de Fourier desempeñan un rol importante en la representación de señales y en el análisis de sistemas lineales. En la actualidad, con los avances de los ordenadores digitales, y el diseño de algoritmos por ordenador, ha sido imprescindible extender el uso de este tipo de herramientas matemáticas al mundo discreto. Como resultado de esto se desarrolló lo que se conoce como la Transformada Discreta de Fourier (TDF). Cooley y Tukey en 1965 publican un trabajo acerca de un algoritmo para el cálculo rápido de la Transformada Discreta de Fourier mediante ordenador [Cooley y Tukey, 1965]. Esto da lugar al resurgimiento de los estudios basados en espectros para aprovechar la flexibilidad y potencialidades de la nueva herramienta.

El análisis de Fourier se dedica al estudio de señales continuas o discretas, ya sea en dominio temporal o frecuencial, y para señales periódicas y no periódicas [Bracewell 1986]; [Sodré 2002]. Todas ellas consisten en la representación de la señal original en función de una base sinusoidal compleja completa. En sus versiones más avanzadas estudia: procesos estocásticos, funciones de distribución, y topologías más complejas, pero sus fundamentos siguen siendo muy simples.

Sea $x(t)$ una señal continua. Se define como la transformada de Fourier de x , $X(\omega)$, la función determinada por la ecuación (2.1) [Oppenheim et. al. 1998]:

$$X(\omega) = \int_{-\infty}^{\infty} x(t)e^{-j\omega t} dt \quad (2.1)$$

La distribución frecuencial, $X(\omega)$, representa el contenido en frecuencia de la señal $x(t)$, y se le denomina el espectro. Otras expresiones alternativas de la ecuación anterior, sustituyen la variable ω por el valor $2\pi f$. [Rubio A. 2002]. Al ser $X(\omega)$ compleja, sus valores tendrán: (a) una parte real y una imaginaria, y (b) una amplitud y una fase. Normalmente se representa en términos de módulo, $|X(\omega)|$ y argumento, $\theta(\omega)$, determinando lo que se denomina el espectro de módulo y de fase respectivamente.

También existe la Transformada inversa de Fourier, que permite obtener la señal en el dominio temporal a partir de su contenido en frecuencia y está definida por la ecuación (2.2)

$$X(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} x(\omega)e^{j\omega t} d\omega \quad (2.2)$$

La ecuación 2.1 se denomina también ecuación de análisis y la ecuación 2.2 ecuación de síntesis. Cabe mencionar que las expresiones 2.1 y 2.2 corresponden al par de transformadas de Fourier en tiempo continuo. Es decir, permite calcular una distribución frecuencial continua a partir de una señal en tiempo continuo.

Cuando la señal a analizar es en tiempo discreto, se denomina la Transformada Discreta de Fourier (DFT), que transforma N puntos en dominio del tiempo en N puntos en el dominio de la frecuencia. Aquí la formulación varía, pudiendo adoptar y formar el par de ecuaciones; de análisis y de síntesis (N es la longitud de $x[n]$) siguientes:

$$X(k) = \sum_{n=0}^{N-1} x(n) \cdot W^{nk} \quad k = 0, 1, \dots, N-1 \quad (2.3)$$

$$x(n) = \frac{1}{N} \sum_{k=0}^{N-1} X(k) \cdot W^{-nk} \quad n = 0, 1, \dots, N-1 \quad (2.4)$$

En donde las constantes “ W ” se conocen como los factores “twiddle” y son definidos como $W = e^{-j2\pi/N}$, que permiten obtener una representación frecuencial discreta a partir de una señal en tiempo discreto. La bondad de la aproximación depende de dos factores fundamentales: *la frecuencia de muestreo y el truncamiento de la señal*. La frecuencia de muestreo es un factor crítico en la digitalización de una señal continua para que no se distorsione la información que conlleva la señal $x(t)$ al transformarse en una serie de valores discretos (x_k). Para un intervalo de muestreo dado (Δt), la frecuencia de Nyquist es dos veces la frecuencia máxima de la señal $F_N = 2F_{Max}$. Si el

espectro de la señal $x(t)$ original tiene componentes superiores a la frecuencia de Nyquist, tiene lugar el fenómeno que se conoce como *solapamiento* o "*aliasing*", es decir, "las componentes de la señal que corresponden a frecuencias superiores a (f_N) se asignan en la TDF a frecuencias inferiores, o sea dentro de la banda de Nyquist ($|f| < f_N$), con la consiguiente distorsión de la TDF".

La siguiente figura muestra la representación de la imagen del espectro de frecuencia, al aplicar la transformada de Fourier a una señal.

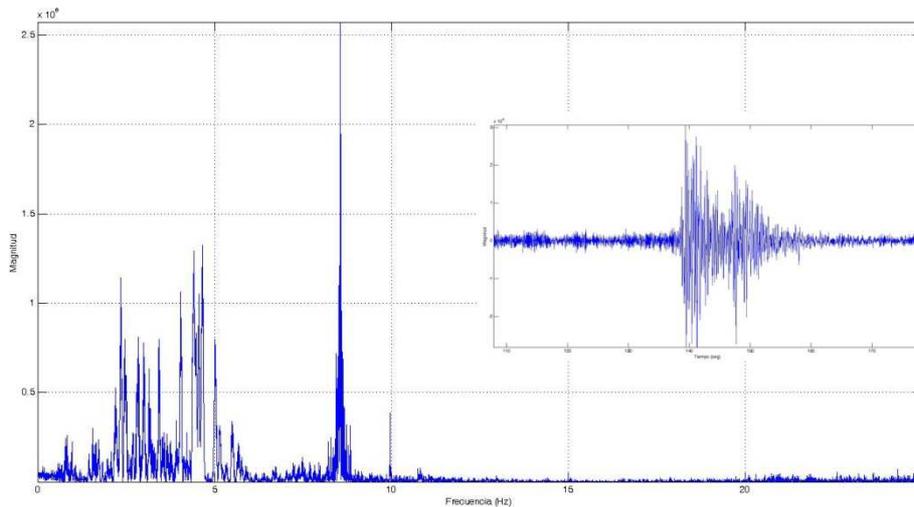


Fig. 2.1 Gráfica de la imagen de la Transformada de Fourier de una señal. En el recuadro se observa la señal en el tiempo.

2.2.2.- El espectrograma

Una correcta clasificación de las señales sísmicas producidas en un volcán, presupone que al ser analizadas no sólo se considere sus formas de onda, sino que, por la gran variedad o complejidad de las fuentes que generan dichas señales y los efectos de las diversas mezclas de que se compone el material volcánico. En las señales sísmico-volcánicas, el contenido espectral no se mantiene a lo largo del tiempo. Es preciso además complementar el análisis realizado en dominio del tiempo, con procedimientos en el dominio de la frecuencia y así aprovechar la información contenida en el espectro de las señales sísmico-volcánicas. Teniendo en cuenta esto, hay dos herramientas de cálculo apropiadas para analizar y representar una señal: el espectro mediante la transformada de Fourier (visto anteriormente) y el denominado espectrograma.

El espectrograma [Nawab y Quatieri 1988]; [Rioul y Flandrin 1992] es el registro gráfico que se obtiene al calcular mediante la transformada de Fourier el espectro de las tramas enventanadas de una señal, o lo que es lo mismo, realizar una estimación del espectro de segmentos consecutivos de la señal solapados entre sí y cuya resultante es una gráfica. Dicha gráfica representa la magnitud de la función calculada que es la evolución del contenido frecuencial de la señal, según va variando a lo largo del tiempo. La resolución que se obtiene en frecuencia depende de la duración del segmento de análisis, de modo que con segmentos de corta duración, se obtiene alta

resolución temporal y poca resolución en frecuencia; a medida que se aumenta la duración del segmento se obtiene menor resolución temporal y mayor resolución en frecuencia. La figura 2.2 representa el espectrograma de una señal sísmica, muestra una gráfica de las variaciones de la frecuencia en el eje vertical (eje y); cada índice del eje x representa un intervalo de tiempo, esto es un segmento de análisis y donde la intensidad de los colores o los grises indican el valor de la energía de la señal (que en el caso del color, las amplitudes más fuertes corresponden al rojo oscuro y las más débiles al violeta/azul claro). En el caso del presente trabajo se utilizará el espectrograma para identificar los diferentes tipos de eventos sísmicos presentes en los volcanes.

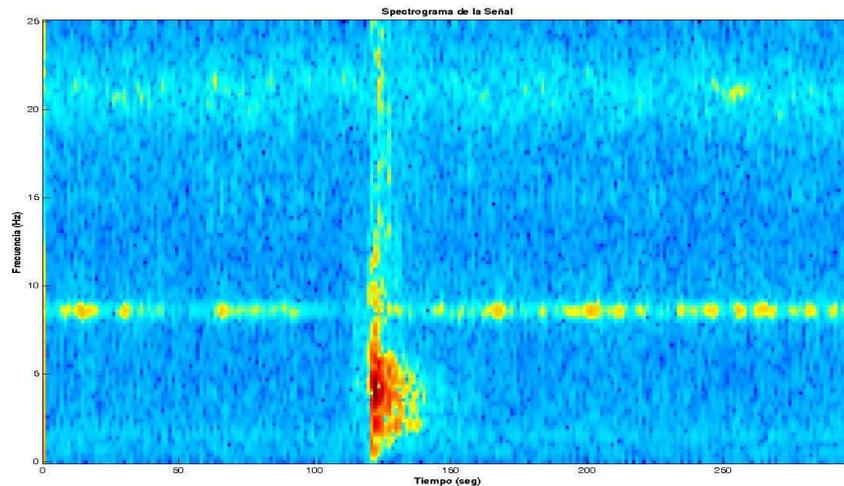


Fig. 2.2 Gráfica del Espectrograma de una señal

Adicionalmente al espectrograma y al análisis de Fourier, podrían utilizarse a futuro análisis mediante el uso “poder espectral” y “transformadas Wavelet” (Ver Apéndice D).

2.3. DESCRIPCIÓN DE LA SEÑAL EN EL DOMINIO CEPSTRAL

En las aplicaciones de reconocimiento y clasificación es conveniente disponer de una representación paramétrica de la señal. El principal objetivo de la caracterización paramétrica de las señales sísmicas es transformar la señal de entrada en un conjunto de vectores que contenga toda la información discriminadora presente en la señal. En otras palabras, es extraer la información relevante contenida en la señal. Aunque existen numerosas posibilidades para ello, esta sección se centrará en una representación de la envolvente espectral de tiempo corto, basada en el análisis del espectro mediante un banco de filtros. Diversos trabajos muestran la utilidad de su uso en el reconocimiento de señales [Nadeu et. al. 1995] [Nadeu et. al. 2001], [Haddad y Park 1995]. Este es un procedimiento alternativo al método de análisis de predicción lineal (*LP*) [Makhoul 1975], [Rabiner y Juang 1993] que ha sido utilizado en numerosas ocasiones en el análisis de señales sísmicas [Del Pezzo et. al. 2003], [Scarpetta et. al. 2005].

Para el diseño del Banco de Filtros, el rango de frecuencias comprendido entre una frecuencia inferior y la frecuencia de Nyquist se divide en un número determinado N de filtros paso-banda, solapados entre sí de anchura L . La forma de los filtros puede variar, pero a excepción del primero y del último, la suma de los filtros es uno. La figura 2.3 muestra un ejemplo de banco de filtros triangulares. A la salida de cada filtro se obtiene el promedio de energía en la banda de frecuencias del filtro.

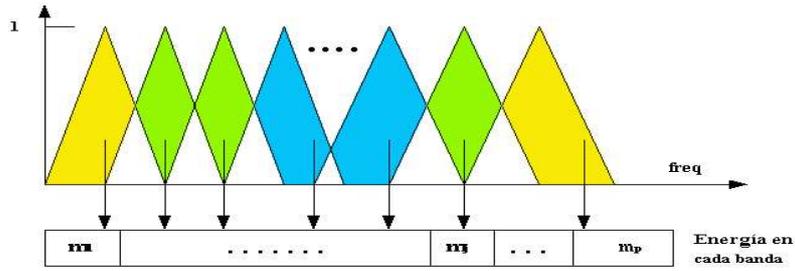


Fig. 2.3 Banco de filtros triangulares. Fuente: [Young et. al. (2002)]

Para obtener una representación paramétrica de la señal, en primer lugar se procede de forma similar a la descrita anteriormente para obtener el espectrograma; esto es, se calcula el espectro en segmentos consecutivos y solapados entre sí de la señal, se obtiene el módulo de la transformada de Fourier para cada segmento y se aplica el banco de filtros. El número de parámetros obtenidos coincide con el número de filtros que definen el banco de filtros. Adicionalmente también se evalúan otros parámetros como la energía de la señal en cada segmento [Young et. al 2002]. A los parámetros así obtenidos se le aplica una compresión logarítmica. Debido al solapamiento de los filtros los parámetros obtenidos están fuertemente correlacionados, para decorrelacionar los coeficientes se aplica la Transformada Discreta del Coseno (DCT) [Ahmed et. al. 1974], [Khayam 2003], [Strang 1999]. Como último paso, para incluir información contextual del segmento se incluyen las derivadas de primer y segundo orden de los parámetros obtenidos dentro del vector de características. La figura 2.4 presenta gráficamente este proceso.

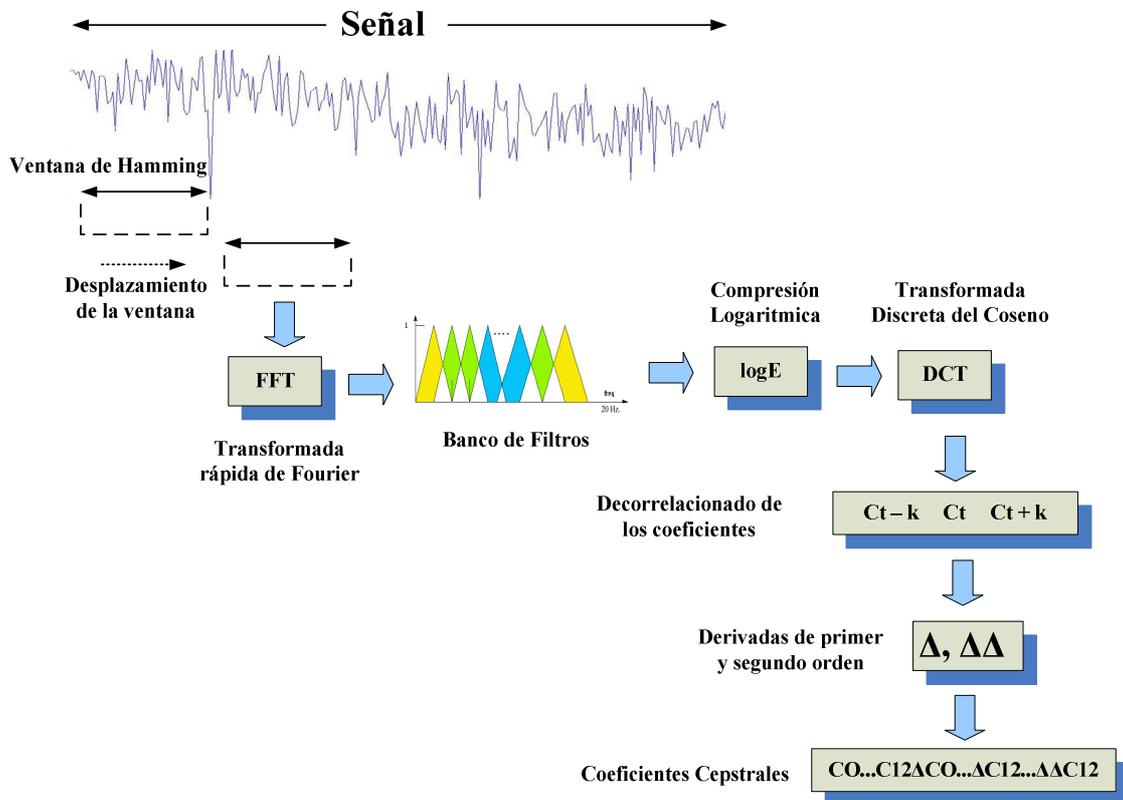


Fig. 2.4 Representación paramétrica de la señal

2.4 LOS MODELOS OCULTOS DE MARKOV. APLICACIÓN A UN SISTEMA DE RECONOCIMIENTO

Desde sus inicios en el tratamiento de reconocimiento de voz los HMMs han sido la técnica más empleada a nivel mundial. En los 70's los primeros trabajos en este sentido dictaminaron esa línea. Tales fueron las aportaciones de Backer [Baker 1975] y Jelinek [Jelinek 1976]. Ya en los 80's esto hace indicar que los HMMs son lo suficientemente potentes como para modelar adecuadamente casi la totalidad o la mayor parte de las fuentes de variabilidad presentes en el habla [Lee 1989]. De la misma forma, los HMMs han sido utilizados en diferentes trabajos para el reconocimiento de señales sísmicas de origen volcánico [Benitez et. al. 2007]; [Cortes et. al. 2009]; [Gutiérrez et. al. 2009]; [Gutiérrez et. al. 2011]; [Ibáñez et. al. 2009].

2.4.1.- Definición de un modelo oculto de Markov

Los modelos ocultos de Markov [Rabiner y Juang 1986] se desarrollan a partir de los procesos de Markov. Éstos se pueden describir mediante un conjunto de estados y un conjunto de probabilidades de transición de un estado a otro. Un proceso oculto de Markov se define como un proceso doblemente estocástico, uno de los cuales, la secuencia de estados, no es directamente observable, el otro proceso, la secuencia de observaciones, sí es observable. Un HMM es, por tanto, un modelo estadístico, que describe la producción de una secuencia de observaciones generada por una secuencia oculta de estados.

Un HMM está definido por un conjunto de estados interconectados entre sí; la matriz $A=\{a_{ij}\}$ define las probabilidades de transición entre dichos estados; la matriz $\boldsymbol{\pi}=\{\pi_i\}$ de probabilidad de ocupación inicial de esos estados y por una matriz B de probabilidad de emisión de observaciones en cada uno de los estados. Por tanto un HMM λ queda definido por la terna $A,\boldsymbol{\pi},B$. ; $\lambda = (A,\boldsymbol{\pi},B)$

- a) La matriz $A = \{a_{ij}\}$ que define la distribución de probabilidades entre los estados se determina por la ecuación 2.5

$$a_{ij} = P[q_{t+1} = j | q_t = i], \quad 1 \leq i, j \leq N \quad \text{con} \quad \sum_{j=1}^N a_{ij} = 1 \quad \forall i \quad (2.5)$$

En donde N es el número de estados del modelo, siendo las etiquetas individuales de los estados $\{1,2,3,\dots,N\}$ y el estado en el momento t como q_t

- b) La matriz $\boldsymbol{\pi}=\{\pi_i\}$ que define la probabilidad de distribución inicial de los estados se determina por la ecuación 2.6

$$\pi_i = P [q_1 = i], \quad 1 \leq i \leq N \quad \text{con} \quad \sum_{i=1}^N \pi_i = 1 \quad (2.6)$$

- c) La matriz $B = \{b_j(k)\}$ que define la distribución de probabilidad de la emisión de observaciones en cada estado se determina por la ecuación 2.7 que define el símbolo de distribución en el estado j , siendo $j = 1,2,\dots,N$

$$b_j(k) = P[o_t = v_k | q_t = j], \quad 1 \leq k \leq M \quad \text{con} \quad \sum_{k=1}^M b_j(k) = 1 \quad \forall j \quad (2.7)$$

En donde M es el número de símbolos de las distintas observaciones por estado, y donde los símbolos individuales están denotados por $V = \{v_1, v_2, \dots, v_M\}$

2.4.2 HMM como generador de secuencias de observaciones

Un ejemplo típico de un Modelo Oculto de Markov es un autómata de cuatro estados presentado en la figura 2.5 en el que se pueden admitir transiciones desde un estado hacia cualquier otro. Para el tratamiento de señales, los *HMMs* son entrenados para que generen una secuencia observable, correspondiente a un segmento determinado en cada tiempo t cuando un estado j admite la transición (una secuencia de observaciones determinadas por vectores O_t). Donde la secuencia de observaciones; $O = \{O_1, O_2, O_3, O_4\}$ y la secuencia de estados $X = [1,2,3,4]$. El proceso comienza en uno o varios estados diseñados para tal efecto, donde se generan secuencias de observaciones al pasar de un estado a otro y se genera un símbolo de salida asociado a ese estado. Así, la elección de la transición y del símbolo es aleatoria. En este sentido, el vector O_t , está gobernado o es generado por distribuciones de probabilidad $b_j(O_t)$. Por lo tanto, los *HMMs* pueden ser considerados como un proceso en donde la secuencia de símbolos "O" de salida, generados a lo largo del tiempo es observable, pero la secuencia de estados X , por los que se ha pasado para generar la anterior serie se desconoce, es decir, permanece oculto, y de ahí es de donde viene el nombre de modelos ocultos.

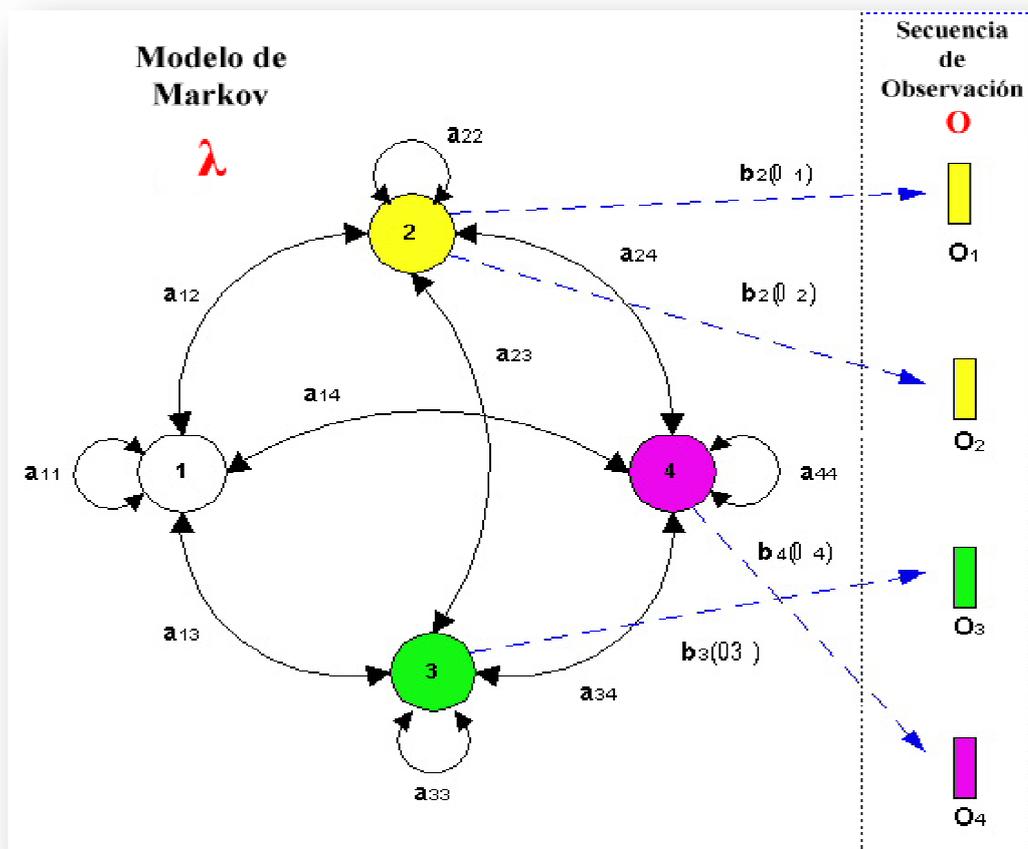


Fig. 2.5 HMM de un Autómata de Estados finitos con 4 estados

2.4.3 Los tres problemas básicos y soluciones propuestas.

Dado un HMM hay tres problemas de interés que hay que resolver para que se puedan usar en aplicaciones reales [Rabiner 1989], [Rabiner y Juang 1993]. Estos problemas son:

- a) **Problema de Evaluación:** dada una secuencia de observaciones $O = (O_1, \dots, O_t)$ y un modelo $\lambda = (A, B, \pi)$, se busca cómo calcular eficientemente la probabilidad $P(O|\lambda)$ de la secuencia observada dado el modelo. La solución a este problema se da mediante el **algoritmo de adelante-atrás (forward-backward)**. Una descripción detallada se presenta en el apéndice I.
- b) **Problema de Decodificación:** dada una secuencia de observaciones y un modelo, se busca cómo elegir una secuencia de estados que genere esas observaciones, que sea óptima en algún sentido. Como solución a este problema se propone el Algoritmo de Viterbi [Viterbi 1967]; [Forney 1973]. Una descripción de dicho algoritmo se encuentra detallada en el apéndice I.

La figura 2.6 muestra el algoritmo de Viterbi, que puede definirse como la búsqueda del mejor camino a través de una matriz. Donde las dimensiones verticales representan los estados de un *HMM* y las dimensiones horizontales representan los segmentos de las señales en el tiempo. Cada punto de convergencia dentro de esta matriz representa la probabilidad logarítmica de la observación del segmento en el tiempo y el arco que une dos puntos, corresponde a la probabilidad logarítmica de la transición entre los estados del modelo. La probabilidad logarítmica de cualquier camino es computada simplemente por la suma de las probabilidades logarítmicas de las transiciones y la salida de las probabilidades logarítmicas a lo largo de ese camino. Los caminos así van creciendo de izquierda a derecha columna a columna. De esta forma, en un tiempo determinado cada camino en particular es conocido por todos los estados.

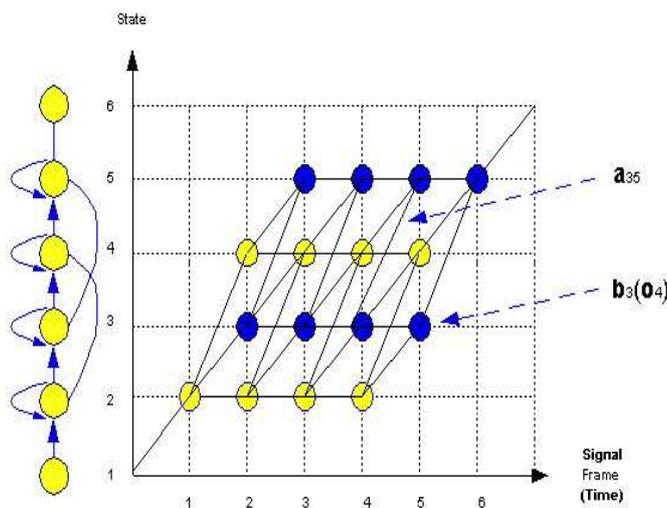


Fig. 2.6 Funcionamiento del algoritmo de Viterbi para el reconocimiento de señales. Donde la métrica de cada rama es "a" y la métrica de cada estado es b. Figura modificada del original tomada de [Young et. al. (2002)]

Por lo tanto, se trata de obtener un algoritmo recursivo que permita encontrar un modelo en el cual los campos posean un máximo valor de probabilidad de reconocimiento, lo que se obtiene mediante la aplicación del algoritmo de Viterbi.

- c) **Problema de aprendizaje:** dada una secuencia de observaciones conocidas se busca cómo determinar los parámetros del modelo de forma óptima. Como solución a este problema se propone el Algoritmo de Baum-Welch. Una descripción detallada se presenta en el apéndice I.

El llamado método de re-estimación Baum-Welch, deriva del algoritmo de máxima expectativa (por sus siglas en inglés EM; *Expectation Maximization*), utilizado en diferentes trabajos [Dempster et. al. 1977]; [Redner y Walter 1984]; [Jordan y Jacobs 1994]; [Bilmes 1998]. En primer lugar, hay que notar que la inclusión de múltiples flujos de datos no altera significativamente el conjunto, desde que se considera que cada flujo de datos es estadísticamente independiente. Además, puede considerarse que los componentes (M) de una mezcla (j) son una forma especial de *sub-estado* ($a_{ij} c_j$) en que las probabilidades de la transición representan el peso de la mezcla (a_{ij}) [Young et. al. (2002)]. Esto se representa gráficamente en la figura 2.7.

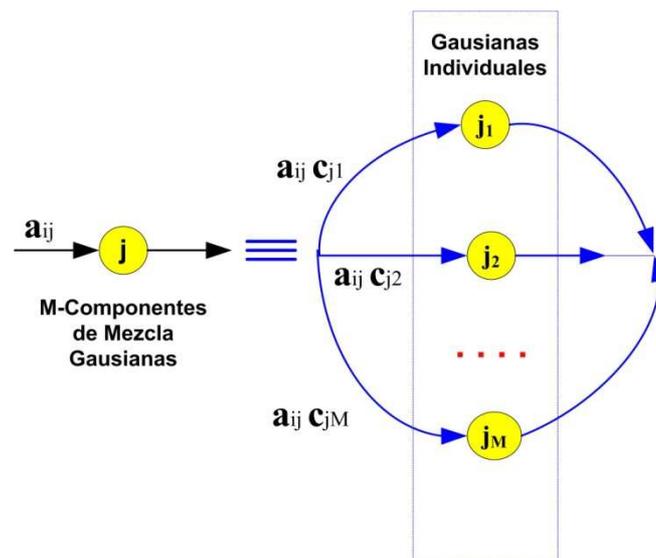


Fig. 2.7 Representación de una mezcla de componentes Gaussianas. Figura modificada del original tomada de [Young et. al. (2002)]

2.5 ESQUEMA DE UN SISTEMA DE RECONOCIMIENTO BASADO EN HMM

Un sistema de clasificación automático generalmente se compone de tres etapas bien definidas que son:

- a) Adquisición de los datos.
- b) Extracción de parámetros o características representativas de los datos
- c) Clasificación de los datos.

Los volcanes generan registros en continuo, por lo que un sistema de reconocimiento automático que se encuentre basado en HMM, necesariamente va a trabajar con registros de datos en continuo, así tiene que realizar dos acciones: la detección del evento y la clasificación de dicho evento. En la figura 2.8 se presenta un diagrama que muestra la arquitectura de un sistema típico de reconocimiento de patrones basado en HMM. En donde se pueden distinguir dos fases principales que son:

- i) Fase de entrenamiento
- ii) Fase de reconocimiento

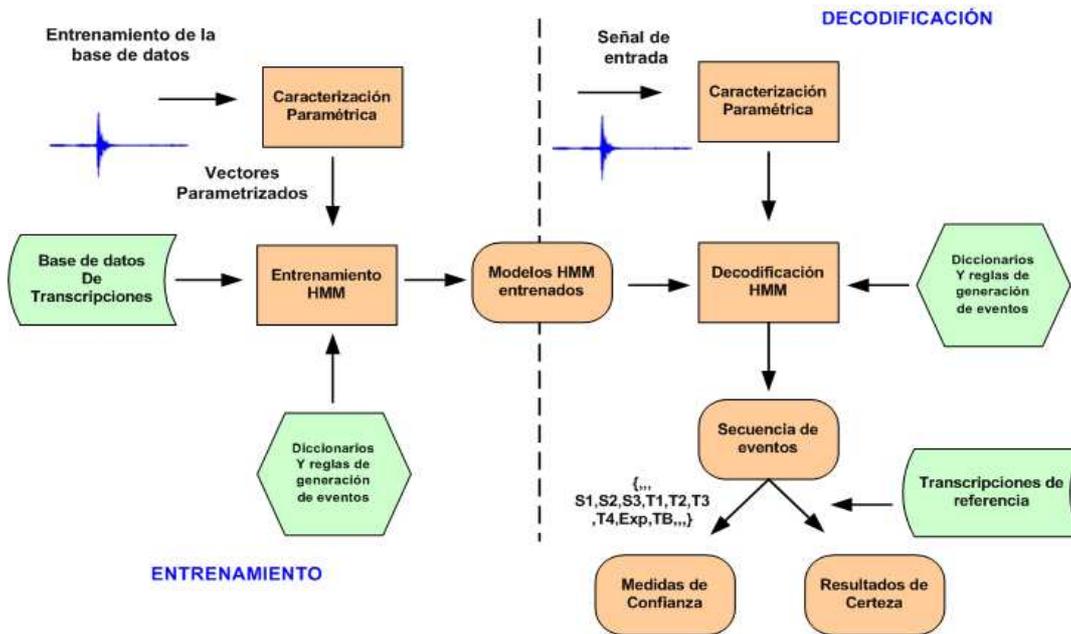


Fig. 2.8 .Arquitectura de un sistema de monitorización sísmico-volcánico basado en modelos HMMs.

En la fase de entrenamiento, a partir de un conjunto de señales sísmicas de origen volcánico que contienen eventos conocidos, esto es a partir de un conjunto de datos etiquetado, se entrena un modelo HMM para cada evento.

En la fase de decodificación, se utilizan los modelos previamente entrenados para “etiquetar” de forma automática una señal de entrada cuyo contenido es desconocido.

2.6.- IMPLEMENTACIÓN CON HTK

2.6.1.- Descripción de HTK

El software que se utiliza para construir y manipular los modelos de Markov es el HTK (Hidden Markov Model Toolkit) [Aguilar 2007], [Young y Young 1993], [Young et. al. 2002]. El HTK es una herramienta que consiste en un denominado “toolkit portable”, compuesto por un conjunto de librerías y herramientas desarrolladas en lenguaje C, utilizado para la construcción y manipulación de modelos de Markov. Dichas herramientas facilitan el análisis de las señales, así como el entrenamiento de

modelos HMM, realizan test de reconocimiento y la extracción de resultados. El software soporta la creación de modelos HMM con distribuciones continuas de mezclas de Gaussianas o por medio de distribuciones discretas, pudiendo así crear diferentes configuraciones de HMM. En un inicio dichas herramientas fueron diseñadas principalmente para construir modelos basados en el procesamiento de señales de habla, posteriormente se han destinado para diversas aplicaciones, como las orientadas al análisis de señales sísmicas producidas en volcanes activos. En el presente trabajo, se han utilizado para construir un sistema de reconocimiento y clasificación de señales sísmicas de origen volcánico.

2.6.2.- Interfaz de usuario

La presente Tesis utiliza HTK-v.3.2 sobre el sistema operativo LINUX. Una vez instalado el HTK dispone de varias herramientas para construir y manipular los HMM. En la figura 2.9 se observan las principales herramientas de acuerdo a la etapa del sistema para la cual están diseñadas. La manipulación de las órdenes y el uso de las herramientas en el presente trabajo, se hace mediante scripts con un intérprete de comandos de UNIX, la C-shell (csh), los principales scripts se encuentran descritos en los apéndices F y H.

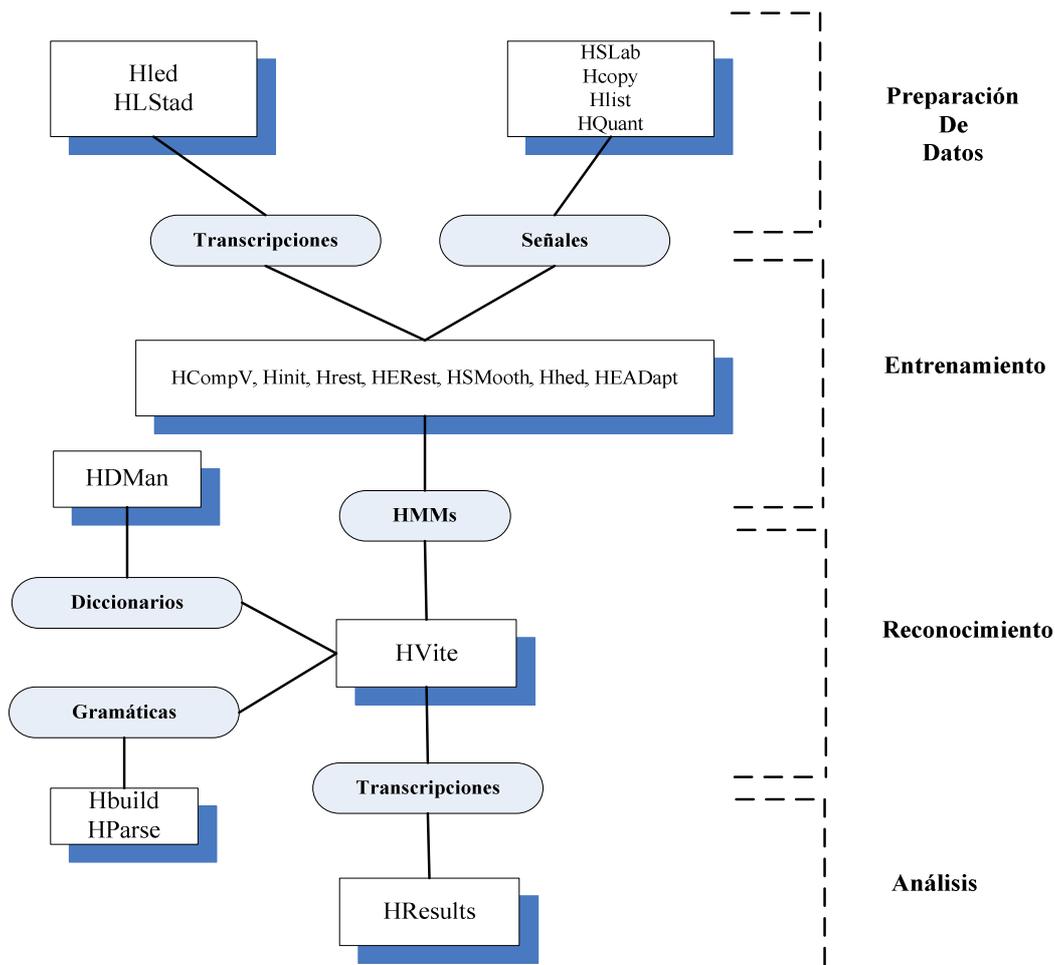


Fig. 2.9 Principales herramientas de HTK de acuerdo a las etapas del sistema. Figura modificada del original tomada de [Young et. al. (2002)]

La figura anterior muestra definidas las cuatro etapas de las que se compone todo el proceso de los HMM que se maneja en este trabajo; preparación de los datos, entrenamiento, reconocimiento y análisis. A continuación se especificarán cada una de ellas.

2.6.3.- Preparación de los datos

En la fase de preparación de los datos, el HTK permite realizar la parametrización que ha sido descrita en la sección 2.3, de los datos de entrenamiento y reconocimiento, lo que se realiza mediante la herramienta HCopy. Este proceso realiza la construcción de transcripciones que asocian los datos del entrenamiento (observables) con los símbolos que se van a reconocer. Los parámetros como la longitud de la ventana a analizar, el número de coeficientes cepstrales a determinar, entre otros puntos se especifica en un fichero de configuración. En el apéndice H se realiza una descripción detallada de dicho fichero. La figura 2.10 presenta un esquema general de los elementos de la parametrización de los datos.

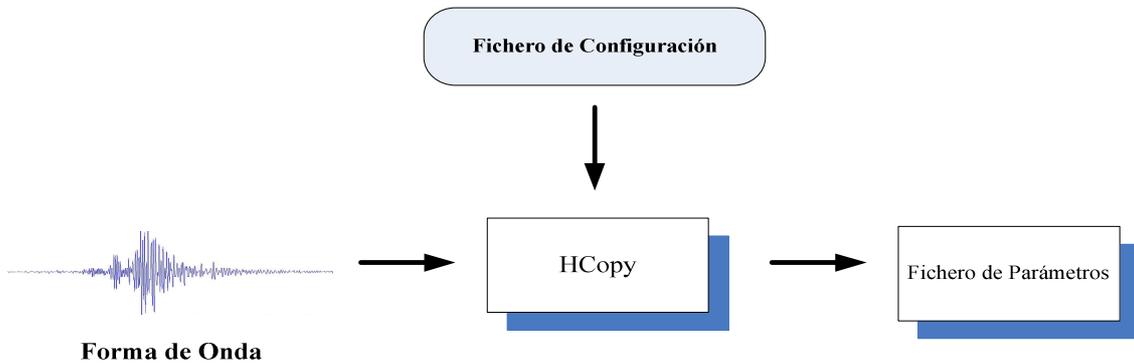


Fig. 2.10 Esquema de la parametrización de datos con HTK

En esta etapa se realiza también la creación y definición de las gramáticas y de los diccionarios de eventos que se van a utilizar. En el apéndice H se realiza una descripción de la creación de estos dos procesos. En el caso de las gramáticas, se fijan las leyes que deben de cumplir con la secuencia de los observables a reconocer. Por ejemplo se define que detrás de cada evento de un tipo denominado “X”, venga otro evento de un tipo denominado “Y”, y que entre ellos exista un tercer tipo denominado “Z”, que puede ser silencio, pausas o en nuestro caso ruido. Sin embargo, inicialmente en el caso que nos ocupa, no existe un reconocimiento previo que permita calcular probabilidades de transiciones entre los eventos, de modo que se asume que todos los eventos son igualmente probables de ser reconocidos, por lo que después de un evento puede ir cualquier otro, es decir se asume lo que se conoce como gramática nula.

En esta etapa también se realiza el etiquetado manual de los datos, que es uno de los procesos más importantes y delicados que requiere más tiempo, debido a que una buena segmentación y etiquetación de la base de datos conduce a buenos resultados. El experto observador que realiza esta tarea debe de tener mucho cuidado al señalar tanto en punto de origen en donde comienza el evento como su punto final. Este cuidado se debe a que, los registros obtenidos representan la base sobre la cual los modelos

tomarán la referencia e indicarán (al hacer uso de las estimaciones de probabilidad), qué tanto se acerca cada registro al modelo de referencia. Así, se ha de proceder a construir una base de datos (segmentada y etiquetada) particular, de cada uno de los volcanes de estudio (Etna, Strómboli, San Cristóbal y Telica). Se crearán ficheros etiquetados (.lab) que contendrán las referencias de inicio y fin de cada evento. Para así, seguidamente, poder construir tanto diccionarios de eventos, como una gramática particular de cada volcán. Entrenando a partir de cero, nuevos modelos de reconocimiento de eventos, con los datos provenientes de los volcanes y posteriormente poder observar los resultados.

Este proceso se ha realizado mediante el uso del programa Transcriber [Barras et. al. 1998] Diversos trabajos sobre reconocimiento han utilizado dicha herramienta en el proceso de etiquetación manual de los datos [Pleva et. al. 2005]; [Ziibert et. al. 2005]; [Barras et. al. 2001]. El proceso de la segmentación y el etiquetado manual de las señales sísmico-volcánicas, puede observarse en la figura 2.11, el resultado que se obtiene es una base de datos propia de cada volcán, almacenadas en ficheros con la extensión “trs”.

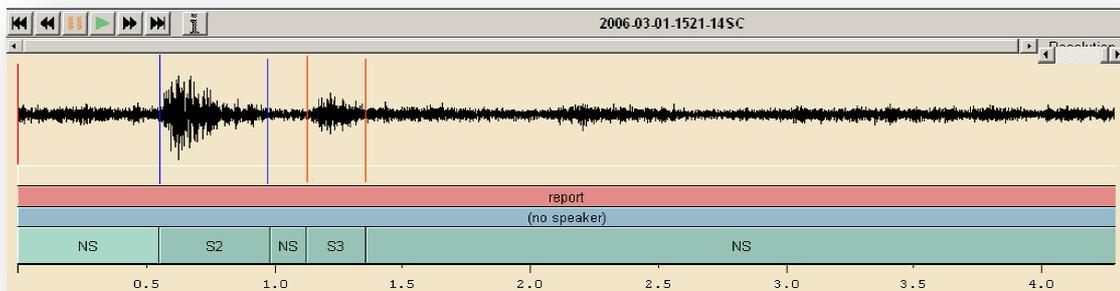


Fig. 2.11 Ejemplo de la segmentación y el etiquetado manual de una señal con la herramienta Transcriber

Además de la herramienta Transcriber, puede realizarse un etiquetado manual mediante la herramienta “HSLab” del conjunto de herramientas HTK. Dicha herramienta crea un fichero de etiquetado simple (.lab) que es el que contiene los tiempos de duración del evento. La figura 2.12 muestra un ejemplo de este tipo de segmentación.

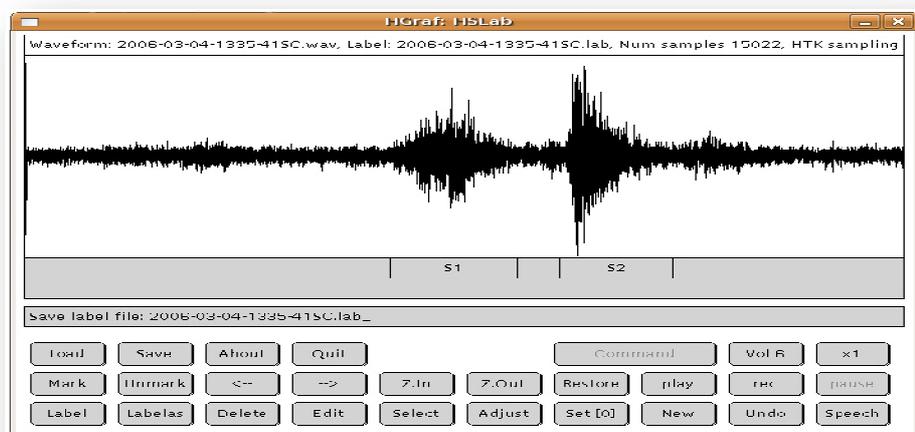


Fig. 2.12 Ejemplo de segmentación de una señal con la herramienta *HSLab* de HTK

Resulta más cómodo y fácil trabajar con la herramienta Transcriber (el proceso de etiquetado es muy largo), por lo que se optó al final, seguir la segmentación con el software Transcriber. Un ejemplo del fichero resultante (.trs) se muestra en la figura 2.13

```

FICHERO: 2006-02-16-2033-38ST.trs

<?xml version="1.0" encoding="ISO-8859-1"?>
<!DOCTYPE Trans SYSTEM "trans-13.dtd">
<Trans scribe="(unknown)" audio filename="2006-02-16-2033-38ST"
version="2" version_date="090105">
<Episode>
<Section type="report" startTime="0" endTime="4.28571428571">
<Turn startTime="0" endTime="4.28571428571">
  <Sync time="0"/>
  NT
  <Sync time="0.986"/>
  T2
  <Sync time="1.243"/>
  NT
  <Sync time="1.76"/>
  T2
  <Sync time="1.979"/>
  NT
  <Sync time="4.162"/>
  T3
</Turn>
</Section>
</Episode>
</Trans>

```

Fig. 2.13 Ejemplo del contenido de un fichero de etiquetado (.lab) mediante la herramienta Transcriber.

2.6.4.- Entrenamiento

El entrenamiento contiene tres elementos que son: la estimación inicial de los modelos, la re-estimación de los modelos y funciones auxiliares para la modificación de los modelos. La figura 2.14 ejemplifica dicha composición.

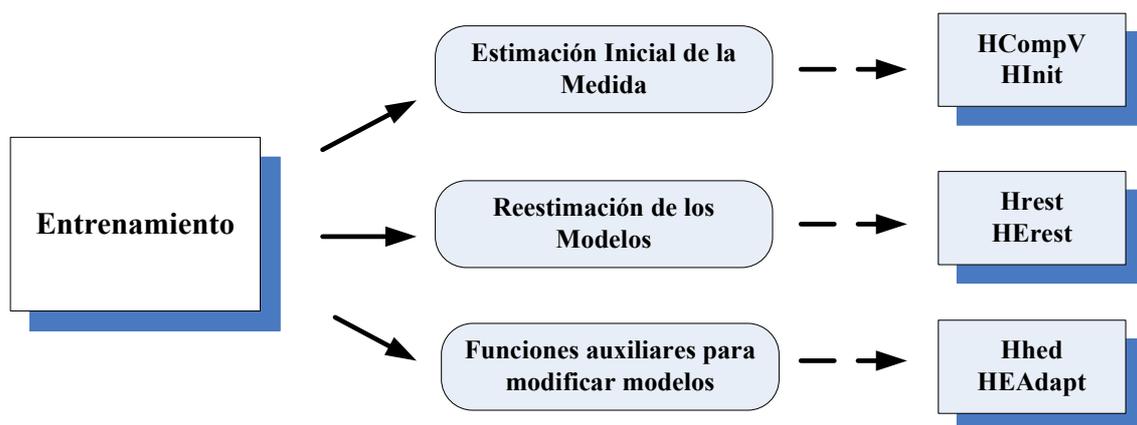


Fig. 2.14 Elementos en un proceso de entrenamiento mediante HTK

El primer paso en el entrenamiento de un HMM es la creación de un modelo prototipo. Los parámetros del modelo no son tan importantes al propósito de definir la

topología del modelo. En el apéndice H se detalla la creación de dicho modelo. La herramienta HCompV contiene un número de opciones específicas para poder configurar el modelo de entrenamiento. Esta herramienta calcula la media y la covarianza del conjunto de datos de entrenamiento

Posteriormente, los modelos iniciales son re-estimados es decir, se comienza el entrenamiento a partir de las segmentaciones hechas utilizando la herramienta integrada Hrest: La herramienta Hrest ejecuta básicamente la re-estimación Baum-Welch de los parámetros de un HMM simple, utilizando un conjunto de secuencias de observación. El principal objetivo de la herramienta es cargar todos los modelos almacenados y re-estimarlos utilizando una lista de datos almacenada en el fichero de configuración, así los nuevos modelos son almacenados en el siguiente directorio. Además existen funciones auxiliares como Hhed y Headapt, que sirven para adaptar o modificar los modelos.

Todo el proceso que involucra el entrenamiento de los modelos se observa en la figura 2.15.

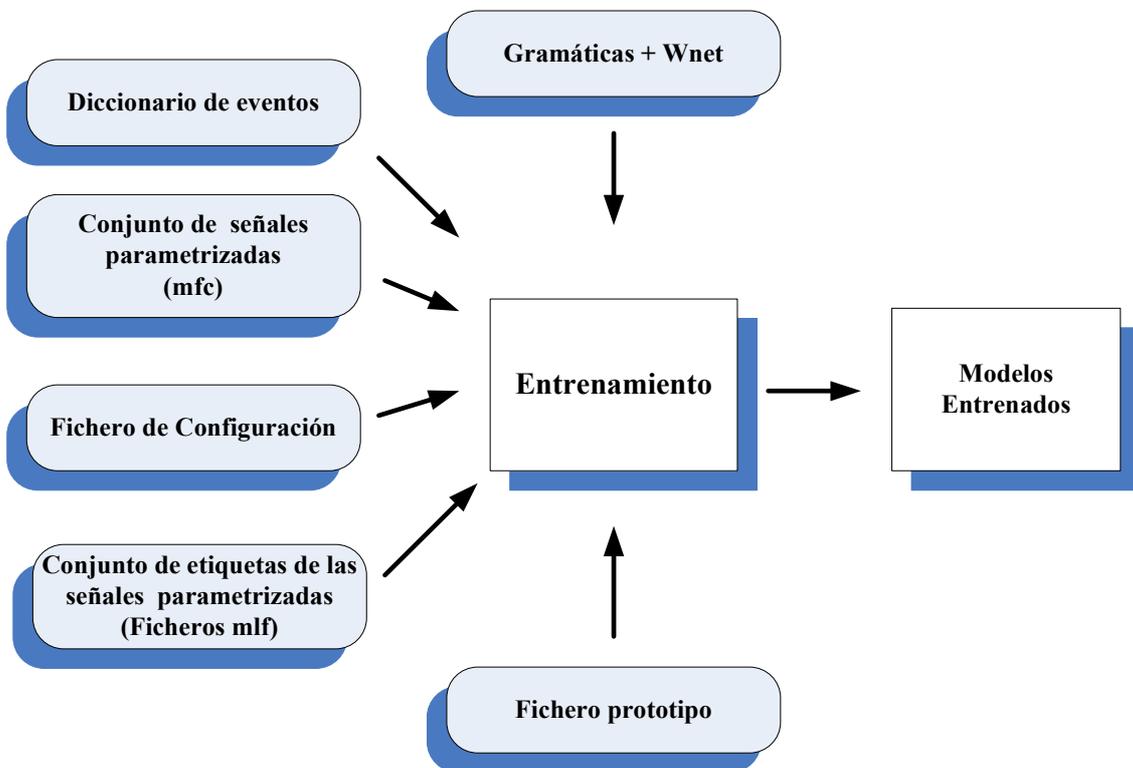


Fig. 2.15 Elementos del proceso de entrenamiento de los HMM con HTK

Como se observa en la figura, para el bloque de entrenamiento hace falta un conjunto de elementos de entrada como son: El fichero prototipo en donde se define como va a ser el modelo, las gramáticas y diccionarios de eventos, el conjunto de señales parametrizadas y el conjunto de etiquetas de dichas señales parametrizadas, además del fichero de configuración. La salida obtiene el conjunto de los modelos entrenados. En el apéndice H se detalla el script que realiza todo el proceso anteriormente mencionado.

2.6.5.- Reconocimiento

Una vez que los HMM han sido entrenados, el reconocimiento se realiza mediante la herramienta de HTK, Hvite. En este proceso participan elementos similares al del reconocimiento, solo que en vez de un fichero prototipo, aquí intervienen los modelos previamente entrenados. En la figura 2.16 se observa un diagrama de los elementos que componen esta fase.

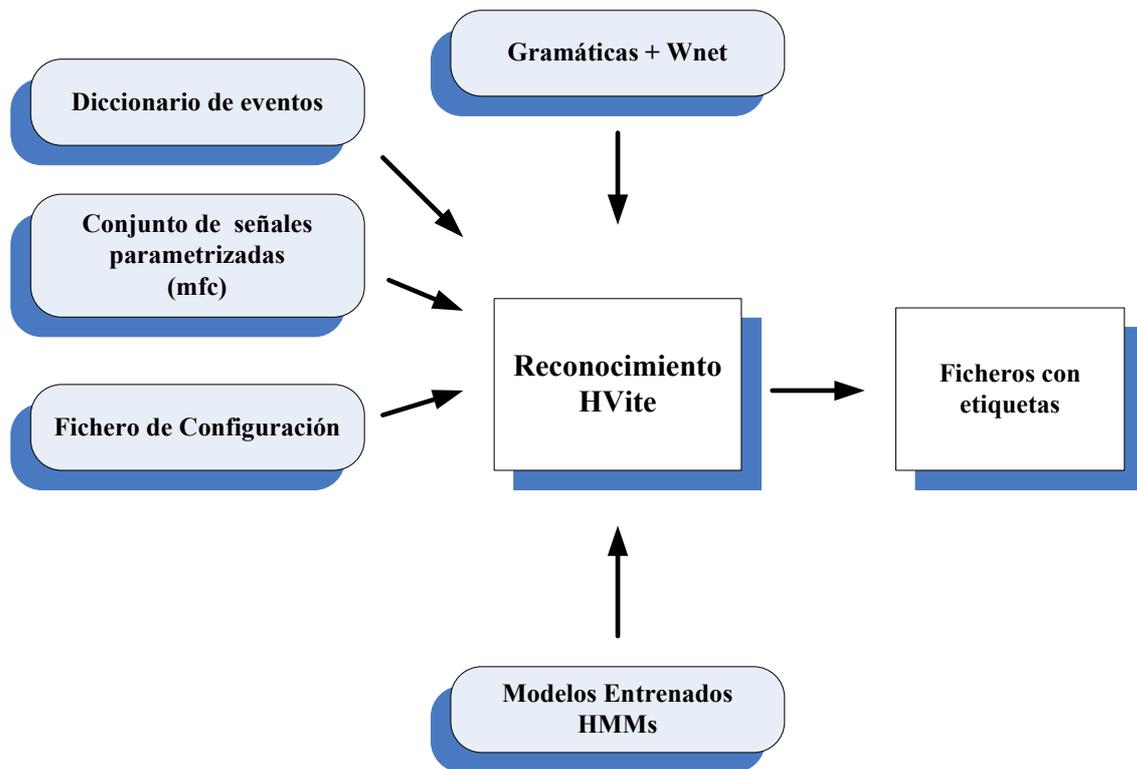


Fig. 2.16 Elementos del proceso de reconocimiento de los HMM con HTK

Mediante la herramienta Hvite se realiza el reconocimiento de señales utilizando el algoritmo de Viterbi. Hvite proporciona el camino de máxima probabilidad para una secuencia óptima de estados, utilizando para ello los modelos almacenados en los “mlf” iniciales, hacia nuevos niveles de transcripción “mlf”. Aquí el reconocedor hace coincidir cada fichero de señales contra la red de modelos HMMs creados y proporciona una salida de transcripción para cada uno de ellos. En el apéndice H se detalla el contenido del script que realiza en reconocimiento mediante Hvite.

2.6.6.- Análisis (Evaluación de los resultados)

Por último, se encuentra la herramienta de desempeño de análisis de HTK que es “Hresults”, la cual lee un conjunto de ficheros de etiquetas, los cuales son generalmente producto de una herramienta de reconocimiento como Hvite y compara su correspondencia de acuerdo a los ficheros de transcripción. Para el análisis del reconocimiento de la salida de señales, la comparación se basa en la programación dinámica (DP). Para el análisis de la visualización de palabras de salida, la comparación utiliza la métrica estándar US NIST. Cuando se utiliza la DP para el cálculo de la

precisión de sentencias, la salida básica es una estadística de reconocimiento que tiene el siguiente formato.

```
----- Overall Results -----
SENT: %Correct=30.25 [H=85, S=196, N=281]
WORD: %Corr=91.24, Acc=71.68 [H=989, D=68, S=27, I=212, N=1084]
```

En donde, la primera línea provee el nivel de precisión de la sentencia basada en el número total de etiquetas que son idénticas a los ficheros de transcripción. La segunda línea es la precisión de palabras basadas en *DP*, entre las coincidencias de los ficheros de etiquetas y los de transcripciones. En la segunda línea la **H** se refiere al número correcto de etiquetas, la **D** es el número de errores de tachaduras o borrados, la **S** es el número de errores de sustituciones, la **I** es el número de errores de inserciones y la **N** es el número total de etiquetas en los ficheros de transcripciones. El porcentaje de etiquetas correctas *%Corr* reconocidas está dado por la ecuación 2.8:

$$\% \text{ Corr} = \frac{H}{N} \times 100\% \quad (2.8)$$

Y el porcentaje de precisión *% Acc* está dado por la ecuación 2.9

$$\% \text{ Acc} = \frac{H - I}{N} \times 100\% \quad (2.9)$$

En donde: $H = N - D - S$

La tabla 2.1 muestra como la herramienta Hresults genera una matriz de confusión.

EJEMPLO MATRIZ CONFUSIÓN SAN CRISTÓBAL - NICARAGUA

```
----- Confusion Matrix -----
      N   S
      S   1 Del [ %c / %e]
NS    610 14   49 [97.8/1.3]
S1    13  379  19 [96.7/1.2]
Ins   145  67
=====
```

Tabla 2.1 Ejemplo de la Matriz de confusión, generada por la herramienta HResults.

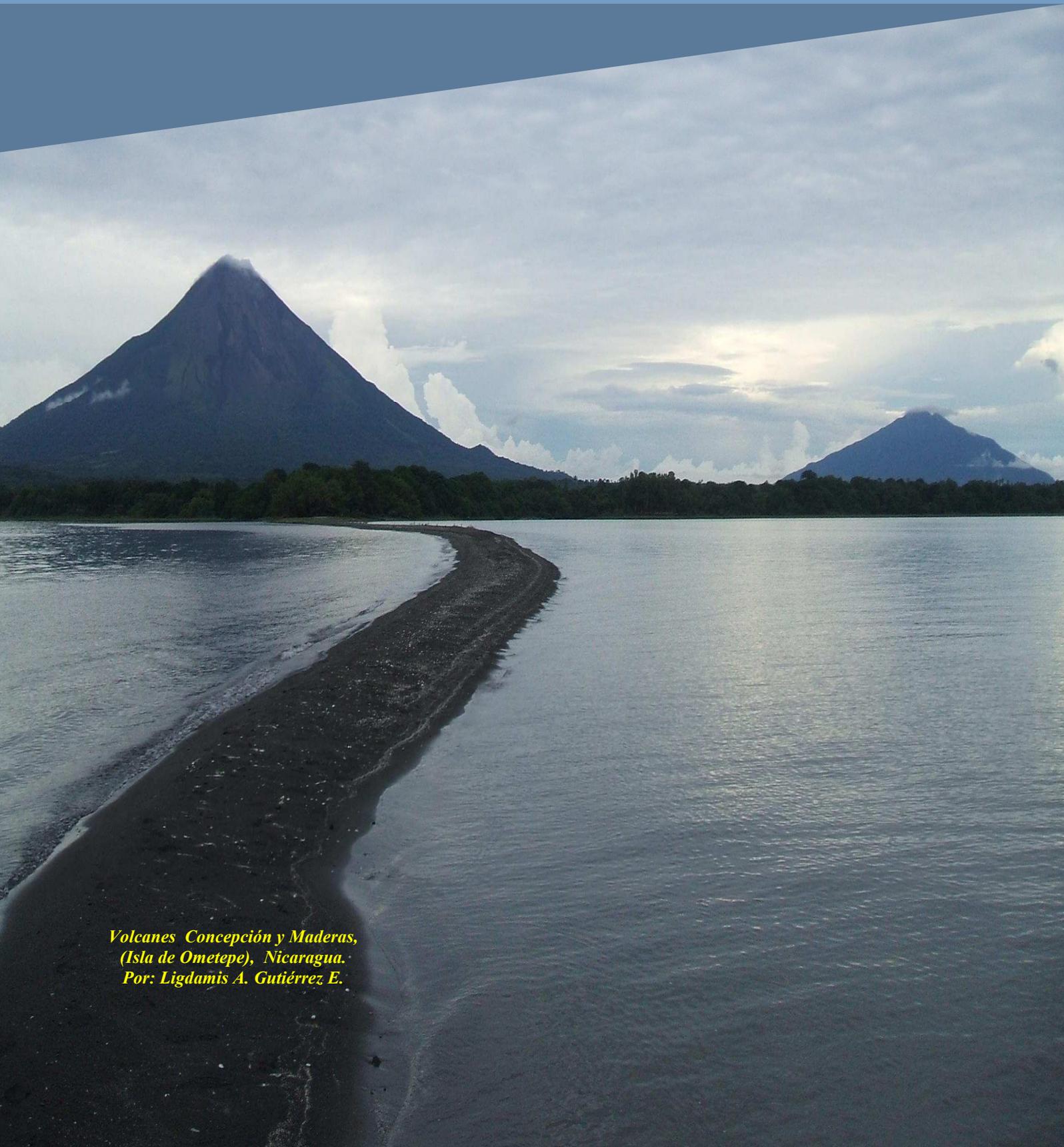
En la tabla anterior, *%c* es el porcentaje correcto en la fila; esto es, cuantas veces una instancia de un evento fue correctamente etiquetado. En cuando a *%e* representa el porcentaje incorrecto de etiquetas en la fila del número total de etiquetas en el conjunto. Al leer a través de las filas, *%c* indica el número de instancias correctas divididas por el número total de instancias en la fila y *%e* indica el número de instancias incorrectas divididas por el número de instancias (*N*). Estos resultados se generan a través de un fichero en cada proceso de la creación de los modelos de entrenamiento HMMs en los directorios *hmm0.0*, quedando en el último directorio las estadísticas finales de reconocimiento al terminar el proceso de entrenamiento y reconocimiento. De esta forma, Hresults genera un fichero paralelo al fichero MLF pero con extensión “.res” para indicar dichos resultados.





SEGUNDA PARTE

CAMPAÑAS, EMPLAZAMIENTOS Y DATOS



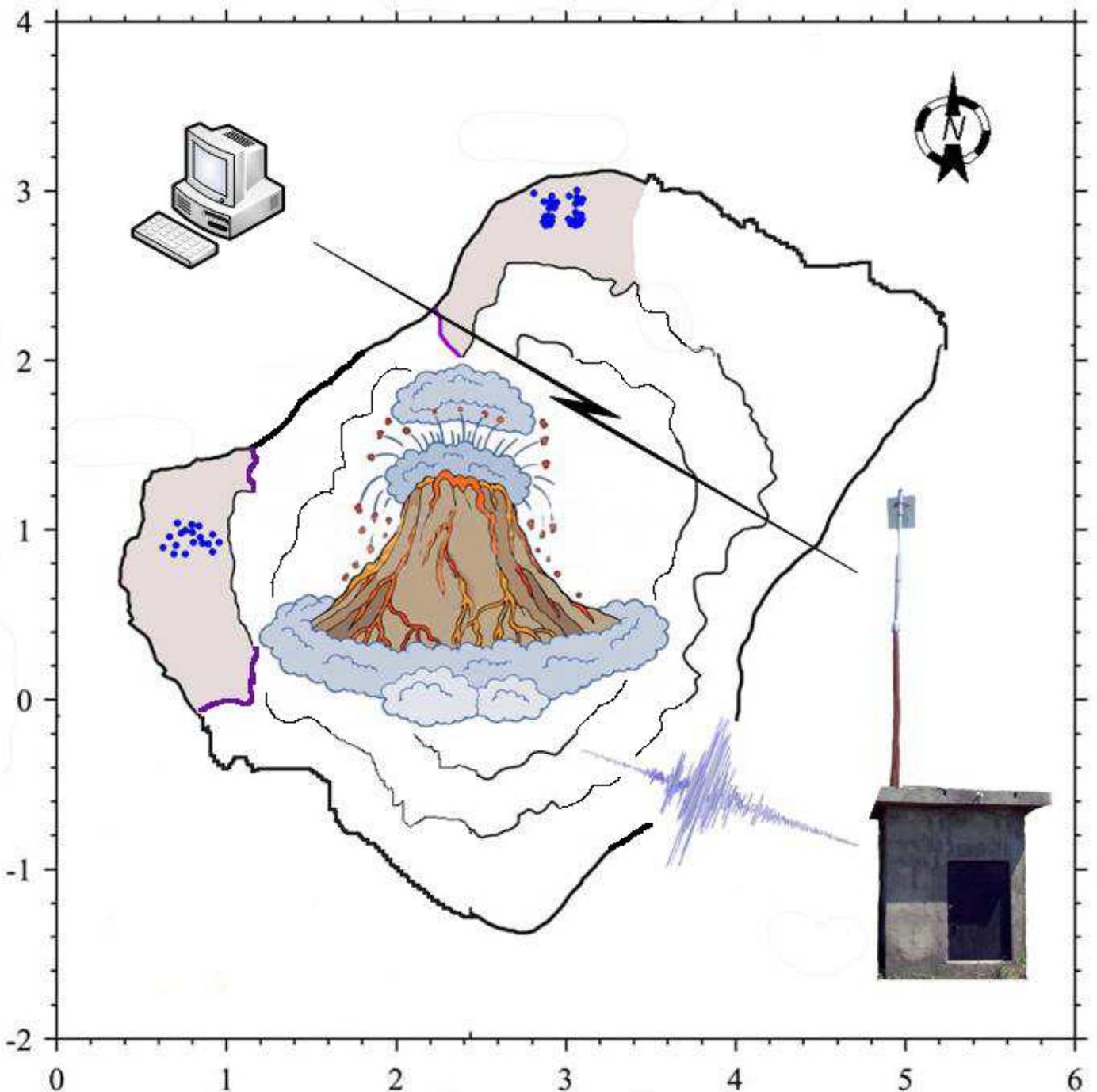
*Volcanes Concepción y Maderas,
(Isla de Ometepe), Nicaragua.
Por: Ligdamis A. Gutiérrez E.*





CAPITULO III

ZONAS DE ESTUDIO, INSTRUMENTOS Y CAMPAÑAS





3.1.- INTRODUCCIÓN

Los dos primeros capítulos del presente trabajo han descrito en términos generales las señales, los volcanes y los sistemas de reconocimiento, también se ha determinado el método a emplear en el presente estudio. En este capítulo se describirá el marco volcánico que comprenden los cuatro volcanes de estudio, asimismo se detallarán las campañas y los instrumentos que determinaron la recopilación de los datos, los cuales se utilizarán para el análisis de las señales sísmico-volcánicas. En el apéndice F se describe detalladamente el proceso de transformación de dichos datos. Una vez realizado esto, podrá concluirse al final del capítulo la cantidad de datos propuesta a analizar. Todo este proceso es previo a la realización del análisis de los datos, al establecimiento de los tipos de eventos presentes en los registros y a la construcción de las bases de datos, procedimientos que se describirán en el siguiente capítulo.

En la elección de los volcanes de estudio, se han tomado en cuenta diversos factores ya que los volcanes se presentan en diferentes áreas geográficas, presentan diferentes ambientes tectónicos, diferentes composiciones magmáticas, diferentes dinanismos eruptivos entre otros y por lo tanto es de esperar diferentes tipos de señales. En este estudio se eligieron cuatro volcanes activos de tipo estratovolcanes, dos italianos (El Etna y el Strómboli) y dos de la cadena cuaternaria del Pacífico de Nicaragua (San Cristóbal y Telica). Debido a que a diferencia de los volcanes de Italia, los procesos eruptivos y los eventos que se producen en los volcanes de Nicaragua no han sido tan extensamente estudiados, ha sido incorporado en los apéndices, un análisis detallado sobre el ámbito sísmico-volcánico de Centro América y Nicaragua (Ver Apéndice B). El análisis de los eventos producidos en volcanes con características geológicas similares, pero ubicados en diferentes áreas geográficas y con diferentes procesos o mecanismos, creemos que va a resultar de mucha utilidad en futuros estudios para poder determinar las similitudes o diferencias entre dichos procesos.

Las campañas de recopilación de la información en Italia han sido establecidas por labores de campo efectuadas por un equipo de investigación ítalo-hispánico durante dos semanas, que en el caso del Etna fueron realizadas en Septiembre de 1999 y en el Strómboli en Septiembre de 1997. Para Nicaragua, el proceso de recopilación de datos abarcó un período desde el 16 de Febrero al 10 de Marzo del 2006 realizado con la ayuda proporcionada con los medios del **INETER** (Instituto Nicaragüense de Estudios Territoriales), entidad gubernamental encargada entre otras funciones del monitoreo, vigilancia y estudio de la cadena volcánica. Como se puede observar, las campañas han sido determinadas en tres diferentes momentos a lo largo de la reciente historia eruptiva de los volcanes de estudio. Invariablemente todo esto tiene que comenzar con un largo proceso, que incluye el poder analizar cada una de las estaciones que han sido utilizadas, para poder definir cuál de ellas es la más idónea en cuanto a la cantidad y calidad de los registros a tratar. Después de un detallado estudio visual por parte de la experiencia otorgada durante las diferentes campañas se llegó a definir dichas estaciones, tanto para Italia como para Nicaragua. Como se ha visto, no todos los observatorios alrededor del mundo manejan la misma terminología y poseen un estándar general sobre los diferentes tipos de señales sísmico-volcánicas que se producen en los volcanes que estudian. En las campañas de Italia, se ha trabajado con datos y eventos ya conocidos y analizados en muchas ocasiones por diferentes autores [Chouet et. al. 1999]; [Cannata et. al. 2009]; [Branca y Del Carlo 2005]; [Mulargia et. al. 1987]; [Ntepe y

Dorel 1990], puesto que se trata de dos de los volcanes más estudiados no solo de Italia sino a nivel global. De la amplia gama de eventos, se han seleccionado aquellos que poseen una mayor cantidad en los registros analizados y que por consiguiente podrán producir resultados mucho más fiables. En el caso de Nicaragua (a diferencia de las de Italia), aun no se ha aplicado a la totalidad de los eventos ninguna de las escuelas de dicha clasificación sísmica. Se maneja por decirlo así la antigua clasificación de Minakami [Minakami 1960]; [Minakami 1969], aunque define algunos eventos de acuerdo con una terminología propia de los sismólogos locales, como es el caso de los llamados eventos dobles. Invariablemente, la actividad sísmica presente en los volcanes activos de la cadena volcánica del pacífico de Nicaragua es constante, existe mucha variedad y cantidad de datos. Por lo tanto, esto ha llevado a que uno de los pasos del presente trabajo sea el poder determinar qué tipos de eventos se producen en cada volcán, para poder ser clasificados y de esta forma, realizar las pruebas de entrenamiento de los modelos, así como la clasificación de los eventos sísmico-volcánicos que se encuentren. Por otra parte, los procesos sísmicos varían en cada volcán, y durante el proceso de recopilación de la información, puede ser que exista o no una crisis volcánica, los cuatro volcanes presentan una actividad constante, por lo que se espera encontrar diversos tipos de eventos presentes en los registros (Ver apéndice C).

Después de un largo proceso de análisis de los registros generados por la recopilación de los datos durante las campañas de 1999 y 1997 en Italia ha sido seleccionado en el Etna un evento característico, al que Sacarroti denominó tremor de tipo burbuja (*Tremor Burst*) [Saccorotti et. al. 2004], además del tremor de fondo y el ruido presente. En el caso del Strómboli, este volcán posee una característica peculiar que se manifiesta en presentar una explosión bastante típica, denominada “*explosión estromboliana*”, y que ocurre constantemente durante cada determinado intervalo de tiempo [Chouet et. al 1997]; [Chouet et. al 1999]; [Chouet et. al 2003]; [Lautze y Houghton 2005], seleccionando así las explosiones, como tipo de evento, más el ruido de fondo presente en el volcán. En las campañas de Nicaragua, como se ha mencionado, no existe una previa clasificación de los eventos. Por esta razón, las primeras clasificaciones visuales de los registros (con base a la experiencia), en una pre-inspección visual se catalogaron más de 12 diferentes tipos de eventos. Los resultados negativos en las primeras pruebas, que arrojaron porcentajes por debajo del 40% de precisión en los reconocimientos, así como la matriz de confusión del sistema en que los eventos tendían mucho a ser catalogados con otro tipo diferente al que había sido etiquetado, indujeron a que algunos de los eventos iniciales en realidad correspondían a uno solo. Se ha tenido que hacer uso de las técnicas de análisis espectrales (que se detallan en este capítulo), para poder detallar y definir aun más dichos eventos y poder agrupar los eventos similares. Adicionalmente, se pueden hacer uso de técnicas auxiliares que podrían ser útiles en futuros trabajos. Un análisis de dichas técnicas se trata en los apéndices (Ver apéndices D y E). De esta forma, al final de otro largo proceso se han llegado a definir, cuatro tipos básicos más el ruido o tremor de fondo, como los tipos de eventos sísmico-volcánicos presentes en el volcán Telica y en la misma línea, se definieron tres tipos de eventos básicos, más el ruido de fondo en el volcán San Cristóbal. Todo este desarrollo implica un largo proceso, que ha comprendido varios meses, en los que se analizaron miles de registros para identificar los eventos entre todos los registros. Este detallado proceso es necesario para poder llegar a conformar la base de datos de los eventos sísmico-volcánicos.

Los métodos para la obtención de datos han sido diferentes en ambas campañas. En Italia se utilizaron técnicas de array sísmico para la toma de los registros, en cambio, en las campañas de Nicaragua se utilizó la red sísmica instalada por el INETER. Análogamente, cada campaña utilizó diversos formatos de almacenamiento, por lo que los parámetros utilizados también son diversos. En el Etna y Strómboli se utilizaron datos de tipo SAD y DTS, en Nicaragua se trabajaron con datos de tipo binario R. Lo que implica que para el desarrollo del trabajo se tendrá que realizar un proceso de transformación de los datos a un formato único. Dicho proceso de transformación se detalla en los apéndices (Ver apéndice F). Otro aspecto es que en todas las campañas, los datos han sido registrados con sensores similares, utilizando la componente vertical. Sobre las características y parámetros de los equipos, los trabajos de [Havskov y Alguacil 2004] y [Ortiz 2000], presentan una extensa documentación de los diversos equipos utilizados en el estudio de la sismología, incluidos los utilizados en este trabajo. En el caso del Strómboli [La Roca et. al. 2004], detalla las características utilizadas por los equipos de la campaña. A continuación se presenta el marco volcánico de cada uno de los ámbitos estudiados.

3.2 MARCO VOLCÁNICO

3.2.1. El volcán Etna

El volcán Etna que se observa en la figura 3.1, se ubica en la provincia de Catania, en la isla de Sicilia, Italia. Entre sus características están el que posee una elevación de 3,330 metros de altitud (10,991 pies), su localización geográfica se encuentra entre las coordenadas de 37,73° N y 15,00° E. La máxima erupción histórica del Etna es un VEI de 6. Su origen se encuentra en el Pleistoceno, siendo un volcán cuaternario compuesto. A nivel global, el Etna es uno de los volcanes más activos, y también es de los más grandes, [Branca et. al 2008], pues a su altura, hay que sumarle la base que mide aproximadamente 36 x 24 millas (60 x 40 Km), con lo que el perímetro calculado de su base es de 250 Km., por lo que se extiende en una superficie de 1,500 Km². Diversos autores sitúan al Etna al pie de la pared de un sistema cuaternario tardío, orientado al este de la corteza de un sistema de escala de falla normal [Ellis y King 1991]; [Monaco et. al 1997]; [Him et. al. 1997]; [Bianca et. al. 1999], cuya parcial reactivación es la escarpa de magma, una discontinuidad del Mesozoico que separa la corteza continental del Bloque Pelagiano [Burollet et. al. 1978] de la corteza oceánica del mar Jónico. [Makris et. al. 1986]. Estos factores hacen suponer que la causa de la deformación que controla la evolución del flujo magmático en el Etna, ha sido la tectónica extensional. De acuerdo a Gilloot y otros, el volumen que presenta el volcán en la actualidad, se debe a la continua acumulación de lavas y piroclastos producto de explosiones que se han dado desde hace más de 18 mil años. [Gilloot et. al. 1994].



Fig. 3.1 Volcán Etna en Italia. Fuente: NASA Astronaut photograph ISS006-E-31042

En cuanto a la historia eruptiva del Etna, éste muestra una continua actividad sobre todo en los cráteres que están situados en su cumbre. Además de que todos los años se muestran frecuentes actividades en sus flancos (Fig. 3.2). Mucha de la actividad en los flancos se sitúa en ciertas áreas, específicamente en grietas situadas al noreste y hacia el sur. Además de dos enjambres de fisuras que se encuentran desde la cumbre hacia el este-noroeste y el sureste [McGuire et. al. 1997].



Fig. 3.2 Mapa de las actividades eruptivas en el volcán Etna en los últimos años. Los círculos amarillos localizan los tres cráteres con mayor actividad. (Fuente Google Earth y McGuire et. al. 1997)

En la historia eruptiva del Etna, después de un largo período de actividad comprendido entre 1991-1993 [Calvari et. al. 1994]; [Tanguy et. al. 1996]; [Stevens et. al. 1997], la actividad eruptiva en el verano de 1995 se manifiesta sobre todo en los alrededores de los cráteres de la cumbre del volcán. En los siguientes 6 años, los eventos eruptivos se concentran en dichos cráteres [Neri y Tomarchio 2000]; [Harris y Neri 2002]; [Calvari et. al. 2002]; [Aloisi et. al. 2002]. En el Etna, la relación entre los productos tectónicos y el volcanismo, ha sido investigada por muchos autores a través de datos que involucran procesos geoquímicos, geofísicos y estructurales [Cristofolini et. al. 1991]. En la actualidad, las erupciones más recientes se caracterizan por la construcción de cráteres en la cima del volcán. Tales erupciones efusivas se producen en su mayoría tanto en la cumbre de los cráteres, como en las rejillas de ventilación laterales, que normalmente se encuentra a lo largo de los sistemas de fracturas relacionadas con la tectónica regional. [McGuire y Pullen 1989]; [Monaco et. al. 1997]. En dos erupciones ocurridas en 2001 y 2002-2003, estudios sobre la sismicidad del Etna han podido recopilar datos sobre la relación que existe entre los procesos eruptivos y los terremotos locales precursores de dichas erupciones. En este sentido, los trabajos de Falsaperla en 2002 relacionaron los procesos eruptivos con los eventos de largo período registrados durante el proceso eruptivo de 1991-1993 [Falsaperla et. al. 2002]. Los dos períodos de erupciones de 2001 a 2003 revelan el contraste entre los mecanismos predictivos ocurridos en un volcán, ya que mientras que en la erupción de 2001 se presentaron 4 días de intensa actividad sísmica, la erupción ocurrida en 2002-2003 se presentó un corto período de 2 horas de sismicidad predictiva [Monaco et. al. 2005]

[Neira et. al. 2005]. Todo lo anterior hace indicar la extensa actividad que se produce en el Etna y por consiguiente la gran cantidad de eventos sísmicos que se generan, los cuales de una u otra forma se encuentran relacionados ya sea previo o posterior a un determinado episodio eruptivo.

3.2.2.- El volcán Strómboli.

El volcán Strómboli que se observa en la figura 3.3, está situado en la isla del mismo nombre, y que pertenece al conjunto de siete islas que conforman el archipiélago Aeoliano, situado en el mar Tyreno al sur de Italia. Las coordenadas geográficas del volcán están en: 38° 47' N y 15° 13' E. La máxima erupción histórica del Strómboli presenta un VEI de 1. A nivel mundial, Strómboli es uno de los volcanes más activos y se caracteriza por la actividad explosiva persistente, denominada, tipo; “estromboliano”, misma que se ha puesto como un referente a las erupciones de los volcanes alrededor del mundo, que presentan similares características. El proceso eruptivo del Strómboli es muy constante y sólo ha sido interrumpido por episodios ocasionales de mayor actividad, acompañados por flujos de lava, ejemplo de ello han sido los sucesos producidos en 1975 [Capaldi et. al. 1978], 1985 [De Fino et. al. 1988], 2003 [Calvari et. al. 2005] y 2007 [Casagli et. al. 2009]; [Landi et. al. 2009]. Esta constante actividad, hace que en la actualidad, la morfología del volcán sea el resultado de varios estados de actividad, que se han producido en períodos a lo largo de 100.000 años.



Fig. 3.3 Mapa de la isla y volcán Strómboli en Italia. (Fuente Google Earth)

El edificio volcánico se levanta sobre los 3,000 m desde el piso oceánico llegando a los 927 m sobre el nivel del mar. De acuerdo a Burton y asociados, la actividad volcánica que presenta en la actualidad el Strómboli, se caracteriza por pequeños e intermitentes chorros de gases acompañados de fragmentos de lava, o una constante desgasificación pasiva, acompañada con pequeñas explosiones de corto período de 5 a 15 seg [Burton et. al. 2007]. Lo típico de estos sucesos es que acontecen

a intervalos de 3 a 10 eventos por hora. La mayoría de los eventos suelen ocurrir sobre los 750 m sobre el nivel del mar, sobre la cima de una zona de graben en el lado nordeste del volcán, denominada “Sciara del Fuoco”, que se caracteriza por una gran depresión, cuya formación se atribuye a un desprendimiento producido hace más de 5.000 años. Los más actuales estudios batimétricos indican que la Sciara del Fuoco se extiende por debajo del nivel del mar, hasta una profundidad de unos 1,700 m. La actividad sísmica en el Strómboli está determinada por señales fusiformes (o en forma de huso) acompañadas con explosiones en la cima del volcán superpuestas por un tembor de fondo. Dichas explosiones son las que se abordan en el presente estudio. El tembor típico del Strómboli se encuentra en un rango de frecuencia de 1 a 5 Hz. con terremotos explosivos con una amplitud sobre un intervalo de frecuencia de 0.1 a 10 Hz. [La Rocca et. al. 2004]; [Neuberg et. al. 1994]; [Langer y Falsaperla 1996]; [Chouet et. al. 1999].

3.2.3.- El volcán San Cristóbal

El volcán San Cristóbal que se observa en la figura 3.4, es un estrato-volcán de forma casi simétrica, que se eleva en forma de cono. Es el volcán más alto del país con 1,745 m sobre el nivel del mar. Su cráter mide 500 x 600 metros de diámetro. La localización geográfica lo sitúan a 12.702° N de latitud y 87.004° O de longitud, a unos 15 Km. al Noreste de la ciudad de Chinandega, en la Cordillera de los Murrubios. El índice de peligrosidad del volcán es de 10 [Yokohama et. al. 1984] y su máxima erupción histórica es un VEI de 1 [Newhall y Self 1982].

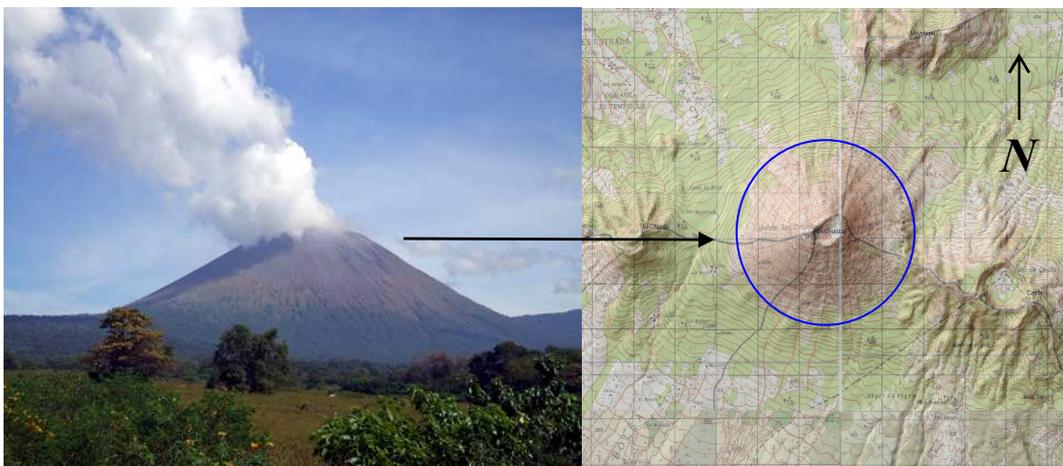


Fig. 3.4 Volcán San Cristóbal (izq.) y mapa geográfico de su ubicación (der.). (Fuente Ligdamis Gutiérrez e INETER)

El San Cristóbal es un volcán cuaternario y es el principal que da nombre a un grupo que forma junto a otros cuatro edificios volcánicos más. Dicho grupo es uno de los cinco grandes complejos volcánicos que forman la cadena volcánica de Nicaragua (Ver apéndice B). Así, el grupo del San Cristóbal es el más occidental y comprende los conos o edificios volcánicos de: El Chonco (1,105 metros), San Cristóbal (1745 metros y 50 Km³), Moyotepe, Casita (1,405 metros) y La Pelona. El San Cristóbal es el volcán más joven del complejo, porque los otros, Casita, Chonco y Moyotepe son del pleistoceno, con una edad de más de un millón de años. Cada estructura del complejo posee un cráter principal y algunos cráteres secundarios o parásitos a los costados. Los estudios geológicos describen a las rocas del complejo del San Cristóbal como basaltos olivínicos, basaltos normales y andesitas [Carr y Rose 198]; [Martínez y Viramente

1973]. La geoquímica del complejo define que los cinco conos están compuestos por basaltos toleíticos a calcoalcalinos y andesita afines con lavas de arcos de islas. También se encuentran grandes domos dacíticos y pómez asociadas otros dos volcanes del complejo [Carr y Rose 198]; [Hazlett 198)]. Respecto a los principales núcleos de población cercanos del departamento de Chinandega como son la cabecera municipal Chinandega (126,000 hab.), la ciudad de Chichigalpa (45,000 hab.), por lo que la peligrosidad del San Cristóbal es evidente [Mooser et al, 1958]; [Lilljequist et al. 1987]. Además, existen en los alrededores una serie de poblados con un promedio de 2,000 habitantes. La población total bajo riesgo supera los 200,000 habitantes. Estos datos son de acuerdo a los aportados por el VIII censo de población del 2005 publicado en Noviembre del 2006 por el INEC (Instituto Nacional de Estadísticas y Censos de Nicaragua). Las amenazas asociadas y el área afectada por ellas se puede observar en los mapas publicados por el INETER que se observan en las figuras 3.5 y 3.6.

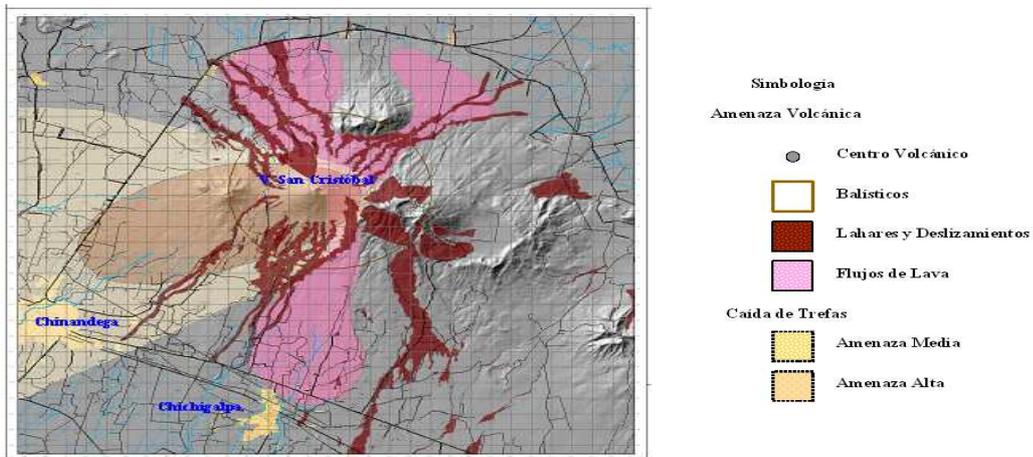


Fig. 3.5 Mapa de amenazas asociadas al volcán San Cristóbal (Fuente: INETER Mapa de Amenazas volcánicas 1995. Escala 1'400.400)

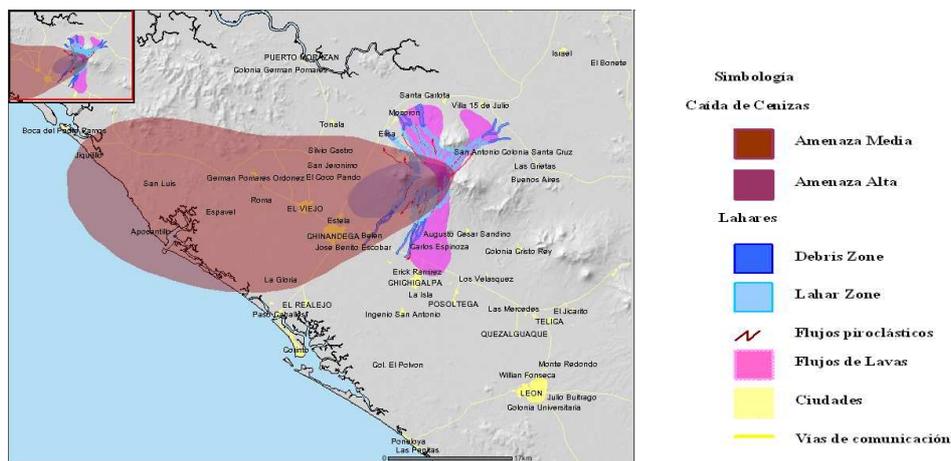


Fig. 3.6 Mapa de poblaciones afectadas por los productos del volcán San Cristóbal (Fuente INETER)

De acuerdo a los estudios elaborados por el INETER, el San Cristóbal a lo largo de su historia eruptiva (ver apéndice C), ha manifestado diferentes tipos de productos, que van desde plinianas a sub-plinianas y lávicas con procesos estrombolianos (expulsión de balísticos). Al San Cristóbal se le considera uno de los volcanes que mayor amenaza representa dentro de toda la cadena volcánica de Nicaragua, por la

probabilidad de producir flujos piroclásticos. Además, los flujos de lodo y la caída de cenizas también forman parte de la historia eruptiva reciente de este volcán y que han afectado principalmente a las ciudades de Chinandega y el Viejo. Otros riesgos asociados a las erupciones que presenta son la lluvia ácida, además de producción de enjambres sísmicos [INETER 1982]; [INETER 2004]. En la tabla 3.1 se muestra la caracterización geológica en el volcán San Cristóbal, teniendo en cuenta los datos recopilados por el INETER.

Tipo	Estado	Productos	Distancia máxima	Composición química
Estrato cono compuesto	Activo	Flujos de lavas	12 Km.	 Basalto-andesita
		Balísticos (Bombas)	5 Km.	
		Flujos piroclásticos	4 Km.	
		Caída de cenizas	1 a 45 Km. al oeste	

Tabla 3.1: Caracterización geológica del volcán San Cristóbal.
Fuente [INETER 2001]; [SINAPRED 2005]

3.2.4.- El volcán Telica

El volcán Telica que se observa en la figura 2.8, es un estrato-volcán basáltico que posee una altura de de 1,061 mts. La localización geográfica lo sitúa a 12.603° N de longitud y 86.845° O de longitud, a 30 Km. de la ciudad de León en la Cordillera de los Marribios. El índice de peligrosidad del volcán es de 10 [Yokohama et al. 1984], siendo su máxima erupción histórica un IEV de 3, mucho mayor que en el San Cristóbal [Newhall y Self 1982]. Debido a las erupciones y a la producción de gases, el Telica no posee vegetación en sus partes altas, y su cono es muy inclinado, cortado por un cráter de 700 metros de diámetro y 120 metros de profundidad. Siendo este cráter por donde se producen las erupciones más recientes (Ver figura 3.7 izquierda).

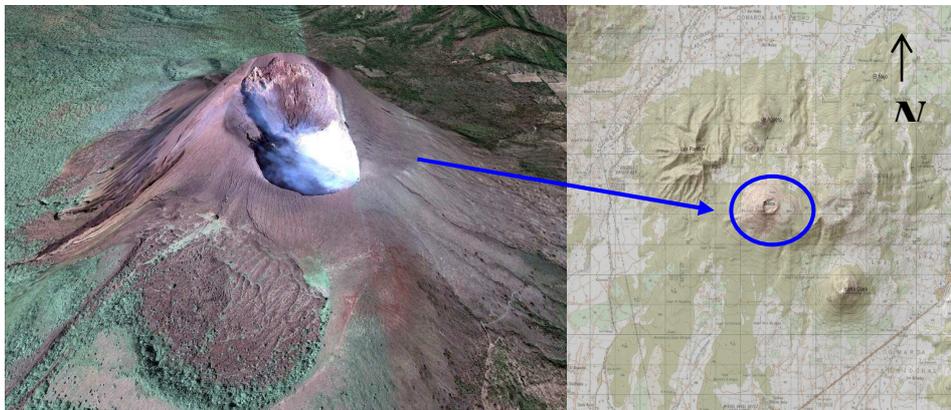


Fig. 3.7. Volcán Telica (izquierda), y mapa geográfico del volcán Telica (derecha).
(Fuente Google Earth e INETER)

De manera similar al San Cristóbal, el Telica forma parte y da nombre a uno de los cinco grupos o complejos volcánicos de la cadena volcánica cuaternaria del Pacífico de Nicaragua. El grupo Telica está compuesto por cuatro edificios más que son: los volcanes de los Portillos, Cerro Agüero, el Listón, y Santa Clara o San Jacinto. Además este grupo posee un campo termal situado a los pies del Santa Clara, llamado “las fumarolas del volcán San Jacinto”. El INETER, ha elaborado un análisis de las amenazas y productos volcánicos del Telica que se observan en la Tabla 3.2. Dicho análisis está basado en información existente en el INETER y en los métodos de evaluación utilizados por esa institución [INETER 1982]; [INETER 2004].

Tipo	Estado	Productos	Distancia máxima	Composición química
Estrato cono volcánico	Activo	Flujos de lavas	10 Km.	
		Balísticos	7 Km.	
		Caída de cenizas	45 Km.	
		Gases	Momentos de Crisis	
				Basalto-andesita

Tabla 3.2: Caracterización geológica del volcán Telica.
Fuente [INETER 2001]; [SINAPRED 2005]

El Telica es aún más explosivo y activo que el San Cristóbal, una descripción de la historia eruptiva y de los procesos volcánicos producidos por los cuatro volcanes se encuentra detallada en los apéndices (Ver apéndices B y C)

3.3.- ESTACIONES SISMICAS E INSTRUMENTOS DE MEDICIÓN EN LOS VOLCANES DE ESTUDIO

3.3.1.- Tipos de estaciones sísmicas

Las señales sísmicas son amplificadas y muestreadas con un convertidor analógico digital *CAD* de 16 bits (ver sección 1.6) y filtros anti-aliasing antes de ser almacenadas. Los datos se recopilan en la estación de campo en donde se ubica el equipo. Atendiendo al diseño de su arquitectura las estaciones sísmicas pueden ser de dos tipos, *tipo A* en la que los instrumentos son enterrados en el suelo sin ninguna estructura y *tipo B* en la que los instrumentos se encuentran dentro de una estructura. La Figura 3.8 presenta el esquema general del diseño de la construcción y los elementos principales de una estación sísmica tipo B, que básicamente es una caseta de concreto o mampostería de 2 m², y que se conecta a la estación central donde se procesan los datos. Para el caso de las estaciones de tipo A, estas no requieren una infraestructura permanente.

Los equipos de las estaciones de campo de tipo B se comunican a las estaciones centrales situadas en los observatorios, en donde se almacena y procesa la información.

Un ejemplo de ello se muestra en la figura 3.9, donde se observan los principales componentes de una estación sísmica de campo tipo B y la conexión a la estación central para la recopilación de datos. Este sistema es similar al empleado en la recopilación de los datos de las campañas de Nicaragua.

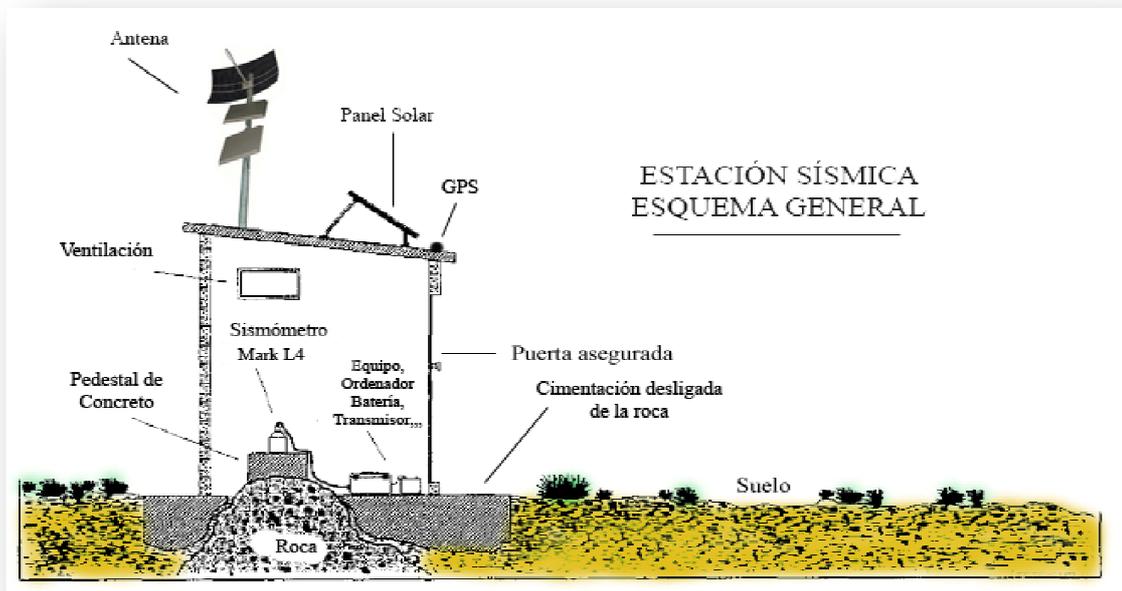


Fig. 3.8 Esquema general de la instalación y el equipo de la estación sísmica de tipo B. Figura modificada del original tomada de [Quaas et. al 2001]

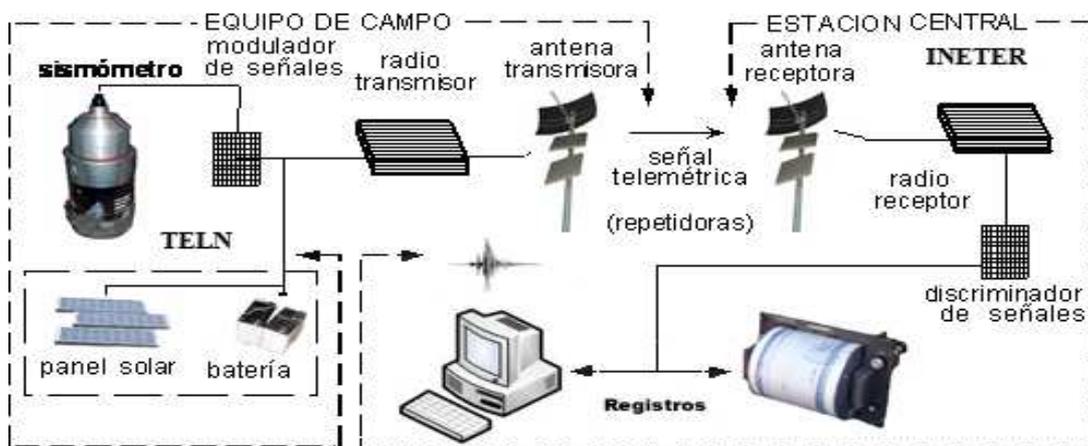


Fig. 3.9 Componentes de una estación sísmica de campo de tipo B, conectados a la estación central. Figura modificada del original, tomado de la página Web del Instituto Geofísico de la Escuela Politécnica Nacional de Ecuador (IG EPN): <http://www.igepn.edu.ec/index.php/nuestro-blog/item/114-estaci%C3%B3n-s%C3%ADsmica.html>

La elección del tipo de estación depende del objetivo de cada campaña. En las campañas en las que se dedica un determinado número de días para la recopilación de datos se eligen la opción de tipo “A”, ya que los instrumentos son llevados y recogidos por el mismo personal que realiza la campaña, además de que los datos son registrados en el sitio por el equipo. En cambio, para equipos que se encuentren de forma permanente en el campo se designa el de tipo “B”, esta opción es ideal para proteger los

equipos tanto de los elementos, animales (ya sean domésticos o salvajes que estén en las cercanías como ganado, etc.), así como la actividad misma del volcán y el vandalismo que pueda surgir. Generalmente, los criterios sobre el área de ubicación de una estación sísmica tipo B se seleccionan en base a los siguientes puntos:

1. El permiso del propietario (particular o público) del terreno donde se ubicará
2. Seguridad ante posibles nuevas erupciones (posición no en las zonas críticas)
3. Seguridad ante Vandalismo y Tránsito de animales;
(Domésticos [vacas, caballos, cabras, etc.] y salvajes en la zona)
4. Drenaje adecuado para evitar inundaciones o deslizamientos del terreno.
5. Exposición Solar (para los paneles) que asegure la alimentación de las baterías
6. Revisión de los niveles de ruido sísmico que puedan interferir en los registros
7. Facilidad de recepción de GPS para medir el tiempo de los registros
8. Fácil acceso para la instalación y el mantenimiento de la estación.

3.3.2.- Estaciones e instrumentos en el Etna

Para las campañas en las que se recopilaban los datos en Italia se utilizaron técnicas de array (ver sección 1.6), que consisten en establecer equipos interconectados en un radio de acción. El tipo de estación para este caso es el A. Esto implicaba enterrar directamente en el terreno el sensor para la captación de datos. Los sensores utilizados en las campañas de Italia han sido de tipo Mark L4C con una frecuencia natural de 1 Hz. Además se utilizaron sensores de banda ancha con tres componentes. La conexión entre los sensores y el sistema de adquisición de datos se realiza mediante cables de par trenzado. El equipo de adquisición de datos es un equipo electrónico que multiplexa, muestrea y digitaliza las señales procedentes de los sismómetros, el sistema muestrea a 200 mps., con un intervalo de 5 ms, en donde cada muestra es convertida a forma digital mediante un convertidor con aproximaciones sucesivas de 16 bits. El modelo de dicho equipo de adquisición de datos es C55016 de CRYSTAL Semiconductor Corporation, con un tiempo mínimo de conversión de 16 μ s. Los datos son controlados y almacenados en un ordenador directamente en el campo. En la figura 3.10 se observan los sitios de la ubicación de los arrays donde se colocaron las estaciones.

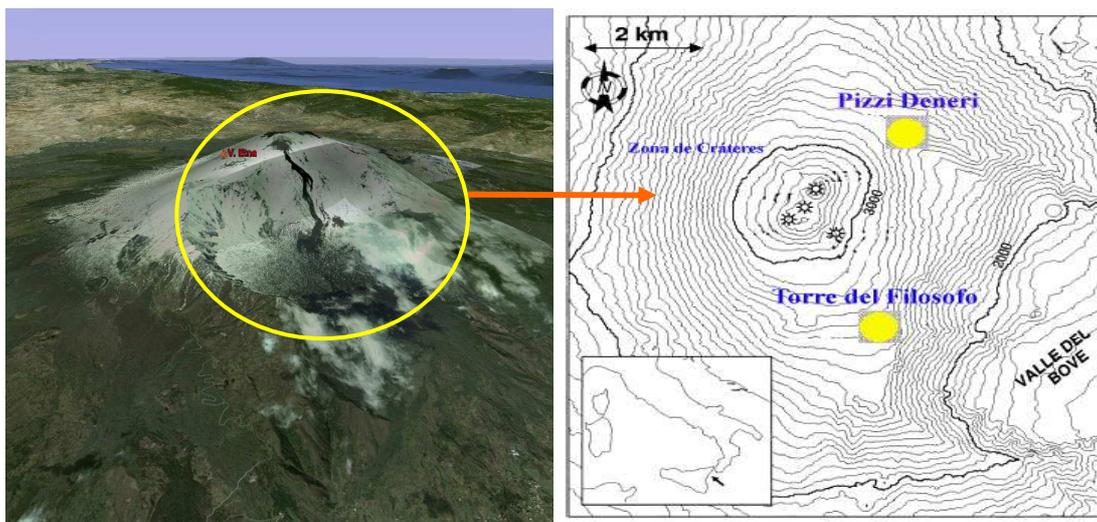


Fig. 3.10 Sitio de ubicación de las estaciones en la parte baja del volcán; Pizzi Deneri y Torre del Filósofo. (Fuente Google Earth y mapa por Ligdamis Gtz.)

La figura 3.11 muestra las estaciones de campo tipo “A” en el volcán Etna, en donde se observa la forma en la que los sensores son enterrados en el campo para la recopilación de datos. En la imagen también se observan los cables que conectan los sensores al resto del equipo de adquisición.

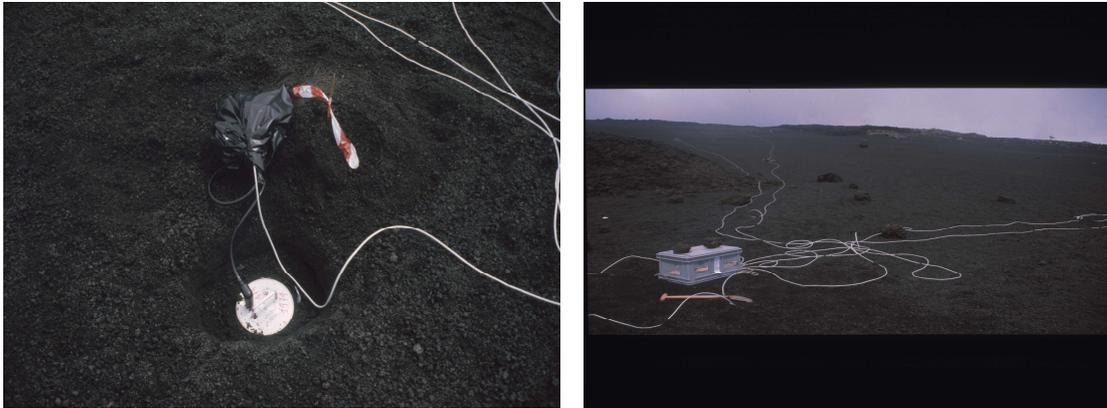


Fig. 3.11 Equipos de las estaciones en el Etna. Fuente campaña Etna

Como se observa en la imagen, para esta campaña se han utilizado varios sensores que han sido colocados en diferentes partes del terreno. Una idea de los eventos que contienen los registros los dan los sismogramas registrados en las estaciones. Un ejemplo de sismograma del Etna se presenta en la figura 3.12, en donde se observa el tremor característico.

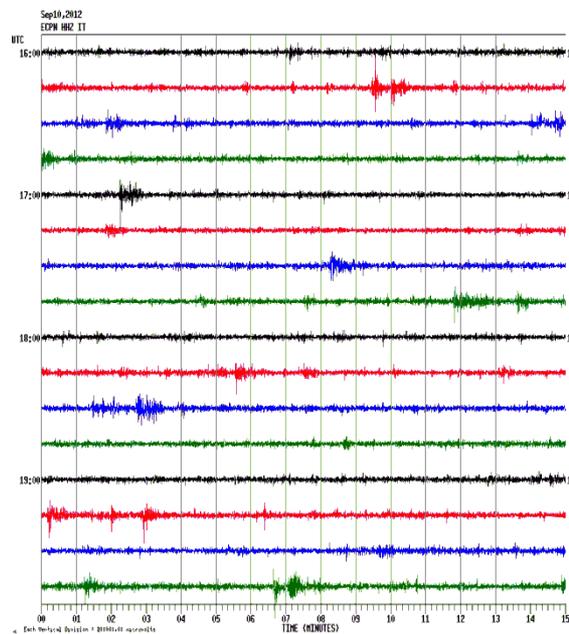


Fig. 3.12 Sismograma típico de los registros en el Etna. Fuente: INGV

Los sismogramas de Italia presentan en cada hoja, datos registrados en un intervalo de cuatro horas, siendo cada línea un registro de 15 minutos de duración.

La figura 3.13 muestra la instalación en el Etna en Septiembre de 1999, de dos arrays (ver sección 1.6) sísmicos y un perfil de estaciones 3-D equipadas con sismómetros de corto período. Para la elección del sitio de instalación se buscaron dos zonas llanas y geológicamente homogéneas, de cerca de 300 metros de ancho, posicionadas alrededor de 90° con respecto a la posición de la fuente, la cual se asumía localizada en las proximidades del área del cráter (Fig. 3.10). Los datos sobre los cuales se realiza el entrenamiento y construcción de los modelos, han sido tomados de eventos producidos durante un período eruptivo acontecido en el volcán Etna durante 1999. Dicho proceso, dio comienzo en Febrero prolongándose hasta Septiembre del mismo año y fue extensamente detallado por Calvari en sus estudios en el 2002 [Calvari et. al. 2002]. Se tomó el período de la producción de eventos acontecido durante el mes de Septiembre 1999, debido a que, durante el mismo se han producido una serie de eventos sísmicos particulares, los denominados eventos “*Tremor bursts (TB)*”, que son un caso particular de tremor que se produce por la explosión de una burbuja de fluidos o gases [Saccorotti et. al. 2004]. En el proceso de la vigilancia volcánica, resulta importante para un observador de un instituto sismológico poder identificar este particular tipo de evento.

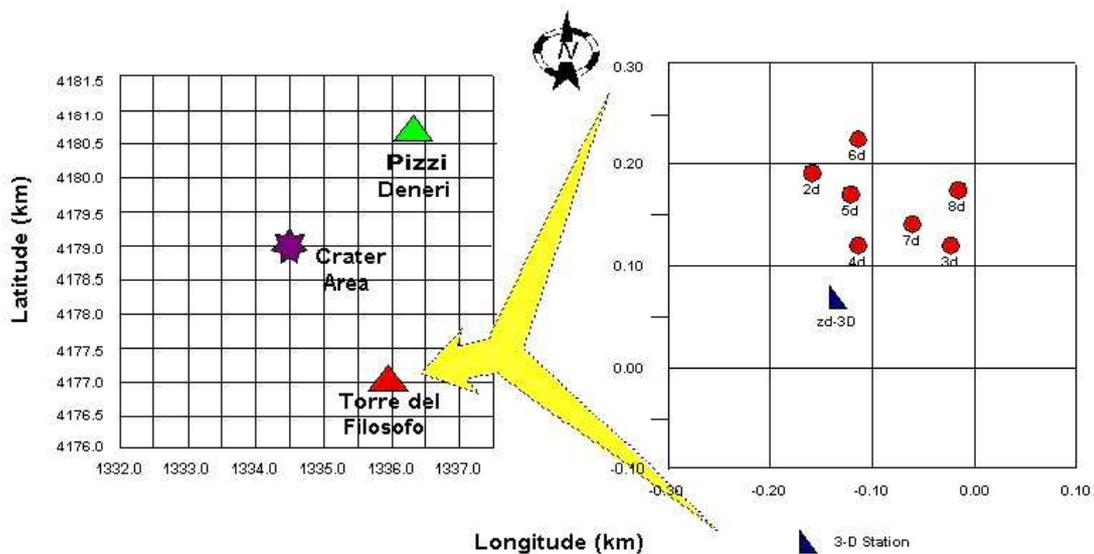


Fig. 3.13. Array sísmico Torre del Filósofo utilizado en el volcán Etna para la recolección de datos en Septiembre 1999. Fuente Gráfica Modificada tomada de la Campaña Etna

Como se observa en la figura anterior, en la campaña del Etna se instalaron los arrays correspondientes a los sitios “Pizzi Deneri” y “Torre del Filósofo”, que se identificaron por las letras “a,b,c,d,e y f” del perfil de estaciones 3-D. El proceso de instalación y recopilación de los datos ocurrido en Septiembre de 1999 fue detallado por Castellano en 2000 [Castellano et. Al, 2000]. De toda la información obtenida en dicha campaña, se estableció un proceso para poder elegir cuál de todas las estaciones que conforman el array sísmico se trabajaría, para con los datos construir las bases de datos para entrenamiento y reconocimiento de los eventos.

En este proceso se ha tomado en cuenta, que de las 6 estaciones, la estación identificada con la letra “d” (que corresponde al array de la “Torre del Filósofo”, Fig. 3.14) es la que muestra una media de datos más completa, presentando la existencia del mayor número de registros continuos en cada uno de los días de la semana a analizar.

Además, al analizar manualmente los registros se ha observado que es en dicha estación, en donde mejor se observan los eventos que se pretenden trabajar en el presente estudio.

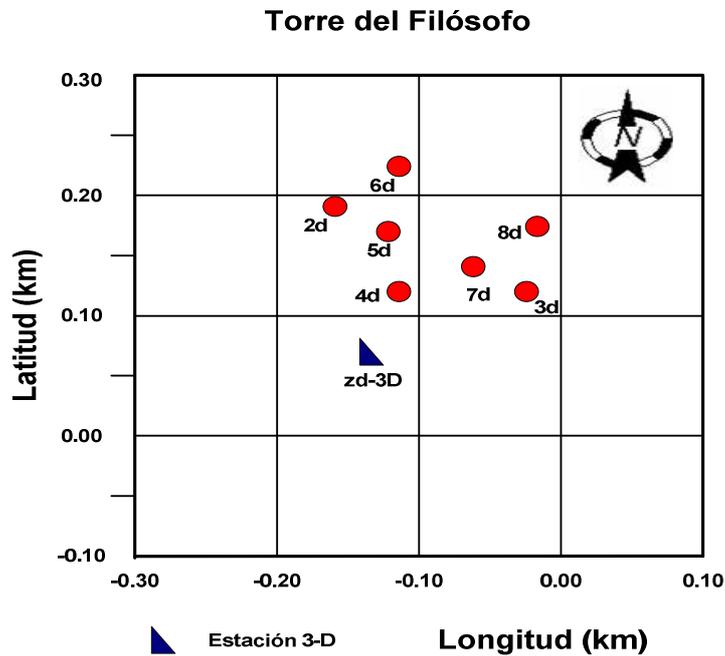


Fig.3.14. Array sísmico “d” en el sitio “Torre del Filósofo” del volcán Etna. Fuente campaña Etna. Los puntos rojos indican la ubicación de los sismómetros de una componente, el triángulo azul indica la posición de la estación de banda ancha de tres componentes. Fuente Gráfica Modificada tomada de la Campaña Etna.

Por los anteriores motivos se ha considerado que dicha estación es la más adecuada, a fin de obtener resultados representativos producidos por eventos continuos sísmico-volcánicos, ocurridos a lo largo de toda la semana. Esto se observa en la distribución de registros que se muestra en la tabla 4.1. En la campaña de recopilación de la información, el formato de cada uno de los ficheros o registros de datos que componen la lectura de eventos es el siguiente: “2540159d”, en donde los tres primeros dígitos, indican el día juliano en que han sido tomados los datos. En el caso del ejemplo anterior, el día es el 254, los siguientes dos dígitos, indican la hora que en este caso es 01. Por último, los dos dígitos restantes, que para el ejemplo son el “59”, indican los minutos. La letra que se muestra al final, indica el sistema que ha tomado las muestras, que para el ejemplo indica que se representa a la estación “d”. En el apéndice F se detallan los formatos de los eventos (Ver apéndice F).

En forma general, los datos originales han sido muestreados por todas las estaciones en períodos de una hora a lo largo de un día, esto indica que el intervalo va en formato de 24 horas, desde “00” hasta “23”, espaciados a intervalos de 3 minutos por cada registro, a lo largo de una hora esto es desde “00” hasta “9”. Cabe recordar que otro dato a tomar en cuenta es que como se ha mencionado, todos los experimentos a trabajar se han realizado con las componentes verticales. En el Etna, las estaciones están clasificadas en un periodo de una semana de muestreo. Esto significa que desde el día 254 hasta el 260, proporciona una media aproximada de recolección de 1,800 ficheros de datos a clasificar por estación, y que en el caso particular de la estación “d”, esto

representa unos 1,919 registros (o ficheros de datos), de acuerdo a lo expresado en la tabla 3.3, en donde en amarillo se indica la fila con la estación seleccionada.

Estación	No. total de registros	Registros existentes por día Juliano						
		254	255	256	257	258	259	260
A	1,680	-	252	216	390	319	370	153
B	1,737	-	252	307	390	269	374	146
C	1,448	-	241	292	391	371	153	-
D	1,919	250	410	429	414	151	259	26
E	1,993	252	408	405	416	393	119	-
F	1,527	190	-	211	409	389	297	31

Tabla 3.3. Número de registros en el volcán Etna por estación en el array sísmico “Torre del Filósofo”, en Septiembre 1999

Como se observa en la tabla anterior, la estación d contiene registros en todos los días de la campaña. El formato de los datos obtenidos en la campaña mediante el uso de sísmógrafos es binario y se almacena en ficheros binarios con extensión “.SAD”. Dichos datos han sido registrados mediante una frecuencia de muestreo de 200Hz a 8 trazas y 11,201 muestras por traza. En los registros, la cabecera de cada uno de los ficheros, viene incorporada en un fichero paralelo de extensión “dts”. Una descripción detallada de los formatos se incluye en el apéndice F de este trabajo. De esta forma, cada fichero con extensión “sad”, debe incorporar su fichero correspondiente con la extensión “dts”. Así, la estación “d” que se muestra en la tabla anterior (con los 1,919 ficheros sad), necesita de 1,919 ficheros dts adicionales. El proceso de conversión de los registros se realizó mediante una serie de programas de Matlab, a través de la creación de una lista (elaborada con los nombres de los ficheros convertidos SAD de la estación “d” muestreados a 200 Hz), a formato binario sin cabecera. Dichos programas, realizan la lectura de los datos binarios de los ficheros sad y al mismo tiempo hacen una lectura y extracción de los valores numéricos de las cabeceras contenidas en los ficheros dts. El fichero de texto con la extensión (dts), contiene la información de cabecera del contenido de los datos. Además, presenta una información complementaria referente al instrumento de medición y las fechas (ver apéndice H).

3.3.3.- Estaciones e instrumentos en el Strómboli

En la figura 3.15 se observa una imagen del Strómboli, en donde de forma similar al Etna se colocaron una serie de estaciones compuestas por una antena o array sísmico (ver sección 1.6). Estos arrays se ubican en dos sitios diferentes denominados Labronzo y Ginostra.



Fig. 3.15 Volcán Strómboli, las estaciones al igual que en el Etna se colocaron en la parte baja del volcán. (Fuente Google Earth y Ligdamis Gtz.)

La figura 3.16., muestra un mapa de localización de dichos arrays, el primer array fue localizado en el antiguo cono de Tampone del Foco, cerca del pueblo de Ginostra; el segundo array se instaló en Labronzo, aproximadamente en el mismo sitio del array descrito por Chouet en su trabajo de 1997 [Chouet et. al. 1997].

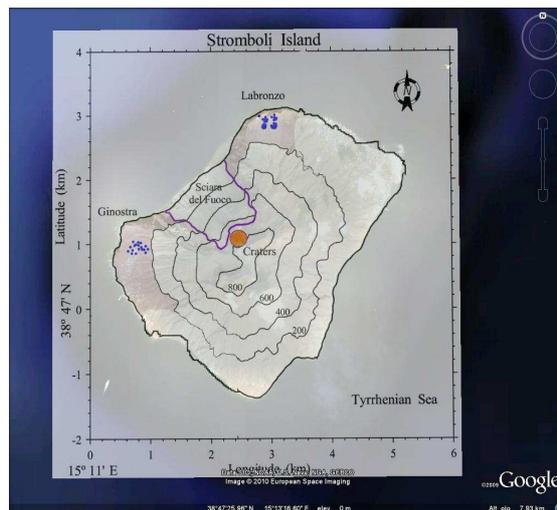


Fig.3.16. Sitios de ubicación de los Arrays sísmicos Labronzo y Ginostra en el Strómboli para la recolección de datos en 1997. (Fuente Google Earth y Ligdamis Gtz.)

En cuanto a los componentes de las estaciones del Strómboli son similares a los del Etna, por lo que son estaciones del tipo “A”, y se componen de equipos sensores (sismómetros), enterrados directamente en el terreno a los que se unen los cables con los

demás equipos de alimentación (energía), adquisición y procesamiento de los datos. En la figura 3.17 se observa una de estas estaciones colocada en el Strómboli



Fig. 3.17 Estación en el volcán Strómboli. Fuente Campaña Strómboli

Los dos arrays sísmicos son de corto período y de pequeña apertura, el array Ginostra incluyó 15 equipos o sismómetros de una componente vertical y tres sismómetros de banda ancha con tres componentes. El array de Labronzo incluyó 26 sismómetros de 1 componente vertical y dos de tres componentes. La Roca y asociados describieron detalladamente en su trabajo, las características electrónicas de los equipos utilizados en dichos arrays sísmicos [La Roca et. al. 2004]. Los eventos registrados presentan características muy definidas, un ejemplo de un sismograma característico en el Strómboli se puede observar en la figura 3.18. Al igual que en el Etna, los datos registrados se muestran en un intervalo de cuatro horas, siendo cada línea un registro de 15 minutos de duración.

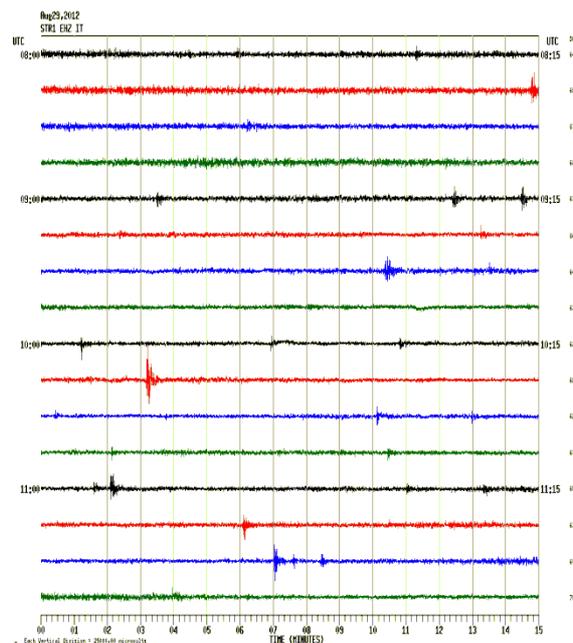


Fig. 3.18 Sismograma típico de los registros en el Strómboli. Fuente: INGV.

La configuración de las estaciones se estableció por medio de letras. Las estaciones ubicadas en el array de Ginostra corresponden a las letras: E, F, y G. Las estaciones ubicadas en el Array de Labronzo, corresponden a las letras: A, B, C, y D. El primer paso fue analizar dichas estaciones para poder elegir cuál sería la más adecuada para el estudio.

De la misma forma que la elección de los datos en el Etna, para el caso del Strómboli, se ha tomado como referencia básica aquellas estaciones en donde se concentra la mayoría de registros. Después de analizar cada una de las estaciones, se ha elegido la estación “G”, que corresponde al array situado en Ginostra (Fig. 3.19). La estación G es la que cuenta con una cantidad suficiente de registros y será utilizada como la fuente de los datos, desde donde se obtendrán para este caso los eventos característicos y definidos en el Strómboli denominados “explosiones Strombolianas”, o simplemente “explosiones” las cuales se analizarán en el presente trabajo.

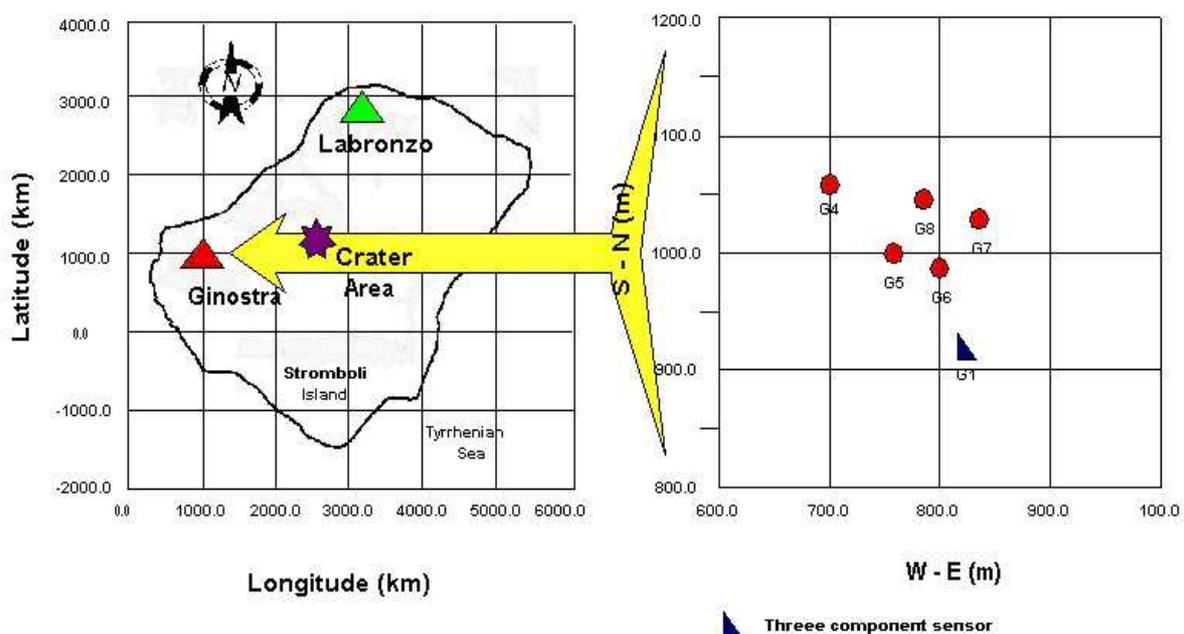


Fig.3.19. Array sísmico Ginostra en el Strómboli. Elección de las estaciones G. Fuente Gráfica modificada tomada de la Campaña Strómboli

De forma similar a la campaña del Etna, el formato de los datos se almacena en ficheros binarios con extensión “.SAD”. Igual que en el Etna los datos han sido registrados mediante una frecuencia de muestreo de 200Hz. Los registros contienen además la cabecera de cada uno de los ficheros, que viene incorporada en un fichero paralelo de extensión “dts”. En el apéndice F se hace una detallada descripción de la composición de los formatos (Ver apéndice F). Una de las ventajas de la obtención de datos mediante el uso de array sísmico, es que se puede hacer una elección de los mejores registros tomados en las diversas estaciones que conforman el array. Así se tendrán como datos recopilados por las estaciones “G” en el Strómboli, los eventos de tipo “Explosiones”, más el ruido de fondo.

3.3.4.- Elección de los sitios de estudio en Nicaragua

En Nicaragua se aprovechó la red sísmológica de monitoreo volcánico existente. Así se disponía de ocho potenciales volcanes que contaban con un monitoreo por parte de la red sísmica que son: El Cosigüina, El San Cristóbal, El Telica, Apoyeque, El Cerro Negro, El Momotombo, El Masaya y El Concepción. El proceso de elección de los volcanes de estudio, contó con la participación de los sísmólogos del observatorio local (INETER), para poder determinar cuál de los ocho edificios activos en Nicaragua, presentan una mayor cantidad y calidad de eventos que se podrían analizar. Una de las variables fue que no todos los volcanes disponían en el momento de la realización de la campaña instrumentos operables en todo el tiempo necesario en una recopilación de eventos para poder construir una adecuada base de datos, algunos de ellos se encontraban con diferentes tipos de averías (tanto de los instrumentos en si, como fallos en la comunicación, baterías, etc., o pérdidas como resultado del vandalismo), las cuales producen lapsos de tiempo, en los que por ende, no es posible registrar ningún tipo de evento. Este fue un primer análisis que llevó varias semanas, pues había que revisar toda la red sísmológica instalada en los alrededores de los volcanes, para poder elegir dos de ellos. En dicho proceso se determinó que volcanes como el Cerro Negro y el Masaya que aunque son de los más activos, presentan un proceso sísmico-volcánico bien definido, que consiste en un tremor constante a lo largo del tiempo, con pocos eventos sísmicos de otro tipo, como pueden ser explosiones, terremotos o LP, que podrían por ejemplo compararse con las explosiones del Strómboli. Además, con el tipo de estaciones sísmicas permanentes de tipo B, existen varios inconvenientes como son: el poder realizar viajes técnicos a corto plazo, para realizar las reparaciones pertinentes en los equipos sísmicos instalados en el campo, también está la dificultad de acceso al sitio, la distancia, falta de recursos como el presupuesto asignado o el personal técnico disponible, variabilidad del clima o el tiempo atmosférico al momento de realizar la campaña, caminos cortados por inundaciones, avalanchas o lahares, etc. Todos estos inconvenientes repercuten en que la programación de los viajes a los sitios se realiza con un intervalo de tiempo muy espaciado, e implican además realizar diferentes labores en las zonas (para aprovechar el viaje al máximo), no limitándose a un solo sitio específico como lo es un determinado volcán. Esto pasa sobre todo con volcanes como El Concepción en la isla de Ometepe, el Momotombo, el San Cristóbal, Telica, Cerro Negro, Cosigüina entre otros. Otro factor a determinar es la red de comunicación entre el sitio y el observatorio, ya que las antenas de tipo Yagi y Tranzeo de las estaciones sísmicas son de telemetría, comunicándose a través de antenas repetidoras al exterior. Dichas estaciones se comunican vía IP al INETER mediante un proveedor de servicios, que es la Empresa Nacional de Telecomunicaciones ENATREL (antes ENTRESA). Lo anterior se realiza en diferentes pasos que son: vía radio desde las repetidoras hacia la cercana ciudad de Chinandega, desde ahí entra a la fibra óptica en una sub-estación de telecomunicaciones situada en la ciudad de “El viejo” y finalmente se comunica por este medio a la planta central del edificio Nacional de Telecomunicaciones (TELCOR) en Managua y de ahí finalmente la señal llega al observatorio de INETER vía radio de 2.4 GHz.

Este proceso de comunicación desde la estación en el campo hasta llegar a los ordenadores del observatorio, presenta en diferentes períodos, varias y a veces severas interrupciones, debido a diversos factores, como son: pérdida de energía por agotamiento de las baterías en el campo o interrupción en el flujo eléctrico de la red nacional, daño en los paneles solares, daños por factores ambientales, por vandalismo o

robo en las instalaciones, o simplemente una mala calidad a la hora de la retransmisión por parte del proveedor de servicios. Por último, otro aspecto determinante en la elección ha sido el geológico, ya que los dos volcanes son representativos de dos de los cinco complejos que conforman la cadena cuaternaria de los Marrabios o Marribios (Ver Apéndice B). Como volcanes representativos de sus respectivos grupos o complejos volcánicos, los dos pueden aportar en su estudio, similitudes o diferencias significativas, lo que se pondrá a prueba al realizar los experimentos descritos en este trabajo.

Por todos estos factores se determinó que los volcanes: San Cristóbal y Telica, disponían de las instalaciones sísmicas adecuadas, que emitían registros constantes y que además producen diferentes tipos de eventos, muy significativos en cantidad y calidad para poder construir una adecuada base de datos para el análisis. Además de que el acceso a dichos volcanes para efectuar reparaciones, visitas técnicas o recopilación de datos in situ es muy óptimo. La posición geográfica de los dos complejos volcánicos se observa en la figura 3.20.

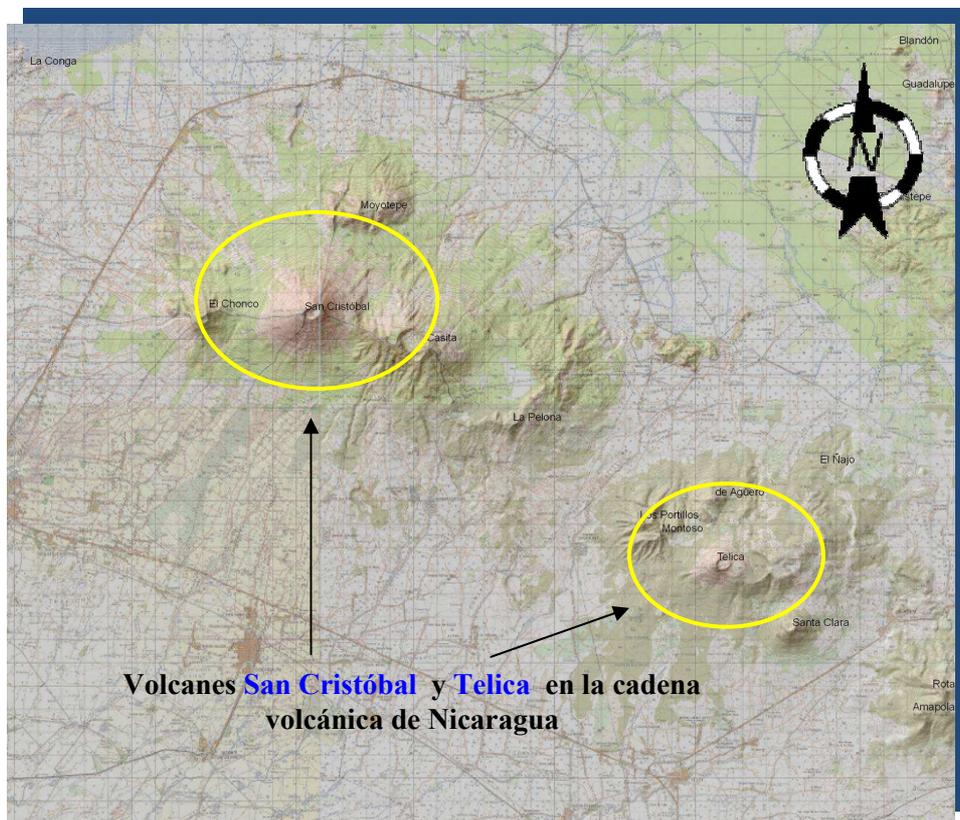


Fig. 3.20 Mapa de la ubicación de los dos complejos volcánicos: San Cristóbal y Telica y sus representativos volcanes (círculos amarillos) en la cadena volcánica de Nicaragua. (Fuente INETER y Ligdamis Gtz.)

Los datos han sido recopilados en las estaciones y enviados vía radio a la central sísmica, ubicada en la capital, Managua. Una vez recopilados en el campo son procesados por el sistema de adquisición de datos existente. Por ejemplo, el sistema Earthworm, que en combinación con el sistema de reconocimiento SEISAN presenta diferentes reportes de dichos registros.

3.3.5.- Estaciones e instrumentos en el San Cristóbal

En Nicaragua, las estaciones sísmicas son del tipo B, constituidas por una caseta de concreto en la que los instrumentos se encuentran dentro. La estación en el volcán San Cristóbal fue instalada el día 05/03/99 y está ubicada a $87^{\circ}03.00'O$ de longitud y $12^{\circ}42.00'N$ de latitud, a una altura de 685 mts. (Figura 2.38). Los datos han sido recopilados por una estación sísmica telemétrica (USA) Tx. El código de la estación es CRIN, el equipo sensor utilizado es un equipo TX marca Monitrón, modelo TR-210, La frecuencia natural del sensor es de 0.970 Hz. Las estaciones poseen un procesador marca SeiscomP que se encarga de la recopilación, procesamiento y transmisión de los datos. El sismómetro es de marca Mark modelo L-4. La estación posee un panel solar. Además, la alimentación hace uso de una o varias baterías. Para transmitir al observatorio, la antena es de tipo Yagui y/o Tranzeo, que se conecta al equipo mediante un cable coaxial con una longitud de 10 mts. El control y almacenamiento de los datos se realiza por medio de PC (Earthworm instalado en el INETER), que corre bajo Windows. Se mantiene un segundo sistema SEISLOG-SEISAN como respaldo para el caso de mal funcionamiento del sistema principal Earthworm. Un ejemplo de los sismogramas de registros en el San Cristóbal se observan en la figura 3.21.

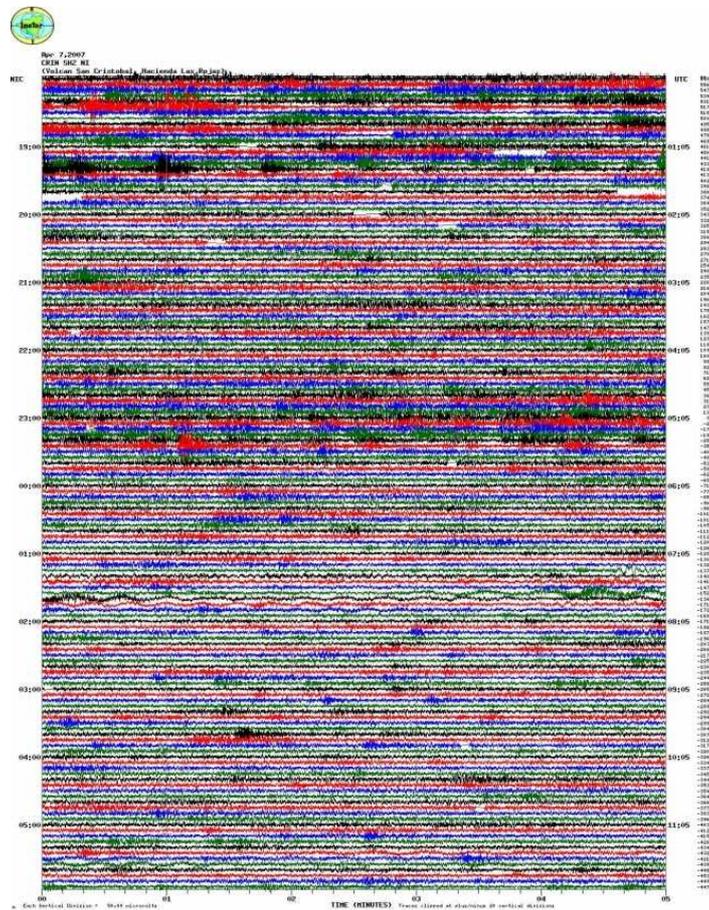


Fig. 3.21 Sismograma típico de los registros en el San Cristóbal. Se observa la presencia de un fuerte tembor de fondo en los registros del San Cristóbal Fuente: INETER

En el San Cristóbal hay cuatro estaciones que estaban siendo operativas, las marcadas con las siglas “MOCN”, “SCRN”, “CRIN”, y “CHIN”, esto se observa en la figura 3.22. Diversos factores, como el costo del mantenimiento, el difícil acceso a las

zonas, vandalismo y por ello la cantidad de registros disponibles, han hecho que no todas las estaciones en el volcán sean las más fiables a la hora de la recopilación de los datos. Con la ayuda de los sismólogos locales y de la revisión de los registros mediante los sismogramas diarios (ver fig. 3.21) que llegan a la central, se ha podido determinar que la estación “CRIN” es la estación más idónea para poder recopilar la información.

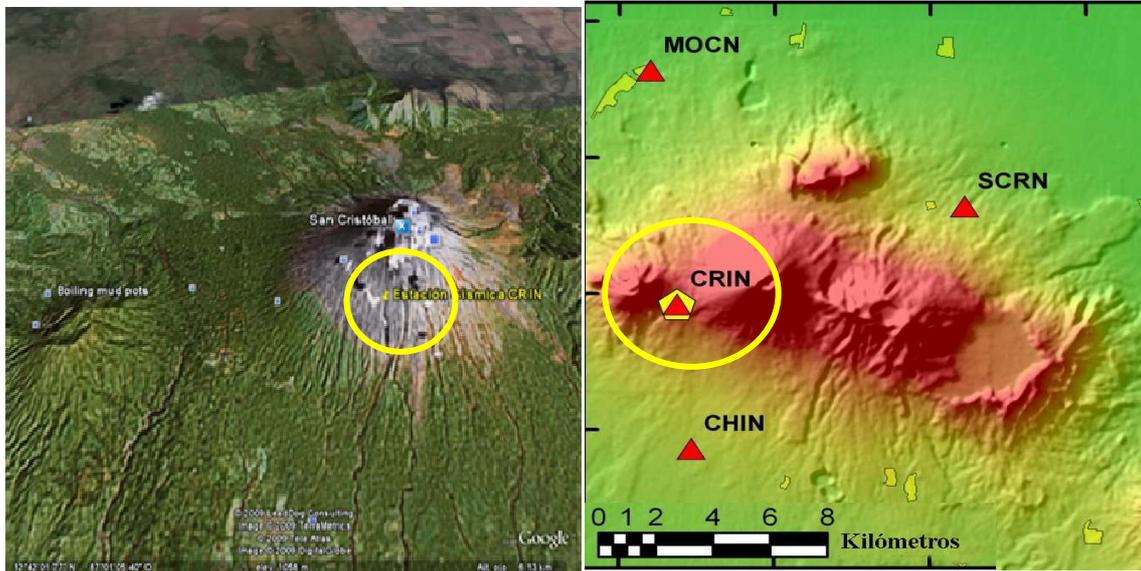


Fig. 3.22 Mapa de la ubicación de las estaciones en el volcán San Cristóbal. Se observa la estación **CRIN** (círculo amarillo) elegida. (Fuente Google Earth e INETER)

La estación “CRIN”, está ubicada en el volcán a unas decenas de metros por debajo de la cumbre y del cráter principal (Fig. 3.23). De todos los registros llegados a la estación central, se almacenarán y procesarán los de dicha estación.



Fig. 3.23 (Izquierda), Autor en la caseta sísmica CRIN, se observa la cumbre del volcán al fondo. En la imagen de la derecha la flecha señala la ubicación de la caseta respecto al cráter en la cima del volcán. Fuente Ligdamis Gtz.

Las estaciones sísmicas de campo y la estación central, poseen varios equipos electrónicos destinados a la recopilación de los datos. Algunos de los instrumentos sísmicos con que cuentan las estaciones sísmicas y la central de datos, pueden observarse en las figuras 3.24 a 3.28.



Fig. 3.24 A la izquierda Sismógrafos verticales MARK serie Cs5016, provistos de sensores L4C-1 Hz, a la derecha paneles solares que alimentan las baterías, GPS para sincronizar el tiempo de registro y antena retransmisora en las casetas sísmicas. Fuente Ligdamis Gtz.



Fig. 3.25 Interior de una caseta sísmica, en donde se pueden observar las baterías que alimentan los equipos. Abajo a la derecha se observa la caja de color negro del sistema SEISCOM que procesa y recopila los datos, el radio transmisor y los convertidores digitales. Fuente Ligdamis Gtz.

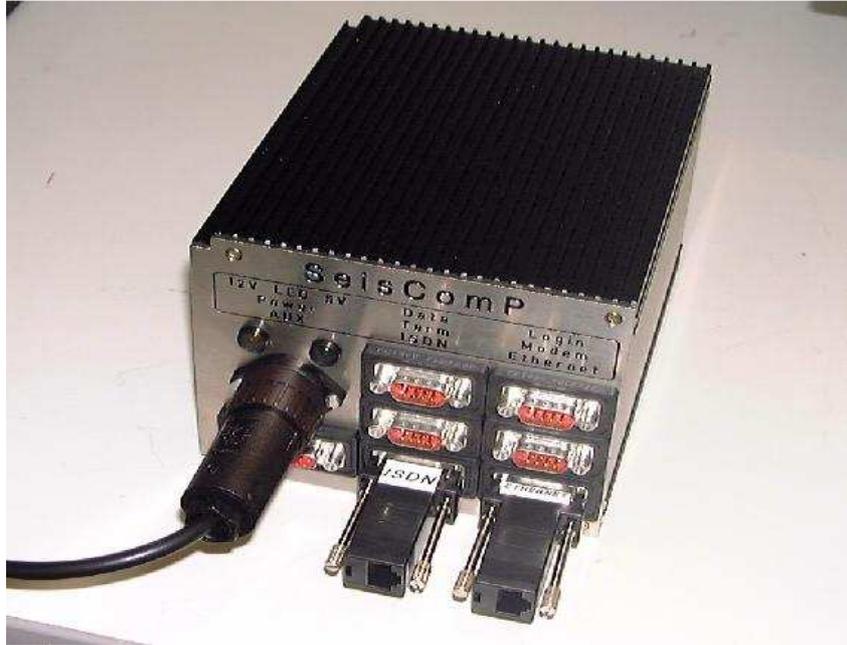


Fig. 3.26 Parte principal del sistema de adquisición y procesamiento y transmisión de datos de las estaciones sísmicas “*Seismological Communication Processor (SeisComP)*” por sus siglas en Inglés (foto fuente GeoForschungsZentrum (Potsdam): SeisComP Software y Hardware Web <http://geofon.gfz-potsdam.de/geofon//seiscomp/pc104.html>).

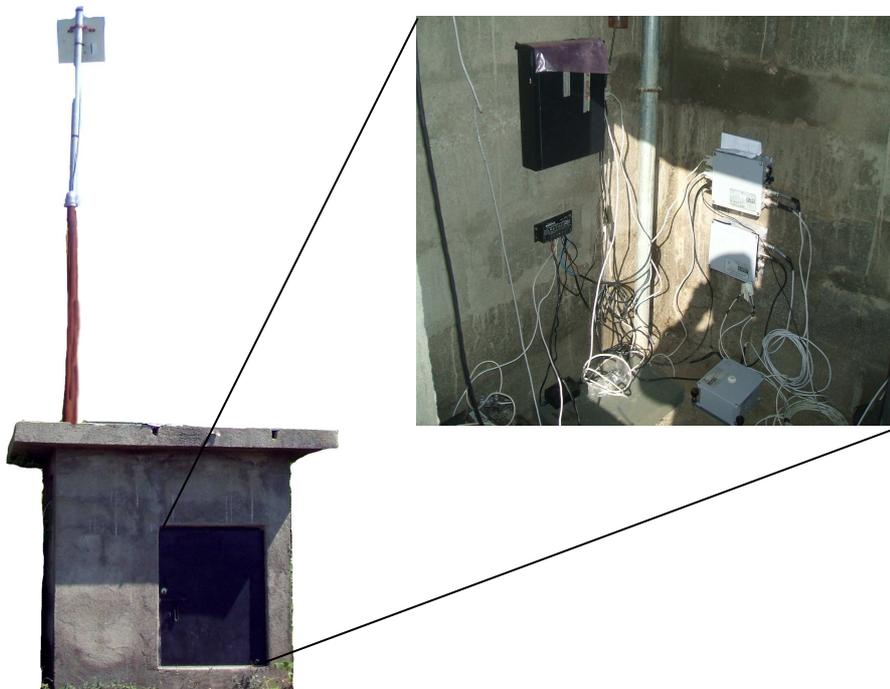


Fig. 3.27 Vista de la caseta sísmica y el interior con los diferentes instrumentos sísmicos. También se observa la torre de comunicación en donde se coloca la antena TRANZEO. Fuente Ligdamis Gtz.



Fig. 3.28 Autor en la Central Sísmica de recepción de datos ubicada en el INETER, Managua.
Fuente Ligdamis Gtz.

3.3.6.- Estaciones e instrumentos en el Telica

La estación fue instalada el día 06/03/99 y está ubicada a $86^{\circ}49.88'O$ de longitud y $12^{\circ}36.25'N$ de latitud, a una altura de 850 mts. Se encuentra ubicada cerca del cráter del volcán (Fig. 3.16 y 3.17). Los datos han sido recopilados por la estación sísmica telemétrica (USA) Tx. El código de la estación es TELN. La estación consta de: a) Un transmisor telemétrico VHF, b) Un sismómetro o geófono Mark L-4, c) Un amplificador sísmico. La composición y característica de los instrumentos de la estación es la misma que en los descritos anteriormente ubicados en el volcán San Cristóbal. Un ejemplo de sismograma que presenta los registros del Telica se observa en la figura 3.29.

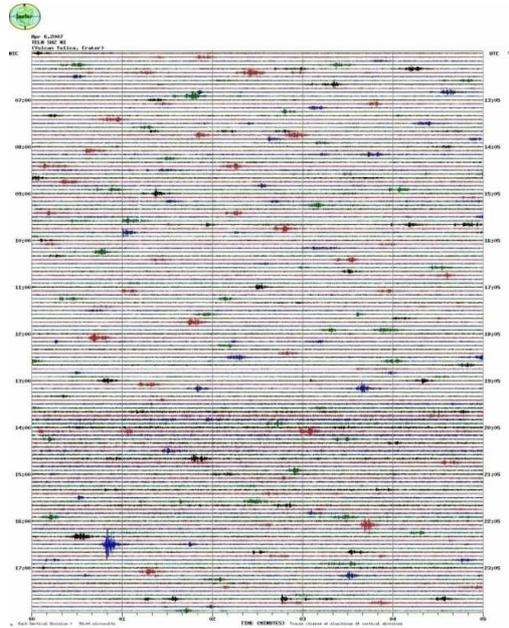


Fig. 3.29 Sismograma típico de los registros en el Telica. A diferencia del San Cristóbal, el Telica presenta un tremor de fondo mucho menor. Fuente: INETER

Los sismogramas de Nicaragua presentan los datos registrados en un período de 12 horas. Cada línea dibuja el sismograma de 5 minutos (un registro almacenado). Al lado izquierdo se da la hora en tiempo local y a la derecha en tiempo mundial (UT). En el registro se observa también el tremor de fondo en cada volcán; es decir, un movimiento constante en el volcán, donde su amplitud puede variar con el tiempo, además de los distintos tipos de eventos generados. Se puede visualizar la complejidad y la cantidad de los registros a analizar. En el volcán Telica en el período desde 1999 hasta 2006, el INETER ubicó tres estaciones sísmicas de tipo B en los alrededores del volcán, denominadas con las siglas “AGUN”, TELN” y TEL3”, que se observan en la figura 3.30

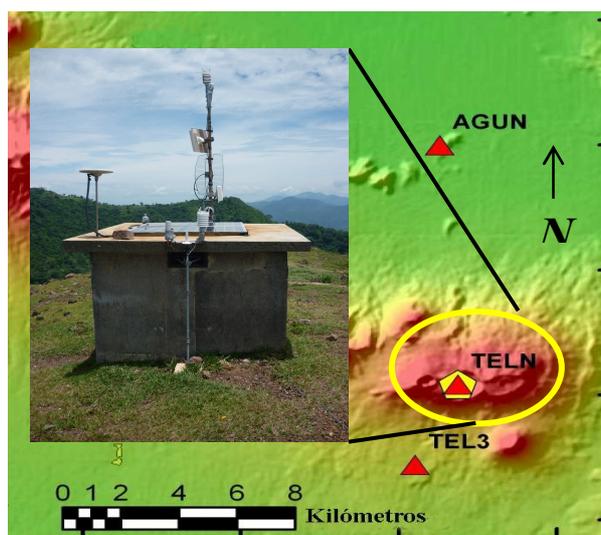


Fig. 3.30 Mapa de la ubicación de las estaciones en el volcán Telica de 1999 al 2006. En el círculo amarillo la estación TELN elegida. Fuente INETER

Los mismos criterios que se siguieron para la elección de la estación en el San Cristóbal se aplicaron para el Telica. Se designó por lo tanto para los estudios la estación denominada “TELN”, situada al borde del cráter (Fig. 3.30 – 3.33), como la más fiable para la recopilación de los registros en la campaña.



Fig. 3.31 Autor en Caseta tipo B, similar a la estación TELN, ubicada cerca del cráter del Telica. Fuente Ligdamis Gtz.

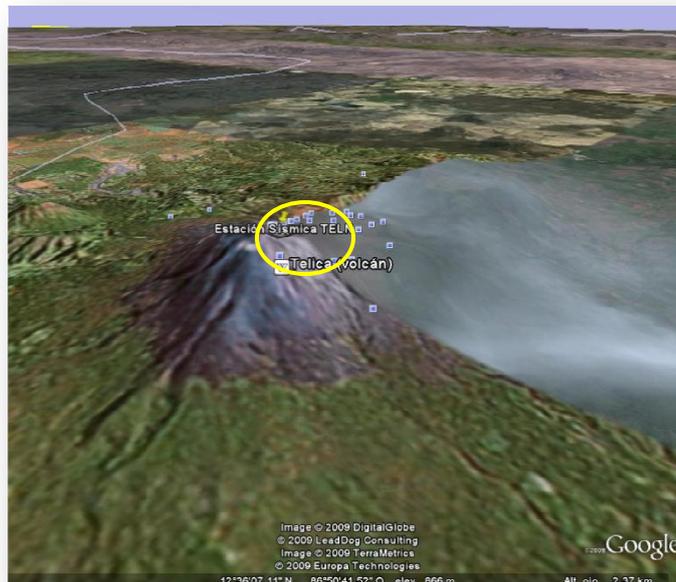


Fig. 3.32 Mapa de la ubicación (círculo amarillo) de la estación **TELN** en el volcán Telica. (Fuente Google Earth).

A diferencia del San Cristóbal, en el Telica la estación se encuentra situada al borde del cráter. En la siguiente figura se observa la estación en el Telica



Fig. 3.33 Estación sísmica **TELN** en el cráter del Telica (Fuente INETER)

En cuanto a los equipos instalados en la estación y los instalados en la central sísmica a donde llegan los datos, estos presentan similares características que los que anteriormente se han descrito para el San Cristóbal.

Por último, referente a las campañas de Nicaragua, hay que hacer hincapié en que los datos han sido tomados en los meses de Febrero y Marzo, fechas en que el país, de acuerdo a la climatología del trópico, se encuentra dentro de lo que se denomina la “estación seca”. Nicaragua, al igual que la mayoría de países que se encuentran entre los trópicos, están dominados por dos tipos de estaciones climáticas; de Noviembre a Mayo una estación con poca precipitación de agua, por lo que es denominada “estación seca”, y de Mayo a Noviembre, una estación con abundante precipitación de agua, llamada “estación lluviosa”. Lo anterior es importante porque hay que tener en cuenta la cantidad de agua presente en los acuíferos (ya sea dentro o alrededor del volcán), lo que podría influir en la generación de los eventos sísmicos, tanto en la cantidad como en el tipo de evento. Una idea de este comportamiento puede encontrarse en el apéndice C, en las estadísticas que presenta el INETER, sobre la generación de sismos a lo largo del 2006 en el San Cristóbal y el Telica. Futuras investigaciones podrían incluir campañas en los meses de la estación lluviosa, y poder comparar los análisis y resultados con los tipos de eventos que se presenten en los volcanes durante dicho período, con los encontrados en este estudio, para así observar la influencia del factor climatológico en la sismología mediante el aumento de los acuíferos en los edificios volcánicos.

3.4.- TRANSFORMACIÓN Y UNIFICACIÓN DE FORMATOS DE LOS DATOS.

En el apéndice F se detalla el proceso de conversión que se realiza con los formatos, en donde los datos tienen un tratamiento similar desde el pre-procesado de los registros, hasta la conversión de los parámetros y la construcción de las bases de datos para el entrenamiento de los modelos de reconocimiento (Ver apéndice F).

Para Italia, en las estaciones “d” de Etna y “G” del Strómboli se tuvieron que analizar la totalidad de los registros en la búsqueda de eventos en cada estación, la tabla 3.3 da una idea de todo el trabajo a realizar con los registros (ver tabla 3.3). Se analizaron las seis estaciones con una totalidad de 10,304 registros, para así tomar la decisión sobre la estación “d”. En dicha estación se procedió a la visualización de cada uno de los 1,919 registros que componen la estación, a fin de detectar cuál de ellos contenía eventos. El caso del Strómboli se cuenta con una cantidad de registros similar a la del Etna, pues se analizaron las estaciones F, G y E, determinando que la mayor cantidad de registros se encontraba en la estación G.

En la campaña de Nicaragua, las listas de registros recabados durante las visitas de campo a los volcanes constituyeron un total de 6,348 registros continuos de 5 minutos cada uno en cada volcán. Esto implicaría tener que analizar de manera visual 12,696 registros totales, lo que corresponde a 1,058 horas de registros en busca de los tipos de eventos previamente catalogados.

Una vez almacenada la información, se ha de proceder a tratarla para poder ser analizada. Para todos los volcanes, como primer paso se ha realizado la recopilación en la estación receptora de los registros en bruto, que en el caso de Italia se almacenaron directamente en formato SAC (mediante el software de adquisición de datos sísmicos), en los ordenadores de campo. Por un momento se dejará a un lado el tratamiento de las señales de las campañas de Italia, para poder detallar todo el procedimiento de las campañas de Nicaragua, ya que en ellas como se ha mencionado, dicho proceso implicó un mayor tratamiento de los registros, desde cero hasta llegar al punto en que todos los registros, incluyendo los de las campañas de Italia, seguirían el mismo procedimiento. Como se ha visto, para las campañas de Nicaragua, los datos enviados por las estaciones sísmicas ubicadas en los volcanes, llegan al observatorio vía FTP. Estos registros en bruto se han recopilado por el sistema Earthworm y son los llamados registros “R”, o registros en crudo o en bruto de acuerdo a los sismólogos locales. El sistema de alerta temprana local analiza dichos registros y si encuentra un evento de una magnitud considerable es analizado por el sismólogo de turno, en caso contrario todos los demás registros no son analizados, solo son contabilizados para las estadísticas. Esto ocurre debido al flujo y el volumen de toda la información que a diario y durante las 24 horas llega procedente de toda la red sísmica. Por ese motivo y ante toda esta información, solo los eventos más significativos son tomados en cuenta, por lo que los datos son filtrados para así solo obtener los registros de mayores magnitudes o aquellos cuya importancia este de acuerdo a las políticas locales. Sin embargo, durante la campaña, se necesitaba captar toda la información, que en continuo entra a través de la red sísmica (procedente de todos los canales y de todas las estaciones en su conjunto), sin descartar ningún registro para poder ser analizado y procesado. El pre-proceso de conversión se realiza mediante el sistema SEISAN. Para ello, se ha debido convertir los registros en bruto (archivos R) a formato binario SEISAN y de ahí hasta el formato de forma de

onda y ficheros r50. Este es un proceso bastante extenso que se realiza en varios pasos, que se detallan en los apéndices (ver apéndice F). Un resumen del resultado de este proceso, presentando los diversos pasos o etapas (en que los registros transitan a diversos formatos), puede observarse de manera general en la tabla 3.4.

PROCESO	FORMATO RESULTANTE	TIPO DE DATO
1	Original	2006-03-01-1411-15SC.R (Binarios en crudo)
2	Red canales	2006-03-01-1411-15S.NIC_064 (seisan 64 canales)
3	Seisan	2006-03-01-1411-15S.TELN_001 (seisan a 1 canal)
4	Pitsa	Pitsa001.001 (Ascii-Pitsa)
5	ASCII	2006-03-01-1411-15SC.001 ASCII
6	Onda	2006-03-01-1411-15SC.wav Wav
7	Onda sin Cabecera	2006-03-01-1411-15SC.raw raw
8	Onda Filtrado	2006-03-01-1411-15SC.r50 Filtrados a 50 Hz

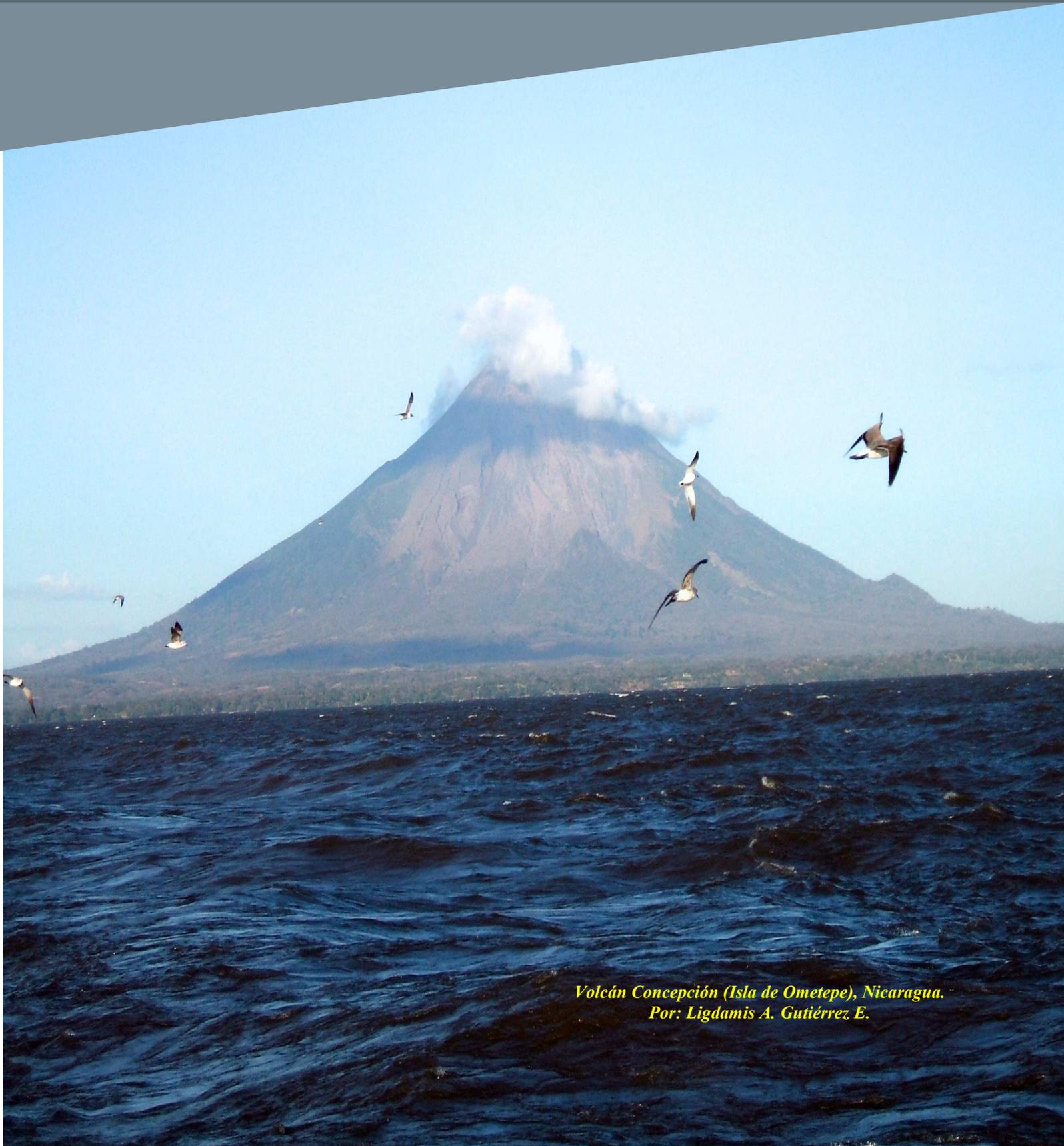
Tabla 3.4 Tipos de formatos a convertir en el desarrollo del proceso de los registros.

En el siguiente capítulo se procederá al análisis de dichos registros resultantes para establecer los tipos de eventos y construir las bases de datos.



CAPÍTULO IV

ANÁLISIS DE LOS DATOS



*Volcán Concepción (Isla de Ometepe), Nicaragua.
Por: Ligdamis A. Gutiérrez E.*



4.1.- INTRODUCCIÓN

En el anterior capítulo se determinaron las zonas de estudio y se describieron las campañas desarrolladas en Italia y Nicaragua. Además se desarrolló el proceso de la obtención de datos y los formatos necesarios en la preparación de la información. En este capítulo se procederá al análisis de los registros recopilados en las campañas, para así establecer los tipos de eventos y construir las bases de datos. Han de definirse los tipos de eventos existentes en cada volcán y se realizará un análisis espectral de dichos eventos a fin de definir sus características. Asimismo, se ha de desarrollar la estadística de la duración de dichos eventos, para hacer notar la complejidad de los mismos a la hora de establecer el reconocimiento y se determinará la segmentación necesaria en cada uno de dichos tipos, los que constituirán las bases de datos para el entrenamiento de los modelos de reconocimiento. Por último, en este capítulo se establecerá una clasificación genérica con los tipos de eventos determinados en los volcanes de estudio de Nicaragua, ya que aún no han sido catalogados, ni corresponden a ninguna clasificación en particular.

4.2.- DETERMINACIÓN DE LOS TIPOS DE EVENTOS.

El determinar los tipos de eventos lleva implícito un largo período, compuesto por varios meses de análisis, debido a que en un primer pre-procesado se debe estudiar de forma visual, cada uno de los registros que se han obtenido en los volcanes. Mediante dicho análisis, se puede catalogar y elaborar un listado de los diferentes eventos sísmico-volcánicos presentes en los registros. Esto se ha realizado en base a la experiencia humana, es decir, de acuerdo a las opiniones de expertos. De esta manera, la experiencia del observador humano es quien construye la primera base del entrenamiento. Es uno de los procesos más largos y delicados, ya que tener un buen catálogo de eventos, respalda una buena base de entrenamiento para los modelos. Como se recordará, para las campañas de Italia, los eventos a tratar han sido bien definidos por diferentes autores. En el caso del Etna es un tipo de evento que se denota como una explosión de burbuja, denominado “Tremor Burst (TB)”, y en el caso del Strómboli, lo más característico y ampliamente conocido son las explosiones de tipo estrombolianas, denominadas “Exp”. Sin embargo, en cuanto a las campañas de Nicaragua, esto ha sido una incógnita hasta la hora de comenzar a analizar y procesar toda la información. Inicialmente, en cada volcán, al revisar de forma visual cada uno de los 6,348 registros de 5 minutos continuos, se catalogaron en un primer instante alrededor de 12 tipos diferentes de eventos en cada volcán. Posteriormente, este proceso fue afinándose con la ayuda de las diferentes técnicas de análisis espectrales. Al observar los primeros experimentos, la confusión entre varios eventos y los pobres resultados, previó el que algunos de ellos presentaban similares características espectrales, así se pudieron delimitar el número de eventos, agrupando los equivalentes. De esta forma, se logró al final poder agrupar un determinado número bastante significativo de eventos en cada volcán con los que poder construir las bases de datos.

Conforme se profundizaban los grupos de señales de acuerdo a su forma, duración en el tiempo y espectro de energía, el número en cada grupo se redujo a 5 en el San Cristóbal y 6 en el Telica, dejando los que han podido establecer al final, el número definitivo de eventos, en el San Cristóbal a 3 más el ruido de fondo y en el Telica a 4 más el ruido de fondo. Con esta clasificación se ha logrado que los tipos de eventos

encontrados, difieran uno de otros de acuerdo a sus características espectrales. Esto hace concluir que los eventos (que conforman la base de datos), pueden ser utilizados sin correr el riesgo de grandes equivocaciones a la hora de confundir un evento con otro. La clasificación final que se ha realizado de los eventos se encuentra en estado óptimo para poder ser procesados en las bases de datos del sistema de reconocimiento. Otra deducción es que al observar tanto los eventos del Telica y los del San Cristóbal también difieren entre ellos. Los de un volcán no son similares a los del otro volcán. Esto se verá más adelante en el siguiente capítulo del presente estudio, al realizar los experimentos cruzados con ambas bases de datos.

Para definir el tipo al cual pertenecen, en el caso de Nicaragua, hasta la fecha, se cataloga la mayoría de los eventos como de tipo LP, pero sin llegar a una clasificación más definida. Por tal motivo, en el presente trabajo se les ha definido, para fines prácticos, etiquetarlos mediante un nombre genérico, tratándose del San Cristóbal, han sido determinados tres tipos de eventos, indicados por números, 1, 2, y 3, con la letra S denotando a San Cristóbal delante, de esta forma se denominan: S1, S2 y S3. En el caso del Telica, se han determinado cuatro diferentes tipos de eventos, nombrados al igual que en San Cristóbal, como 1, 2, 3, y 4 colocando la letra T, de Telica delante, quedando de esta forma: T1, T2, T3, y T4. En estudios y trabajos posteriores, podrán definirse y clasificarse concretamente cada evento, de acuerdo a los tipos existentes de alguna de las escuelas que manejan la clasificación y en consenso con los sismólogos locales.

Un caso especial de evento registrado en los datos del presente estudio, es el tremor continuo o armónico presente en los registros del volcán San Cristóbal. El “*Tremor Armónico*”, en su concepto más general, se define como un tipo de señal sísmica que indica cambios en el estado de la dinámica de un volcán. Generalmente este tipo de tremor se le asocia con movimientos de magma en el interior del volcán y puede catalogarse como evento precursor de erupciones. De ahí la importancia de poder detectarlos y de que en el caso particular del volcán San Cristóbal el análisis de los datos haya podido identificar señales de este tipo. Este tipo de señales suele durar desde varios minutos, hasta varios días en algunos volcanes [Ibáñez et. al. 1997].

En el San Cristóbal el tremor continuo detectado se ha podido establecer con una duración de 1500 segundos o 25 minutos. Dicho tremor es de 5 registros continuos de 5 min. (“2006-02-24-0722-15SC” al “2006-02-24-0742-14SC”). El problema con este tipo de evento es que lamentablemente, en la totalidad de los registros, no existen demasiados casos o eventos de este tipo como para poderlo ubicar en las bases de datos de entrenamiento. Sin embargo, va a ser nombrado en algunos apartados, debido a sus peculiares características. A continuación, se presentarán los diferentes tipos de eventos, obtenidos a partir de catalogar los diferentes tipos de señales sísmico-volcánicas recopiladas.

4.2.1.- Tipos de eventos del volcán Etna

El tipo de evento a analizar en los registros del Etna, se designa en base a los estudios de campo de 1999, es el denominado “*Tremor Burst*” o TB [Saccorotti et. al. 2004]. El ruido de fondo que acompaña a dicho evento ha sido denominado como NE (*Noise Etna*). Las figuras 4.1 y 4.2, muestran un ejemplo típico del evento TB.

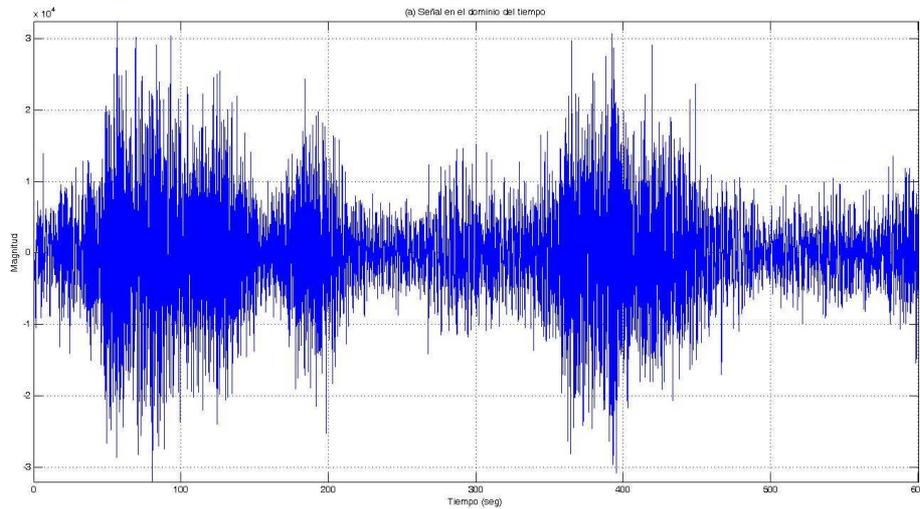


Fig. 4.1 Tipo de evento TB en el volcán Etna.

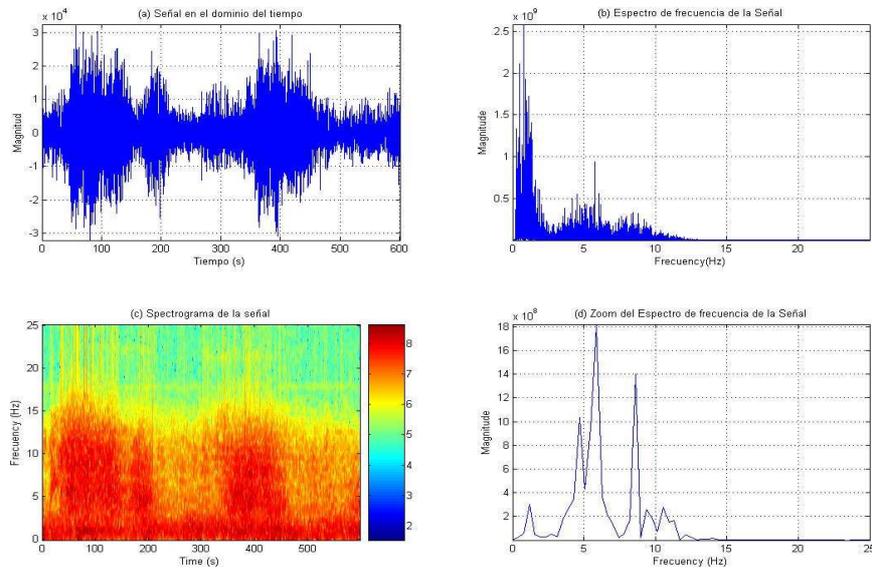


Fig. 4.2 Características del tipo de evento TB en el volcán Etna.

4.2.2.- Tipos de eventos del volcán Strómboli

El evento de estudio son las explosiones a las que se les ha denominado “exp” en el Strómboli. El ruido de fondo es denominado “Sil”. Estos datos son fruto del trabajo de campo de 1997. Las figuras 4.3 y 4.4, muestran dos ejemplos típicos de las explosiones en el Strómboli.

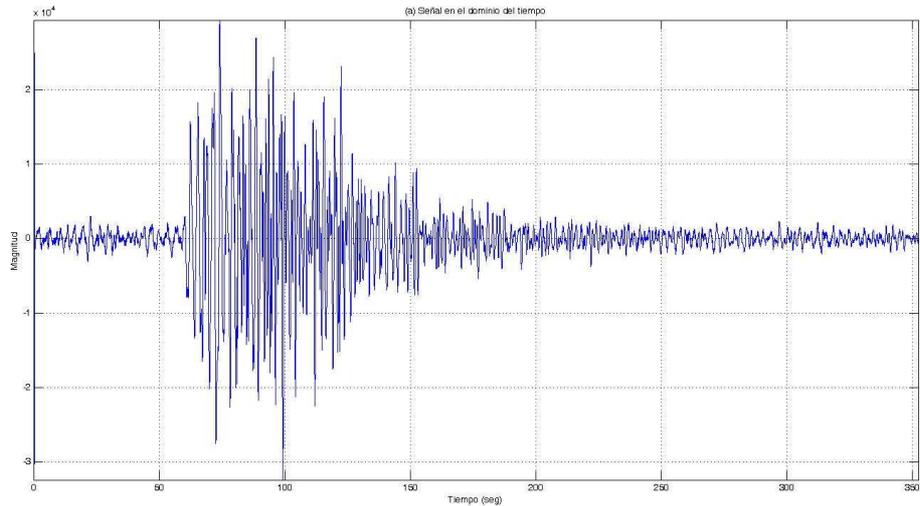


Fig. 4.3 Tipo de evento Exp en el volcán Strómboli.

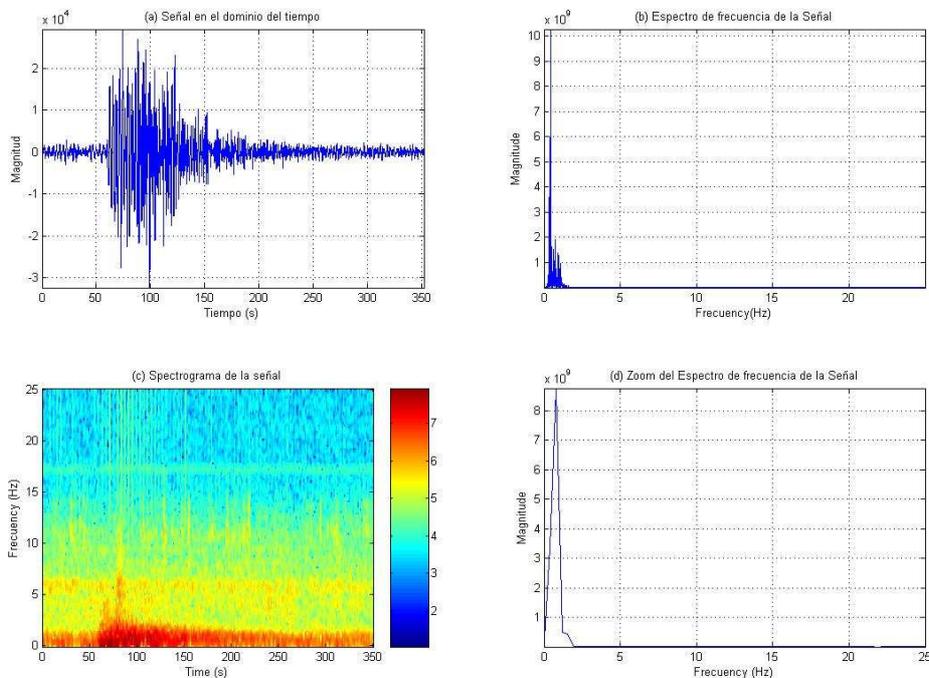


Fig. 4.4 Características del tipo de evento Exp en el volcán Strómboli.

4.2.3.- Tipos de eventos del volcán San Cristóbal

Se nombraron de forma genérica tres tipos en el San Cristóbal, tomando en cuenta la letra de inicio del volcán “S” y la numeración del 1 al 3. Así se tienen los tipos: S1 (Fig. 4.5 y 4.6), S2 (Fig. 4.7 y 4.8), y S3 (Fig. 4.9 y 4.10). El ruido (*noise*) o tremor de fondo que acompaña a los eventos se le designa como NS (*Noise San_Cristóbal*). Las siguientes figuras muestran dichos eventos.

a) Tipo de Evento S1: La figura 4.5 muestra un evento de este tipo

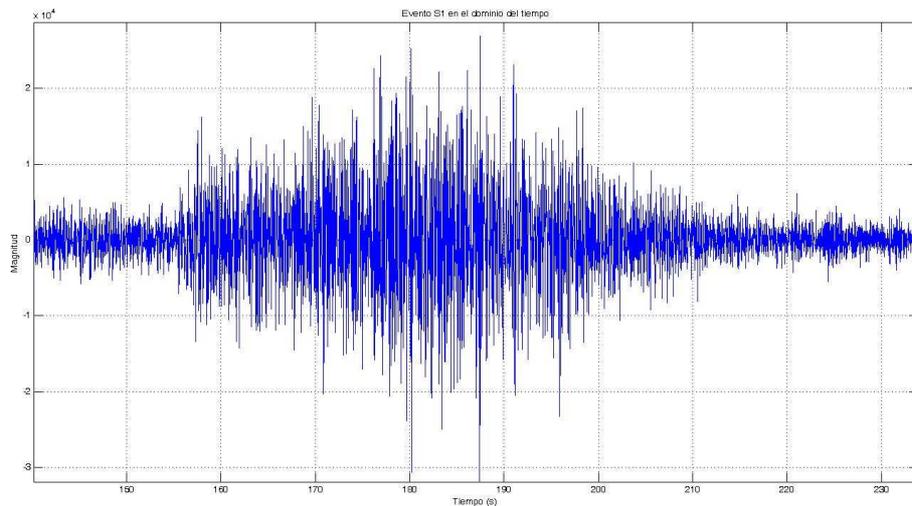


Fig. 4.5 Tipo de evento S1 en el volcán San Cristóbal

La anterior figura es un acercamiento al evento de tipo S1, algunas características, como el espectrograma y el espectro de frecuencia del evento se observa en la figura 4.6.

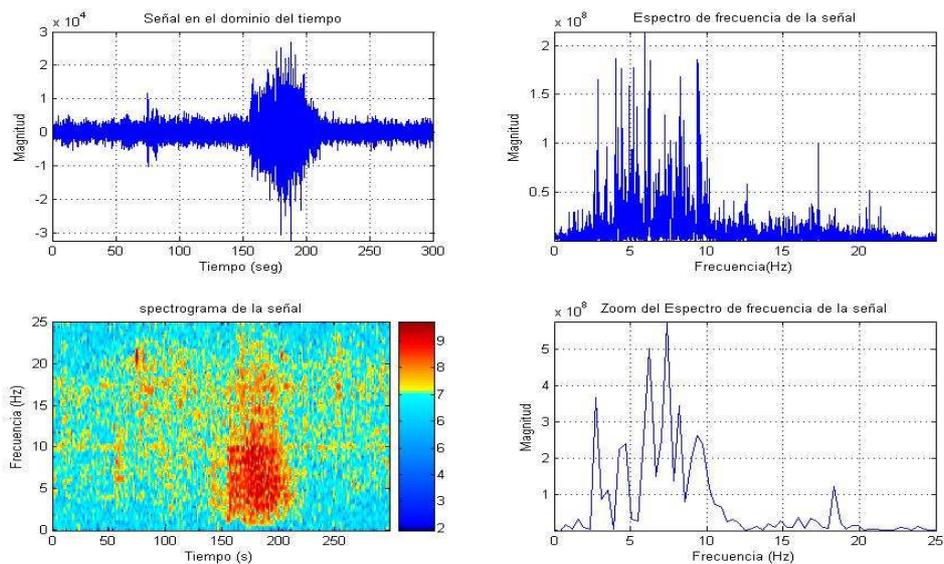


Fig. 4.6 Características del tipo de evento S1 en el volcán San Cristóbal

b) Tipo de evento S2, que se observa en la figura 4.7.

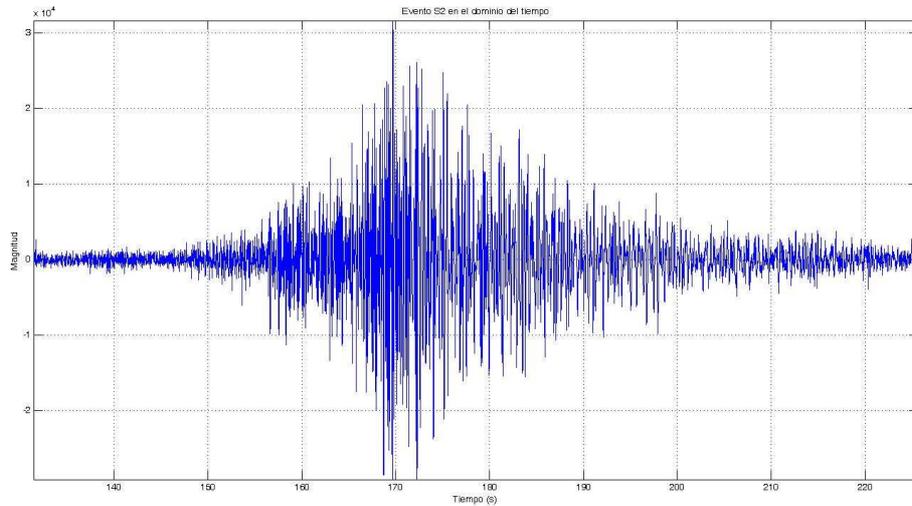


Fig. 4.7 Tipo de evento S2 en el volcán San Cristóbal

Algunas de las características del evento S2 se observan en la figura 4.8. Al comparar los gráficos entre los dos eventos se pueden observar las diferencias espectrales entre el evento S1 y el evento S2

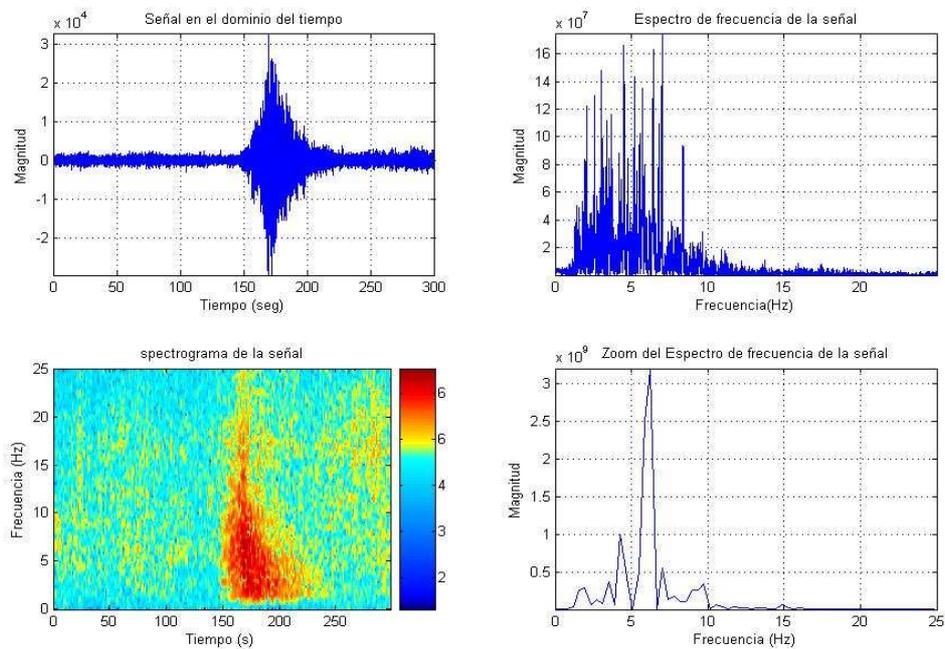


Fig. 4.8 Características del tipo de evento S2 en el volcán San Cristóbal

El tercer evento determinado en el San Cristóbal es el evento denominado S3, que se muestra en las figuras 4.9 y 4.10 siguientes.

c) Tipo de evento S3: Un evento de este tipo se muestra en la figura 4.9

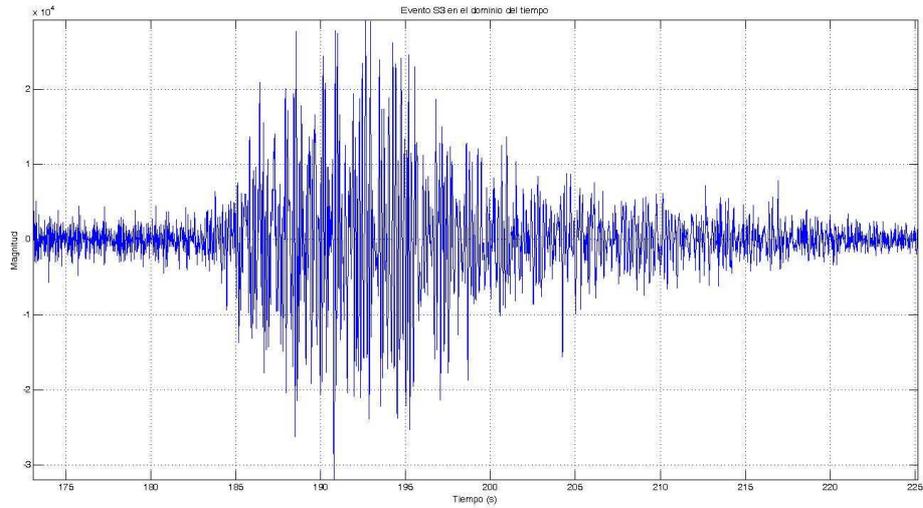


Fig. 4.9 Tipo de evento S3 en el volcán San Cristóbal

Algunas de sus características se observan en la figura 4.10.

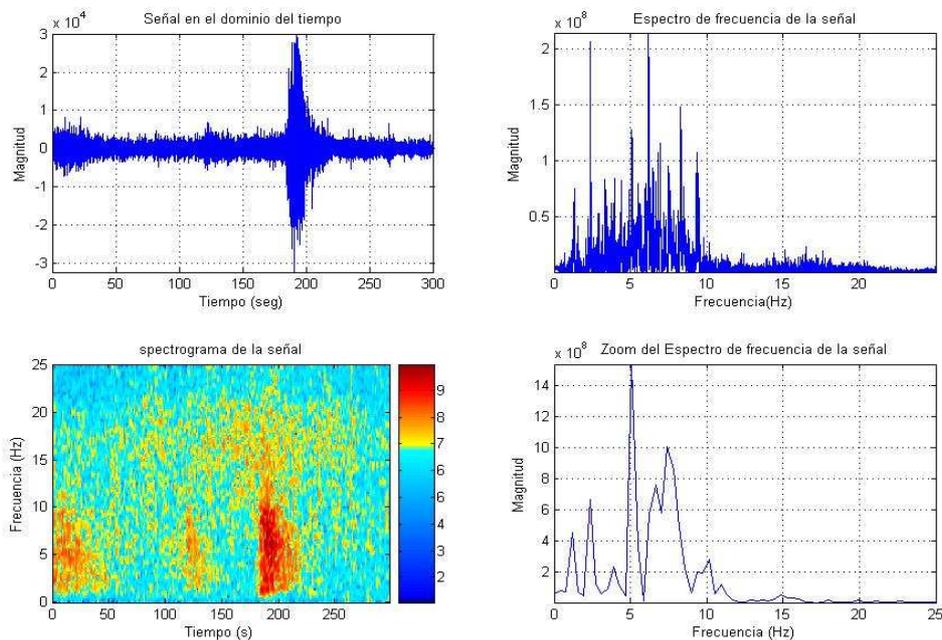


Fig. 4.10 Características del tipo de evento S3 en el volcán San Cristóbal

En los espectrogramas, el evento de tipo S3, se observa que posee una caída de energía más corta que la de los eventos S2 y S1. En cuanto al tremor en el San Cristóbal se observa un ejemplo de este en las figura 4.11 y 4.12.

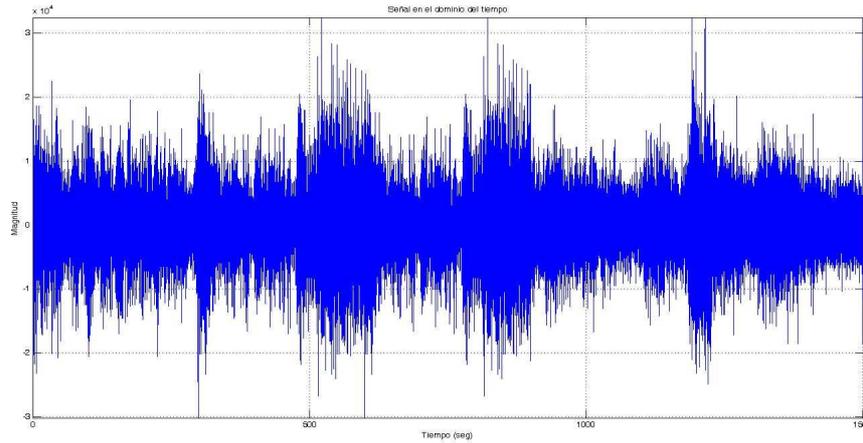


Fig. 4.11 Tipo de evento “Tremor Continuo o armónico” en el volcán San Cristóbal

Este tremor continuo ha sido registrado en un intervalo de tiempo de 1,500 seg., lo que corresponde a 5 registros de 300 segundos, que equivalen a 25 minutos de registro. Algunas características de este tipo de evento se pueden observar en el espectrograma y el espectro de frecuencia que se muestra en la figura 4.12

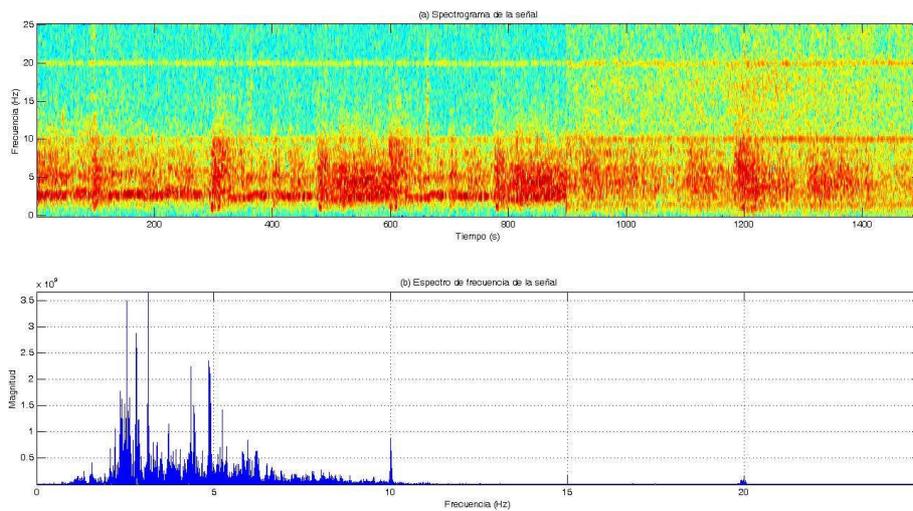


Fig. 4.12 Características del evento “Tremor Continuo o armónico” en el volcán San Cristóbal

Se puede apreciar en las imágenes que el tremor del San Cristóbal está compuesto por al menos cuatro bandas de armónicos, lo que los sismólogos denominan “tremor armónico”, evento que se ha mencionado que generalmente refleja cambios en la dinámica del volcán y puede preceder a episodios eruptivos. En la figura 4.12 se observa que los mayores picos de frecuencia en este evento se encuentran entre los 2 y los 5 Hz, siendo un evento de baja frecuencia. Lamentablemente, como ya se ha dicho en los registros de la campaña existían pocas ocurrencias de este tipo de señal para ser incluido en el presente trabajo. Sin embargo, en futuros trabajos se podrían recopilar muchos más registros de este tipo, para contar con una base de datos más completa que pudiera analizar este particular registro sísmico.

4.2.4.- Tipos de eventos del volcán Telica

De la misma forma que en el San Cristóbal, se nombraron de forma genérica cuatro tipos en el Telica, tomando en cuenta la letra de inicio del volcán “T” y la numeración del 1 al 4. Así se tienen los tipos, T1 (Fig. 4.13 y 4.14), T2 (Fig. 4.15 y 4.16), T3 (Fig. 4.17 y 4.18) y T4 (Fig. 4.19 y 4.20). El ruido (*noise*) o tremor de fondo se le nombre NT (*Noise Telica*). Para ello se representa en forma gráfica un análisis de la señal en el dominio del tiempo, así como el espectro de frecuencia de la señal y su espectrograma. Los eventos catalogados en el Telica son los siguientes:

- a) Tipo de Evento T1. La figura 4.13 muestra un evento de este tipo.

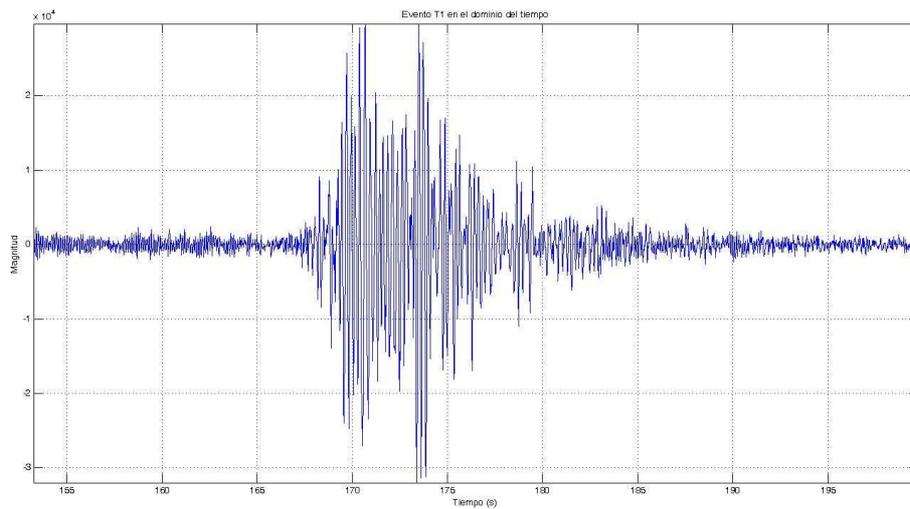


Fig. 4.13 Tipo de evento T1 en el volcán Telica

Algunas características de este tipo de evento se observan en la figura 4.14.

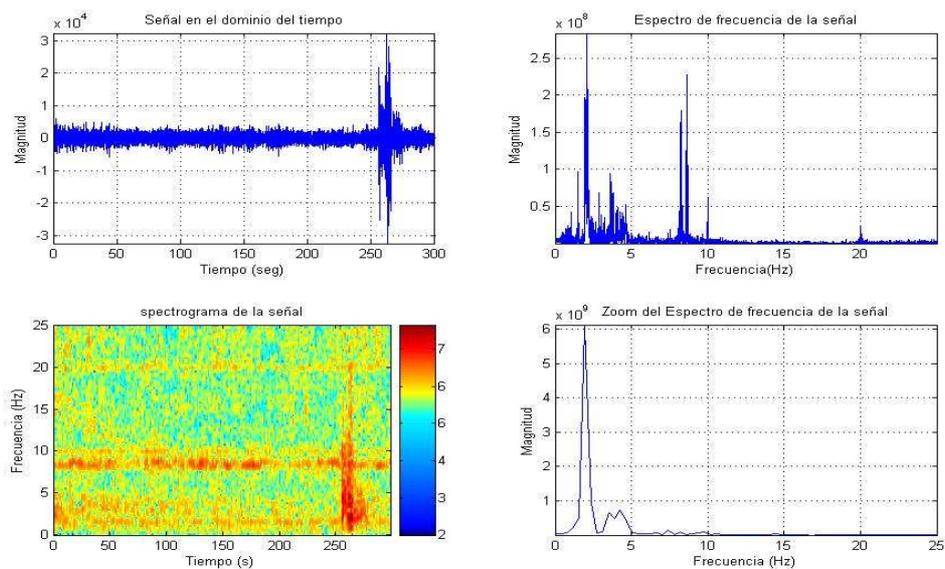


Fig. 4.14 Características del tipo de evento T1 en el volcán Telica

b) Tipo de Evento T2: La figura 4.15 muestra un evento de este tipo

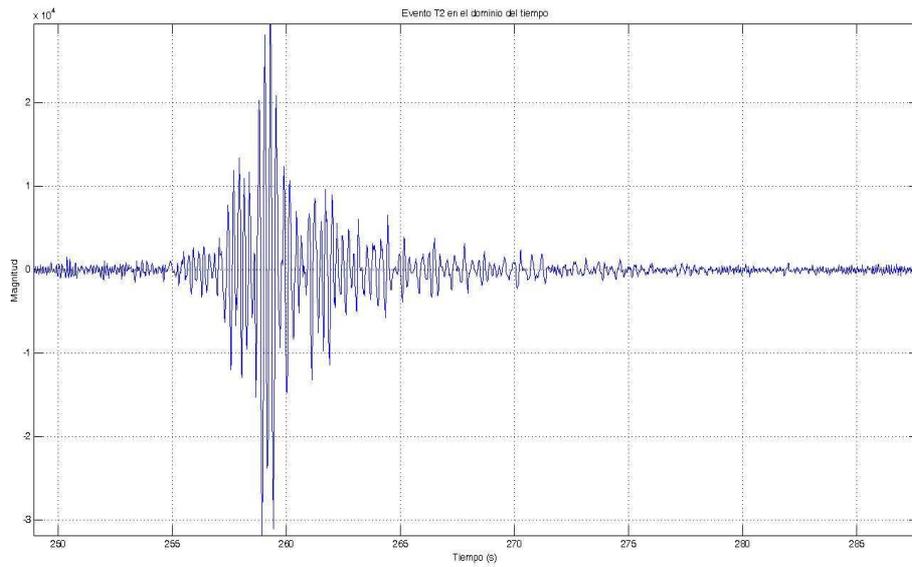


Fig. 4.15 Tipo de evento T2 en el volcán Telica

Algunas características de este tipo de evento se observan en la figura 4.16.

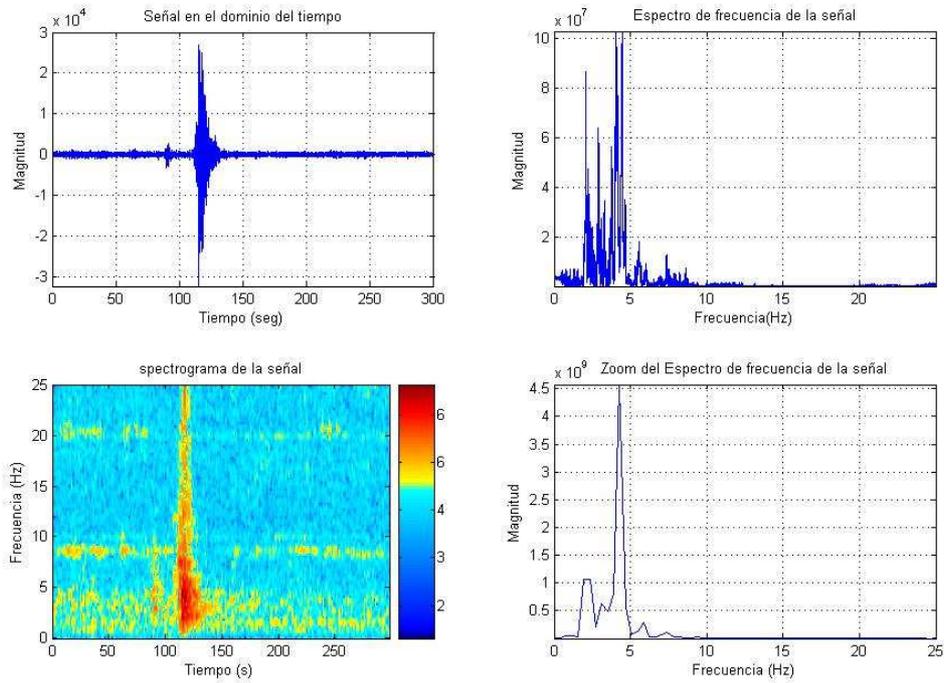


Fig. 4.16 Características del tipo de evento T2 en el volcán Telica

c) Tipo de evento T3: La figura 4.17 muestra un evento de este tipo

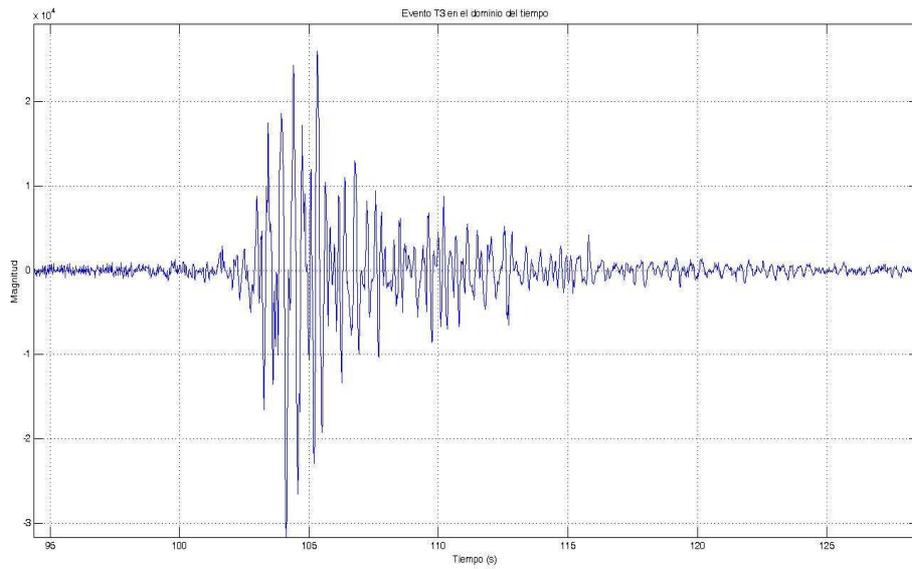


Fig. 4.17 Tipo de evento T3 en el volcán Telica

Algunas características de este tipo de evento se observan en la figura 4.18

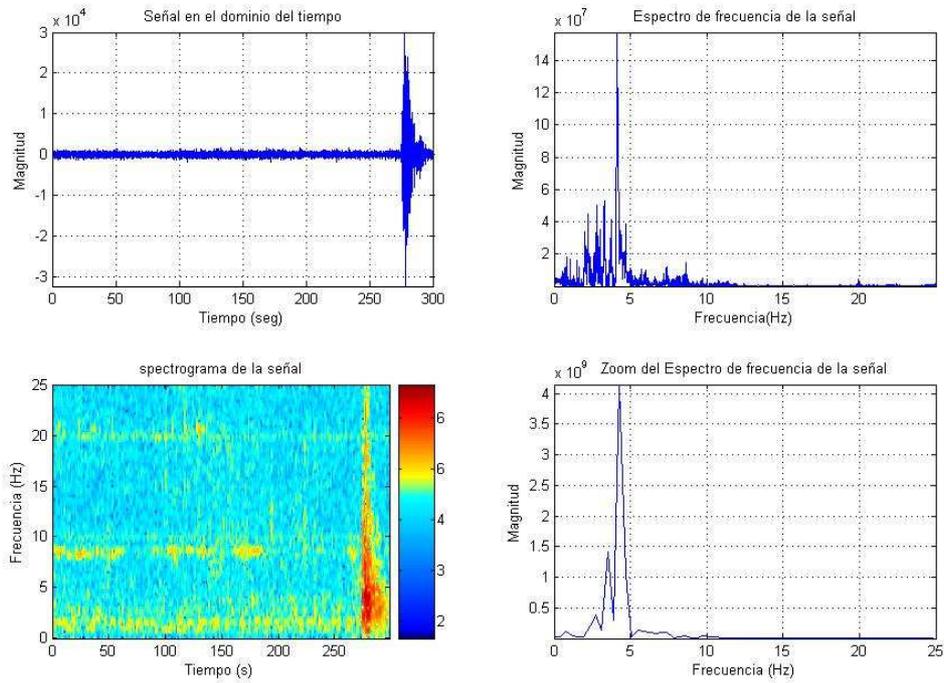


Fig. 4.18 Características del tipo de evento T3 en el volcán Telica.

d) Tipo de Evento T4: La figura 4.19 muestra un evento de este tipo.

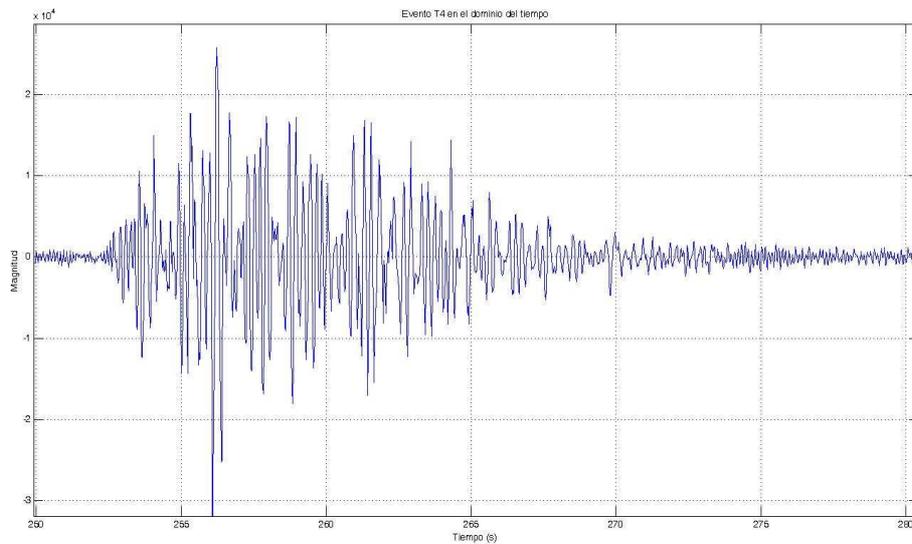


Fig. 4.19 Tipo de evento T4 en el volcán Telica.

Algunas características de este tipo de evento se observan en la figura 4.20.

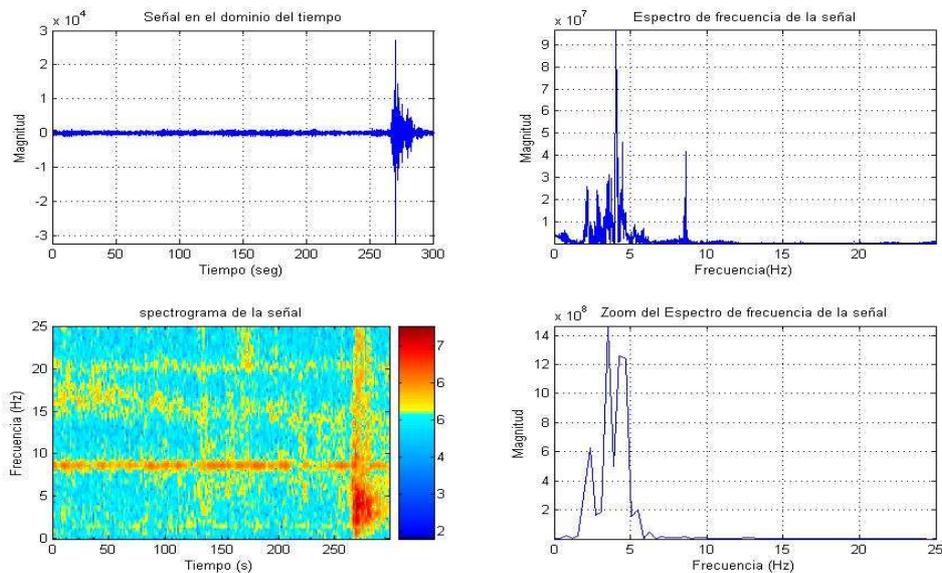


Fig. 4.20 Características del tipo de evento T4 en el volcán Telica.

Algo a destacar en el Telica, es que de acuerdo a las gráficas anteriores, el evento de tipo T1 presenta una primera llegada de energía a baja frecuencia y a continuación una segunda llegada, esto hace que se produzcan dos picos de energías a baja frecuencia característicos de este tipo de señal. Se ha mencionado que a nivel local, los sismólogos del observatorio de Nicaragua lo denominan evento doble, debido a la presencia de estas dos concentraciones de energías, cuando se escucha el sonido que genera este evento, produce un ruido doble seco, similar al latido de un corazón. La señal se concentra en una zona de 2 a 5 Hz. (Ver Fig. 4.12 y 4.13).

4.3.- ANÁLISIS ESPECTRAL DE LOS EVENTOS

Con los diferentes tipos de eventos clasificados en los volcanes se procedió a realizar un análisis mucho más detallado tanto para poder visualizar mejor sus características, como para poder definir tanto la diferencia o semejanza entre ellos. Este proceso es muy importante a la hora de construir las bases de datos para los modelos de entrenamiento y reconocimiento del sistema de clasificación sísmica. Con este fin, se ha contado con el uso de las técnicas de análisis espectrales, basados en las transformadas rápidas de Fourier, para representar los espectros de frecuencia y los espectrogramas de la señal original (En futuros trabajos podría utilizarse además las técnicas de las transformadas Wavelet y la densidad espectral de potencia, para poder definir con mucha más claridad las características de cada tipo al que pertenecen, ver apéndice E). Se han logrado identificar y definir las características de los tipos de eventos que componían las señales en cada volcán. Los resultados conseguidos mediante dichos análisis, aseguran una mejor construcción de las bases de datos en el entrenamiento de los modelos. En las siguientes secciones se van a presentar los análisis espectrales realizados a los tipos de eventos obtenidos.

4.3.1.- Análisis mediante la transformada rápida de Fourier

Se ha indicado que La “*Transformada de Fourier*” es una herramienta que permite transformar una señal expresada en el dominio de tiempo, hacia una expresada en el dominio de frecuencias. Al aplicar la transformada de Fourier a las señales sísmico-volcánicas, será posible ver el contenido espectral que se busca. A continuación en las siguientes figuras, se presentan los tipos de señales en el dominio del tiempo con los espectros de frecuencia, la envolvente y un zoom del espectro para resaltar dichas diferencias. La figura 4.21 muestra dichas características en el evento TB del Etna.

a) Volcán Etna

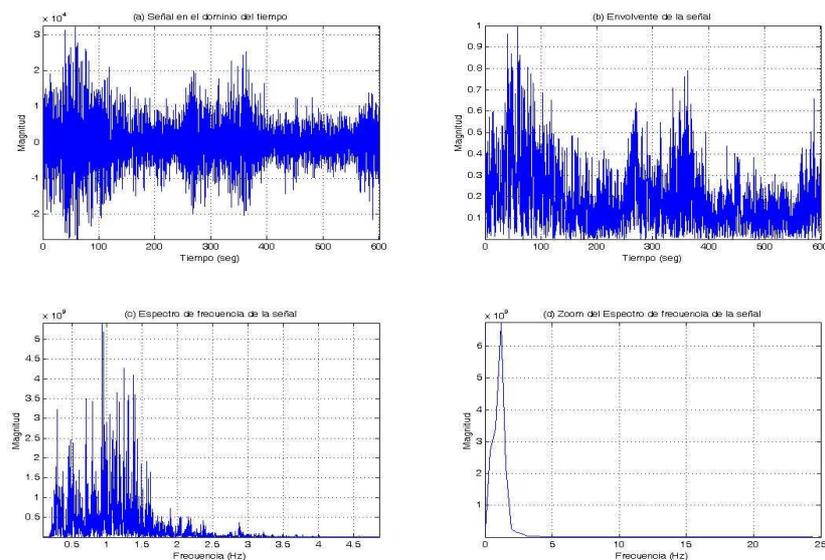


Fig. 4.21 Señal en el Registro 2541132d del Etna (a) En el dominio del tiempo, (b) Envolvente de la señal, (c) espectro de frecuencia de la señal (d) zoom del espectro de frecuencia de la señal

La figura 4.22 presenta las características del evento Exp en el Strómboli.

b) Volcán Strómboli

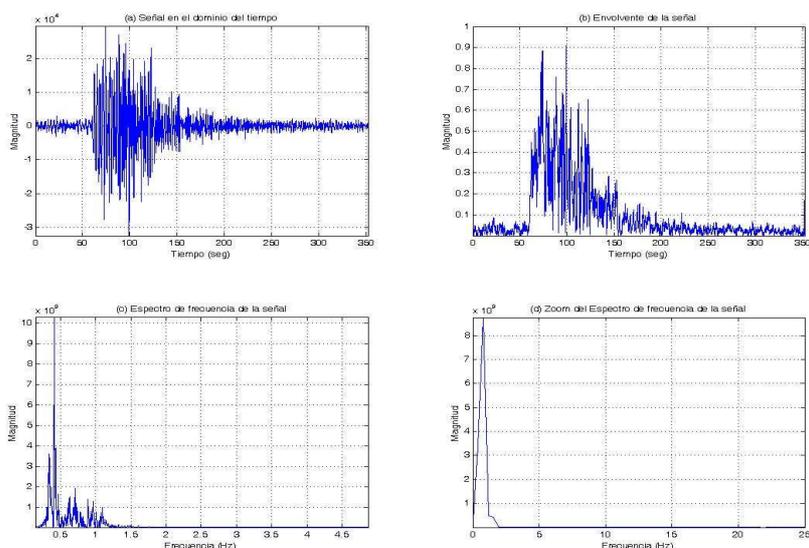


Fig. 4.22 Señal en el Registro 2542227G del Strómboli (a) En el dominio del tiempo, (b) Envolvente de la señal, (c) espectro de frecuencia de la señal (d) zoom del espectro de frecuencia de la señal

La transformada de Fourier que se ha utilizado para definir la señal, su envolvente y el espectro en frecuencia hace definir de forma bastante precisa las características explosiones estrombolianas presentes en los registros de este volcán. Las imágenes 4.23 a 4.25, representan los registros en el San Cristóbal.

c) Volcán San Cristóbal

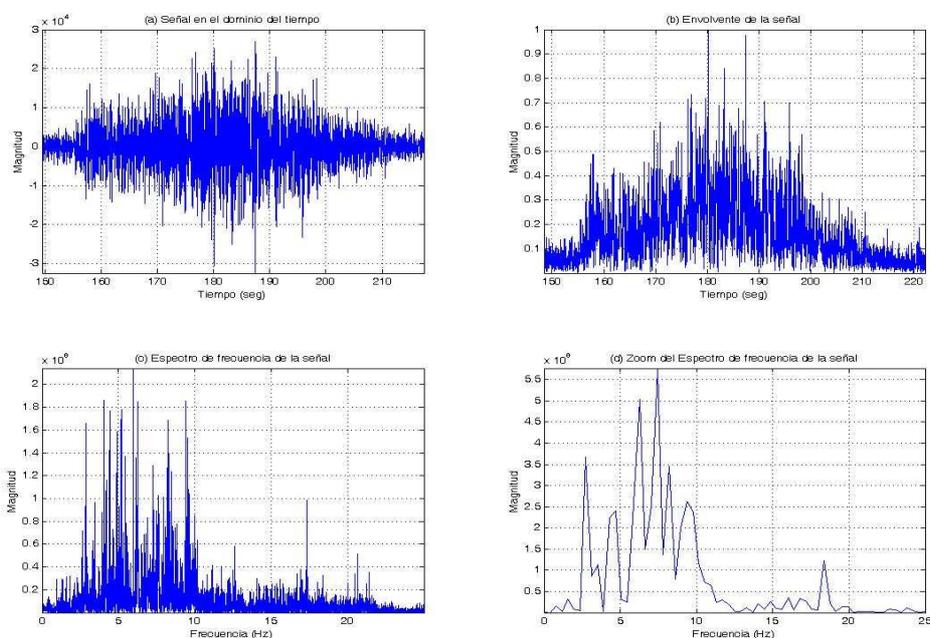


Fig. 4.23 Evento S1, en el Registro del San Cristóbal 2006-02-20-1827-54SC (a) En el dominio del tiempo, (b) Envolvente de la señal, (c) espectro de frecuencia de la señal (d) zoom del espectro de frecuencia de la señal

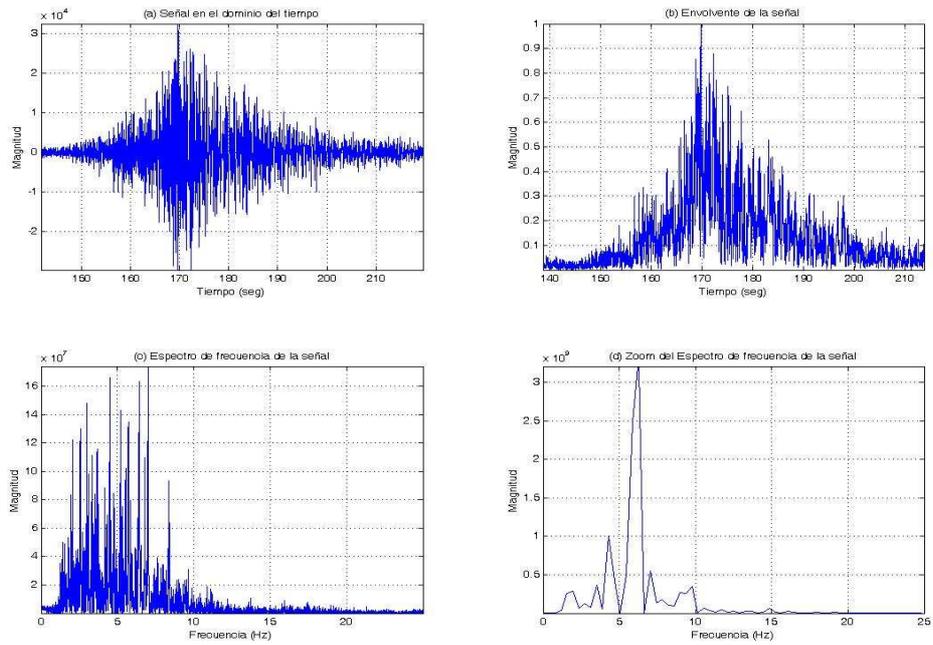


Fig. 4.24 Evento S2, en el Registro 2006-02-18-2018-16SC del San Cristóbal (a) En el dominio del tiempo, (b) Envoltura de la señal, (c) espectro de frecuencia de la señal (d) zoom del espectro de frecuencia de la señal

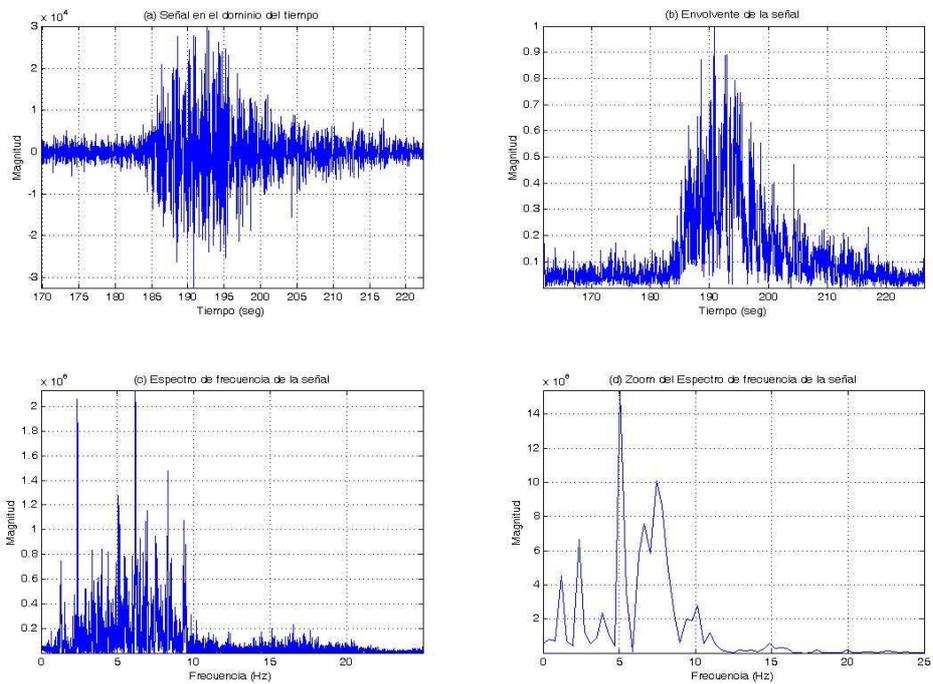


Fig. 4.25 Evento S3, en el Registro 2006-02-18-2253-14SC del San Cristóbal (a) En el dominio del tiempo, (b) Envoltura de la señal, (c) espectro de frecuencia de la señal (d) zoom del espectro de frecuencia de la señal

Mediante la representación gráfica de la transformada de Fourier en las tres imágenes anteriores, puede definirse que en el San Cristóbal, existen al menos tres grupos diferentes de eventos. Las características espectrales de los eventos del Telica se observan en las figuras 4.26 a 4.29.

d) Volcán Telica

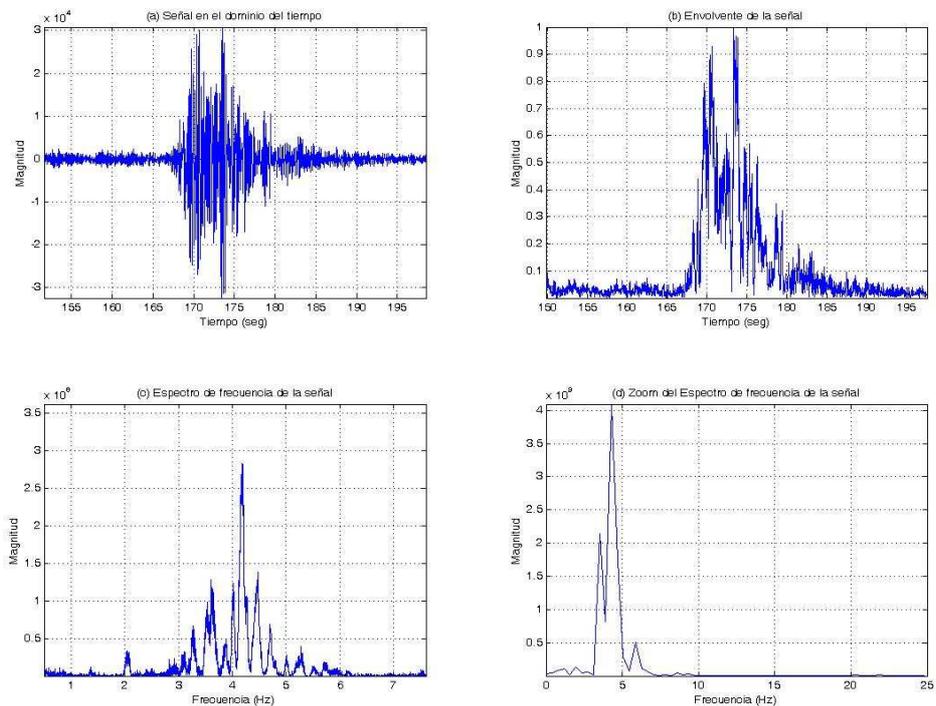


Fig. 4.26 Evento T1, en el Registro 2006-02-17-0048-36ST del Telica (a) En el dominio del tiempo, (b) Envoltura de la señal, (c) espectro de frecuencia de la señal (d) zoom del espectro de frecuencia de la señal.

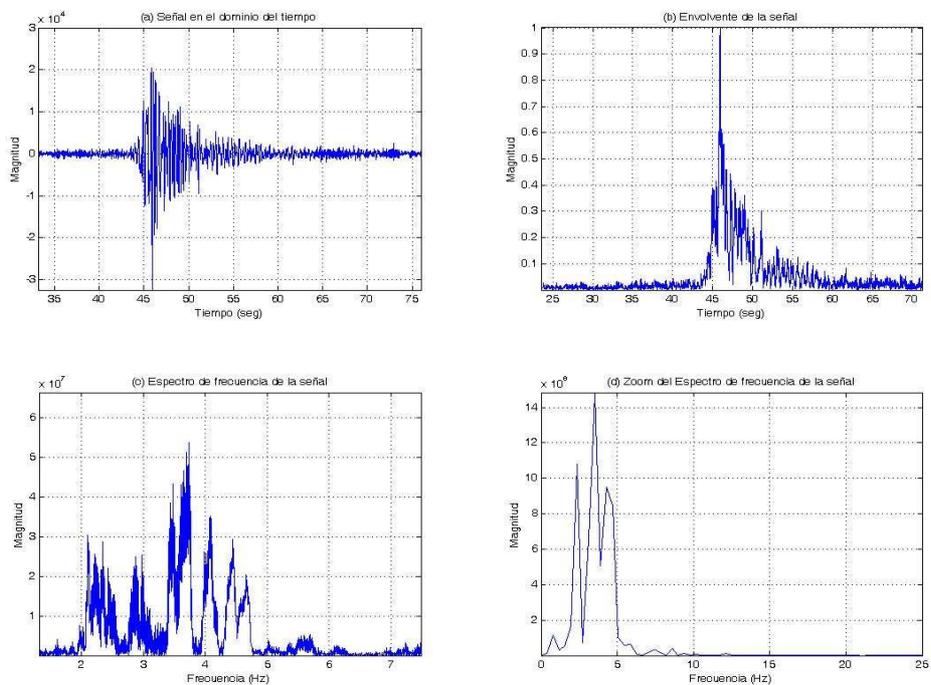


Fig. 4.27 Evento T2, en el Registro 2006-02-18-1048-20ST del Telica (a) En el dominio del tiempo, (b) Envoltura de la señal, (c) espectro de frecuencia de la señal (d) zoom del espectro de frecuencia de la señal.

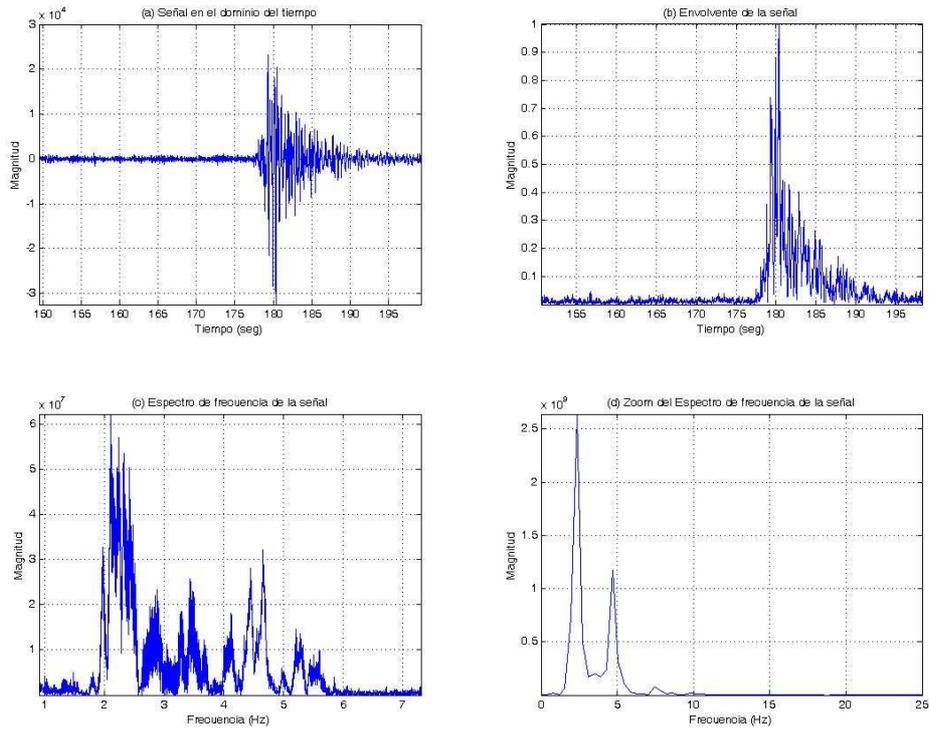


Fig. 4.28 Evento T3, en el Registro 2006-03-07-1940-04ST (a) En el dominio del tiempo, (b) Envoltura de la señal, (c) espectro de frecuencia de la señal (d) zoom del espectro de frecuencia de la señal.

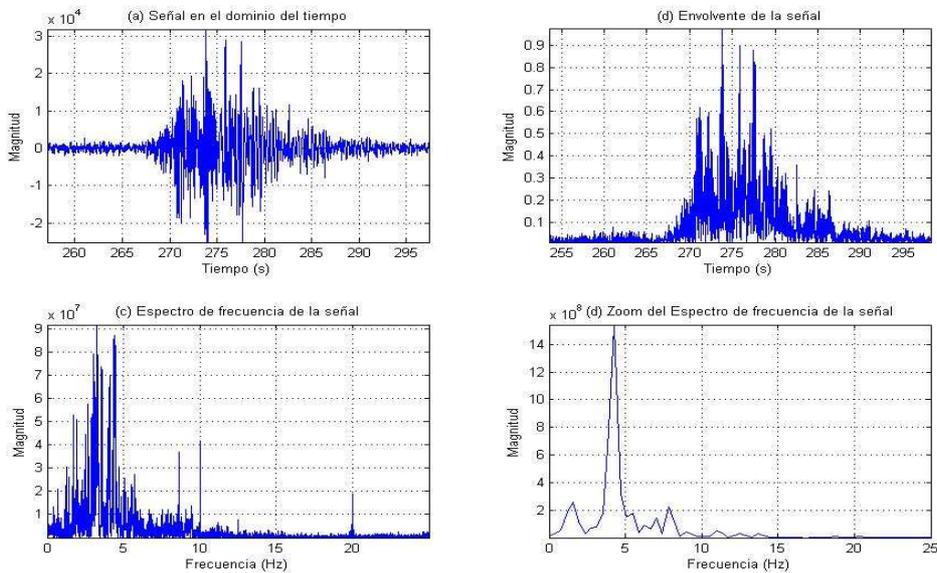


Fig. 4.29 Evento T4, en el Registro 2006-02-17-0633-34ST (a) En el dominio del tiempo, (b) Envoltura de la señal, (c) espectro de frecuencia de la señal (d) zoom del espectro de frecuencia de la señal.

La transformada de Fourier de las cuatro imágenes anteriores en el Telica indica la presencia de al menos cuatro grupos diferentes de eventos en los registros. Si se observa con detalle las gráficas anteriores de los cuatro volcanes, puede deducirse que

todos los eventos en su mayoría, en el espectro de frecuencias están cerca del orden de 5 Hz. Los de Italia incluso bajan de los 3 Hz. En el Etna se encuentran entre 0.5 y 2 Hz, (Fig 4.21) y en el Strómboli entre 0.5 y 1 Hz. (Fig. 4.22). En Nicaragua, entre los eventos se presentan significativos cambios respecto de unos con otros, esto se observa con más detalle en los espectros de frecuencia de cada evento. Por ejemplo, al visualizar los espectros de frecuencia de cada evento del San Cristóbal, el evento S1 de la figura 4.23 presenta un pico de mayor magnitud cercano a 5 Hz. y es mayor a lo largo del tiempo que el evento S2 en la figura 4.24 y el S3 de la figura 4.25. En cambio, en los eventos S1 y S2 de las figuras 4.23 y 4.24, estos presentan varios saltos, siendo mucho más evidentes en el evento S3 de la figura 4.25, que en el S2 de la figura 4.24. El evento S3 en la figura 4.25 es el que menor duración posee a lo largo del tiempo y además contiene un mayor potencial en contenido de frecuencia, pudiendo tratarse de un evento de tipo explosivo.

En el volcán Telica, el evento T1 de la figura 4.26, (llamado evento doble en Nicaragua), presenta una llegada característica de dos picos de energía, cercanos a 5 Hz., pudiendo tratarse de un terremoto de tipo VT. En cambio, el evento T2 de la figura 4.27, presenta un pico muy delineado, pudiendo catalogarse como un evento de tipo explosivo. Los eventos T3 y T4 de las figuras 4.28 y 4.29, presentan picos también cercanos a 5 Hz. con un mayor énfasis en contenido frecuencial en el T4 de la figura 4.29, respecto al T1 de la figura 4.26 y mucho mayor respecto a los eventos T2 y T3 de las figuras 4.27 y 4.28. Siendo, posiblemente el T4, un evento de tipo LP. El de tipo T3, presenta al inicio del evento una caída en baja frecuencia en la llegada de energía (en forma de arco inferior), es el de más corta duración, pudiendo catalogarse como un terremoto. Este tipo de nomenclatura aún es muy prematura, ya que se necesita el consenso de los sismólogos locales. Sin embargo, estos análisis son útiles para poder distinguir los diferentes tipos de eventos presentes en los registros. Aún así, hay que definir todavía más las características espectrales, por lo que se aplican los análisis mediante los espectrogramas que se muestran a continuación.

4.3.2.- Análisis de los espectrogramas de los eventos

En la sección 2.2.2 se mencionó que el espectrograma es el registro gráfico que se obtiene al calcular mediante la transformada de Fourier el espectro de las tramas eventanadas de una señal. La resultante es una gráfica tridimensional que representa la energía del contenido frecuencial de la señal, según va variando ésta a lo largo del tiempo. Las líneas brillantes que se forman en la parte inferior son los fundamentales armónicos de la señal, en conjunto, forman el espectro de frecuencias. Una vez que se ha conseguido el espectrograma se tiene la herramienta adecuada para analizar la variación espectral de las señales. Gracias a esta técnica se ha observado cómo las frecuencias de las señales sísmico-volcánicas tienen fuertes variaciones que permiten distinguir unos u otros tipos de eventos. Los espectrogramas de los eventos del presente estudio, utilizan un número de 256 puntos de la FFT. Para los cálculos se ha utilizado una ventana de tamaño 128 con una frecuencia de muestreo de 50 Hz. = 2.56 seg.

De esta forma, puede obtenerse un espectrograma bastante claro para poder observar la forma y las energías contenidas en la señal que conforma el evento. En esta sección, los análisis en la construcción de las gráficas se han desarrollado mediante tres registros similares, que contienen eventos del mismo tipo, y así apreciar de forma mucho más definida las características, semejanzas y diferencias de cada uno de los

eventos. Las siguientes imágenes que se presentan a continuación, definen los espectrogramas de eventos presentes en los registros. Se han marcado mediante flechas, tanto los máximos de las frecuencias que se observan en cada evento, así como algunas bandas de ruido (sobre todo en el caso de Nicaragua), que están presentes en los registros (debido ya sea al equipo sensor o a una interferencia por la proximidad de las antenas retransmisoras). Las siguientes figuras muestran los espectrogramas de los registros en los volcanes de estudio.

a) Volcán Etna

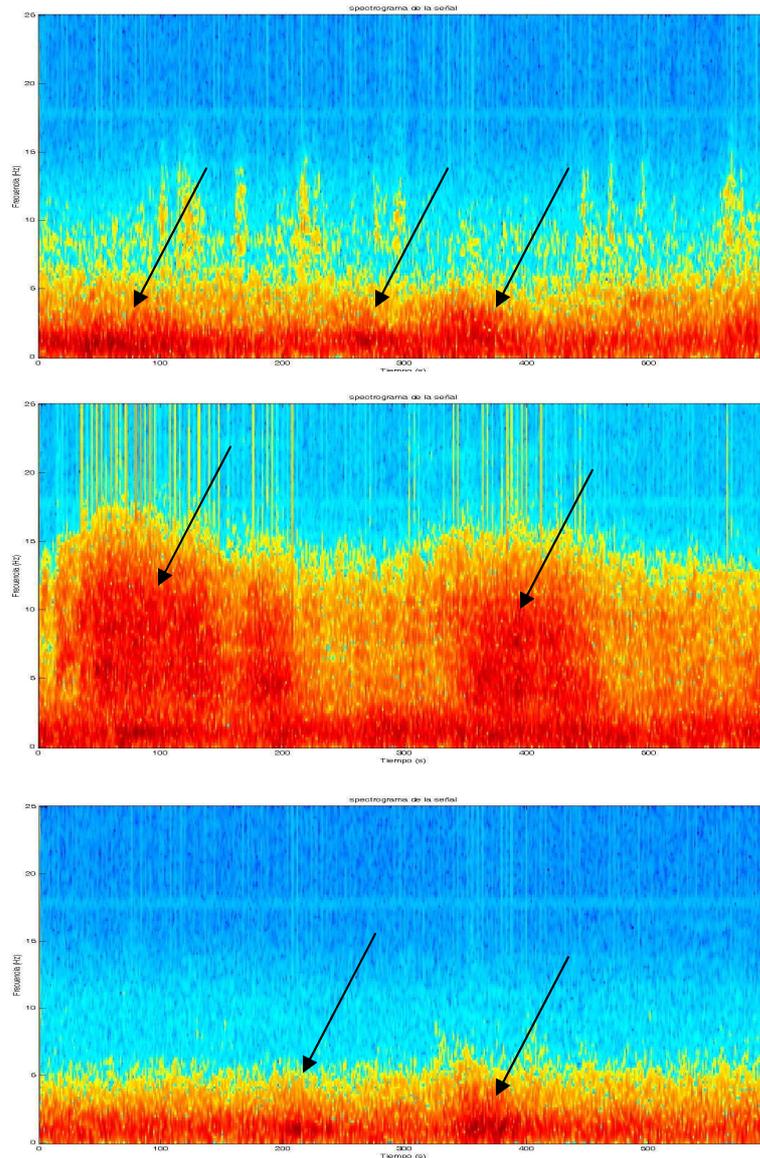


Fig. 4.30 Espectrograma correspondiente a tres registros con eventos de tipo tremor burst con una forma característica entre sí en el Etna. Las flechas indican la presencia de los eventos en el tremor de fondo.

En el tremor del Etna (Fig. 4.30) se pueden observar (dentro del mismo tremor), dos eventos que son los catalogados por Saccorotti como los “Burst” o burbujas, mucho mejor definidos en la figura central. En la mayoría de los registros, los eventos se

encuentran inmersos dentro del mismo tremor de fondo. Este tipo de evento se da a bajas frecuencias, la frecuencia fundamental se encuentra entre los 1 a 10 Hz.

b) Volcán Strómboli

En el caso del Strómboli, el evento está bien definido, son las características explosiones estrombolianas, los espectrogramas que se muestran en la siguiente figura pueden dar una idea sobre las características espectrales de este tipo de evento.

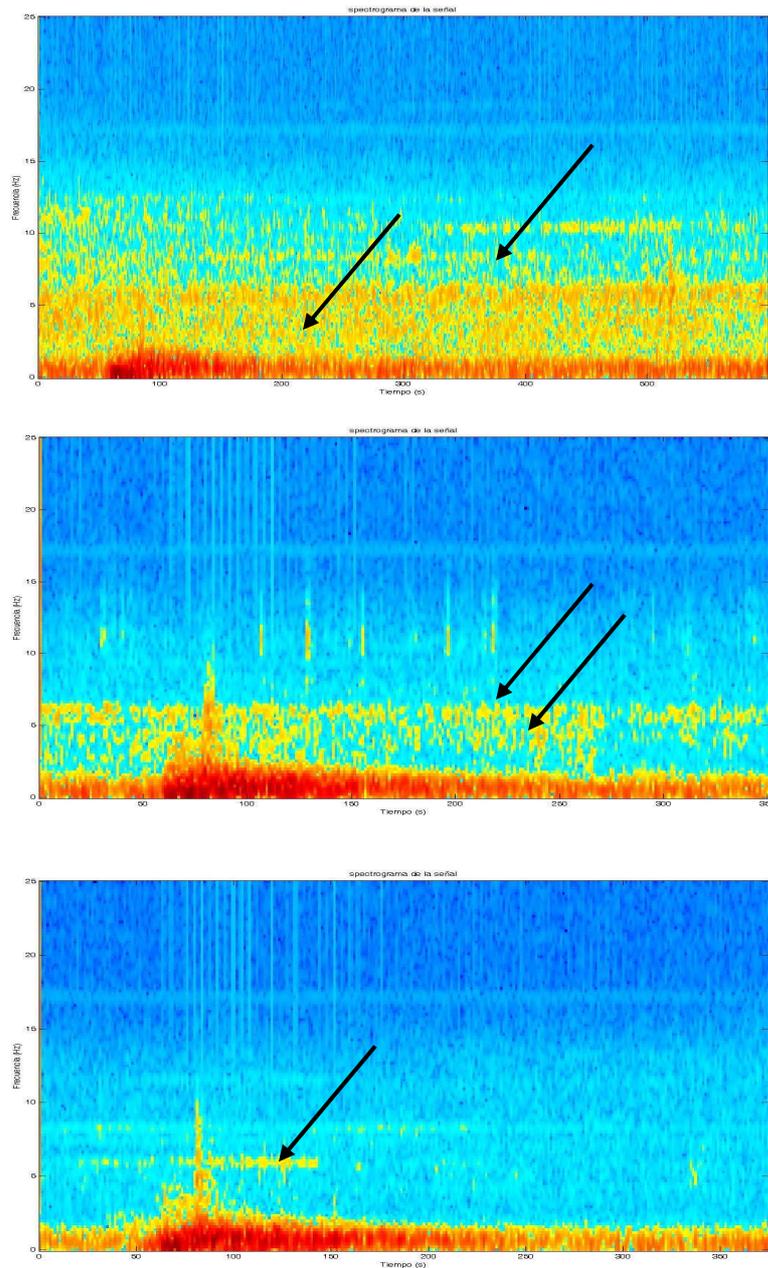


Fig. 4.31 Espectrograma correspondiente a registros con eventos de tipo explosión con una forma característica entre sí en el Strómboli. Las flechas indican la presencia de bandas de ruido, aproximadamente sobre los 7 y 10 Hz.

Las explosiones del Strómboli (Fig. 4.12) presentan un alargamiento en frecuencia a lo largo del tiempo con una llegada en baja frecuencia. Se observan bandas de ruido presentes en los registros (señalizadas mediante flechas). Las explosiones del Strómboli se dan a baja frecuencia como se observa en las gráficas, situándose en una banda de 1 a 5 Hz.

c) Volcán San Cristóbal

En los registros del volcán San Cristóbal, los espectrogramas pueden ayudar a definir los diferentes tipos de eventos que se producen. En este volcán se han determinado tres grupos principales de eventos con similares características espectrales. La mayor concentración de energía de los eventos se encuentra en un intervalo de 2 Hz a 10 Hz. La figura 4.32 presenta el primer grupo de eventos de tipo S1.

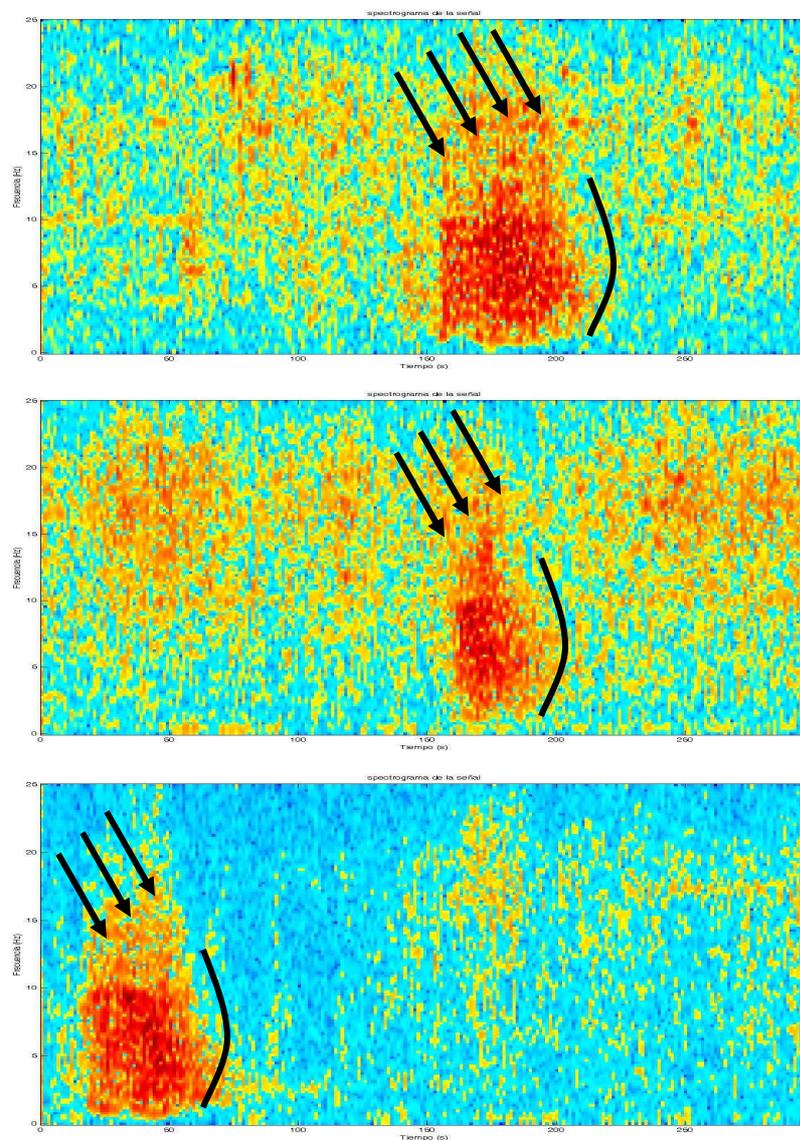


Fig. 4.32 Primer grupo de espectrogramas, que corresponden a tres eventos S1 con una forma característica entre sí en el San Cristóbal. Las flechas indican la presencia de múltiples llegadas en este tipo de evento. La curva indica una caída suave en frecuencia característica de este tipo de evento.

En las imágenes de la figura 4.32 se observan múltiples llegadas y una forma característica en la caída de energía en el espectro de este grupo de eventos S1, el cual se encuentra presente dentro del tremor de fondo del San Cristóbal. Un segundo grupo que representa a los eventos S2 con características diferentes se observa en la figura 4.33.

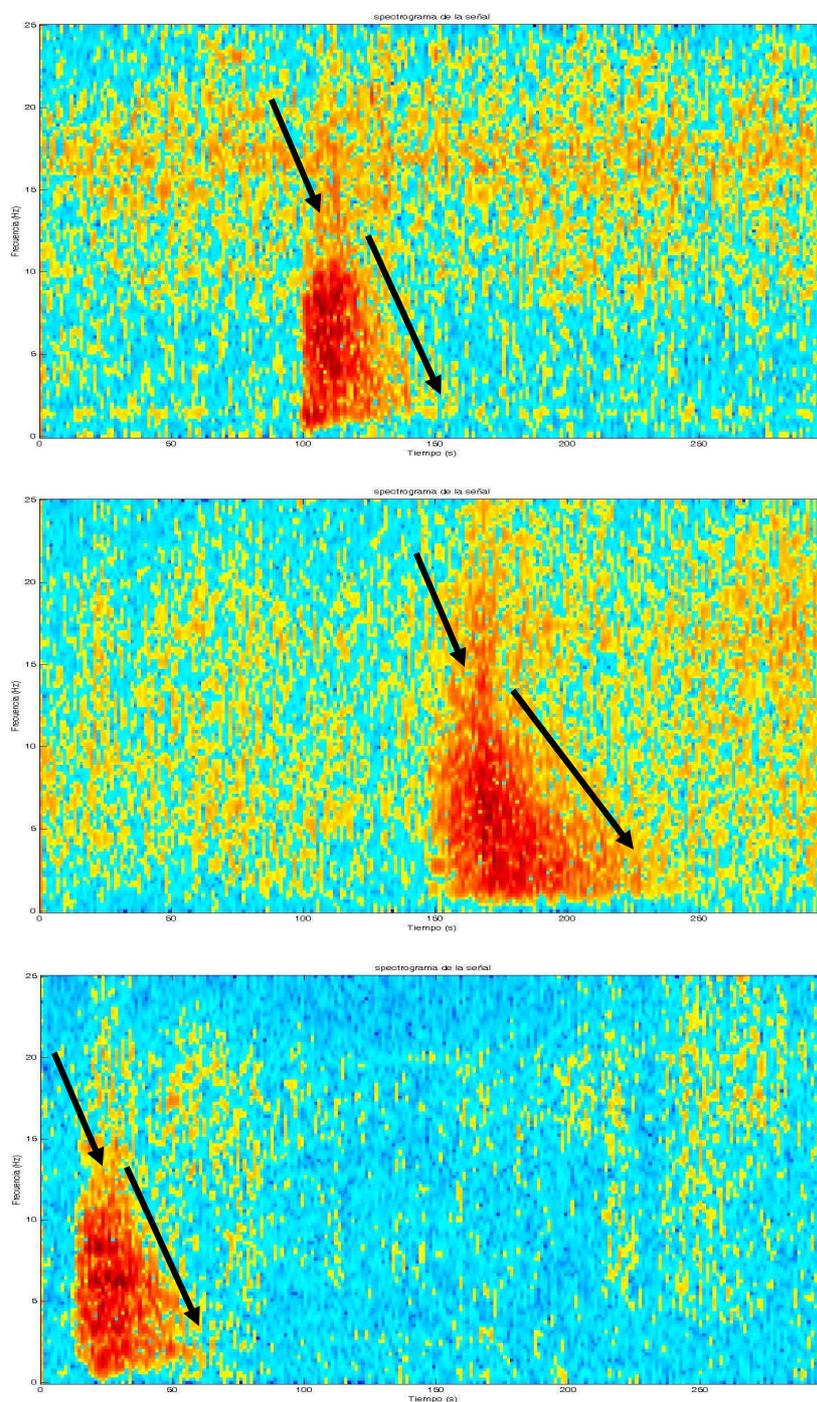


Fig. 4.33 Segundo grupo de espectrogramas, con tres registros con eventos de tipo S2, que poseen una misma forma característica, en el San Cristóbal. La flecha superior indica una sola llegada en frecuencia a diferencia del anterior evento. La flecha inferior indica la forma de la caída en frecuencia de este tipo de evento.

Un análisis del espectrograma de este segundo grupo de eventos, presenta una caída en frecuencia, que se observa en las gráficas mediante la flecha vertical y a diferencia del anterior, presenta un máximo con un pico predominante en frecuencia, en lugar de los múltiples picos que se observaban en el primero evento de tipo S1. La imagen 4.34 presenta el tercer grupo de tres registros, donde se observan eventos de tipo S3.

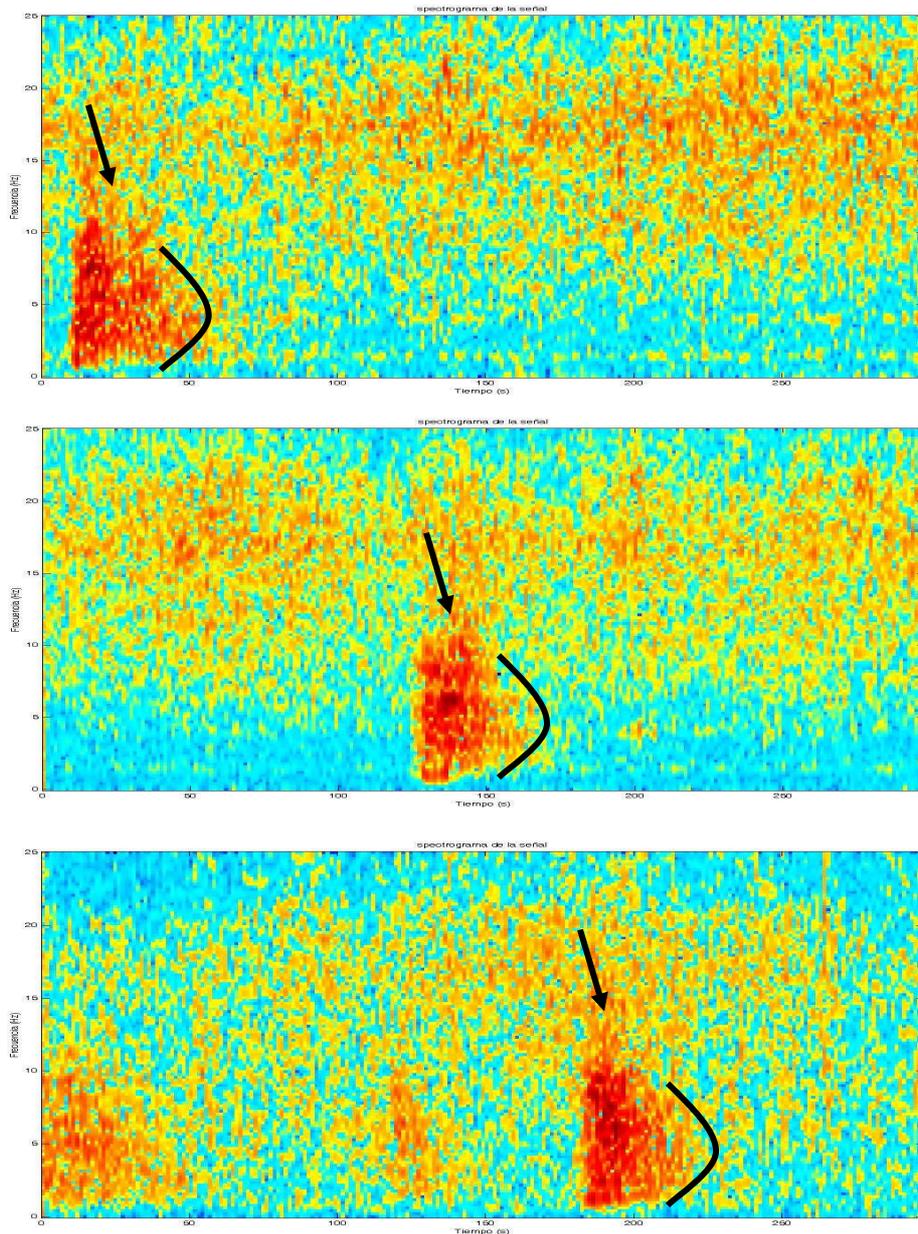


Fig. 4.34 Tercer grupo de espectrogramas, que corresponden a tres registros con eventos de tipo S3, que poseen una misma forma característica a lo largo del tiempo, en el San Cristóbal. La flecha indica una sola llegada similar al evento anterior. Sin embargo, la curva indica que la forma de caída en frecuencia en este tipo de evento es más similar al primer tipo, el S1 en comparación al S2.

Los espectrogramas de este grupo, correspondiente al S3 al igual que el primer grupo del S1, presentan una caída suave en frecuencia con una forma de campana, sin embargo, a diferencia del S1, el S3 presenta un pico predominante en frecuencia en lugar de múltiples picos como presentaba el S1. Por otra parte, el grupo de eventos del S2, al igual que el S1, presenta un inicio brusco en frecuencia, pero con una caída en frecuencia al final del evento, que es diferente a la del primer grupo (S1). Las bandas de frecuencias de los eventos S1 y S2 se encuentran entre 1 a 15 Hz, en cambio el evento S3, es un poco menor situándose en una banda de 1 a 10 Hz. En el tercer grupo del S3, los eventos presentan una menor duración en contenido de frecuencia a lo largo del tiempo. El caso del espectrograma del tremor continuo en el San Cristóbal se observa en la figura 4.35

d) Caso Tremor continuo en el San Cristóbal

En el tremor continuo o armónico del San Cristóbal (fig. 4.35), se pueden apreciar al menos cuatro bandas fundamentales presentes en la señal, las cuales están espaciadas en el tiempo secuencialmente. Dichas bandas fundamentales corresponden aproximadamente a las frecuencia de 2, 4, 6, y 8 Hz. Además se observan dos bandas de ruido a 10 Hz y 20 Hz., y componentes de ruido en alta frecuencia.

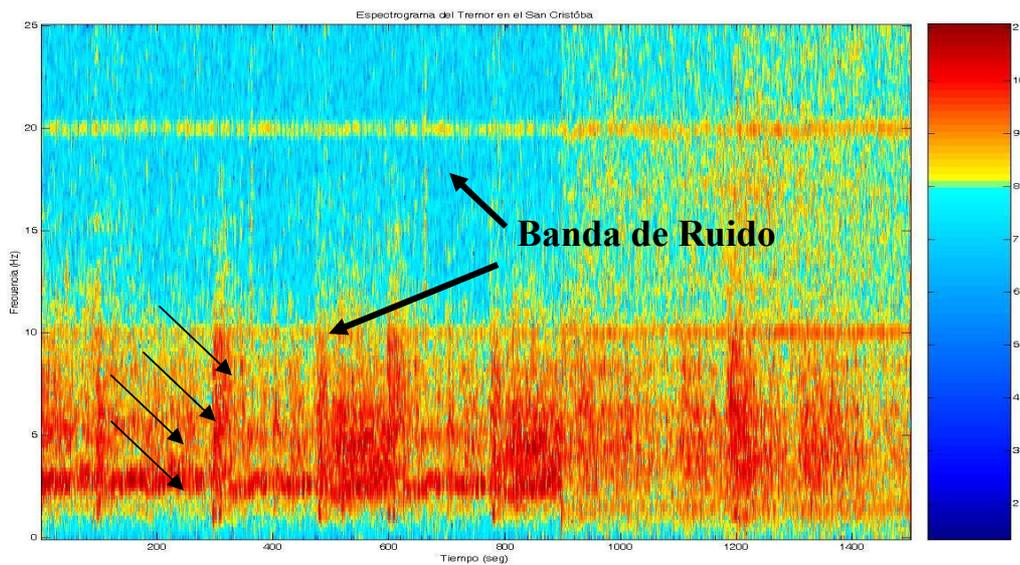


Fig. 4.35 Espectrograma del tremor continuo del San Cristóbal. Las flechas pequeñas a la izquierda indican las bandas de frecuencia presentes en el tremor. Al menos se identifican cuatro armónicos que componen el tremor. Se observan dos bandas de ruido a 10 Hz y a 20 Hz., inducido por los equipos y la antena receptora

Como se ha indicado, en los espectrogramas se puede obtener información que ayude a comprender las características espectrales de los tipos de señales sísmico-volcánicas. Los tres tipos de eventos del San Cristóbal, que se observaron en las figuras 4.32 a 4.34, muestran diferentes formas características. El evento S1, presenta una llegada inicial brusca en frecuencia, con múltiples picos en frecuencia y con una caída suave en forma de campana a lo largo del tiempo. El S2 en cambio muestra una sola llegada con una caída más brusca en frecuencia. El S3 muestra similitud con el S2 al presentar una sola llegada en frecuencia, sin embargo, en relación a su caída en

frecuencia muestra más similitud al S1. A continuación se verán los espectrogramas del Telica y como los eventos de este volcán, difieren entre los encontrados en los anteriores volcanes: San Cristóbal, Etna y Strómboli.

e) Volcán Telica

Mediante los espectrogramas de los eventos pueden diferenciarse cuatro diferentes grupos principales con similares características espectrales en los registros sísmico-volcánicos. La mayor concentración de energía de los eventos se encuentra en un intervalo de 2 Hz a 6 Hz. El primer grupo de tres registros con eventos de tipo T1 se observa en la imagen 4.36.

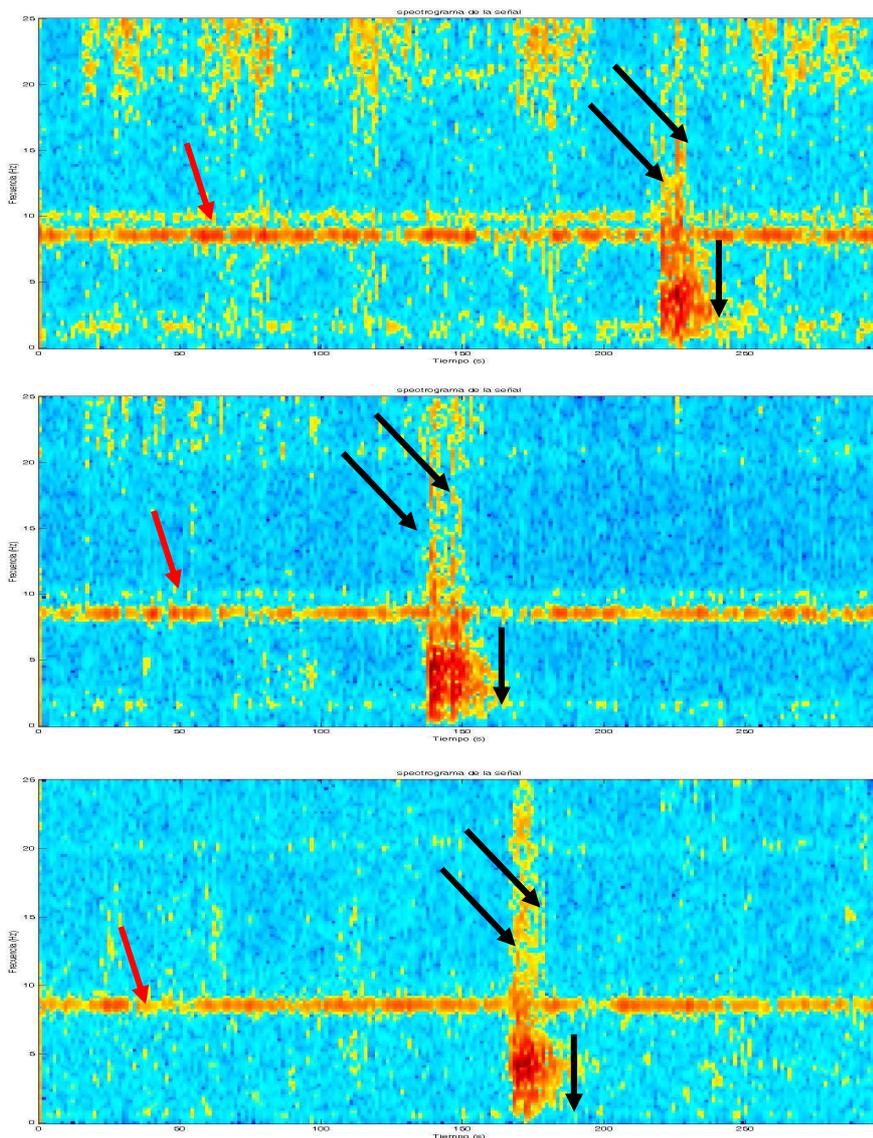


Fig. 4.36 Primer grupo de espectrogramas, que corresponden a registros con tres eventos T1, que poseen una misma forma característica a lo largo del tiempo, en el Telica Las flechas negras inclinadas indican las dos llegadas en frecuencia, por lo que este tipo de evento es catalogado como evento doble. La flecha vertical indica la caída en frecuencia en un intervalo de duración corto a lo largo del tiempo en este tipo de evento. La flecha roja indica la presencia de una banda de ruido a 8 Hz.

En la gráfica anterior se observan las formas características de este tipo de evento, en el que se presentan dos típicas bandas de energía o llegadas de la señal (los eventos dobles de acuerdo con el criterio de los sismólogos locales). También se observan las bandas horizontales (8 Hz) que corresponden al ruido de tipo externo (ya sea ocasionado por la presencia de las antenas en la estación o por el sensor), dicho ruido se representa en la imagen mediante las flechas de color rojo. La caída en energía en este grupo es bastante brusca. La frecuencia de este tipo de evento se encuentra en una banda de 2 a 8 Hz. La imagen 4.37 muestra el segundo grupo de tres registros con características espectrales similares que corresponden a eventos de tipo T2.

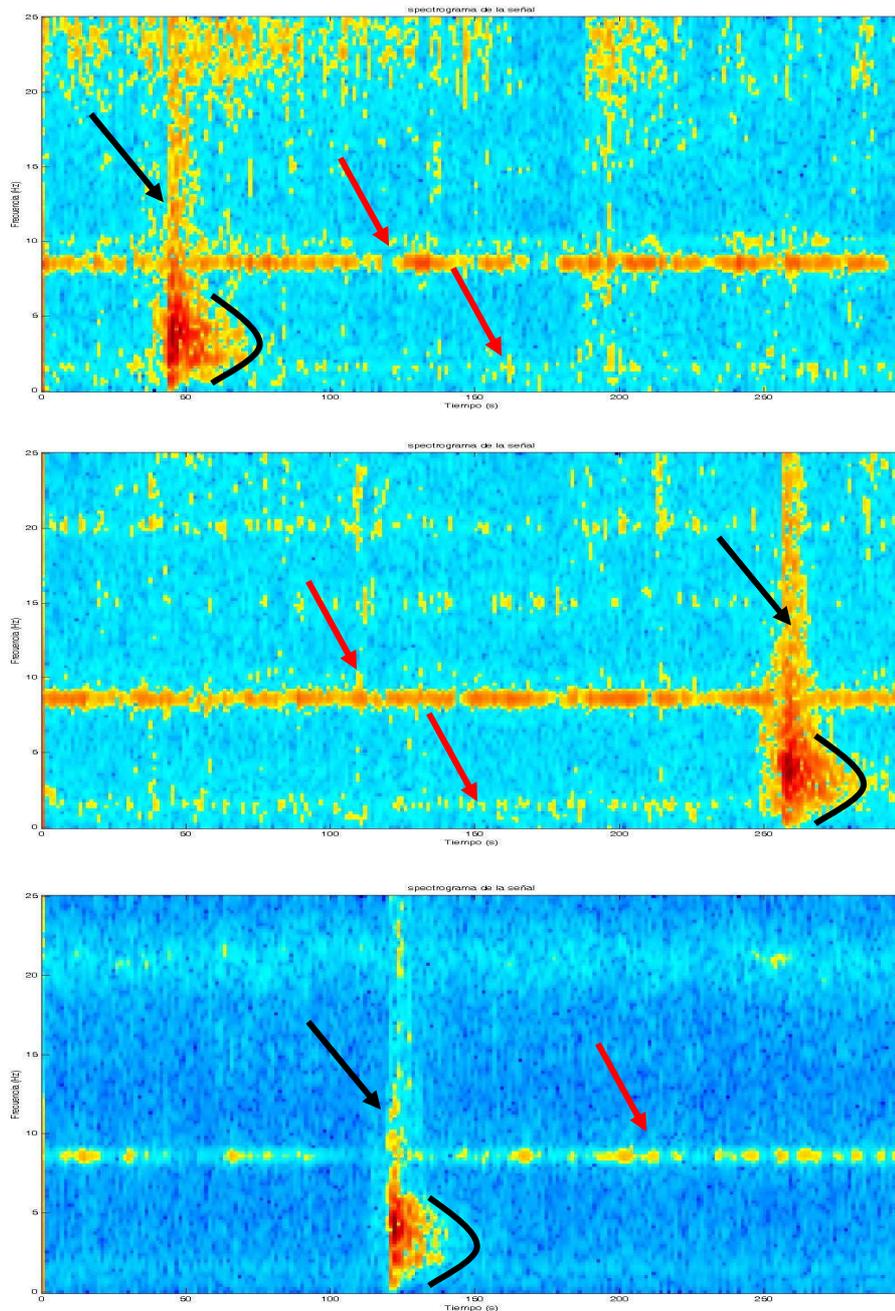


Fig. 4.37 Segundo grupo de espectrogramas, que corresponden a registros con tres eventos T2, que poseen una misma forma característica a lo largo del tiempo, en el Telica. La flecha negra inclinada muestra una única llegada en frecuencia a diferencia del evento anterior. La curva indica la caída suave en frecuencia diferente al evento anterior. Las flechas rojas indican bandas de ruido en torno a los 8 y 3 Hz.

Una componente de llegada en alta energía, con una caída suave en forma de campana, es lo que se observa en los espectrogramas de los eventos T2 en este segundo grupo. Las flechas rojas indican las bandas de ruido presentes alrededor de los 8 Hz. La frecuencia del evento T2 se encuentra en una banda de 2 a 6 Hz. La imagen 4.38 define al tercer grupo de tres registros con eventos de tipo T3.

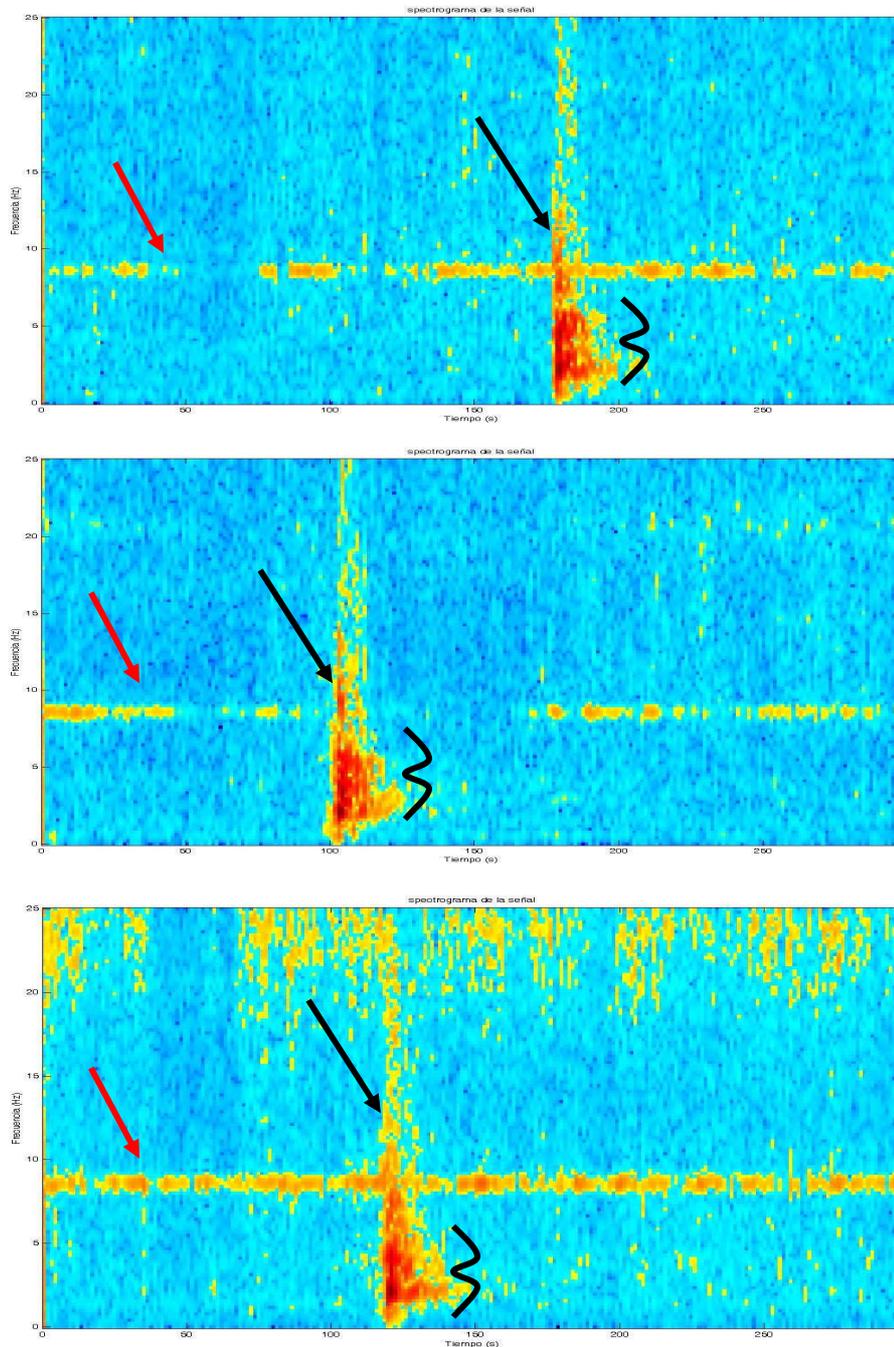


Fig. 4.38 Tercer grupo de espectrogramas, que corresponden a tres registros con eventos de tipo T3, que poseen una misma forma característica a lo largo del tiempo, en el Telica. La flecha negra inclinada indica una única llegada en frecuencia para este tipo de evento. La línea ondulada indica la forma de la caída en frecuencia en este tipo de evento. La flecha roja indica la presencia de una banda de ruido a los 8 Hz.

Al observar los espectrogramas de este tercer grupo, puede definirse al igual que el anterior, en una llegada en alta energía pero con una caída en frecuencia a lo largo del tiempo, diferente respecto a los dos anteriores (forma de un 3 en la figura). La figura 4.39 muestra los espectrogramas correspondientes al tipo de evento T4.

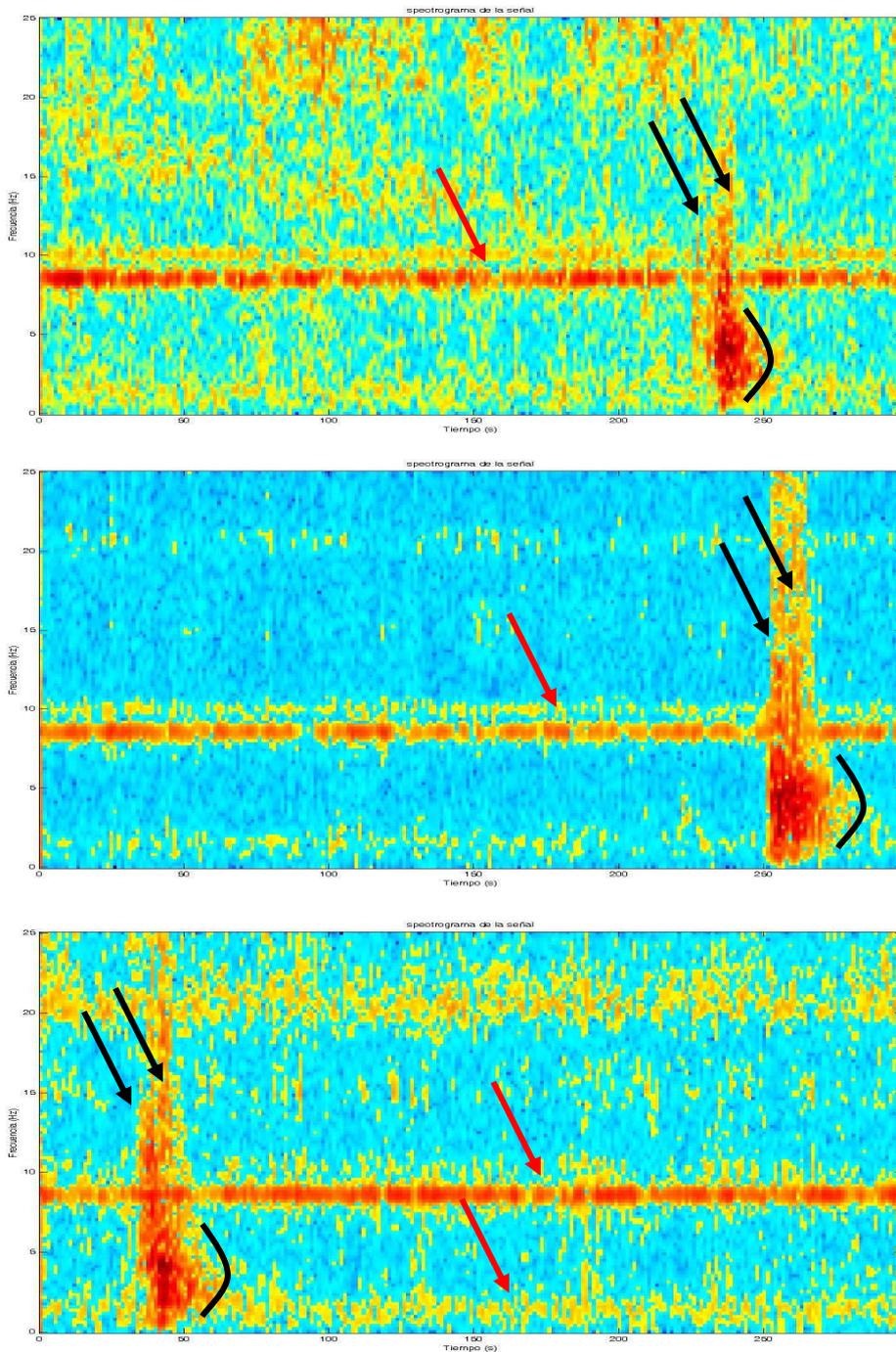


Fig. 4.39 Cuarto grupo de espectrogramas, que corresponden a tres registros con eventos de tipo T4, que poseen una misma forma característica a lo largo del tiempo, en el Telica. Las flechas inclinadas indican la presencia de más de una llegada en frecuencia o múltiples picos en frecuencia para este tipo de evento. La curva indica la caída suave en frecuencia y las flechas rojas la presencia de bandas de ruido.

En este último grupo de eventos de la figura 4.39 anterior, los espectrogramas muestran dos o más componentes en altas energías (con varios picos en frecuencias a lo largo del tiempo) y una caída suave en energía al final del evento (con forma de campana, similar al T2). Al observar los espectrogramas anteriores, un rápido análisis indica que puede notarse una similitud entre las formas de los espectros del segundo grupo de eventos respecto al tercero (T2 y T3). La duración en frecuencia de los eventos T2 y T3 en promedio es muy cercana. De la misma forma, puede observarse algunas características similares entre los espectrogramas del primero y cuarto grupo (T1 y T4), siendo estos eventos en promedio de mayor duración en frecuencia a lo largo del tiempo. De los espectrogramas anteriores se consigue obtener información que ayude a comprender las características espectrales de las señales sísmico-volcánicas. La energía en estos eventos se concentra en bandas de 2 a 7 Hz. Cada evento muestra una forma típica, la cual puede ayudar a identificar el tipo al que pertenece. En las figuras se observan los saltos de frecuencias que presentan la forma en los espectros, en algunos casos múltiples como los eventos del primer y cuarto grupo (T1 y T4), siendo el primer grupo el que posee dos llegadas al inicio (eventos dobles de acuerdo a los sismólogos locales en Nicaragua), con un corte brusco de energía al final del registro a diferencia del cuarto grupo que se prolonga más en el tiempo. De manera similar, en los espectros del segundo y cuarto grupo, se observa una forma de campana, que denota una suave caída en frecuencia a lo largo del tiempo. Tanto el análisis de la transformada de Fourier como el de los espectrogramas, han servido de ayuda en una previa clasificación de los diferentes tipos de eventos. De todo lo anterior y teniendo en cuenta la clasificación general de Minakami [Minakami 1960, 1974], se podría reflexionar (a manera de una primera aproximación de una clasificación general), de que en Nicaragua se presentan eventos de tipo A y B en los volcanes San Cristóbal y Telica. Con respecto a la clasificación de Ibáñez [Ibáñez 1997], podría decirse que se está ante eventos de tipo LP, VT, además del tremor y explosiones típicas en ambos volcanes. La clasificación definitiva de dichos eventos sería un tema a tratar en futuros trabajos. Con los tipos de eventos ya definidos, se construirán las bases de datos necesarias en la aplicación de la metodología, con las que se desarrollarán los modelos de entrenamiento y el reconocimiento de las señales sísmico-volcánicas.

4.4.- ESTADÍSTICAS DE LA DURACIÓN DE LOS EVENTOS

Una vez que se tienen definidos los tipos de eventos a trabajar, resulta importante poder interpretar la representación estadística de las señales sísmico-volcánicas presentes en los volcanes de estudio. Esto refleja la complejidad de las mismas a la hora de poder establecer los modelos de entrenamiento, indica el que un mismo tipo de evento es muy variable en su duración, pues va desde pocos a varios segundos. En el presente estudio, se ha procedido a realizar estadísticas para ver la duración de los eventos en cada volcán. Para los volcanes de Italia, la duración de los eventos se observa en las figuras 4.40 y 4.41.

4.4.1.- Estadísticas en el volcán Etna

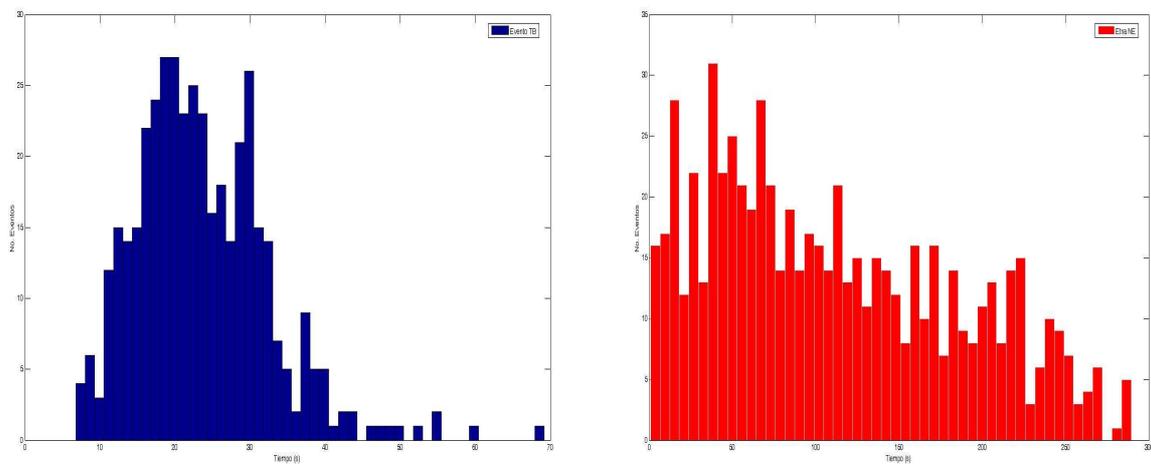


Fig. 4.40 Histograma de duración del evento “TB” (izquierda) y el ruido “NE” (derecha) en el Etna

4.4.2.- Estadísticas en el volcán Strómboli

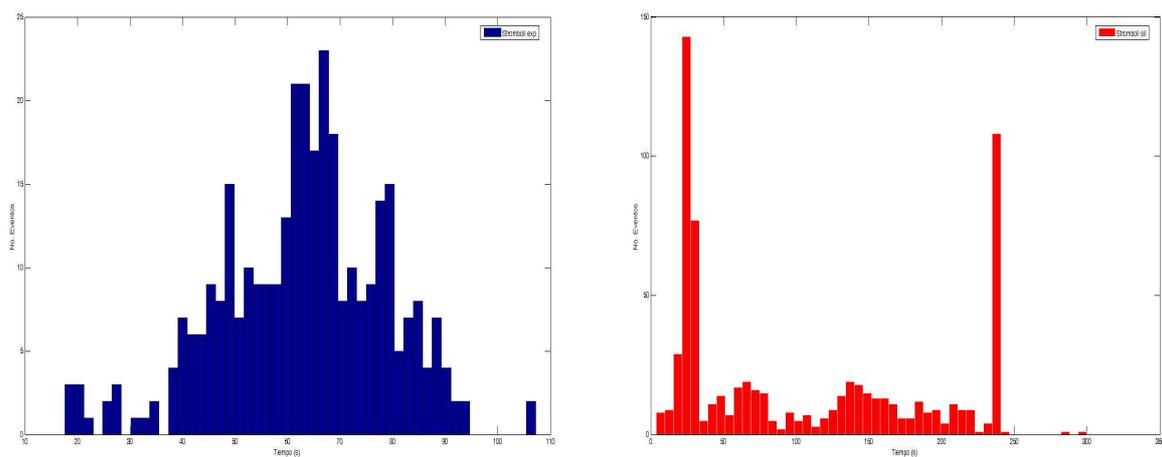


Fig. 4.41 Histograma de duración del evento “Exp” (izquierda) y el ruido “Sil” (derecha) en el Strómboli.

En el Etna se dan eventos que van desde los 8 segundos hasta un poco más del minuto de duración. En el Strómboli las explosiones duran desde unos 18 segundos hasta un poco más de 100 segundos.

4.4.3.- Estadísticas en el volcán San Cristóbal

En el San Cristóbal, los histogramas de los eventos (Fig. 4.42), presentan una duración que se encuentra entre un rango de 10 a 40 segundos, con un ruido de fondo que oscila entre 20 a 120 segundos.

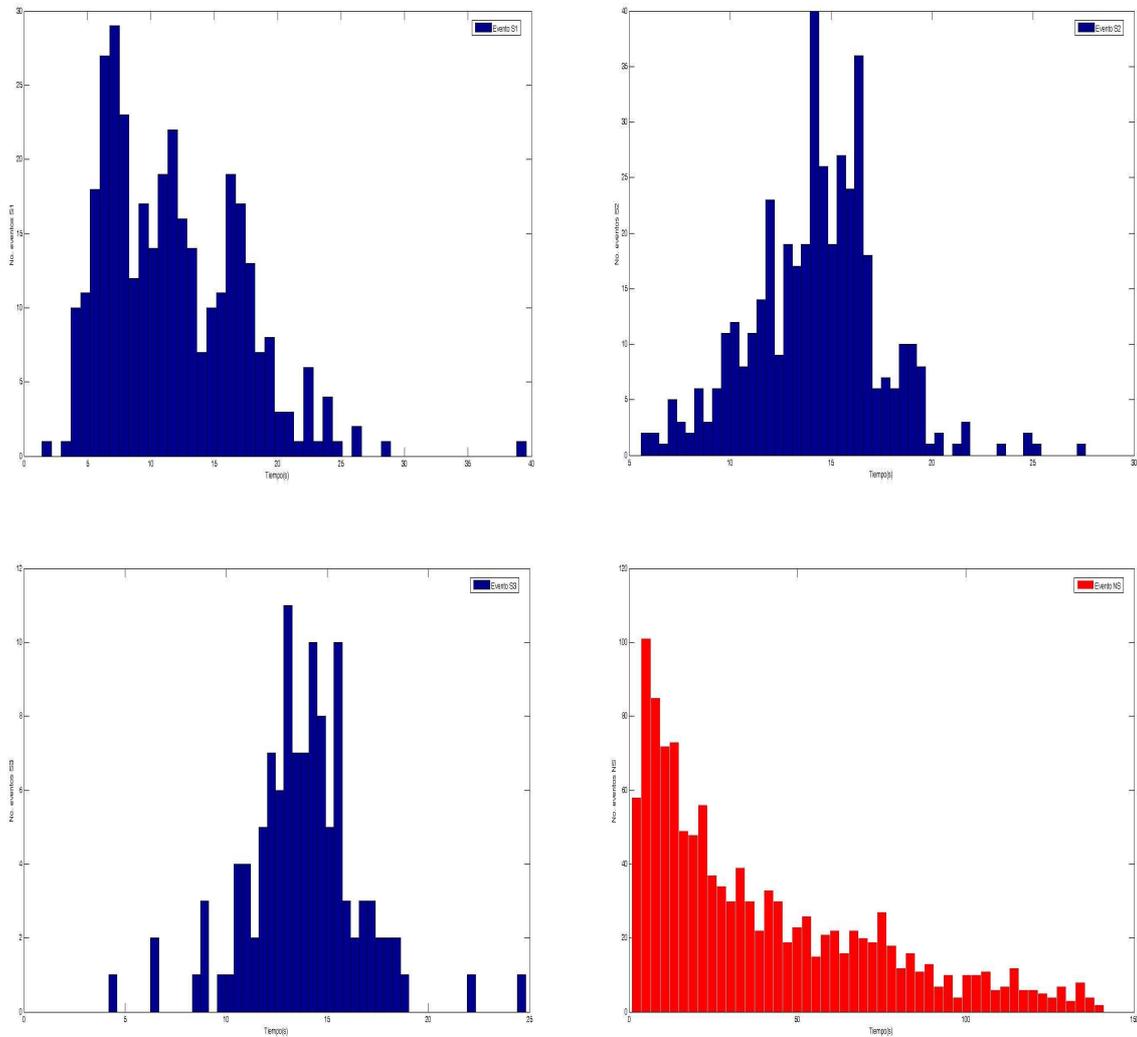


Fig. 4.42 Histograma de duración de los eventos en el volcán San Cristóbal. En orden de izquierda a derecha y arriba hacia abajo se encuentran los eventos S1, S2, S3 y NS

4.4.4.- Estadísticas en el volcán Telica

En cuanto al Telica, los histogramas de los eventos (Fig. 4.43), presentan una duración en un rango que oscila entre los 10 y 60 segundos. El ruido de fondo está entre 20 y 140 segundos.

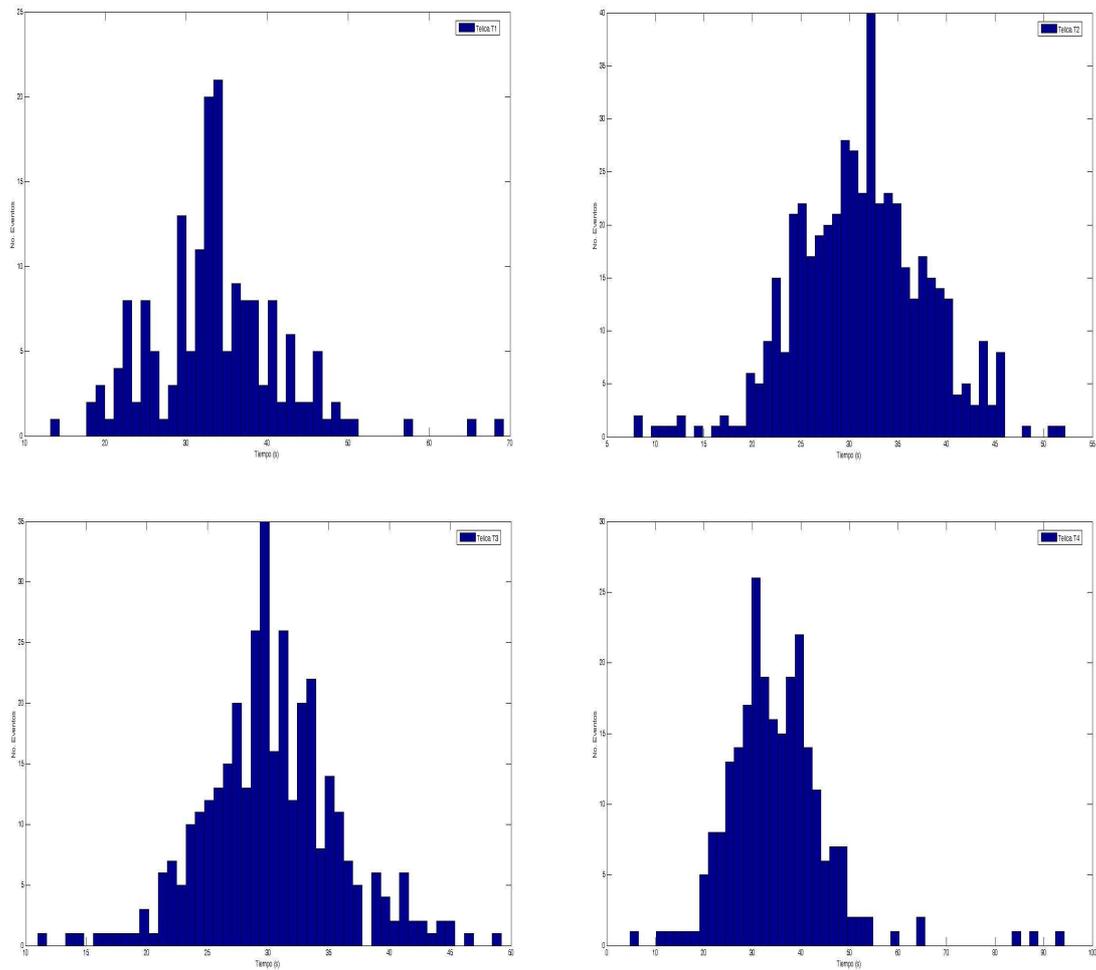


Fig. 4.43 Histograma de duración de los eventos en el volcán Telica. En orden de izquierda a derecha y arriba hacia abajo se encuentran los eventos T1, T2, T3 y T4

El ruido en el Telica presenta un intervalo de duración desde 0.78 seg a 580 seg, que se observan en la figura 4.44

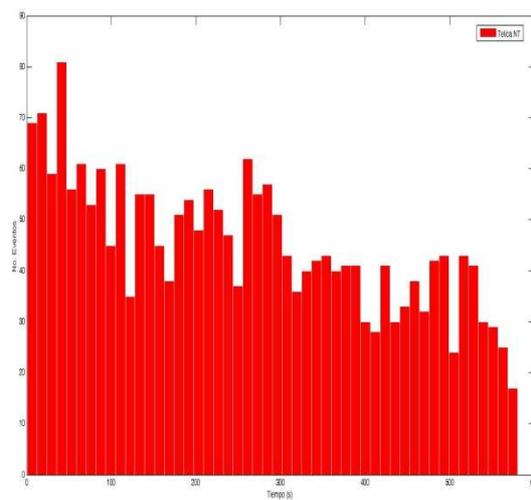


Fig. 4.44 Histograma de duración del ruido o tremor de fondo (NT) en el volcán Telica.

De los gráficos anteriores se puede calcular la duración máxima, mínima, promedio y desviación estándar de los eventos en los volcanes. Por ejemplo, para los de Italia los registros máximos son de 107 seg., y para Nicaragua son de 94 seg. Los valores están representados en las tablas 4.1 a 4.4.

i) Volcán Etna

Evento	Desv. Estándar	Dur. Mínima (seg)	Promedio (seg)	Dur. Máxima (seg)
NE	73.8036	0.8000	110.8275	289.4100
TB	8.9305	6.8000	23.6415	69.2000

Tabla 4.1 Estadísticas correspondientes a la desviación estándar, duración máxima, duración mínima y promedio de los eventos en el volcán Etna de Italia.

ii) Volcán Strómboli

Evento	Desv. Estándar	Dur. Mínima (seg)	Promedio (seg)	Dur. Máxima (seg)
Sil	83.5196	3.8200	106.4755	300.0200
Exp	15.8198	17.6200	62.9303	107.2000

Tabla 4.2 Estadísticas correspondientes a la desviación estándar, duración máxima, duración mínima y promedio de los eventos en el volcán Strómboli de Italia

Para los volcanes de Nicaragua los registros son de 300 seg., es decir de 5 minutos continuos. Se observa en las siguientes tablas.

iii) Volcán San Cristóbal

Evento	Desv. Estándar	Dur. Mínima (seg)	Promedio (seg)	Dur. Máxima (seg)
NS	33.9785	0.8000	39.6115	141.0000
S1	5.3780	1.4000	11.7387	39.6000
S2	3.2139	5.6000	14.2962	27.6000
S3	2.8450	4.2000	13.7722	24.8000

Tabla 4.3 Estadísticas correspondientes a la desviación estándar, duración máxima, duración mínima y promedio de los eventos en el volcán San Cristóbal de Nicaragua.

iv) Volcán Telica

Evento	Desv. Estándar	Dur. Mínima (seg)	Promedio (seg)	Dur. Máxima (seg)
NT	163.8124	0.6700	200.0275	580.2000
T1	8.1718	13.2000	33.7839	69.2000
T2	6.7476	7.80000	31.2161	52.2000
T3	5.5516	11.0000	30.2423	49.2000
T4	10.5786	4.8000	35.1045	94.2000

Tabla 4.4 Estadísticas correspondientes a la desviación estándar, duración máxima, duración mínima y promedio de los eventos en el volcán Telica de Nicaragua.

Como se deduce de la observación de los histogramas y las tablas anteriores, los eventos sísmico-volcánicos representan una gran complejidad. Un mismo tipo de evento fluctúa desde 4 segundos hasta 94 segundos a lo largo del tiempo.

4.5.- PROCESO DE ETIQUETADO MANUAL DE LA BASE DE DATOS

En el desarrollo del presente trabajo, este es uno de los procesos que acarrea más tiempo (meses y años) y a la vez uno de los más delicados, ha sido un proceso bastante largo y riguroso en el que se invirtieron muchos meses de observación, análisis y descarte, para llegar a los tipos de eventos con los que se trabajaría. Una vez que se han definido todos los eventos en los cuatro volcanes, el siguiente paso es la preparación de los registros para la construcción de las bases de datos. De esta forma, se analizaron uno a uno los miles de registros almacenados en las campañas. Este proceso debe de ser muy detallado porque se ha mencionado que el éxito de establecer un buen reconocedor y clasificador de señales sísmico-volcánicas se debe en gran parte a poder establecer una buena base de datos. Por esta razón, llegado a este punto, el proceso de etiquetado es muy importante, ya que además de determinar cada tipo de evento se debe delimitar muy bien, el inicio y el final de dicho evento. Este el paso previo en donde los eventos sísmico-volcánicos van a ser marcados para posteriormente ser parametrizados (ficheros Mfcc). Para esto, se debe de contar una vez más con la experiencia humana, es crucial poder decidir de manera objetiva los intervalos de tiempo, en los cuales se ha de segmentar cada registro, al señalar los eventos que se encuentren en el. La herramienta designada para el etiquetado manual de los eventos es el programa Transcriber [Barras et. al. 1998]. Así pues, se reitera que este proceso se soporta y se realiza en su mayor parte en base a la experiencia y observación humana, conforme a cada tipo seleccionado. Ejemplos del etiquetado manual se observa en las figuras 4.45 a 4.48.

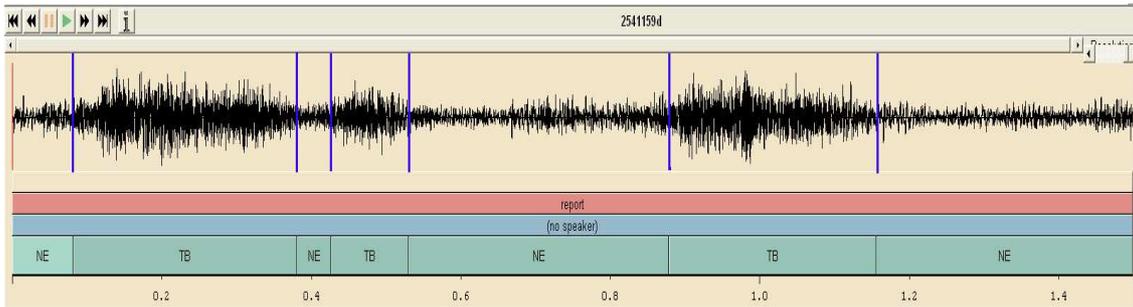


Fig. 4.45 Ejemplo del proceso de etiquetado manual de tres eventos TB en el Etna

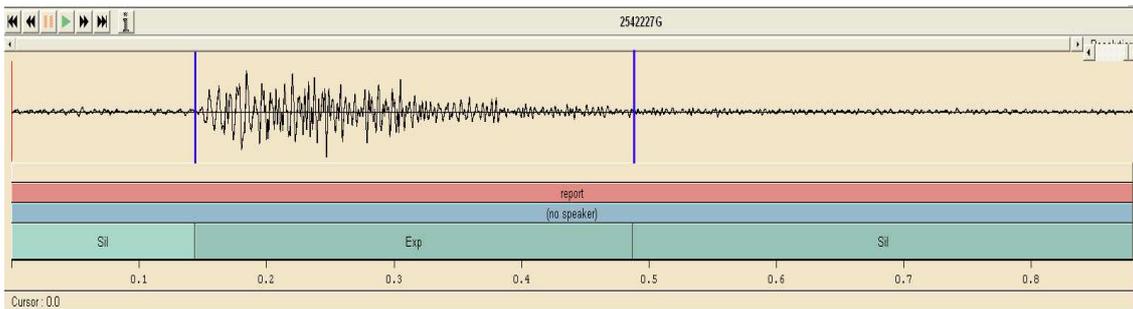


Fig. 4.46 Ejemplo del proceso de etiquetado manual de un evento Exp en el Strómboli

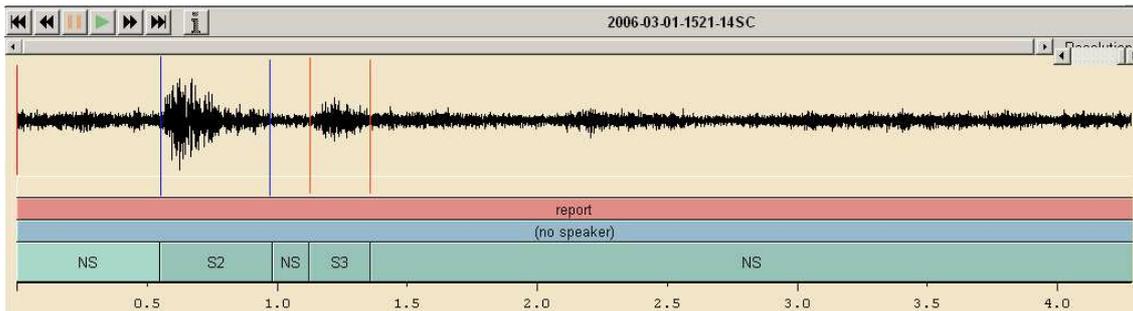


Fig. 4.47 Ejemplo del proceso de etiquetado manual de dos eventos (S2 y S3), en el San Cristóbal.

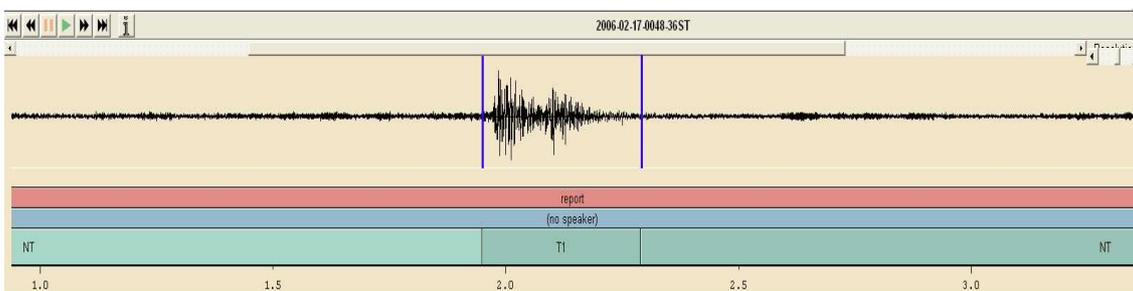


Fig. 4.48 Ejemplo del proceso de etiquetado manual de un evento T1 en el Telica

En la imagen de la figura 4.48 se observa que durante el desarrollo del etiquetado manual, el observador humano la mayor de las veces, tiene que expandir la señal a lo largo del tiempo, para poder ayudarse en identificar y marcar de manera más fiable, tanto el comienzo como el final de un determinado evento.

Este proceso se realiza con cada registro de los cuatro volcanes, en busca de eventos que constituirán la base de datos de entrenamiento. Sin embargo, al ser registros almacenados en continuo, a lo largo de las 24 horas del día, no en todos los registros se da el caso de poder encontrar algún evento, muchos de ellos solo cuentan con ruido de fondo, o se da el caso que incluso algunos no están bien definidos, ya que en algunos se produjeron errores, ya sea a la hora de la grabación del evento por parte del sensor, o por alguna otra causa. Otro aspecto a tomar en cuenta es que algunos eventos están muy enmascarados con ruido de fondo, de esta forma, se deben de seleccionar únicamente aquellos registros en los que uno o varios eventos se encuentran mejor definidos, sin tanta presencia de ruido de fondo.

Se ha llevado a cabo este proceso, segmentando (etiquetado) en el Etna con cada señal de la estación “d” (las 282 previamente seleccionadas), con dos tipos de eventos propios del volcán Etna. Denominamos al primer tipo, Tremor Burst (TB) y el otro lo hemos nombrado como Noise Etna (NE) (atendiendo a la particular forma de presentación del ruido en el Etna). Con estos dos únicos eventos clasificados, procedimos a realizar las pruebas y construir los modelos de entrenamiento. En el Strómboli se estableció un etiquetado con los dos eventos a analizar. Se nombró a las explosiones del Strómboli (Exp) y al segundo como silencio (Sil) que es el equivalente al ruido de fondo. Para los volcanes de Italia, los registros poseen a lo sumo una duración de 180 segundos (3 minutos).

En los volcanes de Nicaragua los registros sísmicos, poseen una duración de 300 segundos (5 minutos). En el volcán San Cristóbal se etiquetaron 3 tipos de eventos nombrándolos como S1, S2 y S3. Al tremor de fondo se le asignaron las siglas (NS) o “Noise San Cristóbal”. En el caso del Telica se etiquetaron cuatro tipos de señales T1, T2, T3 y T4 con un ruido de fondo propio del volcán bajo las siglas de (NT) o Noise Telica. Al finalizar todo este proceso, la totalidad de los registros segmentados en cada volcán se aprecian en la tabla 4.5.

ITALIA		NICARAGUA	
Etna	Strómboli	San Cristóbal	Telica
282	495	401	1240

Tabla. 4.5 Total de registros etiquetados en los cuatro volcanes de estudio

Como se observa en la tabla 4.5, el número de segmentaciones no es similar en todos los volcanes, esto responde a varios factores: primero, que se necesitan un número considerable de eventos (al menos 200), de cada tipo para construir una buena base de datos de entrenamiento, segundo, que el número de los eventos a reconocer no es el mismo en cada volcán y por último, los tipos de eventos que contiene cada registro no son los mismos, y varían en cada volcán, así algunos registros solo contienen un evento y de un solo tipo, otros registros contienen más eventos de ese mismo tipo, o incluso también poseen algún otro evento de diferente tipo. De esta forma, por ejemplo en el Telica se han necesitado más de 1,000 registros para obtener un número adecuado en la base de datos. Por lo tanto, una vez que se han elaborado la totalidad de registros segmentados se ha podido obtener el recuento de los diferentes tipos de eventos presentes en dichos registros, por lo que el número total de los tipos de eventos

encontrados en los registros segmentados de la tabla 4.5 anterior, se pueden observar en la tabla 4.6.

ITALIA				NICARAGUA			
Etna		Strómboli		San Cristóbal		Telica	
TB	411	Exp	337	S1	224	T1	360
NE	673	Sil	711	S2	330	T2	502
				S3	544	T3	429
				NS	1746	T4	298
						NT	2855

Tabla. 4.6 Número total de segmentaciones, de acuerdo a los diferentes tipos de eventos en cada uno de los volcanes de estudio.

Como se observa en la tabla anterior, se ha tratado de dejar en 200 el número mínimo de eventos en cada tipo. En la búsqueda de este número, algunos eventos de otro tipo se han presentado con mucha más regularidad en los registros. Por lo que por ejemplo para el San Cristóbal, el número de eventos de tipo S1 es el parámetro de referencia, lo mismo que para el Telica lo es el número de eventos de tipo T4, pero otros eventos como el caso del S3 en el San Cristóbal, y el T2 en el Telica han pasado de 500, lo que es bueno ya que a mayor número de eventos entrenados, mejores resultados se podrán esperar.

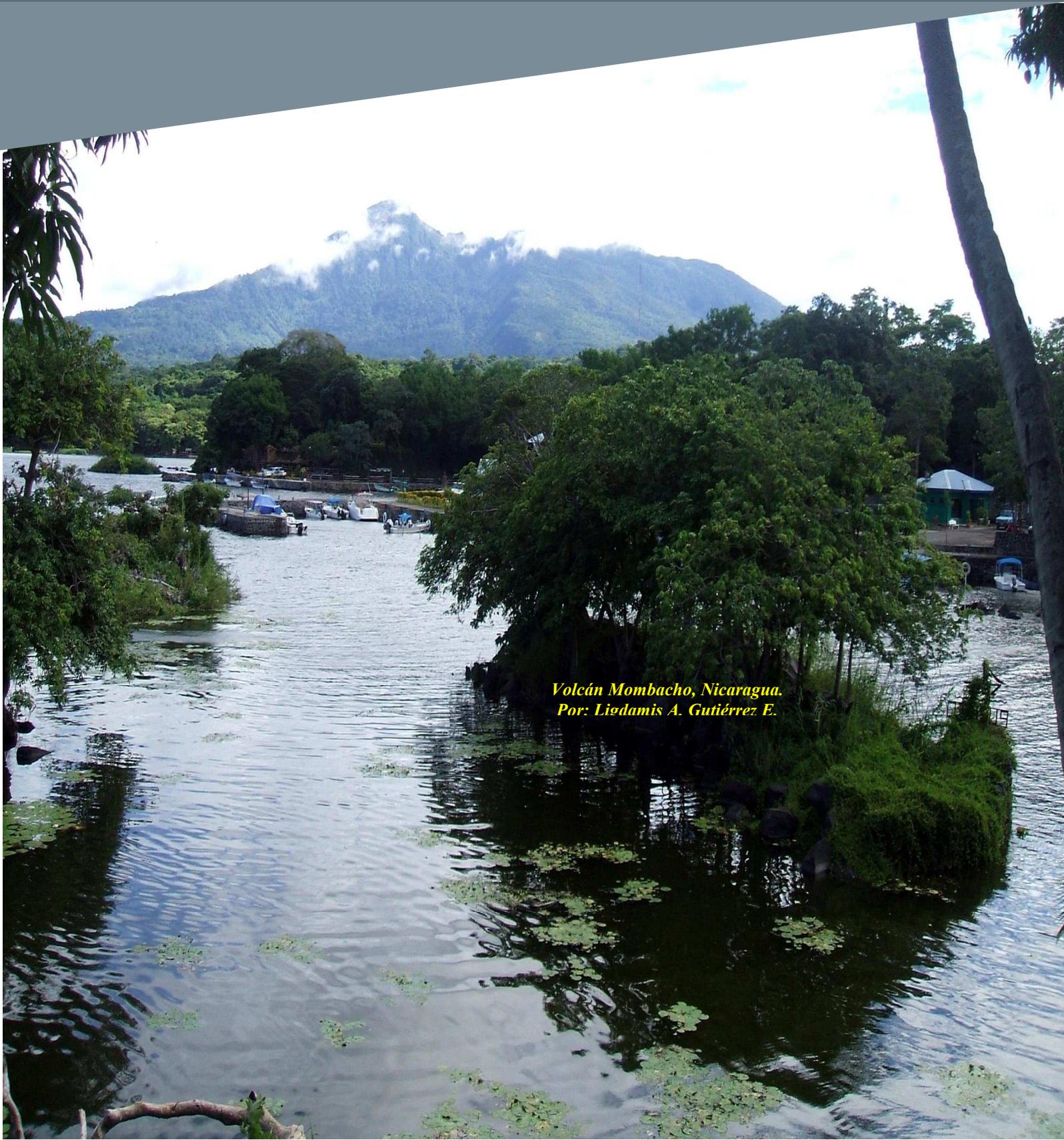
Llegado a este punto de la tesis, han sido analizados miles de eventos de forma individual, se ha realizado la conversión de igual número de ficheros, se ha concluido el proceso de etiquetación manual con los registros seleccionados y se ha establecido la cantidad de registros con los tipos de eventos etiquetados. El siguiente paso consiste en proceder a aplicar la metodología, de modo que con los eventos etiquetados (ver tabla 4.6), se van a constituir las bases de datos, que regirán a todas las demás fases en la aplicación del método, como son la creación y el entrenamiento de los modelos, y así aplicar las pruebas necesarias para el reconocimiento de las señales sísmico-volcánicas. Todo ello se tratará en el siguiente capítulo.





CAPÍTULO V

APLICACIÓN DEL MÉTODO Y RESULTADOS



*Volcán Mombacho, Nicaragua.
Por: Lidamis A. Gutiérrez E.*



5.1.- INTRODUCCIÓN.

El capítulo anterior, concluyó con la designación de los diferentes tipos de eventos a trabajar, a la vez que se procedió a la segmentación o etiquetación manual para la construcción de la base de datos de entrenamiento. El paso a seguir determina que para cada tipo de evento se procederá a la generación de un conjunto de vectores parametrizados. En el apéndice H, se detalla todo lo que involucra este proceso, así como la creación de los diccionarios de eventos y las gramáticas, para posteriormente por medio de listas, crear los ficheros parametrizados “mfcc”, esto se realiza mediante scripts. Seguidamente, se realiza la creación del modelo base o prototipo, y se configuran los parámetros en el script de entrenamiento y reconocimiento (Ver apéndice H). En cada una de las pruebas, el sistema genera los denominados ficheros “mlf”, que corresponden por parte de los modelos de entrenamiento a la obtención de la clasificación y reconocimiento automático de los eventos. Además, resume el proceso de iteración con las gaussianas iniciales a través de una estadística resultante [Young et. al. 2002].

En este capítulo se realizará la aplicación del método de estudio, para ello se desarrollarán cuatro tipos de pruebas, las primeras son las denominadas “**test cerrado o entrenamiento**”, realizadas sobre la base de datos previamente etiquetada. Un resultado exitoso en las pruebas, indicaría un porcentaje de precisión superior al 80%. Lo que significaría que el sistema de cada 100 eventos, está reconociendo con efectividad 80 de ellos. A partir de los resultados iniciales, se introducirá un proceso de mejoras en las pruebas cerradas, que consistirá en realizar una serie de modificaciones a algunos de los parámetros de configuración iniciales, para observar cuál de ellos incrementa los porcentajes de acierto y precisión, y a la vez determinar si se pueden disminuir los valores de las inserciones, borrados y sustituciones en los eventos. Dicho proceso de mejoras constará de cinco pasos; los primeros cuatro, comprenderán la modificación de cuatro de los parámetros de configuración iniciales, el quinto no versará sobre el fichero de configuración, sino que significará realizar un proceso de depuración en el último resultado de los cuatro pasos previos, para eliminar de la base de datos, aquellos registros que obtengan menores porcentajes de aciertos y que estuvieran perjudicando tanto los porcentajes de resultado como el reconocimiento de los eventos. Los cinco puntos son los siguientes:

- a) Cambio en el número de iteraciones
- b) Cambio en el número de Gaussianas
- c) Cambio en el factor de penalización (-p) de la herramienta de reconocimiento Hvite
- d) Cambio en el número de estados *
- e) Proceso de depurado selectivo de los registros, con menores porcentajes de acierto

Llegando al punto del paso “d”, se determina marcar un hito (*), en donde se repetirán las pruebas con el número de gaussianas del paso “b”, para valorar las mejoras que hasta ahora se han obtenido con las modificaciones. Una vez finalizados estos cinco pasos del proceso de mejoras del test cerrado, el siguiente paso a realizar es el segundo tipo de pruebas, denominadas “**test abierto o ciego**”, que esencialmente trata en dividir en dos partes la base de datos. Una primera parte, la formará los modelos entrenados, para pasarlo después a la parte restante y observar los porcentajes de reconocimiento del

modelo entrenado sobre la parte restante. El tercer tipo de pruebas a realizar son las denominadas “**ciego - ciego**”, en las que la base de datos de entrenamiento, reconocerá registros adicionales, que no han sido previamente parametrizados, y que se encuentran sin segmentar (no etiquetados), y en los que la mayoría de ellos posee un contenido de fuerte ruido de fondo. Con esta prueba, podrá constatarse si el modelo de entrenamiento puede reconocer eventos en registros que no han sido manipulados por el observador. Los resultados de estos tres grupos de pruebas podrán medir la eficacia del sistema, para poder reconocer las señales sísmico-volcánicas en cada volcán. El cuarto y último grupo de pruebas, consistirán en “**mezclar las bases de datos de entrenamiento**”, para así poder ver la generalización o no, de los resultados obtenidos. Es decir, si los modelos de entrenamiento de un volcán pueden reconocer eventos en los registros de otro volcán. Para ello se mezclarán las bases de datos y los modelos en dos fases; la primera mezclará las del Etna con las del Strómboli, y las del San Cristóbal con las del Telica. En la segunda fase se realizará la mezcla de las bases de datos y los modelos de entrenamiento de los cuatro volcanes. Todo este proceso de modificaciones indica un desarrollo largo y metódico, que llevará a realizar varios miles de pruebas a lo largo de varios meses, e incluso años, para obtener cada vez mejores porcentajes. Los resultados obtenidos son los que se plasman en este capítulo. Al final del capítulo se atenderá a la discusión de resultados.

Los experimentos se han realizado en diferentes ordenadores con diferentes entornos operativos. En Windows se ha instalado una máquina virtual, para utilizar un servidor Ubuntu. También se han utilizado terminales conectadas a un servidor central con Linux mediante SSH. (Secure Shell), o servidor de órdenes seguras. Esto supone un tiempo de proceso de cálculo en el entrenamiento de aproximadamente 10 minutos (en el reconocimiento es mucho menor), aumentando al incrementar algunos valores como el del número de Gausianas, por lo que se debe de llegar a un consenso en el que la relación entre resultados y el factor tiempo sea el más adecuado. Esto sería un buen punto a tratar en futuras investigaciones, buscar un nuevo o mejor algoritmo, el cual no repercuta en el coste computacional que en el presente estudio se ha obtenido. A continuación se verá el proceso del desarrollo de las pruebas en cada uno de los volcanes.

5.2.- **DESARROLLO DE LAS PRUEBAS EN LOS VOLCANES**

Se realizarán las pruebas de entrenamiento y reconocimiento con las bases de datos etiquetadas en los cuatro volcanes. El procedimiento a seguir será iniciar con un primer ensayo y a partir del resultado, ha de realizarse el proceso de mejoras antes mencionado. A continuación se presentan los resultados obtenidos, en cada uno de los volcanes de estudio.

5.2.1.- **Análisis de resultados en el Etna**

Las primeras pruebas a realizar con el conjunto de herramientas de entrenamiento y reconocimiento de modelos ocultos de Markov “HTK”, son las llamadas pruebas cerradas o “test cerrado”, las cuales se realizan sobre la propia base de datos previamente etiquetada. Aquí se detalla todo el proceso realizado en el Etna, que es el mismo que se repitió con las bases de datos de los tres volcanes restantes. Como ya se ha mencionado, la configuración de los parámetros iniciales en los experimentos se ha tomado en base a: modelos con 11 estados, 6 iteraciones y 16 Gausianas. En esta

primera prueba se utilizó una base de datos con 281 registros segmentados y etiquetados del Etna. El resultado del primer experimento en el Etna se muestra en la tabla 5.1

```
===== HTK Análisis de Resultados =====
SENT: %Correct=28.11 [H=79, S=202, N=281]
WORD: %Corr=90.77, %Acc=69.93 [H=984, D=73, S=27, I=226, N=1084]
----- Matriz de Confusión -----
```

	NE	TB
NE	604	17
TB	10	380
Ins	161	65
Del	52	21

Tabla 5.1 Resultados obtenidos del HTK en el primer experimento en el Etna.
%Corr = Porcentaje correcto de etiquetas reconocidas, %Acc = Porcentaje de precisión
H= Número correcto de etiquetas, D = Número de eliminaciones, S = Número de sustituciones,
I = Número de Inserciones, N = Número total de etiquetas. Ins = Inserciones, Del = Borrados.

La tabla anterior presenta un porcentaje de aciertos del 90.77 % y un porcentaje de efectividad de 69.93%. Para que se considere un éxito, el porcentaje de precisión (Acc) debe ser superior al 80%, lo que indicaría que el sistema es fiable en su reconocimiento. En estas primeras pruebas, el resultado del porcentaje requerido es un 10% menor. La intersección de las columnas con las filas en la matriz de confusión indica si el sistema clasifica eventos de un tipo como otro. En el caso del evento TB, la tabla 5.1 indica que fueron reconocidos por el sistema 380 eventos, hubo confusión en 17 eventos con ruido. Análogamente, el evento NE en el que el sistema reconoció 604 eventos, y hubo confusión con 10 eventos con el evento TB. Otra información que muestra la tabla 5.1 es que el sistema de 1,084 eventos analizados, eliminó 73 eventos constituidos de 52 eventos de ruido, y 21 eventos tipo TB, el sistema además sustituyó 27 eventos e insertó 226 eventos, la mayoría de ruido de fondo con 161, y 65 de tipo TB. Para mejorar estos resultados, ahora se realiza el proceso de mejoras, al modificar los parámetros y observar los efectos.

a) Cambio en el número de iteraciones

En un primer grupo de experimentos, se realizó la modificación del parámetro del número de iteraciones. Al establecer el número de iteraciones en 3, 6, 9, 12 y 15, el resultado se observa en la tabla 5.2. En color amarillo, la fila con mejores resultados

Iteraciones	%Corr	%Acc	Sust	Ins TB	Del TB	Ins NE	Del NE
3	90.22	62.36	42	99	16	203	48
6	90.77	69.93	27	65	21	161	52
9	91.05	72.32	27	65	23	138	47
12	91.33	71.22	31	71	17	147	46
15	91.42	71.40	24	67	19	150	50

Tabla 5.2 Resultados obtenidos en las pruebas cerradas del HTK, con 11 estados y 16 Gausianas al modificar el parámetro del número de iteraciones en el Etna
%Corr = Porcentaje correcto de etiquetas reconocidas, %Acc = Porcentaje de precisión
Ins = Inserciones, Sust = Sustituciones, Del = Borrados

Al observar la tabla anterior, hasta las 9 iteraciones va en aumento el porcentaje de efectividad, disminuyendo a partir de dicho número, sin embargo en cuanto al número de sustituciones, borrados e inserciones, los valores respecto a las 6 iteraciones iniciales no mejoran. Otro punto a tener en cuenta es que a medida que aumente el número de iteraciones en el sistema, aumentará por ende el costo computacional, al tener que pasar mucho más por cada una de las 16 Gausianas. De esta forma, aunque el porcentaje mayor se observa en las 9 iteraciones, esto solo representa un aumento de un 1.47% respecto al de las 6 iteraciones, lo que no compensa el aumento de los otros factores, como sustituciones, borrados e inserciones, sumado al costo computacional. Por lo que se ha dejado el valor del parámetro, en torno a las 6 iteraciones.

b) Cambio en el número de Gausianas

Para este grupo de experimentos, las pruebas se han tomado a partir de ocho gaussianas, incrementando en ocho dichos valores en cada grupo de experimentos, hasta llegar a las 96, para así observar el porcentaje de acierto y efectividad del sistema. El resultado final de este grupo de pruebas se observa en la tabla 5.3, en amarillo se presenta la fila con mejores resultados en relación al costo computacional de tiempo.

Número de Experimento	No Gausianas	% Corr. Obtenido	% Acc. obtenido
1	8	88.56	57.10
2	16	90.77	69.93
3	24	93.08	78.51
4	32	94.56	84.59
5	40	95.85	89.11
6	48	96.59	91.79
7	56	97.05	94.00
8	64	97.42	95.11
9	72	97.60	95.94
10	80	97.97	96.96
11	88	98.06	97.14
12	96	98.43	97.60

Tabla 5.3 Resultados obtenidos en las pruebas cerradas del HTK, con 6 iteraciones y 11 estados, al modificar el parámetro del número de gaussianas en el Etna.

%Corr = Porcentaje correcto de etiquetas reconocidas, %Acc = Porcentaje de precisión
 Ins = Inserciones, Sust = Sustituciones, Del = Borrados

Un análisis gráfico de la variación del porcentaje (respecto al aumento del número de gaussianas en los experimentos), lo observamos en las figuras 5.1 y 5.2.

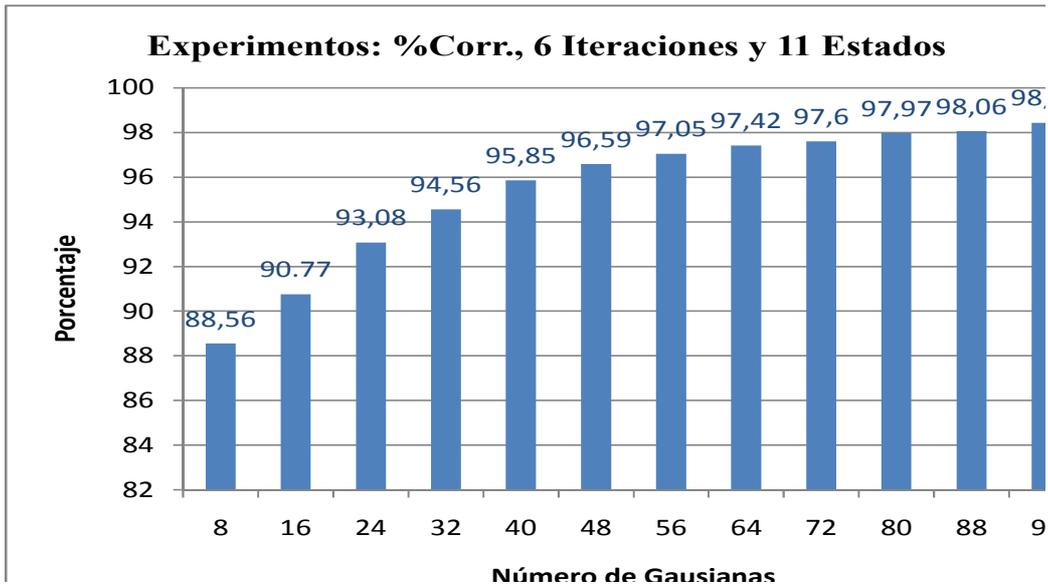


Fig. 5.1 Resultados obtenidos en %Corr en las pruebas cerradas del HTK, en el porcentaje de aciertos al modificar el parámetro del número de gaussianas en el Etna.

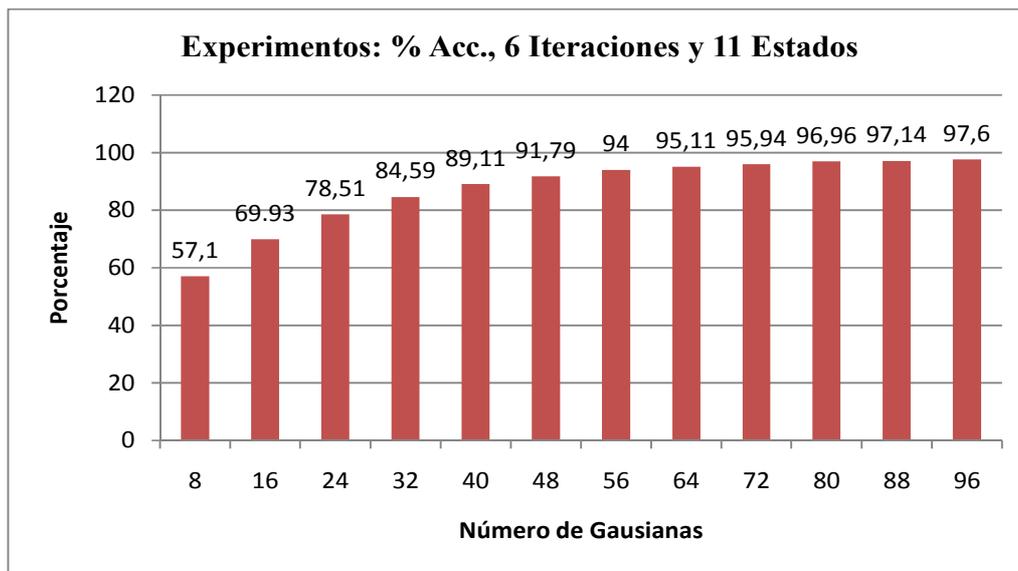


Fig. 5.2 Resultados obtenidos en % Acc en las pruebas cerradas del HTK, en el porcentaje de precisión al modificar el parámetro del número de gaussianas en el Etna.

Al analizar tanto las gráficas anteriores, como la tabla 5.3 se puede observar que al inicio existe una mejora significativa, tanto en el porcentaje correcto, como en el porcentaje de precisión. El sistema tiende a estabilizarse a partir de 72 Gausianas, hasta llegar a 96 Gausianas. El sistema mejoró el porcentaje de aciertos de los 91.24 % iniciales a 98.43 %, lo que representa una mejora bastante significativa. Lo mismo sucede con el porcentaje de precisión, al aumentar de 71.68 % a 97.60 % en 96 Gausianas. La matriz de confusión del experimento con 96 gaussianas, se puede observar en la tabla 5.4.

===== HTK Análisis de Resultados =====
 SENT: %Correct=91.10 [H=256, S=25, N=281]
 WORD: %Corr=98.43, %Acc=97.60 [H=1067, D=17, S=0, I=9, N=1084]
 ----- Matriz de Confusión -----

	NE	TB
NE	656	0
TB	0	411
Ins	9	0
Del	17	0

Tabla 5.4 Resultados obtenidos en las pruebas cerradas del HTK, en la matriz de confusión, al modificar a 96 el parámetro del número de gaussianas en el Etna.

%Corr = Porcentaje correcto de etiquetas reconocidas, %Acc = Porcentaje de precisión
 H= Número correcto de etiquetas, D = Número de eliminaciones, S = Número de sustituciones,
 I = Número de Inserciones, N = Número total de etiquetas. Ins = Inserciones, Del = Borrados

Tal y como se aprecian los resultados de las tablas anteriores, cuanto mayor sea el número de Gaussianas, mejor serán los resultados obtenidos. Si se compara la matriz de confusión de este experimento con el primero, se puede observar el aumento significativo en el número de eventos detectados por el sistema. Por otra parte, el sistema se estabiliza un poco entre los experimentos 8 y 9 (Ver Tabla 5.2), cuando el número de gaussianas está en 72 y 80 respectivamente, pero nuevamente aumenta el porcentaje de aciertos en el experimento 10, al aumentar el número de gaussianas a 88. Sin embargo, el aumento del número de gaussianas implica un costo computacional bastante elevado, ya que al incrementar el número de gaussianas hace que se incremente el tiempo de cálculo por parte del sistema, pues al aumentar el número de gaussianas se eleva el tiempo de procesamiento de los cálculos, ya que el modelo matemático, tiene que realizar las seis iteraciones con cada uno de los indicadores de gaussianas. Así por ejemplo se realizan 6 iteraciones con 1, 2, 3, 4,....16, 17, 18....96 Gaussianas. Este proceso tarda muchas horas dependiendo de las capacidades del ordenador, lo que para análisis en tiempo real no es factible. A fin de buscar un balance entre costo computacional y mejores resultados, se determina dejar el número de gaussianas en 32. Ahora bien, al comparar los resultados entre el primer (tabla 5.1) y último (tabla 5.4) experimento, puede observarse que por una parte, se incrementó el número de eventos TB reconocidos por el sistema, a la vez que disminuyó significativamente el porcentaje de error en los eventos TB.

c) Cambio en el factor de penalización

Teniendo este primer avance en los resultados al modificar las gaussianas, una vez más se procede a mejorar dichos resultados. Para ello, ahora se modificará la línea de comandos en el script del HTK, que corresponde al de la herramienta “Hvite”, que implica el reconocimiento y que se encuentra en el fichero de ejecución “trabajo.csh”. En el capítulo tercero de este trabajo se mencionó que esta herramienta se encarga (dentro del conjunto de herramientas de HTK), del reconocimiento. Así Hvite es la herramienta que maneja el algoritmo de Viterbi para el reconocimiento, la cual proporciona el camino de mayor probabilidad para la secuencia óptima de estados. Por lo que en teoría, al modificar dicho parámetro se podría observar alguna mejora. La modificación implica cambiar el factor (-p) de dicha herramienta. Dicho factor llamado el “factor de penalización de inserción de palabra” (*Word insertion penalty*), es un valor fijo añadido a cada muestra, cuando se transita desde el extremo de una frase al

comienzo de la siguiente, para el caso del presente trabajo, sería desde el comienzo de un evento al siguiente. Tanto el parámetro $-p$ como el parámetro $-s$ podrían tener un efecto significativo en el rendimiento del reconocimiento y, por tanto, su modificación realizando ajustes podría ser muy útil [Young et. al. 2002]. El valor del factor $-p$ por defecto es de cero, por lo que se han realizado experimentos con una diferencia de cinco unidades arriba y abajo hasta el valor 20, para observar los efectos en los porcentajes de acierto y en el número de inserciones y borrados del sistema. Al modificar los parámetros, los resultados finales se muestran en la tabla 5.5., en amarillo se observa la fila que presenta los mejores resultados.

Factor – p	%Corr	%Acc	Sust	Ins TB	Del TB	Ins NE	Del NE
15	96.40	70.11	15	52	5	233	19
10	95.76	78.41	14	37	9	151	23
5	95.30	82.10	8	30	14	113	29
0	94.56	84.59	6	22	16	86	37
-5	93.27	85.24	9	20	18	67	46
-10	92.71	86.81	8	16	18	48	53
-15	92.16	87.64	9	14	18	35	58

Tabla 5.5 Resultados obtenidos con 6 iteraciones, 11 estados y 32 Gaussianas, en la modificación del parámetro “-p” del Hvite en el Etna

%Corr = Porcentaje correcto de etiquetas reconocidas, %Acc = Porcentaje de precisión
Ins = Inserciones, Sust = Sustituciones, Del = Borrados

Los mejores resultados en efectividad, han sido obtenidos con un factor de penalización de -15, por lo que se establecerá en este valor. Hasta ahora, todo el proceso de cambios ha sido: realizar primero experimentos con los datos originales, los cuales arrojaron un porcentaje de efectividad cercano al 72%, lo que había que mejorar. En este sentido se han hecho modificaciones para mejorar los resultados de dichos experimentos, primero se modificó el número de iteraciones, fijándolo en 6, seguidamente se realizaron pruebas con la modificación del número de gaussianas, determinándose en 32, que es el número idóneo entre el costo computacional y los resultados presentados por el sistema. Posteriormente, se ha realizado la modificación del número que establece el factor “-p” (que es el factor de penalización de la herramienta de reconocimiento Hvite), siendo en este caso un valor (-15), el número a fijar en los siguientes ensayos.

d) Cambio en el número de estados

Se ha establecido un nuevo grupo de experimentos, siempre buscando mejoras en el sistema, que consistirá en realizar modificaciones del parámetro que corresponde al número de estados, se ha considerado comenzar a realizar los experimentos a partir de 5 estados, incrementando el número de estados en tres en cada ocasión, hasta 19 estados para observar en cuál de ellos se obtiene los mejores resultados. Dicha modificación, implica el tener que crear mediante el script “crea_prototipo.csh” (Ver apéndice H), los prototipos de modelos iniciales de entrenamiento, para cada grupo de experimentos. Así por ejemplo, se han creado modelos prototipos con 5 a 19 estados y con tamaño de vectores en 39 coeficientes a utilizar (que es el número de coeficientes que se han estimado establecer), teniendo así 8 modelos prototipos de 39 coeficientes con diferentes números de estados (proto.39.5,...proto.39.19). En las siguientes pruebas

se aplicarían los parámetros hasta ahora ya establecidos, es decir, 6 iteraciones, 32 Gaussianas y un factor $-p$ de -15 con esta nueva modificación. Los resultados se observan en la tabla 5.6., en amarillo se presenta la fila con mejores resultados.

No. Estados	%Corr	%Acc	Sust	TB Ins	TB Del	NE Ins	NE Del
5	88.93	72.05	22	95	23	88	75
7	84.50	75.83	15	40	63	54	90
9	91.42	88.01	9	10	45	27	39
11	92.16	87.64	9	14	18	35	58
13	93.27	92.71	1	1	41	5	31
15	95.02	93.73	2	4	32	10	20
17	91.88	89.02	6	17	43	14	39
19	91.97	88.93	5	10	44	23	38

Tabla 5.6 Resultados al modificar el parámetro del número de estados en el Etna, obtenidos en las pruebas cerradas del HTK, con 6 iteraciones, 32 gaussianas y factor $-p$ en -15
 %Corr = Porcentaje correcto de etiquetas reconocidas, %Acc = Porcentaje de precisión
 Ins = Inserciones, Sust = Sustitución, Del = Borrados

Como se puede observar en la tabla anterior, los resultados en el porcentaje de efectividad aumentan hasta los 15 estados, cuando comienzan a disminuir, se observa que entre el modelo con trece estados y el de quince solo hay un pequeño margen de 0.02% de aumento en este último, por lo que habría que observar las variables restantes para determinar cuál es el modelo que indique mejores resultados, es decir, tanto las sustituciones, inserciones, y borrados. El número de sustituciones e inserciones van disminuyendo, desde los modelos con 5 estados, hasta los modelos con trece estados, a partir de los cuales comienzan a aumentar. En cuanto al número de eliminaciones, es bastante variable en relación al número de estados. Esto hace indicar, que el número de estados más favorable para realizar los entrenamientos de los modelos es el de trece. A este punto se marca un hito, que es el repetir las pruebas observadas en la tabla 5.2, para observar los cambios que se obtienen en relación a los parámetros ya establecidos como mejoras. Para ello, se establecerán los valores hasta ahora modificados y comprobados: 6 iteraciones, 13 estados, el factor de penalización $-p$ en (-15), y se va a constituir un nuevo grupo de experimentos, para así comprobar el incremento del porcentaje de precisión del sistema, al aumentar el número de Gaussianas (que obtiene las mayores mejoras). La tabla 5.7, representa los valores obtenidos en dichos experimentos, en amarillo se presenta la fila con mejores resultados

Número de Experimento	No Gaussianas	% Corr. obtenido	% Acc. obtenido
1	8	80.72	73.06
2	16	87.64	84.41
3	24	90.96	89.76
4	32	93.27	92.71
5	40	95.30	95.02
6	48	96.40	96.31

Tabla 5.7 Resultados obtenidos en las pruebas cerradas del HTK, con 6 iteraciones, 13 estados, y factor $-p$ (-15), al modificar el parámetro del número de gaussianas en el Etna. %Corr = Porcentaje correcto de etiquetas reconocidas, %Acc = Porcentaje de precisión.

Al comparar los resultados de la tabla anterior con los de la tabla 5.2, se observan las mejoras en los modelos de entrenamiento con 13 estados. En los modelos con 32 Gausianas, el sistema con 13 estados presenta un porcentaje de precisión (Acc) de 92.71 %, lo que supera a los datos obtenidos con los modelos con 11 estados en 84.59 % de Acc (Cfr. tabla 5.2) El sistema ya se estabiliza a partir del experimento con las 56 Gausianas. Esto indica que en las siguientes pruebas, los experimentos han de tomarse con 13 estados. Por lo tanto, los resultados en las pruebas finales con los modelos de entrenamientos en las pruebas cerradas con 32 Gausianas en el Etna (en amarillo en la tabla 5.7), se puede observar en la tabla 5.8.

===== HTK Análisis de Resultados =====
 SENT: %Correct=79.36 [H=223, S=58, N=281]
 WORD: %Corr=93.27, %Acc=**92.71** [H=1011, D=72, S=1, I=6, N=1084]
 ----- Matriz de Confusión -----

	NE	TB
NE	641	1
TB	0	370
Ins	5	1
Del	31	41

Tabla 5.8 Resultados obtenidos del HTK en el experimento final del Etna. Utilizando 6 Iteraciones, 13 estados, 32 Gausianas y factor $-p = -15$

%Corr = Porcentaje correcto de etiquetas reconocidas, %Acc = Porcentaje de precisión
 H= Número correcto de etiquetas, D = Número de eliminaciones, S = Número de sustituciones,
 I = Número de Inserciones, N = Número total de etiquetas. Ins = Inserciones, Del = Borrados

En la tabla anterior se observa que el sistema, en los 281 registros segmentados, ha logrado un porcentaje de aciertos correctos (Corr) de 93.27% y una efectividad (Acc) de 92.71%. De 1.084 eventos el sistema ha reconocido 1,011, ha eliminado 72, sustituido 1 e insertado seis. Hasta ahora, en cada prueba se ha podido mejorar el porcentaje de efectividad en el reconocimiento por parte del sistema.

e) Proceso de depurado selectivo de los registros

Una vez obtenidos los resultados finales con las modificaciones de parámetros en el proceso de mejoras, el siguiente y último paso en dicho proceso, es observar los resultados obtenidos en cada experimento de forma manual. Con esto se pretende poder identificar los elementos a mejorar en el sistema, como son aquellas posibles señales en las que el valor del porcentaje de efectividad es bajo, y así eliminar dichos registros de la base de datos de entrenamiento. Con esto se logrará una base de datos mucho más completa y fiable. Un ejemplo de la forma manual del proceso de depurado se presenta (en el fragmento del fichero de resultados), en la figura 5.3. En donde se observan los registros reconocidos por el sistema (evento a evento). En dicha lista, los registros que han sido reconocidos por el modelo de entrenamiento están en color rojo y los registros con bajo porcentaje en el reconocimiento (menor a 60%) se presentan en color verde.

Proceso de Depurado Selectivo en el Etna

```

----- Sentence Scores -----
===== HTK Results Analysis =====
Rec : rec.13.32.gram.mlf
----- File Results -----
2541132d.rec: 100.00(100.00) [H= 7, D= 0, S= 0, I= 0, N= 7]
2541138d.rec: 100.00(100.00) [H= 3, D= 0, S= 0, I= 0, N= 3]
2541144d.rec: 100.00(100.00) [H= 3, D= 0, S= 0, I= 0, N= 3]
2541150d.rec: 100.00(100.00) [H= 3, D= 0, S= 0, I= 0, N= 3]
2541226d.rec: 44.44( 44.44) [H= 4, D= 5, S= 0, I= 0, N= 9]
Aligned transcription: 2541226d.lab vs 2541226d.rec
LAB: NE TB NE TB NE TB NE TB NE
REC:      NE TB NE      NE
2550105d.rec: 57.14( 57.14) [H= 4, D= 3, S= 0, I= 0, N= 7]
Aligned transcription: 2550105d.lab vs 2550105d.rec
LAB: NE TB NE TB NE TB NE
REC:      NE TB      TB NE
2550232d.rec: 57.14( 57.14) [H= 4, D= 3, S= 0, I= 0, N= 7]
Aligned transcription: 2550232d.lab vs 2550232d.rec
LAB: NE TB NE TB NE TB NE
REC:      TB NE TB NE
2550256d.rec: 100.00(100.00) [H= 3, D= 0, S= 0, I= 0, N= 3]
2550308d.rec: 100.00(100.00) [H= 3, D= 0, S= 0, I= 0, N= 3]
2550314d.rec: 100.00(100.00) [H= 5, D= 0, S= 0, I= 0, N= 5]
2550317d.rec: 100.00(100.00) [H= 3, D= 0, S= 0, I= 0, N= 3]

```

Fig. 5.3 Ejemplo del Análisis manual en el Proceso selectivo (evento por evento), de los resultados obtenidos en el Etna. Los valores en rojo representan los registros que han tenido un porcentaje de 100 y que no se cambiarán. Los valores en verde son aquellos que han tenido un porcentaje menor a 60 y que se analizarán para eliminarse de la base de datos.

El uso de este tipo de registro mostrado en la figura anterior resulta muy útil, ya que muestran cada uno de los registros como son reconocidos por el sistema. En la figura 5.3 se observan en rojo, los registros en los que el sistema ha coincidido al 100% con el observador, en cuanto al número y tipo de eventos segmentados. Los registros marcados en verde son los que existen discrepancias entre el sistema y el observador. En este caso se van a eliminar de la base de datos aquellos registros en el que el porcentaje sea inferior a 60%. Por ejemplo los registros 2541226d, 2550105d y 2550232d, en los que el porcentaje fue de 44.44 y 57.14, en estos registros, en el 2541226d, de 9 eventos, el sistema solo ha reconocido 4, a eliminado 5, 3 de tipo TB y dos de tipo NE, no ha insertado ni sustituido ninguno. En los otros dos registros restantes, de 7 eventos el sistema ha podido reconocer 4, ha eliminado 3, siendo dos de ellos de tipo TB, y uno de tipo NE, el sistema no ha sustituido ni insertado ningún evento. En las siguientes figuras se presentan dichos registros, en donde se podrá observar cómo se representa la señal en el dominio del tiempo y su espectrograma. Así se analizará porqué el sistema no ha podido reconocer con un porcentaje superior de precisión los eventos presentes en dichos registros.

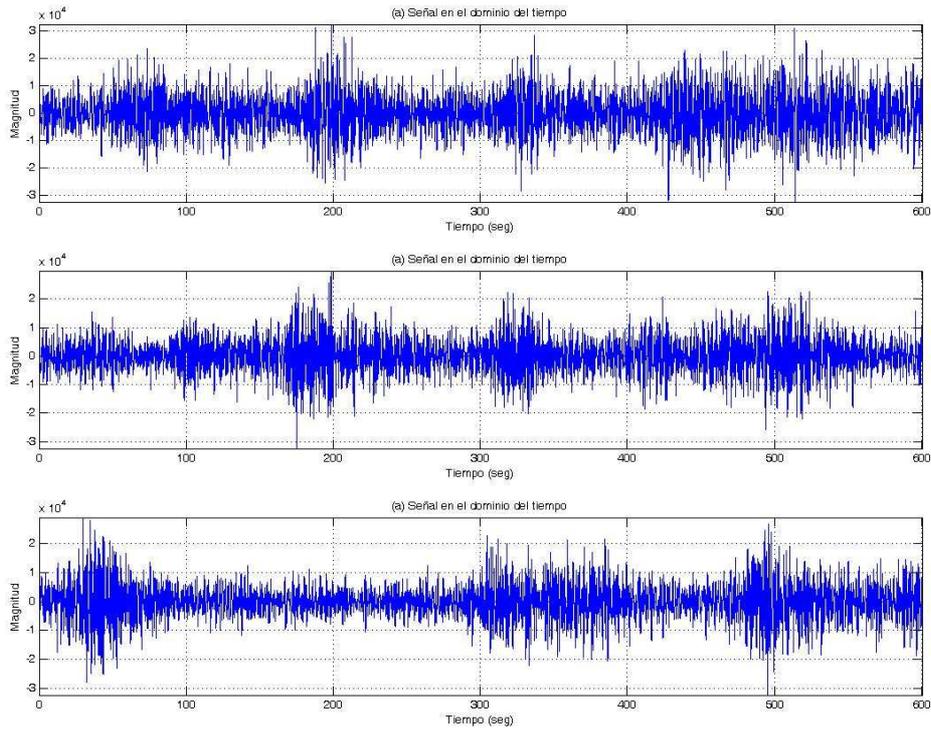


Fig. 5.4 Registros: 2541226d (arriba), 2550105d (centro) y 2550232d (abajo), en el dominio del tiempo

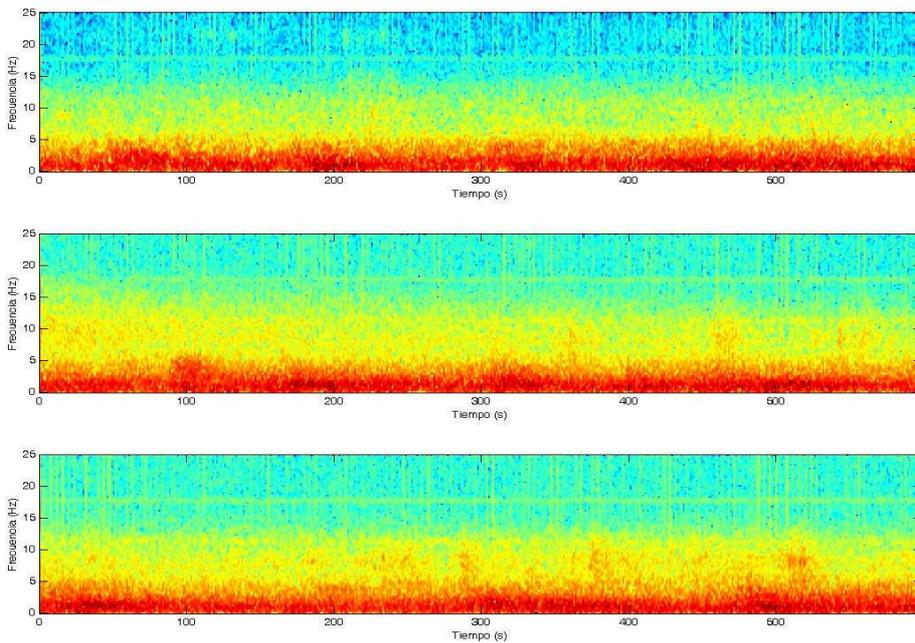


Fig. 5.5 Espectrogramas de los registros: 2541226d (arriba), 2550105d (centro) y 2550232d (abajo).

En los registros anteriores, el sistema no ha podido reconocer eventos marcados por el observador. Al analizar dichos registros, puede notarse que el tremor de fondo enmascara en gran medida los eventos, por lo que ante este tipo de circunstancias, el sistema no acierta a reconocerlos. Por lo tanto, se está de acuerdo con el sistema, al considerar que este tipo de eventos es muy difícil de determinar, porque están

contaminados por mucho ruido de fondo. En los espectrogramas anteriores, puede verse que no se distinguen los eventos dentro del ruido de fondo. Por lo que se procede a eliminar de la base de datos los 12 registros que han resultado por debajo del 60%, y se repite el experimento. El resultado obtenido se observa en la tabla 5.9.

```

===== HTK Análisis de Resultados =====
      SENT: %Correct=79.26 [H=214, S=56, N=270]
      WORD: %Corr=94.07, %Acc=93.29 [H=968, D=59, S=2, I=8, N=1029]
----- Matriz de Confusión -----

```

	NE	TB
NE	619	0
TB	2	349
Ins	8	0
Del	21	38

Tabla 5.9 Resultados obtenidos al eliminar registros con bajo porcentaje del Etna. Utilizando 6 Iteraciones, 13 estados, 32 Gaussianas y factor $-p = -15$

%Corr = Porcentaje correcto de etiquetas reconocidas, %Acc = Porcentaje de precisión
H= Número correcto de etiquetas, D = Número de eliminaciones, S = Número de sustituciones,
I = Número de Inserciones, N = Número total de etiquetas. Ins = Inserciones, Del = Borrados

Si se observa la tabla anterior, con solo 12 registros de diferencia, y a pesar de que hay menos eventos para entrenar, se han obtenido mejores resultados al eliminar de la base de datos, aquellos registros con menor porcentaje de efectividad en el reconocimiento. El porcentaje de efectividad aumentó hasta un 93.29 %, y los elementos eliminados han disminuido a solo 59, las inserciones han sido solo 8. De los porcentajes iniciales en la tabla 5.1 de 91.24 % Corr y 71.68% de Acc, se ha mejorado a 94.07% de Corr y 93.29% de Acc., lo que indica que después del todo el proceso de mejoras y al haber eliminado los registros muy contaminados de ruido, se posee una base de datos mucho más consistente y por lo mismo se ha podido obtener un sistema de entrenamiento y reconocimiento fiable y efectivo.

f) Pruebas abiertas (test ciegos)

En cuanto a las pruebas abiertas, o los llamados “test ciegos”, se van a realizar teniendo ya en cuenta todos los parámetros anteriormente detallados en los experimentos. En este tipo de pruebas se utiliza el fichero de decodificación “decode2.csh” (ver apéndice H), que es el que marca la pauta del reconocimiento. Dichas pruebas van a estar constituidas por dos tipos de experimentos. El primero de ellos se basa en dividir en dos partes la base de datos de entrenamiento, la primera parte la formarán modelos entrenados, para reconocer a la parte restante. El segundo test, es el denominado “ciego-ciego”, en este tipo de test se pasará la base de datos entrenada (de forma completa), a registros no segmentados para observar si reconoce algún tipo de eventos.

i) Test ciego con base de datos entrenada dividida

Para el primer test, de los 270 registros que conforman la base de datos, se han tomado 200 para la parte de entrenamiento y 70 son los que se van a pasar por el reconocedor. El resultado del primer test, se observa en la siguiente tabla.

```

===== HTK Análisis de Resultados =====
SENT: %Correct=68.57 [H=48, S=22, N=70]
WORD: %Corr=91.13, %Acc=89.72 [H=257, D=24, S=1, I=4, N=282]
----- Matriz de Confusión -----
    
```

	NE	TB
NE	160	0
TB	1	97
Ins	3	1
Del	14	10

Tabla 5.10 Resultados obtenidos en el primer tipo de las pruebas abiertas con el Etna
 %Corr = Porcentaje correcto de etiquetas reconocidas, %Acc = Porcentaje de precisión
 H= Número correcto de etiquetas, D = Número de eliminaciones, S = Número de sustituciones,
 I = Número de Inserciones, N = Número total de etiquetas. Ins = Inserciones, Del = Borrados

En la tabla anterior se observa que en las pruebas abiertas, el sistema, de los 70 registros ha reconocido 48 y ha realizado sustituciones en 22 de ellos. Además de 282 eventos, reconoce 257, inserta cuatro, sustituye uno y borra 24. El porcentaje correcto es de un 91.13%, siendo el porcentaje de afectividad de un 89.72%. El sistema ha reconocido la casi totalidad de los eventos, con lo que representa un éxito.

ii) Test ciego-ciego

El segundo tipo de las pruebas abiertas, el test “ciego-ciego”, los 270 registros que constituyen la base de datos de entrenamiento van a reconocer 566 registros adicionales no parametrizados, sin segmentar, la mayoría de ellos conteniendo un fuerte ruido de fondo. El sistema como parámetro de salida ha creado un fichero conteniendo los registros etiquetados. A manera de ejemplo, se han tomado para ser analizados 6 registros de dicho fichero, en donde el sistema reconoce 1, 2 y 3 eventos de tipo TB en cada registro, a continuación se muestra un trozo del resultado en la figura 5.6.

```

Test Ciego-Ciego En el Etna
1 evento tipo TB encontrado
**/2541335d.rec"
0 5000000 NE -4801.775391
5000000 7200000 NE -2188.465332
7200000 14650000 TB -7744.709473

**/2541411d.rec"
0 7450000 NE -7222.188965
7450000 11450000 TB -3883.120361
11450000 14650000 NE -3107.162109

2 eventos tipo TB encontrados
**/2541208d.rec"
0 1800000 NE -1729.400146
1800000 5550000 TB -3810.816895
5550000 7500000 TB -1805.125488
7500000 12950000 NE -5190.973633
12950000 14650000 NE -1673.211670

**/2541223d.rec"
0 1650000 NE -1696.205322
1650000 9850000 TB -7994.345215
9850000 14650000 TB -4826.533203

3 eventos tipo TB encontrados
**/2541908d.rec"
0 3450000 TB -3054.274414
3450000 8400000 NE -4804.175781
8400000 13700000 TB -4993.143066
13700000 14650000 TB -859.768921

**/2550208d.rec"
0 1700000 NE -1821.158447
1700000 6050000 TB -4278.994629
6050000 9650000 TB -3421.294189
9650000 14650000 TB -4959.630371
    
```

Fig. 5.6 Ejemplo de seis registros, en el reconocimiento de eventos tipo TB, por parte del sistema durante las pruebas o test ciego-ciego. En los primeros dos registros el sistema reconoce un evento, en los siguientes dos, el sistema reconoce dos eventos y en los últimos dos, el sistema reconoce tres eventos de tipo TB.

La figura anterior indica que el sistema ha reconocido 1, 2 y 3 eventos de tipo TB en los registros. Lo anterior se puede observar gráficamente mediante la representación de la señal en el dominio del tiempo y su espectrograma. Las imágenes de dichos registros se observan en las figuras 5.7 a 5.9.

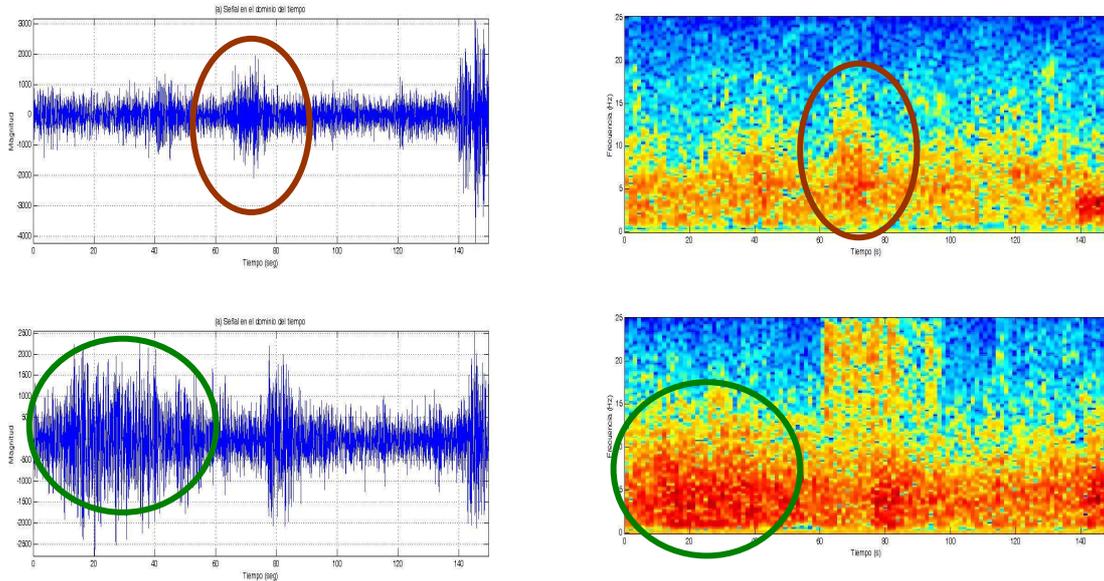


Fig. 5.7 Ejemplo de dos registros en el que el sistema ha reconocido un evento de tipo TB, en el test ciego-ciego

En la figura anterior se observa como el sistema reconoce claramente un evento de tipo TB, aunque en el segundo registro (circulo verde) podría estar compuesto de más de un registro, lo que se observa en los detalles de la imagen de la transformada de Fourier y el espectrograma, el sistema lo ha reconocido como un único evento con las mayores características espectrales. En cuanto a los registros en los que el sistema ha reconocido dos eventos de tipo TB, las gráficas de dichos registros se observan en la figura 5.8.

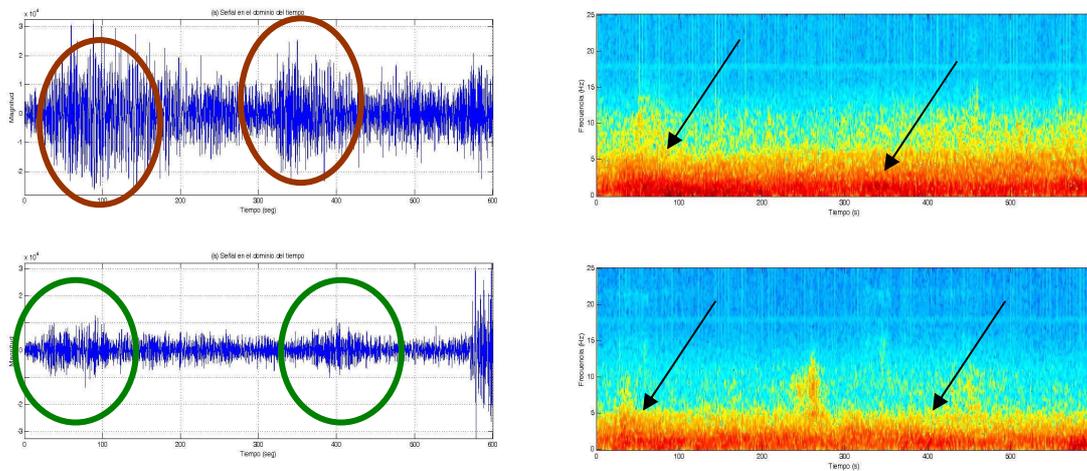


Fig. 5.8 Ejemplo de dos registros en el que el sistema ha reconocido dos eventos de tipo TB, en el test ciego-ciego

La figura anterior muestra los dos eventos que ha podido reconocer el sistema. En el segundo registro (círculo rojo), los eventos están muy enmascarados con el ruido de fondo, sin embargo, el sistema ha sido capaz de poder reconocer eventos en estas circunstancias. El ejemplo de las gráficas de los registros en los cuales el sistema ha podido reconocer tres eventos de tipo TB se observa en la figura 5.9.

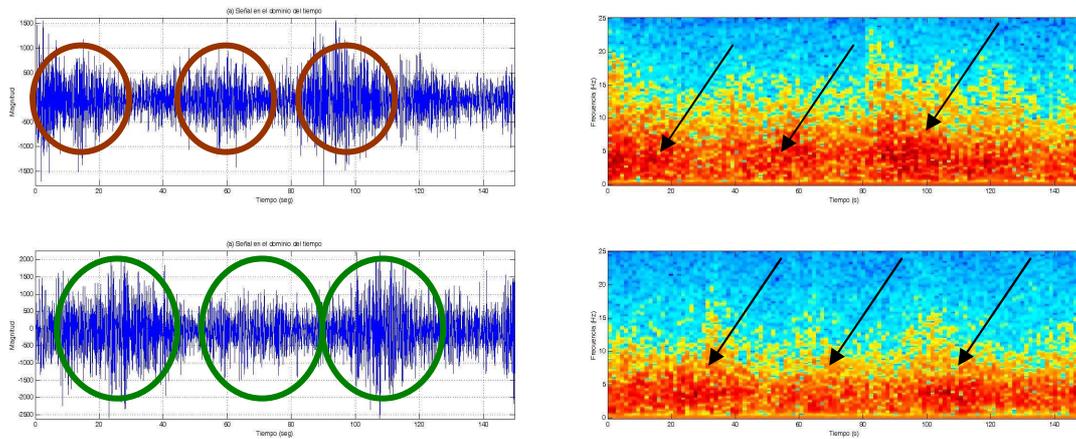


Fig. 5.9 Ejemplo de dos registros en el que el sistema ha reconocido tres eventos de tipo TB, en el test ciego-ciego

Pueden observarse en los registros anteriores los tres eventos de tipo TB reconocidos por el sistema. Sin embargo, puede observarse en el segundo registro (círculos verdes), lo complicado que es delimitar este tipo de evento, en el tremor de fondo que lo acompaña. En este sentido, en cuanto al ruido, el sistema también ha reconocido en los registros eventos de tipo ruido o NE. Un segmento de esto se observa al seleccionar dos de los registros en lo que el sistema catalogó o etiquetó como NE.

Registros NE, en el test ciego-ciego del Etna

```

**/2541159d.rec"
0 11350000 NE -12364.479492
11350000 14650000 NE -3329.124512

**/2550159d.rec"
0 58500000 NE -5524.047363
58500000 146500000 NE -8443.028320
    
```

Fig. 5.10 Ejemplo de dos registros del fichero de resultado, en el reconocimiento de eventos tipo NS, por parte del sistema en el test ciego-ciego

Esto se observa gráficamente en la figura 5.11.

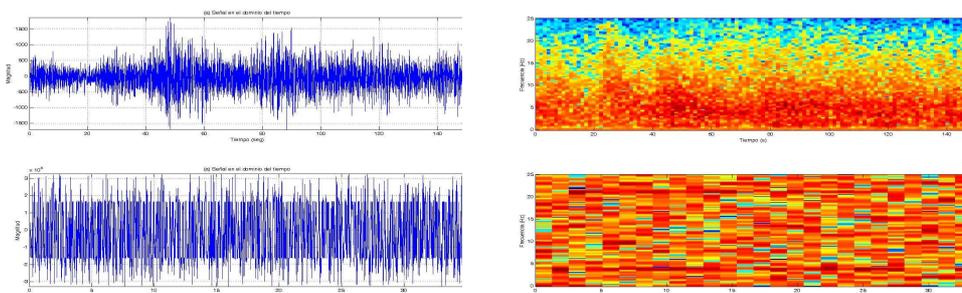


Fig. 5.11 Ejemplo de registros en el que el sistema ha catalogado los registros como de tipo NE, en el test ciego-ciego

En la figura 5.11 anterior, el segundo registro (en la parte inferior) se observa claramente que está conformado por mucho ruido de fondo. Las gráficas de la transformada de Fourier en el dominio del tiempo, así como el espectrograma, no permiten identificar ningún tipo diferente de evento además del ruido. El sistema discrimina muy bien este tipo de registro. En cambio, en el primer registro (en la parte superior), aunque se encuentra muy contaminado por ruido de fondo, podría contener al menos dos tipos de eventos diferentes. Pero para ello sería necesario contar con una base de datos mucho más amplia en la que se lleguen a segmentar muchos más registros de este tipo, para que el sistema pudiera reconocerlos. Pruebas adicionales a futuro con herramientas complementarias (Ver apéndices D y E), podrían definir aun más los resultados. El sistema ha podido reconocer con éxito la mayoría de los eventos presentes. Sin embargo, aún resta afinar, para que no pueda haber alguna duda respecto a la existencia de un cierto tipo de evento. Habría que recordar que el tipo de evento tremor, como se mencionó al inicio de este trabajo, es uno de los más difíciles en reconocer. A continuación se procederá a realizar las pruebas con la base de datos del Strómboli, para observar que resultados producen. Hay que recordar que en la base de datos del Strómboli, el evento a reconocer son las explosiones típicas del Strómboli marcadas como “Exp”, más el ruido de fondo, que ha sido designado como “Sil”.

5.2.2.- Análisis de resultados en el Strómboli

Para las pruebas iniciales en el volcán Strómboli, se ha utilizado una base de datos compuesta de 384 registros segmentados, posteriormente se aumentó el número de segmentaciones hasta llegar a las 495. De los primeros 284 registros segmentados, han sido etiquetados 310 eventos tipo explosión (exp), y 671 de ruido de fondo (sil). Para los primeros ensayos, los parámetros en los ficheros de configuración son similares a los iniciales utilizados en el Etna (anteriormente tratados). Los resultados de las primeras pruebas con dichos parámetros en el reconocimiento, se muestran en la tabla 5.11.

```

===== HTK Análisis de Resultados =====
SENT: %Correct=52.08 [H=200, S=184, N=384]
WORD: %Corr=76.76, %Acc=71.36 [H=753, D=225, S=3, I=53, N=981]
----- Matriz de Confusión -----

```

	Sil	Exp
Sil	557	1
Exp	2	196
Ins	45	8
Del	111	114

Tabla 5.11 Primeros experimentos con 11 estados, 6 iteraciones y 16 Gaussianas en el Strómboli
 %Corr = Porcentaje correcto de etiquetas reconocidas, %Acc = Porcentaje de precisión
 H= Número correcto de etiquetas, D = Número de eliminaciones, S = Número de sustituciones,
 I = Número de Inserciones, N = Número total de etiquetas. Ins = Inserciones, Del = Borrados

En la tabla anterior se observa que el sistema ha reconocido con efectividad 200 registros de los 384, de 981 eventos ha reconocido satisfactoriamente 753, ha eliminado 225 y ha sustituido 2. Además, ha insertado 53 eventos adicionales. El porcentaje de certeza Corr es de 76.76% y el de efectividad se encuentra en 71.36%. En estos experimentos, el sistema ha confundido dos de las explosiones con el ruido de fondo, además ha insertado 8 eventos de tipo explosión y 45 de ruido de fondo. Los eventos de

tipo “exp” eliminados por el sistema han sido 114 y 111 de ruido de fondo. En el Strómboli debido a que existe un menor ruido de fondo (ver Fig. 3.17) que en el Etna se pueden localizar mucho mejor los eventos de tipo explosión, lo que tratándose del Etna era más difícil ya que los eventos tipo TB se enmascaraban con el ruido de fondo, por lo que en el Strómboli se pueden aumentar el número de registros, debido a que se encuentran muchos más eventos de tipo “Exp” en los registros.

Se ha procedido por lo tanto, a realizar una nueva segmentación, con el fin de incrementar el número de eventos y por ende ver si el porcentaje de efectividad del sistema también aumenta. De esta forma, se han incrementado el número de registros segmentados de 384 a 495. El número de los eventos de tipo explosión (exp), aumentó a 385 y el número de de ruido de fondo (NE) quedó en 849 segmentaciones. El resultado de la prueba con los mismos parámetros iniciales se muestra en la tabla 5.12.

```

===== HTK Análisis de Resultados =====
      SENT: %Correct=58.38 [H=294, S=201, N=495]
WORD: %Corr=90.44, %Acc=78.93 [H=1116, D=107, S=11, I=142, N=1234]
----- Matriz de Confusión -----
    
```

	Sil	Exp
Sil	749	5
Exp	6	367
Ins	111	31
Del	94	13

Tabla 5.12 Resultados obtenidos del HTK en los experimentos con 11 estados, 6 iteraciones y 16 Gausianas realizados con el Strómboli, al aumentar los registros segmentados.

%Corr = Porcentaje correcto de etiquetas reconocidas, %Acc = Porcentaje de precisión
 H= Número correcto de etiquetas, D = Número de eliminaciones, S = Número de sustituciones,
 I = Número de Inserciones, N = Número total de etiquetas. Ins = Inserciones, Del = Borrados

En estas pruebas, el sistema, de los 495 registros, ha reconocido 294 que conforman la base de datos. En cuanto a los eventos, el sistema, de 1,234 eventos ha reconocido satisfactoriamente 1,116, eliminado 107, sustituido 11 e insertado 142 eventos adicionales. El porcentaje de aciertos (Corr) es de 90.44% y el de efectividad se sitúa en un 78.93%, lo que representa un 7.57% más que las pruebas con menos registros, por lo que se deduce que al aumentar el número de registros y de eventos, aumenta el valor de efectividad en el reconocimiento por parte del sistema. Otro punto a resaltar es que la mayoría de las inserciones y los borrados de eventos por parte del sistema son de ruido de fondo. El porcentaje de efectividad se sitúa en estas pruebas en torno al 82% Teniendo en cuenta que la mayoría de eventos que han sido eliminados por el sistema, representan ruido de fondo, podrían despreciarse por lo que se puede afirmar, que el porcentaje en cuanto a efectividad real aumentaría en un orden de un 5% más y sería en teoría, cercano al 83% para estas pruebas. Hay que recordar que en cada uno de los volcanes se va a realizar un proceso de mejora al modificar los parámetros de configuración, a fin de optimizar los resultados hasta ahora obtenidos. El primer grupo de modificación a realizar es el cambio en el número de iteraciones y realizar pruebas con 3, 6, 9, 12 y 15 iteraciones.

a) Cambio en el número de iteraciones

El resultado de las pruebas, con el cambio del número de iteraciones en el Strómboli se observa en la tabla 5.13., en amarillo se presenta la fila con mejores resultados.

Iteraciones	%Corr	%Acc	Sust	Ins Exp	Del Exp	Ins sil	Del sil
3	89.71	77.07	10	30	19	126	98
6	90.44	78.93	11	31	13	111	94
9	90.76	78.69	10	29	16	120	88
12	91.25	79.50	10	28	15	117	83
15	91.33	79.09	5	30	17	121	85

Tabla 5.13 Resultados obtenidos en las pruebas cerradas del HTK, con 11 estados y 16 Gaussianas al modificar el parámetro del número de iteraciones en el Strómboli

%Corr = Porcentaje correcto de etiquetas reconocidas, %Acc = Porcentaje de precisión

Ins = Inserciones, Sust = Sustituciones, Del = Borrados

Los resultados entre 6 y 9 iteraciones en los porcentajes de efectividad no varía mucho, hay menos eventos borrados de tipo Exp, menos sustituciones, en cuanto al número de inserciones no difiere mucho de una prueba a otra, además de que el que se inserte elementos no necesariamente represente un error, sino un acierto del sistema, al reconocer eventos que el observador no ha marcado. Por lo que al igual que en el Etna, los resultados indican que se sigue optando por las 6 iteraciones, para tener un balance entre la calidad y el tiempo de cómputo.

b) Cambio en el número de Gaussianas

Ahora se ha procedido a realizar un cambio en el número de Gaussianas empleadas en los modelos. Se establece la modificación del parámetro del número de Gaussianas a utilizar (igual que en el Etna), partiendo de 8 y sumando ocho unidades en cada prueba. Desde este punto, y vistos los resultados en el Etna, el número de Gaussianas solo se aumentará hasta 48, ya que determinar un número mayor se ha comprobado por una parte que incrementa el porcentaje de precisión, sin embargo, por otra parte y como se observó en el Etna, representa un coste computacional en tiempo de proceso bastante alto, por lo que un valor representativo y que estaría aquí caracterizado sería en torno a las 32 gaussianas, que es donde se ha llegado con los experimentos en el Etna. Por lo tanto, en los ensayos siguientes (con los tres volcanes restantes), solo se aumentará dicho número hasta 48 en lugar de hasta 96, como se hizo en el Etna. Los valores que se han encontrado en dichos experimentos se observan en la tabla 5.14., en amarillo se presenta la fila con mejores resultados

Número de Experimento	No Gausianas	% Corr. obtenido	% Acc. obtenido
1	8	88.65	73.10
2	16	90.44	78.93
3	24	91.25	81.28
4	32	91.98	83.87
5	40	92.87	85.49
6	48	93.27	88.41

Tabla 5.14 Resultados obtenidos en las pruebas cerradas del HTK, con 6 iteraciones, 11 estados, y factor $-p$ (0.0), al modificar el parámetro del número de gaussianas en el Strómboli
 %Corr = Porcentaje correcto de etiquetas reconocidas, %Acc = Porcentaje de precisión

Un análisis gráfico de la variación del porcentaje respecto al aumento del número de gaussianas en los experimentos lo observamos en las figuras 5.12 y 5.13

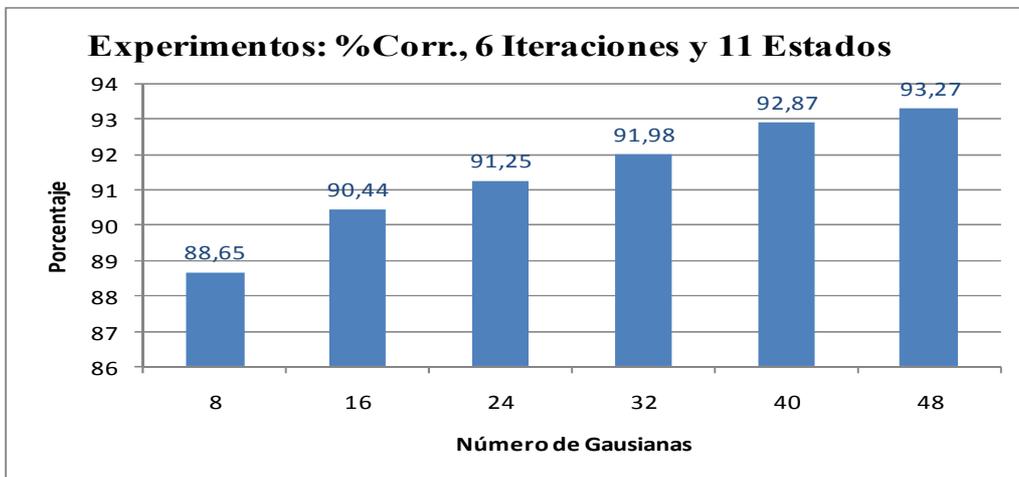


Fig. 5.12 Resultados obtenidos en %Corr en las pruebas cerradas del HTK, en el porcentaje de aciertos al modificar el parámetro del número de gaussianas en el Strómboli.

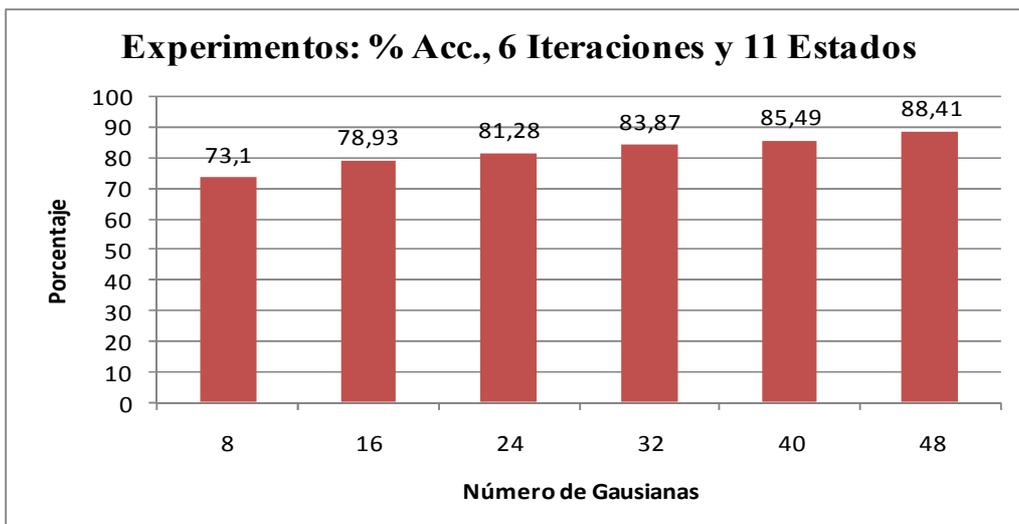


Fig. 5.13 Resultados obtenidos en % Acc en las pruebas cerradas del HTK, en el porcentaje de precisión al modificar el parámetro del número de gaussianas en el Strómboli.

Se observa en las gráficas anteriores que los porcentajes de aciertos y efectividad, presentan la misma tendencia que en las pruebas con el Etna. La diferencia de porcentaje entre 32 y 40 Gaussianas es mínima, aunque el costo en tiempo al aumentar a 40 supera al de 32 en varias horas, por lo que se ha determinado para las siguientes pruebas (al igual que en el Etna), fijar el número de gaussianas en 32.

c) Cambio en el factor de penalización

Las pruebas con la modificación del factor de penalización presentan los resultados que se observan en la tabla 5.15., en amarillo se presenta la fila con mejores resultados

Factor – p	%Corr	%Acc	Sust	Ins Exp	Del Exp	Ins Sil	Del Sil
15	94.33	76.82	4	26	6	190	60
10	93.27	80.06	4	24	7	139	72
5	92.46	81.77	4	23	10	109	79
0	91.98	83.87	4	18	13	82	82
-5	91.49	84.52	4	18	14	68	87
-10	90.92	85.25	4	17	16	53	92
-15	90.11	85.74	5	16	21	38	96

Tabla 5.15 Resultados obtenidos con 6 iteraciones, 11 estados y 32 Gaussianas, en la modificación del parámetro “-p” del Hvite en el Strómboli

%Corr = Porcentaje correcto de etiquetas reconocidas, %Acc = Porcentaje de precisión
Ins = Inserciones, Sust = Sustituciones, Del = Borrados

Como se puede observar en la tabla anterior, el mayor porcentaje de efectividad ha sido el correspondiente a las pruebas con el factor en p -15, al igual que las pruebas realizadas en el Etna, al aumentar el factor, el porcentaje correcto (Corr) aumenta, pero disminuye la efectividad o precisión (Acc) del sistema y al reducir el factor de penalización (-p), el porcentaje (Corr) disminuye, pero aumenta el porcentaje de (Acc), esto se debe a que hay menos inserciones, pero más eliminaciones. Así que se tiene que sopesar, cuál de estos dos factores pesa más a la hora de poder evaluar la efectividad global del sistema. Hay que destacar que el porcentaje de eliminaciones de una prueba respecto a la otra es muy bajo, en cambio hay una diferencia bastante significativa en cuanto al número de inserciones de una prueba respecto a la anterior o a la siguiente, así puede asumirse el número de eliminaciones y valorar más el de inserciones. Por estas razones, se determina optar, por las pruebas que posean menores índices de inserciones, aunque tengan un poco mas de eliminaciones. Por lo tanto, el valor elegido a partir de aquí para los siguientes ensayos, ha sido el factor p = -15

a) Cambio en el número de estados

Los resultados de las pruebas con la modificación del número de estados se observa en la tabla 5.16, en amarillo se presenta la fila con mejores resultados.

No. Estados	%Corr	%Acc	Sustit	Exp Ins	Exp Del	Sil Ins	Sil Del
5	82.17	67.18	39	62	27	123	154
7	78.69	68.64	13	37	70	87	180
9	80.39	73.10	9	26	55	64	178
11	90.11	85.74	5	16	21	38	96
13	84.85	82.33	3	5	34	26	150
15	92.30	89.14	4	12	31	27	60
17	92.46	88.98	2	13	28	30	63
19	91.73	88.49	2	11	28	29	72

Tabla 5.16 Resultados al modificar el parámetro del número de estados en el Strómboli, obtenidos en las pruebas cerradas del HTK, con 6 iteraciones, 32 gaussianas y factor $-p$ en -15
 %Corr = Porcentaje correcto de etiquetas reconocidas, %Acc = Porcentaje de precisión
 Ins = Inserciones, Sust = Sustituciones, Del = Borrados

En los resultados de la tabla anterior se observa, a diferencia del Etna, que el porcentaje de efectividad es mejor en las pruebas con 15 estados de las que se realizan con 11 y 13, el número de inserciones y borrados de los eventos de tipo Exp es bastante menor que en las dos pruebas con 11 y 13, esto hace indicar que la pauta a seguir es asignar las pruebas siguientes con 15 estados.

En este punto se determina el hito de repetir las pruebas observadas en la tabla 5.14, para observar los cambios que se obtienen en relación a los parámetros ya establecidos como mejoras. Así, los valores a configurar son: número de iteraciones de 6, número de estados en 15, factor de penalización de la herramienta de entrenamiento, Hvite $-p$ en (-15). La tabla 5.17, representa los valores obtenidos en dichas pruebas al realizar este nuevo grupo de experimentos, en donde se comprueba el incremento de la efectividad del sistema. En amarillo se presenta la fila con mejores resultados.

Número de Experimento	No Gaussianas	% Corr. obtenido	% Acc. obtenido
1	8	85.09	77.31
2	16	87.84	82.09
3	24	91.00	86.79
4	32	92.30	89.14
5	40	94.25	91.57
6	48	94.89	92.79

Tabla 5.17 Resultados obtenidos en las pruebas cerradas del HTK, con 6 iteraciones, 15 estados, y factor $-p$ (-15), al modificar el parámetro del número de gaussianas en el Strómboli.
 %Corr = Porcentaje correcto de etiquetas reconocidas, %Acc = Porcentaje de precisión

Al comparar los resultados de la tabla anterior con los de la tabla 5.14, se observan las mejoras en los modelos de entrenamiento con 13 estados. El porcentaje de efectividad (Acc) va en aumento con cada prueba, lo que indica que el sistema ha mejorado al implementar el cambio del número de estados, conjuntamente con los parámetros ya establecidos. En las 32 Gaussianas, el porcentaje de precisión (Acc) ha mejorado de un 83.87% a un 89.14 %, lo que representa algo más de un 5% de aumento. Esto indica que en las siguientes pruebas, los experimentos han de tomarse con 13 estados. Los resultados de las pruebas finales de los modelos de entrenamientos en las

pruebas cerradas con 32 Gaussianas en el Strómboli (en amarillo en la tabla 5.17) se puede observar en la tabla 5.18

```

===== HTK Análisis de Resultados =====
          SENT: %Correct=76.77 [H=380, S=115, N=495]
          WORD: %Corr=92.30, %Acc=89.14 [H=1139, D=91, S=4, I=39, N=1234]
          ----- Matriz de Confusión -----
  
```

	Sil	Exp
Sil	785	3
Exp	1	354
Ins	27	12
Del	60	31

Tabla 5.18 Resultados obtenidos en las pruebas del Strómboli. Utilizando 6 Iteraciones, 15 estados, 32 Gaussianas y factor $-p = -15$
 %Corr = Porcentaje correcto de etiquetas reconocidas, %Acc = Porcentaje de precisión
 H= Número correcto de etiquetas, D = Número de eliminaciones, S = Número de sustituciones,
 I = Número de Inserciones, N = Número total de etiquetas. Ins = Inserciones, Del = Borrados

La tabla anterior muestra que el sistema, en los 495 registros segmentados, ha logrado un porcentaje de aciertos correctos (Corr) de 92.30% y una precisión (Acc) de 89.14%. De 1,234 eventos, el sistema ha reconocido 1,139, ha eliminado 91, sustituido 4 e insertado 39. Hasta ahora, cada prueba ha demostrado, que se ha podido mejorar el porcentaje de precisión en el reconocimiento de los eventos por parte del sistema.

e) Proceso de depurado selectivo de los registros

Una vez obtenidos los resultados finales con las modificaciones de parámetros en el proceso de mejoras, el siguiente y último paso en dicho proceso, es observar los resultados obtenidos en cada experimento de forma manual. El fin es determinar cuál de los registros están obteniendo muy bajos porcentajes de reconocimiento y comprobar si pueden ser eliminados de la base de datos de entrenamiento. Un fragmento de dicho fichero se observa en la figura 5.14.

```

Proceso de Depurado Selectivo en el Stromboli
----- Sentence Scores -----
----- HTK Results Analysis -----
Rec : rec.11.32.gram.mlf
----- File Results -----
2541746G.rec: 100.00(100.00) [H= 1, D= 0, S= 0, I= 0, N= 1]
2541753G.rec: 100.00(100.00) [H= 3, D= 0, S= 0, I= 0, N= 3]
2541806G.rec: 100.00(100.00) [H= 3, D= 0, S= 0, I= 0, N= 3]
2542327G.rec: 100.00(-100.00) [H= 1, D= 0, S= 0, I= 2, N= 1]
Aligned transcription: 2542327G.lab vs 2542327G.rec
LAB: sil exp sil
REC: sil exp sil
2561948G.rec: 40.00( 40.00) [H= 2, D= 3, S= 0, I= 0, N= 5]
Aligned transcription: 2561948G.lab vs 2561948G.rec
LAB: sil exp sil exp sil
REC: sil exp sil
2550607G.rec: 100.00(100.00) [H= 1, D= 0, S= 0, I= 0, N= 1]
2550618G.rec: 100.00(100.00) [H= 4, D= 0, S= 0, I= 0, N= 4]
2550630G.rec: 100.00(100.00) [H= 3, D= 0, S= 0, I= 0, N= 3]
2571518G.rec: 100.00( 0.00) [H= 1, D= 0, S= 0, I= 1, N= 1]
Aligned transcription: 2571518G.lab vs 2571518G.rec
LAB: sil exp sil
REC: sil exp sil
2580441G.rec: 100.00(100.00) [H= 1, D= 0, S= 0, I= 0, N= 1]
2580446G.rec: 100.00(100.00) [H= 2, D= 0, S= 0, I= 0, N= 2]
2580452G.rec: 100.00(100.00) [H= 3, D= 0, S= 0, I= 0, N= 3]
2580504G.rec: 100.00(100.00) [H= 3, D= 0, S= 0, I= 0, N= 3]
2580525G.rec: 100.00(100.00) [H= 3, D= 0, S= 0, I= 0, N= 3]
  
```

Fig. 5.14 Ejemplo del Análisis manual en el Proceso selectivo (evento por evento), de los resultados obtenidos en el Strómboli. Los valores en rojo representan los registros que han tenido un porcentaje de 100 y que no se cambiarán. Los valores en verde son aquellos que han tenido un porcentaje menor a 60 y que se analizarán para eliminarse de la base de datos.

Como se puede observar en la figura anterior, algunos de los registros presentaban porcentajes inferiores al 60% de acierto, incluso con valores negativos respecto al valor etiquetado por el observador. Un análisis gráfico de dichos registros se observa en las siguientes figuras.

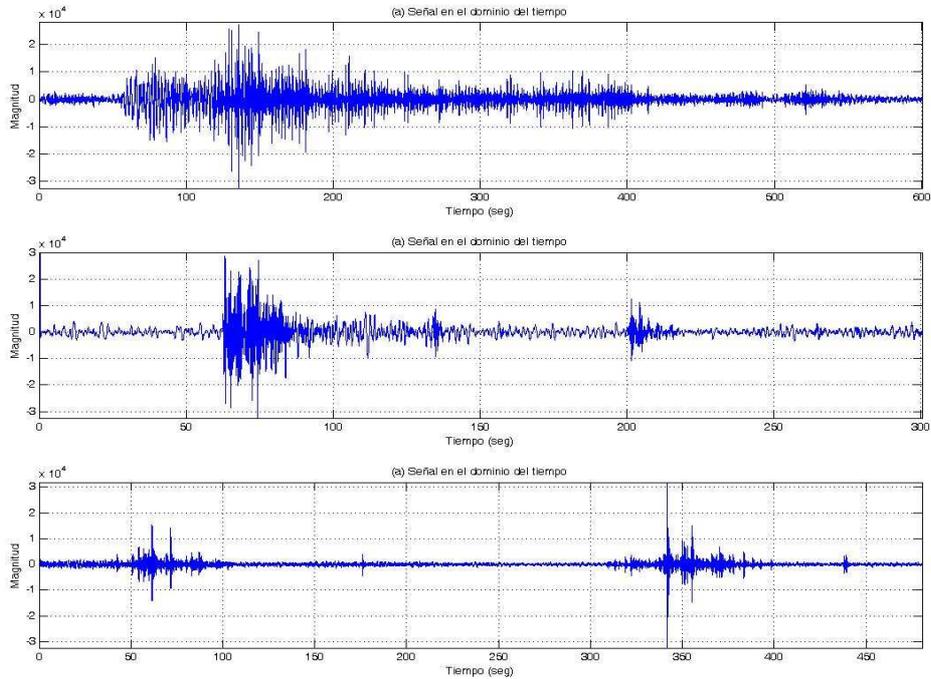


Fig. 5.15 Registros 2542327G (arriba), 2561948G (centro) y 2571518G (abajo), en el dominio del tiempo

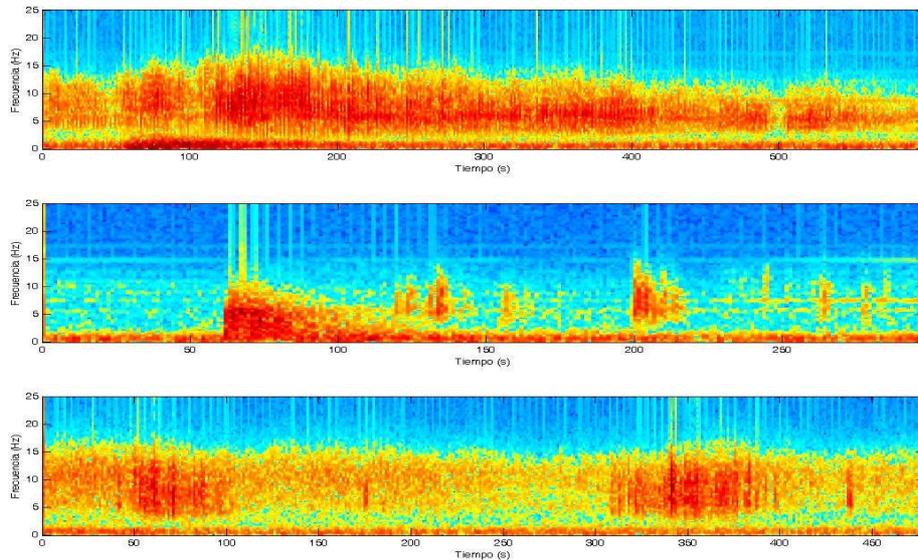


Fig. 5.16 Espectrogramas de los registros: 2542327G (arriba), 2561948G (centro) y 2571518G (abajo)

Al analizar los registros que presentaban los menores porcentajes de acierto, se observa que dichos registros presentan mucho ruido de fondo, causado posiblemente por defectos a la hora del almacenamiento del registro, debido a alguna deficiencia del

sensor. Algunos como el que se observa en la parte inferior, ha tenido un error durante la fase de grabación, por lo que el ruido ha quedado en frecuencias altas y no se puede definir eventos de ningún tipo sino ruido. Otros en cambio, como en el registro que se muestra en la parte superior de las figuras, en donde el evento está muy enmascarado con ruido de fondo en frecuencias altas. En la parte central el observador ha indicado la presencia de dos eventos, pero estos se encuentran muy superpuestos con ruido de fondo, por lo que el sistema los ha eliminado. Se procede a eliminar los registros con menor porcentaje de acierto de la Base de Datos del Strómboli y se repiten las pruebas. El número total de registros eliminados fue de 72. El resultado de la eficiencia del sistema ha aumentado, debido a que dichos registros estaban alterando los resultados reales del sistema al tener tan bajo porcentaje de acierto, esto se observa en la tabla 5.19.

===== HTK Análisis de Resultados =====
 SENT: %Correct=81.52 [H=353, S=80, N=433]
 WORD: %Corr=93.39, %Acc=92.32 [H=1046, D=70, S=4, I=12, N=1120]
 ----- Matriz de Confusión -----

	Sil	Exp
Sil	709	1
Exp	3	337
Ins	10	2
Del	53	17

Tabla 5.19 Resultados obtenidos al eliminar registros con bajo porcentaje del Strómboli. Utilizando 6 Iteraciones, 15 estados, 32 Gausianas y factor $-p = -15$
 %Corr = Porcentaje correcto de etiquetas reconocidas, %Acc = Porcentaje de precisión
 H= Número correcto de etiquetas, D = Número de eliminaciones, S = Número de sustituciones,
 I = Número de Inserciones, N = Número total de etiquetas. Ins = Inserciones, Del = Borrados

Como se puede observar en los resultados al realizar las modificaciones, el número total de registros pasó de 495 a 433. Aun así en las pruebas con 32 Gausianas de la tabla anterior, el sistema ha mejorado y de 1,120 elementos, ha podido reconocer con certeza 964, ha eliminado 70 eventos, de los cuales 53 representan ruido de fondo, solamente ha sustituido a 4 de ellos, y ha realizado 12 inserciones. El porcentaje de efectividad mejoró a un 92.32 %, en relación al 89.14 % anterior, antes de quitar los registros, esto representa casi un 4% más que antes de realizar la modificaciones. Teniendo en cuenta que la mayoría de eventos que han sido eliminados por el sistema corresponden a ruido, y que la mayoría de las inserciones en realidad son éxitos en donde el sistema ha podido reconocer eventos, hace incrementar aproximadamente un 5% más el porcentaje de precisión, lo que daría un porcentaje real cercano al 97% de efectividad en las pruebas del Strómboli. Aun así, el 93.39% de porcentaje correcto de etiquetas reconocidas (Corr) y de 92.32% de precisión (Acc) es mucho mayor que los porcentajes iniciales representados en la tabla 5.11, que marcaban porcentajes de certeza (Corr) de 76.76% y de efectividad o precisión (Acc) de 71.36%., por lo que se concluye que las pruebas cerradas en el Strómboli han tenido éxito.

f) Pruebas abiertas (test ciegos)

Las pruebas abiertas, o los llamados “test ciegos”, igual que en el Etna, se van a realizar teniendo todos los parámetros conformados en las pruebas anteriores. Las pruebas van a estar constituidas por los dos tipos de experimentos antes descritos en el

Etna. El primero de ellos será dividir en dos partes la base de datos. Una primera parte, la formarán los modelos entrenados, para pasarlo después a la parte restante. El segundo test “ciego-ciego”, se pasará la base de datos entrenada, a una parte de registros no segmentados para observar que tipo de eventos puede reconocer el sistema.

i) Test ciego con base de datos entrenada dividida

Para el primer test, de los 433 registros de la última prueba y que conforman la base de datos, se han tomado 359 para la parte de entrenamiento y 74 son los que se van a pasar por el reconocedor. El resultado del primer test, se observa en la tabla 5.20.

```

===== HTK Análisis de Resultados =====
          SENT: %Correct=85.14 [H=63, S=11, N=74]
          WORD: %Corr=90.15, %Acc=88.64 [H=119, D=13, S=0, I=2, N=132]
----- Matriz de Confusión-----
    
```

	Sil	Exp
Sil	97	0
Exp	0	22
Ins	2	0
Del	3	10

Tabla 5.20 Resultados obtenidos en el primer tipo de las pruebas abiertas con el Strómboli
 %Corr = Porcentaje correcto de etiquetas reconocidas, %Acc = Porcentaje de precisión
 H= Número correcto de etiquetas, D = Número de eliminaciones, S = Número de sustituciones,
 I = Número de Inserciones, N = Número total de etiquetas. Ins = Inserciones, Del = Borrados

En la tabla anterior se observa que en las pruebas abiertas, el sistema de los 74 registros ha reconocido 63 y ha realizado sustituciones en 11 de ellos. Además, de 132 eventos, reconoce 119, inserta dos, no sustituye ningún tipo de evento y elimina 13. El porcentaje correcto es de un 90.15%, siendo el porcentaje de efectividad o precisión (Acc) de un 88.64%. El sistema ha podido reconocer la casi totalidad de los eventos, lo que se considera un éxito.

ii) Test ciego-ciego

El segundo tipo de las pruebas abiertas, el test “ciego-ciego”, los 433 registros que constituyen la base de datos de entrenamiento van a reconocer 497 registros adicionales no parametrizados, sin segmentar. En este caso se ha querido probar como reconocen los modelos que han sido entrenados con los registros de la estación G, a los registros de otra estación, en este caso los de la estación F. El sistema, como resultado ha creado un fichero de salida conteniendo en los registros los eventos que ha clasificado y etiquetado. A manera de ejemplo, se han tomado para ser analizados 6 registros de dicho fichero, en donde el sistema reconoce eventos de tipo Exp. Esto se muestra en la figura 5.17.

Test ciego-ciego en el Stromboli

1 evento tipo Exp encontrado

```
"/2541643F.rec"  
0 2050000 sil -2693.313477  
2050000 4750000 exp -3517.940430  
4750000 10600000 sil -8211.021484  
.  
"/2541653F.rec"  
0 2200000 sil -3035.595459  
2200000 5600000 exp -4636.595703  
5600000 14900000 sil -12878.676758
```

2 eventos tipo Exp encontrados

```
"/2541705F.rec"  
0 2150000 sil -3094.713623  
2150000 4750000 exp -3437.228760  
4750000 6150000 sil -1969.279907  
6150000 9400000 exp -4210.433105  
9400000 14900000 sil -7557.620117  
.  
"/2541949F.rec"  
0 2000000 sil -2842.471680  
2000000 4000000 exp -2693.508057  
4000000 5000000 sil -1388.842773  
5000000 10800000 exp -7917.659668
```

3 eventos tipo Exp encontrados

```
"/2550352F.rec"  
0 1500000 sil -2200.222168  
1500000 6900000 exp -7068.249512  
6900000 12100000 exp -7070.397461  
12100000 14900000 exp -3720.062744  
.  
"/2551910F.rec"  
0 1950000 sil -2752.692139  
1950000 4550000 exp -3600.679688  
4550000 5600000 sil -1596.667358  
5600000 11100000 exp -7386.648438  
11100000 12650000 sil -2315.424805  
12650000 14900000 exp -2968.910645
```

Fig. 5.17 Ejemplo de seis registros, en el reconocimiento de eventos tipo Exp, por parte del sistema en el Strómboli, durante las pruebas o test ciego-ciego

En la figura 5.17 se observa que el sistema ha podido detectar en los registros 1, 2 y 3 eventos de tipo Exp. Se pueden analizar y determinar de forma visual a través de una gráfica, los anteriores registros y comprobar si el sistema ha tenido éxito al reconocer dichos eventos en el registro. En la figura 5.18 se observa la señal en el dominio del tiempo, y los espectrogramas en donde los círculos y las flechas sitúan claramente los eventos.

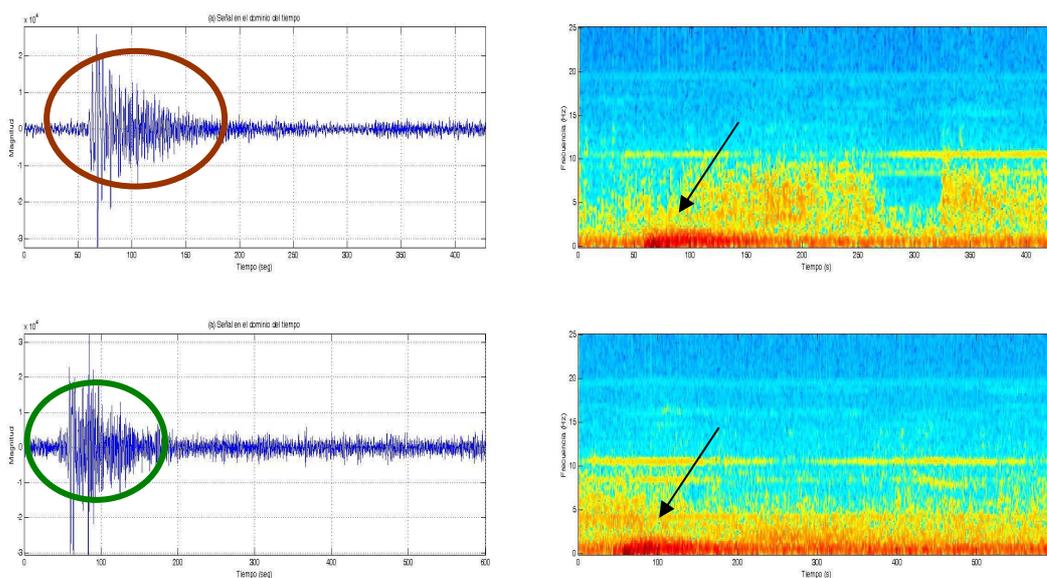


Fig. 5.18 Ejemplo de dos registros en los que el sistema ha reconocido un evento de tipo Exp, en el test ciego-ciego. Los círculos y las flechas indican los eventos.

La figura anterior muestra como el sistema reconoce un evento de tipo Exp, tanto en el primero (círculo rojo) como en el segundo registro (círculo verde). En los detalles de la imagen de la señal en el dominio del tiempo y el espectrograma, el sistema ha reconocido un único evento. Las gráficas de los registros (ver figura 5.17) donde el sistema ha reconocido dos eventos de tipo Exp, se aprecian en la figura 5.19.

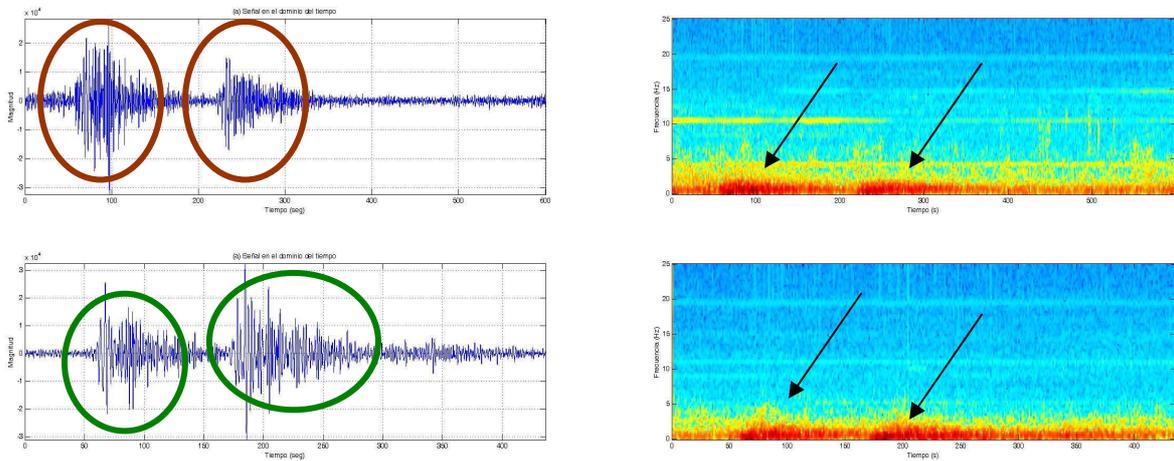


Fig. 5.19 Ejemplo de dos registros en el test ciego-ciego, en el que el sistema ha reconocido dos eventos de tipo Exp. Los círculos y las flechas indican los eventos.

La figura anterior muestra los dos eventos que ha podido reconocer el sistema. En el segundo registro (círculo rojo y verde), tanto en la señal en el dominio del tiempo como los espectrogramas, pueden diferenciarse. Así se observa como los eventos están muy bien delimitados respecto al ruido de fondo, por lo que el sistema ha sido capaz de poder reconocer este tipo de eventos. Las gráficas de los registros en los cuales el sistema ha podido reconocer tres eventos de tipo Exp se observa en la figura 5.20.

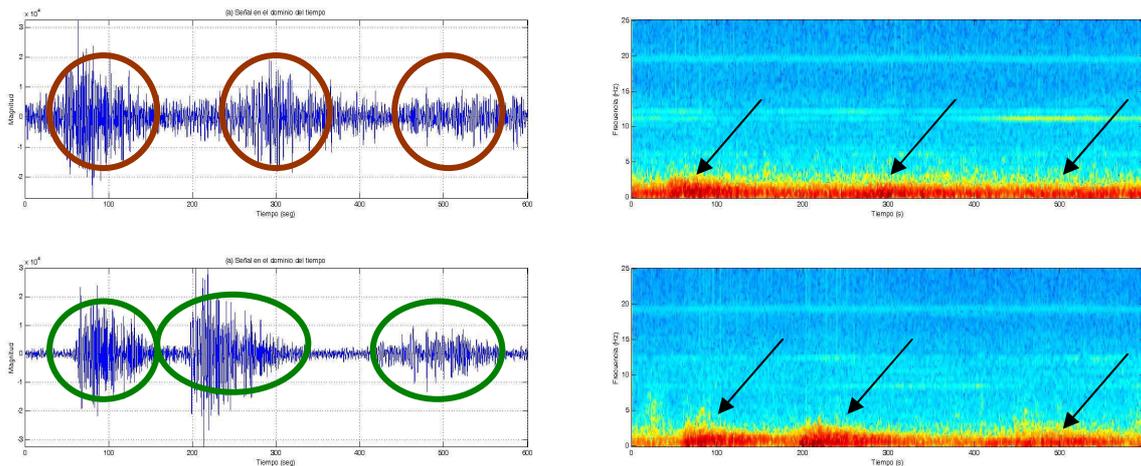


Fig. 5.20 Ejemplo de dos registros en el que el sistema ha reconocido tres eventos de tipo Exp, en el test ciego-ciego. Los círculos y las flechas indican los eventos.

Pueden observarse en los registros anteriores, tres eventos de tipo Exp que han sido reconocidos por el sistema. En el primer registro (círculos rojos), acompaña a los

eventos una mayor concentración de ruido de fondo, sin embargo, el sistema ha podido catalogar tres eventos en dicho registro. El sistema también ha reconocido en los registros, eventos de tipo ruido o Sil. Un segmento de esto se observa en la figura 5.21, en donde se han seleccionado a manera de ejemplo, dos de los registros en lo que el sistema catalogó o etiquetó como Sil.

Registros Sil, en el test ciego-ciego del Stromboli

```

"/2570213F.rec"
0 11900000 sil -16487.509766

"/2541708F.rec"
0 14900000 sil -20622.304688

```

Fig. 5.21 Ejemplo de dos registros del fichero de resultado, en el reconocimiento de eventos tipo Sil, por parte del sistema en el test ciego-ciego

La representación gráfica de la señal en el dominio del tiempo, así como del espectrograma de los registros anteriores se observan en la figura 5.22.

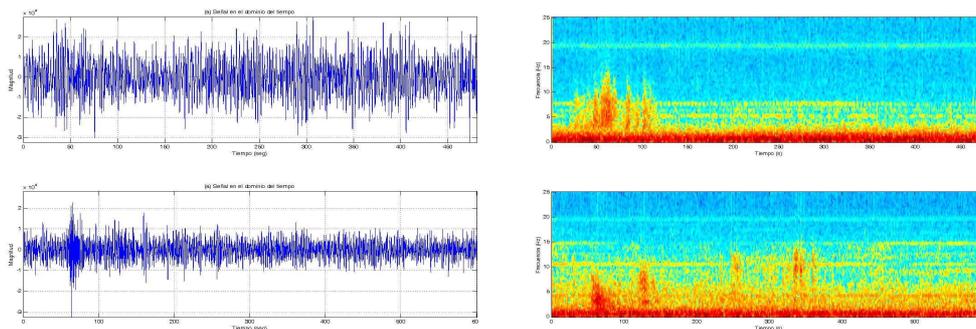


Fig. 5.22 Ejemplo de registros en el que el sistema ha catalogado los registros como de tipo Sil, en el test ciego-ciego

Al analizar la figura anterior, se observa como el ruido de fondo enmascara cualquier tipo de evento que pudiera estar presente en los registros. En la gráfica de la señal en el dominio del tiempo, así como el espectrograma, pudiera determinarse un evento, al menos en la parte inicial de los registros, sin embargo, esto no es posible debido al tremor de fondo. Por lo que el sistema reconoce este tipo de registro solamente como ruido de fondo. Con esto, puede afirmarse que todas las pruebas llevadas a cabo en el Strómboli han tenido éxito en el reconocimiento de los eventos.

5.2.3.- Análisis de resultados en el San Cristóbal

Se procederá a realizar las pruebas de los modelos de entrenamiento con la base de datos segmentada del San Cristóbal. Los parámetros de configuración inicial de los primeros ensayos, al igual que en los volcanes anteriores será de: 6 iteraciones, 16 Gausianas, 11 estados y factor de penalización p por defecto, es decir valor 0.0. Se han segmentado un total de 400 registros. Así, en el San Cristóbal, los primeros reconocimientos con el entrenamiento en pruebas cerradas, presentaron los resultados que se observan en la tabla 5.21.

===== HTK Análisis de Resultados =====
 SENT: %Correct=44.25 [H=177, S=223, N=400]
 WORD: %Corr=84.89, %Acc=80.57 [H=1787, D=262, S=56, I=91, N=2105]
 ----- Matriz de Confusión-----

	NS	S1	S2	S3
NS	994	15	26	0
S1	0	302	7	1
S2	2	3	385	0
S3	0	1	1	106
Ins	55	16	18	2
Del	180	39	32	7

Tabla 5.21 Resultados obtenidos del HTK en el primer experimento con el San Cristóbal.
 %Corr = Porcentaje correcto de etiquetas reconocidas, %Acc = Porcentaje de precisión
 H= Número correcto de etiquetas, D = Número de eliminaciones, S = Número de sustituciones,
 I = Número de Inserciones, N = Número total de etiquetas. Ins = Inserciones, Del = Borrados

Como se puede observar en la tabla anterior, para este experimento se utilizaron 400 registros en el entrenamiento, el sistema arrojó un porcentaje de aciertos de reconocimiento de eventos (Corr) de 84.89 % y un porcentaje de precisión (Acc) de 80.57%, (para que la prueba se considere efectiva y fiable, el sistema debería de presentar un resultado superior al 80% de aciertos), lo que indicaría (de forma parcial) que la prueba inicial ha tenido “éxito”. La intersección de las columnas con las filas en la matriz de confusión indica si el sistema clasifica eventos de un tipo como otro. Por ejemplo, del evento S1 fueron reconocidos por el sistema 302 eventos, hubo confusión en 15 de ellos con ruido, 3 con eventos S2 y 1 con un evento de tipo S3. Análogamente, el evento S2 en el que el sistema reconoció 385 eventos, hubo confusión de 26 eventos con el ruido, 7 con S1 y 1 con S3.

Otra información que muestra la tabla 5.21 es que el sistema de 2,105 eventos analizados, eliminó 262 eventos constituidos de 180 eventos de ruido, 39 eventos tipo S1, 32 eventos tipo S2 y 7 eventos tipo S3. El sistema, además sustituyó 56 eventos e insertó 91, la mayoría (55) han sido de ruido de fondo, 16 son de tipo S1, 18 de tipo S2 y 2 de tipo S3. A partir de estos primeros resultados se va a realizar el mismo proceso de mejoras que ya se ha hecho en los anteriores dos volcanes.

a) Cambio en el número de iteraciones

En la modificación del parámetro del número de iteraciones, los parámetros iniciales son 16 Gaussianas, 11 estados y factor de penalización “p” por defecto (0.0). El resultado de las pruebas, con el cambio del número de iteraciones en el San Cristóbal se observa en la tabla 5.22, en amarillo se presenta la columna con mejores resultados.

Cambio Iter		Número de Iteraciones				
		3	6	9	12	15
% Corr		83.80	84.89	84.66	85.32	85.42
% Acc		79.67	80.57	81.90	82.04	82.28
Sust		71	56	49	46	51
Evento	Unidad	Cantidades				
S1	Ins	18	16	16	12	8
	Del	46	39	42	40	41
S2	Ins	20	18	11	22	19
	Del	32	32	29	25	27
S3	Ins	0	2	0	1	1
	Del	2	7	4	6	8
NS	Ins	49	55	31	34	38
	Del	190	184	199	192	180

Tabla 5.22 Resultados obtenidos en las pruebas cerradas del HTK, con 11 estados y 16 Gausianas al modificar el parámetro del número de iteraciones en el San Cristóbal

%Corr = Porcentaje correcto de etiquetas reconocidas, %Acc = Porcentaje de precisión
H= Número correcto de etiquetas, D = Número de eliminaciones, S = Número de sustituciones,
I = Número de Inserciones, N = Número total de etiquetas. Ins = Inserciones, Del = Borrados

Como se puede observar en la tabla anterior, aunque a mayor número de iteraciones se obtiene un mayor porcentaje de precisión (Acc), dicho porcentaje es solo un 2% superior en las pruebas (máximas) con 15 iteraciones a las pruebas (mínimas) con 6. Lo mismo sucede al comparar en la tabla las sustituciones, borrados e inserciones, el aumento no significa tanto, como lo que implicaría el costo computacional en tiempo al aumentar las iteraciones. Por ese motivo, se ha decidido establecer en 6, el número de iteraciones para los siguientes ensayos.

b) Cambio en el número de Gausianas

Con los 400 registros, al modificar el parámetro del número de Gausianas, esto se representa en la tabla 5.23, en amarillo se presenta la fila con mejores resultados.

Número de Experimento	No Gausianas	% Corr. obtenido	% Acc. obtenido
1	8	79.05	71.31
2	16	84.89	80.57
3	24	88.79	86.56
4	32	91.02	89.50
5	40	93.68	92.97
6	48	95.01	94.58

Tabla 5.23 Resultados obtenidos en las pruebas cerradas del HTK, al modificar el parámetro del número de gaussianas en el San Cristóbal.

Un análisis gráfico de la modificación del número de gaussianas en los experimentos se observa en las figuras 5.23 y 5.24.

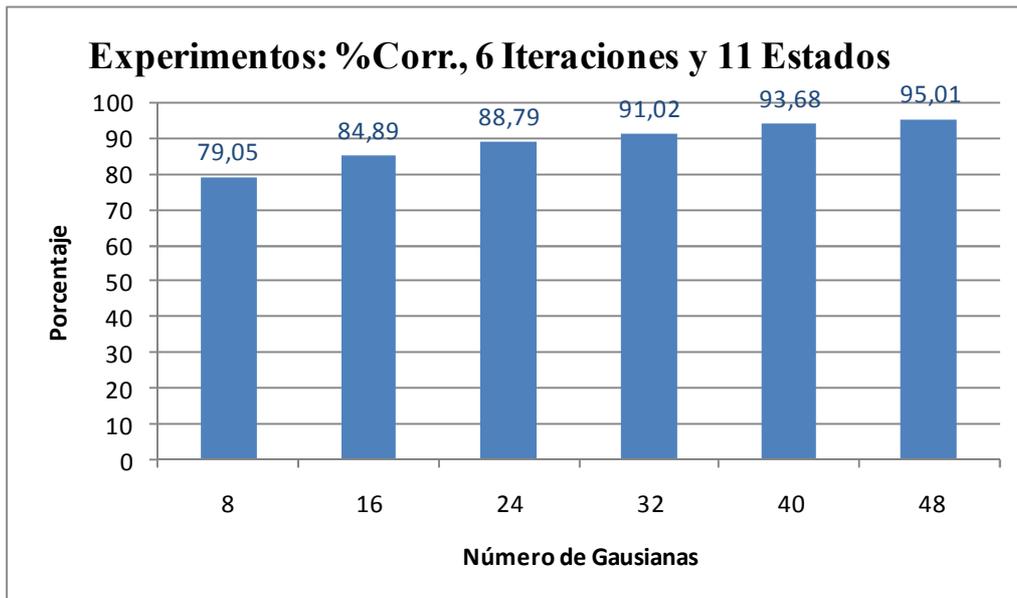


Fig. 5.23 Resultados obtenidos en %Corr en las pruebas cerradas del HTK, en el porcentaje de aciertos al modificar el parámetro del número de gaussianas en el San Cristóbal

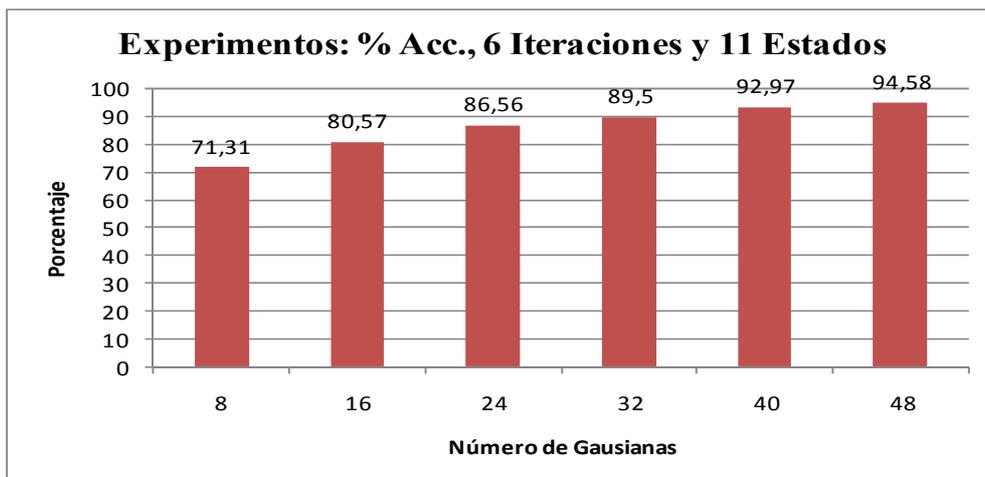


Fig. 5.24 Resultados obtenidos en %Acc en las pruebas cerradas del HTK, en el porcentaje de aciertos al modificar el parámetro del número de gaussianas en el San Cristóbal

De acuerdo a lo que se observa en las figuras 5.23, 5.24, y en la tabla 5.23, existe aproximadamente un 3% de diferencia entre las pruebas con 32 Gaussianas y las realizadas con 40, el costo en tiempo de 32 es mucho menor que en 40, por lo que se determina dejar el parámetro en 32 gaussianas para los siguientes ensayos.

c) Cambio en el factor de penalización

Las pruebas con la modificación del factor de penalización y los parámetros ya establecidos de iteraciones y factor $-p$, presentan los resultados que se observan en la tabla 5.24, en amarillo se presenta la columna con mejores resultados.

Cambio Factor -p		Valor -p						
		-15	-10	-5	0	5	10	15
% Corr		85.99	88.08	89.69	91.02	92.83	93.97	95.01
% Acc		85.70	87.60	88.79	89.50	90.74	90.17	88.08
Sust		15	17	18	18	18	19	19
Evento	Unidad	Cantidades						
S1	Ins	1	1	1	3	5	9	14
	Del	51	37	30	26	18	14	12
S2	Ins	1	1	3	3	3	4	7
	Del	38	35	29	24	21	19	15
S3	Ins	0	0	0	0	0	0	0
	Del	7	4	4	3	2	1	0
NS	Ins	4	8	15	26	36	67	125
	Del	184	158	136	118	92	74	59

Tabla 5.24 Resultados obtenidos con 6 iteraciones, 11 estados y 32 Gaussianas, en la modificación del parámetro “-p” del Hvite en el San Cristóbal.

%Corr = Porcentaje correcto de etiquetas reconocidas, %Acc = Porcentaje de precisión
Ins = Inserciones, Sust = Sustituciones, Del = Borrados

Los resultados representados en la tabla 5.24, presentan los mejores porcentajes con un factor -p de 5, lo que indica que para las posteriores pruebas se asignará este valor.

d) Cambio en el número de estados

Las pruebas con la modificación del factor de penalización, presentan los resultados que se observan en la tabla 5.25, en amarillo se indica la columna con mejores resultados.

Cambio No. Estados		Número de estados							
		5	7	9	11	13	15	17	19
% Corr		94.39	95.30	93.92	92.83	94.16	91.31	88.22	82.00
% Acc		76.29	87.13	90.50	90.74	92.83	90.36	87.41	81.71
Sust		26	12	10	18	7	15	16	38
Evento	Unidad	Cantidades							
S1	Ins	97	44	8	5	1	0	1	1
	Del	11	14	24	18	27	42	90	104
S2	Ins	21	22	13	3	0	0	2	0
	Del	5	8	12	21	24	39	58	82
S3	Ins	4	0	0	0	0	0	0	0
	Del	1	2	5	2	1	6	13	18
NS	Ins	259	106	51	36	27	20	14	5
	Del	75	63	77	92	64	81	71	137

Tabla 5.25 Resultados obtenidos con 6 iteraciones, 11 estados, 32 Gaussianas y factor p 5, en la modificación del número de estados en el San Cristóbal.

%Corr = Porcentaje correcto de etiquetas reconocidas, %Acc = Porcentaje de precisión
Ins = Inserciones, Sust = Sustituciones, Del = Borrados

En los resultados que se muestran en la tabla anterior, se observa que incluso los resultados con 11 estados han mejorado los anteriores resultados, al establecer el

cambio con el factor de penalización, pero los mayores porcentajes de precisión se producen al realizar las pruebas con modelos de 13 estados. Además, el número tanto de sustituciones como de inserciones es el más bajo en dichas pruebas, por lo que en los siguientes ensayos se realizaran con modelos de 13 estados. En este punto se marca el hito de repetir las pruebas observadas en la tabla 5.23, para observar los cambios que se obtienen, en relación a los parámetros ya establecidos como mejoras. Así, se establecen los parámetros de configuración que son: el número de iteraciones en 6, el número de estados en 13, el factor de penalización de la herramienta de entrenamiento, H_{vite} (-p) en 5. Los valores obtenidos en dichas pruebas se observan en la tabla 5.26. En amarillo se observa la fila con mejores resultados.

Número de Experimento	No Gaussianas	% Corr. obtenido	% Acc. obtenido
1	8	77.77	69.07
2	16	85.51	80.86
3	24	91.12	88.88
4	32	94.16	92.83
5	40	96.25	95.20
6	48	97.29	96.72

Tabla 5.26 Resultados obtenidos en las pruebas cerradas del HTK, con 6 iteraciones, 13 estados, y factor -p (5), al modificar el parámetro del número de gaussianas en el San Cristóbal.
%Corr = Porcentaje correcto de etiquetas reconocidas, %Acc = Porcentaje de precisión

Al comparar los resultados de la tabla anterior con los de la tabla 5.14, se observan las mejoras en los modelos de entrenamiento con 13 estados. El porcentaje de efectividad o precisión (Acc) va en aumento con cada prueba. Esto indica que el sistema ha mejorado al implementar el cambio del número de estados conjuntamente con los parámetros ya comprobados. En las 32 Gaussianas, el porcentaje de efectividad ha mejorado de un 83.87% a un 89.14 %, lo que representa algo más de un 5% de aumento. Por lo tanto, los resultados finales con los modelos de entrenamiento en las pruebas cerradas con 32 Gaussianas en el San Cristóbal (en amarillo en la tabla 5.26), se observan en la tabla 5.27

===== HTK Análisis de Resultados =====
 SENT: %Correct=77.50 [H=310, S=90, N=400]
 WORD: %Corr=94.16, %Acc=92.83 [H=1982, D=116, S=7, I=28, N=2105]
 ----- Matriz de Confusión-----

	NS	S1	S2	S3
NS	1149	0	6	0
S1	0	322	0	0
S2	1	0	397	0
S3	0	0	0	114
Ins	27	1	0	0
Del	64	27	24	1

Tabla 5.27 Resultados obtenidos en las pruebas del San Cristóbal.
 Utilizando 6 Iteraciones, 13 estados, 32 Gaussianas y factor (-p) de 5
 %Corr = Porcentaje correcto de etiquetas reconocidas, %Acc = Porcentaje de precisión
 H= Número correcto de etiquetas, D = Número de eliminaciones, S = Número de sustituciones,
 I = Número de Inserciones, N = Número total de etiquetas. Ins = Inserciones, Del = Borrados

La tabla anterior demuestra que el sistema, en los 400 registros segmentados, ha logrado un porcentaje de aciertos correctos (Corr) de 94.16% y una precisión (Acc) de 92.83%. Además, de 2,105 eventos totales, el sistema ha reconocido con éxito 1,982, ha eliminado 116 (la mayoría ruido), sustituido 7 e insertado 28, de los cuales 27 de ellos corresponden a ruido. Hasta ahora, cada prueba en el proceso de mejoras, ha demostrado que se ha podido optimizar el porcentaje de precisión en el reconocimiento por parte del sistema.

e) Proceso de depurado selectivo de los registros

Al realizar una inspección manual, del fichero de resultados que se obtiene en las pruebas anteriores, se detectan registros con porcentaje inferior al 60% en el reconocimiento. A manera de ejemplo, una parte de dicho fichero se observa en la siguiente figura.

Proceso de Depurado Selectivo en el San Cristóbal

```

----- Sentence Scores -----
===== HTK Results Analysis =====
  Rec : rec.13.32.gram.mlf
----- File Results -----
2006-03-01-1416-158C.rec: 100.00(100.00) [H= 7, D= 0, S= 0, I= 0, N= 7]
2006-03-01-1431-158C.rec: 100.00(100.00) [H= 3, D= 0, S= 0, I= 0, N= 3]
2006-03-01-1436-158C.rec: 100.00(100.00) [H= 9, D= 0, S= 0, I= 0, N= 9]
2006-03-01-1451-158C.rec: 100.00(100.00) [H= 3, D= 0, S= 0, I= 0, N= 3]
2006-03-01-1456-148C.rec: 100.00(100.00) [H= 5, D= 0, S= 0, I= 0, N= 5]
2006-03-03-0650-568C.rec: 55.56( 55.56) [H= 5, D= 4, S= 0, I= 0, N= 9]
Aligned transcription: 2006-03-03-0650-568C.lab vs 2006-03-03-0650-568C.rec
LAB: NS S2 NS S1 NS S1 NS S1 NS
REC: NS S2 NS S1 NS
2006-03-03-0700-558C.rec: 100.00(100.00) [H= 4, D= 0, S= 0, I= 0, N= 4]
2006-03-03-0715-558C.rec: 100.00(100.00) [H= 6, D= 0, S= 0, I= 0, N= 6]
2006-03-03-0730-558C.rec: 71.43( 71.43) [H= 5, D= 1, S= 1, I= 0, N= 7]
2006-03-07-2050-048C.rec: 33.33( 33.33) [H= 1, D= 2, S= 0, I= 0, N= 3]
Aligned transcription: 2006-03-07-2050-048C.lab vs 2006-03-07-2050-048C.rec
LAB: NS S2 NS
REC: NS
2006-03-10-0634-368C.rec: 100.00(100.00) [H= 7, D= 0, S= 0, I= 0, N= 7]
2006-03-10-0639-368C.rec: 100.00(100.00) [H= 7, D= 0, S= 0, I= 0, N= 7]
2006-03-10-0729-368C.rec: 100.00(100.00) [H= 3, D= 0, S= 0, I= 0, N= 3]
2006-03-10-0944-358C.rec: 50.00( 50.00) [H= 2, D= 2, S= 0, I= 0, N= 4]
Aligned transcription: 2006-03-10-0944-358C.lab vs 2006-03-10-0944-358C.rec
LAB: NS S2 S1 NS
REC: NS NS
2006-03-10-1849-318C.rec: 100.00(100.00) [H= 5, D= 0, S= 0, I= 0, N= 5]
2006-03-10-1859-308C.rec: 100.00(100.00) [H= 3, D= 0, S= 0, I= 0, N= 3]
2006-03-10-1919-308C.rec: 100.00(100.00) [H= 5, D= 0, S= 0, I= 0, N= 5]

```

Fig. 5.25 Ejemplo del Análisis manual en el Proceso selectivo (evento por evento), de los resultados obtenidos en el San Cristóbal. Los valores en rojo representan los registros que han tenido un porcentaje de 100 y que no se cambiarán. Los valores en verde son aquellos que han tenido un porcentaje menor a 60 y que se analizarán para eliminarse de la base de datos.

En la figura anterior, puede observarse como algunos de los registros presentan porcentajes inferiores al 60% de acierto respecto al valor etiquetado por el observador. Un análisis gráfico de dichos registros se observa en las figuras 5.26 y 5.27.

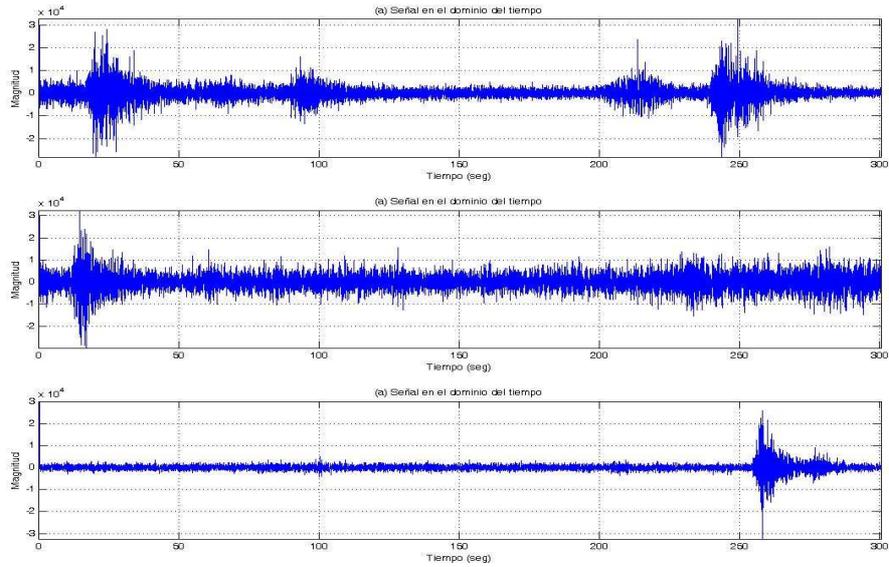


Fig. 5.26 Registros 2006-03-03-0650-56SC (arriba), 2006-03-07-2050-04SC (centro) y 2006-03-10-0944-35SC (abajo), en el dominio del tiempo

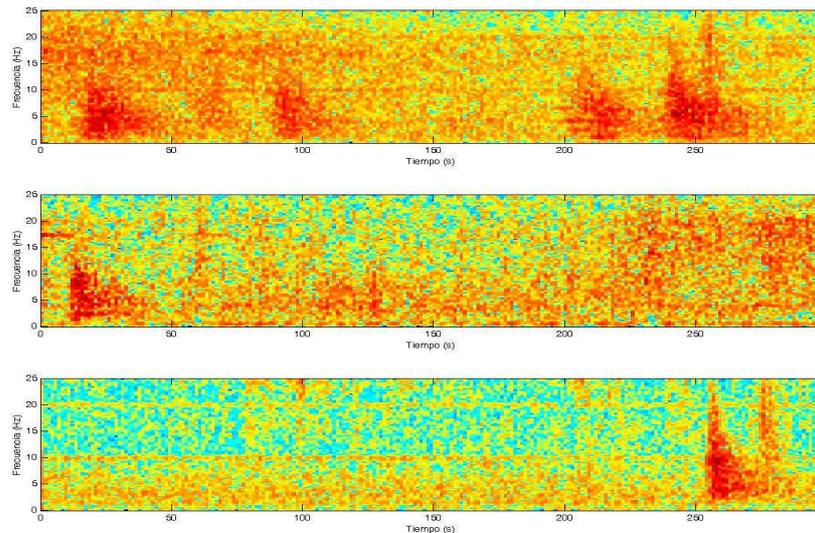


Fig. 5.27 Espectrogramas de los registros: 2006-03-03-0650-56SC (arriba), 2006-03-07-2050-04SC (centro) y 2006-03-10-0944-35SC (abajo)

Puede observarse en las figuras, como el ruido de fondo presente en los registros, hace que no sean reconocidos por el sistema los eventos que se encuentran presentes. Este es el caso de los dos eventos eliminados por el sistema en el registro de la parte superior, el evento de la parte central y los dos eventos de la parte inferior. Aunque dichos eventos se pueden reconocer a simple vista, la presencia de un fuerte ruido en altas energías hace que el sistema no sea eficiente y ante esto, elimina dichos eventos. De esta forma, se procedieron a eliminar los registros similares que contenían porcentajes menores a 60%, en total solamente se han podido detectar 11 de ellos, aún así, se realizará la misma prueba eliminando dichos registros, para observar si el sistema presenta alguna mejoría. En teoría, al eliminar registros y por consiguiente eventos,

podría esperarse que los porcentajes disminuyeran. Los resultados de las pruebas después de haber eliminado los registros se observan en la tabla 5.28.

```

===== HTK Análisis de Resultados =====
          SENT: %Correct=76.86 [H=299, S=90, N=389]
          WORD: %Corr=94.96, %Acc=93.74 [H=1941, D=90, S=13, I=25, N=2044]
          ----- Matriz de Confusión-----
  
```

	NS	S1	S2	S3
NS	1129	2	9	0
S1	0	312	1	0
S2	1	0	388	0
S3	0	0	0	112
Ins	25	0	0	0
Del	45	23	21	1

Tabla 5.28 Resultados obtenidos al eliminar registros con bajo porcentaje del San Cristóbal, utilizando 6 Iteraciones, 13 estados, 32 Gaussianas y factor $-p = 5$

%Corr = Porcentaje correcto de etiquetas reconocidas, %Acc = Porcentaje de precisión
 H= Número correcto de etiquetas, D = Número de eliminaciones, S = Número de sustituciones,
 I = Número de Inserciones, N = Número total de etiquetas. Ins = Inserciones, Del = Borrados

Después de haber eliminado los 11 registros de la base de datos, los porcentajes en las pruebas han mejorado, lo que en este punto indica, que el proceso por el cual se eliminan los registros resulta positivo a la hora de evaluar los resultados. El porcentaje de efectividad (Acc) pasó de un 92.83% a un 93.74% casi un 2% más que en la anterior prueba eliminando solo 11 registros con bajo porcentaje. Otro punto a destacar es que no se han tenido inserciones de ninguno de los tres eventos (S1, S2 y S3), solo concentrándose en eventos de tipo ruido con 25, lo que incrementaría aun más el porcentaje real de precisión (Acc) en las pruebas. También aumentaron un poco las sustituciones de 9 a 13, pero bajaron significativamente las eliminaciones de 116 a 90. Todo ello indica que el sistema ha logrado tener un éxito en las pruebas y mantiene su fiabilidad en el reconocimiento de los eventos.

f) Pruebas abiertas (test ciegos)

En las pruebas abiertas, o “test ciegos (blind test)”, se van a incluir los parámetros ya confirmados en las pruebas anteriores. Como se ha mencionado, las pruebas están constituidas por dos tipos de experimentos, el primero de ellos como se recordará consiste en dividir en dos partes la base de datos. Una primera parte, la formarán los modelos entrenados, para pasarlos después a la parte restante. El segundo test “ciego-ciego”, se pasará la base de datos entrenada, a una parte de registros no segmentados para observar que tipo de eventos puede reconocer el sistema.

i) Test ciego con base de datos entrenada dividida

Para la primera parte de las pruebas abiertas, de la base de datos de 389 registros, han quedado 300 registros para el entrenamiento, y 89 para el reconocimiento. Los resultados de estas pruebas se observan en la tabla 5.29.

=====
 HTK Análisis de Resultados =====
 SENT: %Correct=51.69 [H=46, S=43, N=89]
 WORD: %Corr=90.37, %Acc=84.69 [H=366, D=30, S=9, I=23, N=405]
 ----- Matriz de Confusión-----

	NS	S1	S2	S3
NS	219	1	5	0
S1	0	45	2	0
S2	1	0	79	0
S3	0	0	0	23
Ins	20	2	1	0
Del	20	4	6	0

Tabla 5.29 Resultados obtenidos en las pruebas abiertas del San Cristóbal, utilizando 6 Iteraciones, 13 estados, 32 Gaussianas y factor $-p = 5$

%Corr = Porcentaje correcto de etiquetas reconocidas, %Acc = Porcentaje de precisión
 H= Número correcto de etiquetas, D = Número de eliminaciones, S = Número de sustituciones,
 I = Número de Inserciones, N = Número total de etiquetas. Ins = Inserciones, Del = Borrados

Los resultados indican que en la pruebas, el sistema, de 405 eventos ha podido reconocer 366, ha eliminado 30, sustituido 9 e insertado 23 eventos. El porcentaje de efectividad en dichas pruebas ha sido de 84.69%. En la matriz de confusión se observa que el sistema ha podido reconocer con efectividad casi la totalidad de los eventos con lo que el éxito de esta prueba está comprobado.

ii) Test ciego-ciego

El segundo tipo de las pruebas abiertas, el test “ciego-ciego”, los 270 registros que constituyen la base de datos de entrenamiento van a reconocer 566 registros adicionales no parametrizados, sin segmentar, la mayoría de ellos conteniendo un fuerte ruido de fondo. El sistema, como parámetro de salida ha creado un fichero conteniendo los registros etiquetados. A manera de ejemplo, se han tomado para ser analizados 6 registros de dicho fichero, en donde el sistema reconoce eventos (S1, S2 y S3). A continuación se muestra un trozo del fichero de resultados en la figura 5.28.

Test ciego-ciego en el San Cristóbal

Eventos tipo S1 encontrados

```

**/2006-03-01-1506-148C.rec"
0 750000 NS -721.611267
750000 1850000 S2 -1061.937744
1850000 4150000 NS -2154.983154
4150000 4750000 S1 -525.529236
4750000 6100000 NS -1235.341553
6100000 7150000 S2 -1048.456055
-
**/2006-02-23-1732-218C.rec"
0 1550000 NS -1480.644165
1550000 2400000 NS -797.328369
2400000 3350000 S1 -972.691528
3350000 4100000 S1 -728.222290
4100000 4950000 S1 -917.535400
4950000 5750000 NS -751.073547
5750000 6550000 S1 -700.948364
6550000 7150000 NS -606.859436
    
```

Eventos tipo S2 encontrados

```

**/2006-02-27-1806-358C.rec"
0 2000000 NS -2093.923096
2000000 3100000 S2 -1272.673950
3100000 5100000 NS -2071.823486
5100000 7150000 NS -2217.205566
-
**/2006-02-27-1841-358C.rec"
0 2300000 NS -2244.941162
2300000 3300000 NS -1029.933228
3300000 4150000 S2 -932.130005
4150000 5450000 NS -1296.865112
5450000 6050000 S2 -653.123230
6050000 7150000 NS -1213.076050
    
```

Eventos tipo S3 encontrados

```

**/2006-03-01-1451-158C.rec"
0 2300000 NS -2244.941162
2300000 3250000 S3 -721.914001
3250000 4250000 NS -928.110718
4250000 7150000 NS -2779.465332
-
**/2006-03-01-1601-158C.rec"
0 1200000 NS -1174.665405
1200000 1800000 S3 -525.750305
1800000 2700000 NS -882.972595
2700000 3700000 S1 -996.118042
3700000 4350000 NS -607.729675
4350000 5650000 S3 -1141.839600
5650000 7150000 NS -1332.887817
    
```

Fig. 5.28 Ejemplo de seis registros, en el reconocimiento de eventos tipo S1, S2 y S3, por parte del sistema en el San Cristóbal, durante las pruebas o test ciego-ciego

En la figura anterior, el sistema ha podido clasificar en los registros eventos de tipo S1, S2 y S3. Los anteriores registros se pueden analizar y determinar gráficamente de forma visual si el sistema ha podido reconocer con éxito dichos eventos. En las figuras 5.29, 5.30 y 5.31, pueden observarse las gráficas de dichos registros.

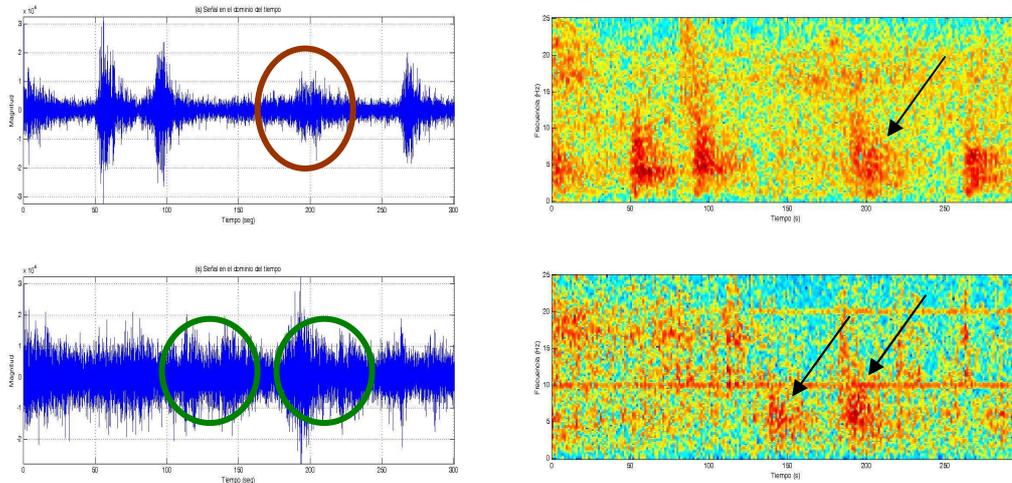


Fig. 5.29 Ejemplo de registros en el test ciego-ciego, en el que el sistema ha reconocido eventos de tipo S1. Los círculos y las flechas indican los eventos.

En la figura 5.29, en el registro superior, el sistema ha reconocido un evento de tipo S1 (círculo rojo), aunque de menor duración, el cual está acompañado de 3 eventos de tipo S2. El segundo registro, en la parte de abajo (círculos verdes), el sistema reconoce dos tipos de eventos S1, uno de menor magnitud y duración que el segundo. Cabe notar que en el segundo registro, el tremor de fondo es mucho mayor, aún así el sistema ha podido reconocer los dos eventos en este ruido de fondo. La siguiente figura muestra el reconocimiento por parte del sistema de eventos de tipo S2.

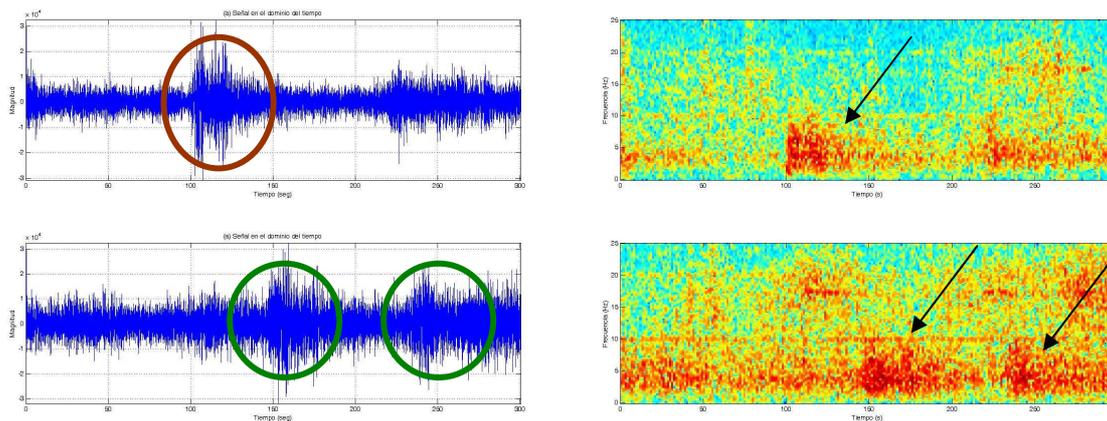


Fig. 5.30 Ejemplo de registros en el test ciego-ciego, en el que el sistema ha reconocido eventos de tipo S2. Los círculos y las flechas indican los eventos.

Puede observarse en la figura anterior, que el sistema en el primer registro (círculo rojo), ha podido reconocer un evento de tipo S2. En cambio en el segundo registro (círculo verde), el sistema ha podido clasificar dos eventos de tipo S2. En el San

Cristóbal, al igual que en el Etna, el tremor de fondo es muy fuerte, por lo que el valor del reconocimiento y clasificación de eventos por parte del sistema en estas circunstancias es aún más valioso. La siguiente figura muestra como el sistema ha podido reconocer eventos de tipo S3.

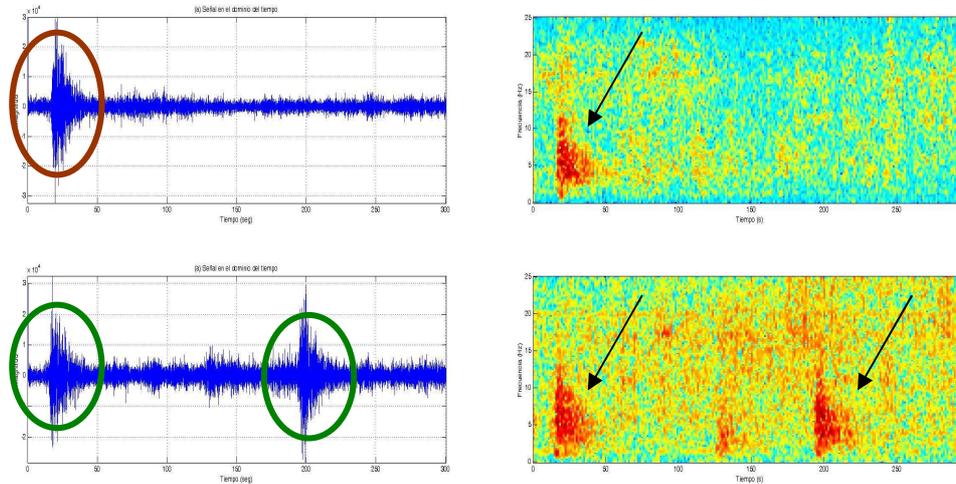


Fig. 5.31 Ejemplo de registros en el test ciego-ciego, en el que el sistema ha reconocido eventos S3. Los círculos y las flechas indican los eventos.

La figura anterior muestra como el sistema reconoce eventos de tipo S3, en el primer registro (círculo rojo), el evento está claramente definido a través de las señales en el dominio del tiempo y el espectrograma. En el segundo registro (círculo verde), se encuentran dos eventos de tipo S3, con un evento de tipo S1 entre ellos. El sistema demuestra así su fiabilidad a la hora de reconocer eventos que el observador no ha marcado como tales. Por otro lado, en cuando a los eventos de tipo NS, la figura 5.32, muestra como el sistema los determina.

Registros NS, en el test ciego-ciego del San Cristóbal

```

"/2006-03-01-1531-14SC.rec"
0 4600000 NS -4585.212891
4600000 7150000 NS -2692.639648
.
"/2006-03-01-1556-15SC.rec"
0 3250000 NS -3273.701660
3250000 4300000 NS -1088.363647
4300000 5850000 NS -1482.213989
5850000 7150000 NS -1347.745850
    
```

Fig. 5.32 Ejemplo de dos registros del fichero de resultado, en el reconocimiento de eventos tipo NS, por parte del sistema en el test ciego-ciego

La figura 5.33, muestra gráficamente la representación de la transformada de Fourier en el dominio del tiempo, así como el espectrograma de dichos registros.

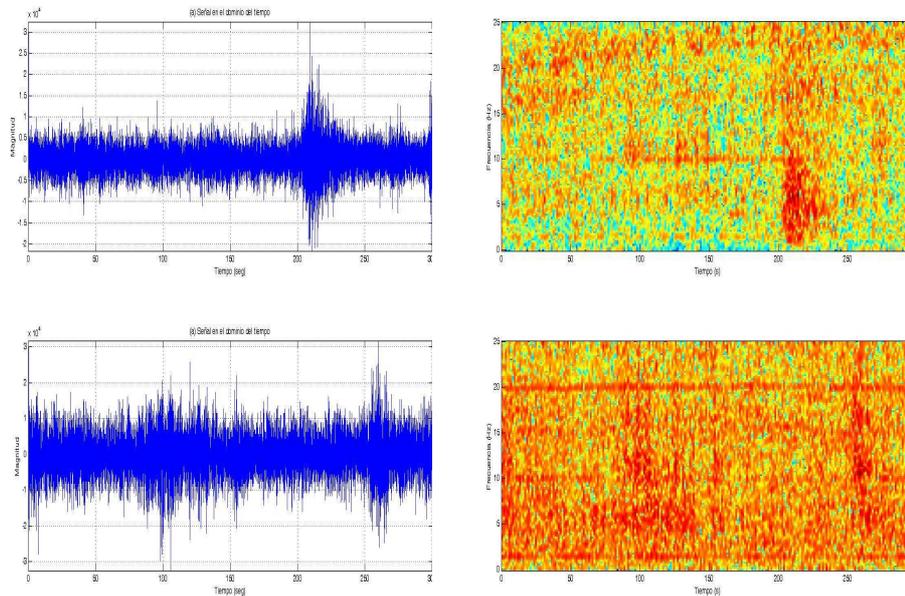


Fig. 5.33 Ejemplo de registros en el que el sistema ha catalogado los registros como de tipo NS, en el test ciego-ciego

En la figura anterior, se puede observar que en el primer registro se aprecia un evento probablemente de tipo S2, que sin embargo el sistema determina marcarlo como NS, y eliminar cualquier evento, debido al ruido de altas energías que lo acompaña. En el segundo registro se aprecia mejor el tipo de tremor de fondo, que abarca la totalidad del registro, aunque existen picos de energía de probables eventos, estos están muy enmascarados por una fuerte presencia de ruido de fondo de altas energías (muy saturados en el caso del segundo registro). Una vez más, en este caso el sistema ha sido fiable a la hora de discernir que este tipo de registro corresponde más a un ruido. Futuros trabajos podrían incluir en la base de datos de entrenamientos, muchos más registros con estas características, para que el sistema pueda ser capaz de reconocer eventos aún en estas condiciones.

5.2.4.- Análisis de resultados en el Telica

Para el Telica, la base de datos está constituida por el segmentado e etiquetado manual de un total de 1,240 registros totales. Con los mismos parámetros originales (6 ite, 11 est, 16 Gauss y $p = 0.0$), se realizarán los primeros ensayos. Posteriormente, se han de buscar los mejores resultados con el proceso de mejoras, ya que no siempre pueden coincidir todos los índices de los parámetros o generalizarlos en todos los volcanes. El motivo podría encontrarse en la diferente forma en que se recopilan los datos, o las diferencias en cuanto a los registros respecto de un volcán a otro, por ello se deben buscar los mejores resultados posibles, mediante la ejecución de la serie de pruebas establecidas, destinadas a dejar un balance entre los resultados y el tiempo o costo computacional. Así se van a realizar las mismas acciones que ya se han hecho en los anteriores tres volcanes. Con los 1,240 registros segmentados se han realizado las primeras pruebas, que han dejado los resultados que se observan en la tabla 5.30.

```
===== HTK Análisis de Resultados =====
SENT: %Correct=33.79 [H=419, S=821, N=1240]
WORD: %Corr=85.66, %Acc=71.39 [H=3799, D=400, S=236, I=633, N=4435]
```

----- Matriz de Confusión-----

	NT	T1	T2	T3	T4
NT	2473	36	44	28	37
T1	1	294	3	10	6
T2	4	7	407	15	7
T3	1	11	7	373	3
T4	2	8	3	3	252
Ins	130	115	159	127	102
Del	230	45	62	34	29

Tabla 5.30 Resultados obtenidos del HTK en las pruebas con 1,240 registros y los parámetros iniciales en el Telica.
 %Corr = Porcentaje correcto de etiquetas reconocidas, %Acc = Porcentaje de precisión
 H= Número correcto de etiquetas, D = Número de eliminaciones, S = Número de sustituciones,
 I = Número de Inserciones, N = Número total de etiquetas. Ins = Inserciones, Del = Borrados

La interpretación de la tabla muestra que fueron analizados 1,240 registros obteniéndose un porcentaje correcto (Corr) de 85.66% y una fiabilidad o precisión (Acc) de 71.39%. El sistema, de 4,435 eventos, ha reconocido 3,799, eliminó 400 sustituyó 236, e insertó 633. Fueron reconocidos efectivamente: 2,473 de tipo NT, 294 eventos de tipo T1, 407 eventos de tipo T2, 373 eventos de tipo T3 y 252 eventos de tipo T4. En la matriz de confusión, por ejemplo, se señala que en el evento T1, un evento fue catalogado como ruido, 3 como T2, 10 como T3 y 6 como T4. El análisis de los demás eventos es similar. A continuación se realizará el proceso de mejoras, a fin de conseguir optimizar el porcentaje de precisión (Acc), mediante la modificación de los parámetros.

a) Cambio en el número de iteraciones

Al modificar el parámetro del número de iteraciones, se obtienen los resultados que se presentan en la tabla 5.31, en amarillo se presenta la fila con mejores resultados.

Cambio Iter		Número de Iteraciones				
		3	6	9	12	15
% Corr		84.69	85.66	86.25	85.89	85.71
% Acc		67.44	71.39	72.36	73.20	76.06
Sust		896	821	787	776	701
Evento	Unidad	Cantidades				
T1	Ins	140	115	106	99	87
	Del	49	45	42	41	44
T2	Ins	200	159	141	130	94
	Del	67	62	55	55	67
T3	Ins	166	127	122	130	81
	Del	39	34	43	41	45
T4	Ins	132	102	107	69	72
	Del	38	29	26	31	46
NT	Ins	127	130	140	136	95
	Del	214	230	212	226	237

Tabla 5.31 Resultados obtenidos con 11 estados, 16 Gaussianas, $p = 0.0$ en la modificación del parámetro del número de iteraciones en el Telica

%Corr = Porcentaje correcto de etiquetas reconocidas, %Acc = Porcentaje de precisión
Ins = Inserciones, Sust = Sustituciones, Del = Borrados

De acuerdo a los resultados presentados en la tabla anterior, puede observarse que los mejores resultados se obtienen al realizar los ensayos con modelos que poseen 15 iteraciones, por lo que se ha determinado fijar este número para las siguientes pruebas.

b) Cambio en el número de Gaussianas

Al modificar el parámetro del número de Gaussianas, los resultados de las pruebas cerradas se muestran en la tabla 5.32, en amarillo se observa la fila con los mejores resultados.

Número de Experimento	No Gaussianas	% Corr. obtenido	% Acc. obtenido
1	8	79.01	63.47
2	16	85.71	76.06
3	24	89.20	82.92
4	32	91.25	86.54
5	40	92.91	89.36
6	48	93.90	91.16

Tabla 5.32 Resultados obtenidos en las pruebas cerradas del HTK con 15 iteraciones, al modificar el parámetro del número de gaussianas en el Telica.

%Corr = Porcentaje correcto de etiquetas reconocidas, %Acc = Porcentaje de precisión

En la tabla se muestra que los porcentajes con 16 gaussianas, han cambiado respecto a los valores iniciales de la tabla 5.30, ya que las pruebas se han realizado con 15 iteraciones en lugar que con 6.

En las figuras 5.34 y 5.35, puede observarse el análisis gráfico, al realizar la modificación del parámetro del número de gaussianas en las pruebas.

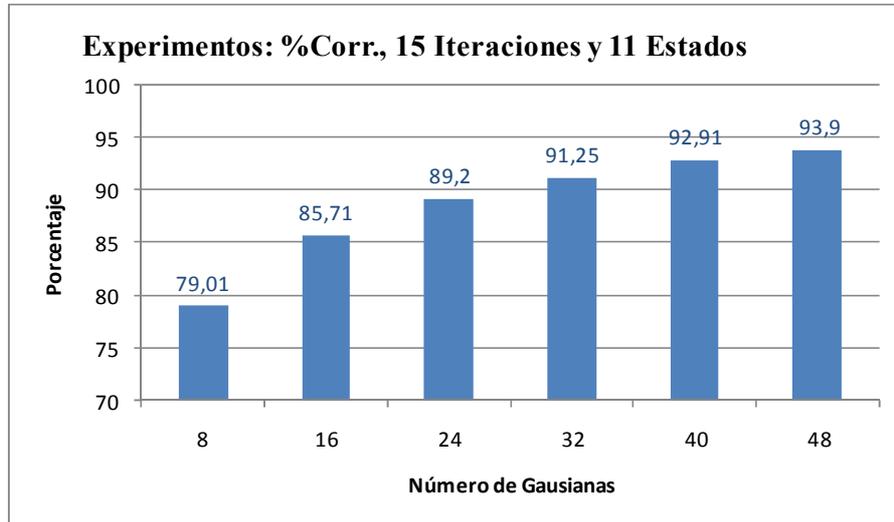


Fig. 5.34 Resultados obtenidos en %Corr en las pruebas cerradas del HTK, en el porcentaje de aciertos al modificar el parámetro del número de gaussianas en el Telica

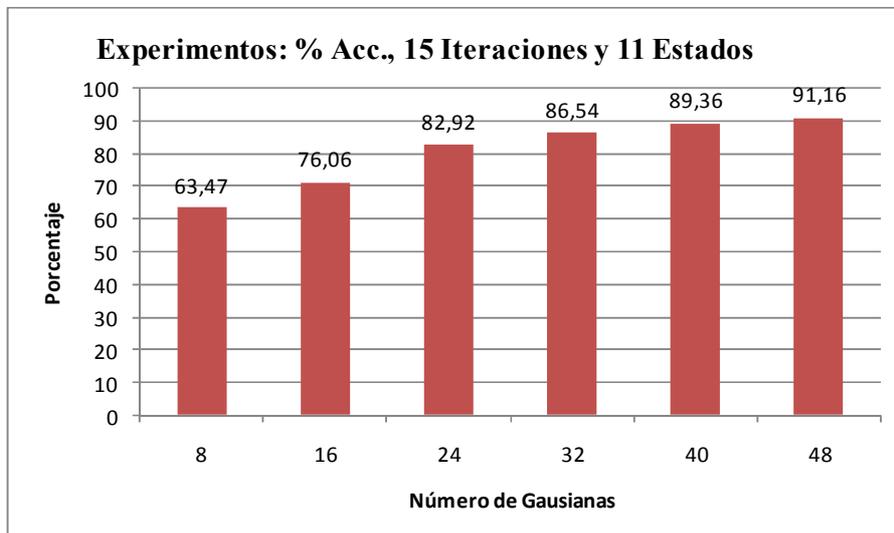


Fig. 5.35 Resultados obtenidos en %Acc en las pruebas cerradas del HTK, en el porcentaje de aciertos al modificar el parámetro del número de gaussianas en el Telica

c) Cambio en el factor de penalización (-p)

El siguiente paso en las mejoras de los resultados es realizar la modificación del factor de penalización (-p) de la herramienta de reconocimiento Hvite. Los resultados se observan en la tabla 5.33, en amarillo se observa la columna con los mejores porcentajes.

Cambio Factor -p		Valor -p						
		-15	-10	-5	0	5	10	15
% Corr		87.56	88.91	90.03	91.25	92.42	93.84	94.87
% Acc		85.89	86.41	86.66	86.54	85.87	84.00	79.59
Sust		36	37	49	54	58	67	75
Evento	Unidad	Cantidades						
T1	Ins	5	11	15	25	33	54	80
	Del	63	55	44	36	31	28	22
T2	Ins	11	14	17	30	49	81	128
	Del	80	66	50	39	29	19	13
T3	Ins	8	13	16	21	36	56	82
	Del	57	49	46	37	30	22	18
T4	Ins	12	15	18	23	28	42	61
	Del	38	31	26	23	16	10	7
NT	Ins	38	58	84	110	145	203	238
	Del	279	255	228	200	173	129	93

Tabla 5.33 Resultados obtenidos con 15 iteraciones, 11 estados y 32 Gaussianas, en la modificación del parámetro “-p” del Hvite en el Telica,
%Corr = Porcentaje correcto de etiquetas reconocidas, %Acc = Porcentaje de precisión,
Ins = Inserciones, Sust = Sustituciones, Del = Borrados

De acuerdo a la tabla anterior, los mejores resultados se obtienen con el parámetro -p con un valor de -5, por lo que las siguientes pruebas se realizarán con dicho valor.

d) Cambio en el número de estados

Cambio No Estados		Número de estados							
		5	7	9	11	13	15	17	19
% Corr		84.41	88.59	90.21	90.03	92.37	91.86	89.83	89.53
% Acc		69.94	82.04	86.34	86.66	90.35	90.08	87.75	87.46
Sust		346	165	72	49	21	20	16	31
Evento	Unidad	Cantidades							
T1	Ins	159	46	32	15	4	2	4	5
	Del	29	23	23	44	31	31	43	39
T2	Ins	137	81	38	17	12	5	3	7
	Del	43	44	45	50	35	45	63	69
T3	Ins	154	55	24	16	15	8	4	5
	Del	24	35	41	46	38	35	44	38
T4	Ins	120	60	14	18	1	7	3	4
	Del	34	29	35	26	24	28	36	48
NT	Ins	73	49	64	84	58	56	77	67
	Del	217	211	219	228	190	199	242	221

Tabla 5.34 Resultados obtenidos con 15 iteraciones, 11 estados, 32 Gaussianas y factor p -5, en la modificación del número de estados en el Telica,
%Corr = Porcentaje correcto de etiquetas reconocidas, %Acc = Porcentaje de precisión,
Ins = Inserciones, Sust = Sustituciones, Del = Borrados

La tabla anterior, muestra que los mejores resultados se han obtenido en las pruebas realizadas con modelos de 13 estados. Por lo que los siguientes ensayos se realizarán tomando como parámetro dicha cantidad. En este punto se determina el hito de repetir las pruebas observadas en la tabla 5.32, para observar los cambios que se obtienen, en relación a los parámetros ya establecidos como mejoras. Los valores de configuración son: número de iteraciones 15, el número de estados 13, y el factor de penalización de la herramienta de entrenamiento, Hvite (-p) en -5. Los resultados obtenidos en dichas pruebas se observan en la tabla 5.35. En amarillo se observa la fila con los mejores resultados.

Número de Experimento	No Gausianas	% Corr. obtenido	% Acc. obtenido
1	8	79.60	71.31
2	16	87.15	81.96
3	24	90.12	87.22
4	32	92.37	90.35
5	40	93.77	92.53
6	48	94.82	93.88

Tabla 5.35 Resultados obtenidos en las pruebas cerradas del HTK, con 15 iteraciones, 13 estados, y factor -p (-5), al modificar el parámetro del número de gaussianas en el Telica,
%Corr = Porcentaje correcto de etiquetas reconocidas, %Acc = Porcentaje de precisión

Al comparar los resultados de la tabla anterior con los de la tabla 5.32, se observan las mejoras en los modelos de entrenamiento con 13 estados. El porcentaje de efectividad (Acc) va en aumento con cada prueba. Esto indica que el sistema ha mejorado al implementar el cambio del número de estados y con los parámetros ya comprobados. En las 32 Gaussianas, el porcentaje de efectividad ha mejorado de un 83.87% a un 89.14 %, lo que representa algo más de un 5% de aumento, se determina dejar fijo el número de estados en 15, para los siguientes ensayos. Los resultados finales en las pruebas cerradas con 32 Gaussianas en el Telica (en amarillo en la tabla 5.35), se observan en la tabla 5.36.

```

===== HTK Análisis de Resultados =====
SENT: %Correct=77.60 [H=963, S=278, N=1241]
WORD: %Corr=92.37, %Acc=90.35 [H=4105, D=318, S=21, I=90, N=4444]
----- Matriz de Confusión-----

```

	NT	T1	T2	T3	T4
NT	2650	7	3	3	2
T1	0	328	0	1	0
T2	0	1	464	0	2
T3	0	1	1	389	0
T4	0	0	0	0	274
Ins	58	4	12	15	1
Del	190	31	35	38	24

Tabla 5.36 Resultados obtenidos en las pruebas utilizando 15 Iteraciones, 13 estados, 32 Gaussianas y factor (-p) a -5, en el Telica.

%Corr = Porcentaje correcto de etiquetas reconocidas, %Acc = Porcentaje de precisión,
H= Número correcto de etiquetas, D = Número de eliminaciones, S = Número de sustituciones,
I = Número de Inserciones, N = Número total de etiquetas. Ins = Inserciones, Del = Borrados

Como se observa en las tablas 5.34, 5.35 y 5.36, al dejar el número de estados en 13, el porcentaje de precisión aumentó a 90.35% en relación al 86.66% de las pruebas anteriores y al 71.39% obtenido al inicio (ver tabla 5.30). Esto representa un éxito al lograr un porcentaje superior al 90%. De 4,444 eventos, el sistema ha reconocido 4,105, ha eliminado 318, la mayoría (190) corresponden a ruido, ha sustituido 21, e insertado 90. Cada prueba en el proceso de mejoras ha demostrado, que se ha podido mejorar el porcentaje de precisión (Acc) en el reconocimiento por parte del sistema. Una vez obtenidos los resultados finales (con las modificaciones de parámetros en el proceso de mejoras), el siguiente paso en este proceso, es el depurado selectivo de los registros, en donde se observan de forma manual los resultados obtenidos en el último experimento.

e) Proceso de depurado selectivo de los registros

De acuerdo a los resultados obtenidos hasta ahora en las pruebas, el siguiente paso es identificar de forma manual los errores del sistema y/o aquellas posibles señales en las que el valor del porcentaje de efectividad (Acc) es bajo, procediendo así a eliminarlos de la base de datos y así poder lograr mejorar el porcentaje de precisión (Acc) en el reconocimiento. En la siguiente figura se presenta, a manera de ejemplo un trozo del último fichero de resultados obtenido, en donde se observan los eventos reconocidos por el sistema (evento a evento) comparados con los etiquetados por el observador. En verde se visualizan los registros con bajo porcentaje (menor de 60%).

```

Proceso de Depurado Selectivo en el Telica

----- Sentence Scores -----
===== HTK Results Analysis =====
Rec : rec.13.32.gram.mlf
----- File Results -----
2006-02-16-2038-38ST.rec: 100.00(100.00) [H= 4, D= 0, S= 0, I= 0, N= 4]
2006-02-16-2048-38ST.rec: 100.00(100.00) [H= 5, D= 0, S= 0, I= 0, N= 5]
2006-02-16-2103-38ST.rec: 100.00(100.00) [H= 3, D= 0, S= 0, I= 0, N= 3]
2006-02-16-2108-38ST.rec: 100.00(100.00) [H= 3, D= 0, S= 0, I= 0, N= 3]
2006-02-16-2118-38ST.rec: 100.00(100.00) [H= 3, D= 0, S= 0, I= 0, N= 3]
2006-02-17-1438-30ST.rec: 33.33( 33.33) [H= 1, D= 2, S= 0, I= 0, N= 3]
Aligned transcription: 2006-02-17-1438-30ST.lab vg 2006-02-17-1438-30ST.rec
LAB: NT T3 NT
REC: NT
2006-02-17-1948-27ST.rec: 25.00( 25.00) [H= 1, D= 3, S= 0, I= 0, N= 4]
Aligned transcription: 2006-02-17-1948-27ST.lab vg 2006-02-17-1948-27ST.rec
LAB: NT T4 NT T4
REC: T4
2006-02-17-1953-27ST.rec: 100.00(100.00) [H= 3, D= 0, S= 0, I= 0, N= 3]
2006-02-17-1958-27ST.rec: 100.00(100.00) [H= 3, D= 0, S= 0, I= 0, N= 3]
2006-02-17-2008-27ST.rec: 100.00(100.00) [H= 5, D= 0, S= 0, I= 0, N= 5]
2006-02-17-2013-27ST.rec: 100.00(100.00) [H= 3, D= 0, S= 0, I= 0, N= 3]
2006-02-17-2028-27ST.rec: 100.00(100.00) [H= 3, D= 0, S= 0, I= 0, N= 3]
2006-03-03-1930-50ST.rec: 33.33( 33.33) [H= 1, D= 2, S= 0, I= 0, N= 3]
Aligned transcription: 2006-03-03-1930-50ST.lab vg 2006-03-03-1930-50ST.rec
LAB: NT T2 NT
REC: NT
2006-03-03-1935-50ST.rec: 100.00(100.00) [H= 4, D= 0, S= 0, I= 0, N= 4]
2006-03-03-1945-49ST.rec: 100.00(100.00) [H= 3, D= 0, S= 0, I= 0, N= 3]

```

Fig. 5.36 Sección del fichero de resultados del Análisis manual, en el Proceso selectivo (evento por evento), obtenidos en el Telica. Los valores en rojo representan los registros que han tenido un porcentaje de 100 y que no se cambiarán. Los valores en verde son aquellos que han tenido un porcentaje menor a 60 y que se analizarán para eliminarse de la base de datos.

En la figura anterior se observan tres registros (en verde), en donde se han obtenido porcentajes inferiores al 60% de reconocimiento, por parte del sistema respecto al valor etiquetado por el observador. Los registros en rojo muestran los registros con un reconocimiento al 100%, que han sido definidos por parte del observador, y que el sistema ha coincidido en reconocer los eventos presentes. Un análisis gráfico de dichos registros se observa en las figuras 5.37 y 5.38.

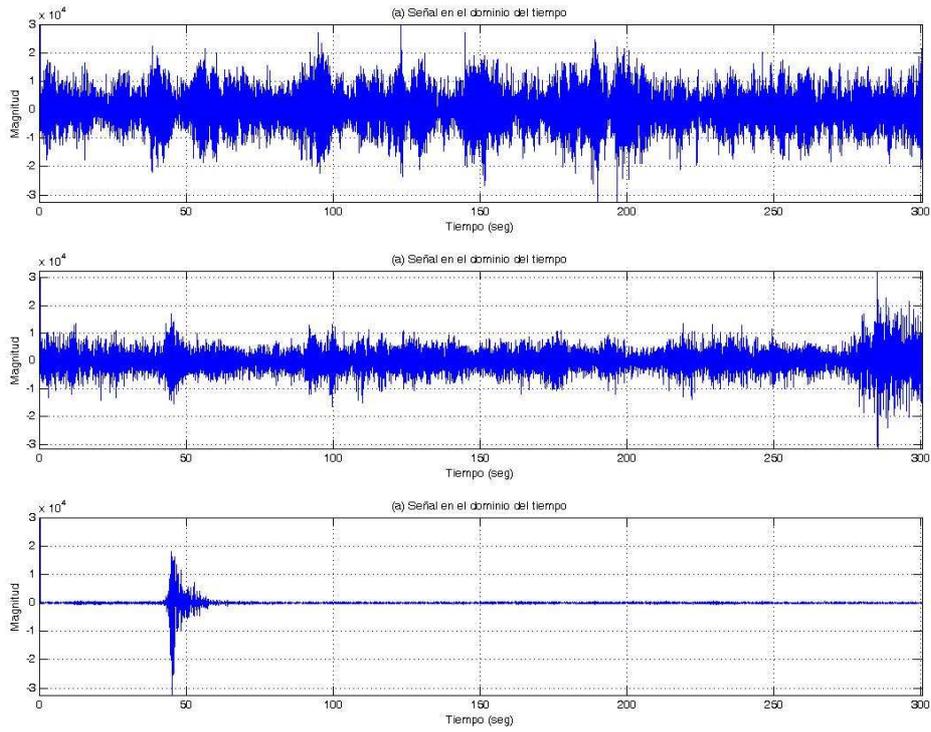


Fig. 5.37 Registros: 2006-02-17-1348-30ST (arriba), 2006-02-17-1948-27ST (centro) y 2006-03-03-1930-50ST (abajo), en el dominio del tiempo.

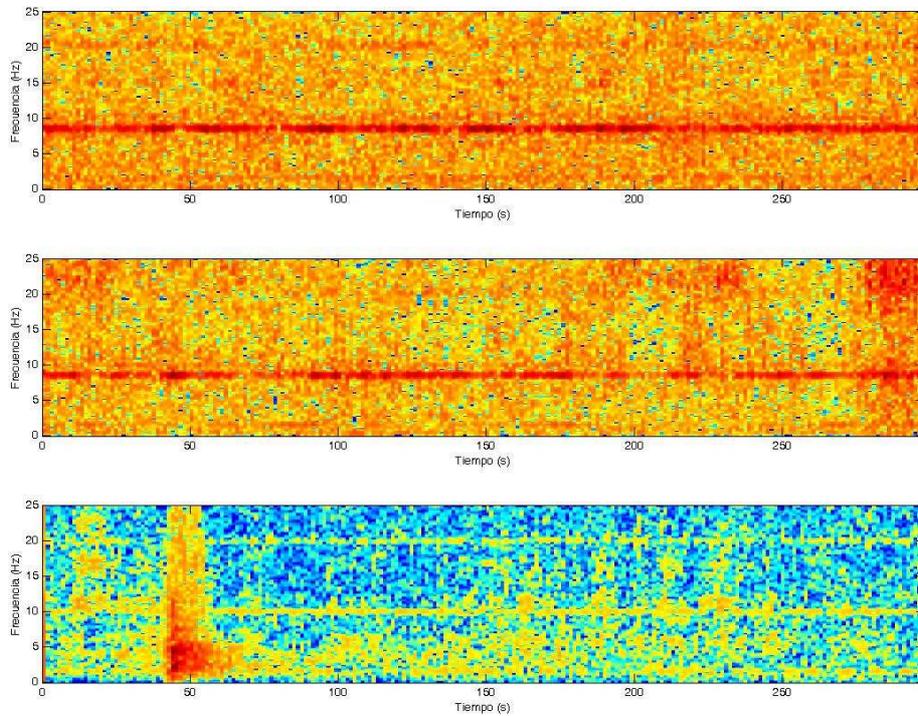


Fig. 5.38 Espectrogramas de los registros: 2006-02-17-1348-30ST (arriba), 2006-02-17-1948-27ST (centro) y 2006-03-03-1930-50ST (abajo).

Las figuras anteriores muestran los tres ejemplos de registros, en los que el sistema no ha podido reconocer eventos marcados por el observador. Un análisis de dichos registros denota como en los dos primeros, el tremor de fondo cubre la totalidad del registro, haciendo imposible el poder reconocer eventos, por lo que ante este tipo de circunstancias el sistema no acierta a reconocer ningún evento y elimina los posibles eventos marcados por el observador. Se está de acuerdo con el sistema, al considerar que este tipo de eventos es muy difícil de determinar. En cuanto al tercer tipo de registro, situado en las figuras en la parte inferior, se observa claramente un evento, sin embargo, hay un ruido de fondo presente que ha sido introducido en el registro debido a un error de grabación por parte del sensor. Este tipo de circunstancia, tampoco es manejada por el sistema, ya que elimina un evento aunque se encuentre claramente definido dentro del registro. Seguidamente, siguiendo la pauta ejercida en los tres anteriores volcanes, se han quitado de la base de datos los registros que han resultado por debajo del 60%, y se repiten las pruebas con los parámetros ya establecidos. El resultado se observa en la tabla 5.37.

```

===== HTK Análisis de Resultados =====
          SENT: %Correct=83.28 [H=986, S=198, N=1184]
WORD: %Corr=94.86, %Acc=93.18 [H=3953, D=200, S=14, I=70, N=4167]
----- Matriz de Confusión -----

```

	NT	T1	T2	T3	T4
NT	2548	1	4	4	1
T1	1	316	1	1	0
T2	0	0	457	0	0
T3	1	0	0	384	0
T4	0	0	0	0	248
Ins	48	4	11	7	0
Del	127	22	21	12	18

Tabla 5.37 Resultados obtenidos al eliminar registros con bajo porcentaje del Telica. Utilizando 6 Iteraciones, 13 estados, 32 Gaussianas y factor $-p = -15$
 %Corr = Porcentaje correcto de etiquetas reconocidas, %Acc = Porcentaje de precisión,
 H= Número correcto de etiquetas, D = Número de eliminaciones, S = Número de sustituciones,
 I = Número de Inserciones, N = Número total de etiquetas. Ins = Inserciones, Del = Borrados

Al observar la tabla anterior, con solo 12 registros de diferencia, y a pesar de que hay menos eventos para entrenar, se han obtenido mejores resultados al eliminar de la base de datos, aquellos registros con menor porcentaje de precisión en el reconocimiento. El porcentaje de efectividad (Acc) aumentó de un 90.35% a un 93.18 %, los elementos eliminados han disminuido de 318 a 200, las sustituciones de 21 a 14, y las inserciones de 90 a 70. De los porcentajes iniciales de 91.24 % (Corr) y 71.68% (Acc), se ha mejorado a 94.86% (Corr) y 93.18% (Acc), lo que indica el éxito y la fiabilidad del sistema en el reconocimiento de los eventos en las pruebas cerradas.

f) Pruebas abiertas (test ciegos)

Las pruebas abiertas, o “test ciegos (blind test)”, están constituidas por los dos tipos de experimentos antes realizados en los tres anteriores volcanes. Recordar que el primero consiste en dividir en dos partes la base de datos. Una primera parte, la formarán los modelos entrenados, para pasarlos después a la parte restante. El segundo

test “ciego-ciego”, se pasará la base de datos entrenada (completa), a una parte de registros no segmentados para observar los tipos de eventos que reconoce el sistema.

i) Test ciego con base de datos entrenada dividida

Para la primera parte de las pruebas abiertas, de la base de datos de 1184 registros, han quedado 784 registros para el entrenamiento, y 400 para el reconocimiento. Los resultados de estas pruebas se observan en la tabla siguiente.

```

===== HTK Análisis de Resultados =====
      SENT: %Correct=73.00 [H=292, S=108, N=400]
      WORD: %Corr=92.07, %Acc=88.78 [H=1288, D=92, S=19, I=46, N=1399]
----- Matriz de Confusión -----
    
```

	NT	T1	T2	T3	T4
NT	847	0	4	4	2
T1	1	132	0	1	0
T2	0	0	117	3	0
T3	0	3	1	142	0
T4	0	0	0	0	50
Ins	20	8	8	9	1
Del	50	18	13	8	3

Tabla 5.38 Resultados obtenidos en las pruebas abiertas del San Cristóbal, utilizando 6 Iteraciones, 13 estados, 32 Gausianas y factor $-p = 5$

%Corr = Porcentaje correcto de etiquetas reconocidas, %Acc = Porcentaje de precisión,
 H= Número correcto de etiquetas, D = Número de eliminaciones, S = Número de sustituciones,
 I = Número de Inserciones, N = Número total de etiquetas. Ins = Inserciones, Del = Borrados

Los resultados en las pruebas indican que el sistema, de 400 eventos ha podido reconocer 292, ha eliminado 92, sustituido 19 e insertado 46. El porcentaje de precisión (Acc) en dichas pruebas ha sido de 88.78%. La mayoría de los eventos eliminados e insertados por el sistema han sido ruido. En la matriz de confusión se observa que el sistema ha podido reconocer con efectividad casi la totalidad de los eventos con lo que el éxito de esta prueba está comprobado.

ii) Test ciego-ciego

El segundo tipo de las pruebas abiertas, el test “ciego-ciego”, los 1184 registros que constituyen la base de datos de entrenamiento, van a reconocer 538 registros adicionales no parametrizados, sin segmentar. El sistema como parámetro de salida ha creado un fichero conteniendo los registros etiquetados. A manera de ejemplo, se han tomado para ser analizados 4 registros de dicho fichero, en donde el sistema reconoce eventos de tipo T1, T2, T3 y T4. A continuación en la figura 5.39 se muestra parte del fichero de resultados.

Test ciego-ciego en el Telica

Evento tipo T1, T2, T3 y T4 encontrados

```
"*/2006-02-27-1721-36ST.rec"  
0 4500000 NT -4567.684082  
4500000 6350000 T3 -1659.446167  
6350000 7150000 NT -850.984619  
.  
"*/2006-02-27-1726-36ST.rec"  
0 650000 NT -642.871643  
650000 1500000 T1 -816.641296  
1500000 4750000 NT -3201.396240  
4750000 7150000 T3 -2249.921631  
.  
"*/2006-02-27-2026-34ST.rec"  
0 1200000 NT -1142.868652  
1200000 5350000 T2 -3901.895264  
5350000 7150000 NT -1841.560059  
.  
"*/2006-02-27-2036-34ST.rec"  
0 2450000 NT -2558.766357  
2450000 3200000 T4 -768.495605  
3200000 7150000 NT -3969.297363
```

Fig. 5.39 Ejemplo de cuatro registros, en el reconocimiento de eventos tipo T1, T2, T3 y T4, por parte del sistema en el Telica, durante las pruebas o test ciego-ciego

La figura anterior, muestra como el sistema ha podido reconocer en los registros eventos de los diferentes tipos clasificados (T1, T2, T3 y T4) en el Telica. Se procede a analizar y determinar gráficamente de forma visual, si el observador coincide con el sistema. En las figuras 5.40 y 5.41, pueden observarse las gráficas de dichos registros.

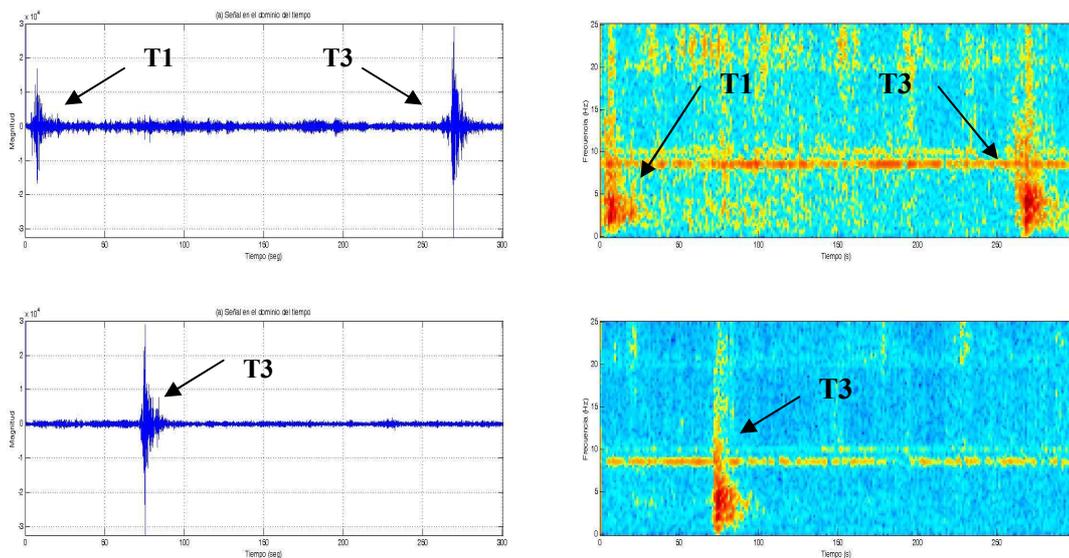


Fig. 5.40 Reconocimiento por parte del sistema, de eventos de tipo T1 y T3 en las pruebas ciego-ciego

La figura anterior muestra a través de la señal en el dominio del tiempo y del espectrograma, como el sistema ha podido reconocer eventos de tipo T1 y T3 presentes en los registros. Aunque de menor magnitud, se puede observar la forma alargada característica del tipo T1. La siguiente figura muestra como el sistema ha reconocido eventos de tipo T2 y T4

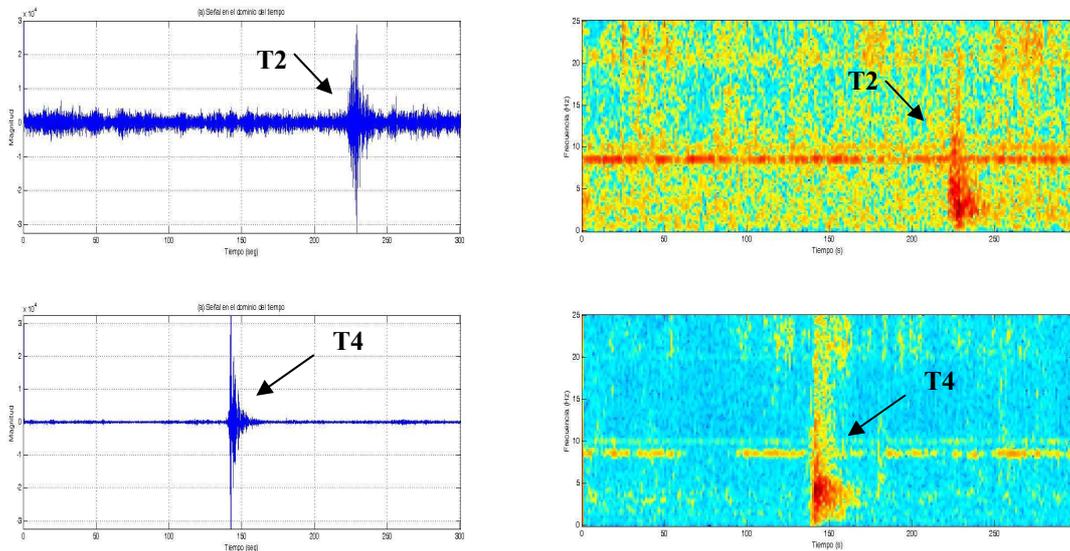


Fig. 5.41 Reconocimiento por parte del sistema, de eventos de tipo T2 y T4 en las pruebas ciego-ciego

La figura anterior muestra como el sistema ha podido reconocer eventos de tipo T2 y T4 en los registros. Puede apreciarse en las gráficas en el espectrograma, la característica forma de cono del evento T2 y la de pico en el T4. En cuanto al ruido, la figura 5.42 muestra una parte del fichero de resultados del test ciego-ciego, en el que el sistema marca como eventos NS, los dos registros.

Registros NT, en el test ciego-ciego del Telica

```

"/2006-03-01-1101-16ST.rec"
0 3100000 NT -4283.853516
3100000 7150000 NT -4630.634766
.
"/2006-03-01-1156-16ST.rec"
0 7150000 NT -7229.134277
    
```

Fig. 5.42 Parte del fichero de resultado en el que se muestran dos registros marcados por el sistema como NT, en el test ciego-ciego.

En la figura 5.43, se presenta gráficamente la señal en el dominio del tiempo y el espectrograma de los registros anteriores, en donde el sistema ha catalogado dichos registros como NT. Así se podrá comprobar porqué el sistema los ha designado como ruido.

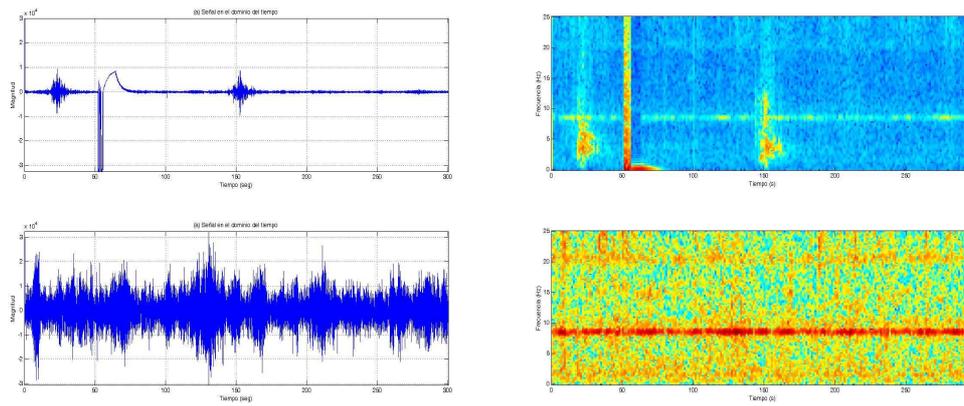


Fig. 5.43 Registros en los que el sistema ha etiquetado eventos NT en el test ciego-ciego.

Se observa que los registros que poseen errores de grabación, como es el caso del primero que se muestra en la gráfica (parte superior), es catalogado como un ruido, de esta forma el sistema desprecia cualquier evento que se encuentre en el, lo que representa un acierto en el reconocimiento, ya que este tipo de registros no ha podido ser almacenado correctamente por parte del sensor. De la misma forma, el sistema cataloga como NT, aquellos registros en los que el ruido de fondo presente se concentra en altas energías. Este es el caso del evento de la parte inferior, en donde el espectrograma de ese registro presenta los diferentes picos de frecuencia característicos de este tipo de ruido, con una banda (en color rojo) en torno a los 10 Hz, que es un tipo de ruido inducido por la antena o por alguna interferencia. El sistema una vez más ha acertado en el reconocimiento, demostrando así su fiabilidad y el éxito en las pruebas de reconocimiento.

5.3.- PRUEBAS CRUZADAS CON LAS BASES DE DATOS

Las pruebas cruzadas tienen la finalidad de poder comprobar si los modelos de entrenamientos de un determinado volcán pueden reconocer los eventos presentes en los registros de otro volcán diferente. Así, con las bases de datos entrenadas, se ha procedido a realizar pruebas, primero con los dos volcanes de Italia, seguidamente con los dos de Nicaragua, para al final presentar las pruebas cruzadas con los cuatro volcanes. De esta forma, se podrá observar si es posible generalizar los modelos de entrenamiento para los cuatro volcanes, o si por el contrario cada uno de ellos necesita un entrenamiento y reconocimiento particular.

5.3.1.- Pruebas cruzadas entre el Etna y el Strómboli

Se procedió a observar si el sistema podría reconocer en el Etna algún tipo de eventos con los modelos entrenados del Strómboli y viceversa. Esto indicaría si las explosiones del Strómboli pueden reconocer algún evento similar en los registros del Etna y si los Tremor Bursts del Etna, podrían reconocer algún tremor de características similares en el Strómboli. Posteriormente se realizará un tercer experimento en donde se mezclarían ambas bases de datos y observar el resultado obtenido. Los resultados de los experimentos combinados se observan en las tablas 5.39 y 5.40.

===== HTK Análisis de Resultados =====
 SENT: %Correct=65.07 [H=434, S=233, N=667]
 WORD: %Corr=90.21, %Acc=84.32 [H=1870, D=192, S=11, I=122, N=2073]

----- Matriz de Confusión -----

	Strómboli		Etna	
	Exp	Sil	TB	NE
Exp	244	2	7	0
Sil	0	601	0	0
TB	0	0	390	5
NE	0	0	4	635
Ins	3	15	18	86
Del	69	73	16	34

Tabla 5.39 Resultados obtenidos del HTK, con el Etna y el Strómboli, en las pruebas combinadas de ambas bases de datos.
 %Corr = Porcentaje correcto de etiquetas reconocidas, %Acc = Porcentaje de precisión,
 H= Número correcto de etiquetas, D = Número de eliminaciones, S = Número de sustituciones,
 I = Número de Inserciones, N = Número total de etiquetas. Ins = Inserciones, Del = Borrados

La figura anterior representa la combinación de las dos bases de datos y el resultado obtenido a través de la Matriz de confusión. Como se puede observar, no se reconocen datos del Etna en las explosiones del Strómboli y tampoco se reconocen explosiones en los eventos TB del Etna. El resultado de las tres pruebas se presenta en la tabla 5.40

Pruebas Cruzadas	Strómboli Registros	Etna Registros	Strómboli + Etna Registros
Modelos del Strómboli	Éxito Corr = 85.73%, Acc = 83.28%	No éxito Menos del 10%	-----
Modelos del Etna	No éxito Menos del 10%	Éxito Corr = 95.02%, Acc = 85.70%	-----
Modelos Strómboli + Etna	-----	-----	Éxito Corr = 90.21% Acc = 84.32% Pero, con reconocimientos individuales

Tabla 5.40 Resultados de la combinación de los modelos en ambos volcanes de Italia.

La tabla 5.40 muestra el éxito en la clasificación de conjuntos independientes de eventos tanto en el Etna como en el Strómboli y también se observa como los modelos entrenados del Strómboli no reconocen eventos en los registros del Etna. El porcentaje de precisión en este caso ha sido menor al 10%, lo mismo sucede con los modelos entrenados del Etna, al no poder reconocer eventos en los registros del Strómboli. Similarmente a los experimentos anteriores en el Strómboli, el porcentaje de precisión en las pruebas ha sido mucho menor al 10%. En las pruebas combinadas al sumar las

dos bases de datos y tener los modelos del Etna y Strómboli, han tenido éxito pero con los reconocimientos individuales que se precisaron en la tabla 5.39. Por los resultados anteriores se podría llegar parcialmente a una premisa inicial que es: Los resultados del entrenamiento de los modelos del Strómboli con su base de datos fueron exitosos. Igualmente, han sido los resultados del Etna con su propia base de registros. Sin embargo, al combinar datos no se han obtenido resultados satisfactorios. Por lo tanto, se puede mencionar que no se debería de generalizar el entrenamiento de modelos con diferentes señales provenientes de distintos volcanes. Aunque estos fueran de características muy similares o correspondan a un mismo tipo geológico.

Esto llevaría seguidamente a mencionar que no sería posible establecer un modelo de entrenamiento universal, el cual fuera posible aplicar a la generalidad de los volcanes, sino más bien se debe de tomar a cada volcán como un caso muy particular y único y realizar de esta forma un análisis y construcción de modelos de entrenamiento particular para cada estructura volcánica. Habría por lo tanto, siguiendo esta premisa que tratar a las señales sísmicas generadas del volcán Etna de una forma particular y realizar el proceso de entrenamiento del Etna con modelos propios creados a partir de una segmentación particular y lo mismo aplicable al Strómboli. En este sentido, y en la búsqueda de poder tener una mejor explicación de por qué se obtuvieron este tipo de resultados, se procedió a realizar un análisis comparativo de las señales de las explosiones ya previamente clasificadas y etiquetadas del Strómboli y compararlas con los eventos identificados en el Etna. Como se ha visto, las explosiones tanto en el Etna como en el Strómboli podrían deberse a movimientos de gases y magma. Por ello, se podría esperar características similares en los resultados de las pruebas. Sin embargo, los resultados obtenidos por los experimentos indicaron que no se identificaron eventos, el sistema, en el Strómboli clasificó la totalidad de las señales de entrada de la estación “d” del Etna como ruido (Sil). Un ejemplo de ello se observa en la figura 5.43, en donde claramente se puede observar cómo en la parte superior de la figura (a) del Etna se acentúa mucho más en el espectro las altas energías que en el caso de la parte inferior (b) que corresponde al Strómboli. Esto nos indica nuevamente como en el caso de las explosiones del Strómboli, la energía está concentrada a mucha más baja frecuencia que en el caso del Etna. De ahí que el sistema de entrenamiento para los modelos del Strómboli, diera como resultado una clasificación de ruido, al encontrarse los modelos del Strómboli entrenados para reconocer los eventos a más baja frecuencia que en el Etna.

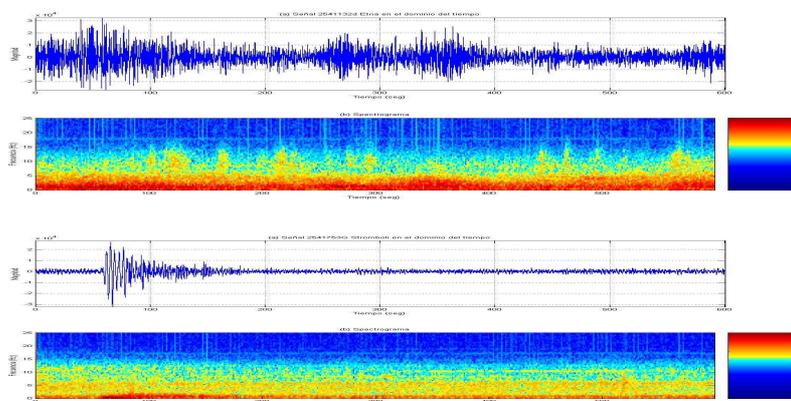


Fig. 5.44 Señal y Espectrograma de eventos 2541132d del Etna en parte superior y 2541753G del Strómboli en la parte inferior.

Al observar la magnitud de potencia del espectro de la FFT de las dos señales representadas en la figura 5.45, puede determinarse que ambas señales no comparten características similares.

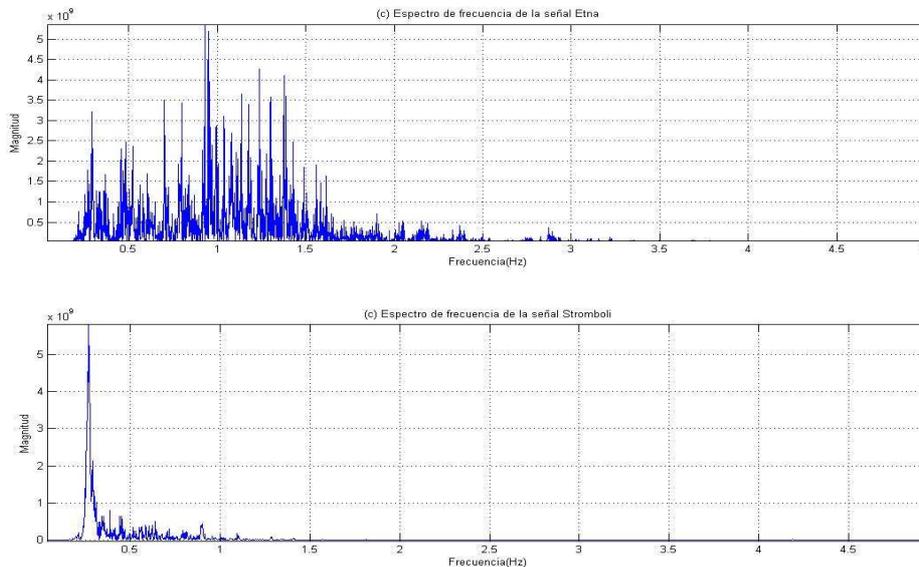


Fig. 5.45 Magnitud de la FFT de los Espectros de potencia de explosiones 2541705d del Etna en parte superior y 2541753G del Strómboli en la parte inferior.

En la figura anterior, se observa como difiere la señal del Strómboli en relación a la del Etna en la magnitud de la FFT del espectro de potencia. Entonces, podemos observar en los resultados tanto de las gráficas anteriores como del experimento, que estando las energías de los eventos del Strómboli concentrados en un espectro de frecuencia mucho menor que la del Etna todo lo que se encuentra por encima del espectro al que han sido entrenados los modelos iniciales, el sistema de entrenamiento lo reconoce como ruido. Por otro lado, como también se puede contemplar en las gráficas del Etna, visualmente se perciben que existen eventos sísmicos claramente definidos y por lo tanto, no se podría decir que ante los resultados de los modelos entrenados del Strómboli se debe de catalogar a todos los eventos del Etna como ruido.

5.3.2.- Pruebas cruzadas entre el San Cristóbal y el Telica

Al realizar las pruebas cruzadas con las bases de datos de los volcanes Etna y Strómboli, estas no se combinaban. Es decir que los modelos de entrenamiento del tremor del Etna no reconocían los eventos de las explosiones del Strómboli aunque se tratase de dos volcanes con características geológicas similares. Como se dispone de cuatro volcanes, ahora se trata de ver qué pasa con los dos volcanes de Nicaragua. También se han realizado pruebas para saber, si los modelos de entrenamiento de los eventos del San Cristóbal pueden reconocer los eventos del Telica y viceversa. De esta forma, poder generalizar el conjunto de bases de datos, obteniendo un modelo de entrenamiento estándar que sea aplicable a los dos volcanes. Se han obtenido los resultados que se presentan en la tabla 5.41 y 5.42.

===== HTK Análisis de Resultados =====
 SENT: %Correct = 75.66 [H=917, S=295, N=1212]
 WORD: %Corr=93.52, %Acc=91.41 [H=4531, D=274, S=40, I=102, N=4845]

----- Matriz de Confusión -----

	SAN CRISTÓBAL				TELICA				
	NS	S1	S2	S3	NT	T1	T2	T3	T4
NS	837	2	8	0	7	0	0	1	1
S1	0	248	1	0	0	0	0	0	0
S2	0	0	298	0	1	0	0	0	0
S3	0	0	0	84	0	0	0	0	0
NT	10	0	0	0	1964	4	3	0	0
T1	0	0	0	0	0	232	0	1	0
T2	0	0	0	0	0	0	351	0	0
T3	0	0	0	0	0	0	0	308	0
T4	0	0	0	0	0	0	0	0	209
Ins	19	0	3	0	67	1	6	5	1
Del	57	11	16	0	118	22	24	13	13

Tabla 5.41 Resultados obtenidos del HTK los experimentos de pruebas cruzadas con las bases de datos de los modelos de entrenamiento del Telica y del San Cristóbal.

%Corr = Porcentaje correcto de etiquetas reconocidas, %Acc = Porcentaje de precisión,
 H= Número correcto de etiquetas, D = Número de eliminaciones, S = Número de sustituciones,
 I = Número de Inserciones, N = Número total de etiquetas. Ins = Inserciones, Del = Borrados

Un resumen final de los resultados de efectividad obtenidos con los diferentes experimentos realizados en ambos volcanes puede observarse en la tabla 5.42.

MODELOS	REGISTROS		
	TELICA	SAN CRISTOBAL	REGISTROS COMBINADOS
ENTRENAMIENTO	83.29%	87.99%	--
TEST	74.48%	88.92%	--
BD ENT TELICA	<u>Éxito</u>	<u>No éxito</u>	--
BD ENT SAN CRISTÓBAL	<u>No éxito</u>	<u>Éxito</u>	--
BD COMBINADA TELICA Y SAN CRISTÓBAL	--	--	<u>Éxito ENT</u> 91.41% Pero solo con reconocimientos individuales

Tabla 5.42 Resumen de resultados obtenidos en las diferentes pruebas del Telica y San Cristóbal.
 BD = Base de datos, ENT = Entrenamiento.

La tabla 5.41 muestra que en la prueba combinada o cruzada (al construir una base de datos combinada), se emplearon 1,212 registros de los dos volcanes. De ellos fueron reconocidos correctamente 917 registros. En cuanto a eventos, fueron

introducidos 4,845 eventos, de los cuales el sistema reconoció correctamente 4,531. Además el sistema eliminó 274, sustituyó 40 e insertó 102. La mayoría de los eventos eliminados e insertados se trataron de ruido de fondo NS y NT. El porcentaje de eventos correctamente reconocidos (Corr) ha sido de 93.52% y el porcentaje de fiabilidad o precisión (Acc) de 91.41%.

En la tabla 5.41 se puede observar que del mismo modo que los resultados obtenidos en las pruebas con el Etna y el Strómboli, los modelos del San Cristóbal solo han podido reconocer sus propios eventos y lo mismo sucede con los modelos del Telica. Es decir, que la confusión entre eventos por parte del sistema es mínima. Solamente al tratarse del ruido de fondo se presentaron unos pocos elementos de confusión, pero al tratarse de los eventos sísmicos ninguno de ellos se confundió con los del otro volcán. Esto corrobora la teoría anteriormente mencionada de que para cada volcán se precisa una base de entrenamiento particular o propio, la cual debe de provenir de sus propios eventos. En la tabla 5.42, se observa que se ha logrado obtener un sistema fiable en los entrenamientos de los eventos del volcán Telica 74.48% y un sistema bastante más fiable en los entrenamientos de los eventos del San Cristóbal 88.92%. Las bases de datos de forma individual reconocen y clasifican los eventos presentes en los registros del volcán al que pertenecen. Las pruebas han demostrado, que los modelos de entrenamiento funcionan mucho mejor con el San Cristóbal que con el Telica. Por otra parte, las pruebas cruzadas han tenido éxito 91.41% en el entrenamiento pero solo con reconocimiento individual de cada base de datos. Análogamente, los modelos de entrenamiento del Telica no reconocen los registros del San Cristóbal y viceversa. Esto comprueba nuevamente la no compatibilidad de los eventos sísmicos presentes en el San Cristóbal y el Telica para el reconocimiento cruzado. Los volcanes, independientemente de sus características geológicas, poseen además muchos factores adicionales como son: geográficos, de forma, de composición del suelo y contenido del mismo, fallamiento, existencia o no de depósitos acuíferos, etc., que rigen e influyen en la formación y características de sus eventos (o su particular constitución fonadora, o de la forma en que produce los eventos sísmicos), haciéndolos únicos a la hora de constituir una base de datos de modelos de entrenamiento, aunque los eventos sean similares a otros volcanes y queden dentro de alguna escuela de clasificación sísmica, poseen una particularidad que es intrínseca a sí mismo.

5.3.3.- Resultado de las pruebas cruzadas con los cuatro volcanes (Etna, Strómboli, San Cristóbal, Telica)

Ahora bien, un experimento final consistió en combinar las cuatro bases de datos. Esto para poder observar cómo se comportaban los eventos en los entrenamientos. La pregunta ahora será, ¿Podrían los modelos de entrenamiento de Nicaragua reconocer los eventos de Italia y viceversa?

Para estos experimentos con las cuatro bases de datos, los tipos de eventos encontrados en los registros segmentados difieren un poco de los presentados en las tablas iniciales. Esto es debido a que para estos experimentos finales se incluyeron registros adicionales en los volcanes de Italia en función de un mayor incremento en el número de eventos en dichas bases de datos. Esto se observa en la tabla 5.43.

NICARAGUA				ITALIA			
Telica		San Cristóbal		Etna		Strómboli	
T1	360	S1	224	TB	388	Exp	339
T2	502	S2	330	NE	639	Sil	746
T3	429	S3	544				
T4	298	NS	1746				
NT	2855						

Tabla. 5.43 Número de eventos incluidos en las pruebas cruzadas con los cuatro volcanes.

Esto indica que vamos a tener para los entrenamientos de los modelos un total de 9,400 eventos incluyendo los ruidos de fondo. Con estos datos se ha procedido a realizar los experimentos. El resultado de estos se muestra en la tabla 5.44.

```

===== HTK Análisis de Resultados =====
SENT: %Correct = 63.20 [H=1673, S=974, N=2647]
WORD: %Corr = 87.22, %Acc = 84.52 [H=8199, D=1000, S=201, I=254, N=9400]

```

----- Matriz de Confusión -----

		NICARAGUA										ITALIA			
		SAN CRISTÓBAL				TELICA						ETNA		STROMBOLI	
		NS	S1	S2	S3	NT	T1	T2	T3	T4	NE	TB	SIL	EXP	
NICARAGUA	SAN CRISTÓBAL	NS	1498	1	1	40	30	0	0	3	0	0	1	1	0
		S1	0	211	0	1	0	0	0	0	0	0	0	0	0
		S2	0	0	300	1	0	0	0	0	0	0	0	0	0
		S3	0	0	0	497	1	0	0	0	0	0	0	0	0
	TELICA	NT	58	2	1	4	2409	9	7	13	3	0	0	0	0
		T1	1	0	0	0	2	293	0	0	0	0	0	0	0
		T2	0	0	0	0	1	1	445	1	2	0	0	0	0
		T3	2	0	0	0	2	2	1	391	0	0	0	0	0
ITALIA	ETNA	T4	0	0	0	0	0	0	1	254	0	0	0	0	
		NE	0	0	0	0	0	0	0	0	0	624	0	1	0
	STROMBOLI	TB	0	0	0	0	0	0	0	0	0	0	364	0	0
		SIL	0	0	0	0	0	0	0	0	0	0	0	611	2
		EXP	0	0	0	0	0	0	0	0	0	0	4	302	
		INS	71	0	3	4	77	9	7	9	4	18	7	39	6
		DEL	171	12	29	46	349	64	52	31	43	14	23	133	33

Tabla 5.44 Resultados obtenidos del HTK los experimentos de pruebas cruzadas con las bases de datos de los modelos de entrenamiento de Italia y Nicaragua.

%Corr = Porcentaje correcto de etiquetas reconocidas, %Acc = Porcentaje de precisión,
H= Número correcto de etiquetas, D = Número de eliminaciones, S = Número de sustituciones,
I = Número de Inserciones, N = Número total de etiquetas. Ins = Inserciones, Del = Borrados

En los entrenamientos y reconocimientos de los modelos, se obtuvo una eficacia o precisión (Acc) aproximada de un 85%. Si se toma en cuenta que la mayoría de los eventos eliminados e insertados por el sistema corresponden a ruidos, tanto en

Nicaragua como en Italia (NS, NT, NE, SIL) esto indicaría que el porcentaje en efectividad (Acc) estaría cercano al 90%. Lo más importante a destacar como se puede observar en la tabla anterior, ha sido que ni los modelos de Italia reconocen a los eventos de Nicaragua, ni los modelos de Nicaragua reconocen a los eventos de Italia. La matriz de confusión, no presenta eventos significativos que hayan sido mezclados en los cuatro volcanes. De hecho, ni entre ellos, como se ha visto en las anteriores pruebas, han presentado confusiones de eventos. Esto indica lo que ya se había destacado, que para el entrenamiento de los modelos, cada volcán precisa de una base de datos de entrenamiento propia. En este sentido, el entrenamiento y clasificación de los eventos debe de ser particular para cada volcán que se va a estudiar. Con esto se sentó una premisa de que cada volcán, precisa de una base de datos de entrenamiento y modelos particulares, ya que al igual que la voz en el humano poseen características únicas en cada individuo, aun tratándose de los mismos fonemas, sucede lo mismo en los edificios volcánicos. Cada volcán se debe de entrenar con una base de datos propia de eventos producidos por él. A futuro por ejemplo para Nicaragua, esto ha de tomarse muy en cuenta, ya que se debe preparar una base de datos propia en cada uno de los edificios de la cadena volcánica.

5.4.- DISCUSIÓN DE RESULTADOS

El llegar a los resultados finales ha involucrado un largo proceso, que ha pasado por analizar uno a uno más de 15,000 registros sísmicos de cuatro volcanes, de donde se han seleccionado más de 2,300. Con ellos se han segmentado y etiquetado de forma manual más de 7,800 eventos distribuidos en 13 diferentes tipos, que fueron previamente catalogados y analizados, incluyendo el ruido de fondo. Una vez hecho esto, con las bases de datos creadas y los modelos entrenados se realizaron cerca de 10,000 diferentes experimentos.

Las diferentes pruebas o experimentos que se han realizado con los volcanes han determinado que no en todos los casos, los parámetros a establecer van a ser los mismos. Esto es, que no se puede determinar una generalización en cuanto al establecimiento de parámetros para todos los volcanes. Esto se observa en la tabla 5.45

Parámetro	Etna	Strómboli	San Cristóbal	Telica
Iteraciones	6	6	6	15
Gaussianas	32	32	32	32
-p	-15	-15	5	-5
estados	13	15	13	13

Tabla 5.45 Variación del valor de los parámetros en las pruebas en cada volcán

Por ejemplo, los mejores resultados en el número de iteraciones ha sido de 6 en los primeros tres volcanes no así en el Telica, que el mayor porcentaje se dio en 15. El factor de penalización (-), ha sido diferente, pues mientras que para el Etna y Strómboli se ha situado en -15, para el San Cristóbal ha sido de 5 y para el Telica de -5. El número de estados de los modelos de entrenamiento también ha variado en el Strómboli respecto a los otros tres volcanes.

Respecto a las pruebas iniciales, las modificaciones hechas en el proceso de mejora han resultado muy efectivas. Esto se refleja en la tabla 5.46, donde se presenta que los porcentajes finales han mejorado un promedio casi de un 8 % respecto a los iniciales en los resultados correctos, y aumentó un promedio cercano al 20 % en la precisión respecto a los valores iniciales.

Parámetro		Etna		Strómboli		San Cristóbal		Telica	
		Inicial	Final	Inicial	Final	Inicial	Final	Inicial	Final
Pruebas Cerradas	% Corr	90.77	94.07	76.76	93.39	84.89	94.96	85.66	94.86
	%Acc	69.93	93.29	71.36	92.32	80.57	93.74	71.39	93.18

Tabla 5.46 Resultados iniciales y finales obtenidos en las pruebas en cada volcán
%Corr = Porcentaje correcto de etiquetas reconocidas, %Acc = Porcentaje de precisión.

De la tabla anterior, puede calcularse el aumento positivo en los porcentajes individuales en cada volcán al realizar el proceso de mejoras. Los valores en los resultados son los siguientes:

Etna	% Corr	3.30	% Acc	23.36
Strómboli	% Corr	16.63	% Acc	20.96
San Cristóbal	% Corr	10.07	% Acc	13.17
Telica	% Corr	9.20	% Acc	21.79

Esto indica el éxito del proceso de mejoras que se ha desarrollado. La tabla 5.47 muestra que en las pruebas cerradas, han podido obtenerse porcentajes de efectividad superiores al 90%, con un margen +5% (respecto al valor de las eliminaciones de ruido y las inserciones de eventos), lo que podría aumentar el valor real de las pruebas. En las “pruebas abiertas” o también denominadas “test ciegos”, el resultado general en los cuatro volcanes que se observa en la tabla 5.47, ha superado el 88% en cada uno de los volcanes, lo que demuestra la efectividad del sistema. De la misma forma, y siguiendo la premisa de aumentar un 5% respecto al valor de las eliminaciones de ruido y las inserciones de eventos, el valor de los resultados en las pruebas abiertas está cercano o superior al 90%. El sistema ha sabido clasificar y reconocer los eventos aun en circunstancias donde el ruido es mucho mayor, sin embargo, un elemento a mejorar se presenta ante el ruido en altas energías, en donde el sistema no puede reconocer eventos aun y cuando se encuentren presentes en los registros. Esto podría determinarse en trabajos futuros, al disponer de una base de datos mucho mayor, y la cual contenga más elementos con estas características. Lo que no pudo hacerse en el presente trabajo, ya que únicamente se dispone de registros que corresponden únicamente como máximo a dos semanas de almacenamiento, producto de las cuatro campañas realizadas en Italia y Nicaragua.

Parámetro		Etna	Strómboli	San Cristóbal	Telica
Pruebas Cerradas	% Corr	94.07	93.39	94.96	94.86
	%Acc	93.29	92.32	93.74	93.18
Pruebas Abiertas	% Corr	91.13	90.15	90.37	92.07
	% Acc	89.72	88.64	84.69	88.78

Tabla 5.47 Resultados finales obtenidos en las pruebas en cada volcán
%Corr = Porcentaje correcto de etiquetas reconocidas, %Acc = Porcentaje de precisión.

Por otra parte, el sistema, en las pruebas ciego-ciego, ha podido reconocer cada uno de los diferentes tipos de eventos presentes en los registros (sin parametrizar) de los cuatro volcanes. Por ejemplo, en las pruebas de los modelos, en el reconocimiento sobre el resto de los registros no segmentados, fue característico encontrar lo siguiente:

- a) El sistema realizó la clasificación automática de todas las restantes señales, identificando eventos y señalando segmentos de ruido. En este sentido, identificó eventos en donde el observador no los había clasificado.
- b) Una de las limitaciones observadas durante los primeros experimentos, fue que el sistema no podía discriminar las altas energías presentes en señales ruidosas, o en registros que por alguna causa no se almacenaron correctamente. A manera de ejemplo, se presentan los registros del Etna: 2591214d y 2591541d, que se presentan en las figuras 5.46 y 5.47. Sin embargo, lo anterior fue superado a medida que se realizó el proceso de mejoras en el sistema.

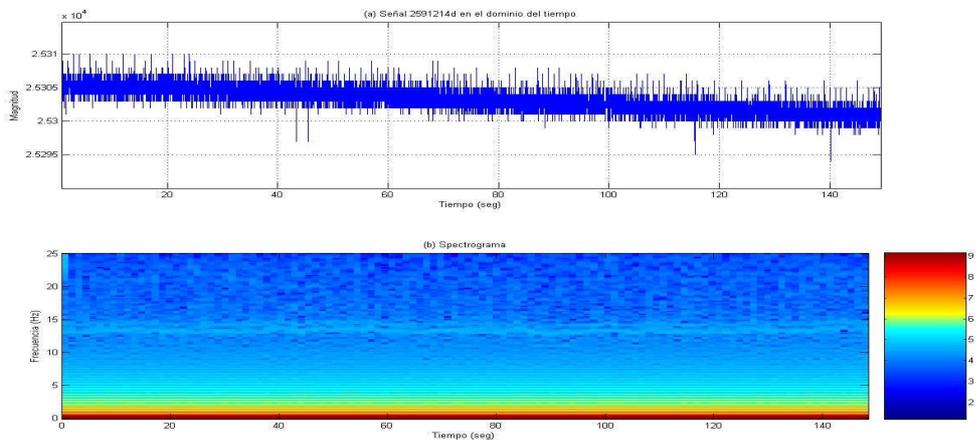


Fig. 5.46 Señal ruidosa 2591214d en el Etna.

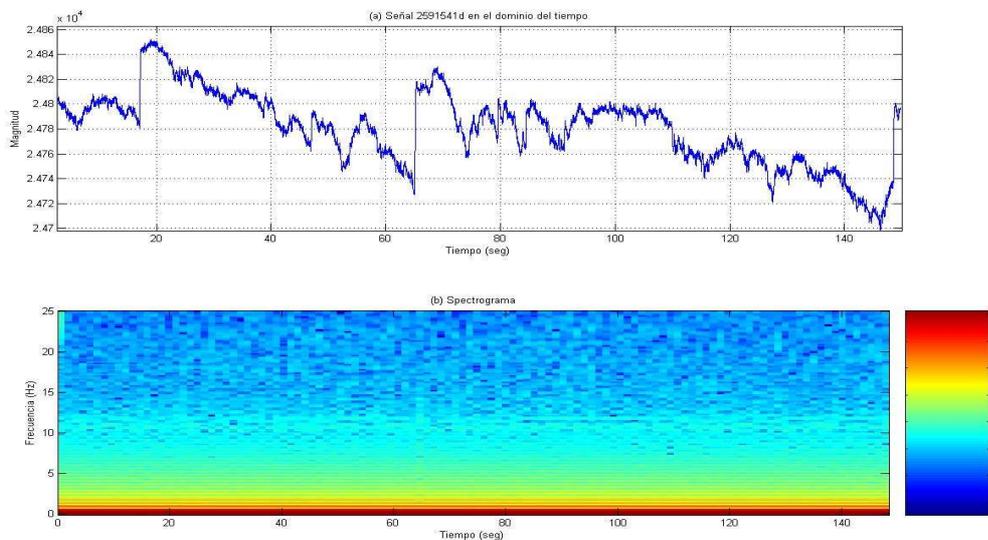


Fig. 5.47 Señal ruidosa 2591541d en el Etna.

En el tipo de señal que se muestra en las figuras anteriores, el sistema inicialmente detecta posibles eventos de tipo TB en el Etna, lo que es incorrecto. En las siguientes pruebas, al realizar los procesos de mejoras, el sistema podía discernir y catalogar dichas señales como ruido (NT). En este tipo de señales, el ruido permanece en componentes de baja energía, y el sistema tendía a confundirlo con eventos de tipo TB, lo que se observa tanto en la forma de la señal en el tiempo (superiores en las figuras) como en el espectrograma (inferiores en las figuras). Este tipo de registros podría deberse a un error de medición de los datos al ser almacenados en la fuente. Al realizar las mejoras con las modificaciones del parámetro de las gaussianas, las siguientes pruebas mostraron que se superó este tipo de confusión. El sistema mejoró, al descartar la existencia de eventos TB en este tipo de señales, clasificándolas únicamente como ruido (NE). Esto significó un acierto en el entrenamiento de los modelos y reconocimiento por parte del sistema.

Otro aspecto que se presentó en las pruebas abiertas, por ejemplo en el Etna, en donde se indicó inicialmente que podría haber un error por parte del sistema, y que sin embargo, en este estudio se ha considerado como un éxito, ha sido la forma en la que el sistema clasificó algunas explosiones presentes en las señales que correspondían a registros con componentes de altas energías, las que podrían catalogarse como explosiones o terremotos y no propiamente eventos del tipo Tremor (hay que recordar que en el Etna, al igual que en todos los volcanes, se producen diferentes tipos de eventos). Esto se observa en la figura 5.48 y 5.49, y ocurrió debido a que el sistema no contaba con ningún otro parámetro de referencia, al haberse entrenado los modelos únicamente con dos tipos de eventos (TB y NE). En los últimos experimentos, el sistema logró descartar la mayoría de este tipo de señales, aunque a riesgo de confusión identificó como eventos tipo TB, señales con componentes de energía elevados. Por ejemplo, en el registro sin previo procesamiento 2541705d, el sistema reconoció un evento de tipo TB (figura 5.48). Sin embargo, en la señal 2541708d con componentes de altas frecuencias, el sistema pudo descartar un evento TB y lo asignó como NE (figura 5.49).

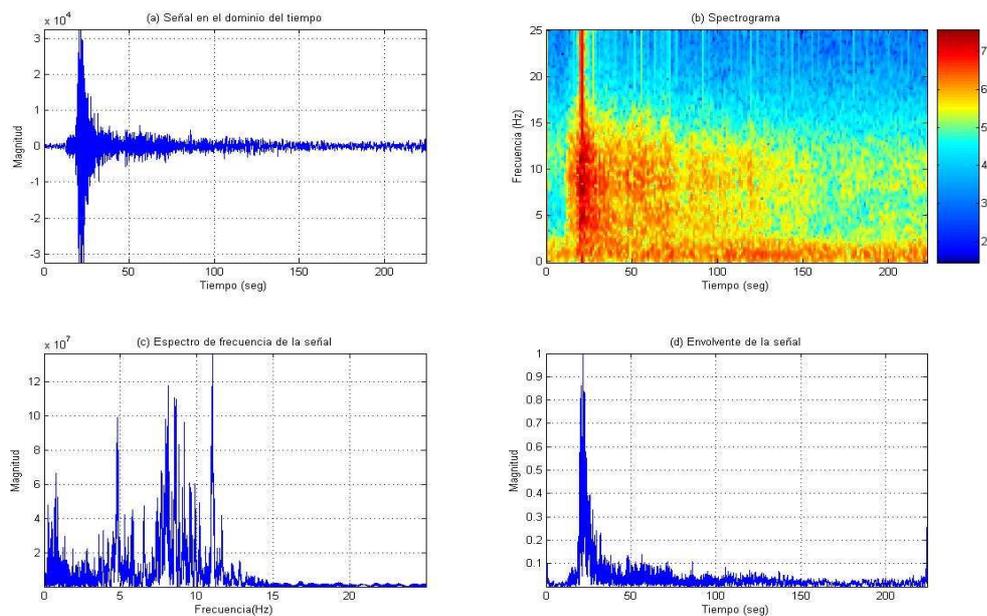


Fig. 5.48 Señal del registro 2541705d en el Etna. El sistema reconoce un evento TB.

Esto sucede en cada uno de los volcanes, al presentarse en los registros algunos tipos de eventos, que han sido excluidos por el observador en el proceso de la construcción de la base de datos, debido a la poca cantidad existente.

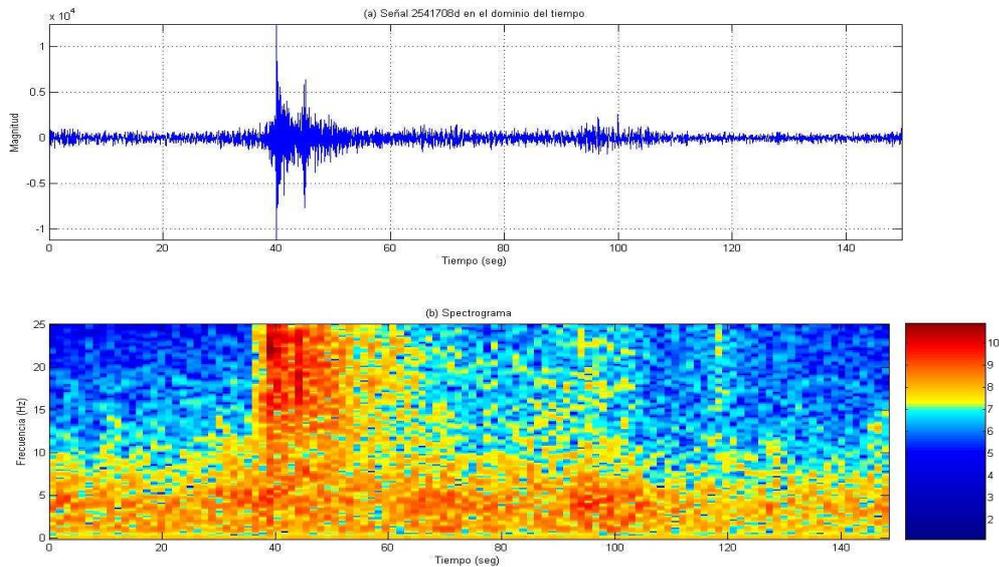


Fig. 5.49 Señal y Espectrograma del registro 2541708d en el Etna

Estos tipos de registro en el Etna, así como algunas explosiones del San Cristóbal, fueron pasados a los modelos de entrenamiento del Strómboli en las pruebas cruzadas, para observar si los modelos entrenados del Strómboli reconocían a los registros como explosiones. Sin embargo, los modelos de entrenamiento del Strómboli catalogaron a dichos registros como “sil”, o ruido. Esto es una de las mejoras a ser incluidas en posteriores trabajos, cuando se puedan incluir en la construcción de las bases de datos y en el entrenamiento de los modelos, muchos más eventos de diversos tipos, lo que no se pudo realizar en el presente estudio, ya que en la recopilación de datos, no se disponía de la suficiente cantidad y variedad, para poder realizar una base de datos lo suficientemente amplia, con más tipos de eventos, y así poder realizar las pruebas y mejoras adicionales.

Otro de los aciertos del sistema, incluso desde las primeras pruebas cerradas, ha sido el poder reconocer eventos, en prácticamente la totalidad de registros en donde el observador los había indicado. Esto se observó, mediante un análisis de los registros en los que originalmente se etiquetaron uno, dos o más eventos. En los cuales, después de las pruebas, se pudo comprobar que el sistema reconoció de forma automática el mismo número de eventos que el observador.

Otro aspecto a tener en cuenta, es la forma en que el observador y el sistema han marcado el inicio y final de cada evento. Esto se puede observar en las figuras 5.50 a 5.53.

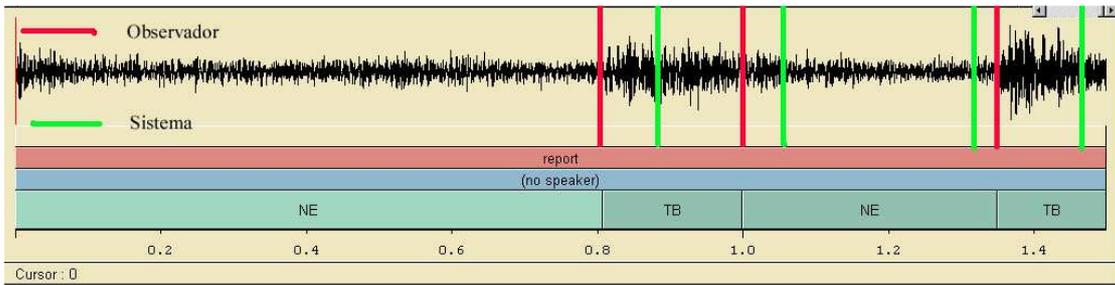


Fig. 5.50 Determinación de parte del observador y el sistema, de los dos eventos TB presentes en el registro 2541344d.

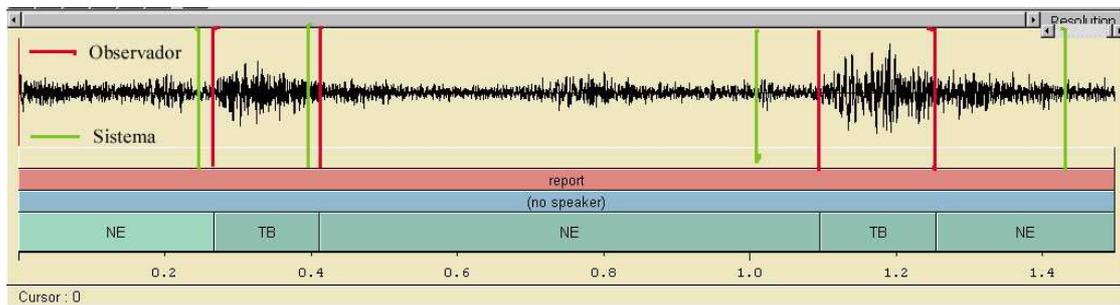


Fig. 5.51 Determinación de parte del observador y el sistema de los dos eventos TB presentes en el registro 2570114d.

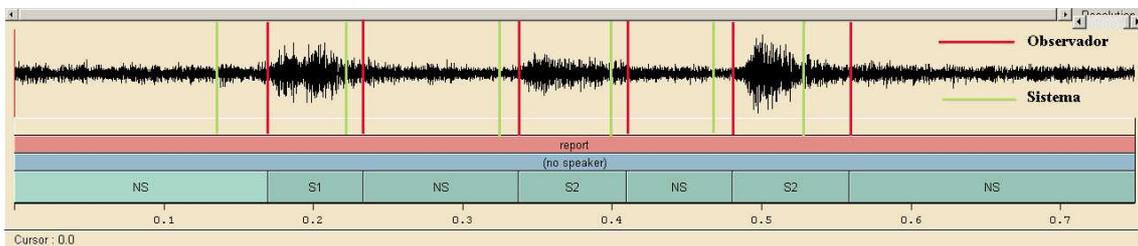


Fig. 5.52 Marcación por parte del Observador y el Sistema de eventos en el San Cristóbal

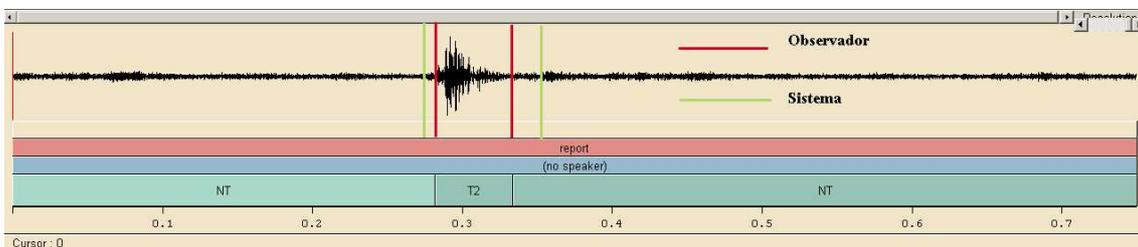


Fig. 5.53 Marcación por parte del Observador y el Sistema de eventos en el Telica

Como se puede analizar en las anteriores figuras (fig. 5.50 a 5.53), el sistema se acerca mucho al criterio del observador al marcar los eventos. Difiere algo en cuanto al inicio y el final, pero no a la marcación del evento propiamente dicho. Este

comportamiento se repite en casi la totalidad de los resultados. En los registros en los que hay más de dos eventos presentes, el sistema y el observador coinciden en el número de eventos encontrados. El sistema de forma automática reconoce eventos donde el observador también los ha identificado.

Otro aspecto que representa un éxito por el sistema, ha sido en las pruebas cerradas, y se produjo al reconocer eventos en donde no habían sido marcados originalmente por el observador. El sistema ya entrenado, puede reconocer todos los tipos de los eventos clasificados en los registros sin etiquetar, lo que constituye uno de los éxitos. De esta forma, el sistema ha podido identificar, uno, dos o más eventos en cada registro.

Un aspecto más a destacar, ha sido determinar que el número de inserciones en los resultados, no necesariamente se trata de un fallo en el sistema, sino al contrario, indica que el sistema está reconociendo eventos donde el observador no los había marcado, por lo que el porcentaje de inserciones de eventos, tendría que aumentar el porcentaje de precisión por parte del sistema. Por lo tanto, se determina como un éxito, la mayoría de las inserciones que el sistema ha realizado en las pruebas, ya que establece que el sistema sabe reconocer eventos y así lo indican los resultados.

Otro de los aspectos a enfatizar en las pruebas, ha sido el ruido de fondo presente en los volcanes. En los registros del Etna y Strómboli, los del Strómboli en comparación de los del Etna están menos acompañados de ruido o tremor de fondo. Lo mismo sucede con los del San Cristóbal y el Telica, en donde los del San Cristóbal están mucho más enmascarados por el tremor de fondo que los del Telica. Esto indicaba además, que muchos de los eventos estaban acompañados de un tremor de fondo en altas y bajas energías, lo que hacía difícil el poder catalogarlos. Podría decirse que de los cuatro volcanes, dos de ellos, el Etna y el San Cristóbal presentan una característica común: los eventos los acompaña un fuerte tremor de fondo. En cambio, tanto el Strómboli como el Telica, presentan señales mucho más definidas, como son: el Strómboli con las explosiones y el Telica, con los cuatro tipos de eventos clasificados. Aún así, los modelos de entrenamiento han sabido diferenciar este tipo de registros, al reconocer los diferentes tipos de eventos presentes en los cuatro volcanes, aun con la presencia del ruido de fondo.

Otra particularidad que se ha encontrado al realizar las pruebas, ha sido la cercanía de eventos de características similares. Este tipo de eventos se han descartado a la hora de la construcción de la base de datos, ya que eventos muy cercanos se prestan a confusión por parte del sistema, al menos en los encontrados en los registros analizados. Posibles futuros trabajos también podrían incluir más registros con estas características, a fin de que el sistema pueda ser entrenado con estas diferencias.

Ahora bien, en todas las pruebas, los resultados estuvieron dados en base a los parámetros que se configuraron en los modelos, conforme a la experiencia del entrenamiento realizado durante los primeros experimentos y teniendo como premisa el hecho de que no todas las bases de datos y los volcanes se comportan de la misma manera, además de que los instrumentos en la recopilación de datos difieren ligeramente al almacenar los registros y que las pruebas son realizadas en registros continuos. Así, con todo ello, los resultados obtenidos se catalogan de bastante exitosos. El proceso de mejoras que se realizó con los volcanes, con el objetivo de afinar los modelos y el

reconocimiento de los eventos, conforme a los parámetros específicos de cada volcán, produjo como se observó en la tabla 5.46, éxitos en precisión (Acc) superiores al 90%.

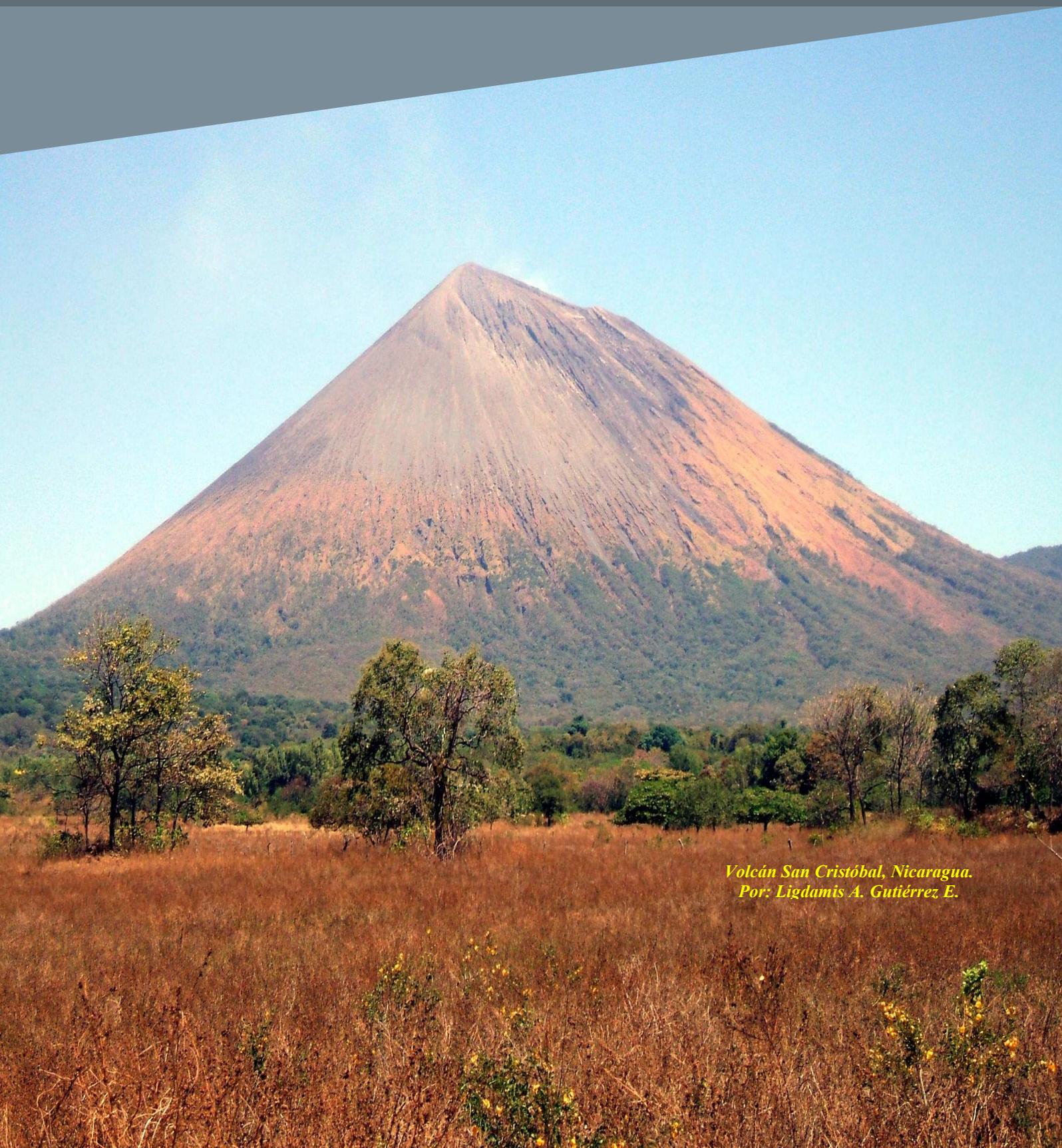
Por otro lado, uno de los objetivos fundamentales de este trabajo es universalizar el proceso de reconocimiento automático para cualquier ambiente volcánico. Como se ha observado en los anteriores cuatro análisis de resultados con los volcanes, el proceso de clasificación automático se ha realizado en base a la existencia de una base de datos completa, grande y de alta calidad. Sin embargo, esta situación no siempre es posible en otros volcanes o en otros períodos de los mismos volcanes. A veces la actividad sísmica es escasa o el registro sísmico no ha sido lo suficientemente grande como para mantener una base de datos adecuada. En este sentido, los resultados expresados en las pruebas cruzadas, como la primera parte lo indica, los modelos no son exportables y como la segunda lo reafirma, habría que utilizar más volcanes y muchas más bases de datos a fin de poder observar a futuro la universalización que se pretende establecer.

Por lo tanto, se necesita a futuro la recopilación de más datos, con una mayor cantidad de estaciones, y con instrumentos de mucha mayor precisión, lo que permitiría mejorar aún más los resultados que se han obtenido en el presente trabajo. También, trabajos futuros deberán incluir mejoras en los algoritmos para poder delimitar aun más el inicio y el fin de cada evento a reconocer. El sistema sin embargo, como se ha demostrado, ha podido reconocer la existencia de varios tipos de eventos presentes en un determinado registro, incluso en aquellos que el observador no ha etiquetado. Esto se presentará en el siguiente capítulo en referencia a las conclusiones y mejoras que trabajos a futuro deberán de incluir.



CAPÍTULO VI

CONCLUSIONES Y TRABAJOS A FUTURO



*Volcán San Cristóbal, Nicaragua.
Por: Ligdamis A. Gutiérrez E.*



6.1.- CONCLUSIONES Y LOGROS

Con los resultados recabados en el presente trabajo, han podido definirse las siguientes conclusiones metodológicas y experimentales:

6.1.1.- Conclusiones Metodológicas

- a) Esta tesis ha demostrado que el proceso de reconocimiento automático de señales sísmico-volcánicas, puede realizarse de manera fiable y precisa mediante el uso de las técnicas de reconocimiento basadas en modelos HMM.
- b) A partir de la realización de la aplicación del conjunto de datos, hemos demostrado que con el fin de tener un resultado fiable, es crucial un adecuado proceso de segmentación.
- c) Ha sido posible generar modelos de entrenamiento con éxito en dos volcanes activos de Italia, utilizando cuatro tipos de eventos incluyendo el ruido de fondo. De la misma forma, se han generado modelos de entrenamiento efectivos, en los volcanes San Cristóbal y Telica de Nicaragua, utilizando 9 tipos diferentes de eventos incluyendo el ruido de fondo.
- d) Se determina la utilidad como herramienta para establecer los tipos de eventos sísmico-volcánicos, del análisis espectral basado en la transformada rápida de Fourier y el espectrograma, ya que dan una información útil en primera aproximación.
- e) En esta tesis se han explorado otras herramientas para complementar el análisis de la caracterización de las señales presentes en volcanes, fundamentalmente análisis basados en densidad de espectro de potencia y la transformada Wavelet.
- f) Se establece que los parámetros de configuración, utilizados en las diferentes pruebas al mejorar el porcentaje de los modelos difieren en cada volcán, lo que determina que no necesariamente puede establecerse un conjunto único de configuraciones para todos los volcanes, necesitando entrenar cada base de datos, de acuerdo a unas características específicas de las señales sísmico-volcánicas presentes en cada volcán.
- g) Se ha podido estimar la valía de la caracterización espectral de las señales sísmico-volcánicas, para poder constituir una base de datos fiable. Esto ha sido particularmente útil en eventos desconocidos, generados por los dos volcanes activos de Nicaragua, y que aún no han sido clasificados ni estudiados a fondo a diferencia de los volcanes de Italia, en los que diversos autores en trabajos previos han catalogado y estudiado profusamente los eventos producidos por ellos.
- h) Se ha presentado un extenso análisis, de cómo las componentes espectrales de las señales que constituyen los eventos sísmico-volcánicos en los cuatro volcanes de estudio, presentan características únicas, las cuales se diferencian significativamente en todos los volcanes. Esto establece que aunque los

volcanes sean de características geológicas muy similares y estén situados geográficamente muy cercanos, los eventos que producen, difieren.

6.1.2.- Conclusiones Experimentales

- i) Las pruebas cerradas concluyeron con modelos de entrenamiento que mostraron una eficiencia superior al 90 % en los cuatro volcanes. En cuanto a las pruebas abiertas, estas presentaron porcentajes de efectividad superiores al 88%, lo que indica la eficacia en el reconocimiento por parte del sistema.
- j) El sistema ha podido discriminar bien el ruido de fondo presente en los registros de los volcanes para que este no influya en los resultados.
- k) Se ha determinado que el sistema difiere un poco del observador, al marcar el inicio y el fin de cada evento en la clasificación. Sin embargo, esto no afecta el resultado final de las pruebas y los reconocimientos.
- l) Se ha podido comprobar en el San Cristóbal la presencia de eventos como el tremor armónico, estos eventos acompañan al tremor de fondo normal que presentan los volcanes y que precisan de un análisis de sus características y el papel que desempeñan en el ámbito sísmico. Esto debe de representar un trabajo mucho más detallado para futuras investigaciones.
- m) Se ha logrado consolidar el catálogo de los eventos que predominan en la actividad de los dos volcanes Telica y San Cristóbal como aporte al estudio sísmológico local de Nicaragua. Este catálogo no existía al realizar el presente estudio, lo que sumado al consenso que se debe de realizar entre los sísmólogos a cargo del estudio de los volcanes de Nicaragua, debe de llevar a que se establezca el tipo correcto, al cual pertenecen cada evento de acuerdo a la escuela que se juzgue conveniente seguir.
- n) En los entrenamientos y el reconocimiento de los modelos, las matrices de confusión de los eventos no describieron muchas sustituciones entre los eventos. Por otra parte, si se toma en cuenta que la mayoría de los eventos eliminados e insertados por el sistema, corresponden a ruidos tanto en Nicaragua como en Italia, esto indicaría que el porcentaje en precisión en todos los volcanes estaría cercano al 94 %.
- o) El sistema reconoce eventos que no han sido etiquetados por el observador, pudiendo reconocer más de un tipo en un registro o catalogar más de dos eventos del mismo tipo dentro de un mismo registro. En este sentido, se acepta como un acierto o éxito los eventos que el sistema ha insertado ya que corresponden a eventos que el observador no había catalogado y que sin embargo el sistema reconoce.
- p) Las pruebas de entrenamiento cruzadas permiten reafirmar la teoría, de que para entrenar adecuadamente un volcán se necesita una base de datos particular de él mismo. Por lo tanto, el entrenamiento de los modelos en cada volcán, precisa de una base de datos de entrenamiento propia.

- q) Se ha logrado desarrollar una implementación tecnológica a partir de los análisis de los eventos sísmico-volcánicos. Dicho software se podrá integrar con el sistema local de adquisición de datos en Nicaragua, brindando así una importante ayuda a los sismólogos locales en el reconocimiento de los eventos y sus posibles consecuencias de riesgo y prevención sísmico-volcánica.
- r) Esta tesis doctoral ha tenido como aportación la publicación de capítulos en dos libros, además de tres publicaciones en revistas de alto impacto y varias ponencias en diferentes congresos científicos a nivel internacional.
- s) Los resultados de esta tesis son un avance fundamental para la gestión y vigilancia de volcanes activos. El hecho de poder incorporar como herramienta de rutina un sistema automático y fiable de reconocimiento de señales sísmico-volcánicas, permite instaurar protocolos de seguimiento y alerta temprana de erupciones volcánicas, basados en la cuantificación y clasificación en tiempo real de la actividad sísmica. Este sistema complementa al trabajo de cualquier operador, lo que permite dedicar el esfuerzo a la evaluación de la actividad y no a su identificación.
- t) En la presente tesis se ha observado la no posibilidad de combinar las bases de datos, este hecho refleja la necesidad de que para tener un sistema robusto y fiable, exportable a cualquier tipo de sistema volcánico, sería necesaria la integración del mayor número posible de bases de datos de diferentes volcanes.



6.2.- TRABAJOS A FUTURO

Como parte de las mejoras o aspectos a desarrollar en futuras investigaciones se detallan las siguientes:

- a) Una línea de investigación futura pasaría por constituir una mayor base de datos, con una cantidad equilibrada en el número de eventos, por cada tipo encontrado en cada volcán. De esta forma, podrían obtenerse mejores resultados que los conseguidos durante el desarrollo del presente trabajo.
- b) Otra de las líneas de investigación pasaría por poder desarrollar campañas durante el desarrollo de crisis sísmicas y/o eruptivas en los volcanes, para analizar la presencia de los diferentes tipos de eventos que se determinaron en este trabajo, y descubrir su influencia en la actividad sísmica del volcán, así como determinar si podrían catalogarse de eventos premonitores de posibles futuras erupciones.
- c) Debido a la corta duración de las campañas empleadas en este estudio, el número de eventos a analizar ha sido limitado. Durante el desarrollo de este trabajo se detectaron algunos eventos que por su poca cantidad o por estar asociados a mucho ruido de fondo, fueron eliminados del proceso de análisis. Sin embargo, a futuro, sería interesante poder incorporar en los análisis la mayor combinación de situaciones posibles, al contar con una mayor cantidad de registros a analizar.
- d) Otra línea de investigación pasaría por obtener más datos referentes al temblor continuo del San Cristóbal (detectado durante el desarrollo de este trabajo). Así este particular tipo de evento podría incluirse dentro del proceso de estudio, al analizar a fondo sus características y su relación con futuros procesos eruptivos en el volcán.
- e) Otro aspecto a tratar a futuro, es poder contar con más estaciones instaladas alrededor de los volcanes (array sísmicos y estaciones de más de una componente), y poder de esta forma establecer con mucha mayor precisión, si los tipos de eventos aquí analizados, puedan pertenecer a grupos de familias o subfamilias sísmicas, y así poder hacer un análisis más detallado, con muchas más fuentes de información de las que se contaban en este estudio. Por lo tanto, podrían mejorarse los resultados de los análisis de los tipos de eventos y su relación con los períodos eruptivos.
- f) Otro aspecto básico en lo referente a Nicaragua, sería a futuro ampliar la base del conocimiento y la metodología utilizada en la tesis, para incluir los siete volcanes activos restantes de la cadena volcánica cuaternaria de Nicaragua, y que representen una amenaza o riesgo natural latente para los núcleos de poblaciones cercanas.
- g) Otro punto a futuro sería poder conocer la relación entre los procesos explosivos o eruptivos en el volcán y la actividad de la cadena volcánica. De acuerdo a ello se podría saber si algunos tipos de eventos de un volcán, determinan una actividad premonitoria en otro o no. Por lo tanto, determinar si hay eventos

precursores de similares características entre los volcanes de Italia y Nicaragua. Así como para determinar si la actividad en la cadena volcánica, influye en la activación o no de un volcán en particular.

- h) Otro aspecto a investigar en futuros trabajos para Nicaragua, referente a la geología de la cadena volcánica, es analizar y conocer las velocidades de propagación y los efectos de sitio en los volcanes de Nicaragua, y su relación con los diferentes tipos de eventos determinados en el presente estudio.
- i) En la aplicación del método utilizado en el presente trabajo (al realizar el análisis mediante bancos de filtros), futuras pruebas y estudios podrían incluir en los entrenamientos, diferentes tipos de análisis de predicción. De esta forma, se podrían comparar los resultados con los obtenidos en este documento.
- j) Otro aspecto a futuro sería poder consolidar y desarrollar más a fondo la aplicación de los scripts y programas, tanto de análisis y clasificación de los eventos, e incorporar mejoras en los algoritmos utilizados en la creación de los HMMs en el presente trabajo.
- k) En la presente tesis se realizaron las modificaciones de cuatro tipos de parámetros para mejorar los porcentajes de eficiencia en el reconocimiento de los modelos HMMs, a futuro queda incluir dentro de las pruebas, el establecer modificaciones de otros diversos parámetros y poder observar los resultados.
- l) En este trabajo se han establecido las bases de dar un tratamiento individual a los eventos ocurridos en un determinado volcán, al trabajar con un sistema de reconocimiento particular, capaz de generar modelos de entrenamiento adecuados a las particularidades geológicas y geofísicas de una determinada estructura volcánica. Una línea de investigación futura sería el poder integrar una mayor cantidad de bases de datos entrenadas de diferentes volcanes, para establecer una posible generalización y exportación del método aquí empleado a volcanes no entrenados.
- m) Futuras investigaciones podrían incluir campañas en Nicaragua durante la estación lluviosa, para comparar análisis y resultados de los tipos de eventos en este período con los realizados en este estudio, y observar la influencia del aumento de los acuíferos en la sismología volcánica.
- n) Futuros trabajos podrían mejorar el diseño y el alcance de la implementación tecnológica desarrollada en este estudio, a fin de que pueda abarcar y enriquecerse con diferentes análisis espectrales adicionales, así como el poder incluir en ella el reconocedor HMMs, determinado en el presente trabajo.
- o) Por último, el siguiente paso es integrar los resultados del presente trabajo, así como los programas obtenidos, con el sistema de adquisición de datos local en Nicaragua, brindando así una importante ayuda a los sismólogos locales, en las labores de vigilancia, ofreciendo una medición técnica al reconocimiento visual de los eventos y sus posibles consecuencias de riesgo en la prevención sísmico-volcánica.

6.3. - CONCLUSIONS AND SUCCESSES

With the results of this work, we have obtained the following methodological and experimental conclusions:

6.3.1 Methodological conclusions

- a) This thesis has proved that the automatic recognition process in volcanic seismic signals can be performed reliably and accurately, by using techniques based on HMM recognition models.
- b) From the application performed over our data set, we have demonstrated that in order to have a reliable result, a careful and adequate segmentation process is crucial.
- c) It has been possible to generate successful training models on two active volcanoes in Italy, using four types of seismic-volcanic events including the background noise. In the same way, we have generated effective training models in San Cristobal and Telica volcanoes of Nicaragua. To do this, we used 9 different seismic-volcanic events including the background noise.
- d) It determines the usefulness as a tool to establish the types of seismic volcanic events, spectral with the analysis based on the fast Fourier transform and spectrogram, since they give useful information in the first approximation.
- e) This thesis has explored other tools to complement the analysis of the characterization of the seismic-volcanic signals, based on power spectral density and wavelet transform.
- f) This thesis has determined that the configuration parameters, used in the different tests to improve the percentage of the models are different in each volcano, which determines that may not necessarily establish a single set of configurations for all volcanoes, need to train each database according to specific characteristics of the seismic-volcanic signals present in each volcano.
- g) It was possible to estimate the value of the spectral characterization of the seismic-volcanic signals in order to provide a reliable database. This has been particularly useful in unknown events, generated by the two active volcanoes in Nicaragua, and that have not yet been classified and studied thoroughly unlike the volcanoes of Italy, in which several authors in previous works have been cataloged and studied extensively the events produced by them.
- h) There has been an extensive analysis of how the spectral components of the signals which forming the seismic- volcanic events in the four study volcanoes, have unique characteristics, which differ significantly in all volcanoes. This establishes that although geological features in the volcanoes are very similar, and are located very close geographically, the events are different.

6.3.2 Experimental conclusions

- i) Closed testing concluded with training models that showed greater than 90% accuracy in the four volcanoes, and the open tests showed accuracy rates exceeding 88%, indicating the effectiveness of the recognition by the system.
- j) The system was able to discriminate very well the background noise present in the records of the volcanoes, so that this does not influence the results.
- k) It has been determined that the system differs slightly from the observer in the labeling, to mark the beginning and end of each event, However, this does not affect the final result of the tests.
- l) It has been found in the San Cristóbal the presence of harmonic tremor, these events accompany the normal background tremor that occurs in volcanoes, and require an analysis of its features and the role they play in the seismic ambit. This must be presented in much more detail in future research.
- m) It has consolidated the catalog of events in the activity of the two volcanoes: Telica and San Cristobal, as a contribution to local seismological study in Nicaragua. This catalog does not exist when performing this study, this added to the consensus that should be performed among seismologists in charge in Nicaragua, should lead to the establishment of the right type, it belongs each event according to the school at be established.
- n) In the training and recognition models, the confusion matrix of the events did not describe many substitutions between events. Moreover, the majority of events removed and inserted by the system, are noise both in Nicaragua and in Italy, this would indicate that the accuracy in all the volcanoes would be nearly 94%.
- o) The system acknowledges events, which are not have labeled by the observer, and can recognize more than one type of event in a record, or appoint more than two events of the same type within a record. In this sense, is designated the insertions like a success because correspond to events what the observer had not classified yet, but that have been recognized by the system.
- p) Cross-training tests allow reaffirm the theory, that to adequately train a volcano requires a particular own database. Therefore, the training models in each volcano requires an own training database.
- q) It has succeeded in developing a technology implementation from the analysis of seismic volcanic events. Such software can be integrated with the local data acquisition system in Nicaragua, providing substantial assistance to local seismologists, in recognition of the events and the possible consequences of risk and seismic-volcanic prevention.
- r) This thesis has been to contribute to publishing chapters in two books, and three publications in high impact journals and several dissertations in international scientific conferences.

- s) The results of this thesis are a major advance in the management and monitoring of active volcanoes. The fact to incorporate as tool routine an automatic and reliable system recognition of volcanic seismic signals, let establish protocols for monitoring and early warning of volcanic eruptions, based on the quantification and real-time classification of the seismic activity. This system complements the work of any operator allowing the effort devoted to the evaluation of the activity and not at its identification.

- t) In this thesis there has been no possibility of combining the databases, this reflects the need to have a robust and reliable system, exportable to any volcanic system, would require the integration of as many different volcanoes databases.



6.4. - FUTURE WORKS

Improvements or aspects to be developed in future research are the following:

- a) A future research would be of greater database, with a balanced amount in the number of events for each type found in each volcano. In this way, could be obtained better results than those achieved during the development of this work.
- b) Another future research would be able to develop campaigns for the development of seismic volcanic crisis, to analyze the presence of different types of events that were determined in this work, and discover its influence on seismic activity in the volcano, and to determine whether events could be considered as possible premonitory of future eruptions.
- c) Because the short duration of campaigns employed in this study, the number of events to be analyzed has been limited. During the course of this work some events were detected by their low amount or to be associated with high background noise were eliminated from the analysis process. However, in future it would be interesting to incorporate into the analysis the highest combination of possible situations, to have a greater number of records to analyze.
- d) Another future research would go through more data regarding the San Cristobal continuous tremor (detected during the development of this work). So this special type of event could be included in the study process to thoroughly analyze their characteristics and their relationship to future eruptive processes in the volcano.
- e) Another aspect to try in the future, is to have more stations installed around volcanoes (array seismic stations and more than one component), and thus to establish much more precisely, if the types of events analyzed here, can belong to groups of families or subfamilies seismic, so we can do a more detailed analysis, with many more sources of information that were counted in this study. Therefore, could be improved analysis results of event types and their relation to eruptive periods.
- f) In Nicaragua, would be looking to expand the knowledge base and the methodology used in the thesis, to include the seven remaining active volcanoes Nicaragua quaternary volcanic chain, and represent a threat or latent natural hazard the nuclei of nearby towns.
- g) Another point in future research would be able to know the relationship between explosive or eruptive processes in the volcano and the volcanic chain activity. According to it you could know if some types of events from a volcano, determine an activity to another or not prescient. Therefore, to determine whether precursor events of similar characteristics between the Italy and Nicaragua volcanoes. So as to determine whether the activity in the volcanic chain, influences the activation or not in a particular volcano.

- h) Another aspect to investigate in future work to Nicaragua concerning the geology of the volcanic chain, is to analyze and understand the speed of propagation and site effects in the volcanoes of Nicaragua, and its relation to the different types of events identified in this study
- i) The method used in this work (to perform the analysis using filter banks), future studies could include testing and in practice, different types of predictive analysis. Thus, the results could be compared with those obtained in this document.
- j) Another aspect would be to consolidate and further develop the application of scripts and programs, both analysis and classification of events, and incorporate improvements in the algorithms used in the creation of the HMMs in this work
- k) In this thesis modifications were performed four types of parameters to improve efficiency percentages recognition HMMs models, the future is include in testing, other modifications set various parameters and observe the results.
- l) In this work we have laid the groundwork to give individual treatment to the events of a particular volcano, when working with a particular recognition system, capable of generating adequate training models to geological and geophysical peculiarities of a particular volcanic structure. A line of future research would be to integrate a greater number of trained databases of different volcanoes, to establish and export a possible generalization of the method here employed untrained volcanoes.
- m) Future research could include campaigns in Nicaragua during the rainy season, to compare results and analysis of the types of events in this period with those obtained in this study, and to observe the influence of rising groundwater in volcanic seismology.
- n) Future work could improve the design and scope of technology implementation developed in this study, so it can cover and enriched with different additional spectral analysis and power by including the HMM recognizer, determined in this work.
- o) Finally, the next step is to integrate the results of this work and the programs obtained with the Nicaragua data acquisition system, providing significant assistance to local seismologists, in surveillance tasks, providing a measurement visual recognition of technical events and their potential impact on prevention risk seismic-volcanic.

BIBLIOGRAFÍA

Abril M. E. Ibáñez J.M., (2000) Uso de antenas sísmicas en ambientes volcánicos. En Curso internacional de volcanología y geofísica volcánica. Edición 2000. Editado por Mar Astiz y Alicia García. Nº 7 de la serie Casa de los Volcanes, Cabildo Insular de Lanzarote, pp. 283-295. ISBN 84-87021-74.3

Aguilar, R. C. (2007). Diseño y manipulación de modelos ocultos de markov, utilizando herramientas HTK. Una tutoría *design and manipulation of hidden markov models using htk tools. A tutorial*. Ingeniare. Revista chilena de ingeniería, 15(1), 18-26.

Ahmed, N., Natarajan, T., & Rao, K. R. (1974). Discrete cosine transform. Computers, IEEE Transactions on, 100(1), 90-93.

Aki, K., & Ferrazzini, V. (2000). Seismic monitoring and modeling of an active volcano for prediction. *Journal of Geophysical Research*, 105(B7), 16617-16.

Albrecht, J.L., Walker, J.A., Patino, L.C., and Carr, M.J., (2005). Geochemical history of the Telica volcanic complex, Nicaraguan volcanic front: implications of new data. The Geological Society of America (GSA) North-Central Section - 39th Annual Meeting (May 19–20, 2005) Minneapolis, Minnesota. U.S.A

Almendros, J., Ibáñez, J. M., Alguacil, G., & Del Pezzo, E. (1999). Array analysis using circular-wave-front geometry: An application to locate the nearby seismo-volcanic source. *Geophysical Journal International*, 136(1), 159-170.

Almendros, J., Ibáñez, J. M., Carmona, E., & Zandomenghi, D. (2007). Array analyses of volcanic earthquakes and tremor recorded at Las Cañadas caldera (Tenerife Island, Spain) during the 2004 seismic activation of Teide volcano. *Journal of volcanology and geothermal research*, 160(3), 285-299.

Aloisi, M., D'Agostino, M., Dean, K.G., Mostaccio, A., Neri, G., (2002). Satellite analysis and PUFF simulation of the eruptive cloud generated by the Mount Etna paroxysm of 22 July 1998. *J. Geophys. Res.* 107.

Alvarado, G.E., Barquero, R., Boschini, I., Chiesa, S. y Can, M., (1986). Relación entre la neotectónica y el vulcanismo en Costa Rica. -Bogotá, Simposio internacional sobre neotectónica y riesgos volcánicos, 19 p.

Álvarez-Gómez, J. A., Meijer, P. T., Martínez-Díaz, J. J., & Capote, R. (2008). Constraints from finite element modeling on the active tectonics of northern Central America and the Middle America Trench. *Tectonics*, 27(1).

Aristizábal A. G., Kumagai H., Nakano M., (2004). Source Process of Volcanic Tremor Inferred From Waveform Inversion. American Geophysical Union, Fall Meeting 2004, abstract #V14B-03. San Francisco. U.S.A. 13-17

Baker J., (1975) The Dragon System - An overview IEEE Transactions on Acoustics, Speech and Signal Processing, vol 23, no 1, pp. 24-29.

Barberi, F., Rosi, M., Sodi, A., (1993). Volcanic hazard assessment at Stromboli based on review of historical data. *Acta Vulcanol.* 3, 173–187

- Barras, C., Geoffrois, E., Wu, Z., & Liberman, M. (1998). Transcriber: a free tool for segmenting, labeling and transcribing speech. In Proceedings of the First International Conference on Language Resources and Evaluation (LREC).
- Barras, C., Geoffrois, E., Wu, Z., & Liberman, M. (2001). Transcriber: Development and use of a tool for assisting speech corpora production. *Speech. Communication*, 33(1–2), 5–22
- Bartosch,; Seidl Seidl, (1999) Spectrogram analysis of selected tremor signal using short-time Fourier transform and continuous wavelet transform. *Ann Geofis.* 42 (3): 497-506
- Behncke Boris and Tanguy Jean-Claude 1999 Summary of June-September activity; powerful eruption from the Voragine on 4 September 09/1999 (BGVN 24:09). Global Volcanism Program. Smithsonian Institution National Museum of Natural History.
- Benítez, M. C., Ramírez, J., Segura, J. C., Ibáñez, J. M., Almendros, J., García-Yeguas, A., & Cortés, G. (2007). Continuous HMM-based seismic-event classification at Deception Island, Antarctica. *Geoscience and Remote Sensing, IEEE Transactions on*, 45(1), 138-146.
- Bertagnini, A., Coltelli, M., Landi, P., Pompilio, M., Rosi, M., (1999). Violent explosions yield new insights into dynamic of Stromboli volcano. *EOS Trans. AGU* 80 (52), 633–636
- Bertucco, L., Coltelli, M., Nunnari, G., Occhipinti, L., (1999). Cellular neural networks for real-time monitoring of volcanic activity. *Computer and Geosciences* 25, 101– 117
- Bezdek J.C. (1981). *Pattern Recognition with fuzzy Objective Function Algorithms*. Plenum Press, New York.
- Bianca, M., Monaco, C., Tortorici, L., Cernobori, L., (1999). Quaternary normal faulting in southeastern Sicily (Italy): a seismic source for the 1693 large earthquake. *Geophys. J. Int.* 139, 370– 394.
- Bice, David C., (1985), Quaternary volcanic stratigraphy of Managua, Nicaragua: Correlation and source assignment for multiple overlapping plinian deposits. *Geological Society of American Bulletin*, 96: 553-566.
- Bilmes, J. (1998). A gentle tutorial on the EM algorithm and its application to parameter estimation for Gaussian mixture and Hidden Markov Models. Tech. Rep. ICSI-TR-97-021, University of California Berkeley
- Blong, R. J. (1984). *Volcanic Hazards: a source book on the effect of eruptions*. Academic Press. 424pp.
- Boashash, B., and Black, P. (1987). An efficient real-time implementation of the Wigner-Ville distribution. *Acoustics, Speech and Signal Processing, IEEE Transactions on*, 35(11), 1611-1618.
- Borrero, H., Baquero, Y., & Alezones, Z. (2011). Reconocimiento de Palabras Aisladas Utilizando LPC Y DTW, para control de navegación de un mini-robot. In Congreso Internacional de Ingeniería Mecatrónica-UNAB (Vol. 1, No. 1).
- Bracewell, R. N., & Bracewell, R. N. (1986). *The Fourier transform and its applications* (Vol. 31999). New York: McGraw-Hill
- Branca, S., and Del Carlo, P. (2005). Types of eruptions of Etna volcano AD 1670–2003: implications for short-term eruptive behaviour. *Bulletin of Volcanology*, 67(8), 732-742.

- Branca, S., Coltelli, M., De Beni, E., & Wijbrans, J. (2008). Geological evolution of Mount Etna volcano (Italy) from earliest products until the first central volcanism (between 500 and 100 ka ago) inferred from geochronological and stratigraphic data. *International Journal of Earth Sciences*, 97(1), 135-152.
- Bridle, John S. Alpha-Nets: (1990). A recurrent 'neural' network architecture with a hidden Markov model interpretation. *Speech Communication*, vol. 9, no 1, p. 83-92.
- Brinton DG (1887) On an ancient human footprint from Nicaragua. *Am Philos Soc Proc* 24:437–444
- Burke, K., J. F. Cooper, P. Dewey, P. Mann and J. L. Pindell, (1984). Caribbean tectonics and relative plate motions. *In: W. Boini & R. Hargraves (Eds.): Caribbean - South America Plate. Geol. Soc. Amer. Mem., 162, 31-63; New York.*
- Burrollet, P.F., Mugniot, G.M., Sweeney, P., (1978). The geology of the Pelagian Block: the margins and basins of Southern Tunisia and Tripolitania. *In: Nairn, A., Kanes, W., Stelhi, F.G. (Eds.), The ocean basins and margins. Plenum Press, New York, NY, pp. 331– 339.*
- Burrus, C. S., Gopinath, R. A., Guo, H., Odegard, J. E., & Selesnick, I. W. (1998). Introduction to wavelets and wavelet transforms: a primer (Vol. 23). Upper Saddle River: Prentice hall. 400pp.
- Burton, M.R., Mader, H.M., Polacci, M., (2007). The role of gas percolation in quiescent degassing of persistently active basaltic volcanoes. *Earth Planet. Sci. Lett.* Doi:10.1016/j.epsl.2007.08.028.
- Calvari, S., Coltelli, M., Neri, M., Pompilio, M., Scribano, V., (1994). The 1991–1993 Etna eruption: chronology and lava flow-field evolution. *Acta Vulcanol.* 4, 1– 14.
- Calvari, S., INGV-CT scientific staff, (2002). Update on the eruptive activity at Mt. Etna: multidisciplinary evidence of magmarefilling (abs). *Eos, Trans. Am. Geophys. Union* 83 (47), F1483 (Fall Meet. Suppl.).
- Calvari, S., Neri, M. Pinkerton, H., (2002). Effusion rate estimations during the 1999 summit eruption on Mount Etna, and growth of two distinct lava flow fields. *J. Volcano. Geotherm. Res.* 119, 107-123.
- Calvari, S., Neri, M., Pinkerton, H., (2003). Effusion rate estimations during the 1999 summit eruption on Mount Etna, and growth of two distinct lava flow fields. *Journal of Volcanology and Geothermal Research* 119 (1–4), 107– 123
- Calvari, S., Spampinato, L., Lodato, L., Harris, A. J., Patrick, M. R., Dehn, J., ... & Andronico, D. (2005). Chronology and complex volcanic processes during the 2002–2003 flank eruption at Stromboli volcano (Italy) reconstructed from direct observations and surveys with a handheld thermal camera. *Journal of Geophysical Research: Solid Earth* (1978–2012), 110(B2).
- Cameron I., (2001), Young volcanic Tephra stratigraphy near the Nejapa crater of Managua, Nicaragua. Internal Report, Department of Geoscience, University of Iowa, Iowa City.
- Cannata, A., Hellweg, M., Di Grazia, G., Ford, S., Alparone, S., Gresta, S., ... & Patanè, D. (2009). Long period and very long period events at Mt. Etna volcano: Characteristics, variability and causality, and implications for their sources. *Journal of Volcanology and Geothermal Research*, 187(3), 227-249.

- Capaldi, G., Guerra, I., Bascio, A. L., Luongo, G., Pece, R., Rapolla, A., ... & La Volpe, L. (1978). Stromboli and its 1975 eruption. *Bulletin Volcanologique*, 41(3), 259-285.
- Carniel Roberto 1998. Larger explosions in January, August, and September 1998. 10/1998 (BGVN 23:10). Global Volcanism Program. Smithsonian Institution National Museum of Natural History.
- Carr, M.J. & Rose, W., (1987). Cetam-A data base of central American Volcanic Rocks. – *Journal of Volcanol. Geother. Res*, 33, p. 239-240
- Casagli, N., Tibaldi, A., Merri, A., Del Ventisette, C., Apuani, T., Guerri, L., ... & Tarchi, D. (2009). Deformation of Stromboli Volcano (Italy) during the 2007 eruption revealed by radar interferometry, numerical modelling and structural geological field data. *Journal of volcanology and geothermal research*, 182(3), 182-200.
- Castellano, M., Capello, M., Del Pezzo, E., Giudicepietro, F., La Rocca, M., Martini, M., ... & Gresta, S. (2000). A double seismic array experiment on Mt. *Etna*. *Osservatorio Vesuviano Open file report*, (2-2000), 30.
- Chastin, S. F., & Main, I. G. (2003). Statistical analysis of daily seismic event rate as a precursor to volcanic eruptions. *Geophysical research letters*, 30(13).
- Chavez, O., Millan-Almaraz, J. R., Pérez-Enríquez, R., Arzate-Flores, J. A., Kotsarenko, A., Cruz-Abeyro, J. A., & Rojas, E. (2010). Detection of ULF geomagnetic signals associated with seismic events in Central Mexico using Discrete Wavelet Transform. *Nat. Hazards Earth Syst. Sci*, 10, 2557-2564.
- Chouet B., (1992). A Seismic Model for the Source of Long-Period Event and Harmonic Tremor. *IAVCEI Proceedings in Volcanology*, Vol. 3, Springer-Verlag, Berlin, 3-10
- Chouet, B. A., Page, R. A., Stephens, C. D., Lahr, J. C., & Power, J. A. (1994). Precursory swarms of long-period events at Redoubt Volcano (1989–1990), Alaska: Their origin and use as a forecasting tool. *Journal of volcanology and geothermal research*, 62(1), 95-135.
- Chouet, B., Saccorotti, G., Martini, M., Dawson, P., De Luca, G., Milana, G., Scarpa, R., (1997) Source and path effects in the wave fields of tremor and explosions at Stromboli Volcano, Italy. *Journ. Geoph-vs. Res*, 102, B7. 15 129-15150.
- Chouet, B.A., Saccorotti, G., Dawson, P., Martini, M., Scarpa, R., De Luca, G., Milana, G., Cattaneo, M. (1999). Broadband measurement of the sources of explosions at Stromboli volcano, Italy. *Geophys. Res. Lett.* 26, 1937-1940.
- Chouet B. (2003) *Volcano Seismology*. *Pure and Applied Geophysics*. 160(2003) 739-788
- Chouet, B., Dawson, P., Ohminato, T., Martini, M., Saccorotti, G., Giudicepietro, F., ... & Scarpa, R. (2003). Source mechanisms of explosions at Stromboli Volcano, Italy, determined from moment-tensor inversions of very-long-period data. *Journal of Geophysical Research: Solid Earth* (1978–2012), 108(B1), ESE-7.
- Coltelli, M., Pompilio, M., Del Carlo, P., Calvari, S., Pannucci, S., Scribano, V. (1998). Mt. Etna: the eruptive activity. *Acta Vulcanologica* 10, 141–148.
- Coltelli, M., Del Carlo, P., Pompilio, M. (2000). Mt. Etna: eruptive activity in 1996. *Acta Vulcanologica* 12 (1), 63– 67

- Connor, C.B., Williams, S.N., (1989). Interpretation of gravity anomalies, Masaya Caldera complex, Nicaragua. Preprint: Transaction of the 12th Caribbean Conference, St Croix, Virgin Islands.
- Cooley J.W. and Tukey J.W. (1965). An algorithm for the machine calculation of complex Fourier series. *Maths Comput* 19, 297.
- Corsaro R.A., Pompilio M. (2004), Magma dynamics in the shallow plumbing system of Mt. Etna as recorded by compositional variations in volcanics of recent summit activity (1995–1999). *Journal of Volcanology and Geothermal Research* 137 (2004) 55– 71
- Cortés, G., Arámbula, R., Gutiérrez, L. A., Benitez, C., Ibaez, J., Lesage, P., ... & García, L. (2009). Evaluating robustness of a HMM-based classification system of volcano-seismic events at Colima and Popocatepetl volcanoes. In *Geoscience and Remote Sensing Symposium, 2009 IEEE International, IGARSS 2009* (Vol. 2, pp. II-1012). IEEE.
- Costantini, L., Bonadonna C., Houghton B. F., and Wehrmann H. (2008). New physical characterization of the Fontana Lapilli basaltic Plinian eruption, Nicaragua, *Bull Volcanol* DOI 10.1007/s00445-008-0227-9
- Cristofolini, R., Corsaro, R.A., Ferlito, C. (1991). Variazioni petrochimiche nella successione etnea: un riesame in base a nuovi dati da campioni di superficie e da sondaggi. *Acta Vulcanol.* 1, 25–37.
- Cruden, A. R. (1989). The structure of south-western Nicaragua - a preliminary assessment. - Swedish Geological SGAB International AB, project: INMINE/SAREC, IDnr.: URAP 89001, - INETER-Archiv 28 pp; Managua.
- Dames & Moore-Lamsa (1978). Informe final del Estudio Geológico de las Ciudades del Sistema Metropolitano, Vice Ministerio de Planificación Urbana, Gobierno de la República de Nicaragua.
- Dampier W. (1927). A new voyage around the world (1697). Project Gutenberg Australia. eBook No.: 0500461.txt
Dirección Web: <http://gutenberg.net.au>
- Darce, M & Kondakov L.A.. (1993). Mapa Cosmogéolico de la República de Nicaragua, Esc 1:500,000 Managua, Nicaragua
- Darce M., (2002). Amenaza y Zonificación sísmica en Nicaragua. Instituto Nicaragüense de Energía INE, Managua, Nicaragua.
- Daubechies, I. (1988). Orthonormal bases of compactly supported wavelets. *Communications on pure and applied mathematics*, 41(7), 909-996.
- Daubechies, I. (1992). *Ten Lectures on Wavelets*. Society for Industrial and Applied Mathematics. Philadelphia. 377pp
- De Bruijn, L. M., Verheijen, E., Hasman, A., Van Nes, F. L., & Arends, J. W. (1995). Speech interfacing for diagnosis reporting systems: an overview. *Computer methods and programs in biomedicine*, 48(1), 151-156.
- Declaración de la Antigua (2003). Consulta Hemisférica sobre Alerta Temprana. Antigua Guatemala

- De Fino, M., La Volpe, L., Falsaperla, S., Frazzetta, G., Neri, G., Francalanci, L., ... & Sbrana, A. (1988). The Stromboli eruption of December 6, 1985–April 25, 1986: volcanological, petrological and seismological data. *Rend. Soc. Ital. Mineral. Petrol*, 43, 1021-1038.
- De Oliveira, H. M. (2007). *Análise de sinais para engenheiros*. Brasport.
- De Oviedo, G. F., & de los Ríos, A. (1851). *Historia general y natural de las Indias: Islas y tierra-firme del mar océano*. Real Academia de la Historia.
- Del Pezzo, E., Guerra, I., Lo Bascio, A., Luongo, G., Nappi, G., Scarpa, R. (1974) Microtremors and volcanic explosions at Stromboli. part 2. *Bull. Int. Geophys. Union*, 38, 1023-1036
- Del Pezzo, E., Esposito, A., Giudicepietro, F., Marinaro, M., Martini, M., & Scarpetta, S. (2003). Discrimination of earthquakes and underwater explosions using neural networks. *Bulletin of the Seismological Society of America*, 93(1), 215-223.
- Dempster A.P., Laird N.M., and Rubin D.B. (1977). Maximum-likelihood from incomplete data via the em algorithm. *J. Royal Statist. Soc. Ser. B.*, 39
- Dengo, G., Bohnenberger, O. and Bonis, S. (1970). Tectonics and volcanism along the Pacific Marginal Zone of Central America. *-Geol. Rundsch.*, 59: 1215-1232.
- Devijver PA, Kittler J. (1982) *Pattern recognition: a statistical approach*. Prentice-Hall, Englewood Cliffs
- Dhir, Nitin; Shkreli, Djovana; Foster, Thorp. (2003). *System and methods for call decisioning in a virtual call center integrating telephony with computers*. U.S. Patent No 6,553,113, 22
- Dietel, C., Chouet, B., Kleinman, J., De Luca, G., Martini, M., Milana, G., Power, J., Harlow, D., Scarpa, R. (1994). Arraytracking tremor sources at Stromboli volcano. Italy. (US Geol. Surv. Open-File rep. 94-142, 1-86.
- Duda R.O., and Hart P.E. (1973). *Pattern Classification and Scene Analysis*. J. Wiley and Sons, New York.
- Ellis, M., King, G., (1991). Structural control of flank volcanism in continental rifts. *Science* 254, 839–842.
- Espíndola, J. M., Zamora-Camacho, A., & Jiménez, Z. (2005). Earthquake families in the seismicity of Popocatepetl volcano. *Geofísica Internacional*, 44(2), 169-176.
- Falsaperla, S., Langer, H., Spampinato, S. (1998). Statistical analysis and characteristics of volcanic tremor on Stromboli volcano (Italy). *Bull. Volcanol.* 60, 75-88.
- Falsaperla, S., Langer, H. (2000). Long-term dynamics of eruptive activity at Stromboli volcano (Italy): Insights from volcanic tremor data. *EOS Trans. AGU Fall Meet. Suppl.* 81, V52F.
- Falsaperla S., Privitera E., Chouet B., Dawson P. (2002) Analysis of long-period events recorded at Mount Etna (Italy) in 1992, and their relationship to eruptive activity. *Journal of Volcanology and Geothermal Research* 114 (2002) 419-440
- Flint, E., (1884), Human footprints in Nicaragua: *American Antiquarian*, v. 6, p. 112-113. [Holocene; Central America, Nicaragua; hominid]
- Forney Jr, G. D. (1973). The viterbi algorithm. *Proc. IEEE*, vol. 61(3), pp.268-278.

- Fremont, M. J., & Malone, S. D. (1987). High precision relative locations of earthquakes at Mount St. Helens, Washington. *Journal of Geophysical Research*, 92(B10), 10223-10.
- Frischbutter A. (2002). Structure of the Managua graben, Nicaragua, from remote sensing images. *Geofísica Internacional*, Vol. 41, Num. 2, pp. 87-102
- Furui, S. (1986). Speaker-independent isolated word recognition using dynamic features of speech spectrum. *Acoustics, Speech and Signal Processing, IEEE Transactions on*, 34(1), 52-59.
- Garcés M. A. and Hansen R. A. (1997). Waveform analysis of seismoacoustic signals radiated during the Fall 1996 eruption of Pavlof volcano, Alaska, (Submitted to *Geophysical Research Letters* on June 23, 1997).
- Giesecke, A. (1989). Riesgo volcánico, evaluación y mitigación en América Latina: aspectos sociales, institucionales y científicos. -Lima, Centro Regional de sismología para América del Sur, 298 p.
- Gillot, P.Y., Kieffer, G., Romano, R. (1994). The evolution of Mount Etna in the light of potassium-argon dating. *Acta Vulcanol.* 5, 81– 87.
- Godoy Calderón, J. Fco. Martínez Trinidad, and Manuel Lazo Cortés, (2006). Proposal for a Unified Methodology for Evaluating Supervised and Non Supervised Classification Algorithms. *Progress in Pattern Recognition, Image Analysis and Applications Lecture Notes in Computer Science Volume 4225*, pp 674-685
- Guada Barráez, C., & Silva Trujillo, J. (2002). Data acquisition and processing system for local seismologic networks. *Revista Técnica de la Facultad de Ingeniería Universidad del Zulia*, 25(2), 59-67.
- Gutiérrez, L., Ibaez, J., Cortés, G., Ramirez, J., Benítez, C., Tenorio, V., & Isaac, A. (2009). Volcano-seismic signal detection and classification processing using hidden Markov models. Application to San Cristóbal volcano, Nicaragua. In *Geoscience and Remote Sensing Symposium, 2009 IEEE International, IGARSS* (Vol. 4, pp. IV-522). IEEE.
- Gutiérrez, Ligdamis, Ramírez, Javier, Ibañez, Jesús and Benítez, Carmen (2011). Volcano-Seismic Signal Detection and Classification Processing Using Hidden Markov Models - Application to San Cristóbal and Telica Volcanoes, Nicaragua, *Hidden Markov Models, Theory and Applications*, Dr. Przemyslaw Dymarski (Ed.), ISBN: 978-953-307-208-1, InTech, DOI: 10.5772/14312.
- Haddad, R. A., & Park, K. (1995). Modeling, analysis, and optimum design of quantized M-band filter banks. *Signal Processing, IEEE Transactions on*, 43(11), 2540-2549.
- Hamaguchi, H.; Hasegawa, (1975). A. Recurrent occurrence of the earthquakes with similar wave forms and its related problems. *J. Seismol. Soc. Japan*, vol. 28, p. 153-169.
- Harris, A., Neri, M. (2002). Volumetric observations during paroxysmal eruptions at Mount Etna: pressurized drainage of a shallow magma chamber or pulsed supply? *J. Volcanol. Geotherm. Res.* 116, 79– 95.
- Havskov, J., & Ottemoller, L. (1999). SEISAN earthquake analysis software. *Seismological research letters*, 70(5), 532-534.
- Havskov, J., & Alguacil, G. (Eds.). (2004). *Instrumentation in earthquake seismology* (Vol. 22). Springer.

- Hayes C. Willard The Nicaragua canal route. (1899). *Science*, Volume 10, Issue 239, pp. 97-104 doi: 10.1126/science.10.239.97
- Hazlett R.W. (1987). Geology of San Cristóbal volcanic complex, Nicaragua. *J. Volcanol Geotherm Res. Special Issue, R.E. Stoiber 75th Birthday* 33:223- 230
- Hirn, A., Nicolich, R., Gallart, J., Laigle, M., and ETNASEIS Scientific Group. (1997). Roots of Etna volcano in faults of great earthquakes. *Earth Planet. Sci. Lett.* 148, 171–191.
- Hradecky P., Havliceck P., Navarro M., Novak Z., Stanik E., Sebesta J. (1997). Estudio para el Reconocimiento de la Amenaza Geológica en el Área de Managua, Nicaragua. CGU/INETER. Praha-Managua.
- Hwang, J. N., Vlontzos, J. A., & Kung, S. Y. (1989). A systolic neural network architecture for hidden Markov models. *Acoustics, Speech and Signal Processing, IEEE Transactions on*, 37(12), 1967-1979.
- Ibáñez J.M., Carmona E. (1997). *Sismicidad Volcánica*. Instituto Andaluz de Geofísica, Universidad de Granada,
- Ibáñez, J. M., Morales, J., Alguacil, G., Almendros, J., Ortiz, R., & Del Pezzo, E. (1997). Intermediate - focus earthquakes under South Shetland Islands (Antarctica). *Geophysical research letters*, 24(5), 531-534.
- Ibáñez, J. M., Benítez, C., Gutiérrez, L. A., Cortés, G., García-Yeguas, A., & Alguacil, G. (2009). The classification of seismo-volcanic signals using Hidden Markov Models as applied to the Stromboli and Etna volcanoes. *Journal of Volcanology and Geothermal Research*, 187(3), 218-226.
- Incer J. (1993) Nicaragua; Viajes, Rutas, Encuentros 1502-1838, Colección V Centenario, 2da. Edición.
- Incer, Jaime (1998). *Geografía Dinámica de Nicaragua*. Managua: Editorial Hispamer.
- INETER, (1982). *Geología y Riesgos volcánicos de Nicaragua*. Informe Interno, INETER, Managua, Nicaragua, p55
- INETER, (2001). *Amenazas Naturales de Nicaragua*. INETER, 1ª Edición. Managua 310p
- INETER, (2004). *Evaluación de las Amenazas Geológicas e Hidrometeorológicas para Sitios de Urbanización Sep 2004 Proyecto "Estudios para detectar amenazas de fenómenos naturales en sitios previstos para la construcción de viviendas"*, Tipitapa, Informe N° 6.
- Jelinek, F. (1976). "Continuous Speech Recognition by Statistical Methods", in *Proc. of the IEEE*, vol. 64, pp. 532-556.
- Jordan M. and Jacobs R. (1994). Hierarchical mixtures of experts and the em algorithm. *Neural Computation*, 6:181–214.
- Kaiser, G. (1994). *A friendly guide to Wavelets*. Birkhäuser. Boston. 300 pp.
- Kaufman, Steven; Moser, James; Parente, Ronald N. (1992). *Integrated voice controlled report generating and communicating system*. U.S. Patent No 5,168,548.

- Khayam, S. A. (2003). The discrete cosine transform (DCT): theory and application. Michigan State University.
- Kingsbury, N. (2001). Complex wavelets for shift invariant analysis and filtering of signals. *Applied and computational harmonic analysis*, 10(3), 234-253.
- Konstantinou K. I., Lin C.H. (2004). Nonlinear Time Series Analysis of Volcanic Tremor Events Recorded at Sangay Volcano, Ecuador, *Pure appl. geophys.* 161 (2004) 145–163
- Kutterolf S., Freundt A., Pérez W., Wehrmann H., and Schmincke H.-U.. (2007) Late Pleistocene to Holocene temporal succession and magnitudes of highly-explosive volcanic eruptions in west-central Nicaragua. *Journal of Volcanology and Geothermal Research* Volume 163, Issues 1-4, 1, Pages 55-82
- Kutterolf, S., A. Freundt, W. Pérez, T. Mörz, U. Schacht, H. Wehrmann, and H.-U. Schmincke (2008), Pacific offshore record of plinian arc volcanism in Central America: 1. Along-arc correlations, *Geochem. Geophys. Geosyst.*, 9, Q02S01, doi:10.1029/2007GC001631.
- La Rocca Mario, Saccorotti Gilberto, Del Pezzo Edoardo, Ibáñez Jesús. (2004). Probabilistic source location of explosion quakes at Stromboli volcano estimated with double array data. *Journal of Volcanology and Geothermal Research* 131. 123-142
- Lamel, L., Rabiner, L., Rosenberg, A., & Wilpon, J. (1981). An improved endpoint detector for isolated word recognition. *Acoustics, Speech and Signal Processing, IEEE Transactions on*, 29(4), 777-785.
- Landi, P., Corsaro, R. A., Francalanci, L., Civetta, L., Miraglia, L., Pompilio, M., & Tesoro, R. (2009). Magma dynamics during the 2007 Stromboli eruption (Aeolian Islands, Italy): mineralogical, geochemical and isotopic data. *Journal of Volcanology and Geothermal Research*, 182(3), 255-268.
- Langer, H., Falsaperla, S. (1996). Long-term observation of volcanic tremor on Stromboli Volcano. *Pure Appl. Geophys.* 147, 57-82.
- Langer H., Falsaperla S. (2003), Seismic monitoring at Stromboli volcano (Italy); a case study for data reduction and parameter extraction. *Journal of Volcanology and Geothermal Research* 128 233-245
- Langer H., Falsaperla S., Powell T. and Thompson G. (2006). Automatic Classification and a-Posteriori Analysis of Seismic Events Identification at Soufrière Hills Volcano, Monserrat. *Journal of Volcanology and Geothermal Research*. Vo. 153. pp. 1-10
- Lautze, N. C., & Houghton, B. F. (2005). Physical mingling of magma and complex eruption dynamics in the shallow conduit at Stromboli volcano, Italy. *Geology*, 33(5), 425-428.
- Lee K.F. (1989). "Hidden Markov Models: Past, Present and Future". *Proc. of the European Conference on Speech Communication and Technology (Eurospeech)*, 1989, pp. 148-155.
- Lee, Chin-Hui; Soong, Frank K. & Paliwal, Kuldeep K. (Eds.). (1996). *Automatic speech and speaker recognition: advanced topics*. (Vol. 355). Springer.
- Leemann, W. P., M. J. Carr and J. D. Morris, (1994). Boron geochemistry of the Central American Volcanic Arc: Constraints on the genesis of subduction-related magmas. *Geochimica et Cosmochimica Acta*, 58, 149-168.

- Lermo, J., J. Cuenca, F. J. Chávez-García, A. Huidobro, M. L. Bermúdez, M. Rodríguez, R. Quaas, F. J. Sánchez-Sesma, E. Nava, H. Mijares y E. Briones. (1995). Características espectrales de temblores tipo-A, tipo-B y tremores asociados a la erupción del volcán Popocatepetl México en diciembre de 1994. En: Volcán Popocatepetl. Estudios realizados durante la crisis de 1994-1995. SNPC, Cenapred, UNAM, julio 1995, p. 139-186.
- Lilljequist. R., Andersson, L.C. & Astrand. (1987) Structural Interpretation of Landsat Images over Tertiary Volcanism in Nicaragua. Swedish Geological Survey-Inmine
- Lockley, M. G., Vasquez, R. G., Espinoza, E., & Lucas, S. G. (2009). America's most famous human footprints: History, context and first description of mid-Holocene tracks from the shores of Lake Managua, Nicaragua. *Ichnos*, 16(1-2), 55-69.
- Makhoul, J. (1975). Linear prediction: A tutorial review. *Proceedings of the IEEE*, 63(4), 561-580.
- Makris, J., Nicolich, R., Weigel, W., (1986). A seismic study in the western Ionian Sea. *Ann. Geophys.* 6, 665– 678.
- Manrique Ruiz L., Gutiérrez Franco E. (2007). Aplicación de redes neuronales para el pronóstico de sismos. Centro de Investigación en tecnologías avanzadas de decisión. Universidad de la Sabana, Colombia
- Martínez M & Viramente J. (1973): Estudio Geológico de la Cordillera de los Marrabios, Nicaragua.- Pub. Geol.. de ICAITI, Guatemala, C.A. 4: 139-148
- McBirney, A.R and Williams, H., 1965. Volcanic history of Nicaragua. *Univ. Calif. Publ. Geol. Sci.*, 55: 1—65
- McGuire, W.J., Pullen, A.D. (1989). Location and orientation of eruptive fissures and feeder-dykes at Mount Etna: influence of gravitational and regional tectonic stress regimes. *J. Volcanol. Geotherm. Res.* 38, 325–344.
- McGuire, W.J., Stewart, I.S., Saunders, S.J. (1997). Intra-volcanic rifting at Mount Etna in the context of regional tectonics. *Acta Vulcanol.* 9, 147– 156.
- McNutt, S. R. (1996). Seismic monitoring and eruption forecasting of volcanoes: a review of the state-of-the-art and case histories. In *Monitoring and mitigation of volcano hazards* (pp. 99-146). Springer Berlin Heidelberg.
- Melhem, H., & Kim, H. (2003). Damage detection in concrete by fourier and wavelet analyses. *Journal of Engineering Mechanics*, 129(5), 571-577.
- Minakami, T., (1960). Fundamental research for predicting volcanic eruptions (I). *Bull. Earthquake Res. Inst.*,38, 497-544
- Minakami, T. (1969). Earthquakes originating from Volcanoes. *Tai XVIII.Conv. Ass. Geof. tal*, 1-4, Napoli. Pág 1-38.
- Minakami, T. (1974) Seismology of volcanoes in Japan. In *Physical Volcanology*. Civetta L.; Gasparini P.; Luongo, G.; Rapolla, A. editores, Elsevier. Amsterdam: 1-27
- Mizukoshi, I., & Moriya, T. (1980). Broadband and wide dynamic range observation of Usu volcano earthquake swarm. *J. Seismol. Soc. Japan*, 33, 479-491.

- Monaco, C., Catalano, S., Cocina, O., De Guidi, G., Ferlito, C., Gresta, S., ... & Tortorici, L. (2005). Tectonic control on the eruptive dynamics at Mt. Etna Volcano (Sicily) during the 2001 and 2002–2003 eruptions. *Journal of volcanology and geothermal research*, 144(1), 211-233.
- Monaco, C., Tapponnier, P., Tortorici, L., Gillot, P.Y. (1997). Late Quaternary slip rates on the Acireale–Piedimonte normal faults and tectonic origin of Mt. Etna (Sicily). *Earth Planet. Sci. Lett.* 147, 125– 139.
- Mooser, F., Meyer-Abich, H., & McBirney, A. R. (1958). Catalogue of the active volcanoes of the world including solfatara fields: Part VI. Central America. International Volcanological Association, pp. 109-111
- Morales A. y Strauch, W. (1999): Sismicidad histórica de Managua, Informe para el proyecto de la Microzonificación Sísmica de Managua, INETER, Septiembre 1999
- Mulargia, F., Gasperini, P., & Tinti, S. (1987). Identifying different regimes in eruptive activity: an application to Etna volcano. *Journal of volcanology and geothermal research*, 34(1), 89-106.
- Nadeu, C., Hernando, J., Gorricho, M., (1995). On the decorrelation of filter-bank energies in speech recognition. In: Proc. Eurospeech, pp. 1381±1384
- Nadeu, C., Macho, D., & Hernando, J. (2001). Time and frequency filtering of filter-bank energies for robust HMM speech recognition. *Speech Communication*, 34(1), 93-114.
- Nawab, S. H., & Quatieri, T. F. (1988). Short-time Fourier transform. Advanced topics in signal processing, Lim J.S. and Oppenheim A.V., Eds. Englewood Cliffs, N.J. Prentice-Hall.
- Neri, M., Tomarchio, G., (2000). Mount Etna: eruptions welcoming the Third Millennium (Movie). RAI Radiotelevisione Italiana - Sistema Poseidon Production. Presented at: Geoexpo 31st Int. Geol. Cong. Rio de Janeiro, Brazil.
- Neria M., Acocellab V., Behnckea B., MaiolinoV., UrsinoA, Velarditaa R. (2005). Contrasting triggering mechanisms of the 2001 and 2002–2003 eruptions of Mount Etna (Italy). *Journal of Volcanology and Geothermal Research* 144 (2005) 235– 255
- Neuberg, J. W., Tuffen, H., Collier, L., Green, D., Powell, T., & Dingwell, D. (2006). The trigger mechanism of low-frequency earthquakes on Montserrat. *Journal of Volcanology and Geothermal Research*, 153(1), 37-50.
- Neuberg, J., Luckett, R., Ripepe, M., Braun, T., (1994). Highlights from a seismic broadband array on Stromboli Volcano. *Geophys. Res. Lett.* 21, 749-752.
- Newhall, C. G., & Self, S. (1982). The Volcanic Explosivity Index (VEI) an estimate of explosive magnitude for historical volcanism. *Journal of Geophysical Research: Oceans (1978–2012)*, 87(C2), 1231-1238.
- Niles, L. T., & Silverman, H. F. (1990). Combining hidden Markov model and neural network classifiers. In *Acoustics, Speech, and Signal Processing, 1990. ICASSP-90., 1990 International Conference on* (pp. 417-420). IEEE.
- Nishimura, Y., H. Y. Mori and H. Okada, (1992). Earth-quake Families Observed at Tarumi Volcano, Hokkaido, Japan, during January and February 1988, *J. Fac. Sci., Hokkaido University, Series VII*, 9, 2

- Ntepe, N., and Dorel, J. (1990). Observations of seismic volcanic signals at Stromboli volcano (Italy). *Journal of volcanology and geothermal research*, 43(1), 235-251.
- Okada, H., Watanabe, H., Yamashita, H., & Yokoyama, I. (1981). Seismological significance of the 1977–1978 eruptions and the magma intrusion process of Usu volcano, Hokkaido. *Journal of volcanology and geothermal research*, 9(4), 311-334.
- Oppenheim, A. V., Willsky, A. S., Nawab, S. H., & Fernández, A. S. (1998). *Señales y sistemas* (Vol. 2). México: Prentice Hall.
- Oppenheim, A. V., Schafer, R. W., & Buck, J. R. (1999). Discrete-time signal processing (Vol. 5). Upper Saddle River: Prentice Hall.
- Orozco, M., García, M. E., Duin, R. P., & Castellanos, C. G. (2006). Dissimilarity-based classification of seismic signals at Nevado del Ruiz Volcano. *Earth Sciences Research Journal*, 10(2), 57-66.
- Ortiz, R. (2000). Técnicas instrumentales para la vigilancia de volcanes activos. Serie Casa Volcanes, Cabildo Insular Lanzarote, 7, 227-242.
- Patrick, M.R., Harris, A.J.L., Ripepe, M., Dehn, J., Rothery, D., Calvari, S. (2007). Strombolian explosive styles and source conditions: insights from thermal (FLIR) video. *Bull. Volcanol.* 69, 769–784. doi:10.1007/s00445-006-0107-0.
- Peña, J., (2006), Espectrogramas basados en funciones wavelets, España, Dpto. Volcanología, Museo nacional de Cs. Naturales, Madrid.
- Pérez, N. M., & Hernández, P. A. (2008). La vigilancia volcánica en España: una apuesta crucial para la reducción del riesgo volcánico. El estudio y la gestión de los riesgos geológicos. Publicaciones del Instituto Geológico y Minero de España. Serie Medio Ambiente. Riesgos Geológicos Nº12. Madrid, 159-174.
- Pérez W., Freundt A., Schmincke H.-U., and Kutterolf S. (2006), Highly explosive eruptions from Masaya Caldera Complex (central Nicaragua) during the past 6000 years: Stratigraphy and hazard aspects. *Geophysical Research Abstracts*, Vol. 8, 07483, 2006 SRef-ID: 1607-7962/gra/EGU06-A-07483
- Pleva, M., Juhár, J., & Čižmár, A. (2005). Speech Detection in the Broadcast News Processing.
- Plinio el Joven, (2005), *Cartas*, Madrid: Editorial Gredos. ISBN 9788424927912
- Power J.A., Murray T.L., Marso J.N. and Laguerta P.E. 1991: Preliminary Observations of Seismicity at Mount Pinatubo by use of the Seismic Spectral Amplitude Measurement (SSAM) System, May 13- June 18, 1991; in *Fire and Mud*, pp. 269-283
- Qian, S. (2002). Introduction to time-frequency and wavelet transforms (Vol. 68). Prentice Hall PTR.
- Quaas R, Guevara E., González R., Ramos E., De la Cruz S., (2001). Monitoreo volcánico: Instrumentación y Métodos de Vigilancia,, en el libro "El volcán Popocatepetl Estudios Durante la crisis de 1994-1995", editores O. Zepeda y T. Sánchez, publicado por el Centro Nacional de Prevención de Desastres y la Universidad Nacional Autónoma de México.
- Rabiner, L., & Juang, B. (1986). An introduction to hidden Markov models. *ASSP Magazine, IEEE*, 3(1), 4-16.

- Rabiner, L.R. (1989). A tutorial on hidden Markov models and selected applications in speech recognition, Proc. of the IEEE, vol. 77, no. 2, Feb 1989.
- Rabiner L.R and. Juang B, (1993). Fundamentals of Speech Reconignition, Prentice – Hall.
- Redner R. and Walker H. (1984). Mixture densities, maximum likelihood and the em algorithm. *SIAM Review*, 26(2).
- Ren, Z. Justin; Zhou, Yong-Pin. (2008). Call center outsourcing: Coordinating staffing level and service quality. *Management Science*, vol. 54, no 2, p. 369-383.
- Restrepo B. J., Alvarez M., Henao R., (2007) Comparison between generative and discriminative approaches for classification of seismic events. *Scientia et Technica Año XIII, No 37. Universidad Tecnológica de Pereira. ISSN 0122-1701*
- Rioul, O., & Flandrin, P. (1992). Time-scale energy distributions: A general class extending wavelet transforms. *Signal Processing, IEEE Transactions on*, 40(7), 1746-1757.
- Rogers, J. A., & Stephens, C. D. (1995). SSAM: Real-Time seismic spectral amplitude measurement on a PC and its application to volcano monitoring. *Bulletin of the Seismological Society of America*, 85(2), 632-639.
- Rojas, W., H. Bungum and C. D. Lindholm, (1993). Historical and recent earthquakes in Central America. *Rev. Geol. Amer. Central*, 1993, 128-135.
- Rosi, M., Bertagnini, A., Landi, P., (2000). Onset of the persistent activity at Stromboli volcano (Italy). *Bulletin of Volcanology* 62, 294– 300.
- Rowe, C. A., Aster, R. C., Kyle, P. R., Dibble, R. R., & Schlue, J. W. (2000). Seismic and acoustic observations at Mount Erebus Volcano, Ross Island, Antarctica, 1994–1998. *Journal of volcanology and geothermal research*, 101(1), 105-128.
- Rubio, A. J. (2002). Procesado digital de señales. Ugr.
- Saccorotti, G., Almendros, J., Carmona, E., Ibáñez, J. M., & Del Pezzo, E. (2001). Slowness anomalies from two dense seismic arrays at Deception Island Volcano, Antarctica. *Bulletin of the Seismological Society of America*, 91(3), 561-571.
- Saccorotti Gilberto, Zuccarello Luciano, Del Pezzo Edoardo, Ibanez Jesus, Gresta Stefano, (2004). Quantitative analysis of the tremor wavefield at Etna Volcano, Italy. *Journal of Volcanology and Geothermal Research* 136 223– 245.
- Sapper, Karl. (1925), *Los Volcanes de la América Central: Halle (Saale) : Max Niemeyer*, VI p., 116 p.
- Scarpetta S., Giudicepetro F., Ezin E., Petrosino S., Pezzo E., Martín M. and Marinaro M. (2005). Automatic Classification of Seismic Signals at Mt. Vesuvius, Italy, using Neural Networks. *Bulletin of the Seismology Society of America*. Vol. 95(1). pp. 185-196
- Schmincke Hans-Ulrich, Kutterolf Steffen, Perez Wendy, Rausch Juanita, Freundt Armin, and Strauch Wilfried. (2008) Walking through volcanic mud: the 2,100-year-old Acahualinca footprints (Nicaragua). *Bull Volcanol DOI 10.1007/s00445-008-0235-9*
- Schmincke, H. U., Rausch, J., Kutterolf, S., & Freundt, A. (2010). Walking through volcanic mud: the 2,100 year-old Acahualinca footprints (Nicaragua) II: the Acahualinca people,

environmental conditions and motivation. *International Journal of Earth Sciences*, 99(1), 279-292.

Segura, F., Bungum, H., Lindholm, C., Hernández, Z. (1999). Estudio de Amenaza Sísmica de Managua, Nicaragua, NORSAR-INETER, Proyecto de la Microzonificación Sísmica de Managua, Septiembre de 1999

Seyfried, H. (1991). Depositional sequences and sequence boundary in fore-arc costal embayments. *In: McDonald, D.M. (Ed.) - Sedimentation, tectonics and eustasy; sealevel changes at active margins. Spec. Pub. Int. Ass. Sed., 12, 241-258; Oxford*

Siebert L, Simkin T (2002). *Volcanoes of the World: an Illustrated Catalog of Holocene Volcanoes and their Eruptions*. Smithsonian Institution, Global Volcanism Program Digital Information Series, GVP-3, (<http://www.volcano.si.edu/world/>).

Simkin, T. and Siebert, Lee. (1994). *Volcanoes of the world*. -Arizona, Geoscience Press, Inc., Smithsonian Institution, x + 3419p

SINAPRED 2005, Reporte sobre las Amenazas, Vulnerabilidad y Riesgos ante Inundaciones, Deslizamientos, Actividad Volcánica y Sismos. República de Nicaragua. Sistema Nacional para la Prevención, Mitigación y Atención de Desastres (SINAPRED). Secretaría ejecutiva.

Smithsonian Intitution (Bull), (1994); Global Volcanism Netwok-Summary of Recent Volcanic Activity, Bull Volcanol 56. 75-76

Sodré, Ulysses. (2002) Transformadas de Fourier.

Sparks, R. S. J. (2003). Forecasting volcanic eruptions. *Earth and Planetary Science Letters*, 210(1), 1-15.

Squier, E. G. (1973). *Nicaragua; its people, scenery, monuments, resources, condition, and proposed canal*. Book (ISBN 0404062202). AMS Press (New York)

Sten.-Åke. Elming. (1998). A palaeomagnetic study and K-Ar age determinations of tertiary rocks in Nicaragua. *Geoscientific research in Nicaragua : a Swedish-Nicaraguan joint project during the period 1981-1991*

Sten.-Åke. Elming, P. Layer, K. Ubieta (2001). A palaeomagnetic study and age determinations of Tertiary rocks in Nicaragua, Central America. *Geophysical Journal Internacional* Volume 147, Issue 2 , Pages294 – 309 The Royal Astronomical Society

Stevens, N.F., Murray, J.B., Wadge, G. (1997). The volume and shape of the 1991–1993 lava flow field at Mount Etna, Sicily. *Bull. Volcanol.* 58, 449–454.

Stoica, P., and Moses, R. L. (1997). *Introduction to spectral analysis (Vol. 89)*. New Jersey: Prentice hall.

Strang, G. (1999). The discrete cosine transform. *SIAM review*, 41(1), 135-147.

Strauch, W. (ed.) (2000) Estudio de la Microzonificación Sísmica de Managua, Informe INETER 2000

Tanguy, J.C., Kieffer, G., Patane, G., (1996). Dynamics, lava volume and effusion rate during the 1991–1993 eruption of Mount Etna. *J. Volcanol. Geotherm. Res.* 71, 259–265.

- Tárraga M., Abella R., Ortiz R, Astiz M, (2000). Seguimiento de la actividad volcánica: Estaciones multiparamétricas y Sistemas de Comunicaciones. Curso Internacional de Volcanología y Geofísica Volcánica (CIVGV) pp. 243-263
- Tenorio Virginia y Strauch Wilfried. (2000) Alerta temprana de erupciones volcánicas en nicaragua con información de la red sísmica. Geofísica/INETER.
- Tilling, R. (1989). Volcanic Hazards and their mitigation progress and problems. Rev Geophysics, 27: 237-269
- Torrence, C. and Compo, G.P (1998). A practical Guide to Wavelet Analysis. Bull. Amer Meteor. Soc. 79: 61-78
- Valdés Gonzáles, C. y G. J Gonzáles Pomposo. (1994). Sismicidad del volcán Popocatepetl noviembre 1993 a marzo 1994. Reporte Interno 94-2 Instituto de Geofísica UNAM, pp. 13 México D.F.
- Valrey, G., (1988). Desplazamiento de actividades volcánicas y su relación entre paleo - arcos antiguos y depósitos minerales en Nicaragua. -Quito, Memorias VI Curso Internacional de Metalogenia, Universidad Central del Ecuador, Facultad de Ingeniería en Geología, Minas y Petróleo, 241 -251.
- Verdejo, J. D., Luna, J. S., Herreros, A. P., Ayuso, A. R., & López, J. G. (1993). Entrenamiento discriminativo para HMM utilizando redes neuronales recurrentes. *Procesamiento de Lenguaje Natural*, vol 13.
- Viterbi A. D. (1967). "Error bounds for convolutional codes and an asymptotically optimum decoding algorithm," IEEE Trans. Inform. Theory, vol.IT-13, pp. 260-269.
- Walker, J.A., Williams, S.N., Kalamarides, R.I., Feigenson, M.D. (1993). Shallow open system evolution of basaltic magma beneath a subduction zone volcano: the Masaya Caldera Complex, Nicaragua. *J Volcanol Geotherm Res*, 56: 379-400.
- Wassermann, J. (2002). Volcano seismology. *IASPEI New Manual of Seismological Observatory Practice*, 1(13), 662-703.
- Williams, J. R., & Amaratunga, K. (1994). Introduction to wavelets in engineering. *International journal for numerical methods in engineering*, 37(14), 2365-2388.
- Yokoyama, I, Tilling,R.I & Scarpa R. (1984). International Mobile Early-Warning System (S) For Volcanic eruptions and Related Seismic Activities - UNESCO (Paris), EP/2106-8201 (2286), 102p
- Young, S. J., & Young, S. (1993). The HTK hidden Markov model toolkit: Design and philosophy. University of Cambridge, Department of Engineering.
- Young, S., Evermann, G., Kershaw, D., Moore, G., Odell, J., Ollason, D., ... & Woodland, P. (2002). The HTK book. Cambridge University Engineering Department, vol 3.
- Zibert J., Mihelic F., Martens J. P., Neto J., Meinedo H., Mateo C. G., Docio L., Zdansky J., David P., Pleva M., Ižmar A. Č, Zgank A., Kacic Z., Teleki C., Vicsi K. (2005). The COST278 Broadcast News Segmentation and Speaker Clustering Evaluation - Overview, Methodology, Systems, Results, Proc. Interspeech 2005 (Eurospeech) – 9th European Conference on Speech Communication and Technology



LISTADO DE FIGURAS

CAPÍTULO 1

- Fig. 1.1. Muestreo y reconstrucción de una señal continua en el tiempo
- Fig. 1.2. Elementos que intervienen en el procesamiento de la señal.
- Fig. 1.3. Esquema general de un Sistema de reconocimiento de patrones.
- Fig. 1.4. Esquema de un volcán y sus principales productos
- Fig. 1.5 Placas litosféricas y la relación con la distribución de volcanes y terremotos. Fuente: Kramer S., Geotechnical Earthquake Engineering., Ed. Prentice Hall., 1996
- Fig. 1.6 Eventos eruptivos que caracterizan la amenaza volcánica
- Fig. 1.7 Evento Volcano-Tectónico (VT) en el Monte Merapi, Indonesia. Fuente: [Wassermann 2002]
- Fig. 1.8 Evento Híbrido en el Monte Merapi, Indonesia. Fuente: [Wassermann 2002]
- Fig. 1.9 Evento LP en el Monte Merapi, Indonesia. Fuente: [Wassermann 2002]
- Fig. 1.10 Evento Tremor en el Monte Merapi, Indonesia. Fuente: [Wassermann 2002]
- Fig. 1.11 Evento Explosión en el Strómboli, Italia. Fuente: [Ibáñez et. al. 1997]

CAPÍTULO 2

- Fig. 2.1 Gráfica de la imagen de la Transformada de Fourier de una señal. En el recuadro se observa la señal en el tiempo
- Fig. 2.2 Gráfica del Espectrograma de una señal
- Fig. 2.3 Banco de filtros triangulares. Fuente: [Young et. al. (2002)]
- Fig. 2.4 Representación paramétrica de la señal.
- Fig. 2.5 HMM de un Autómata de Estados finitos con 4 estados.
- Fig. 2.6 Funcionamiento del algoritmo de Viterbi para el reconocimiento de señales. Donde la métrica de cada rama es "a" y la métrica de cada estado es b. Figura modificada del original tomada de [Young et. al. (2002)].
- Fig. 2.7 Representación de una mezcla de componentes Gaussianas. Figura modificada del original tomada de [Young et. al. (2002)]
- Fig. 2.8 .Arquitectura de un sistema de monitorización sísmico-volcánico basado en modelos HMMs.
- Fig. 2.9 Principales herramientas de HTK de acuerdo a las etapas del sistema. Figura modificada del original tomada de [Young et. al. (2002)]
- Fig. 2.10 Esquema de la parametrización de datos con HTK
- Fig. 2.11 Ejemplo de la segmentación y el etiquetado manual de una señal con la herramienta Transcriber
- Fig. 2.12 Ejemplo de segmentación de una señal con la herramienta *HSLab* de HTK
- Fig. 2.13 Ejemplo del contenido de un fichero de etiquetado (.lab) mediante la herramienta Transcriber
- Fig. 2.14 Elementos en un proceso de entrenamiento mediante HTK
- Fig. 2.15 Elementos del proceso de entrenamiento de los HMM con HTK
- Fig. 2.16 Elementos del proceso de reconocimiento de los HMM con HTK

CAPÍTULO 3

- Fig. 3.1 Volcán Etna en Italia. Fuente: NASA Astronaut photograph ISS006-E-31042
- Fig. 3.2 Mapa de las actividades eruptivas en el volcán Etna en los últimos años. Los círculos amarillos localizan los tres cráteres con mayor actividad. (Fuente Google Earth y McGuire et. al. 1997)
- Fig. 3.3 Mapa de la isla y volcán Strómboli en Italia. (Fuente Google Earth)
- Fig. 3.4 Volcán San Cristóbal (izq.) y mapa geográfico de su ubicación (der.). (Fuente Ligdamis Gutiérrez e INETER)

- Fig. 3.5 Mapa de amenazas asociadas al volcán San Cristóbal (Fuente: INETER Mapa de Amenazas volcánicas 1995. Escala 1'400.400)
- Fig. 3.6 Mapa de poblaciones afectadas por los productos del volcán San Cristóbal (Fuente INETER)
- Fig. 3.7. Volcán Telica (izquierda), y mapa geográfico del volcán Telica (derecha). (Fuente Google Earth e INETER)
- Fig. 3.8 Esquema general de la instalación y el equipo de la estación sísmica de tipo B. Figura modificada del original tomada de [Quaas et. al 2001]
- Fig. 3.9 Componentes de una estación sísmica de campo de tipo B, conectados a la estación central. Figura modificada del original, tomado de la página Web del Instituto Geofísico de la Escuela Politécnica Nacional de Ecuador (IG EPN): <http://www.igepn.edu.ec/index.php/nuestro-blog/item/114-estaci%C3%B3n-s%C3%ADsmica.html>
- Fig. 3.10 Sitio de ubicación de las estaciones en la parte baja del volcán; Pizzi Deneri y Torre del Filósofo. (Fuente Google Earth y mapa por Ligdamis Gtz.)
- Fig. 3.11 Equipos de las estaciones en el Etna. Fuente campaña Etna
- Fig. 3.12 Sismograma típico de los registros en el Etna. Fuente: INGV
- Fig. 3.13. Array sísmico Torre del Filósofo utilizado en el volcán Etna para la recolección de datos en Septiembre 1999. Fuente Gráfica Modificada tomada de la Campaña Etna
- Fig.3.14. Array sísmico “d” en el sitio “Torre del Filósofo” del volcán Etna. Fuente campaña Etna. Los puntos rojos indican la ubicación de los sismómetros de una componente, el triángulo azul indica la posición de la estación de banda ancha de tres componentes. Fuente Gráfica Modificada tomada de la Campaña Etna.
- Fig. 3.15 Volcán Strómboli, las estaciones al igual que en el Etna se colocaron en la parte baja del volcán. (Fuente Google Earth y Ligdamis Gtz.)
- Fig.3.16. Sitios de ubicación de los Arrays sísmicos Labronzo y Ginostra en el Strómboli para la recolección de datos en 1997. (Fuente Google Earth y Ligdamis Gtz.)
- Fig. 3.17 Estación en el volcán Strómboli. Fuente Campaña Strómboli
- Fig. 3.18 Sismograma típico de los registros en el Strómboli. Fuente: INGV.
- Fig.3.19. Array sísmico Ginostra en el Strómboli. Elección de las estaciones G. Fuente Gráfica modificada tomada de la Campaña Strómboli
- Fig. 3.20 Mapa de la ubicación de los dos complejos volcánicos: San Cristóbal y Telica y sus representativos volcanes (círculos amarillos) en la cadena volcánica de Nicaragua. (Fuente INETER y Ligdamis Gtz.)
- Fig. 3.21 Sismograma típico de los registros en el San Cristóbal. Se observa la presencia de un fuerte tremor de fondo en los registros del San Cristóbal Fuente: INETER
- Fig. 3.22 Mapa de la ubicación de las estaciones en el volcán San Cristóbal. Se observa la estación **CRIN** (círculo amarillo) elegida. (Fuente Google Earth e INETER).
- Fig. 3.23 (Izquierda), Autor en la caseta sísmica CRIN, se observa la cumbre del volcán al fondo. En la imagen de la derecha la flecha señala la ubicación de la caseta respecto al cráter en la cima del volcán. Fuente Ligdamis Gtz.
- Fig. 3.24 A la izquierda Sismógrafos verticales MARK serie Cs5016, provistos de sensores L4C-1 Hz, a la derecha paneles solares que alimentan las baterías, GPS para sincronizar el tiempo de registro y antena retransmisora en las casetas sísmicas. Fuente Ligdamis Gtz.
- Fig. 3.25 Interior de una caseta sísmica, en donde se pueden observar las baterías que alimentan los equipos. Abajo a la derecha se observa la caja de color negro del sistema SEISCOM que procesa y recopila los datos, el radio transmisor y los conversores digitales. Fuente Ligdamis Gtz.
- Fig. 3.26 Parte principal del sistema de adquisición y procesamiento y transmisión de datos de las estaciones sísmicas “*Seismological Communication Processor (SeisComP)*” por sus siglas en Inglés (foto fuente GeoForschungsZentrum (Potsdam): SeisComP Software y Hardware Web <http://geofon.gfz-potsdam.de/geofon//seiscomp/pc104.html>).
- Fig. 3.27 Vista de la caseta sísmica y el interior con los diferentes instrumentos sísmicos. También se observa la torre de comunicación en donde se coloca la antena TRANZEO. Fuente Ligdamis Gtz.
- Fig. 3.28 Autor en la Central Sísmica de recepción de datos ubicada en el INETER, Managua. Fuente Ligdamis Gtz.
- Fig. 3.29 Sismograma típico de los registros en el Telica. A diferencia del San Cristóbal, el Telica presenta un tremor de fondo mucho menor. Fuente: INETER

- Fig. 3.30 Mapa de la ubicación de las estaciones en el volcán Telica de 1999 al 2006. En el círculo amarillo la estación TELN elegida. Fuente INETER
- Fig. 3.31 Autor en Caseta tipo B, similar a la estación TELN, ubicada cerca del cráter del Telica. Fuente Ligdamis Gtz.
- Fig. 3.32 Mapa de la ubicación (círculo amarillo) de la estación TELN en el volcán Telica. (Fuente Google Earth).
- Fig. 3.33 Estación sísmica TELN en el cráter del Telica (Fuente INETER).

CAPÍTULO 4

- Fig. 4.1 Tipo de evento TB en el volcán Etna.
- Fig. 4.2 Características del tipo de evento TB en el volcán Etna.
- Fig. 4.3 Tipo de evento Exp en el volcán Strómboli.
- Fig. 4.4 Características del tipo de evento Exp en el volcán Strómboli.
- Fig. 4.5 Tipo de evento S1 en el volcán San Cristóbal.
- Fig. 4.6 Características del tipo de evento S1 en el volcán San Cristóbal
- Fig. 4.7 Tipo de evento S2 en el volcán San Cristóbal.
- Fig. 4.8 Características del tipo de evento S2 en el volcán San Cristóbal.
- Fig. 4.9 Tipo de evento S3 en el volcán San Cristóbal.
- Fig. 4.10 Características del tipo de evento S3 en el volcán San Cristóbal.
- Fig. 4.11 Tipo de evento “Tremor Continuo o armónico” en el volcán San Cristóbal
- Fig. 4.12 Características del evento “Tremor Continuo o armónico” en el volcán San Cristóbal
- Fig. 4.13 Tipo de evento T1 en el volcán Telica.
- Fig. 4.14 Características del tipo de evento T1 en el volcán Telica
- Fig. 4.15 Tipo de evento T2 en el volcán Telica
- Fig. 4.16 Características del tipo de evento T2 en el volcán Telica.
- Fig. 4.17 Tipo de evento T3 en el volcán Telica.
- Fig. 4.18 Características del tipo de evento T3 en el volcán Telica.
- Fig. 4.19 Tipo de evento T4 en el volcán Telica.
- Fig. 4.20 Características del tipo de evento T4 en el volcán Telica.
- Fig. 4.21 Señal en el Registro 2541132d del Etna (a) En el dominio del tiempo, (b) Envoltiente de la señal, (c) espectro de frecuencia de la señal (d) zoom del espectro de frecuencia de la señal
- Fig. 4.22 Señal en el Registro 2542227G del Strómboli (a) En el dominio del tiempo, (b) Envoltiente de la señal, (c) espectro de frecuencia de la señal (d) zoom del espectro de frecuencia de la señal.
- Fig. 4.23 Evento S1, en el Registro del San Cristóbal 2006-02-20-1827-54SC (a) En el dominio del tiempo, (b) Envoltiente de la señal, (c) espectro de frecuencia de la señal (d) zoom del espectro de frecuencia de la señal.
- Fig. 4.24 Evento S2, en el Registro 2006-02-18-2018-16SC del San Cristóbal (a) En el dominio del tiempo, (b) Envoltiente de la señal, (c) espectro de frecuencia de la señal (d) zoom del espectro de frecuencia de la señal.
- Fig. 4.25 Evento S3, en el Registro 2006-02-18-2253-14SC del San Cristóbal (a) En el dominio del tiempo, (b) Envoltiente de la señal, (c) espectro de frecuencia de la señal (d) zoom del espectro de frecuencia de la señal.
- Fig. 4.26 Evento T1, en el Registro 2006-02-17-0048-36ST del Telica (a) En el dominio del tiempo, (b) Envoltiente de la señal, (c) espectro de frecuencia de la señal (d) zoom del espectro de frecuencia de la señal.
- Fig. 4.27 Evento T2, en el Registro 2006-02-18-1048-20ST del Telica (a) En el dominio del tiempo, (b) Envoltiente de la señal, (c) espectro de frecuencia de la señal (d) zoom del espectro de frecuencia de la señal.
- Fig. 4.28 Evento T3, en el Registro 2006-03-07-1940-04ST (a) En el dominio del tiempo, (b) Envoltiente de la señal, (c) espectro de frecuencia de la señal (d) zoom del espectro de frecuencia de la señal.
- Fig. 4.29 Evento T4, en el Registro 2006-02-17-0633-34ST (a) En el dominio del tiempo, (b) Envoltiente de la señal, (c) espectro de frecuencia de la señal (d) zoom del espectro de frecuencia de la señal.

- Fig. 4.30 Espectrograma correspondiente a tres registros con eventos de tipo tremor burst con una forma característica entre sí en el Etna. Las flechas indican la presencia de los eventos en el tremor de fondo.
- Fig. 4.31 Espectrograma correspondiente a registros con eventos de tipo explosión con una forma característica entre sí en el Strómboli. Las flechas indican la presencia de bandas de ruido, aproximadamente sobre los 7 y 10 Hz.
- Fig. 4.32 Primer grupo de espectrogramas, que corresponden a tres eventos S1 con una forma característica entre sí en el San Cristóbal. Las flechas indican la presencia de múltiples llegadas en frecuencia en este tipo de evento. La curva indica una caída suave en frecuencia característica de este tipo de evento.
- Fig. 4.33 Segundo grupo de espectrogramas, con tres registros con eventos de tipo S2, que poseen una misma forma característica, en el San Cristóbal. La flecha superior indica una sola llegada en frecuencia a diferencia del anterior evento. La flecha inferior indica la forma de la caída en frecuencia de este tipo de evento.
- Fig. 4.34 Tercer grupo de espectrogramas, que corresponden a tres registros con eventos de tipo S3, que poseen una misma forma característica a lo largo del tiempo, en el San Cristóbal. La flecha indica una sola llegada similar al evento anterior. Sin embargo, la curva indica que la forma de caída en frecuencia en este tipo de evento es más similar al primer tipo, el S1 en comparación al S2
- Fig. 4.35 Espectrograma del tremor continuo del San Cristóbal. Las flechas pequeñas a la izquierda indican las bandas de frecuencia presentes en el tremor. Se observan dos bandas de ruido a 10 Hz y a 20 Hz., inducido por los equipos y la antena receptora.
- Fig. 4.36 Primer grupo de espectrogramas, que corresponden a registros con tres eventos T1, que poseen una misma forma característica a lo largo del tiempo, en el Telica. Las flechas negras inclinadas indican las dos llegadas en frecuencia, por lo que este tipo de evento es catalogado como evento doble. La flecha vertical indica la caída en frecuencia en un intervalo de duración corto a lo largo del tiempo en este tipo de evento. La flecha roja indica la presencia de una banda de ruido a 8 Hz.
- Fig. 4.37 Segundo grupo de espectrogramas, que corresponden a registros con tres eventos T2, que poseen una misma forma característica a lo largo del tiempo, en el Telica. La flecha negra inclinada muestra una única llegada en frecuencia a diferencia del evento anterior. La curva indica la caída suave en frecuencia diferente al evento anterior. Las flechas rojas indica bandas de ruido en torno a los 8 y 3 Hz.
- Fig. 4.38 Tercer grupo de espectrogramas, que corresponden a tres registros con eventos de tipo T3, que poseen una misma forma característica a lo largo del tiempo, en el Telica. La flecha negra inclinada indica una única llegada en frecuencia para este tipo de evento. La línea ondulada indica la forma de la caída en frecuencia en este tipo de evento. La flecha roja indica la presencia de una banda de ruido a los 8 Hz.
- Fig. 4.39 Cuarto grupo de espectrogramas, que corresponden a tres registros con eventos de tipo T4, que poseen una misma forma característica a lo largo del tiempo, en el Telica. Las flechas inclinadas indican la presencia de más de una llegada en frecuencia o múltiples picos en frecuencia para este tipo de evento. La curva indica la caída suave en frecuencia y las flechas rojas la presencia de bandas de ruido
- Fig. 4.40 Histograma de duración del evento “TB” (izquierda) y el ruido “NE” (derecha) en el Etna.
- Fig. 4.41 Histograma de duración del evento “Exp” (izquierda) y el ruido “Sil” (derecha) en el Strómboli.
- Fig. 4.42 Histograma de duración de los eventos en el volcán San Cristóbal. En orden de izquierda a derecha y arriba hacia abajo se encuentran los eventos S1, S2, S3 y NS.
- Fig. 4.43 Histograma de duración de los eventos en el volcán Telica. En orden de izquierda a derecha y arriba hacia abajo se encuentran los eventos T1, T2, T3 y T4.
- Fig. 4.44 Histograma de duración del ruido o tremor de fondo (NT) en el volcán Telica.
- Fig. 4.45 Ejemplo del proceso de etiquetado manual de tres eventos TB en el Etna
- Fig. 4.46 Ejemplo del proceso de etiquetado manual de un evento Exp en el Strómboli
- Fig. 4.47 Ejemplo del proceso de etiquetado manual de dos eventos (S2 y S3), en el San Cristóbal.
- Fig. 4.48 Ejemplo del proceso de etiquetado manual de un evento T1 en el Telica

CAPÍTULO 5

- Fig. 5.1 Resultados obtenidos en %Corr en las pruebas cerradas del HTK, en el porcentaje de aciertos al modificar el parámetro del número de gaussianas en el Etna
- Fig. 5.2 Resultados obtenidos en % Acc en las pruebas cerradas del HTK, en el porcentaje de precisión al modificar el parámetro del número de gaussianas en el Etna
- Fig. 5.3 Ejemplo del Análisis manual en el Proceso selectivo (evento por evento), de los resultados obtenidos en el Etna. Los valores en rojo representan los registros que han tenido un porcentaje de 100 y que no se cambiarán. Los valores en verde son aquellos que han tenido un porcentaje menor a 60 y que se analizarán para eliminarse de la base de datos.
- Fig. 5.4 Registros: 2541226d (arriba), 2550105d (centro) y 2550232d (abajo), en el dominio del tiempo.
- Fig. 5.5 Espectrogramas de los registros: 2541226d (arriba), 2550105d (centro) y 2550232d (abajo).
- Fig. 5.6 Ejemplo de seis registros, en el reconocimiento de eventos tipo TB, por parte del sistema durante las pruebas o test ciego-ciego. En los primeros dos registros el sistema reconoce un evento, en los siguientes dos, el sistema reconoce dos eventos y en los últimos dos, el sistema reconoce tres eventos de tipo TB.
- Fig. 5.7 Ejemplo de dos registros en el que el sistema ha reconocido un evento de tipo TB, en el test ciego-ciego.
- Fig. 5.8 Ejemplo de dos registros en el que el sistema ha reconocido dos eventos de tipo TB, en el test ciego-ciego.
- Fig. 5.9 Ejemplo de dos registros en el que el sistema ha reconocido tres eventos de tipo TB, en el test ciego-ciego.
- Fig. 5.10 Ejemplo de dos registros del fichero de resultado, en el reconocimiento de eventos tipo NS, por parte del sistema en el test ciego-ciego.
- Fig. 5.11 Ejemplo de registros en el que el sistema ha catalogado los registros como de tipo NE, en el test ciego-ciego.
- Fig. 5.12 Resultados obtenidos en %Corr en las pruebas cerradas del HTK, en el porcentaje de aciertos al modificar el parámetro del número de gaussianas en el Strómboli.
- Fig. 5.13 Resultados obtenidos en % Acc en las pruebas cerradas del HTK, en el porcentaje de precisión al modificar el parámetro del número de gaussianas en el Strómboli.
- Fig. 5.14 Ejemplo del Análisis manual en el Proceso selectivo (evento por evento), de los resultados obtenidos en el Strómboli. Los valores en rojo representan los registros que han tenido un porcentaje de 100 y que no se cambiarán. Los valores en verde son aquellos que han tenido un porcentaje menor a 60 y que se analizarán para eliminarse de la base de datos.
- Fig. 5.15 Registros 2542327G (arriba), 2561948G (centro) y 2571518G (abajo), en el dominio del tiempo.
- Fig. 5.16 Espectrogramas de los registros: 2542327G (arriba), 2561948G (centro) y 2571518G (abajo).
- Fig. 5.17 Ejemplo de seis registros, en el reconocimiento de eventos tipo Exp, por parte del sistema en el Strómboli, durante las pruebas o test ciego-ciego
- Fig. 5.18 Ejemplo de dos registros en el que el sistema ha reconocido un evento de tipo Exp, en el test ciego-ciego. Los círculos y las flechas indican los eventos.
- Fig. 5.19 Ejemplo de dos registros en el test ciego-ciego, en el que el sistema ha reconocido dos eventos de tipo Exp. Los círculos y las flechas indican los eventos.
- Fig. 5.20 Ejemplo de dos registros en el que el sistema ha reconocido tres eventos de tipo Exp, en el test ciego-ciego. Los círculos y las flechas indican los eventos.
- Fig. 5.21 Ejemplo de dos registros del fichero de resultado, en el reconocimiento de eventos tipo Sil, por parte del sistema en el test ciego-ciego
- Fig. 5.22 Ejemplo de registros en el que el sistema ha catalogado los registros como de tipo Sil, en el test ciego-ciego.
- Fig. 5.23 Resultados obtenidos en %Corr en las pruebas cerradas del HTK, en el porcentaje de aciertos al modificar el parámetro del número de gaussianas en el San Cristóbal
- Fig. 5.24 Resultados obtenidos en %Acc en las pruebas cerradas del HTK, en el porcentaje de aciertos al modificar el parámetro del número de gaussianas en el San Cristóbal
- Fig. 5.25 Ejemplo del Análisis manual en el Proceso selectivo (evento por evento), de los resultados obtenidos en el San Cristóbal. Los valores en rojo representan los registros que han

- tenido un porcentaje de 100 y que no se cambiarán. Los valores en verde son aquellos que han tenido un porcentaje menor a 60 y que se analizarán para eliminarse de la base de datos
- Fig. 5.26 Registros 2006-03-03-0650-56SC (arriba), 2006-03-07-2050-04SC (centro) y 2006-03-10-0944-35SC (abajo), en el dominio del tiempo.
 - Fig. 5.27 Espectrogramas de los registros: 2006-03-03-0650-56SC (arriba), 2006-03-07-2050-04SC (centro) y 2006-03-10-0944-35SC (abajo).
 - Fig. 5.28 Ejemplo de seis registros, en el reconocimiento de eventos tipo S1, S2 y S3, por parte del sistema en el San Cristóbal, durante las pruebas o test ciego-ciego.
 - Fig. 5.29 Ejemplo de registros en el test ciego-ciego, en el que el sistema ha reconocido eventos de tipo S1. Los círculos y las flechas indican los eventos.
 - Fig. 5.30 Ejemplo de registros en el test ciego-ciego, en el que el sistema ha reconocido eventos de tipo S2. Los círculos y las flechas indican los eventos.
 - Fig. 5.31 Ejemplo de registros en el test ciego-ciego, en el que el sistema ha reconocido eventos S3. Los círculos y las flechas indican los eventos.
 - Fig. 5.32 Ejemplo de dos registros del fichero de resultado, en el reconocimiento de eventos tipo NS, por parte del sistema en el test ciego-ciego.
 - Fig. 5.33 Ejemplo de registros en el que el sistema ha catalogado los registros como de tipo NS, en el test ciego-ciego.
 - Fig. 5.34 Resultados obtenidos en %Corr en las pruebas cerradas del HTK, en el porcentaje de aciertos al modificar el parámetro del número de gaussianas en el Telica
 - Fig. 5.35 Resultados obtenidos en %Acc en las pruebas cerradas del HTK, en el porcentaje de aciertos al modificar el parámetro del número de gaussianas en el Telica
 - Tabla 5.36 Resultados obtenidos en las pruebas utilizando 15 Iteraciones, 13 estados, 32 Gaussianas y factor (-p) a -5, en el Telica. %Corr = Porcentaje correcto de etiquetas reconocidas, %Acc = Porcentaje de precisión H= Número correcto de etiquetas, D = Número de eliminaciones, S = Número de sustituciones
 - Fig. 5.36 Sección del fichero de resultados del Análisis manual, en el Proceso selectivo (evento por evento), obtenidos en el Telica. Los valores en rojo representan los registros que han tenido un porcentaje de 100 y que no se cambiarán. Los valores en verde son aquellos que han tenido un porcentaje menor a 60 y que se analizarán para eliminarse de la base de datos.
 - Fig. 5.37 Registros: 2006-02-17-1348-30ST (arriba), 2006-02-17-1948-27ST (centro) y 2006-03-03-1930-50ST (abajo), en el dominio del tiempo.
 - Fig. 5.38 Espectrogramas de los registros: 2006-02-17-1348-30ST (arriba), 2006-02-17-1948-27ST (centro) y 2006-03-03-1930-50ST (abajo).
 - Fig. 5.39 Ejemplo de cuatro registros, en el reconocimiento de eventos tipo T1, T2, T3 y T4, por parte del sistema en el Telica, durante las pruebas o test ciego-ciego.
 - Fig. 5.40 Reconocimiento por parte del sistema, de eventos de tipo T1 y T3 en las pruebas ciego-ciego.
 - Fig. 5.41 Reconocimiento por parte del sistema, de eventos de tipo T2 y T4 en las pruebas ciego-ciego.
 - Fig. 5.42 Parte del fichero de resultado en el que se muestran dos registros marcados por el sistema como NT, en el test ciego-ciego.
 - Fig. 5.43 Registros en los que el sistema ha etiquetado eventos NT en el test ciego-ciego.
 - Fig. 5.44 Señal y Espectrograma de eventos 2541132d del Etna en parte superior y 2541753G del Strómboli en la parte inferior.
 - Fig. 5.45 Magnitud de la FFT de los Espectros de potencia de explosiones 2541705d del Etna en parte superior y 2541753G del Strómboli en la parte inferior.
 - Fig. 5.46 Señal ruidosa 2591214d en el Etna.
 - Fig. 5.47 Señal ruidosa 2591541d en el Etna.
 - Fig. 5.48 Señal del registro 2541705d en el Etna. El sistema reconoce un evento TB.
 - Fig. 5.49 Señal y Espectrograma del registro 2541708d en el Etna.
 - Fig. 5.50 Determinación de parte del observador y el sistema, de los dos eventos TB presentes en el registro 2541344d
 - Fig. 5.51 Determinación de parte del observador y el sistema de los dos eventos TB presentes en el registro 2570114d.
 - Fig. 5.52 Marcación por parte del Observador y el Sistema de eventos en el San Cristóbal
 - Fig. 5.53 Marcación por parte del Observador y el Sistema de eventos en el Telica

APÉNDICE B

- Fig. B.1 Placas interoceánicas y Sismicidad mundial de 1990 a 2000 Fuente: USGS: U.S. Geological Survey. Web: http://earthquake.usgs.gov/earthquakes/world/seismicity_maps/ Los círculos rojos representan los sismos. Los triángulos amarillos los volcanes, y las líneas amarillas, la división entre las placas.
- Fig. B.2. Mapa geológico regional de Nicaragua. Se muestra la cadena de volcanes activos (línea de puntos), la zona de subducción, así como el promedio de velocidades de deriva para posiciones diferentes. Fuente: Frischbutter 2002 [8]
- Fig. B.3. Interacción entre la placa de los Cocos y la Placa Caribe. Se muestra la zona de subducción que forma la cadena volcánica
- Fig. B.4. Huellas de Acahualinca sobre toba volcánica en Managua. Fuente: Instituto nicaragüense de la Cultura. Web: [http://www.inc.gob.ni/images/stories/Huellas%20de%20Acahualinca\(2\).pdf](http://www.inc.gob.ni/images/stories/Huellas%20de%20Acahualinca(2).pdf)
- Fig. B.5 Mapa de Sismicidad y riesgo sísmico en Centro América y Nicaragua de 1990 a 2006 Fuente: National Earthquake Information Center, U.S. Geological Survey. Web: http://earthquake.usgs.gov/regional/world/central_america/seismicity.php
- Fig. B.6. Mapas de densidad de población y Riesgo sísmico normalizado de Nicaragua (fuente INETER)
- Fig. B.7 Mapa de las estaciones sísmicas instaladas por el INETER en la cadena volcánica de Nicaragua (fuente INETER) Página Web: <http://webserver2.ineter.gob.ni/geofisica/sis/redsis.html>
- Fig. B.8 Cadena volcánica cuaternaria en la zona Pacífico de Nicaragua (Fuente INETER)
- Fig. B.9 Caldera y laguna de Apoyo con 6.5 Km., de diámetro. Vista desde el mirador de Catarina. Al fondo el lago Cocibolca o Nicaragua.
- Fig. B.10 Caldera Masaya vista desde el espacio (izq.), conjunto de cráteres de la caldera (der.)
- Fig. B.11 Zonas de amenazas de la cadena volcánica de Nicaragua (Fuente INETER)
- Fig. B.12 Izquierda, Distribución espacial de la producción de sismos durante 2006 en Nicaragua. a) Vista oblicua b) Corte perpendicular a la zona de subducción. Derecha, Número de sismos de 2003 a 2006 (Fuente INETER)
- Fig. B.13 Porcentaje de sismos en las diferentes zonas durante 2006 en Nicaragua (Fuente INETER).
- Fig. B.14 Modelo de capas de la Sismicidad durante 2006 en Nicaragua. (Fuente INETER)
- Fig. B.15 Rango de profundidad de la Sismicidad durante 2006 en Nicaragua (Fuente INETER)
- Fig. B.16 Localización en 2006 de los sismos superficiales (a), Intermedios (b), profundos (c) y los sismos de acuerdo a su magnitud (ML) en Nicaragua. (Fuente INETER)
- Fig. B.17 Localización del número de eventos producidos en profundidad durante 2006 en Centro América (izquierda) y en Nicaragua (derecha) (Fuente INETER)
- Fig. B.18 Mapa de sismicidad superficial en Nicaragua período 1975-2000. En el círculo, la ubicación de la capital, Managua (derecha). Fuente: INETER
- Fig. B.19 Número de eventos producidos por la cadena volcánica de Nicaragua en rango de magnitud y profundidad durante 2006 (Fuente INETER)

APÉNDICE C

- Fig. C.1 Mapa de la ubicación de los cráteres en la cima del Etna a Enero de 1999. Las áreas en gris representan las coladas de lava producidas. Fuente Corsaro y Pompilio [2.21]
- Fig. C.2. Explosión ocurrida el 13 de Noviembre del 2000. Fuente INETER
- Fig. C.3. Explosión ocurrida el 11 de Noviembre del 2004. Foto tomada por la cámara Web del INETER
- Fig. C.4. Sismicidad en Italia durante el período de 1983 a 2002. De acuerdo al Catálogo del INGV (Istituto Nazionale di Geofisica e Vulcanologia) de Italia
- Fig. C.5. Actividad en los cráteres de la cumbre del volcán Etna entre el 7 de Septiembre y el 1 de Octubre de 1999. Mapa Tomado y modificado del original de [Behncke y Tanguy 1999]
- Fig. C.6. Actividad de eventos por día ocurridas durante 1997 en el volcán Strómboli. Fuente: 10/1998 (BGVN 23:10) Larger explosions in January, August, and September 1998. Roberto Carniel, Dipartimento di Georisorse e Territorio, Università di Udine, Via Cotonificio

- Fig. C.7 Número de microsismos ocurridos en el volcán San Cristóbal en el mes de Febrero del 2006 (Fuente INETER)
- Fig. C.8 Número de microsismos ocurridos en el volcán San Cristóbal en el mes de Marzo del 2006 (Fuente INETER)
- Fig. C.9 Número de sismos ocurridos en el volcán San Cristóbal durante el 2006 (Fuente INETER)
- Fig. C.10 Temperaturas en los alrededores del cráter del San Cristóbal de Noviembre del 2005 a Febrero del 2006 (Fuente INETER)
- Fig. C.11 Temperaturas en los alrededores del cráter del San Cristóbal de Enero a Marzo del 2006 (Fuente INETER)
- Fig. C.12 Temperaturas en los diferentes puntos de observación alrededor del cráter del volcán San Cristóbal en el 2006. (Fuente INETER)
- Fig. C.13 Anomalías térmicas RSAM en el volcán San Cristóbal de Noviembre del 2005 a Febrero del 2006 (Fuente INETER)
- Fig. C.14 Anomalías térmicas RSAM en el volcán San Cristóbal en Febrero del 2006 (Fuente INETER)
- Fig. C.15 Anomalías térmicas RSAM en el volcán San Cristóbal en el mes de Marzo del 2006 (Fuente INETER)
- Fig. C.16 Anomalías térmicas y promedio RSAM del San Cristóbal ocurridas en el 2006 (Fuente INETER)
- Fig. C.17 RSAM del volcán San Cristóbal en el mes de Mayo del 2006 cuando se produjeron emanaciones de gas, ceniza y explosiones freáticas. (Fuente INETER)
- Fig. C.18 Número de sismos ocurridos en el volcán Telica en el mes de Febrero del 2006 (Fuente INETER)
- Fig. C.19 Número de sismos ocurridos en el volcán Telica en el mes de Marzo del 2006 (Fuente INETER)
- Fig. C.20 Número de sismos ocurridos en el volcán Telica en el 2006 (Fuente INETER)
- Fig. C.21 Temperaturas en el cráter del volcán Telica durante el 2006 (Fuente INETER)
- Fig. C.22 Anomalías Térmicas y promedio de RSAM del volcán Telica en el período de 01/07/2005 al 28/02/2006 (Fuente INETER)
- Fig. C.23 RSAM del volcán Telica durante el mes de Febrero del 2006 (Fuente INETER)
- Fig. C.24 RSAM del volcán Telica durante el mes de Marzo del 2006 (Fuente INETER)
- Fig. C.25 Anomalías Térmicas y promedio de RSAM del volcán Telica en los años 2005 y 2006 (Fuente INETER)

APÉNDICE D

- Fig. D.1 Imagen de la Densidad Espectral de Potencia de una señal estimada vía Periodograma
- Fig. D.2 Gráfica de la transformada Wavelet de una señal sísmico-volcánica.

APÉNDICE E

- Fig. E.1 Representación de la estimación del poder espectral PSD por medio de Welch del evento TB, (Señal 2541132d) (izq.) y Zoom a los 10 Hz. (der.) en el volcán Etna
- Fig. E.2 Representación de la estimación del poder espectral PSD por medio de Welch del evento TB, (Señal 2541159d) (izq.) y Zoom a los 10 Hz. (der.) en el volcán Etna
- Fig. E.3 Representación de la estimación del poder espectral PSD por medio de Welch del evento Exp, (Señal 2542227G) (izq.) y Zoom a los 10 Hz. (der.) en el volcán Strómboli
- Fig. E.4 Representación de la estimación del poder espectral PSD por medio de Welch del evento Exp, (Señal 2550049G) (izq.) y Zoom a los 10 Hz. (der.) en el volcán Strómboli
- Fig. E.5 Representación de la estimación del poder espectral PSD por medio de Welch del evento S1 (izq.) y Zoom a los 10 Hz. (der.) en el volcán San Cristóbal
- Fig. E.6 Representación de la estimación del poder espectral PSD por medio de Welch del evento S2 (izq.) y Zoom a los 10 Hz. (der.) en el volcán San Cristóbal
- Fig. E.7 Representación de la estimación del poder espectral PSD por medio de Welch del evento S3 (izq.) y Zoom a los 10 Hz. (der.) en el volcán San Cristóbal

- Fig. E.11 Representación de la estimación del poder espectral PSD por medio de Periodograma del Tremor continuo en el volcán San Cristóbal
- Fig. E.12 Representación de la estimación del poder espectral PSD por medio de Welch del Tremor continuo en el volcán San Cristóbal
- Fig. E.13 Representación de la estimación del poder espectral PSD por medio de Periodograma del Tremor continuo en el volcán San Cristóbal con dos componentes de frecuencia (izq.) y Zoom a los 10 Hz. (der.)
- Fig. E.14 Representación de la estimación del poder espectral PSD por medio de Welch del evento T1 (izq.) y Zoom a los 10 Hz. (der.) en el volcán Telica
- Fig. E.15 Representación de la estimación del poder espectral PSD por medio de Welch del evento T2 (izq.) y Zoom a los 10 Hz. (der.) en el volcán Telica
- Fig. E.16 Representación de la estimación del poder espectral PSD por medio de Welch del evento T3 (izq.) y Zoom a los 10 Hz. (der.) en el volcán Telica
- Fig. E.17 Representación de la estimación del poder espectral PSD por medio de Welch del evento T4 (izq.) y Zoom a los 10 Hz. (der.) en el volcán Telica
- Fig. E.18 Wavelet continua (CWT) de eventos de tipo TB en tres registros del Etna. Los círculos representan los eventos TB y las flechas, los puntos donde se concentra la mayor componente energética.
- Fig. E.19 Wavelet 3D de eventos de tipo TB en tres registros del Etna. Las formas de ondas muestran diversos aspectos de las burbujas por cómo es llamado y que conforman este tipo de evento
- Fig. E.20 CWT 3D valor absoluto (abs.) de eventos de tipo TB en tres registros del Etna. Se puede observar la complejidad de este tipo de evento en las formas de sus picos de energía
- Fig. E.21 Wavelet continua (CWT) de eventos de tipo TB en tres registros del Strómboli. Las explosiones marcan diversas llegadas a muy baja frecuencia.
- Fig. E.22 Wavelet 3D de eventos de tipo TB en tres registros del Strómboli. Puede verse la forma de onda de las explosiones a baja frecuencia del Strómboli.
- Fig. E.23 CWT 3D valor absoluto (abs.) de eventos de tipo TB en tres registros del Strómboli. Los picos de energía denotan la forma de las explosiones del Strómboli.
- Fig. E.24 Wavelet continua (CWT) de eventos tipo S1 en tres registros del San Cristóbal. Las flechas indican múltiples picos en frecuencia, característicos de este tipo de evento. La caída en frecuencia en el tiempo es más prolongada.
- Fig. E.25 Wavelet 3D de eventos tipo S1 en tres registros del San Cristóbal. Puede observarse la forma irregular de las ondas que conforman el evento.
- Fig. E.26 CWT 3D valor absoluto (abs.) de eventos de tipo S1 en tres registros del San Cristóbal. La forma de un pico predominante de energía al cual confluyen el resto de llegadas es característica en este tipo de evento.
- Fig. E.27 Wavelet continua (CWT) de eventos tipo S2 en tres registros del San Cristóbal. Al igual que el evento anterior, las flechas indican que este tipo de evento posee múltiples picos de llegadas en energía. A diferencia del evento anterior la caída en frecuencia a lo largo del tiempo es mucho menor.
- Fig. E.28 Wavelet 3D de eventos tipo S2 en tres registros del San Cristóbal. La forma de ondas en las llegadas de energías de este tipo de evento es aún más irregular que del tipo S1.
- Fig. E.29 CWT 3D valor absoluto (abs.) de eventos de tipo S2 en tres registros del San Cristóbal. En este tipo de evento se pueden observar muchos más picos energéticos en llegadas. Cuanto más irregular sea la forma de onda, más picos energéticos produce.
- Fig. E.30 Wavelet continua (CWT) de eventos tipo S3 en tres registros del San Cristóbal. Las flechas indican que posee también múltiples llegadas en frecuencia, la caída en frecuencia es más suave que el evento anterior S2, aunque más corta que la del evento S1.
- Fig. E.31 Wavelet 3D de eventos tipo S3 en tres registros del San Cristóbal. La forma de onda es similar al de tipo S1, solo que este evento es mucho mayor en el tiempo presentando una caída en energía diferente, lo que se observa en la figura.
- Fig. E.32 CWT 3D valor absoluto (abs.) de eventos de tipo S3 en tres registros del San Cristóbal. Este tipo de evento presenta un componente máximo en energía al que confluyen todas las llegadas, se asemeja al del tipo S1, con la diferencia de la caída en frecuencia al final del evento.
- Fig. E.33 Wavelet continua CWT del tremor continuo en el volcán San Cristóbal. Se observa como el tremor está compuesto por múltiples bandas de energía.

- Fig. E.34 Wavelet 3D del tremor continuo presente en el volcán San Cristóbal. Se observa la multiplicidad de ondas en las diversas llegadas de energía de los armónicos presentes en el tremor.
- Fig. E.35 Wavelet 3D valor absoluto (abs.) del tremor continuo presente en el volcán San Cristóbal. En la gráfica se observan los diversos picos de energía que componen el tremor del San Cristóbal.
- Fig. E.36 Wavelet continua (CWT) de eventos tipo T1 en tres registros del Telica. En el círculo, se pueden observar las dos llegadas energéticas que caracterizan a este tipo de evento como evento doble. Además la flecha indica la forma de la caída en frecuencia de este tipo de evento.
- Fig. E.37 Wavelet 3D de eventos tipo T1 en tres registros del Telica. Se pueden observar las dos formas de ondas que corresponden a las dos llegadas en energía de este tipo de evento.
- Fig. E.38 CWT 3D valor absoluto (abs.) de eventos de tipo T1 en tres registros del Telica. Aunque los tres eventos aquí mostrados difieren a lo largo del tiempo (ver estadísticas de los tipos de eventos), es en el tercero que se observa con mayor precisión, la llegada de las dos componentes en energía.
- Fig. E.39 Wavelet continua (CWT) de eventos tipo T2 en tres registros del Telica. La flecha indica a diferencia del T1, una única componente mayor de llegada en energía y la curva indica la forma de caída en frecuencia de este tipo de evento.
- Fig. E.40 Wavelet 3D de eventos tipo T2 en tres registros del Telica. Una única onda en frecuencia muestra la llegada de la componente mayor en energía.
- Fig. E.41 CWT 3D valor absoluto (abs.) de eventos de tipo T2 en tres registros del Telica. Se puede observar cómo se muestra el pico en frecuencia.
- Fig. E.42 Wavelet continua (CWT) de eventos tipo T3 en tres registros del Telica. Al igual que el de tipo T2 este tipo de evento no presenta múltiples llegadas en frecuencia, la flecha superior lo indica. Sin embargo, como indica la flecha más a la derecha, la forma de caída en frecuencia es mucho más abrupta en este tipo de evento.
- Fig. E.43 Wavelet 3D de eventos tipo T3 en tres registros del Telica. La onda energética en la forma de las componentes mayores de este tipo de evento no es tan senoidal como la del evento T2, es más aplanada.
- Fig. E.44 CWT 3D valor absoluto (abs.) de eventos de tipo T3 en tres registros del Telica. Se puede observar que el pico en frecuencia en este tipo de evento es más pronunciado, similar al del tipo T1
- Fig. E.45 Wavelet continua (CWT) de eventos tipo T4 en tres registros del Telica. En el círculo se aprecian múltiples llegadas de energía para este tipo de evento, asimismo la flecha indica una caída en frecuencia mucho más abrupta y espaciada a lo largo del tiempo, por lo que este tipo de evento podría tratarse un evento de tipo LP.
- Fig. E.46 Wavelet 3D de eventos tipo T4 en tres registros del Telica. La forma de onda energética es más alargada a lo largo del tiempo.
- Fig. E.47 CWT 3D valor absoluto (abs.) de eventos de tipo T4 en tres registros del Telica. El pico predominante en frecuencia en este tipo de evento es similar al del T1, solo que no posee dos picos mayores, con lo que los diferentes picos menores quedan integrados en el pico predominante.

APÉNDICE F

- Figura F.1 Contenido del fichero de cabecera “dts” en la campaña del Etna

APÉNDICE G

- Fig. G.1 Diagrama de flujo con las interfaces del sistema
- Fig. G.2 Interfaz de Inicio del Sistema de Análisis y Reconocimiento de Eventos Sísmico-volcánicos
- Fig. G.3 Interfaz de Menú Principal de la Aplicación
- Fig. G.4 Interfaz de Ayuda
- Fig. G.5 Interfaz del bloque I, Ejemplos de los tipos de eventos con análisis de Fourier
- Fig. G.6 Interfaz del bloque I, Ejemplos de los tipos de eventos con análisis Wavelet

- Fig. G.7 Interfaz del bloque II, Conversión de Ficheros
- Fig. G.8 Pantalla de éxito en la creación de un fichero
- Fig. G.9 Pantalla de éxito en la creación de un fichero
- Fig. G.10 Interfaz del bloque II, Ejemplo de Conversión de Ficheros
- Fig. G.11 Pantalla de Interfaz para el análisis espectral de los ficheros
- Fig. G.12 Pantalla de selección de los ficheros a analizar
- Fig. G.13 Pantalla con un ejemplo de la Interfaz para el análisis espectral. Se observa el menú de análisis
- Fig. G.14 Pantalla de análisis espectral. Selección de análisis por espectro de frecuencia
- Fig. G.15 Pantalla de análisis espectral. Selección de análisis Wavelet
- Fig. G.16 Pantalla de análisis espectral. Selección de análisis Wavelet 3D
- Fig. G.17 Pantalla de error en la Conversión de Ficheros
- Fig. G.18 Pantalla de error al seleccionar un fichero
- Fig. G.19 Pantalla de error al graficar un fichero
- Fig. G.20 Pantalla de error al reproducir un fichero
- Fig. G.21 Pantalla de confirmación para eliminar un fichero
- Fig. G.22 Pantalla que confirma que el fichero ha sido eliminado
- Fig. G.23 Pantalla de error al seleccionar la ruta de los ficheros
- Fig. G.24 Pantalla de error en la entrada de los ficheros.



LISTADO DE TABLAS

CAPÍTULO 2

- Tabla 2.1 Ejemplo de la Matriz de confusión, generada por la herramienta HResults.

CAPÍTULO 3

- Tabla 3.1: Caracterización geológica del volcán San Cristóbal. Fuente [INETER 2001]; [SINAPRED 2005]
- Tabla 3.2: Caracterización geológica del volcán Telica. Fuente [INETER 2001]; [SINAPRED 2005]
- Tabla 3.3. Número de registros en el volcán Etna por estación en el array sísmico “Torre del Filósofo”, en Septiembre 1999.
- Tabla 3.4 Tipos de formatos a convertir en el desarrollo del proceso de los registros

CAPÍTULO 4

- Tabla 4.1 Estadísticas correspondientes a la desviación estándar, duración máxima, duración mínima y promedio de los eventos en el volcán Etna de Italia
- Tabla 4.2 Estadísticas correspondientes a la desviación estándar, duración máxima, duración mínima y promedio de los eventos en el volcán Strómboli de Italia.
- Tabla 4.3 Estadísticas correspondientes a la desviación estándar, duración máxima, duración mínima y promedio de los eventos en el volcán San Cristóbal de Nicaragua.
- Tabla 4.4 Estadísticas correspondientes a la desviación estándar, duración máxima, duración mínima y promedio de los eventos en el volcán Telica de Nicaragua.
- Tabla. 4.5 Total de registros etiquetados en los cuatro volcanes de estudio
- Tabla. 4.6 Número total de segmentaciones, de acuerdo a los diferentes tipos de eventos en cada uno de los volcanes de estudio

CAPÍTULO 5

- Tabla 5.1 Resultados obtenidos del HTK en el primer experimento en el Etna. %Corr = Porcentaje correcto de etiquetas reconocidas, %Acc = Porcentaje de precisión H= Número correcto de etiquetas, D = Número de eliminaciones, S = Número de sustituciones, I = Número de Inserciones, N = Número total de etiquetas. Ins = Inserciones, Del = Borrados.
- Tabla 5.2 Resultados obtenidos en las pruebas cerradas del HTK, con 11 estados y 16 Gausianas al modificar el parámetro del número de iteraciones en el Etna %Corr = Porcentaje correcto de etiquetas reconocidas, %Acc = Porcentaje de precisión Ins = Inserciones, Sust = Sustituciones, Del = Borrados.
- Tabla 5.3 Resultados obtenidos en las pruebas cerradas del HTK, con 6 iteraciones y 11 estados, al modificar el parámetro del número de gaussianas en el Etna. %Corr = Porcentaje correcto de etiquetas reconocidas, %Acc = Porcentaje de precisión Ins = Inserciones, Sust = Sustituciones, Del = Borrados.
- Tabla 5.4 Resultados obtenidos en las pruebas cerradas del HTK, en la matriz de confusión, al modificar a 96 el parámetro del número de gaussianas en el Etna. %Corr = Porcentaje correcto de etiquetas reconocidas, %Acc = Porcentaje de precisión H= Número correcto de etiquetas, D = Número de eliminaciones, S = Número de sustituciones, I = Número de Inserciones, N = Número total de etiquetas. Ins = Inserciones, Del = Borrados.
- Tabla 5.5 Resultados obtenidos con 6 iteraciones, 11 estados y 32 Gausianas, en la modificación del parámetro “-p” del Hvite en el Etna %Corr = Porcentaje correcto de etiquetas reconocidas, %Acc = Porcentaje de precisión Ins = Inserciones, Sust = Sustituciones, Del = Borrados
- Tabla 5.6 Resultados al modificar el parámetro del número de estados en el Etna, obtenidos en las pruebas cerradas del HTK, con 6 iteraciones, 32 gaussianas y factor -p en -15 %Corr = Porcentaje correcto de etiquetas reconocidas, %Acc = Porcentaje de precisión Ins = Inserciones, Sust = Sustitución, Del = Borrados.

- Tabla 5.7 Resultados obtenidos en las pruebas cerradas del HTK, con 6 iteraciones, 13 estados, y factor $-p$ (-15), al modificar el parámetro del número de gaussianas en el Etna. %Corr = Porcentaje correcto de etiquetas reconocidas, %Acc = Porcentaje de precisión.
- Tabla 5.8 Resultados obtenidos del HTK en el experimento final del Etna. Utilizando 6 Iteraciones, 13 estados, 32 Gaussianas y factor $-p = -15$ %Corr = Porcentaje correcto de etiquetas reconocidas, %Acc = Porcentaje de precisión H= Número correcto de etiquetas, D = Número de eliminaciones, S = Número de sustituciones, I = Número de Inserciones, N = Número total de etiquetas. Ins = Inserciones, Del = Borrados
- Tabla 5.9 Resultados obtenidos al eliminar registros con bajo porcentaje del Etna. Utilizando 6 Iteraciones, 13 estados, 32 Gaussianas y factor $-p = -15$ %Corr = Porcentaje correcto de etiquetas reconocidas, %Acc = Porcentaje de precisión H= Número correcto de etiquetas, D = Número de eliminaciones, S = Número de sustituciones, I = Número de Inserciones, N = Número total de etiquetas. Ins = Inserciones, Del = Borrados.
- Tabla 5.10 Resultados obtenidos en el primer tipo de las pruebas abiertas con el Etna %Corr = Porcentaje correcto de etiquetas reconocidas, %Acc = Porcentaje de precisión H= Número correcto de etiquetas, D = Número de eliminaciones, S = Número de sustituciones, I = Número de Inserciones, N = Número total de etiquetas. Ins = Inserciones, Del = Borrados.
- Tabla 5.11 Primeros experimentos con 11 estados, 6 iteraciones y 16 Gaussianas en el Strómboli %Corr = Porcentaje correcto de etiquetas reconocidas, %Acc = Porcentaje de precisión H= Número correcto de etiquetas, D = Número de eliminaciones, S = Número de sustituciones, I = Número de Inserciones, N = Número total de etiquetas. Ins = Inserciones, Del = Borrados.
- Tabla 5.12 Resultados obtenidos del HTK en los experimentos con 11 estados, 6 iteraciones y 16 Gaussianas realizados con el Strómboli, al aumentar los registros segmentados. %Corr = Porcentaje correcto de etiquetas reconocidas, %Acc = Porcentaje de precisión H= Número correcto de etiquetas, D = Número de eliminaciones, S = Número de sustituciones, I = Número de Inserciones, N = Número total de etiquetas. Ins = Inserciones, Del = Borrados.
- Tabla 5.13 Resultados obtenidos en las pruebas cerradas del HTK, con 11 estados y 16 Gaussianas al modificar el parámetro del número de iteraciones en el Strómboli %Corr = Porcentaje correcto de etiquetas reconocidas, %Acc = Porcentaje de precisión Ins = Inserciones, Sust = Sustituciones, Del = Borrados.
- Tabla 5.14 Resultados obtenidos en las pruebas cerradas del HTK, con 6 iteraciones, 11 estados, y factor $-p$ (0.0), al modificar el parámetro del número de gaussianas en el Strómboli %Corr = Porcentaje correcto de etiquetas reconocidas, %Acc = Porcentaje de precisión.
- Tabla 5.15 Resultados obtenidos con 6 iteraciones, 11 estados y 32 Gaussianas, en la modificación del parámetro “-p” del Hvite en el Strómboli %Corr = Porcentaje correcto de etiquetas reconocidas, %Acc = Porcentaje de precisión Ins = Inserciones, Sust = Sustituciones, Del = Borrados.
- Tabla 5.16 Resultados al modificar el parámetro del número de estados en el Strómboli, obtenidos en las pruebas cerradas del HTK, con 6 iteraciones, 32 gaussianas y factor $-p$ en -15 %Corr = Porcentaje correcto de etiquetas reconocidas, %Acc = Porcentaje de precisión Ins = Inserciones, Sust = Sustituciones, Del = Borrados.
- Tabla 5.17 Resultados obtenidos en las pruebas cerradas del HTK, con 6 iteraciones, 15 estados, y factor $-p$ (-15), al modificar el parámetro del número de gaussianas en el Strómboli. %Corr = Porcentaje correcto de etiquetas reconocidas, %Acc = Porcentaje de precisión.
- Tabla 5.18 Resultados obtenidos en las pruebas del Strómboli. Utilizando 6 Iteraciones, 15 estados, 32 Gaussianas y factor $-p = -15$ %Corr = Porcentaje correcto de etiquetas reconocidas, %Acc = Porcentaje de precisión H= Número correcto de etiquetas, D = Número de eliminaciones, S = Número de sustituciones, I = Número de Inserciones, N = Número total de etiquetas. Ins = Inserciones, Del = Borrados.
- Tabla 5.19 Resultados obtenidos al eliminar registros con bajo porcentaje del Strómboli. Utilizando 6 Iteraciones, 15 estados, 32 Gaussianas y factor $-p = -15$ %Corr = Porcentaje correcto de etiquetas reconocidas, %Acc = Porcentaje de precisión H= Número correcto de etiquetas, D = Número de eliminaciones, S = Número de sustituciones, I = Número de Inserciones, N = Número total de etiquetas. Ins = Inserciones, Del = Borrados.
- Tabla 5.20 Resultados obtenidos en el primer tipo de las pruebas abiertas con el Strómboli %Corr = Porcentaje correcto de etiquetas reconocidas, %Acc = Porcentaje de precisión H= Número correcto de etiquetas, D = Número de eliminaciones, S = Número de sustituciones, I = Número de Inserciones, N = Número total de etiquetas. Ins = Inserciones, Del = Borrados.

- Tabla 5.21 Resultados obtenidos del HTK en el primer experimento con el San Cristóbal. %Corr = Porcentaje correcto de etiquetas reconocidas, %Acc = Porcentaje de precisión H= Número correcto de etiquetas, D = Número de eliminaciones, S = Número de sustituciones, I = Número de Inserciones, N = Número total de etiquetas. Ins = Inserciones, Del = Borrados.
- Tabla 5.22 Resultados obtenidos en las pruebas cerradas del HTK, con 11 estados y 16 Gaussianas al modificar el parámetro del número de iteraciones en el San Cristóbal %Corr = Porcentaje correcto de etiquetas reconocidas, %Acc = Porcentaje de precisión H= Número correcto de etiquetas, D = Número de eliminaciones, S = Número de sustituciones, I = Número de Inserciones, N = Número total de etiquetas. Ins = Inserciones, Del = Borrados
- Tabla 5.23 Resultados obtenidos en las pruebas cerradas del HTK, al modificar el parámetro del número de gaussianas en el San Cristóbal.
- Tabla 5.24 Resultados obtenidos con 6 iteraciones, 11 estados y 32 Gaussianas, en la modificación del parámetro “-p” del Hvite en el San Cristóbal. %Corr = Porcentaje correcto de etiquetas reconocidas, %Acc = Porcentaje de precisión Ins = Inserciones, Sust = Sustituciones, Del = Borrados.
- Tabla 5.25 Resultados obtenidos con 6 iteraciones, 11 estados, 32 Gaussianas y factor p 5, en la modificación del número de estados en el San Cristóbal. %Corr = Porcentaje correcto de etiquetas reconocidas, %Acc = Porcentaje de precisión Ins = Inserciones, Sust = Sustituciones, Del = Borrados.
- Tabla 5.26 Resultados obtenidos en las pruebas cerradas del HTK, con 6 iteraciones, 13 estados, y factor -p (5), al modificar el parámetro del número de gaussianas en el San Cristóbal. %Corr = Porcentaje correcto de etiquetas reconocidas, %Acc = Porcentaje de precisión.
- Tabla 5.27 Resultados obtenidos en las pruebas del San Cristóbal. Utilizando 6 Iteraciones, 13 estados, 32 Gaussianas y factor (-p) de 5 %Corr = Porcentaje correcto de etiquetas reconocidas, %Acc = Porcentaje de precisión H= Número correcto de etiquetas, D = Número de eliminaciones, S = Número de sustituciones, I = Número de Inserciones, N = Número total de etiquetas. Ins = Inserciones, Del = Borrados.
- Tabla 5.28 Resultados obtenidos al eliminar registros con bajo porcentaje del San Cristóbal, utilizando 6 Iteraciones, 13 estados, 32 Gaussianas y factor -p = 5 %Corr = Porcentaje correcto de etiquetas reconocidas, %Acc = Porcentaje de precisión H= Número correcto de etiquetas, D = Número de eliminaciones, S = Número de sustituciones, I = Número de Inserciones, N = Número total de etiquetas. Ins = Inserciones, Del = Borrados.
- Tabla 5.29 Resultados obtenidos en las pruebas abiertas del San Cristóbal, utilizando 6 Iteraciones, 13 estados, 32 Gaussianas y factor -p = 5 %Corr = Porcentaje correcto de etiquetas reconocidas, %Acc = Porcentaje de precisión H= Número correcto de etiquetas, D = Número de eliminaciones, S = Número de sustituciones, I = Número de Inserciones, N = Número total de etiquetas. Ins = Inserciones, Del = Borrados
- Tabla 5.30 Resultados obtenidos del HTK en las pruebas con 1,240 registros y los parámetros iniciales en el Telica %Corr = Porcentaje correcto de etiquetas reconocidas, %Acc = Porcentaje de precisión H= Número correcto de etiquetas, D = Número de eliminaciones, S = Número de sustituciones, I = Número de Inserciones, N = Número total de etiquetas. Ins = Inserciones, Del = Borrados.
- Tabla 5.31 Resultados obtenidos con 11 estados, 16 Gaussianas, p = 0.0 en la modificación del parámetro del número de iteraciones en el Telica %Corr = Porcentaje correcto de etiquetas reconocidas, %Acc = Porcentaje de precisión Ins = Inserciones, Sust = Sustituciones, Del = Borrados.
- Tabla 5.32 Resultados obtenidos en las pruebas cerradas del HTK con 15 iteraciones, al modificar el parámetro del número de gaussianas en el Telica. %Corr = Porcentaje correcto de etiquetas reconocidas, %Acc = Porcentaje de precisión.
- Tabla 5.33 Resultados obtenidos con 15 iteraciones, 11 estados y 32 Gaussianas, en la modificación del parámetro “-p” del Hvite en el Telica %Corr = Porcentaje correcto de etiquetas reconocidas, %Acc = Porcentaje de precisión Ins = Inserciones, Sust = Sustituciones, Del = Borrados
- Tabla 5.34 Resultados obtenidos con 15 iteraciones, 11 estados, 32 Gaussianas y factor p -5, en la modificación del número de estados en el Telica %Corr = Porcentaje correcto de etiquetas reconocidas, %Acc = Porcentaje de precisión Ins = Inserciones, Sust = Sustituciones, Del = Borrados.

- Tabla 5.35 Resultados obtenidos en las pruebas cerradas del HTK, con 15 iteraciones, 13 estados, y factor $-p$ (-5), al modificar el parámetro del número de gaussianas en el Telica. %Corr = Porcentaje correcto de etiquetas reconocidas, %Acc = Porcentaje de precisión
- Tabla 5.36 Resultados obtenidos en las pruebas utilizando 15 Iteraciones, 13 estados, 32 Gaussianas y factor $(-p)$ a -5, en el Telica. %Corr = Porcentaje correcto de etiquetas reconocidas, %Acc = Porcentaje de precisión H= Número correcto de etiquetas, D = Número de eliminaciones, S = Número de sustituciones, I = Número de Inserciones, N = Número total de etiquetas. Ins = Inserciones, Del = Borrados.
- Tabla 5.37 Resultados obtenidos al eliminar registros con bajo porcentaje del Telica. Utilizando 6 Iteraciones, 13 estados, 32 Gaussianas y factor $-p = -15$ %Corr = Porcentaje correcto de etiquetas reconocidas, %Acc = Porcentaje de precisión, H= Número correcto de etiquetas, D = Número de eliminaciones, S = Número de sustituciones, I = Número de Inserciones, N = Número total de etiquetas. Ins = Inserciones, Del = Borrados.
- Tabla 5.38 Resultados obtenidos en las pruebas abiertas del San Cristóbal, utilizando 6 Iteraciones, 13 estados, 32 Gaussianas y factor $-p = 5$ %Corr = Porcentaje correcto de etiquetas reconocidas, %Acc = Porcentaje de precisión, H= Número correcto de etiquetas, D = Número de eliminaciones, S = Número de sustituciones, I = Número de Inserciones, N = Número total de etiquetas. Ins = Inserciones, Del = Borrados.
- Tabla 5.39 Resultados obtenidos del HTK, con el Etna y el Strómboli, en las pruebas combinados de ambas bases de datos. %Corr = Porcentaje correcto de etiquetas reconocidas, %Acc = Porcentaje de precisión, H= Número correcto de etiquetas, D = Número de eliminaciones, S = Número de sustituciones, I = Número de Inserciones, N = Número total de etiquetas. Ins = Inserciones, Del = Borrados.
- Tabla 5.40 Resultados de la combinación de los modelos en ambos volcanes de Italia.
- Tabla 5.41 Resultados obtenidos del HTK los experimentos de pruebas cruzadas con las bases de datos de los modelos de entrenamiento del Telica y del San Cristóbal. %Corr = Porcentaje correcto de etiquetas reconocidas, %Acc = Porcentaje de precisión, H= Número correcto de etiquetas, D = Número de eliminaciones, S = Número de sustituciones, I = Número de Inserciones, N = Número total de etiquetas. Ins = Inserciones, Del = Borrados.
- Tabla 5.42 Resumen de resultados obtenidos en las diferentes pruebas del Telica y San Cristóbal. BD = Base de datos, ENT = Entrenamiento.
- Tabla. 5.43 Número de eventos incluidos en las pruebas cruzadas con los cuatro volcanes.
- Tabla 5.44 Resultados obtenidos del HTK los experimentos de pruebas cruzadas con las bases de datos de los modelos de entrenamiento de Italia y Nicaragua. %Corr = Porcentaje correcto de etiquetas reconocidas, %Acc = Porcentaje de precisión, H= Número correcto de etiquetas, D = Número de eliminaciones, S = Número de sustituciones, I = Número de Inserciones, N = Número total de etiquetas. Ins = Inserciones, Del = Borrados.
- Tabla 5.45 Variación del valor de los parámetros en las pruebas en cada volcán.
- Tabla 5.46 Resultados iniciales y finales obtenidos en las pruebas en cada volcán %Corr = Porcentaje correcto de etiquetas reconocidas, %Acc = Porcentaje de precisión.
- Tabla 5.47 Resultados finales obtenidos en las pruebas en cada volcán %Corr = Porcentaje correcto de etiquetas reconocidas, %Acc = Porcentaje de precisión

APÉNDICE E

- Tabla E.1 Relación de los componentes en frecuencia y magnitud de los eventos en los volcanes Etna e Strómboli de Italia. Análisis mediante Periodograma Welch
- Tabla E.2 Relación de los componentes en frecuencia y magnitud de los eventos en los volcanes San Cristóbal y Telica de Nicaragua. Análisis mediante Periodograma Welch



APÉNDICES



*Volcán Maderas (Isla de Ometepe), Nicaragua.
Por: Ligdamis A. Gutiérrez E.*



APÉNDICE A: Referencias a trabajos publicados

Fruto del desarrollo del presente trabajo ha sido la publicación en los siguientes medios:

A) REVISTAS

2009

Ibáñez, J. M., Benítez, C., Gutiérrez, L. A., Cortés, G., García-Yeguas, A., & Alguacil, G. (2009). Classification of seismo-volcanic signals using Hidden Markov Models: an application to Stromboli and Etna volcanoes. *Journal of Volcanology and Geothermal Research*. Volume 187, Issues 3-4, 10 November 2009, Pages 218-226.

2009

Gutiérrez, L., Ibáñez, J., Cortés, G., Ramírez, J., Benítez, C., Tenorio, V., & Isaac, A. (2009). Volcano-seismic signal detection and classification processing using Hidden Markov Models. Application to San Cristóbal volcano, Nicaragua. *Geoscience and Remote Sensing Symposium, 2009 IEEE International, IGARSS 2009*, Volume: 4 On page(s): IV-522 Conference Location: Cape Town E-ISBN: 978-1-4244-3395-7 Print ISBN: 978-1-4244-3394-0 INSPEC Accession Number: 11246276 Digital Object Identifier: 10.1109/IGARSS.2009.5417428

2006

Gutiérrez, L., Ramírez, J., Benítez, C., Ibáñez, J., Almendros, J., & García-Yeguas, A. (2006). HMM-based classification of seismic events recorded at Stromboli and Etna Volcanoes. *IEEE International, IGARSS 2006* on page(s): 2765 Conference Location: Denver, CO, USA Print ISBN: 0-7803-9510-7 Digital Object Identifier: 10.1109/IGARSS.2006.711

B) CAPÍTULOS DE LIBROS

2011

INTECH Open Access Publisher 2011

Book title: *Hidden Markov Models, Theory and Applications* (ISBN 978-953-307-208-1)

Chapter title: Volcano-seismic signal detection and classification processing using Hidden Markov Models. Application to San Cristóbal and Telica volcanoes, Nicaragua

Authors: Ligdamis Gutierrez, Javier Ramirez, Jesús Ibanez and Carmen Benitez Book edited by: Dr. Przemyslaw Dymarski

2009

VOLUME Project Consortium, Dublin, 2009.

Book title: *Volcanoes, understanding subsurface mass movement*. Proyecto VOLUME, EU PF6 (No. 018471)

Chapter title: Analysis of volcanic seismicity at Deception Island, Stromboli and Etna Volcanoes using an automatic HMM Based Recognition Method

Authors: M. Carmen Benítez, Jesús Ibáñez, Luz García, Ligdamis Gutiérrez, Guillermo Cortés, and Isaac Álvarez.

C) CONGRESOS Y PONENCIAS NACIONALES E INTERNACIONALES:

2009

Gutiérrez, L., Ibáñez, J., Cortés, G., Ramírez, J., Benítez, C., Tenorio, V., & Isaac, A. (2009). Volcano-seismic signal detection and classification processing using Hidden Markov Models. Application to San Cristóbal volcano, Nicaragua. Geoscience and Remote Sensing Symposium, 2009 IEEE International, IGARSS 2009.

Tipo de Participación: Póster

2009

Ligdamis A. Gutiérrez-Espinoza, Jesús M. Ibáñez, Carmen Benítez, Javier Ramírez, and Virginia Tenorio (2009). Volcano-seismic signal detection and classification processing using Hidden Markov Models. Application to Telica and San Cristóbal volcanoes, Nicaragua. XI International Meeting of Volcán de Colima. Colima México- 2009. Página Web: <http://www.ucol.mx/volcan/reunion2009/programa.htm>

Tipo de Participación: Póster

2009

Ibáñez, Jesús M. Carmen Benítez, Javier Ramírez, Ligdamis A. Gutiérrez, Guillermo Cortés and Araceli García-Yeguas (2009). Automatic discrimination of seismic signals using Hidden Markov Models using data of Stromboli and Etna volcanoes. XI International Meeting of Volcán de Colima. Colima México- 2009.

Página Web: <http://www.ucol.mx/volcan/reunion2009/programa.htm>

Tipo de Participación: Póster

2007

Gutiérrez, L., Strauch, W., Tenorio, V., Ibáñez, J., Ramírez, J., And Benítez, C. (2007). Classification of volcano-seismic signals using Hidden Markov Models Application to Telica and San Cristóbal volcanoes, Nicaragua. 20th Colloquium on Latin American Earth Sciences. Kiel, Alemania.

Tipo de Participación: Póster.

2007

Ligdamis A. Gutiérrez (2007) Classification of volcano-seismic signals using Hidden Markov Models Application to Telica and San Cristóbal volcanoes, Nicaragua. XVI Jornada Geocientífica “Enrique Campbell” In Memoriam. Asociación Nacional de Geólogos y Profesionales Afines (ANGPA). Managua, Nicaragua

Tema: Geociencias Eje del Desarrollo Sostenible.

Tipo Participación: Ponencia

2006

Ligdamis A. Gutiérrez-Espinoza, Jesús M. Ibáñez, Carmen Benítez, Javier Ramírez, Javier Almendros, And Araceli García-Yeguas (2006). HMM-based classification of seismic events recorded at Stromboli and Etna Volcanoes. 2006 IEEE International Geoscience and Remote Sensing Symposium & 27th Canadian Symposium on Remote Sensing. Denver Colorado, U.S.A.

Tipo de Participación: Paper Exposición.

2006

Ligdamis A. Gutiérrez E. (2006) Clasificación de eventos sísmicos en los volcanes San Cristóbal y Telica Taller científico-técnico: Avances y Resultados en Geofísica/INETER. Ponencia organizada por el INETER, 13 de Febrero, 2007, con el objetivo de presentar los avances y resultados obtenidos en esta Dirección en el año 2006. En el taller participaron técnicos de otras Áreas del INETER, de SINAPRED (Sistema Nacional para la Prevención, Mitigación y Atención de Desastres), y Universidades de Nicaragua. Tema: Red de monitoreo y alerta temprana
Tipo Participación: Ponencia

2005

Ligdamis A. Gutiérrez-Espinoza, Jesús M. Ibáñez, Carmen Benítez, Javier Ramírez, Javier Almendros, Araceli García-Yeguas (2005). Discrimination between different volcanic signals using Hidden Markov Model: An example of Stromboli and Etna volcanoes. Congreso Internacional de Vulcanología, Colima, México.
Tipo de Participación: Póster.



APÉNDICE B: *Á*mbito sísmico-volcánico de Centro América y Nicaragua

1.- INTRODUCCIÓN

Este apéndice pretende situar a los volcanes de la región Centro Americana y particularmente a la república de Nicaragua, en el marco geodinámico al cual pertenecen. En el capítulo primero del presente trabajo se abordaba el tema del ámbito de los volcanes y las señales sísmico-volcánicas que se producen en ellos. Los cuatro volcanes de estudio, están comprendidos en dos zonas geográficas, la italiana y la nicaragüense, de estas dos, la primera de ellas ha sido muy documentada y los volcanes ahí presentes han sido protagonistas de múltiples estudios, abordado por diferentes autores, por lo mismo, los eventos sísmico-volcánicos que se manifiestan son muy conocidos y característicos. Esta profusa difusión no ha sido la misma en la región centro americana y concretamente en Nicaragua. Es por ello que no es fácil encontrar gran abundancia en cuanto a bibliografía e información, en particular sobre cada uno de los edificios volcánicos, así como de sus productos. Por lo que se ha creído conveniente incluir en este apéndice, un compendio de todo lo existente, con la finalidad de poder facilitar información y encuadrar a Nicaragua en el margo geodinámico que posee, así como los procesos sísmicos que acontecen, su historia y la influencia sobre la población actual y futura de Nicaragua en la región Centroamericana.

2.- SITUACIÓN DE NICARAGUA: UN PAÍS SÍSMICO-VOLCÁNICO

Nicaragua se sitúa dentro del llamado cinturón de fuego del Pacífico (Figura B1). Adicionalmente, el país se encuentra ubicado en el borde la colisión de dos placas interoceánicas; la placa de los Cocos, situada en el Pacífico con la placa del Caribe [Burke et. al. 1984], (figuras B1 y B2). Esta particular distribución y situación geográfica provoca en el territorio una amplia producción sísmica (figura B1). Además de haber conformado tanto la orografía como la geografía del país a lo largo de la historia, creando a lo largo de diferentes eras geológicas, una gran cantidad y variedad de estructuras volcánicas, cuya actividad se remonta desde el período cretácico hasta la actualidad. Diferentes autores han descrito y detallado tanto la morfología, la geología, la química y la composición geológica de las rocas del denominado Arco Volcánico Centro Americano y de Nicaragua [Cruden A. 1989]; [Leemann y Morris 1994]; [Schmincke et. al. 2008]; [Valrey 1988]; [Costantini et. al. 2008]. En la actualidad en Nicaragua, de la gran variedad de estructuras de origen volcánico (unas 200 a 600) existentes [Lilljequist et. al. 1987], se pueden localizar al menos 8 edificios volcánicos activos (INETER 1982), además de otros diversos cuya actividad se encuentra latente.

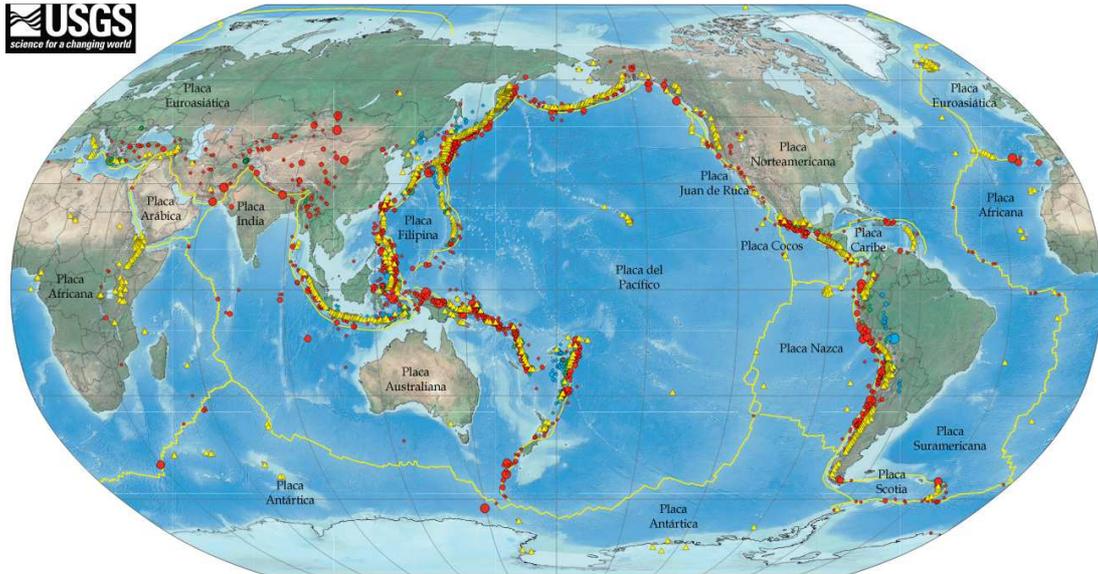


Fig. B.1 Placas interoceánicas y Sismicidad mundial de 1990 a 2000

Fuente: USGS: U.S. Geological Survey.

Web: http://earthquake.usgs.gov/earthquakes/world/seismicity_maps/

Los círculos rojos representan los sismos. Los triángulos amarillos los volcanes, y las líneas amarillas, la división entre las placas.

La geológica en Centro América es compleja, debido a la interacción de las placas Cocos, Nazca y Caribe entre las placas de Norte y Sur América, las cuales producen una diversidad estructural en la corteza de la tierra. Varios estudios han indicado que en el pasado se han producido una gran cantidad de eventos geotectónicos, que dieron comienzo desde tiempos del Jurásico (incluso desde el Cretácico). En un contexto geológico, Nicaragua se encuentra situada en la parte meridional del bloque los Chortís (Fig. B.2), que limita al noroeste por la Sinistral² Motagua-Polochic-Faults (incluyendo la Caimán Trough) y al Sureste por la Dextral³ Hess-Escarpment. [Álvarez et. al. 2008], [Frischbutter 2002]. Su interacción hace que se presenten una gran variedad de eventos sísmico-tectónicos y sísmico-volcánicos. La mayor consecuencia entre la acción de la placa Caribe y la de Cocos es la llamada zona de subducción, que hace que la placa de los cocos se interne por debajo de la placa Caribe, y creando así la cadena volcánica que atraviesa Nicaragua (Figura B.3). Esto provoca que la mayor concentración de volcanes se encuentre en la región del pacífico. De hecho, una de las consecuencias de la actividad volcánica a lo largo de las épocas, ha dispuesto que esta región sea una zona muy fértil, en la que se ha asentado la mayor parte de la población de la zona. A lo largo de la historia, los pobladores han tenido que convivir con las diversas manifestaciones volcánicas de los alrededores. Un ejemplo de ello representan las denominadas “huellas de Acahualinca” (Fig. B.4), descubiertas accidentalmente en 1874 [Flint 1884]; [Brinton 1887], datan del Holoceno, situadas en el noreste de Managua la capital de Nicaragua, con más de 6,000 años de antigüedad, fueron formadas por antiguos habitantes y animales en toba o tufa volcánica, después de la erupción de algunos de los cinco volcanes que contiene la capital [Lockley et. al. 2009]; [Schmincke et. al. 2008]; [Schmincke et. al. 2010]. Actualmente, en la capital los

² Movimiento o deslizamiento lateral izquierdo de la corteza terrestre

³ Movimiento o deslizamiento lateral derecho de la corteza terrestre

espacios dejados por estos cráteres se encuentran ocupados por las lagunas: Asososca, Nejapa, Tiscapa, Xiloa, y Apoyeque.

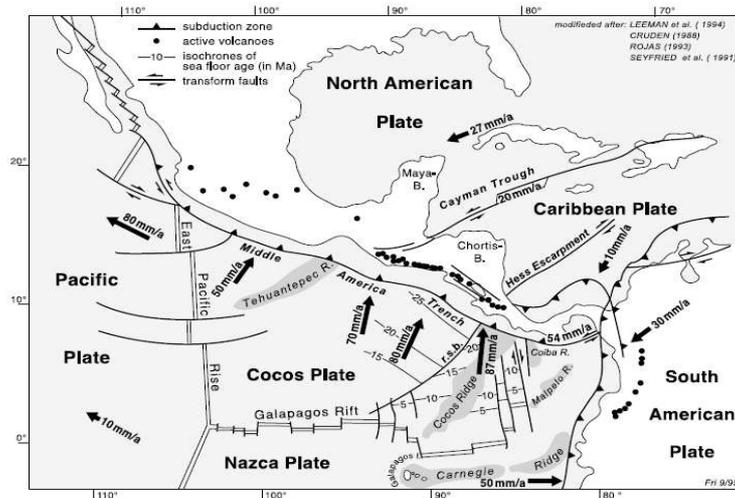


Fig. B.2. Mapa geológico regional de Nicaragua. Se muestra la cadena de volcanes activos (línea de puntos), la zona de subducción, así como el promedio de velocidades de deriva para posiciones diferentes. Fuente: Frischbutter 2002

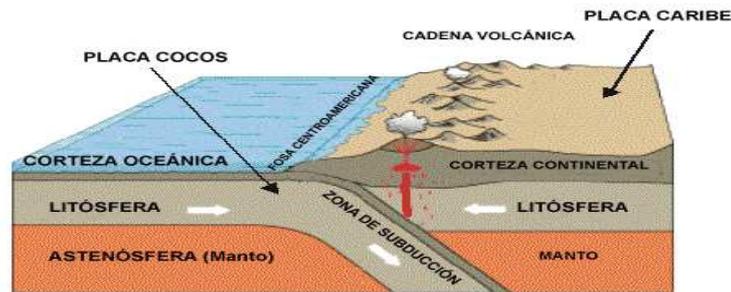


Fig. B.3. Interacción entre la placa de los Cocos y la Placa Caribe. Se muestra la zona de subducción que forma la cadena volcánica. Figura modificada del original de USGS



Fig. B.4. Huellas de Acahualinca sobre toba volcánica en Managua.

Fuente: Instituto nicaraguense de la Cultura.

Web: [http://www.inc.gob.ni/images/stories/Huellas%20de%20Acahualinca\(2\).pdf](http://www.inc.gob.ni/images/stories/Huellas%20de%20Acahualinca(2).pdf)

2.1 La Sismicidad en Centroamérica y Nicaragua

Centro América en una región con un alto grado de sismicidad presentando en la zona pacífica un elevado riesgo sísmico (Fig. B.5). Tanto riesgo como sismicidad se manifiesta en torno a dos zonas concretas; la zona de subducción entre la placa de los Cocos y la del Caribe y en la zona de la cadena volcánica (Ver figuras B.2 y B.5).

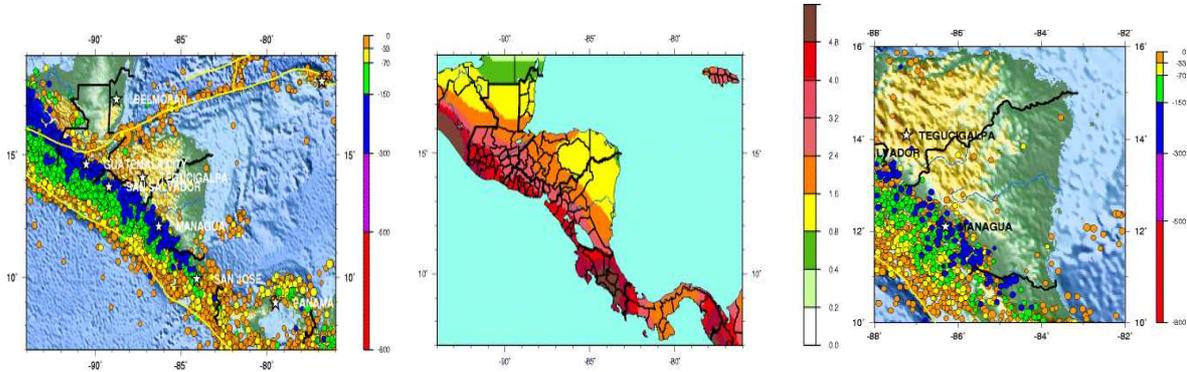


Fig. B.5 Mapa de Sismicidad y riesgo sísmico en Centro América y Nicaragua de 1990 a 2006

Fuente: National Earthquake Information Center, U.S. Geological Survey. Web: http://earthquake.usgs.gov/regional/world/central_america/seismicity.php

Los mapas de la figura anterior fueron elaborados con datos sísmicos de 1990 a 2006 (2006 es el año de la recopilación de datos para el presente estudio). Como se puede observar en las gráficas, la producción sísmica se centra en la zona costera del Pacífico, en donde confluyen las placas tectónicas y sobre la cadena volcánica que recorre la región, antes mencionadas. El peligro radica si se observa el mapa de riesgo sísmico (Fig. B.5. Centro), en que sobre esta misma zona en donde se encuentra asentada la mayor parte de la población. Este mismo precedente posee la población de Nicaragua que se concentra en su mayoría en la región del Pacífico. Esto debido a que desde tiempos remotos, las zonas agrícolas abonadas por los minerales de los volcanes son ocupadas por poblaciones. Por eso, el riesgo sísmico y la población van de la mano en Nicaragua. Esto se observa en el mapa elaborado por el INETER de la figura B.6, donde se determina con claridad la densidad de población y el riesgo sísmico normalizado.

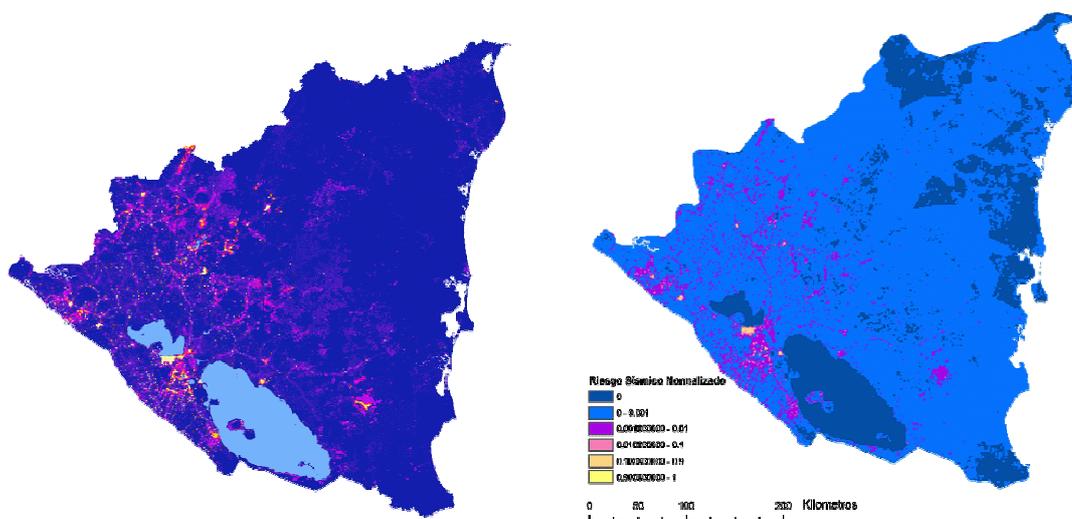


Fig. B.6. Mapas de densidad de población y Riesgo sísmico normalizado de Nicaragua (fuente INETER)

De acuerdo a los mapas anteriores, los mayores riesgos se encuentran situados en los núcleos de población en la zona Pacífico. Es en esta región del Pacífico de Nicaragua la que posee una extensa variedad de estructuras volcánicas, de ahí que sea prioritaria la vigilancia sísmico-volcánica. Dicha vigilancia es ejercida por el INETER, entidad encargada de la vigilancia sísmica y volcánica, a través del departamento de Geofísica en colaboración con varios proyectos internacionales. La figura B.7 muestra el mapa de las estaciones sísmicas (57 en total), instaladas por el INETER, a lo largo de la cadena de volcanes del Pacífico nicaragüense, durante las últimas décadas (1992-2010).

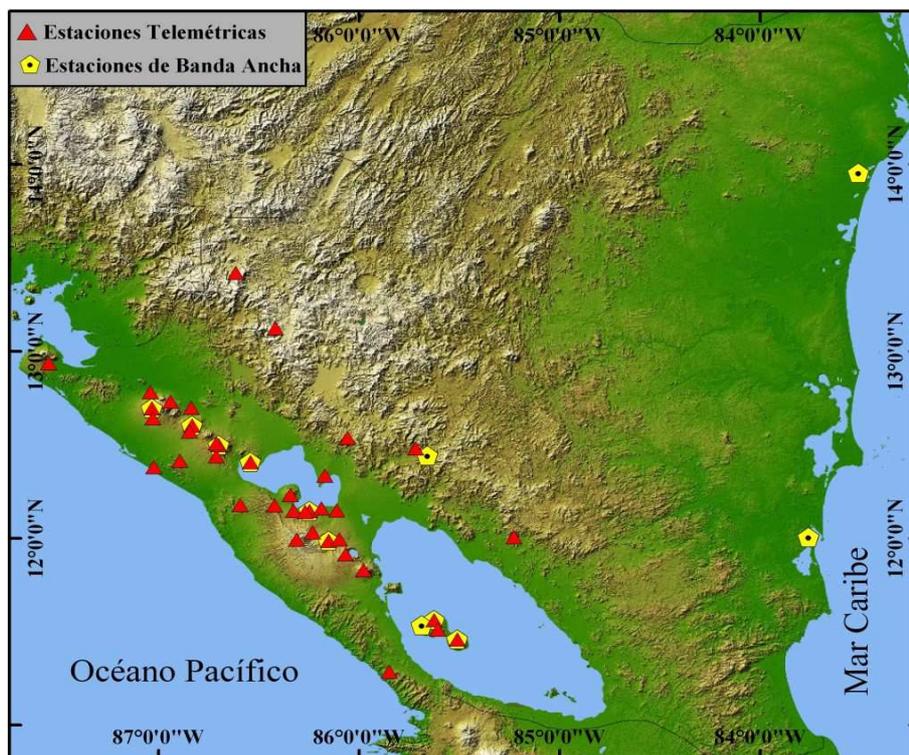


Fig. B.7 Mapa de las estaciones sísmicas instaladas por el INETER en la cadena volcánica de Nicaragua (fuente INETER)

Página Web: <http://webserver2.ineter.gob.ni/geofisica/sis/redsis.html>

Los datos emitidos por los volcanes son recopilados por sismógrafos (geófonos) ubicados en las cercanías de los volcanes y transmitidos mediante diversos medios (Radios, Internet), al observatorio situado en la capital del país para su posterior análisis.

3.- HISTORIA SÍSMICA-ERUPTIVA DE NICARAGUA

Rojas y Frischbutter en sus trabajos indican que los volcanes de Nicaragua se consideran parte de la provincia volcánica de Centroamérica, la cual se extiende desde el norte de Guatemala hasta Chiriquí en Panamá y se compone de más de 100 conos situados a lo largo de la costa del Pacífico. Aunque en el pasado, sólo 25 de ellos han estado activos, en Nicaragua se pueden observar al menos 18 edificios volcánicos. Sin embargo, existen muchos otros que se encuentran inactivos, en reposo o son conos en proceso de evolución. Toda la estructura volcánica de Nicaragua es el resultado de la

iteración de las placas que componen la plataforma nicaragüense [Rojas et. al. 1993], [Frischbutter 2002]. La fricción entre las placas y el calor que generan movimientos subterráneos, derriten porciones de la placa Cocos, de esta forma, los materiales volcánicos buscan una salida a la superficie formando los edificios volcánicos que se observan en la actualidad en la zona del Pacífico [Dames & Moore-Lamsa, 1978]. De acuerdo a diversos estudios, la distribución geológica de Nicaragua tuvo su origen en el Paleozoico y culminó con una intensa actividad a principios del Cuaternario [Costantini et. al. 2008]; [Giesecke 1989]; [Seyfried 1991]. En la zona central de Nicaragua se encuentran rocas volcánicas del Terciario y del Cretácico, que se distribuyen en al menos doce paleoarcos curvos volcánicos (magmáticos). Esto también incluye un desplazamiento en la actividad volcánica acontecido en el Terciario de este a oeste y una ligera relación entre las estructuras, especialmente de caldera y zonas de falla prominentes. *La disposición de los paleoarcos volcánicos coincide grandemente con la geomorfología del área central de Nicaragua* [Costantini et. al. 2008]; [Seyfried 1991]; [Valrey 1988]. El origen de los volcanes posiblemente ocurrió durante el pleistoceno, cuando se formó la cordillera de los Marribios, y que contiene los volcanes que se encuentran en ella, los cuales son conocidos como “*volcánicos del cuaternario*”. El nombre de esta cordillera se debe a un grupo aborígen, los llamados “Marribios”, que ocupaban la llanura cuando llegaron los conquistadores españoles, cuyos descendientes son el grupo denominado los Subtiavas.

De esta forma, en el territorio Nicaragüense, las estructuras volcánicas incluyen por lo menos 10 estratovolcanes (9 activos desde la conquista española), A pesar de no poseer un gran territorio, en Nicaragua se catalogan más de 200 estructuras volcánicas; entre domos, complejos maáricos, caldericos, estratovolcanes, conos de escoria, etc. De todas estas estructuras volcánicas, al menos 12 han mostrado históricamente alguna actividad significativa. Asimismo, diferentes estudios realizados describen, en la región de Centroamérica, grupos individuales de conos que están localizados a lo largo de fallas que van de norte a sur. [Alvarado et. al. 1986]; [Dengo et. al. 1970]; [Simkin y Lee, 1994].

En este sentido, algunas de las calderas volcánicas que se encuentran en Nicaragua tienen grandes dimensiones, como es la caldera del volcán Masaya, en la que destacan las fallas verticales, que van desde los 0 a los 250 metros en relieve, y cuyas dimensiones son aproximadamente de 12 por 5 kilómetros [Connor y Williams 1989]. Las erupciones que se produjeron en la caldera Masaya hace 2,250 - 6,500 años, cubrieron con nubes piroclásticas un área de 1,800 Km². Siendo una extensión mucho mayor que las producidas por el volcán Krakatoa, en 1928 ó el volcán Atitlán en Guatemala hace 40,000 años. Desde la llegada de los españoles que lo describieron en 1,524, se han producido 19 erupciones importantes. Algunos estudios han revelado que en dicha caldera, han podido distinguirse cinco principales depósitos piroclásticos del Holoceno [Kutterolf et. al. 2007]; [Pérez et. al. 2006]; [Walker et. al., 1993]. De hecho, en Centro América, desde tiempos históricos los volcanes presentan una gran historia de eventos eruptivos, por lo que el arco pliniano volcánico de Centro América, ha dejado amplios sedimentos de tephas, derivados de al menos 26 grandes erupciones, depositadas desde el pleistoceno en el fondo marino. La costa centroamericana contiene al menos 213 capas de cenizas. De dichas cenizas, Kutterolf en su estudio identificó capas de 11 erupciones de volcanes de Nicaragua. [Kutterolf et. al. 2008]. Los mecanismos eruptivos presentes en el arco volcánico centro americano y más concretamente en la cadena volcánica de Nicaragua, varían ampliamente en magnitud y

duración, no solo de un volcán a otro, sino también en el mismo volcán, por ejemplo, un caso típico es la misma caldera Masaya, que en su formación presenta variaciones, desde grandes erupciones freatoplínicas, basálticapliniana, subpliniana, freatomagmáticas a coladas lávicas [Pérez et. al. 2006]; [Walker et. al. 1993]. La cadena volcánica de Nicaragua (Fig. B.8), ha producido casi todos los mecanismos y productos volcánicos conocidos en el mundo, se pueden mencionar entre otros los siguientes:

- Lava de baja viscosidad (Santiago, Nindirí, Masaya, Cerro Negro).
- Lava que fluye fácilmente y puede esparcirse sobre grandes áreas (Masaya).
- Lavas de alta viscosidad que se superponen unas sobre otras para formar domos o flujos cortos y densos (Santiago Cristóbal, Mombacho, Maderas etc.).
- Además de esta alta viscosidad, los volcanes están sometidos en su interior a grandes presiones debido a la presencia de gas, de tal manera que cuando se libera, el gas se expande de manera muy explosiva y arrastra grandes cantidades de lavas (fundidas o sólidas) que se encuentran en suspensión.

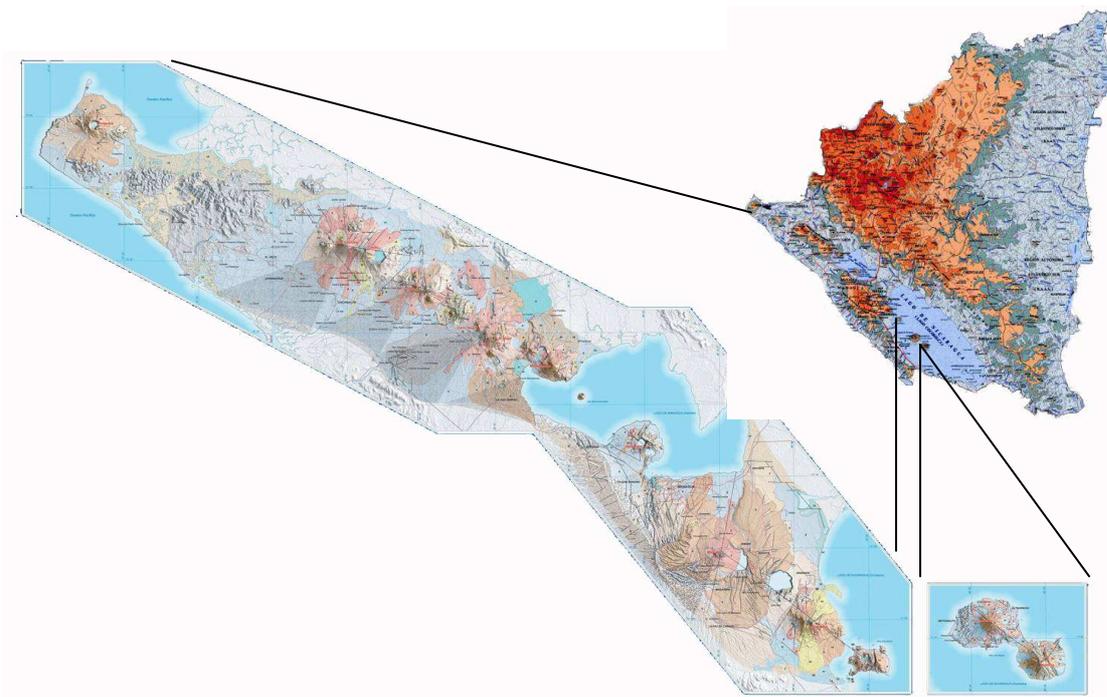


Fig. B.8 Cadena volcánica cuaternaria en la zona Pacífico de Nicaragua (Fuente INETER)

Como se observa en la figura anterior, los volcanes de la cadena volcánica del cuaternario se encuentran alineados en torno a la línea de la zona del Pacífico, y a su vez conforman cinco principales grupos o complejos volcánicos, separados por angostos pasajes, salvo el Momotombo que se levanta a orillas del lago de Managua, el Zapatera, ubicado en la isla del mismo nombre y los volcanes Concepción y Maderas que se encuentran en la isla de Ometepe, tanto la isla Zapatera como la de Ometepe están en el lago Cocibolca o lago de Nicaragua. Los cinco grupos o complejos volcánicos son los siguientes:

1. El grupo San Cristóbal es el más occidental y comprende los conos de El Chonco (1,105 metros), San Cristóbal (1,745 metros), Cerro Moyotepe, Casita (1,405 metros) y La Pelona
2. El grupo Telica abarca los volcanes de Los Portillos, Cerro Agüero, Telica, El Listón, Santa Clara y tiene a sus pies un campo termal (las fumarolas del volcán San Jacinto).
3. El Volcán Orata es un cerro aislado muy gastado por la erosión; su actividad cesó desde hace muchos milenios.
4. El grupo Las Pilas comprende el Cerro Negro y conos adventicios, el cerro La Mula, Cerro Grande, Cabeza de Vaca, Cerro El Hoyo y el cono de Ajusco, en cuyos pies se abre la laguna cráter llamada del Tigre, o Asososca de León.
5. El grupo Momotombo abarca el Cerro Montoso, la caldera de Monte Galán, el Volcán Momotombo y el cerro La Guatusa.

Hacia la costa Caribe, el único volcán registrado es el volcán Azul, y que en la actualidad se encuentra inactivo o en reposo. Desde el siglo XIX, en la cadena volcánica se han presentado algunas erupciones consideradas violentas, ejemplo de estas es la ocurrida en el Cosigüina el 22 de Enero de 1835, en esta ocasión, la erupción abarcó una gran área de Centro América y México, cubriendo extensas zonas con cenizas y piedra pómez. La erupción principal depositó cenizas a sitios tan lejanos como en la ciudad de México hacia el norte y Jamaica hacia el este. La nube de cenizas bloqueó el sol en un radio de 150 Km. [Mooser et. al. 1958]; [Simkin y Lee 1994]. En los volcanes de Nicaragua, a lo largo de su historia eruptiva se han observado la mayoría de los principales tipos de erupción volcánica. De acuerdo al INETER, la lista de dichas erupciones así como de los volcanes que las han producido es la siguiente:

- **Plinianas:** Apoyo, Apoyeque, Caldera Masaya, Concepción y Cosigüina;
- **Hawaiana:** Masaya;
- **Strombolianas:** Cerro Negro, Telica, San Cristóbal;
- **Freatomagmáticas:** Masaya,
- **Lineamiento** Nejapa/ Miraflores, Telica;
- **Dirigidas:** Mombacho y Momotombo.

En la lista anterior, se observa que algunos volcanes presentan diferentes tipos de erupciones, ocurridas en diferentes épocas, lo que produjo diferentes estructuras que han llegado hasta la época actual. Estas erupciones han producido gran variedad de cráteres y estructuras, Lilljequist en un estudio de imágenes por satélite, encontró un número considerable de estructuras circulares con cierta relación de alineamiento. En dicho estudio fueron definidas aproximadamente unas 572, que variaban en diámetro desde 100 metros hasta 52 Km. Como ya se ha mencionado, lo más probable es que la mayoría, sean el resultado o el producto de una intensa actividad volcánica en el terciario. *“Alineadas en arcos que parecen migrar de Este a Oeste, estas calderas marcan la posición de viejos alineamientos de la cadena volcánica nicaragüense, alguna vez ubicados muy al este de su posición moderna”*. [Lilljequist et. al. 1987]. Este agrupamiento de antiguos cráteres ha permitido poder reconocer erupciones de épocas remotas del terciario, en al menos tres grandes períodos de gran actividad volcánica. El primer período se establece durante el Oligoceno, que produjo la aparición de los terrenos de las Tierras Altas del Norte o grupo Matagalpa; el segundo período de actividad se estima que tuvo lugar en el Mioceno, y que produjo depósitos de mapas

basálticos a lo largo de la costa del Pacífico, en lo que se le ha nombrado como “*Arco volcánico Coyol*”; y el tercer y último período se establece durante el Plioceno-Pleistoceno, dicho período produjo la migración del arco volcánico en dirección Suroeste hacia el alineamiento donde se encuentran actualmente los volcanes activos de Nicaragua. [Dames y Moore 1978]; [Sten.-Åke 1998]; [Sten.-Åke et. al. 2001]. Durante todo el período del terciario, la actividad volcánica rellenó suelos y formó grandes porciones de territorio. Hayes a finales del siglo XIX en estudios geológicos, constató que aún en la zona del Pacífico, terrenos como los de la formación de Brito en el sureño departamento de Rivas, fueron probablemente creados por erupciones de volcanes del Oligoceno, hoy inactivos, entonces situados hacia el Oeste [Hayes 1899]. Durante el período cuaternario, la actividad volcánica siguió siendo intensa durante su fase temprana. Se supone que los depósitos de piedra pómez, arenas y cenizas presentes en los terrenos del Pacífico fueron arrojados por los volcanes modernos en grandes masas casi desde el principio de su actividad. Ya en el siglo XX, Sapper, indicó que se han identificado calderas de grandes dimensiones, que en períodos remotos estaban muy activas, cerca del sitio denominado “el Sauce”, municipio del departamento de León en la zona del Pacífico de Nicaragua. Incluso la misma caldera del Masaya, pertenece a un cono ya extinto, que tuvo en el pasado, gigantescas erupciones y cuyos efectos, aun en la actualidad son apreciables en toda el área de las Sierras de Managua [Sapper 1925]. Gran parte de la Depresión Nicaragüense fue rellenada por material de conos volcánicos del grupo denominado “*Las Sierras*” [Darce y Kondakov 1993]; [Sapper 1925]. Sapper supone que este grupo estuvo muy activo a principios del cuaternario, depositando enormes masas de tobas de un espesor de 400 metros en las sierras de Managua y en Masaya. Otros investigadores como Darce, creen que estos depósitos de piedra cantera, provienen de un gigantesco lahar o flujo de lodo de origen todavía desconocido [Darce 2002]. Kutterolf incluso afirma que del Pleistoceno al Holoceno la acumulación sucesiva de capas de tephra en la región central-occidental de Nicaragua se dio por la acción explosiva de tres complejos volcánicos: los complejos de chiltepe, Masaya y Apoyo y que alcanzó un volumen de 37 Km^3 . [Kutterolf et. al. 2007]; [Kutterolf et. al. 2008]. En esta línea, han sido documentadas erupciones de hace 23.000 años, una de estas ha sido la explosión que formó la caldera, que actualmente ocupa la laguna de Apoyo (Fig. B.9). Por su dimensión actual (6.5 Km. de diámetro), se sugiere que la magnitud de la erupción puede haber sido una de las mayores a escala mundial en ese tiempo. Los flujos de pómez depositados por el Apoyo, alcanzaron en áreas próximas, espesores de 37 metros y a distancias mayores alcanzaron a sepultar con una capa de 15 metros de espesor, el terreno donde hoy se asienta la ciudad de Granada. El proceso de colapso todavía continúa aunque muy lentamente.

Procesos eruptivos similares han moldeado lo que actualmente es la caldera Masaya, que con sus 46 Km^2 . es la más grande del país (Fig. B.10), donde a lo largo de casi 40 siglos, las diversas erupciones han creado varios cráteres entre los que se destacan: San Fernando, Masaya, Sapper, Oviedo, San Pedro, Nindirí y Santiago, siendo este último el que en la actualidad se encuentra activo. Cada uno de dichos cráteres en su momento eruptivo, han dejado tras de sí flujos de lava que han transformado la geografía de la zona. Una de dichas coladas de lava cubrió parcialmente la laguna Masaya, llegando a las cercanías de la ciudad del mismo nombre (Ver Fig. B.10). Por otra parte, explosiones del complejo Nejapa-Miraflores cubrieron un área de 5 a 5.5 Km. con una capa de Tephra con espesores desde 50 cm. a 1 metro de grosor [Cameron 2001]; [Hradecky et. al. 1997].



Fig. B.9 Caldera y laguna de Apoyo con 6.5 Km., de diámetro. Vista desde el mirador de Catarina. Al fondo el lago Cocibolca o Nicaragua. Foto: Ligdamis A. Gutiérrez



Fig. B.10 Caldera Masaya vista desde el espacio (izq.), conjunto de cráteres de la caldera (der.).
(Fuente Google Earth e INETER)

Otra gran explosión ocurrida por la misma época fue la del Apoyeque. Esta erupción que formó el cráter y la laguna del mismo nombre (una de las seis que rodean a la ciudad de Managua), cubrió el volcán con una gruesa capa de pómez y flujos de escorias que se extendieron hacia el occidente de Managua.

El resultado de la actividad reciente de la cadena volcánica cuaternaria puede verse en los cráteres actuales que forman las lagunas de los volcanes históricos. Se han producido hundimientos en fallas sísmicas, que dejaron los volcanes que hace miles de años estuvieron activos [Incer 1993]. Este es el caso de las lagunas de Nejapa, Asososca, Acahualinca, Jiloá (Xiloá), Tiscapa, así como los cráteres de Ticomo,

Refinería, Miraflores y de los pequeños conos de Las Jinotepes, Motastepe, Javier, San Judas, Embajada, Chico Pelón y Veracruz [Bice 1985]; [Incer 1998]. Se atribuye a una única explosión, que dejó por una parte capas de ceniza con un espesor de 1 a 3 metros y que formó las huellas de Acahualinca, y que por otro lado formó la caldera de Masaya, hace aproximadamente 6,500 años. Las huellas de Acahualinca (Ver Fig. B.4) es uno de los registros más vivos de la actividad eruptiva de la antigüedad que ha llegado hasta el presente. En esta misma línea, en Octubre de 1919, fueron encontrados en la península de Chiltepe antiguos restos humanos con signos de haber sido carbonizados durante una erupción [Morales y Strauch 1999].

Cerca del nivel donde quedaron impresas las huellas de Acahualinca, se encuentra una capa de pómez depositada por una explosión posterior que se cree fue producida por el volcán Apoyeque (hoy laguna de Apoyeque). La última erupción de este volcán fue hace 3,500 años. De acuerdo a Incer, la espesa capa que entonces depositó sobre la tierra de la ciudad de Mateare, pudo haber sido la causa de un lahar que se dio en el flanco sur oeste de la sierra y que sepultó durante la estación lluviosa el sitio conocido por los antiguos pobladores como Taguistepe que significa “pueblo enterrado” [Incer 1998]. Desde esa fecha, la actividad sísmico-volcánica ha tendido a disminuir.

Todo esto lleva a Squier en su trabajo a deducir que el frente volcánico nicaragüense es muy activo, afirmaba que “*No hay en todo el continente americano, ni quizás en el mundo entero una extensión de tierra que como ésta contenga tantos volcanes en actividad o apagados, ni que presente tantos vestigios de convulsiones plutónicas... se ven en Nicaragua catorce volcanes de un solo vistazo*” [Squier 1973]. En cuanto a las crónicas de los eventos volcánicos en el pasado, destaca la expuesta por los aborígenes y descrita por De Oviedo sobre la erupción del 1,500 en el Masaya [De Oviedo 1,851]. En total, de los volcanes activos de Nicaragua se han evidenciado cerca de 190 erupciones a partir de 1,520, en esta lista documentada tanto por el INETER como por McBirney y Williams se encuentran: El Masaya con 36 y el Momotombo con 27, el Telica con 58 episodios, de los cuales 51 ocurrieron durante el siglo XX, El Concepción registra al menos 22 erupciones. El Cerro Negro en 18 ocasiones y el San Cristóbal, 14. Con menos eventos eruptivos se encuentran, el Hoyo con 5 y el Cosigüina con 4. Los volcanes Pilas, Rota, Santa Clara y Cerro Montoso registran solamente una erupción [INETER 2001]; [McBirney y Williams 1965].

No toda la actividad de los volcanes es negativa, ya que la extensa actividad eruptiva ha jugado un papel benéfico para la formación de Nicaragua. De hecho, casi la totalidad de las tierras del país, partiendo de las bases formadas por rocas del fondo oceánico, han sido rellenadas sucesivamente por capas superpuestas de material volcánico. La costa del Pacífico, en épocas anteriores a los volcanes modernos se sugiere que estaba ubicada más hacia el Este y una ensenada grande del océano fue separada del mismo por depósitos volcánicos formando los lagos. Erupciones posteriores separaron los dos lagos y formaron sus penínsulas e islas [Hayes 1899]. La llanura del Atlántico sumergida bajo el océano en el pasado remoto, fue emergiendo y ganando terreno como producto de materiales sedimentarios volcánicos terciarios arrastrados por los ríos.

Otro hecho ha sido que la misma altura de los volcanes ha diferenciado la climatología y trasladado el límite entre el clima húmedo y seco hacia la Costa del

Pacífico. Los volcanes también han influido en la diversidad y la distribución de plantas y animales en el territorio. Los suelos se han beneficiado por las emisiones de cenizas ricas en nutrientes estableciendo además planicies, cuencas y lagunas propicias para asentamientos humanos. El aspecto negativo lo continúa ejerciendo, la amenaza latente que representan los volcanes para las poblaciones asentadas en sus márgenes y alrededores. Las amenazas actuales de la cadena volcánica del pacífico de Nicaragua se observan en el mapa elaborado por el INETER (figura B.11). Como ya se ha mencionado, la región del Pacífico concentra más del 80% de la población del país.

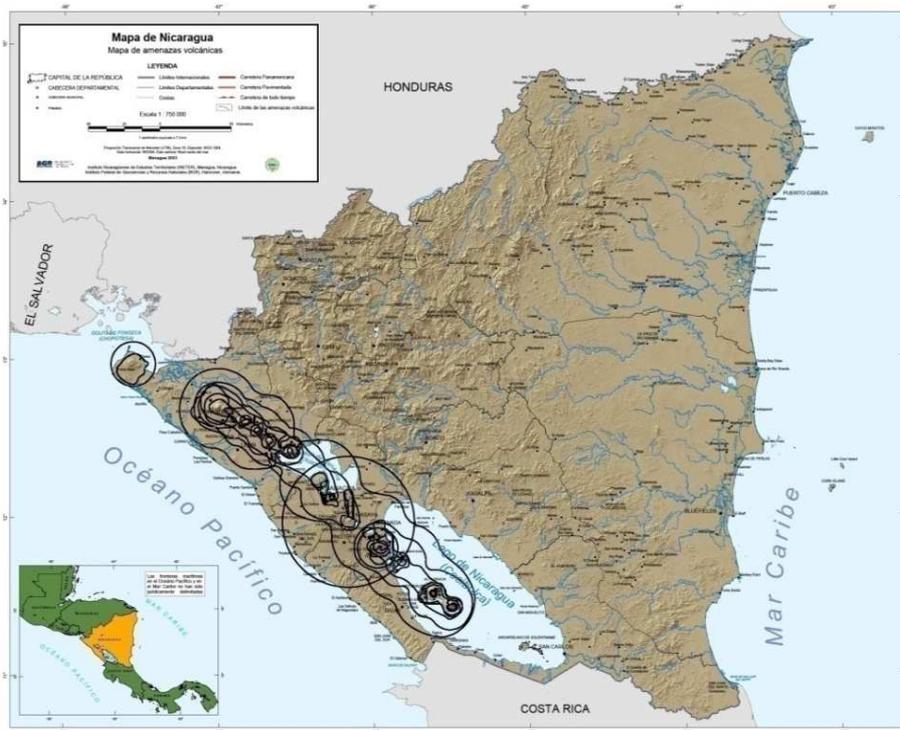


Fig. B.11 .Zonas de amenazas de la cadena volcánica de Nicaragua (Fuente INETER)

4- ACTIVIDAD SÍSMICA EN LA CADENA VOLCÁNICA DE NICARAGUA

Durante el período de recopilación de los datos (Febrero a Marzo) en el 2006, el INETER⁴ reportó el registro de 1,974 sismos. Aunque este número fue menor que el del año anterior (Fig. B.12 derecha), se mantuvo básicamente en el promedio de los últimos cuatro años. De la totalidad de sismos 1,517 ocurrieron en Nicaragua, 318 en el resto de Centro América, 35 fueron localizados distantes y 104 no se pudieron localizar. La cadena volcánica de Nicaragua, mantuvo la tendencia a producir la mayoría de los sismos superficiales. Esto se puede apreciar en las gráficas que elaboró el INETER sobre la distribución espacial. Los sismos que se establecieron en un rango de 0 a 30 Km. de profundidad correspondieron a los mecanismos originados por la cadena volcánica (Fig. B.12 izquierda).

⁴ Los datos y las gráficas del San Cristóbal y Telica que aparecen en este apéndice, han sido tomados de los boletines mensuales (Febrero y Marzo) y anuales (2006) que el INETER elabora.

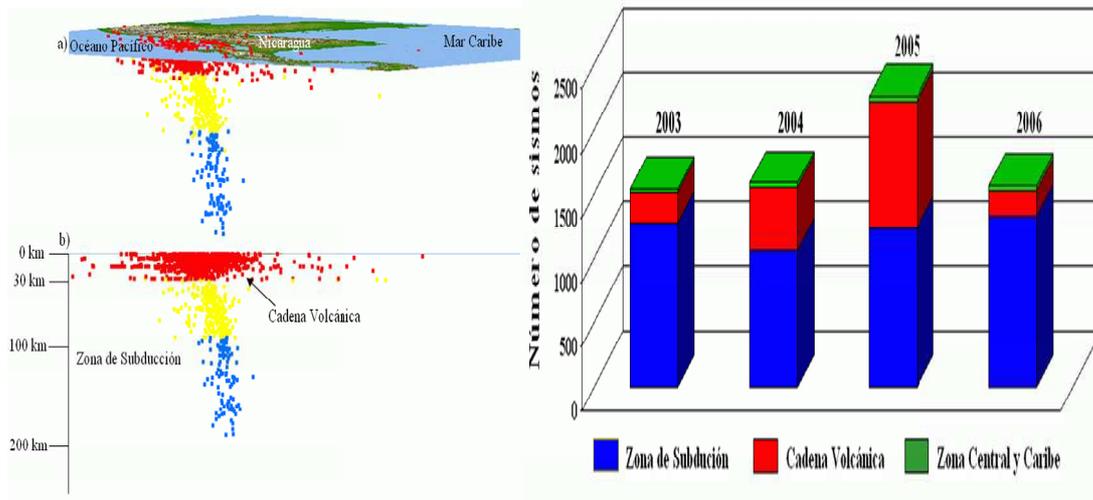


Fig. B.12 .Izquierda, Distribución espacial de la producción de sismos durante 2006 en Nicaragua. a) Vista oblicua b) Corte perpendicular a la zona de subducción. Derecha, Número de sismos de 2003 a 2006 (Fuente INETER)

En cuanto a zonas de ocurrencia, el porcentaje de la totalidad de eventos ocurridos durante 2006 correspondieron en un 13% a la cadena volcánica. La mayoría con un 85% fueron producto de la zona de subducción de las placas interoceánicas y el restante 2% ocurrió en la región norte y atlántica (norte y sur) del país (Fig. B.13).

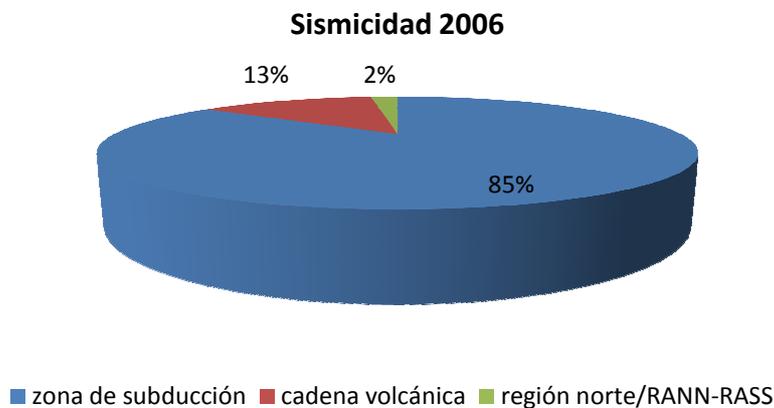


Fig. B.13 Porcentaje de sismos en las diferentes zonas durante 2006 en Nicaragua (Fuente INETER).

De acuerdo a la lista de sismos ocurridos durante 2006 el INETER elaboró un modelo de capas que se observa en la figura 13, en la que se muestra que el moho⁵ se establece en torno a 30 Km. de profundidad con una Vp de 8.10 Km./s (fig. B.14 izquierda). El rango de profundidad de la mayoría de los eventos en ese año se produjeron de 0 a 30 Km. (figura B.14 derecha).

⁵ Capa límite entre la corteza y el manto

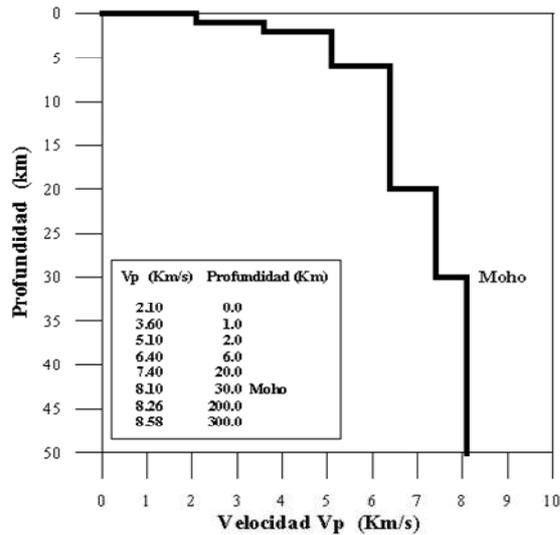


Fig. B.14 Modelo de capas de la Sismicidad durante 2006 en Nicaragua. (Fuente INETER)

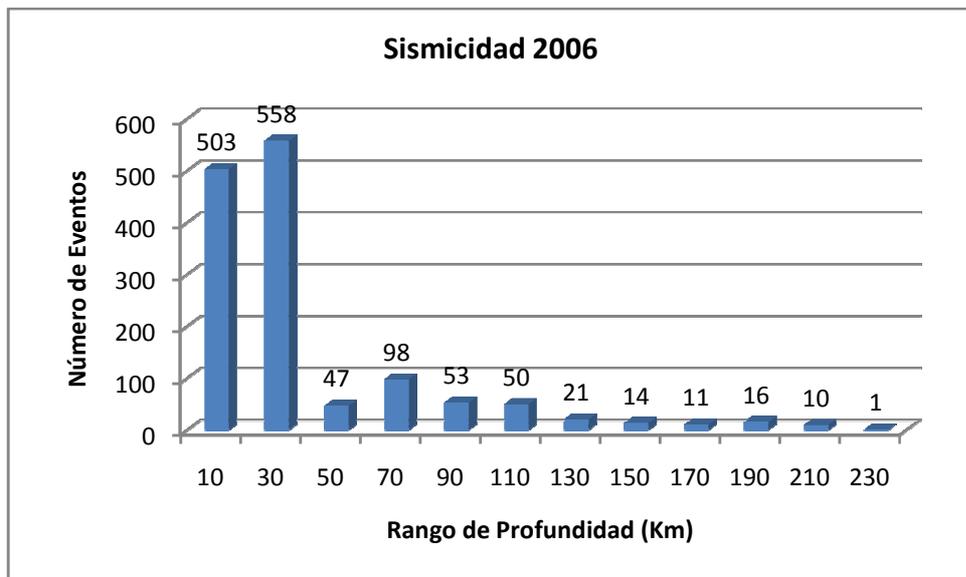


Fig. B.15 Rango de profundidad de la Sismicidad durante 2006 en Nicaragua (Fuente INETER)

Los sismos superficiales (0 a 30 Km.) se concentraron principalmente en dos zonas, claramente establecidas como son: la zona de la cadena volcánica y la zona de subducción. Los sismos intermedios (30 a 100 Km.) se manifestaron en la zona de subducción con mayor énfasis en la parte central. En cuanto a los sismos profundos (más de 100 Km.) estos se dieron en una franja estrecha a lo largo de la zona costera del pacífico (figura B.16).

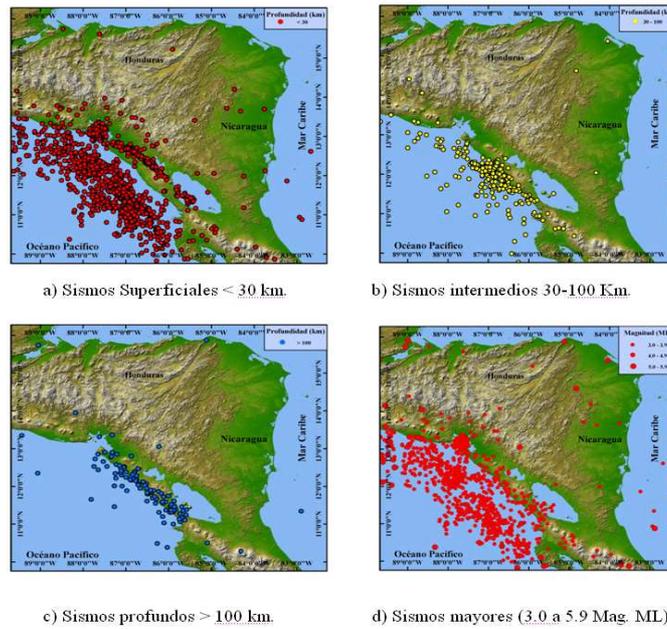


Fig. B.16 Localización en 2006 de los sísmos superficiales (a), Intermedios (b), profundos (c) y los sísmos de acuerdo a su magnitud (ML) en Nicaragua. (Fuente INETER)

De acuerdo a lo anterior, los sísmos tanto en el área Centroamericana como en particular en Nicaragua abarcaron principalmente dos zonas que fueron la zona de subducción entre la iteración de la placa pacífica y Caribe y la zona de la cadena volcánica concentrándose en su mayoría en Nicaragua (fig. B.17). La cadena volcánica es un factor importante a tener en cuenta en cuanto a la producción de sísmos de menor profundidad.

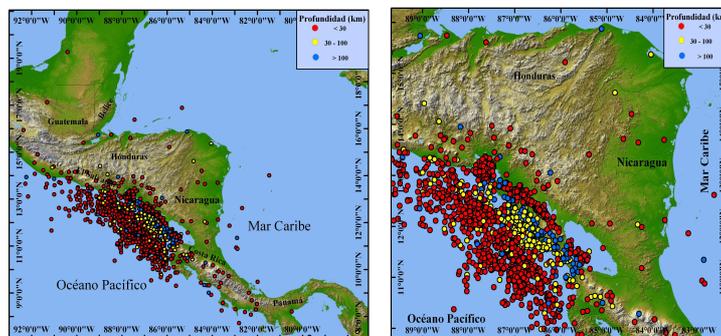


Fig. B.17 Localización del número de eventos producidos en profundidad durante 2006 en Centro América (izquierda) y en Nicaragua (derecha) (Fuente INETER)

Esto también se corrobora con estudios realizados por el INETER en un margen mucho mayor, al hacer referencia al período que comprende desde 1975 al 2000. Dicho estudio indica que la sismicidad superficial en Nicaragua con una profundidad menor de 30 Km., y que se concentra en una estrecha banda de 20 Km. de diámetro (figura B.18). Dicha banda coincide con la cadena volcánica de Nicaragua, por lo que se puede suponer que tanto sísmos como volcanes tienen una causa común [INETER 2004]; [Segura et al., 1999]; [Strauch 2000].

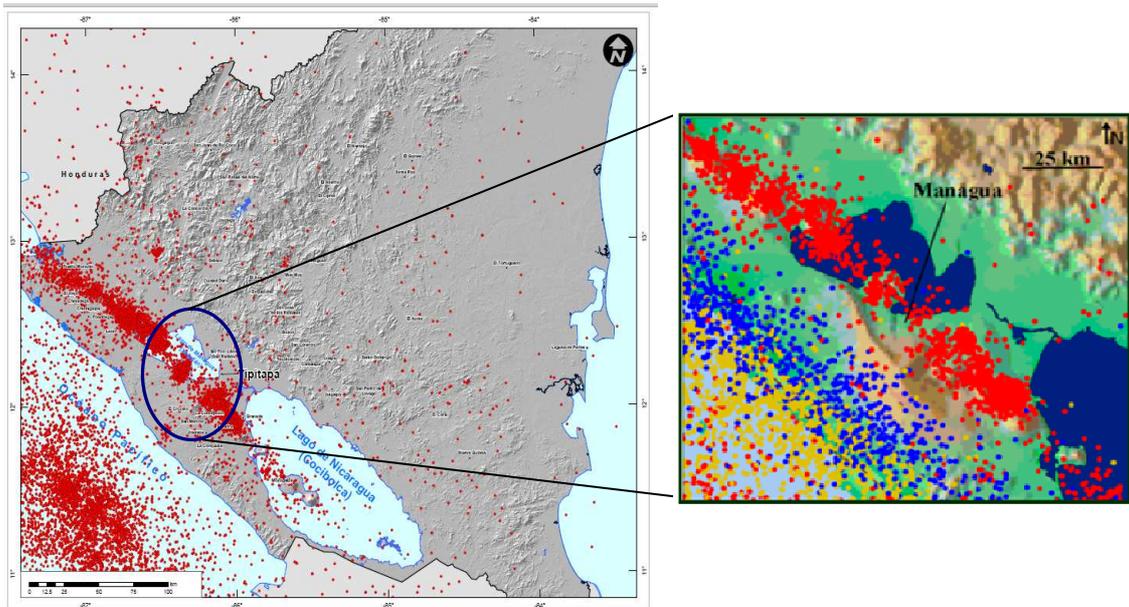


Fig. B.18 Mapa de sismicidad superficial en Nicaragua período 1975-2000. En el círculo, la ubicación de la capital, Managua (derecha). (Fuente: INETER)

En el círculo que se observa en la figura anterior se presenta un acercamiento sobre la capital Managua (Fig. B.18 derecha), que muestra los sismos superficiales (puntos rojos) localizados por el INETER en las últimas décadas, dichos sismos se concentran en la banda antes mencionada y a lo largo de la misma ubicación de los volcanes. Por lo que se deduce que el origen en su mayoría es probablemente sísmico-volcánico más que tectónico. Managua y varias de las principales ciudades como Granada, Masaya, León y Chinandega se ubican directamente en el eje de esta banda. Managua al ser la más poblada (1,500,000 hab.) es la ciudad más peligrosa en Nicaragua. De ahí la importancia de poder analizar y deducir tanto los tipos de eventos que los volcanes producen. La actividad de los volcanes ha producido un fallamiento local, lo que genera en gran variedad de sismos de origen tectónico. Algunos de ellos de magnitud considerable, como los producidos en Managua en 1931 y 1972 derivaron en la destrucción parcial de la ciudad, causando más de 10,000 víctimas. En cuanto a los sismos de mayor profundidad (puntos amarillos y azules), estos se mencionaban que se concentran en torno al área costera del Pacífico en la zona de subducción.

Prácticamente la totalidad de ellos son producidos por la acción de la interacción de las dos placas que chocan, donde la placa Cocos se hunde por debajo de la placa Caribe, en la zona de subducción. Sin embargo, los sismos de mayor profundidad son menos peligrosos que los superficiales y los intermedios debido a que la distancia espacial supera los 100 km. (Ver figura B.12) [Strauch 2000].

En cuanto a los sismos producidos por la cadena volcánica, el principal peligro no lo representan los sismos en sí mismos, sino los diferentes productos que emanan de los volcanes, en las constantes actividades que se producen en la zona y que podrían ocasionar daños a las poblaciones asentadas en sus alrededores. Por otra parte, al observar la profundidad de la mayoría de ellos, puede deducirse que la fuente de origen de dichos sismos no se encuentra a mucha profundidad. Habría que realizar posteriores análisis y estudios para poder confirmar dicho planteamiento y poder ubicar con exactitud el origen de los mismos. En los análisis periódicos registrados en el 2006 por

el INETER, de los sismos que produce la cadena volcánica, se observa que la mayoría de ellos, son de poca magnitud de 1° a 3.9° (Fig. B.19 izq.), concentrándose en una profundidad de 5 a 10 Km. (Fig. B.19 derecha). Salvo que ocurran eventos explosivos mayores, los volcanes en relación a los sismos no presentan mucho peligro a la población o las construcciones en los alrededores.

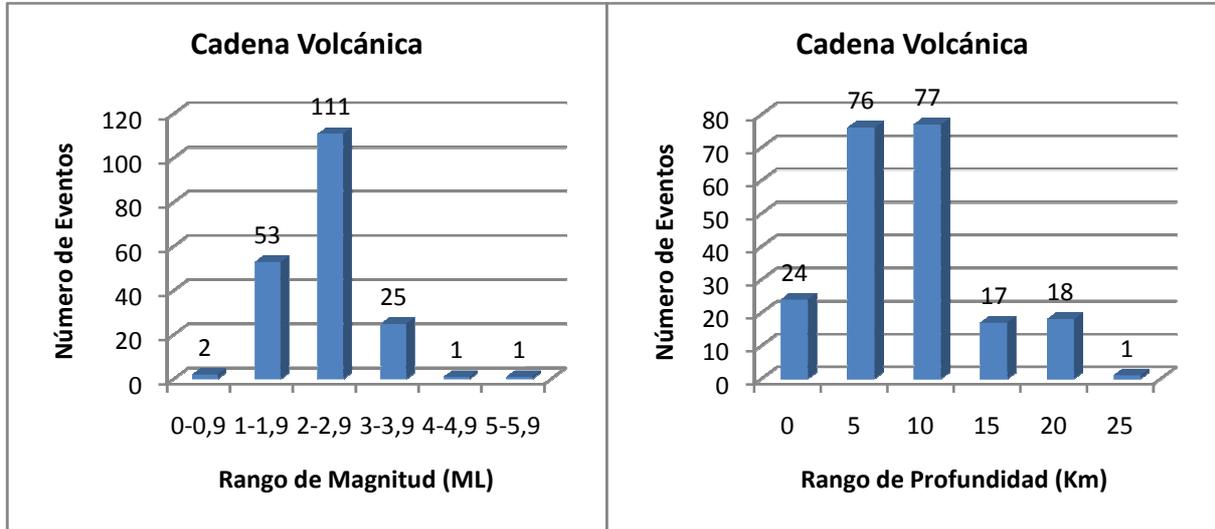


Fig. B.19 Número de eventos producidos por la cadena volcánica de Nicaragua en rango de magnitud y profundidad durante 2006 (Fuente INETER)

La anterior información representa como se está comportando la cadena volcánica durante el período que se realizaron las campañas de recopilación de la información. En general se observa que la mayoría de los sismos de la cadena volcánica durante este período, son superficiales y de poca magnitud, ya que se encuentran en el rango de magnitud de 1 a 3 grados y a una profundidad de 5 a 10 kilómetros. Esto puede indicar que en los volcanes, los eventos se producen en torno al edificio central, y determina que al ubicarse las estaciones en las cercanías de los cráteres, los datos que se recopilen de ellas serán apropiados y fiables. En el siguiente apéndice, se detallará el comportamiento de los volcanes durante el período de recopilación de la información.



APÉNDICE C: Historia eruptiva y actividad sísmico-volcánica durante el período de recopilación de datos

La actividad sísmica relacionada con la base de datos realizada en la presente Tesis está recogida en un intervalo temporal, limitado en cada volcán, por tanto dicha actividad sísmica es un reflejo de la actividad volcánica que en ese momento estaba aconteciendo. Es evidente que un cambio en las características volcánicas podría llevar asociado un cambio en el tipo de señal sísmica. Este apéndice sirve para contextualizar las señales sísmicas con su dinámica volcánica, por lo que se incluye una recopilación de información sobre la historia eruptiva de los volcanes, así como de la actividad sísmico-volcánica, presente en los cuatro volcanes de estudio, durante el período que se realizaron las campañas de recopilación de los datos.

1.- Actividad sísmica-eruptiva en el Etna

Después de la llamada “más larga erupción lateral”, ocurrida en los últimos tres siglos en el Etna de 1991 a 1993 [Calvari et. al. 1994], la mayor parte de la actividad volcánica en el Etna, estuvo limitada más a la parte superior del volcán. La cima de Etna comprende un cono principal compuesto de dos cráteres principales Bocca Nuova (Bn) y Vorágine (Vr) y dos conos parasitarios cráter NE (nor-este) y cráter SE (sur este), localizados respectivamente al norte y sur del borde este del cono principal (Figura C1), [Corsaro and Pompilio 2004].

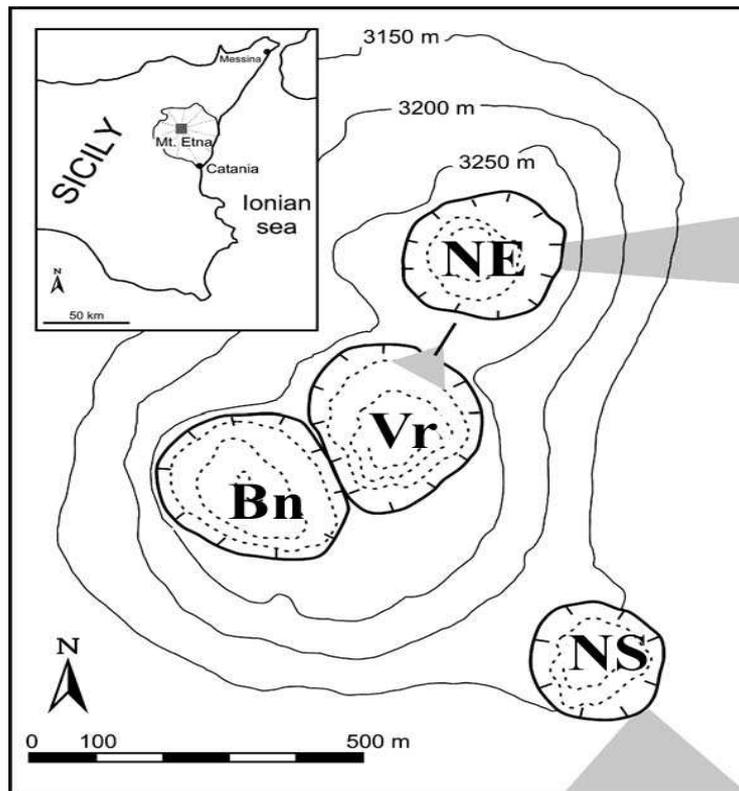


Fig. C.1 Mapa de la ubicación de los cráteres en la cima del Etna a Enero de 1999. Las áreas en gris representan las coladas de lava producidas. Fuente [Corsaro and Pompilio 2004]

Este período eruptivo del Etna de 1991 a 1993, se destaca como el más largo de los tres últimos siglos en la historia del volcán. Detalles extensos de la observación de este proceso fueron descritos por Calvari y asociados, en el trabajo publicado en 1994 [Calvari et. al. 1994]. Así como la gran erupción subterminal que comenzó en Febrero de 1999 [Calvari et. al. 2003], durante dicho período se realizó la campaña de la recopilación de los datos del presente trabajo.

Después de la erupción de 1991 a 1993, la actividad volcánica durante la segunda mitad de 1995 se sitúa en el cráter Bocca Nuova (Bn). Fueron observados fenómenos ya sea de forma directa como de forma remota a través de cámaras de video localizadas en el flanco sureste cercano a los 2,500 m. [Bertuccio et. al. 1999].

Se reportaron colapsos en el piso del cráter, mientras que comenzaba una progresiva actividad Stromboliana en el cráter Bn. En Noviembre de 1996 se observó una actividad Stromboliana en el sureste del cráter SE. También fue observada desde el verano de 1997 una actividad Stromboliana en el piso del cráter Voragine (Vr). Al final del verano de 1997 todos los cráteres manifestaron alguna actividad de distintas intensidades.

De Octubre de 1997 a Mayo de 1998 fueron observadas rápidas oscilaciones en la actividad de todos los cráteres. El cráter Ne finalizó dicha actividad con un episodio de una fuente de fuego. De Junio de 1998 a Febrero de 1999 se dio un incremento de la actividad en torno a tres de los cuatro cráteres. Explosiones Strombolianas y flujos en los intracrater se intensificaron progresivamente en los cráteres Bn y Vr y continuó siendo elevada la actividad después de un episodio de fuente de fuego ocurrida el 22 de Julio de 1998.

En Septiembre 15, después de una pausa de seis semanas, se presentó un nuevo episodio eruptivo con características diferentes en el cráter NS. Esto produjo 22 episodios de fuentes de fuego de baja a mediana intensidad que fueron observadas hasta Febrero del 1999. Durante el último período de las fuente de fuego se formó un corte de una fisura en el flanco del cráter SE la cual produjo una cortina de fuego. Esta fisura alimentó varios flujos que cubrieron la parte occidental del muro superior del Valle del Bove [Calvari et. al. 2003].

La actividad en el Etna de 1995 a 1999 produjo bastantes productos volcánicos variados; tales como flujos de lavas, bombas, caída de cenizas entre otros. Los episodios violentos de fuentes de fuego causaron caída o lluvia de Tephra que cubrieron los flancos oriental y meridional. [Coltelli et. al. 1998]; [Coltelli et. al. 2000].

El evento del Etna en el que se centra este trabajo es el tremor volcánico. Anteriormente se ha mencionado que el tremor volcánico es uno de los más difíciles casos de estudio para la sismología. Una falta de claridad en la fase de la llegada de las ondas y una rápida pérdida de la señal de coherencia con el incremento del espaciamiento estacionario hacen imposible recuperar el origen mediante las técnicas de determinación del hipocentro manejado por la sismología clásica. A esto se suma la complejidad y heterogeneidad de las estructuras volcánicas, las cuales afectan las señales sísmico-volcánicas en gran medida, de manera que la separación entre la fuente y el sitio de recolección de datos representa un difícil esfuerzo. Sin embargo, a pesar de todo esto, la correspondencia entre tremor volcánico y procesos eruptivos seguirá

determinando los esfuerzos a analizar las fuentes que originan el temblor y las señales sísmicas adicionales que se producen.

2.- Actividad sísmica-eruptiva en el Strómboli

Desde cerca de 1,800 años la actividad del volcán Strómboli se ha caracterizado por las denominadas explosiones “Strombolianas”, las cuales tienen lugar desde varios respiraderos a intervalos variables. Las características de dichas emanaciones explosivas son: efusión scoriaceous (comúnmente denominadas bombas), con clastos o material juveniles (procedentes de la cámara magmática) de escoria negra, con rocas de naturaleza shoshonítica⁶ con alto contenido en potasio (high-K) de composición basáltica. El Strómboli se caracteriza por tener una numerosa y gran actividad sísmica con una larga variedad de señales. Langer y Falsaperla en 2003, documentaron un gran número de terremotos explosivos (más de 200 por día), pocas secuencias explosivas en promedio de dos por cada año, extraños terremotos de corteza o superficiales y una persistente radiación sísmica mejor conocida como temblor volcánico [Langer y Falsaperla 2003]. El temblor ha sido asociado a determinados fenómenos volcánicos en la superficie, y que se han dado por largos períodos de tiempo como las actividades de 1990 a 1998 [Langer y Falsaperla 1996]; [Falsaperla et. al. 1998]; [Falsaperla y Langer 2000].

Así, el volcán Strómboli es el único volcán en el mundo en que aproximadamente cada 15 a 20 minutos tiene un evento eruptivo o explosión típicamente llamado “*explosión de tipo Estromboliano*” [Bertagnini et. al. 1999]; [Patrick et. al. 2007]; [Rosi et. al. 2000]. Esto hace posible monitorizar todo el sistema explosivo de principio a fin para de esta forma, poder realizar una investigación del proceso que precede, prepara, se desarrolla y finalmente explota. Cada explosión es de esta forma el inicio y el fin de una nueva observación para un nuevo experimento. De hecho, el Strómboli es conocido como el faro del Mediterráneo, por su frecuente actividad explosiva, que ha sido constante al menos durante los últimos 2,500 años [Rosi et. al. 2000]. Además, la historia eruptiva de Strómboli también se caracteriza por la ocurrencia de más hechos de violencia que exhiben explosividad mucho mayor con respecto a las explosiones normales. [Barberi et al 1993].

En los últimos 20 años se han hecho muchos esfuerzos para poder caracterizar las propiedades físicas de la fuente de las explosiones observadas en las señales sísmicas. [Chouet et. al. 1997]; [Del Pezzo et. al. 1974]; [Dietel et. al. 1994].

3.- Actividad sísmica-eruptiva en el San Cristóbal

En su historia eruptiva, una de las primeras informaciones sobre las erupciones del San Cristóbal, fueron documentadas por De Oviedo en las crónicas que en 1,520 observó sobre este volcán nicaragüense [De Oviedo 1,851]. En el informe presentado en 2,005 y generado por el gobierno de Nicaragua a través del Sistema Nacional para la Prevención, Mitigación y Atención de Desastres (SINAPRED), indica que a lo largo de su historia, en los períodos de actividad, el tipo de erupción que ha tenido el volcán San Cristóbal ha llegado a ser muy explosiva, emitiendo flujos de lava y piroclastos, que

⁶ Serie rica en K, solo aparece en zonas de subducción. En volcanes como; Absaroke (Wyoming) y Stromboli (Italia).

alcanzaron una altura de hasta 12 kilómetros. Lo que se ha comprobado mediante los depósitos que han sido descubiertos [SINAPRED 2005].

Salvo la erupción de 2,012 el volcán no había manifestado mucha actividad en los últimos años. Sin embargo, se han podido identificar al menos nueve erupciones importantes ocurridas en los últimos cuatro siglos, desde el tiempo de la conquista española, generalmente siendo el tipo de erupciones de Stromboliana a Sub-pliniana [Smithsonian Intitution, Bull, 1994]. Una de las erupciones más grandes ocurrió en el siglo XVI, en la que el volcán emitió flujos de lava que alcanzaron los 15 kilómetros [SINAPRED 2005].

El informe del SINAPRED de 2005 también revela, que el registro histórico de flujos de lava en el San Cristóbal se concentran en la parte norte y sur del edificio volcánico, alcanzando distancias desde el cráter cercanas a 20 Km. hacia el sur. El informe manifiesta que las erupciones desde 1,600 hasta la época actual, produjeron procesos freáticos que afectaron a la ciudad de Chinandega y sus alrededores [SINAPRED 2005].

En 1,685 el pirata Dampier que se encontraba cerca de las costas de Nicaragua, lo encontró en plena actividad, y le sirvió como un faro guía, para así poder alcanzar el puerto del Realejo y asaltar la ciudad de León [Dampier 1927]. El informe del SINAPRED del 2,005 y el INETER manifiestan que en 1,971 el volcán emitió cenizas durante casi un mes. Durante el año 1,992 volvió a emitir cenizas durante un tiempo de quince días. Entre 1,997 a 1,999 presentó una actividad que se prolongó durante dos años con períodos de calma intercalados [INETER 2001]; [SINAPRED 2005].

Otro período eruptivo de este volcán corresponde de fines de 1,999 a principios del año 2,000. De acuerdo a INETER, se considera que este volcán puede producir erupciones violentas de tipo plinianas, formando columnas compuestas de escoria, gases y ceniza que alcanzarían varios kilómetros de altitud [INETER 2001]; [SINAPRED 2005].

4.- Actividad sísmica-eruptiva en el Telica

La historia eruptiva del volcán se remonta desde 1,527 a la fecha. Ha sido documentada por el Smithsonian Intitution Bull en los años 1,989, 1,994, 1,998 y 1,999. De acuerdo a esta institución y al INETER, el volcán Telica posee características eruptivas similares a las del San Cristóbal, ya que ha producido erupciones de tipo estromboliano a subpliniano. Las erupciones de mayor importancia se han presentado en el 2,000 y es considerado una de las estructuras más activas dentro de la cadena volcánica cuaternaria nicaragüense.

En el informe presentado en 2,005 por el SINAPRED, se indica que durante el año 1,982, en el Telica se produjo una erupción que generó productos de tipo proyectiles balísticos (bombas), los que se acumularon en zonas cercanas al cráter. Además la erupción produjo columnas de cenizas y gases con una altura de 4.3 Km. Las cenizas de dicha erupción cubrieron zonas hacia el oeste del país, afectando a varios municipios de las ciudades de Chichigalpa, Posoltega, Quezalguaque y León. Depósitos de casi un metro de espesor de tipo escoria, producto de la acción de las bombas volcánicas alcanzaron zonas hasta una distancia de 7 Km del cráter [SINAPRED 2005].

El informe también indica que ya desde el año 1,859 se informaba de erupciones que presentaban columnas de gases y cenizas que afectaban a las plantaciones agrícolas ubicadas en las cercanías del volcán. Estos acontecimientos se repitieron en erupciones ocurridas en los años 1,937, 1,973 y 1,991. En el informe se indica que durante la historia eruptiva del volcán, los flujos de lava producida por las erupciones han alcanzado distancias considerables de hasta 10 Km desde el cráter, siendo los últimos depósitos fechados alrededor de 1,000 años atrás, llegando hasta los lindes de la carretera que conduce de la ciudad de León a la de Chinandega casi a los 20 Km. de distancia [SINAPRED 2005].

El informe del SINAPRED también manifiesta que en la historia eruptiva del volcán Telica se han producido de 1,527 a 1,982 doce erupciones de tipo freático, con salida de gases y emisión de importantes volúmenes de cenizas [SINAPRED 2005]. Aunque no existen registros históricos de los flujos de lava en el volcán, las explosiones del Telica, han producido expulsiones de tephra y flujos piroclásticos con algunas emanaciones de lava, que han llegado hasta las cercanías del municipio de Telica situado al sur del volcán. Se calcula que una erupción de características violentas en el Telica produciría espesores de ceniza de aproximadamente 1 centímetro en toda la costa del Pacífico nicaragüense [SINAPRED 2005]; [Albrecht et. al. 2005].

Paralelamente, el INETER elabora resúmenes de la actividad volcánica que publica mensualmente, en dichos resúmenes en el período de 1,984 a 1,999, destaca la erupción del 31 de Julio de 1,984, que produjo columnas de cenizas y gases a una altura de 800 metros y cuyas cenizas abarcaron un área de 17 km. En 1,994 desde el 31 de Julio, el volcán comenzó a tener manifestaciones explosivas con emanaciones de cenizas y gases. El 12 de Agosto ocurrió una explosión [Tenorio y Strauch 2000]. Mediciones de SO₂ en Marzo del 1,997 hecho con sistema CONSPEC, dieron como resultado un promedio de 41 ± 20 toneladas al día. En esa misma fecha, las mediciones de temperatura estuvieron en orden de 300° a 350° C., y se registraron además sismos con intervalos de 30 a 40 seg.

Para Febrero de 1,998 de acuerdo al INETER se dio un aumento en la actividad fumarólica del volcán y las temperaturas en el cráter alcanzaron los 550° C. En Mayo de 1,999 se dio un aumento en la actividad sísmica, reportándose más de 200 eventos por día. En ese mismo mes comenzó una erupción freática y la pluma de gases y cenizas alcanzó una altura de 500 m por sobre el borde del cráter. Para Junio de ese mismo año corrieron nuevas explosiones freáticas, ceniza volcánica en pequeñas cantidades cayó a 15 Km. al Oeste suroeste del volcán. En Agosto del mismo año después de una erupción de otro volcán (El Cerro Negro) se dio un súbito aumento en la actividad sísmica y volcánica en el volcán. Se produjeron pequeñas explosiones con emanaciones sucesivas de gases y cenizas. El magma observado en el interior del cráter aumentó de diámetro y se produjo una desgasificación acompañada por un ruido como de motor de jet de acuerdo a los lugareños. La actividad que comenzó en Mayo se prolongó hasta Diciembre de ese año.

En Mayo de 1,999 el INETER reporta que ocurrió una nueva actividad que produjo cenizas y gases. Para el 10 de Agosto de 1,999 el volcán entró nuevamente en actividad presentando significativas emanaciones de cenizas y gases. La diferencia con respecto a Mayo fue que la cantidad emanada fue mucho mayor y se prolongó durante

todo el mes. La temperatura que se observó en el intercráter fue de 192° C que fue mucho mayor que la registrada en el mes anterior de 137° C.

En Enero del 2000 el INETER reportó un aumento en la sismicidad y actividad volcánica, culminando El 13 de Enero del 2000 cuando se produjo una explosión con emanación de cenizas y gases (figura C.2). Durante varias horas se reportaron moderadas explosiones que acontecían cada cinco minutos, con ocasionales períodos de fuertes explosiones cada tres minutos. Asimismo, se presentaron colapsos alrededor del cráter principal. Durante todo el mes de Enero el número de terremotos volcánicos fue de 3,950 y los registros de RSAM (Amplitud sísmica) oscilaron entre 40 a 120 unidades.



Fig. C.2. Explosión ocurrida el 13 de Noviembre del 2000. (Fuente INETER)

En Noviembre del 2004 se presentó una nueva explosión acompañada previamente por un aumento en el temblor del volcán (figura C.3).



Fig. C.3. Explosión ocurrida el 11 de Noviembre del 2004. Foto tomada por la cámara Web del INETER

5.- Actividad sísmico-volcánica en el Etna durante el período de estudio

Italia es una zona que presenta un grado muy alto en materia de sismicidad, esto se refleja en el mapa que se presenta en la figura C.4. En este sentido el Etna y el Strómboli son dos de los volcanes con mayor actividad.

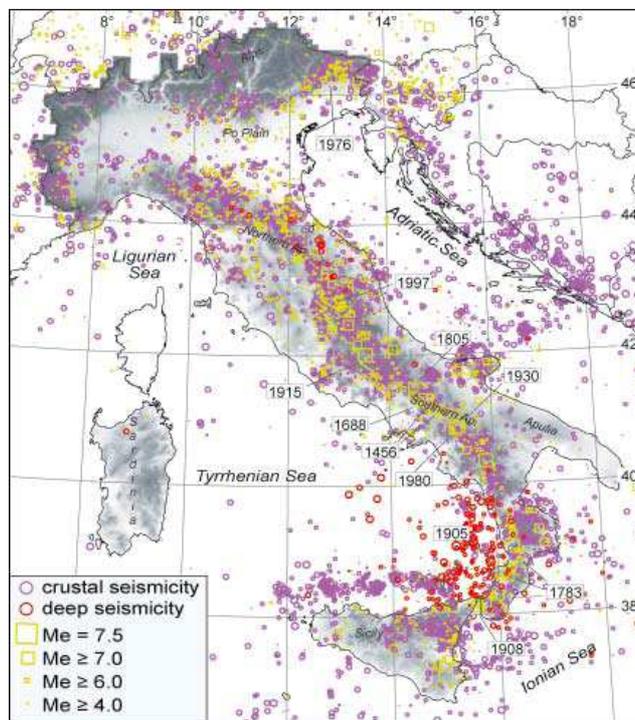


Fig. C.4. Sismicidad en Italia durante el período de 1983 a 2002. Fuente: De acuerdo al Catálogo del INGV (Istituto Nazionale di Geofisica e Vulcanologia) de Italia

Durante el desarrollo de la campaña el volcán Etna ha presentado mucha actividad. Un resumen de dicha actividad se puede determinar de acuerdo al boletín informativo del departamento de ciencias de la Universidad de Catania, elaborado por Boris Behncke y publicado por el Museo Nacional de Historia Natural del Instituto Smitsonian, a través de la página Web del Programa de Vulcanismo Global, en donde se manifiesta que el volcán Etna presentó una elevada actividad de Septiembre a Octubre de 1999. Dicha actividad se concentró alrededor de los cráteres de la cumbre y se refleja en el mapa de la figura C.5. Siguiendo un resumen de la cronología de los eventos principales narrados por Behncke: del 1 de Agosto al 3 de Septiembre, hubo poca actividad en el cráter Bocca Nuova (BN), el 15 de Agosto, se presentó desgasificación tranquila en la salida del cráter NE (nor-este), mientras que nubes de cenizas diluidas fueron emitidos a intervalos de varias horas en la salida del cráter SE (sur-este). En la Vorágine (Vr) sólo había emisión de gases del respiradero de NE, pero erupciones menores se produjeron a intervalos de 5-60 minutos de la ventilación central. Tremores, ruidos y emisiones de gases densos se produjeron de las profundidades de NE. El 4 de Febrero, varios pequeños flujos de lava se emitieron desde la zona del extremo superior de la fisura. Se observó una efusiva actividad ocurriendo una explosión visible entre nubes de gas y ceniza que afectaron varias poblaciones cercanas.

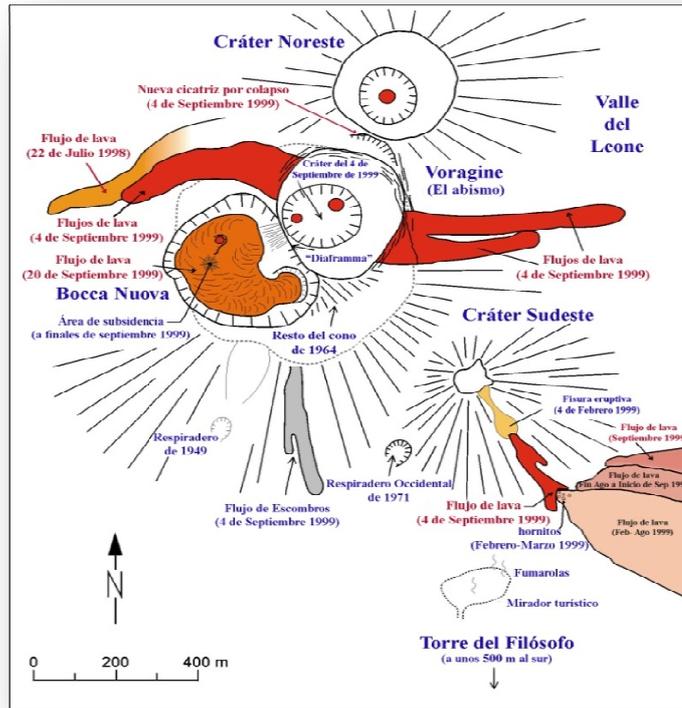


Fig. C.5. Actividad en los cráteres de la cumbre del volcán Etna entre el 7 de Septiembre y el 1 de Octubre de 1999. Mapa Tomado y modificado del original de [Behncke y Tanguy 1999]

Durante las semanas siguientes al 4 Septiembre, la actividad continuó en los cráteres de la cumbre, pero las observaciones se vieron interrumpidas por el mal tiempo. La intensa actividad explosiva ocurrió cada día en la BN, y en ocasiones las bombas fueron expulsadas a las pistas exteriores del cono de la cumbre principal. La actividad se mantuvo y la sismicidad indicó que la actividad más intensa se produjo entre 0100 y 0400 metros. Leves explosiones magmáticas se observaron por guías cada pocos minutos el 18 de Septiembre. En el día siguiente, se observaron salpicaduras de lava desde el cono NO (nor-oeste) de BN. La actividad estromboliana fue relativamente débil hasta la mañana siguiente. Las coladas de lava continuaron durante el 20 de Septiembre. En la figura C5 se aprecian los flujos de lava que se produjeron alrededor de los cráteres durante el mes de Septiembre. La dinámica en el Etna continuó con una actividad eruptiva el día 28 de Septiembre, la cual se desarrolló con frecuentes explosiones. La actividad eruptiva se reanudó en el cráter BN, el 30 de Septiembre, expulsando grandes bombas a cientos de metros más allá del borde del cráter. Al mismo tiempo, la actividad aumentó en el cráter NE. Del 29 al 30 de Septiembre, continúa la actividad estromboliana, expulsando gran cantidad de bombas sobre el borde del cráter, con grandes ráfagas que alcanzaron alturas de hasta 150 m, lanzando bombas por todo el suelo del cráter y en los flancos del cono NE [Behncke y Tanguy 1999].

6.- Actividad sísmico-volcánica en el Strómboli durante el período de estudio

De acuerdo a un informe elaborado por Roberto Carniel y publicado por el Museo Nacional de Historia Natural del Instituto Smitsonian, a través de la página Web del Programa de Vulcanismo Global, la actividad en el volcán Strómboli a principios de 1997 fue levemente moderada y continuó aproximadamente hasta Mayo de 1997, con una tendencia ligeramente decreciente en la intensidad del temblor volcánico y un

número ligeramente cada vez mayor de los eventos registrados. De Enero a Mayo, el número de eventos de saturación, es decir, los eventos que llegan al nivel de saturación para una estación sísmica, siempre fue muy bajo (la mayoría del tiempo se mantuvo por debajo de 10 por día).

También en el informe se indica que se produjo un aumento en el número total de eventos en los meses de Junio y Julio de 1997, a veces más de 300 al día y eventos de saturación (máximo: 92 el 8 de Junio de 1997). Después de una disminución de la actividad en Agosto, nuevamente comenzó la actividad en Septiembre (mes de la campaña) y continuó hasta Noviembre de 1997, durante los cuales la actividad sísmica fue casi continua durante varios días, en los que la intensidad de los sismos alcanzó valores muy altos (8 - 9 Vs). Los observadores indicaron que este tipo de comportamiento sísmico se registra habitualmente en presencia de salpicaduras casi continua en los respiraderos del volcán [Carniel 1998]. El informe del Smitsonian (BGVN 23:10), publicado en 1998, muestra en la figura C.6, la actividad de los eventos sísmicos por día durante el año 1997, en donde se observa el incremento de los sismos desde Septiembre mes en que se desarrolló la campaña a Noviembre.

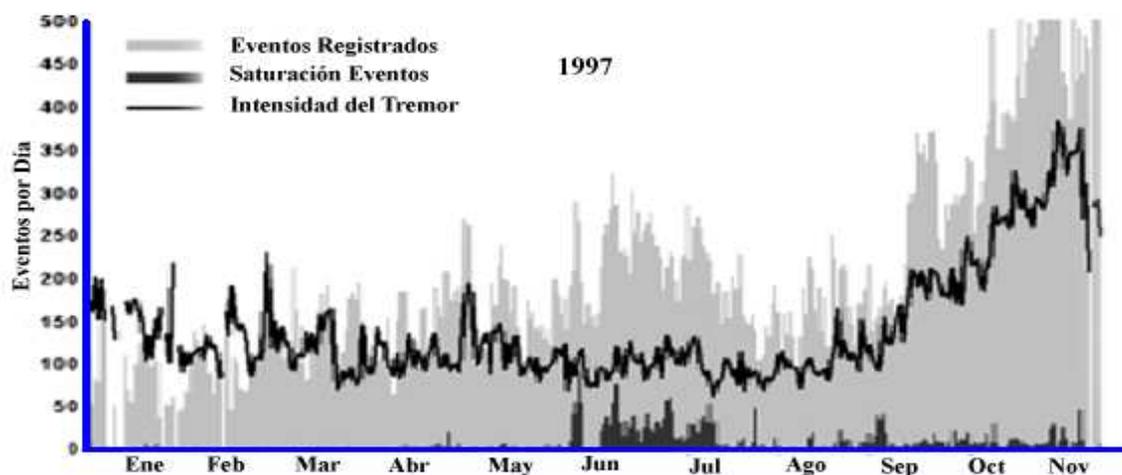


Fig. C.6. Actividad de eventos por día ocurridas durante 1997 en el volcán Strómboli. Fuente: Figura Tomada y modificada de la original de [Carniel 1998]

Hay que recordar que las explosiones en el Strómboli generalmente ocurren con una secuencia de 3 a 10 eventos por hora [Chouet et. al. 1999] con una actividad ocasional de lo que se conoce como enjambre sísmico, alcanzando de 20 a 30 eventos por hora. De acuerdo al informe del Smitsonian, la actividad estromboliana en Septiembre de 1997 se describe como un enjambre sísmico. Desafortunadamente, no existen datos sísmicos que fueron registrados entre el 24 de Noviembre de 1997 y 09 de Enero 1998.

7.- Actividad sísmico-volcánica del San Cristóbal durante el período de estudio

El INETER⁷ en Nicaragua desarrolla y publica mensualmente la vigilancia en los volcanes (entre ellos el San Cristóbal), mediante tres aspectos principales.

⁷ Los datos y las gráficas del San Cristóbal y Telica que aparecen en este apéndice, han sido modificados de los originales, que se encuentran en los boletines mensuales (Febrero y Marzo) y anuales publicados por el INETER en 2006.

- a) El número de microsismos
- b) La temperatura en el interior del cráter
- c) Las anomalías térmicas RSAM

A continuación, se desarrollan cada uno de estos aspectos ocurridos durante la campaña de 2006 en el San Cristóbal.

- a) El número de microsismos.

Durante el período de recopilación de datos del presente estudio (Febrero a Marzo del 2006), el volcán San Cristóbal de acuerdo al INETER presentó un número de sismos alto. En el mes de Febrero se registraron 681 microsismos (figura C.7), de estos el sistema del observatorio localizó 3 eventos sísmicos significativos. Al referirse a significativos, el observatorio los considera como tal cuando exceden los 3 o más grados de magnitud. En los primeros días del mes el número de microsismos oscilaron entre 70 y 50 por día. Entre el 5 y 17 de Febrero el número de microsismos bajó al igual que la amplitud de los mismos.

De acuerdo al INETER, el tremor en el volcán ocurrió en forma de bandas con duración de 2, 3 y 5 por día. Además, en este mes el volcán continuó con la actividad volcánica en menor escala que el mes anterior. Se registraron aproximadamente 47 pequeñas explosiones con salida de ceniza en pocas cantidades y abundantes emanaciones de gas de forma permanente, afectando a los poblados aledaños al volcán. Los días 15, 21, 25 y 27 de Febrero, ocurrieron pequeñas explosiones de gas y ceniza. De estas explosiones la más fuerte ocurrió el 21 de Febrero. En ese día el INETER emitió un comunicado vulcanológico sobre la explosión ocurrida en el volcán.

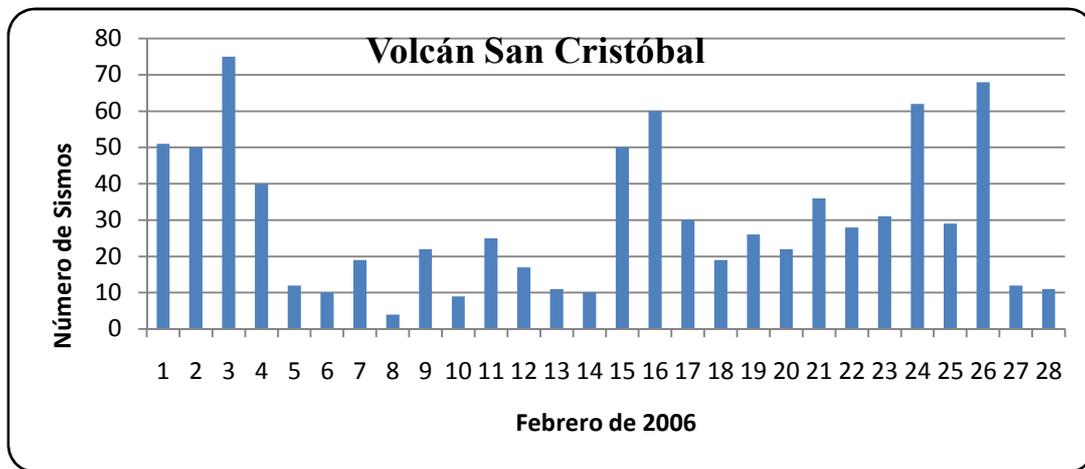


Fig. C.7 Número de microsismos ocurridos en el volcán San Cristóbal en el mes de Febrero del 2006 (Fuente INETER)

Durante el mes de Marzo, de acuerdo al boletín mensual emitido por el INETER, el número de microsismos continuó en aumento (Fig. C.8). En este mes se registraron 1,647 microsismos, de estos se localizaron por su magnitud 2. En los primeros días, el número de microsismos se mantuvo entre 90 y 70 por día. Del 7 al 23 de Marzo, oscilaron entre 30 y 40 microsismos. Posterior a este día, el número de microsismos fueron en aumento, manteniéndose entre 70 y 90 por día. El tremor se presentó en forma de bandas con una hora de duración.

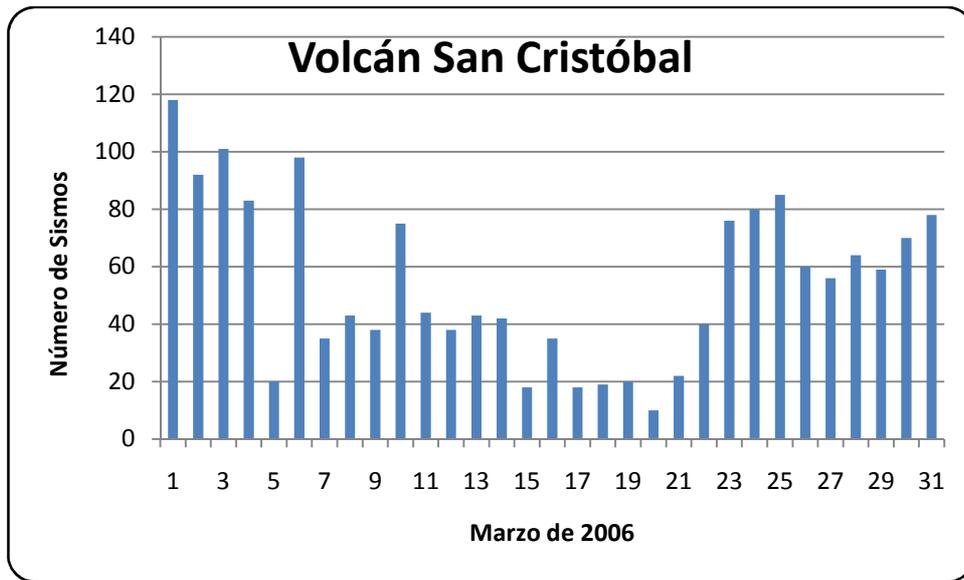


Fig. C.8 Número de microsismos ocurridos en el volcán San Cristóbal en el mes de Marzo del 2006 (Fuente INETER)

Durante el año 2006 la actividad sísmica del volcán fue relativamente alta. La tendencia al aumento de los sismos que se observó en Febrero y continuó en Marzo posteriormente concluyó en una erupción en el mes de Abril. Los mayores períodos de actividad estuvieron entre los meses de Marzo y Junio. El número de microsismos reportados por el INETER durante todo el 2006 fue de más de 15,735. Durante este período se destaca la explosión ocurrida el 23 de Abril (Fig. C.9).

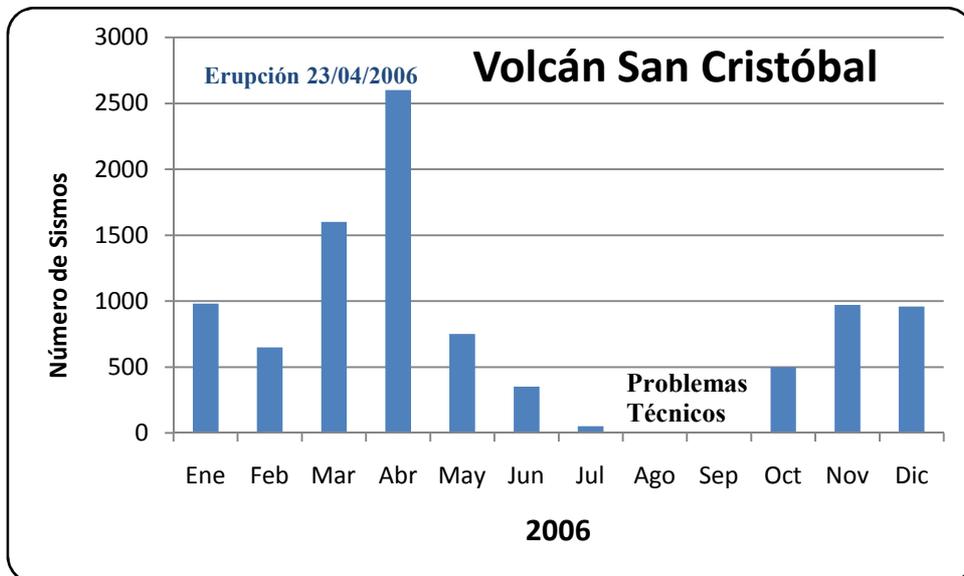


Fig. C.9 Número de sismos ocurridos en el volcán San Cristóbal durante el 2006 (Fuente INETER)

b) La temperatura en el interior del cráter.

La temperatura en el cráter es medida por un observador que reporta dichas mediciones al observatorio. No siempre es posible subir a la cumbre del cráter debido a las condiciones meteorológicas o por alguna otra causa. Se disponen de cinco puntos alrededor del cráter desde los cuales se recopilan los valores de temperatura. Dichos sitios han sido denominados como: El Conejo, Derecha, El Muñeco, El Zopilote, y Punto Sur. La temperatura reportada en los diferentes puntos de observación alrededor del cráter en el mismo período se puede observar en la figura C.10.

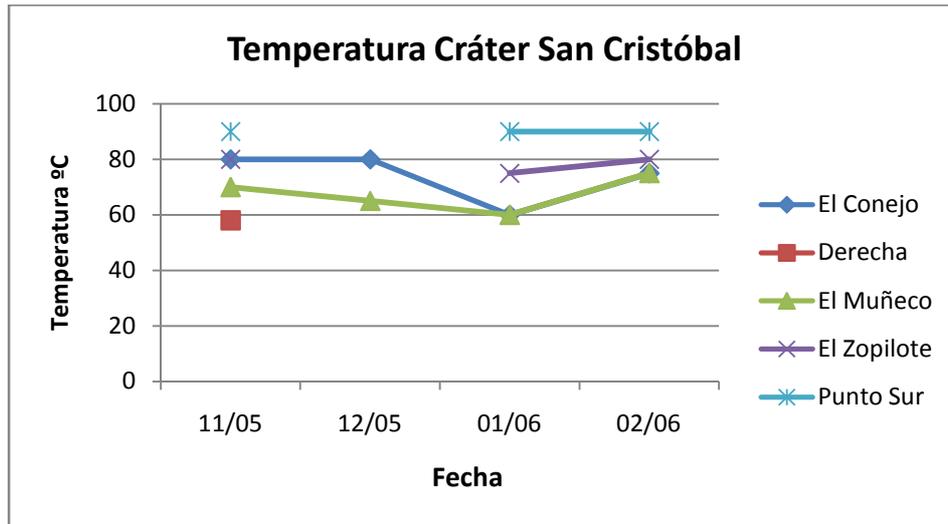


Fig. C.10 Temperaturas en los alrededores del cráter del San Cristóbal de Noviembre del 2005 a Febrero del 2006 (Fuente INETER)

Aunque no en todas las fechas se pudo tomar la temperatura, se pudo observar que hubo una disminución de la temperatura en el mes de Enero con un sucesivo aumento para el mes de Febrero del 2006. En el día 19 de Marzo, se realizaron mediciones de temperaturas, las cuales no presentaron variaciones significativas con respecto a los meses anteriores, lo que significó (figura C.11) una baja en los niveles con respecto a Febrero.

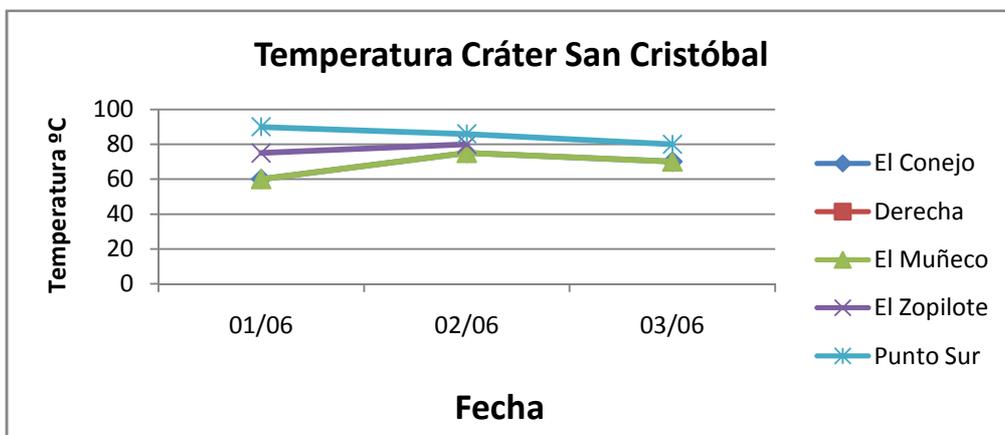


Fig. C.11 Temperaturas en los alrededores del cráter del San Cristóbal de Enero a Marzo del 2006 (Fuente INETER)

Como se observa en la figura C.11, de Enero a Marzo, la temperatura varió un poco en alguno de los puntos medidos. En el “Punto Sur”, mantuvo un leve descenso, en el punto “El Muñeco”, aumentó de Enero a Febrero, pero disminuyó hacia Marzo. Esa misma tendencia podría haberse dado en el punto “El Zopilote”, pero en Marzo no fue tomada la temperatura, al igual que ocurrió en otros puntos.

A lo largo del año 2006, la temperatura en los diferentes puntos de control de la zona del cráter del volcán del San Cristóbal fue monitorizada y se logró recopilar los datos que se muestran en la figura C.12.

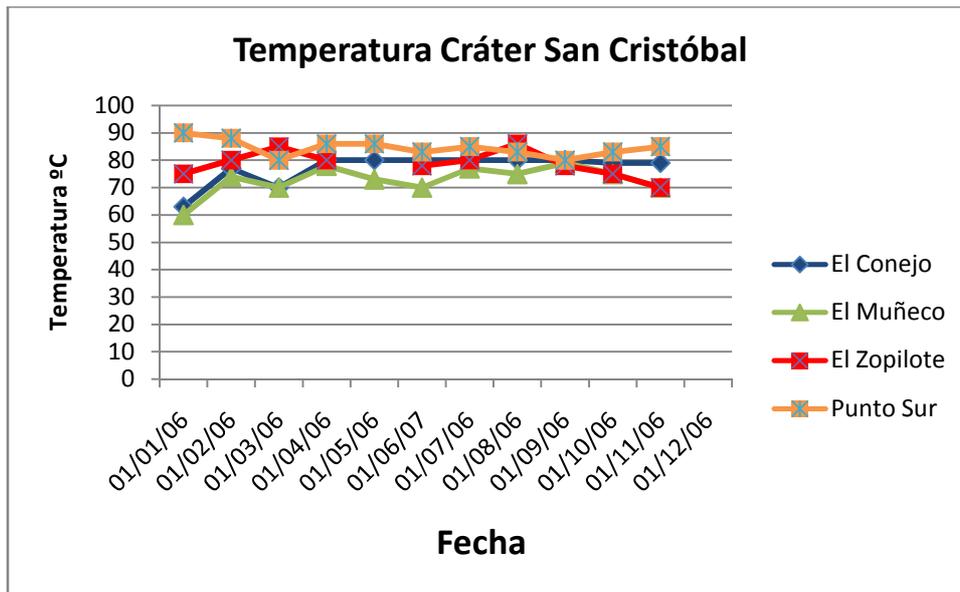


Fig. C.12 Temperaturas en los diferentes puntos de observación alrededor del cráter del volcán San Cristóbal en el 2006. (Fuente INETER)

Los índices de temperatura que se observan en la figura anterior, señalan que el promedio de temperatura en el interior del cráter del San Cristóbal, oscila entre valores cercanos a 70 °C y 90 °C aproximadamente.

c) Las anomalías térmicas RSAM

El nivel normal del tremor en el volcán San Cristóbal es de aproximadamente 10 unidades RSAM (amplitud sísmica promedia en un minuto). Sin embargo, se produjo un incremento en las anomalías térmicas RSAM reportadas en el volcán de Noviembre del 2005 a Febrero del 2006, de acuerdo al incremento de los microsismos en ese período como se aprecia en la figura C.13.

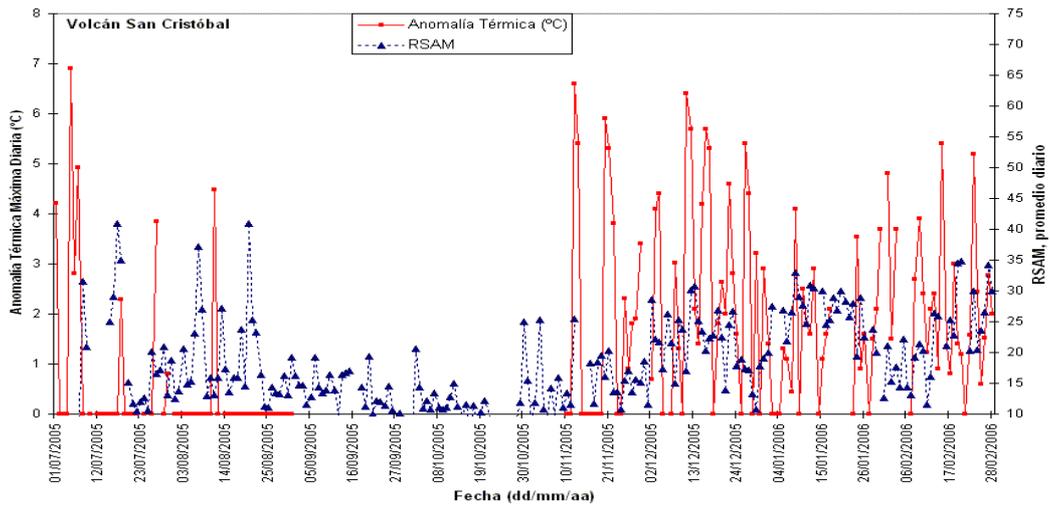


Fig. C.13 Anomalías térmicas RSAM en el volcán San Cristóbal de Noviembre del 2005 a Febrero del 2006 (Fuente INETER)

En los meses de recopilación de datos de Febrero a Marzo del 2006 se mantuvo la calma. Durante el mes de Febrero del 2006 la gráfica del RSAM se observa en la figura C.14.

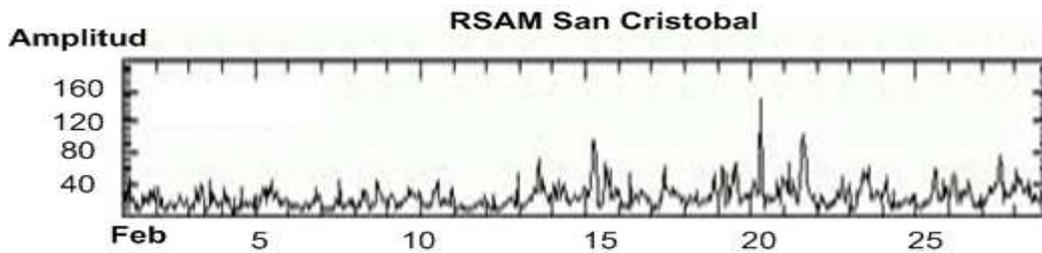


Fig. C.14 Anomalías térmicas RSAM en el volcán San Cristóbal en Febrero del 2006 (Fuente INETER)

Hubo unos picos de aumento de amplitud en torno al 15 y 20 de Febrero, por lo demás los registros observados en el volcán continuaron siendo normales.

Durante el mes de Marzo el RSAM del volcán San Cristóbal se mantuvo normal experimentando un leve aumento en los días 6 y 17 que se observa en la figura C.15.

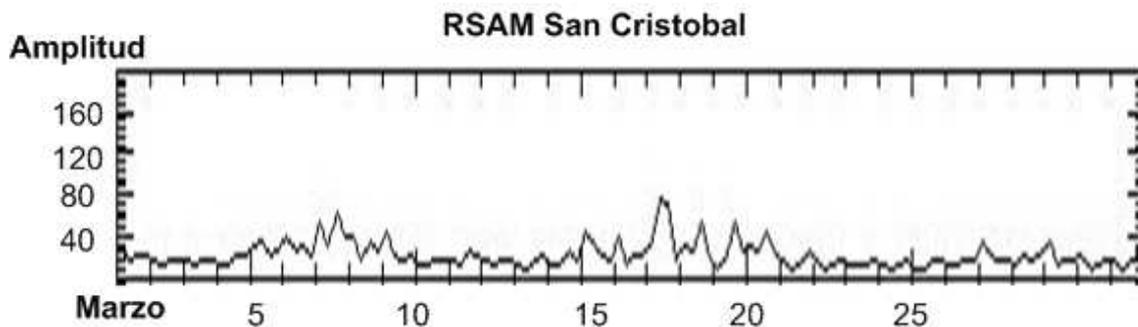


Fig. C.15 Anomalías térmicas RSAM en el volcán San Cristóbal en el mes de Marzo del 2006 (Fuente INETER)

El tremor sísmico se mantuvo entre 10 y 100 unidades RSAM. Tanto la figura anterior como el registro RSAM del mismo período que se muestra en la figura C.16, presentan un aumento significativo del número de sismos y las anomalías térmicas en el volcán antes de la erupción del día 23.

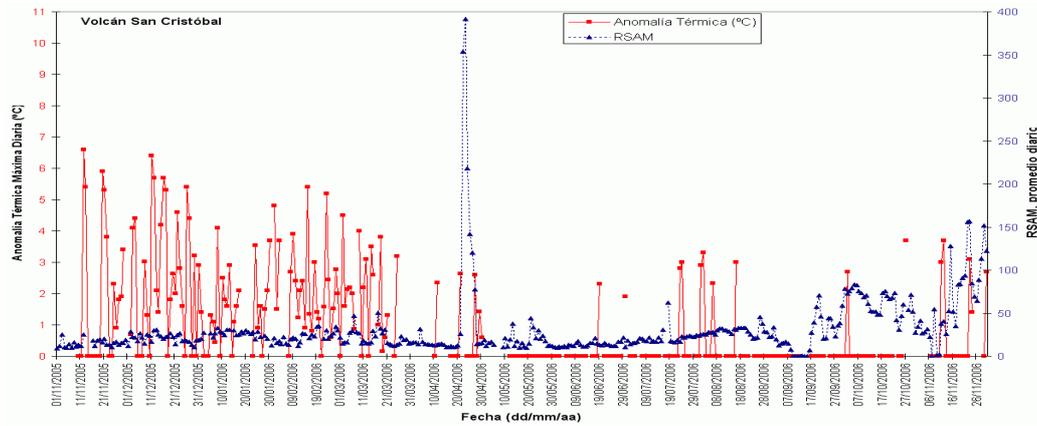


Fig. C.16 Anomalías térmicas y promedio RSAM del San Cristóbal ocurridas en el 2006 (Fuente INETER)

En el mes de Mayo del 2006 ocurrieron en el volcán sucesivas emanaciones de gas y cenizas así como algunas explosiones freáticas. Unas de mayor magnitud que otras. Esto se observa en el registro RSAM del San Cristóbal del mes de Mayo figura C.17.

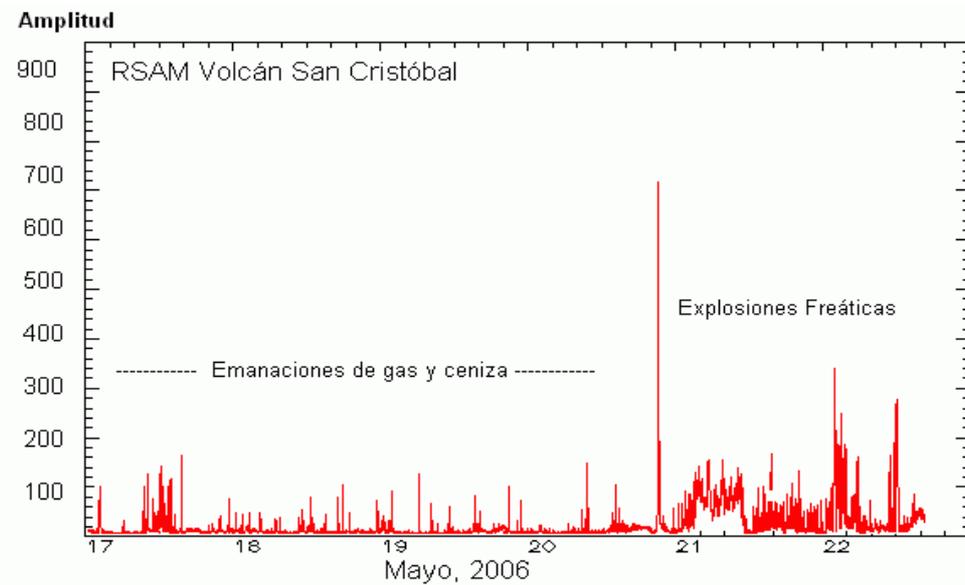


Fig. C.17 RSAM del volcán San Cristóbal en el mes de Mayo del 2006 cuando se produjeron emanaciones de gas, ceniza y explosiones freáticas. (Fuente INETER)

8.- Actividad sísmico-volcánica en el Telica durante el período de estudio

Al igual que en el San Cristóbal, en el Telica la vigilancia del INETER está caracterizada por tres aspectos principales:

- a) El número de microsismos
- b) La temperatura en el interior del cráter
- c) Las anomalías térmicas RSAM

Durante la campaña de 2006, en el Telica los boletines mensuales del INETER reflejan la siguiente actividad:

- a) El número de microsismos

De acuerdo a los datos recopilados por el INETER en el mismo período de recopilación de datos que en el volcán San Cristóbal (Febrero a Marzo del 2006), el volcán Telica presentó un número de sismos alto. En el mes de Febrero se registraron 3,568 microsismos, de estos el observatorio pudo localizar 1 evento sísmico significativo. (Figura C.18). La frecuencia predominante de los sismos se mantiene entre 2.0 y 5.0 Hz.

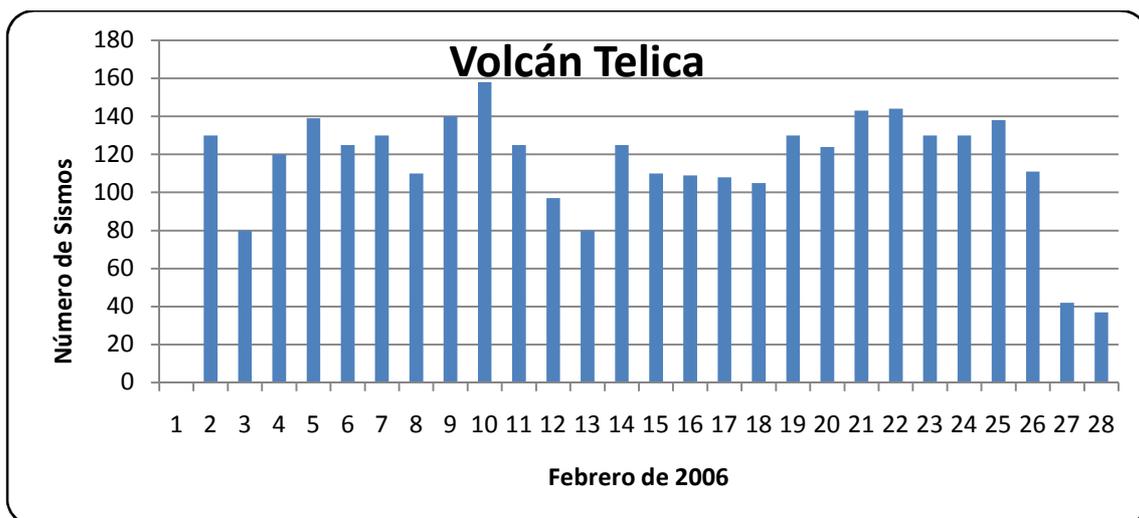


Fig. C.18 Número de sismos ocurridos en el volcán Telica en el mes de Febrero del 2006
(Fuente INETER)

La figura C19, muestra de acuerdo al INETER el número de microsismos ocurridos durante el mes de Marzo en el Telica. El comportamiento siguió siendo alto, aún más que en el mes anterior. De acuerdo al INETER, se registraron 4,739, de estos, el observatorio localizó (de acuerdo a su magnitud) un evento sísmico significativo. La frecuencia predominante de los sismos se mantuvo entre 2.0 y 5.0 Hz.

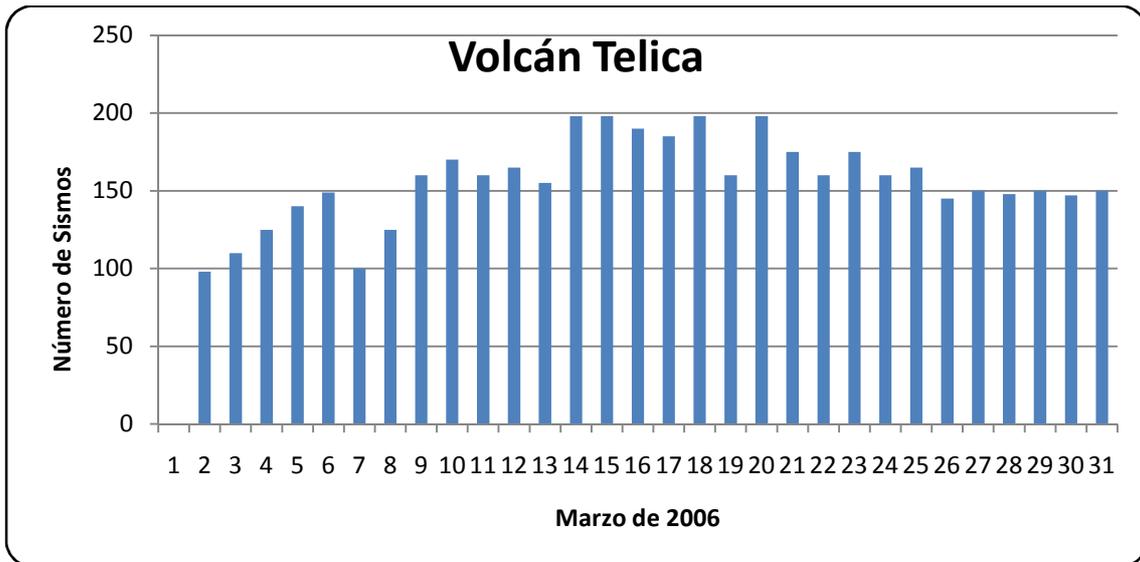


Fig. C.19 Número de sismos ocurridos en el volcán Telica en el mes de Marzo del 2006 (Fuente INETER)

A lo largo del año 2,006, la producción de sismos en el volcán Telica se mantuvo con un promedio de casi 4,000 sismos por mes. Este número se elevó a casi 5,000 o más durante los meses Abril y Septiembre (figura C.20).

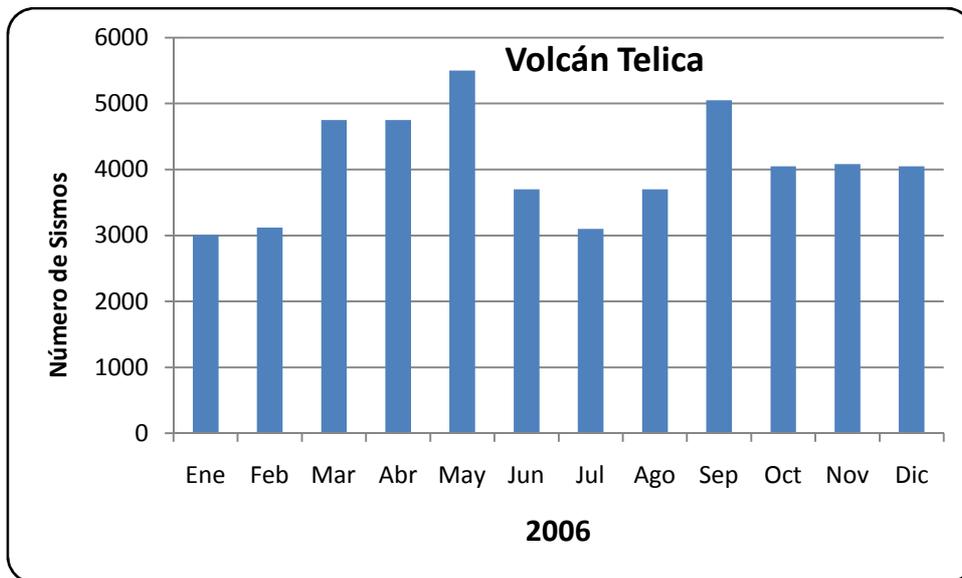


Fig. C.20 Número de sismos ocurridos en el volcán Telica en el 2006 (Fuente INETER)

b) La temperatura en el interior del cráter.

Al ubicarse los volcanes San Cristóbal y Telica en complejos relativamente cercanos, la temperatura en el cráter del Telica también es medida por el mismo observador, y que reporta dichas mediciones al observatorio. En el Telica se disponen de seis puntos alrededor del cráter, desde los cuales se recopilan los valores de

temperatura. Dichos sitios han sido denominados como: Intercráter Norte, Intercráter Oeste, Pared Sur, Pared NE, Intercráter NO e Intercráter Este

De manera similar que en el San Cristóbal se presentaron algunos inconvenientes en la medición de la temperatura, así que por problemas de traslados y/o condiciones ambientales, no se pudo realizar la medición de temperatura en los meses de Enero a Marzo del 2006, que fue el período cuando se realizó la campaña de recopilación de la información, lo mismo sucedió en Diciembre de ese año. Sin embargo, la figura C21, muestra el resumen de la temperatura que pudo ser reportada en los diferentes puntos de observación alrededor del cráter, en el 2006.

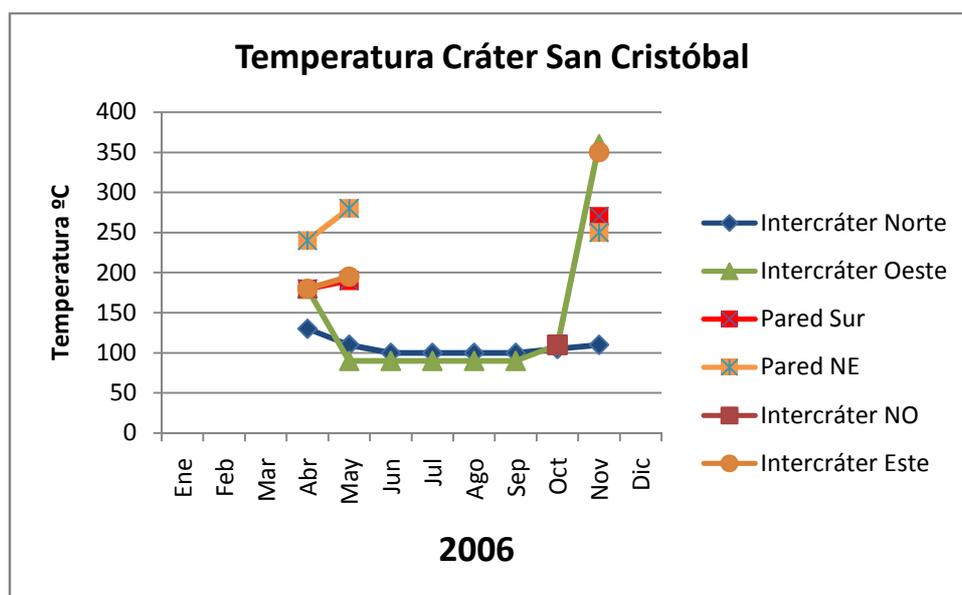


Fig. C.21 Temperaturas en el cráter del volcán Telica durante el 2006 (Fuente INETER)

Como se observa en la figura anterior, en el 2006 las temperaturas promedio en los puntos de observación en el Telica, experimentaron aumentos en Mayo y Octubre. La temperatura promedio de Abril a Mayo estuvo en torno a 200 °C, manteniéndose durante los siguientes meses, hasta aumentar de Octubre a Noviembre llegando a un poco más de 350 °C. Con estos valores se puede observar que a comparación del San Cristóbal, el Telica en su cráter mantiene un promedio de temperatura mucho más elevada.

c) Las anomalías térmicas RSAM

El tremor sísmico a diferencia del San Cristóbal de 10 unidades se mantiene en 15 unidades RSAM (figura C.22)

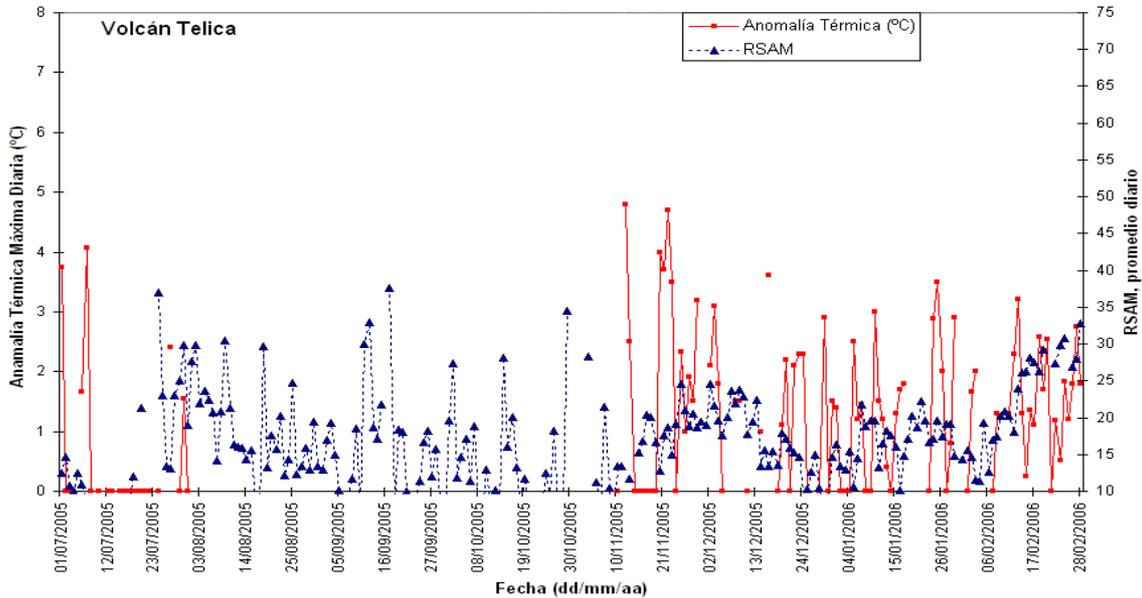


Fig. C.22 Anomalías Térmicas y promedio de RSAM del volcán Telica en el período de 01/07/2005 al 28/02/2006 (Fuente INETER)

Como se puede observar en la figura anterior hubo aumento tanto de la anomalía térmica como del índice RSAM a partir del mes de Noviembre del 2006. Este aumento continuó siendo variable hasta Febrero del 2006. En los meses de recopilación de datos, durante el mes de Febrero de ese año (fig. C.23), el RSAM experimentó pico los días 5 y 20, por lo demás su amplitud se mantuvo constante.

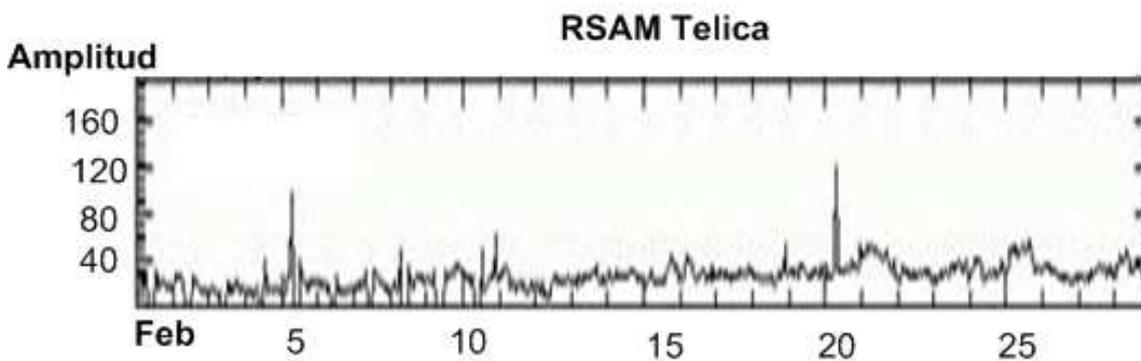


Fig. C.23 RSAM del volcán Telica durante el mes de Febrero del 2006 (Fuente INETER)

Para el mes de Marzo, en la gráfica del RSAM (figura C.24), puede observarse que a pesar del aumento de los sismos, no se presentaron picos significativos como en el mes anterior. La amplitud se mantuvo constante. El tremor sísmico continuó manteniéndose en 15 unidades RSAM.

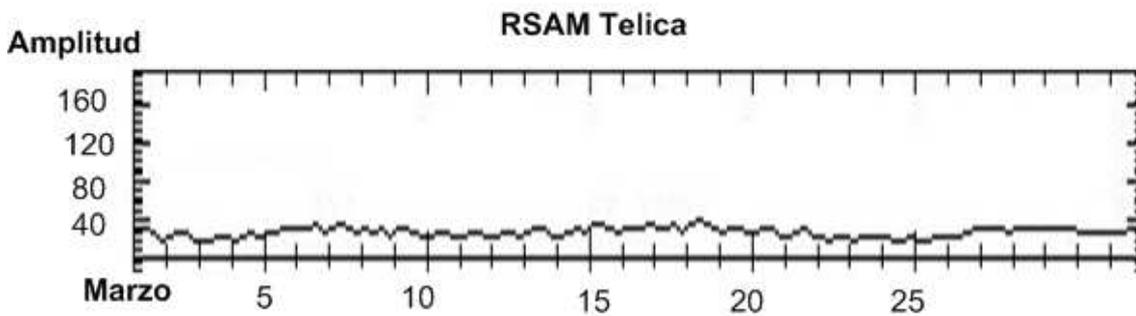


Fig. C.24 RSAM del volcán Telica durante el mes de Marzo del 2006 (Fuente INETER)

El comportamiento de las anomalías térmicas y el promedio RSAM del volcán entre Noviembre del 2005 a Noviembre del 2006 comenzó a ser constante a partir de Enero del 2005 y se mantuvo en ese mismo nivel hasta Noviembre del 2006 (figura C.25).

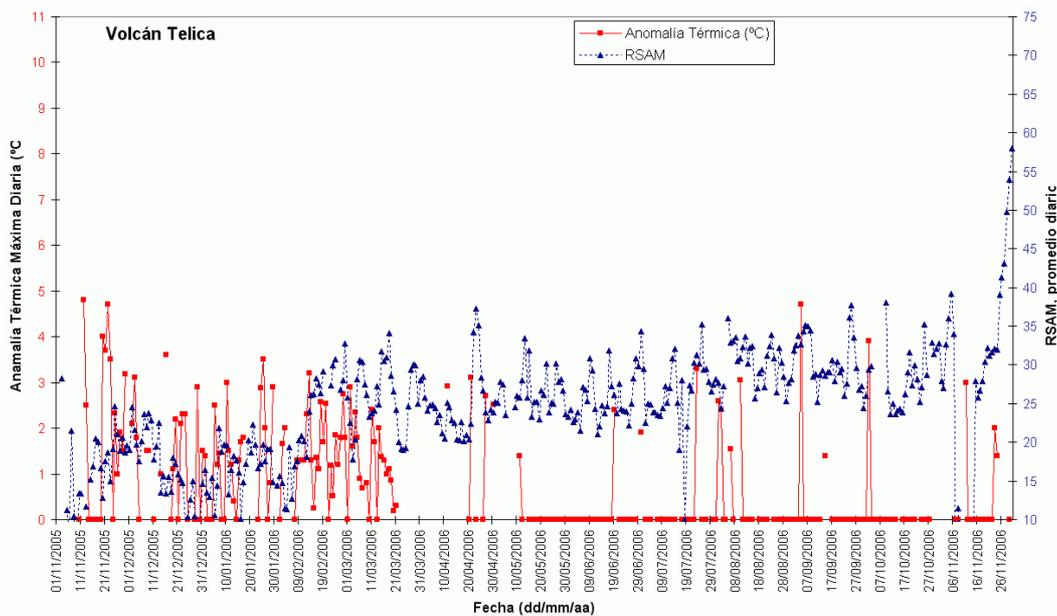


Fig. C.25 Anomalías Térmicas y promedio de RSAM del volcán Telica en los años 2005 y 2006 (Fuente INETER)

Toda la información anterior refleja la intensa actividad que se produjo en los cuatro volcanes de estudio durante el período de recopilación de la información. En Nicaragua, durante el 2006 de acuerdo a las gráficas, puede apreciarse que el volcán Telica mantiene una actividad sismo-volcánica mucho mayor que el San Cristóbal, en cuanto a la producción de sismos y al aumento de temperaturas. En cambio, tanto las anomalías térmicas como el promedio RSAM durante el 2006, ha sido mucho menor en el Telica que en el San Cristóbal.

APÉNDICE D: Herramientas complementarias en el análisis de señales

1. Introducción.

Tal y como se ha indicado en el capítulo dos y en una de las conclusiones, es posible incorporar otras herramientas que las utilizadas en el presente trabajo, a fin de completar el análisis de la caracterización de las señales sísmico-volcánicas. En este apéndice se presenta un complemento a dicho análisis, mediante el uso de algunas herramientas, que pueden ayudar en el desarrollo de las componentes espectrales. Dichas herramientas pueden incorporarse en futuras investigaciones, a fin de que puedan reforzar los análisis ya vistos en este presente trabajo, en busca de obtener mejores resultados. A continuación se presentan dos de ellas. La densidad espectral de potencia y la Transformada Wavelet.

2.- Densidad espectral de potencia de las señales

La **Densidad Espectral** (Spectral Density) representada en la figura D.1, es una medida de la energía de la señal en función de la frecuencia. Dicho de otra forma es una función matemática que informa de cómo está distribuida la potencia o la energía (según el caso) de dicha señal sobre las distintas frecuencias de las que está formada, es decir, su espectro. En la práctica es conveniente, además, expresar el espectro de frecuencias de esta señal (originalmente analógica) en el ancho de banda de 0 a la mitad de la frecuencia de muestreo. M.S. Bartlett a partir de 1953 inició el estudio de promediar los periodogramas para la estimación de los espectros de potencia de una señal. A medida que fueron perfeccionándose los algoritmos para el cálculo de la transformada discreta de Fourier, P.D. Welch (1970) combinó esos algoritmos con el uso de una ventana de datos $w[n]$, para desarrollar el método de lo que se denomina método modificado de promedios de periodogramas. En este apéndice, se analiza esta técnica mediante el uso del periodograma por medio de Welch, para encontrar las componentes frecuenciales propias de las señales sísmicas en los volcanes a fin de caracterizarlas y permitir su detección.

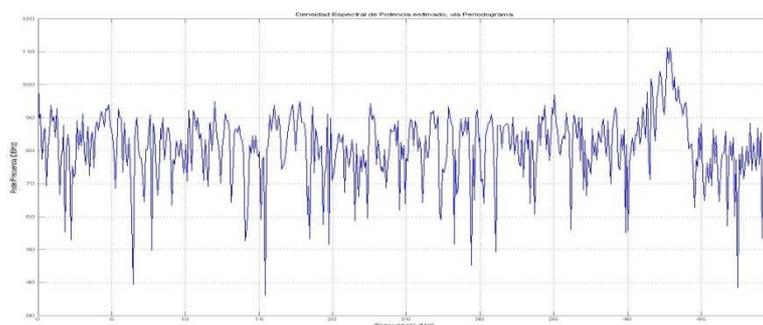


Fig. D.1 Imagen de la Densidad Espectral de Potencia de una señal estimada vía Periodograma

Para un proceso estocástico estacionario “ $x(t)$ ”, la Densidad Espectral de Potencia se define como

$$G_x(f) = \lim_{T \rightarrow \infty} \frac{1}{T} E\{|X_T(f)|^2\} \quad (1)$$

Donde, $X_T(f)$ es la Transformada de Fourier del proceso truncado

$$x_T(t) = \begin{cases} x(t) & |t| < T/2 \\ 0 & |t| > T/2 \end{cases} \quad (2)$$

y donde queda por lo tanto:

$$X_T(f) = \int_{-\infty}^{+\infty} X_T e^{-j2\pi ft} dt = \int_{-\frac{1}{2}T}^{\frac{1}{2}T} x(t) e^{-j2\pi ft} dt \quad (3)$$

Como estimador de la densidad espectral, se suele emplear en la práctica la versión discretizada y se denomina: "El Periodograma" [Oppenheim et. al. 1999], que en una secuencia $[X_1, \dots, X_n]$ se determina por la ecuación 4.

$$S(e^{j\omega}) = \frac{1}{n} \left| \sum_{l=1}^n x_l e^{-j\omega l} \right|^2 \quad (4)$$

En caso de que el valor de la secuencia de la señal se encuentra dado por una ventana de $[W_1, \dots, W_n]$, entonces, el periodograma ponderado o modificado se encuentra definido mediante la ecuación 5.

$$S(e^{j\omega}) = \frac{\frac{1}{n} \left| \sum_{l=1}^n \omega_l e^{-j\omega l} \right|^2}{\frac{1}{n} \sum_{l=1}^n |\omega_l|^2} \quad (5)$$

En ambos casos, el periodograma utiliza una FFT para calcular la densidad espectral de potencia como $S(e^{j\omega})/F$, donde F corresponde a los siguientes valores:

- . 2π , cuando se desconoce la frecuencia de muestreo
- . fs, cuando se conoce la frecuencia de muestreo

El periodograma por lo tanto, se calcula dividiendo la señal en un número determinado de segmentos, posiblemente traslapados, y evaluando la transformada de Fourier en cada uno de estos segmentos. Dicho de otra forma, resulta cuando el espectro de potencia se calcula vía transformada de Fourier y posteriormente se promedia.

3.- La transformada Wavelet

El término Wavelet, se originó del vocablo francés "ondelettes", que posteriormente derivó a la expresión inglesa de "wavelets" (wave = onda; let = pequeña) o pequeñas ondas. [De Oliveira 2007]. Es por ello que la principal función de la Transformada Wavelet es la de descomponer la señal en unas determinadas

componentes llamadas Wavelets (o onditas), y que forman una base de espacio de funciones con algunas propiedades como el tamaño, la ortogonalidad, la suavidad, duración entre otras. Las Wavelets pueden interpretarse como transformaciones lineales locales generadas por un banco de filtros de factor de calidad constante [De Oliveira 2007]. Una definición formal ofrecida por Peña, podría definirse a las wavelet como “*funciones matemáticas que filtren series temporales de datos para analizar la variación de su contenido espectral*” [Peña 2006]. Así, la transformada Wavelet analiza la señal con distintas resoluciones para diferentes frecuencias, lo que implicaría una representación en tiempo-frecuencia mucho más exacta para las señales no estacionarias (como por ejemplo los sismos de origen tectónico y volcánico), que un análisis mediante la transformada de Fourier [Melhem y Kim 2003]. El uso de la Transformada Wavelet en el análisis de señales es relativamente nuevo y aún se sigue definiendo, aunque su importancia se hace cada vez más evidente. Diversos trabajos, presentan una definición de las wavelet, su papel en la ingeniería y en el análisis de señales [Burrus et. al. 1998]; [Daubechies 1988]; [Daubechies 1990]; [Holschneider et. al. 1990]; [Kaiser 1994]; [Kingsbury 2001]; [Williams y Amaratunga 1994]. El análisis Wavelet es similar al análisis de Fourier, ya que descompone la señal en versiones dilatadas y desplazadas de una función madre. Tanto la transformada de Fourier como la Transformada Wavelet pueden ser utilizadas para analizar señales periódicas y ambas son reversibles, es decir que pueden pasar del dominio del tiempo al de la frecuencia y viceversa. Sin embargo, a diferencia de la transformada de Fourier, la transformada Wavelet permite proporcionar de manera simultánea información tanto en el dominio del tiempo como en el de la frecuencia, lo que no ocurre con la transformada de Fourier que pierde la información de un dominio al pasar al otro [Peña 2006].

Al buscar la evolución temporal de un espectro de frecuencias, generalmente se utilizan algoritmos como la transformada rápida de Fourier (FFT) y la transformada localizada de Fourier (STFT). Dichos algoritmos utilizan el solapamiento de ventanas para así localizar frecuencias en el tiempo. Sin embargo, el que la resolución se encuentre determinada por el tamaño fijo de dichas ventanas puede ser un problema. Por lo tanto, al utilizar la FFT ventanas de longitudes fijas, acarrea la pérdida de resolución a frecuencias mayores que la longitud de la propia ventana. Una posible solución a este problema sería el utilizar técnicas como el solapamiento de las ventanas, con lo que se reduciría, pero no por completo [Peña 2006]. Aunque existen otros métodos que trabajan a escalas temporales más grandes, como el SSAM (*Seismic Spectral Amplitude Measurement*) [Power et. al. 1991]; [Rogers et. al. 1995], indudablemente de acuerdo a Bartosch, el análisis con técnicas de funciones wavelets, consigue analizar los datos a una escala más fina que la que hace la transformada localizada de Fourier (STFT) [Bartosch 1999]. Esto sucede porque se utilizan ventanas de longitud variable en función de la búsqueda deseada, descartando la información que se da para frecuencias mayores. El uso de la transformada Wavelet facilitaría esta operación, porque en lugar de calcular un promedio, identificaría la frecuencia de un determinado ciclo de la serie temporal y le asignaría la posición. Con esto, los espectrogramas resultantes podrían utilizarse para detectar los patrones de comportamiento en las variaciones del espectro. Lo que en el caso de las señales sísmicas, sería especialmente útil para agrupar e identificar de acuerdo a estos criterios los diferentes tipos de eventos que se producen en un volcán [Peña 2006]. Peña describe dos tipos de transformadas Wavelet, una de ellas es la transformada discreta (DWT) que separa los contenidos de frecuencias altas y bajas de forma equivalente a un banco de filtros pasa-altos y pasa-bajos. Este tipo de transformada se utiliza generalmente para las series de datos que presentan grandes

variaciones y discontinuidades, como es el caso del tratamiento de imágenes. El otro tipo es la transformada continua (CWT), esta actúa como un filtro paso-banda, dejando sólo las componentes de frecuencia deseadas [Peña 2006]. En ambos casos las transformadas son utilizadas para resolver el problema de resolución que presenta la STFT, sin embargo, la CWT es más eficaz al tratar el estudio de series temporales de datos con una variación más suave, que es el caso particular de los eventos sísmicos (tectónicos y volcánicos). La transformada Wavelet continua (DWT) que se observa en la figura D.2, es la suma para todo tiempo (t) de la señal multiplicada por la escala de una versión desplazada de una función prototipo, denominada la función madre $\Psi(t)$. La DWT pretende así determinar una señal $x(t)$ que es continua en el tiempo, a través de una expansión de términos o coeficientes que son proporcionales al producto interno entre la señal y diferentes versiones escalonadas y trasladadas de la función madre $\Psi(t)$. El resultado lo determinan los coeficientes Wavelet que son escala y posición de la función $\Psi(t)$. Aquí se asume que tanto la señal original como la función $\Psi(t)$ son de energía finita definiéndose por lo tanto, la función Wavelet a través de la ecuación 6.

$$\Psi_{a,b}(t) = \frac{1}{\sqrt{a}} \Psi\left(\frac{t-b}{a}\right) \quad a, b \in \mathbb{R}, a \neq 0 \quad (6)$$

Por lo que la Transformada Continua Wavelet (DWT) se expresa por medio de la ecuación 7

$$CWT(a, b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} x(t) \Psi\left(\frac{t-b}{a}\right) dt \quad (7)$$

Donde “b” es el factor de desplazamiento y “a” es el factor de escala de la wavelet, que determina su tamaño relativo y es proporcional al inverso de la frecuencia, por lo tanto, contiene la información relativa a las componentes espectrales de la señal. Cuanto más pequeño sea el factor de escala, mucho mayor será la comprensión de la Wavelet. En general, no se puede dar una relación más precisa entre la escala y la frecuencia, debido a que esto depende de la función que se utiliza y se debe de averiguar en cada caso. [De Oliveira 2007]; [Torrence y Compo 1998].

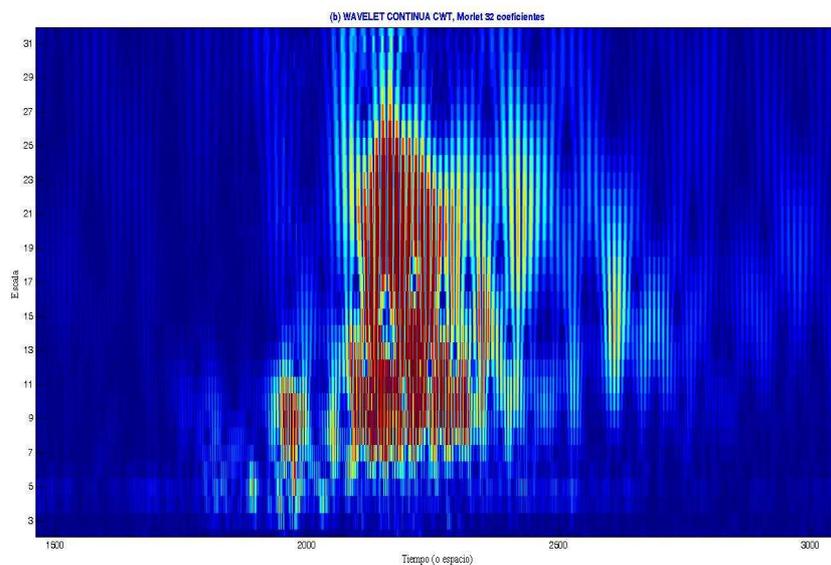


Fig. D.2 Grafica de la transformada Wavelet de una señal sísmico-volcánica.

En cuanto a las funciones wavelet, la variedad disponible es muy amplia, en función del detalle con que se quiera analizar el registro, de esta forma: pueden proporcionar información de la frecuencia de cada ciclo (función Mexhat o sombrero mexicano), también pueden estudiar la señal a una escala más amplia si sólo interesa la tendencia general del espectro (función de Morlet [Torrence y Compo 1998]). Existen otros métodos, como los filtros gaussianos, por ejemplo en los trabajos de Lermo en 1995 que producen representaciones tiempo-frecuencia parecidas puesto que la CWT también utiliza funciones gaussianas para filtrar la señal [Lermo et. al. 1995]. Con la transformada de Wigner-Ville también se obtienen resultados parecidos puesto que se consigue resolver el espectrograma a una escala muy fina [Qian 2002]; [Boashash y Black 1987]. Además de la ventaja que proporciona el análisis de la señal a pequeña escala, la principal ventaja práctica de la transformada continua wavelets es que puede resolverse con la ayuda de los algoritmos tradicionales de la transformada de Fourier (FFT), que han sido métodos lo suficientemente probados y de uso común en el análisis de señales.

Por medio de esta técnica se ha podido observar como las frecuencias de señales volcánicas poseen fuertes variaciones que permiten distinguir la ocurrencia de unos u otros tipos de eventos [Chavez et. al. 2010]. Lo que para futuros trabajos podría ser muy útil, al analizar los eventos de los volcanes, ya que así se podría definir aun más dichos eventos, al diferenciar entre eventos similares (esto sería un paso previo a la construcción de bases de datos más fiables con eventos reconocidos por los expertos). Estas variaciones o cambios se estudian en el plano temporal, de forma que se pueden averiguar el momento de ocurrencia de un determinado pulso. De la misma forma pueden detectarse pequeñas variaciones en el contenido espectral de la señal que en ciertos casos podrían interpretarse como consecuencia del establecimiento de fracturas (pulsos de alta frecuencia sobre un fondo de bajas frecuencias), y en el momento del establecimiento de las posibles fracturas respecto del comienzo del evento.

Al utilizar las variaciones de los espectrogramas pueden establecerse con mayor precisión los algoritmos de clasificación automática de los eventos sísmico-volcánicos presentes en las bases de datos que se construyan. Así estos algoritmos que en el presente trabajo están basados en los HMMs (Modelos Ocultos de Markov), permiten identificar de forma efectiva eventos como los Vulcano-tectónicos, el tremor y algunos eventos de otros tipos como los de Tipo-B. También pueden identificarse algunos tipos de explosiones en base a su contenido espectral, sin embargo, en muchos casos al analizar con los espectrogramas, la similitud entre los eventos de señales de tipo-B hace complicada una efectiva clasificación. Este problema se da sobre todo con los eventos del Tipo-B y como lo definen las clasificaciones actuales, que lo asemejan a explosiones. Por lo tanto, para resolver este y otros casos se tendría que optimizar la precisión de dichos algoritmos, al añadir ya sea más eventos o modelos. Para el caso de las señales sísmicas de los volcanes se podría hacer uso de la Wavelet continua (CWT) de tipo Morlet, a fin de tener una imagen mucho más clara de las componentes de las energías de la señal.

En el siguiente apéndice, se hará uso de estas dos técnicas para realizar un análisis complementario de los eventos sísmico-volcánicos, a fin de determinar su posible utilidad en futuros trabajos, de las semejanzas y diferencias entre los eventos.



APENDICE E: Aplicaciones a las señales sísmicas de las herramientas complementarias de análisis de poder espectral y componentes Wavelet

1. Introducción.

El apéndice anterior mostraba los componentes teóricos de las herramientas que podrían utilizarse para complementar el análisis de la caracterización de las señales sísmicas. En este apéndice, se presenta un detallado proceso de dicho análisis realizado a diversos registros presentes en los cuatro volcanes de estudio. Se podrá determinar la posible importancia de incorporar el análisis espectral de potencia y de la transformada Wavelet en los registros sísmicos presentes en los volcanes.

2.- Estimación del poder espectral “PSD”, de los eventos

En futuras investigaciones pueden tomarse en cuenta diversos métodos de análisis con las señales sísmico-volcánicas. Uno de ellos es el método de periodograma de Welch, que consiste en segmentar la señal (en forma solapada o no). En función de la forma que tenga la Densidad Espectral de Potencia, PSD (Power Spectral Density) [Stoica y Moses 1997], se puede entonces ayudar a caracterizar el tipo de evento ante el que se encuentra. La utilidad de esta técnica ha sido presentada en diversos trabajos, en donde se manifiesta su utilidad en la clasificación y análisis de las señales sísmicas producidas en los volcanes [Chouet et. al. 1994]; [Orozco et. al. 2006]; [Rowe et. al. 2000]. A continuación se presentan las formas características de los distintos tipos de eventos analizados en los volcanes de estudio.

a) Volcán Etna

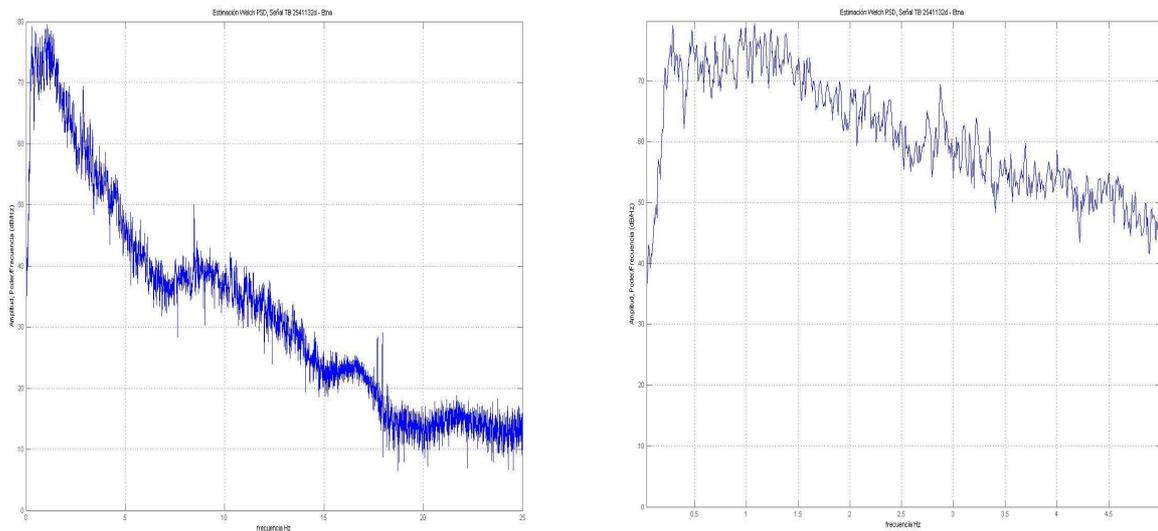


Fig. E.1 Representación de la estimación del poder espectral PSD por medio de Welch del evento TB, (Señal 2541132d) (izq.) y Zoom a los 5 Hz. (der.) en el volcán Etna

Picos predominantes entre los 0.5 y 1.5 Hz. La señal comienza a decaer con una suave pendiente a partir de dicha medida hasta los 5 Hz.

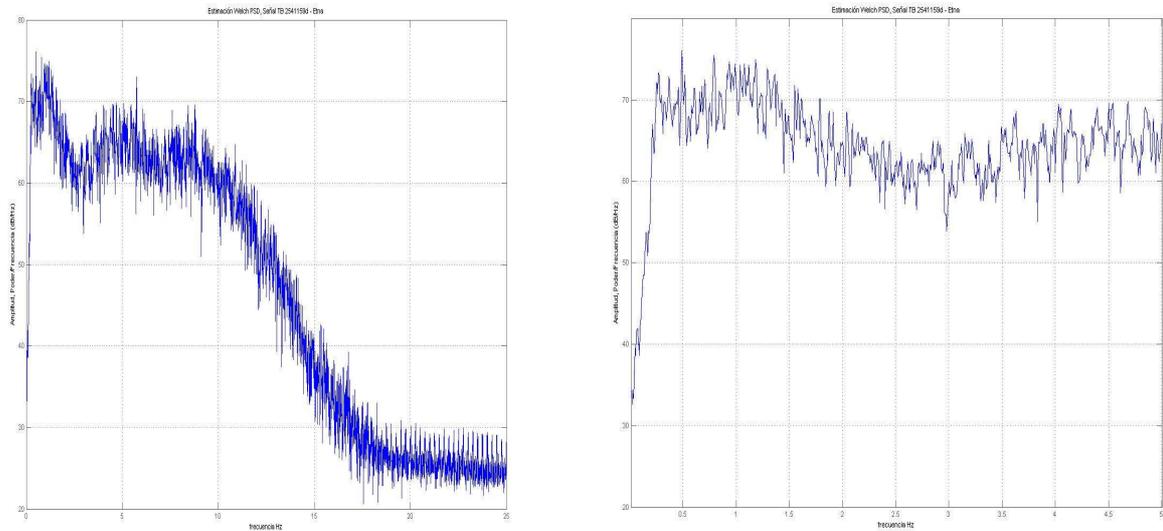


Fig. E.2 Representación de la estimación del poder espectral PSD por medio de Welch del evento TB, (Señal 2541159d) (izq.) y Zoom a los 5 Hz. (der.) en el volcán Etna

Picos predominantes entre los 0.5 y 1.5 Hz. La señal comienza a decaer con una suave pendiente a partir de dicha medida hasta los 5 Hz.

b) Volcán Strómboli

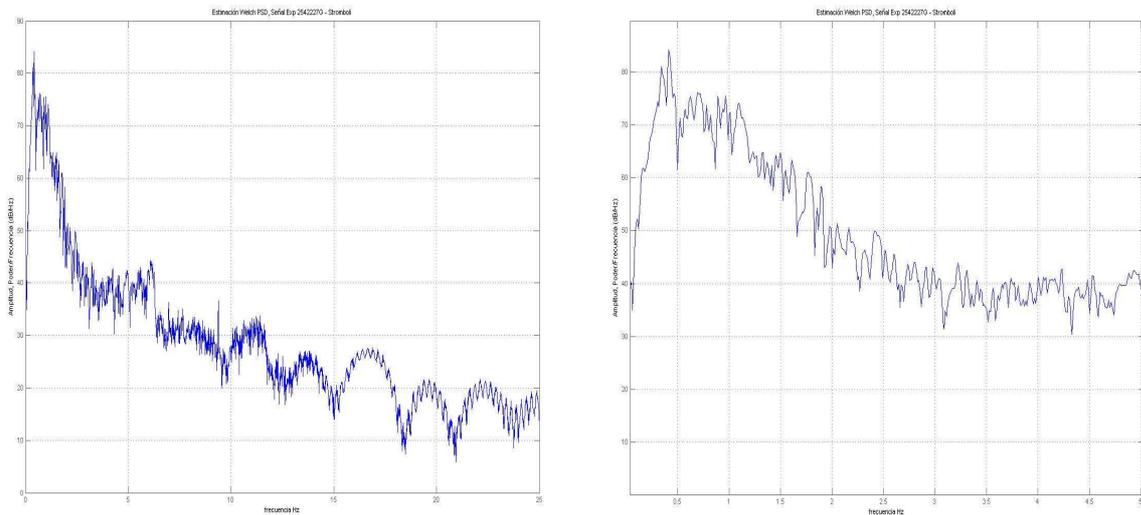


Fig. E.3 Representación de la estimación del poder espectral PSD por medio de Welch del evento Exp, (Señal 2542227G) (izq.) y Zoom a los 5 Hz. (der.) en el volcán Strómboli

Picos predominantes entre los 0.5 y 1.3 Hz. La señal comienza a decaer a partir de dicha medida hasta los 5 Hz. La pendiente de caída es mucho mayor que los TB del Etna.

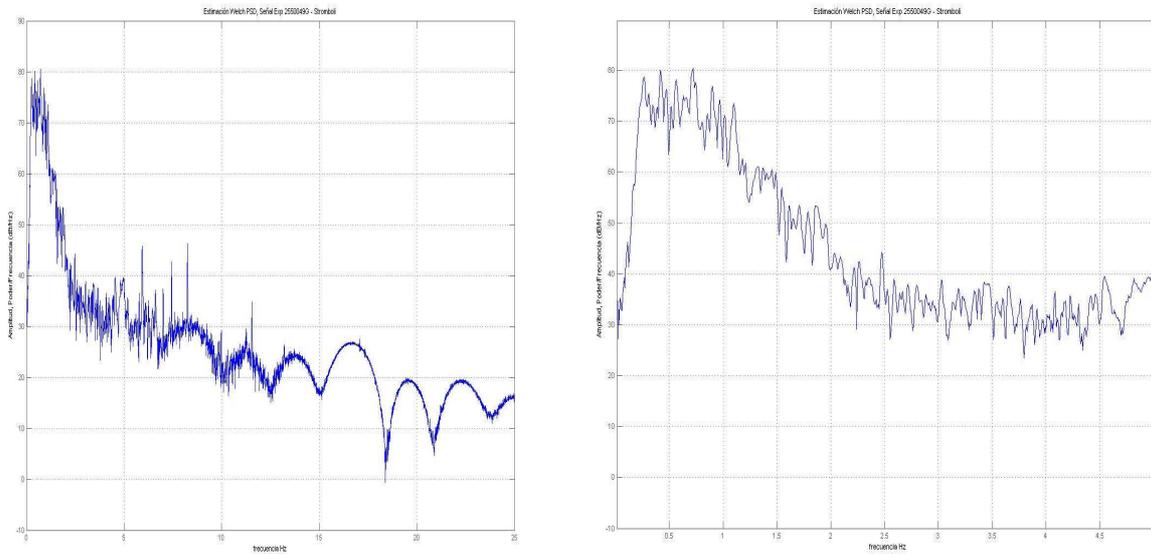


Fig. E.4 Representación de la estimación del poder espectral PSD por medio de Welch del evento Exp, (Señal 2550049G) (izq.) y Zoom a los 10 Hz. (der.) en el volcán Strómboli

Picos predominantes entre los 0.5 y 1 Hz. La señal comienza a decaer a partir de dicha medida hasta los 5 Hz.

c) Volcán San Cristóbal

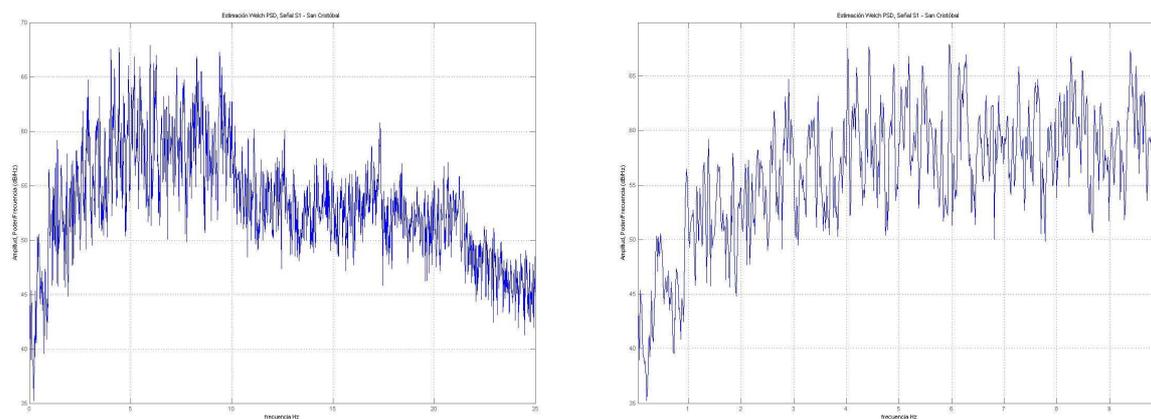


Fig. E.5 Representación de la estimación del poder espectral PSD por medio de Welch del evento S1 (izq.) y Zoom a los 10 Hz. (der.) en el volcán San Cristóbal

Se observa que los picos predominantes de energía se encuentran entre los 4 y 6 Hz, la señal comienza a decaer a los 10 Hz. La curva de energía se concentra entre 1 y 10 Hz. Mucho mayor que los eventos de Etna y Strómboli.

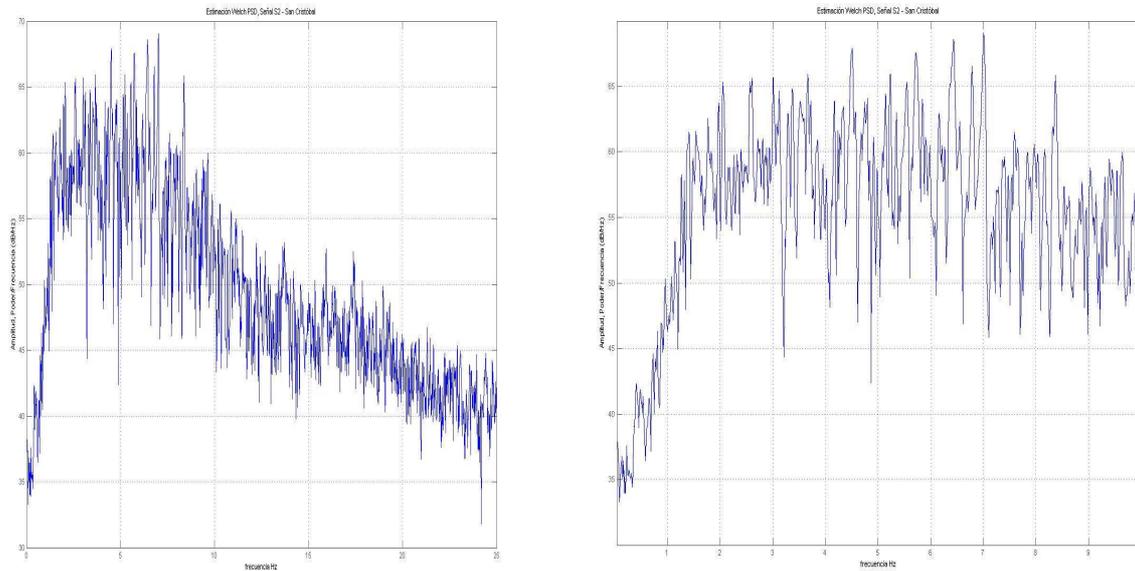


Fig. E.6 Representación de la estimación del poder espectral PSD por medio de Welch del evento S2 (izq.) y Zoom a los 10 Hz. (der.) en el volcán San Cristóbal

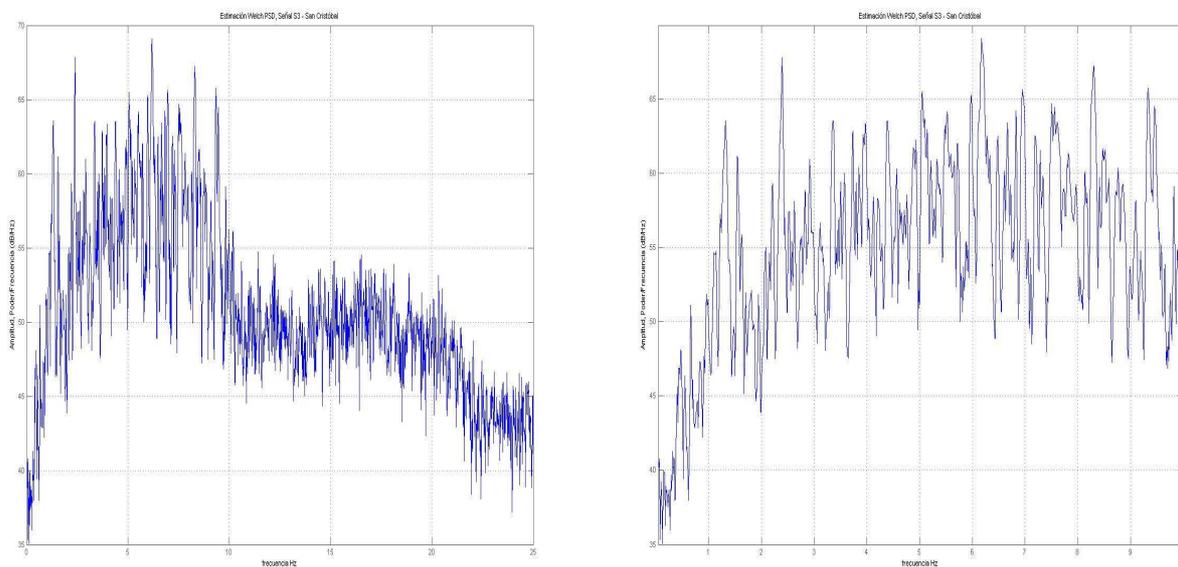


Fig. E.7 Representación de la estimación del poder espectral PSD por medio de Welch del evento S3 (izq.) y Zoom a los 10 Hz. (der.) en el volcán San Cristóbal

En los eventos del San Cristóbal la curva de energía se concentra entre 1 y 10 Hz, comenzando a caer a partir de dicho valor.

d) Caso particular del Tremor continuo del San Cristóbal

Las tres figuras siguientes, muestran un análisis del cálculo del poder espectral de la señal del tremor en el volcán San Cristóbal utilizando esta técnica. Por medio de Periodograma (Fig. E11), por medio de Welch (Fig. E.12) y en la figura E.13 se utiliza el periodograma, para estimar la densidad de poder espectral de la señal con dos componentes de frecuencia y un zoom de 0 a 10 Hz. para ver con mayor claridad el pico máximo y las componentes en esta banda de frecuencia donde se concentra la señal.

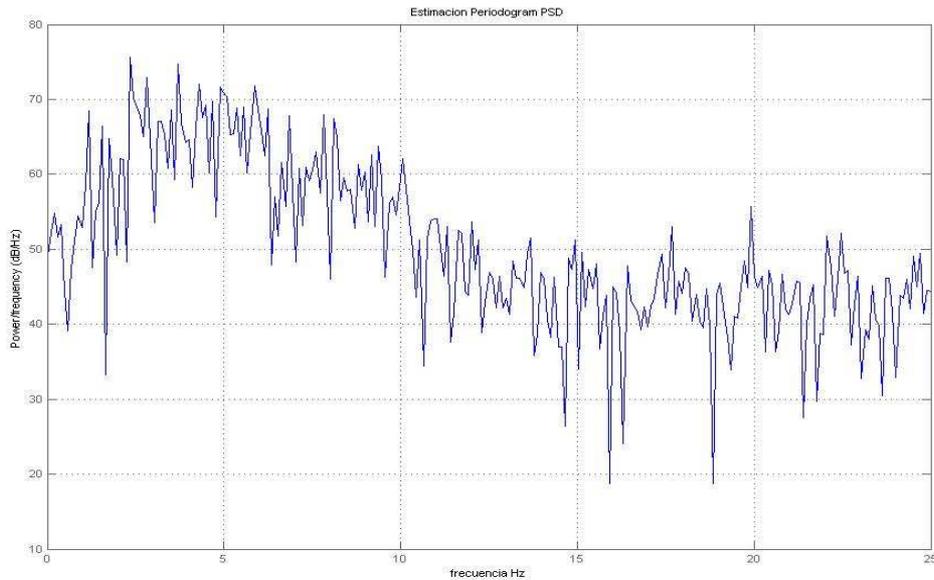


Fig. E.11 Representación de la estimación del poder espectral PSD por medio de Periodograma del Tremor continuo en el volcán San Cristóbal.

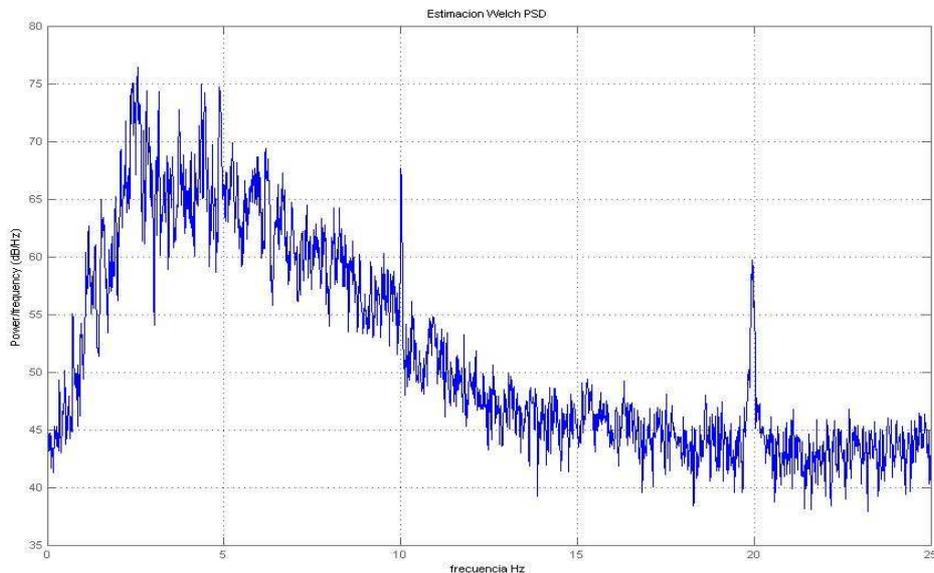


Fig. E.12 Representación de la estimación del poder espectral PSD por medio de Welch del Tremor continuo en el volcán San Cristóbal

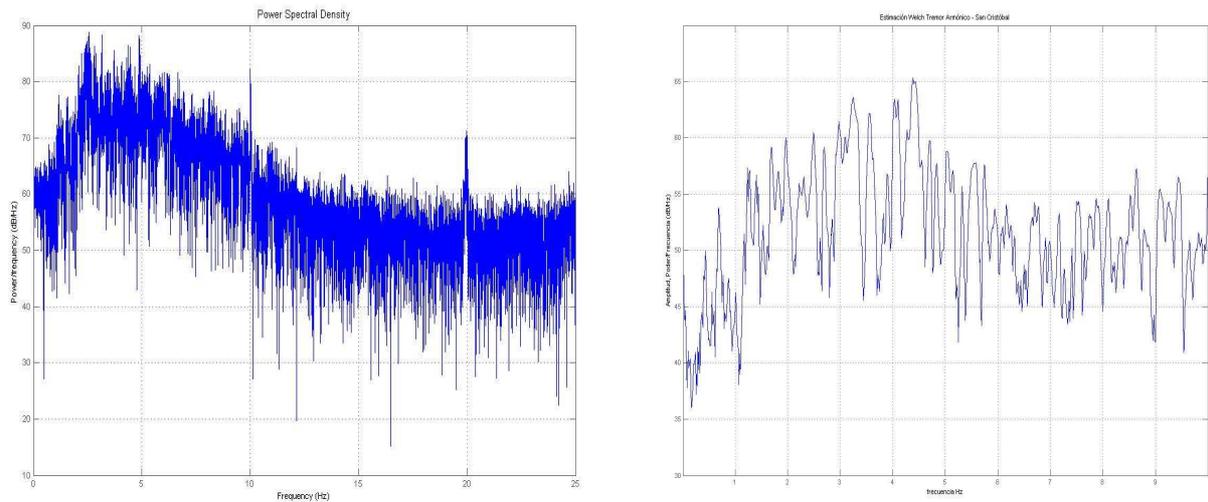


Fig. E.13 Representación de la estimación del poder espectral PSD por medio de Periodograma del Tremor continuo en el volcán San Cristóbal con dos componentes de frecuencia (izq.) y Zoom a los 10 Hz. (der.)

Una observación detallada de las imágenes anteriores del tremor en el San Cristóbal, nos indica que en este tipo de evento en los volcanes de Nicaragua las amplitudes se sitúan en magnitudes de 65 dB, a baja frecuencia en un rango máximo de bandas de 0.2 Hz. a 4.5 Hz. Teniendo un pico máximo predominante en el tremor del San Cristóbal sobre los 4.5 Hz.

e) Volcán Telica

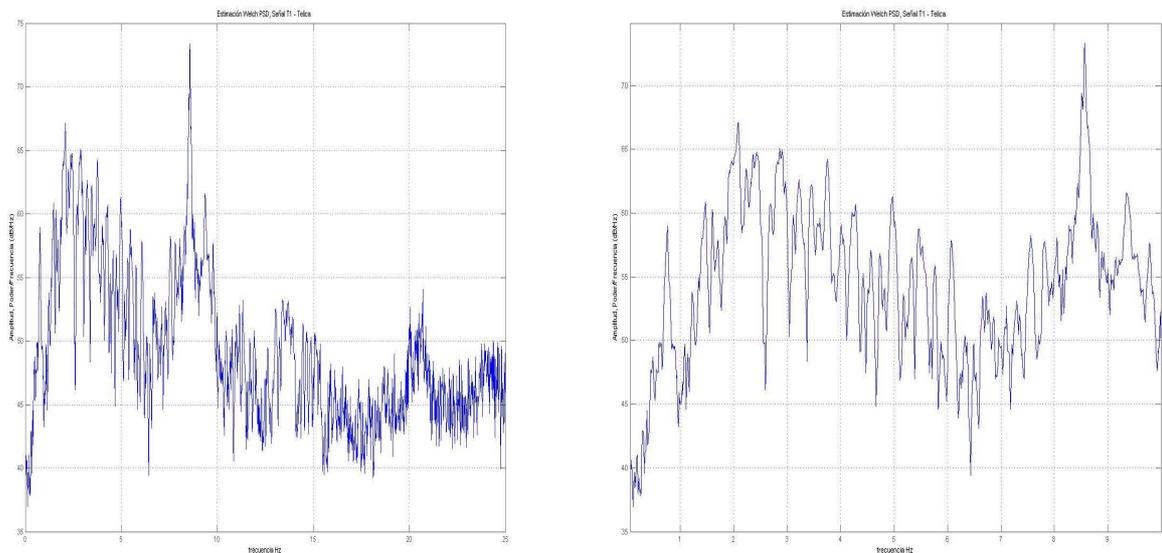


Fig. E.14 Representación de la estimación del poder espectral PSD por medio de Welch del evento T1 (izq.) y Zoom a los 10 Hz. (der.) en el volcán Telica

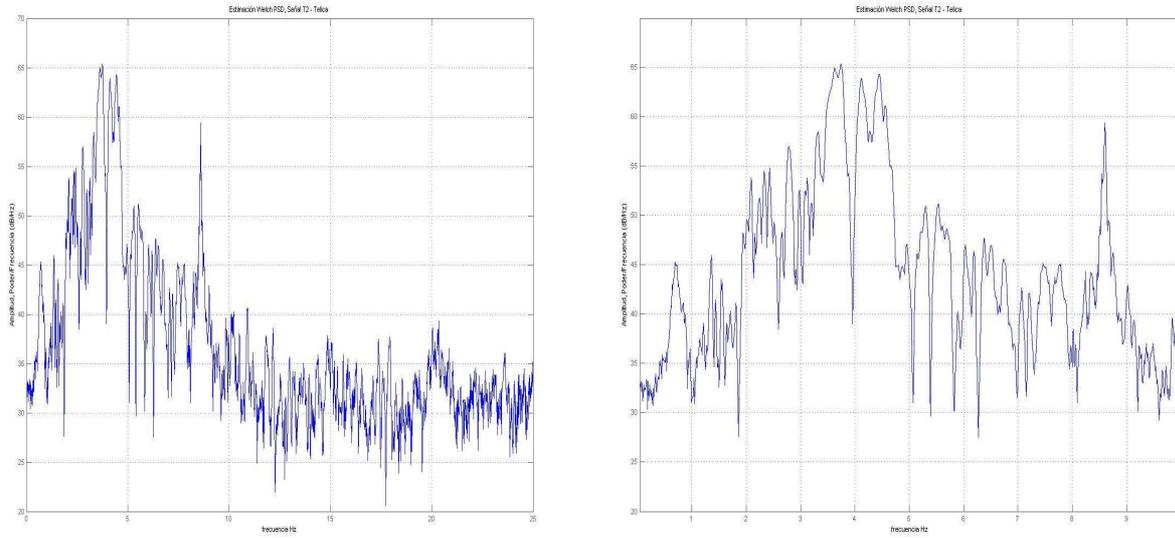


Fig. E.15 Representación de la estimación del poder espectral PSD por medio de Welch del evento T2 (izq.) y Zoom a los 10 Hz. (der.) en el volcán Telica

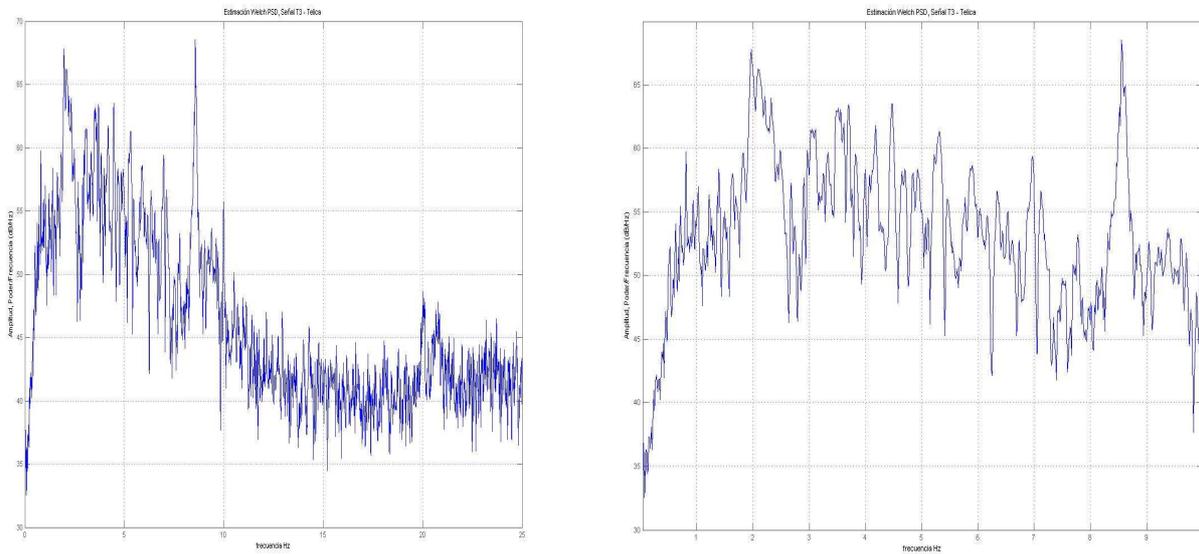


Fig. E.16 Representación de la estimación del poder espectral PSD por medio de Welch del evento T3 (izq.) y Zoom a los 10 Hz. (der.) en el volcán Telica

Los eventos en el Telica se concentran en bandas entre los 3 y 5 Hz. Con pendiente de caída más pronunciada que en el San Cristóbal.

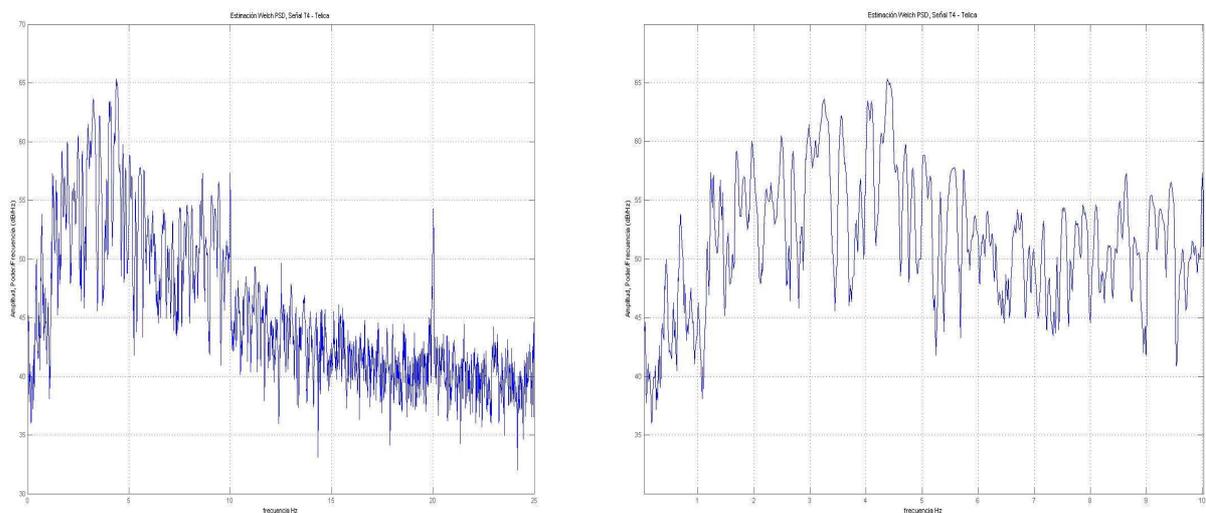


Fig. E.17 Representación de la estimación del poder espectral PSD por medio de Welch del evento T4 (izq.) y Zoom a los 10 Hz. (der.) en el volcán Telica

La información que sobre los eventos se obtiene de las gráficas anteriores se puede representar en las tablas siguientes.

Tipo de Evento		Frecuencia	Magnitud
Etna	Strómboli		
TB		0.2 Hz - 1.3 Hz	75 dB – 80 dB
	Exp	0.4 Hz – 0.6 Hz	80 dB – 84 dB

Tabla E.1 Relación de los componentes en frecuencia y magnitud de los eventos en los volcanes Etna e Strómboli de Italia. Análisis mediante Periodograma Welch.

Tipo de Evento		Frecuencia	Magnitud
San Cristóbal	Telica		
S1		6 Hz.	67 dB
S2		7 Hz.	69 dB
S3		6.2 Hz.	69dB
Tremor armónico		0.2 Hz. – 4.5 Hz	65 dB
	T1	2.1 Hz.	67 dB
	T2	3.7 Hz – 4.5 Hz	65 dB
	T3	2 Hz	67 dB
	T4	4.4 Hz.	65 dB

Tabla E.2 Relación de los componentes en frecuencia y magnitud de los eventos en los volcanes San Cristóbal y Telica de Nicaragua. Análisis mediante Periodograma Welch.

Al realizar una comparación entre la tabla de los volcanes de Nicaragua y la de los de Italia, se observa que las magnitudes de los eventos en los volcanes de Nicaragua son mucho menores que en los volcanes de Italia. Sin embargo, los eventos generados en los volcanes italianos, presentan contenidos más bajos en frecuencia a comparación de los de la cadena volcánica cuaternaria de Nicaragua. Esto también se repite con el tremor. Si se observa la comparación entre el tremor del Etna con el del San Cristóbal, el evento TB del Etna se mantiene a una baja frecuencia hasta 1.3 Hz. Sin embargo, el del San Cristóbal posee varios picos de frecuencia hasta los 4.5 Hz. De igual forma, la magnitud del tremor del Etna es mucho mayor que la del Tremor del San Cristóbal.

En las señales de Italia, las explosiones en el Strómboli se encuentran a más baja frecuencia que el tremor TB en el Etna. Esto es producto del mismo tipo de señales que como antes referíamos es producida por una especie de estallido de una burbuja en la cámara magmática, lo que lo diferencia a las señales de las explosiones en el Strómboli aquí estudiadas. Por otra parte, al observar las formas que presentan la Densidad Espectral de Potencia, se puede observar como las señales entre los cuatro tipos de volcanes: el Etna, el Strómboli, el San Cristóbal y el Telica difieren entre ellas. Por lo que podría deducirse que los eventos difieren entre sí, son eventos con características diferentes. Además, los picos de las señales en Italia se encuentran por debajo de los 5Hz mientras que los eventos en Nicaragua se encuentran sobre los 4 Hz a 6 Hz.

En Nicaragua, los eventos del volcán Telica se encuentran a menores frecuencias que en los del San Cristóbal, influyendo en este sentido el tremor del San Cristóbal que se encuentra presente en casi todas las señales. No así en el San Cristóbal en la que las señales se producen con un bajo tremor de fondo. Esto podría deberse a la composición geológica del San Cristóbal, ya que en el San Cristóbal existen depósitos acuíferos en el edificio volcánico, que de hecho sirven para irrigación de cosechas en la zona. Dichos mantos acuíferos influyen en el tipo de señal que genera el volcán. Mientras que en el Telica no existe este tipo de situación. Esto aún tratándose de volcanes con características geológicas similares y ubicándose en la misma cadena volcánica a relativamente poca distancia entre ellos. Las siguientes secciones realizarán un análisis similar, pero con las diferentes características espectrales del espectro de la señal. Esto va a permitir definir mucho mejor, las características espectrales de las señales sísmico-volcánicas de los volcanes y poder así catalogarlas con mucha mayor facilidad.

3.- Análisis wavelet de los eventos

Otra herramienta complementaria de análisis es hacer uso de las transformadas Wavelets. Esto podría ser de mucha utilidad a la hora de poder definir los diferentes tipos de eventos al estar construyendo las bases de datos de entrenamiento, así por ejemplo al utilizar las variaciones observadas en los espectrogramas, puede realizarse con mucha mayor confianza, el establecimiento de los algoritmos de clasificación automática de los eventos sísmico-volcánicos presentes en las bases de datos que se construyan. También estos algoritmos que en el presente trabajo están basados en los HMMs (Modelos Ocultos de Markov), permiten identificar de manera satisfactoria los eventos Vulcano-tectónicos, tremor y eventos de otros tipos como Tipo-B. En algunos casos se consiguen identificar ciertos tipos de explosiones en función de su contenido espectral, pero en otras, el parecido entre los eventos con los espectrogramas de sismos tipo-B hace bastante difícil esta clasificación. Este problema es intrínseco a la naturaleza de los eventos Tipo-B y a lo que en la actualidad se entiende por ellos, que en muchos casos no poseen mucha diferencia con un evento explosivo. Para resolver muchos más casos, habría que mejorar la precisión de dichos algoritmos, ya sea añadiendo más eventos o modelos. En este caso para las señales sísmico-volcánicas podría utilizarse la Wavelet continua (CWT) de tipo Morlet. Para este análisis, se ha utilizado una escala de coeficientes de 1:32 a lo largo del tiempo de duración de la señal (15022 para los tres minutos).

Los espectros y las energías observadas en las figuras representadas por las transformadas Wavelet, indican la diferencia que existe al observar cada señal. En las señales sismo-volcánicas que aun al observarlas, tienden a confusión al utilizar solo análisis con las transformadas de Fourier, los espectrogramas y con los espectros de potencia. Al utilizar transformadas Wavelet dichas señales, podrían quedar más claramente definidas. Por ejemplo, podría decirse que en el San Cristóbal, el evento S1 es más constante en el tiempo, pudiendo tratarse tanto de un evento de tipo LP (largo período o de un terremoto), y de manera similar, el evento S3 es de mucha más corta duración en el tiempo, aunque posee energía con concentraciones más altas que el S1. Asimismo, el evento S2 difiere de los dos anteriores, destacando la forma típica de una explosión. Si además se observan las transformadas Wavelet 3D, estas diferencias se pueden resaltar aun más. Definiendo así las diferencias entre los eventos y pudiéndolos agrupar de acuerdo a un específico tipo en cada caso.

De la misma forma que los eventos del San Cristóbal, los del Telica presentan significativas diferencias en sus formas espectrales. Este patrón se sigue al observar la representación en 3D. Por lo tanto, la ventaja respecto al espectrograma, es que ofrece información en tiempo-frecuencia instantánea para cada punto de la señal. Así se puede analizar los eventos localizando discontinuidades o picos de corta duración, que son más difíciles de observar únicamente analizando los resultados con la FFT y el espectrograma.

A continuación, se presentan varios ejemplos de la aplicación de la CWT y su representación en 3D, en donde se observan las discontinuidades en frecuencia de los distintos tipos de eventos sísmico-volcánicos, pertenecientes a los cuatro volcanes estudiados en el presente trabajo.

a) Wavelet de los eventos en el Etna

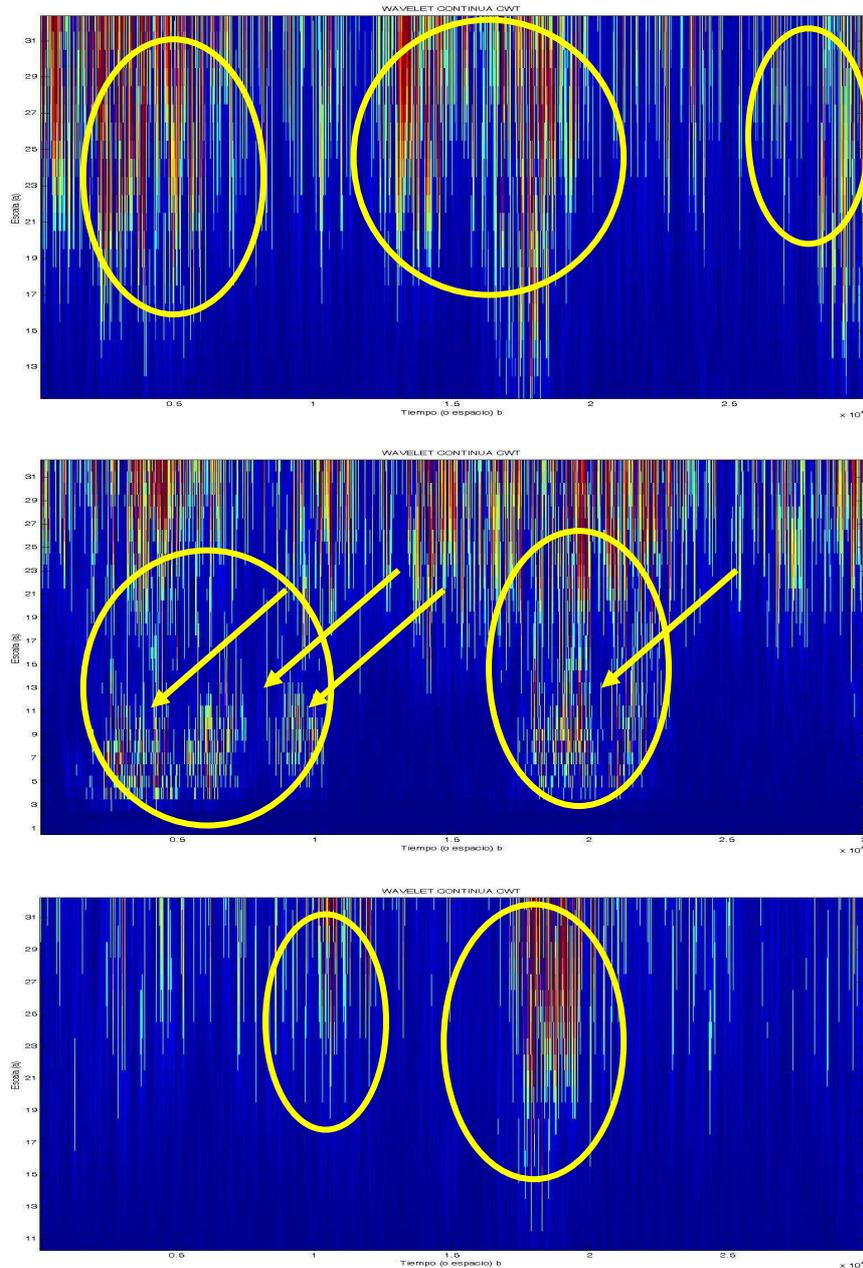


Fig. E.18 Wavelet continua (CWT) de eventos de tipo TB en tres registros del Etna. Los círculos representan los eventos TB y las flechas, los puntos donde se concentra la mayor componente energética.

En las gráficas se pueden observar dentro del tremor los eventos. La imagen del centro presenta más detalladamente las energías de un evento a más baja amplitud que los anteriores que se encuentran enmascarados dentro del mismo tremor de fondo del Etna. Las líneas indican las discontinuidades presentes en dicho evento. Una representación en 3D de estos mismos eventos se observa en la siguiente imagen.

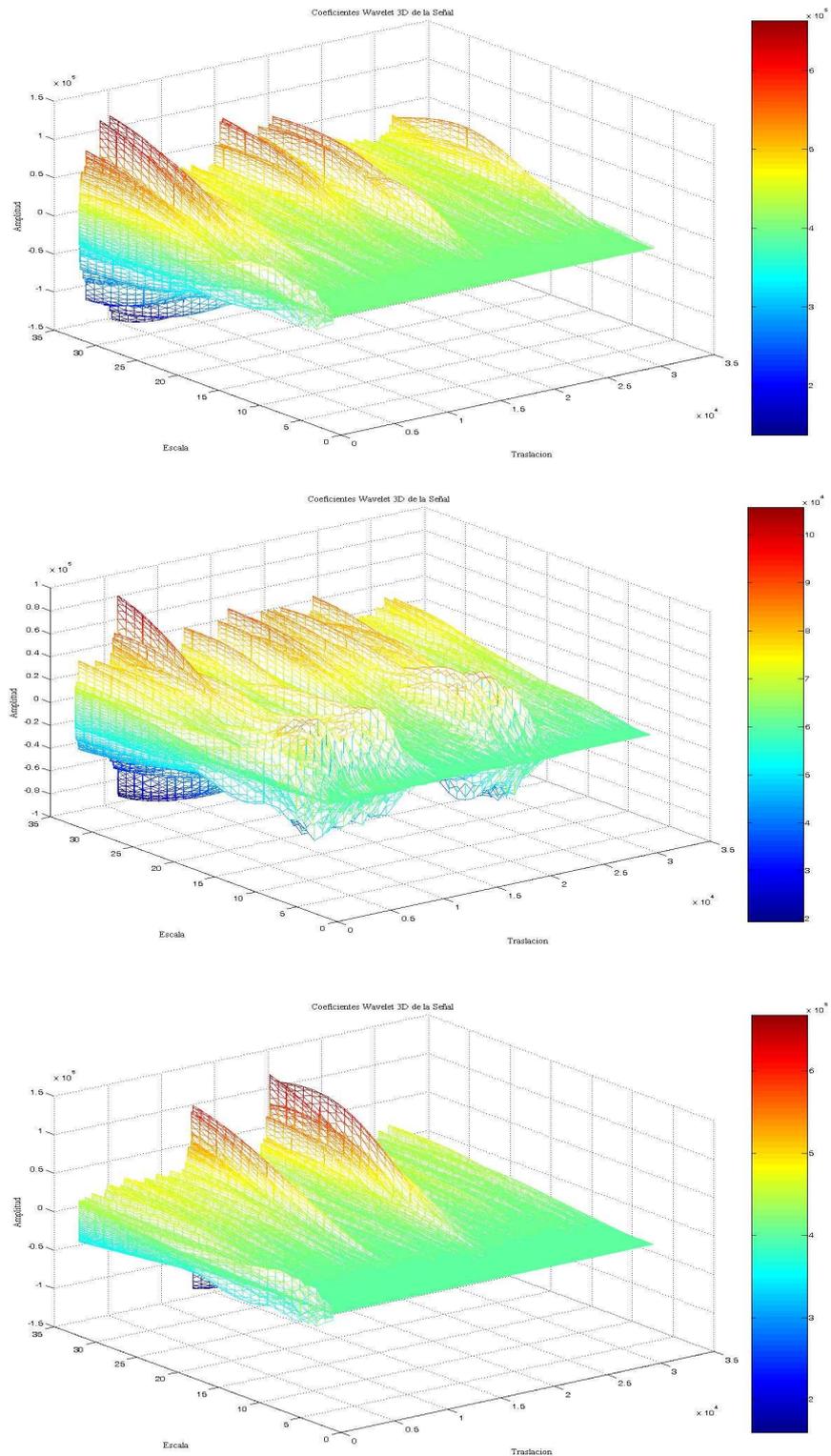


Fig. E.19 Wavelet 3D de eventos de tipo TB en tres registros del Etna. Las formas de ondas muestran diversos aspectos de las burbujas por cómo es llamado y que conforman este tipo de evento.

Se observa en las gráficas anteriores los eventos asociados en el tremor. La gráfica central muestra como las discontinuidades se ven mucho mejor definidas en este registro que en los dos anteriores. Sin embargo, la mayoría de los eventos en el Etna son

eventos asociados al tremor de fondo. De ahí la dificultad de poder identificarlos. A continuación, se observa el valor absoluto de estas gráficas, en donde se pueden detallar mucho más cada evento.

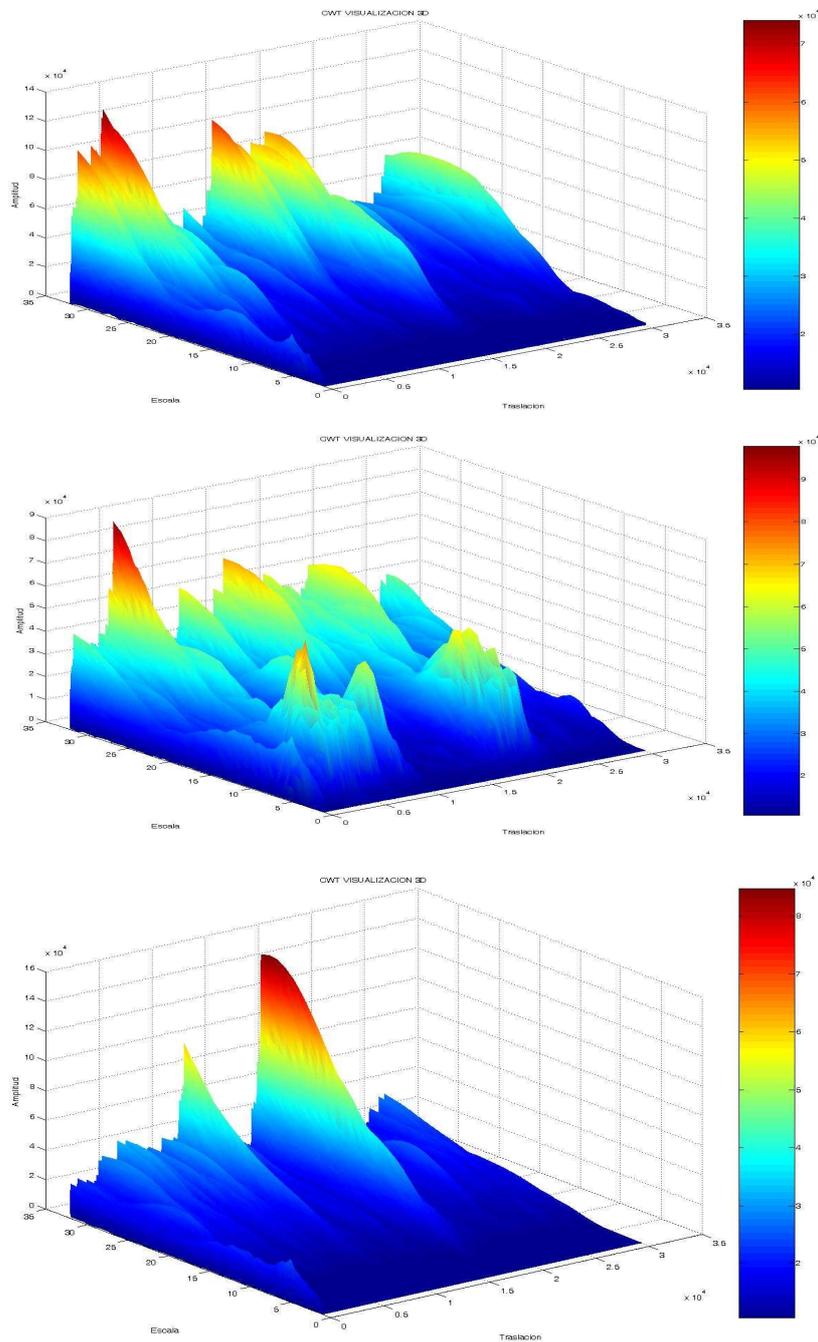


Fig. E.20 CWT 3D valor absoluto (abs.) de eventos de tipo TB en tres registros del Etna. Se puede observar la complejidad de este tipo de evento en las formas de sus picos de energía.

Se observan claramente definidos los eventos en las gráficas anteriores. Comparándolos con las graficas de espectrogramas, puede observarse que un análisis Wavelet podría contribuir a definir con mucha más claridad las energías contenidas en un evento y diferenciarlo de otro.

b) Wavelet de los eventos en el Strómboli.

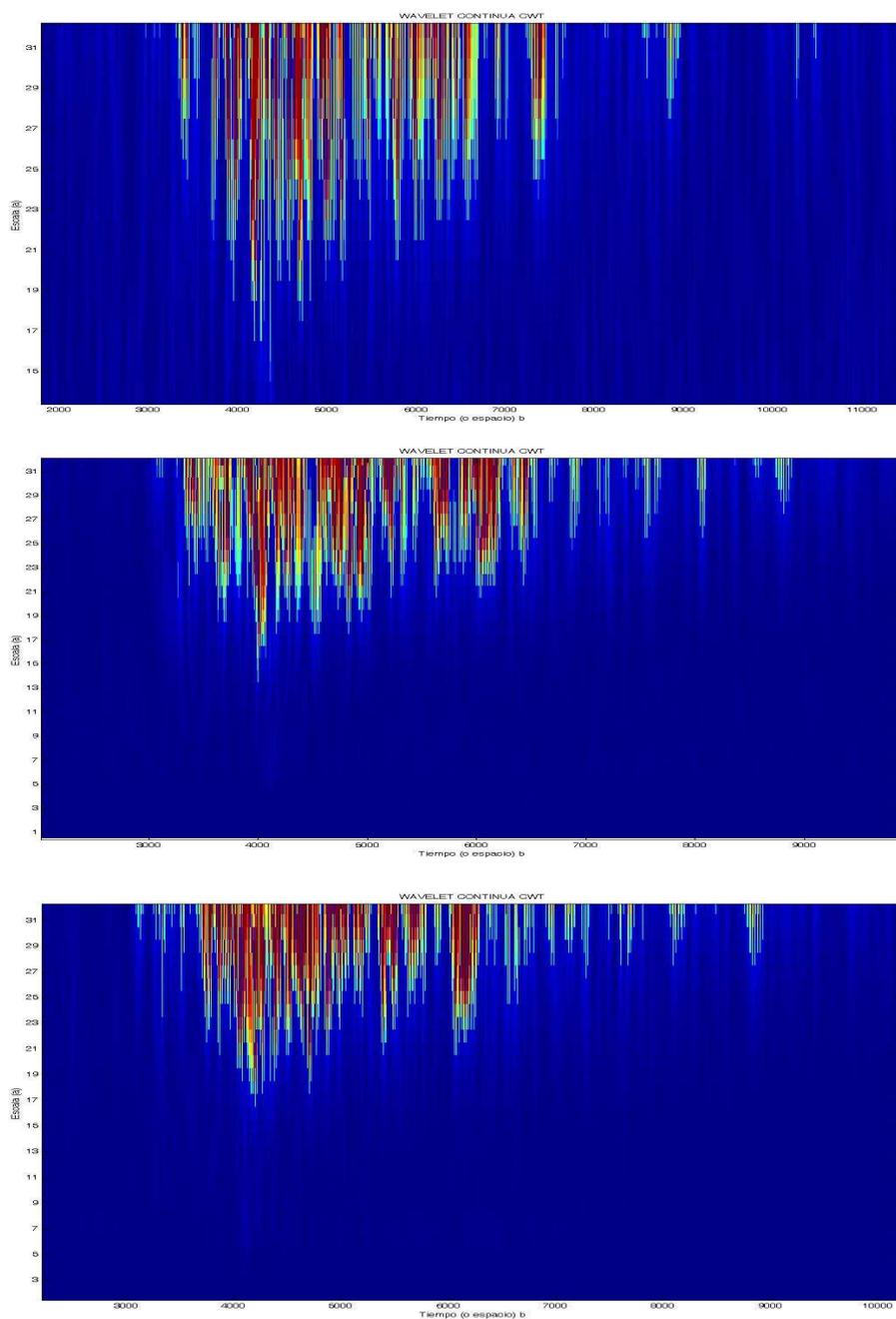


Fig. E.21 Wavelet continua (CWT) de eventos de tipo TB en tres registros del Strómboli. Las explosiones marcan diversas llegadas a muy baja frecuencia.

Al observar las transformadas Wavelet 3D de los anteriores eventos, se obtienen las gráficas E.22 y E.23.

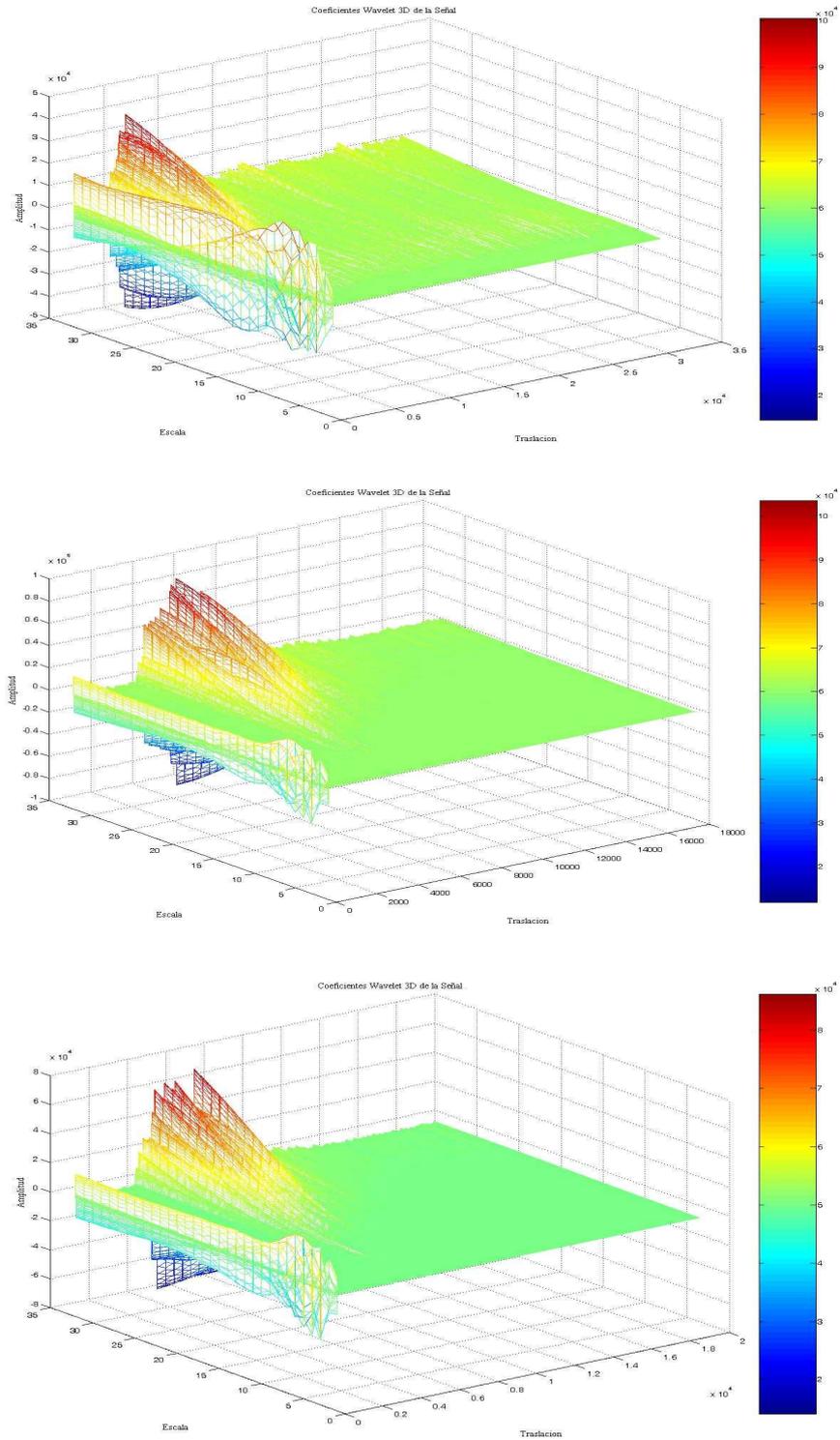


Fig. E.22 Wavelet 3D de eventos de tipo TB en tres registros del Strómboli. Puede verse la forma de onda de las explosiones a baja frecuencia del Strómboli.

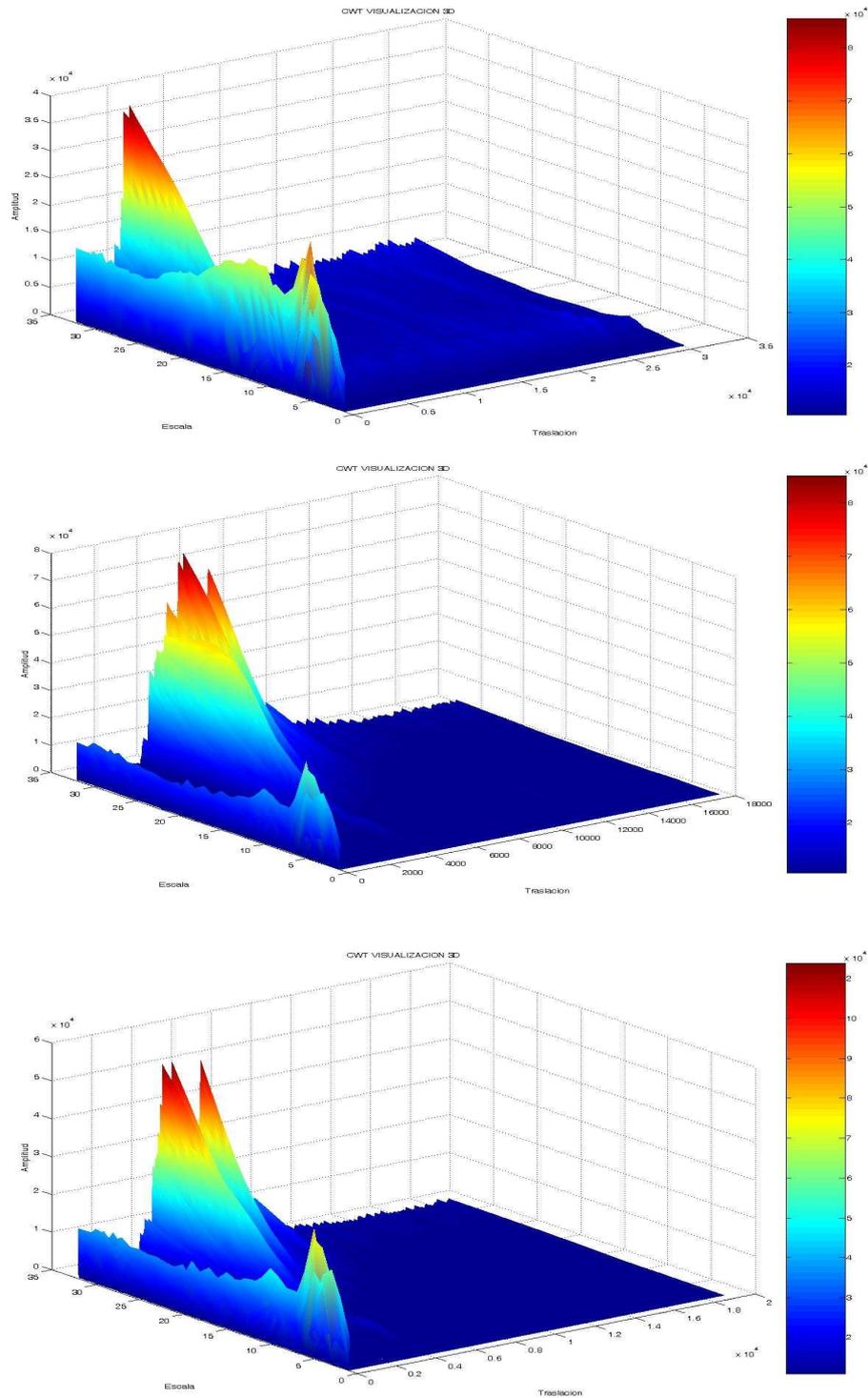


Fig. E.23 CWT 3D valor absoluto (abs.) de eventos de tipo TB en tres registros del Strómboli. Los picos de energía denotan la forma de las explosiones del Strómboli.

Observando las anteriores gráficas, los espectros de las señales de las explosiones en el Strómboli son bastante homogéneos. Esto indica que las explosiones en el Strómboli no difieren mucho entre ellas. Sin embargo, esto mismo no sucede con

el tremor del Etna. Ya se ha comentado la dificultad que radica el estudiar el tremor y más tratándose de un caso especial, como lo es el TB que se ocupa en el presente trabajo. Las imágenes muestran diferentes características del evento TB en el Etna. El tremor del Etna presenta bastante más contenido de frecuencia a lo largo del tiempo que las explosiones en el Strómboli. Por otro lado, comparando este tipo de eventos con los volcanes de Nicaragua, se observa que las señales son claramente diferentes. Esto debido a que son producidas por volcanes con características diferentes. De hecho, esto ya indica algo, que el poder generalizar la clasificación e identificación de estas señales en una sola base de datos va a resultar muy difícil. Esto se puede demostrar, al establecer la combinación de dichas bases de datos en el sistema de clasificación, mediante el uso de los modelos matemáticos que trata la Tesis. Los análisis Wavelet con los eventos de Nicaragua se observan en las siguientes figuras

c) Wavelet de los eventos en el San Cristóbal

- i) Evento tipo S1: La figura E.36 muestra la representación de la Wavelet continua (CWT), del evento S1.

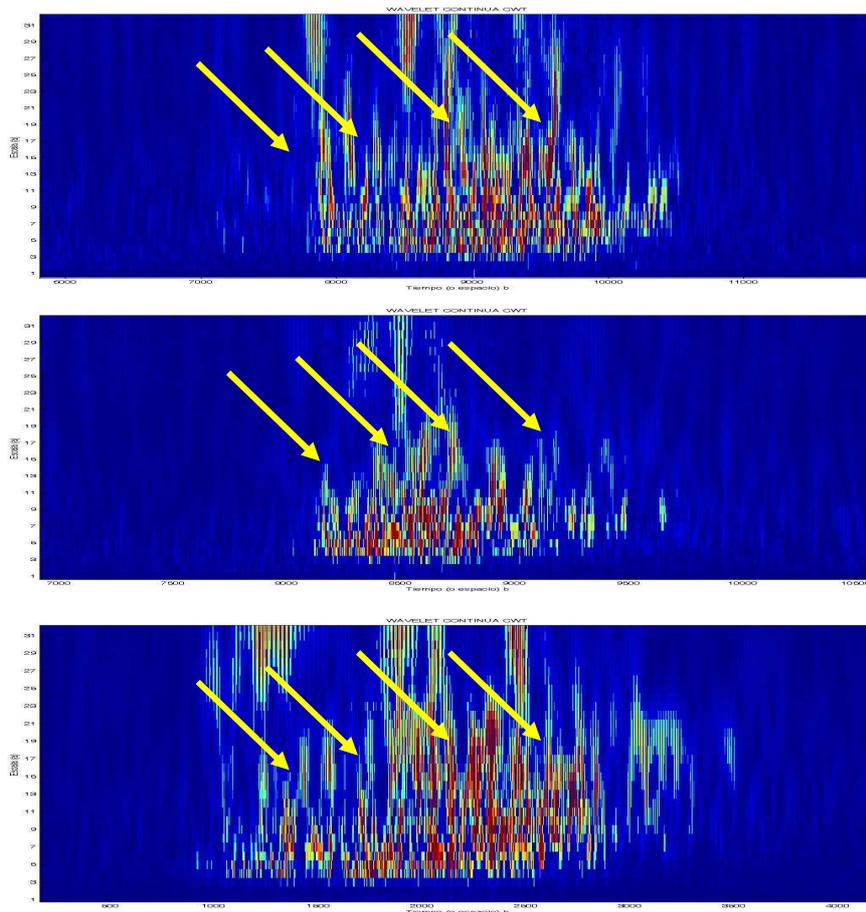


Fig. E.24 Wavelet continua (CWT) de eventos tipo S1 en tres registros del San Cristóbal. Las flechas indican múltiples picos en frecuencia, característicos de este tipo de evento. La caída en frecuencia en el tiempo es más prolongada.

En la figura anterior, pueden observarse las diversas discontinuidades presentes en este tipo de evento. Otro aspecto que puede ser resaltado es la representación tridimensional (3D) del análisis de las Wavelet., y su valor absoluto. Las figuras E.25 y E.26, muestran las señales con una representación de este análisis.

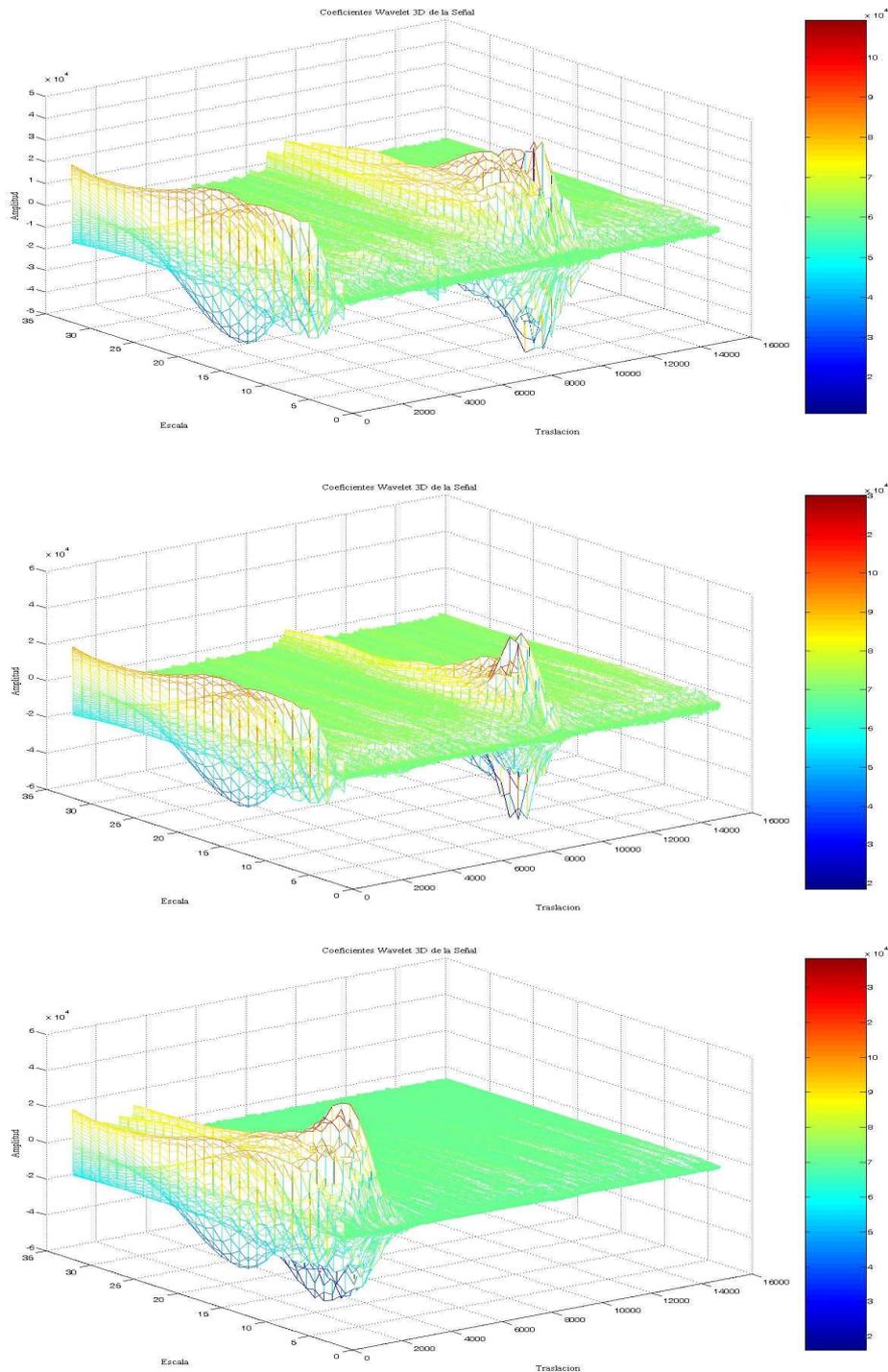


Fig. E.25 Wavelet 3D de eventos tipo S1 en tres registros del San Cristóbal. Puede observarse la forma irregular de las ondas que conforman el evento.

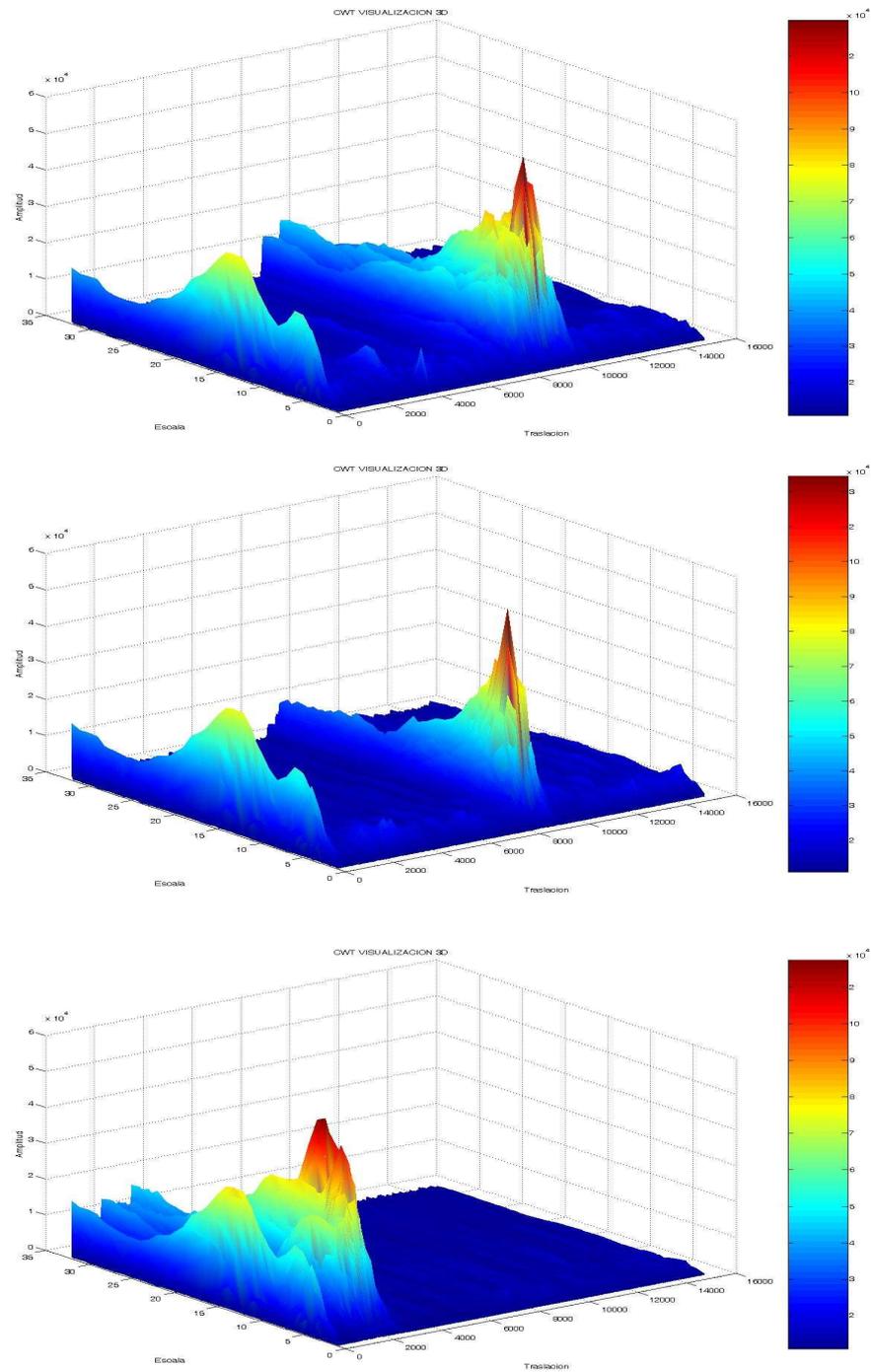


Fig. E.26 CWT 3D valor absoluto (abs.) de eventos de tipo S1 en tres registros del San Cristóbal. La forma de un pico predominante de energía al cual confluyen el resto de llegadas es característica en este tipo de evento.

Obsérvese en la gráfica, la semejanza en la forma (punta) de los máximos energéticos de este evento.

- ii) Evento tipo S2: La figura E.27 muestra la representación de la Wavelet continua (CWT), del evento S2.

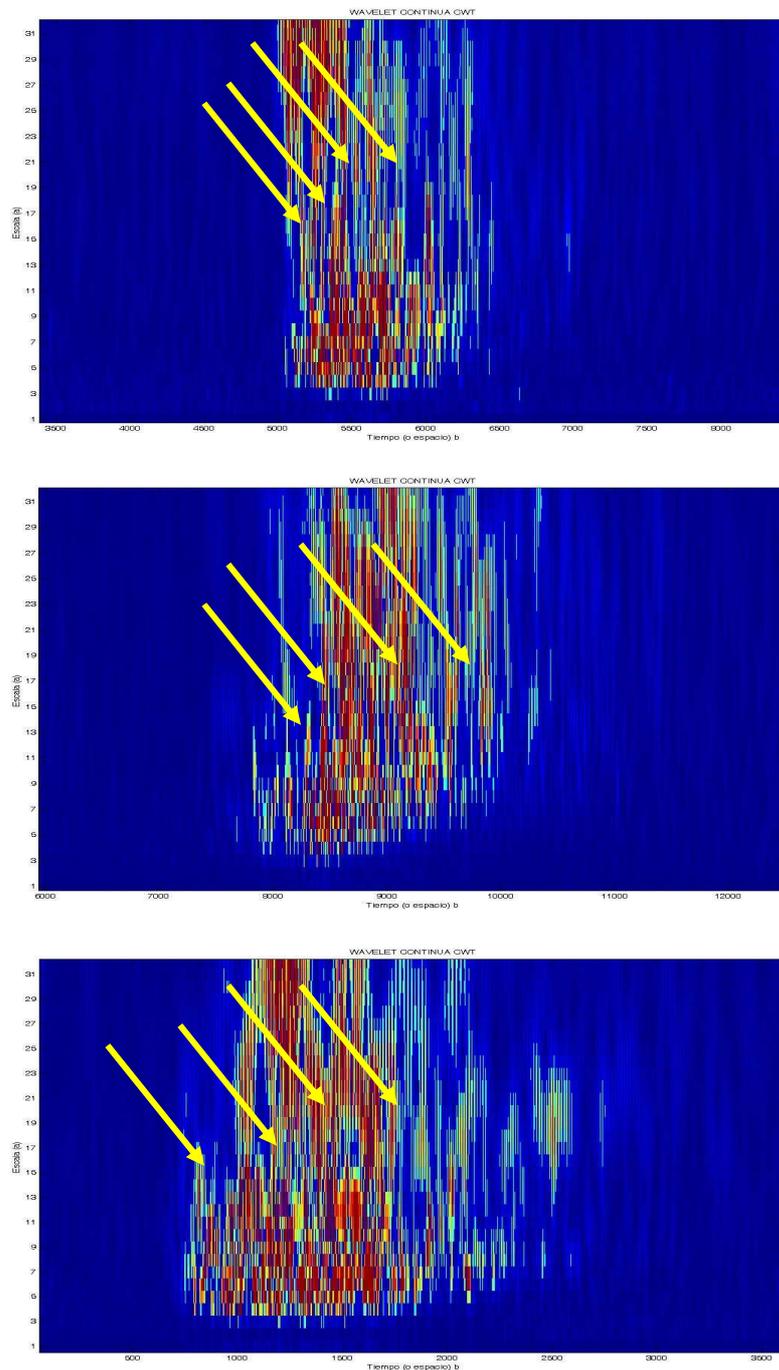


Fig. E.27 Wavelet continua (CWT) de eventos tipo S2 en tres registros del San Cristóbal. Al igual que el evento anterior, las flechas indican que este tipo de evento posee múltiples picos de llegadas en energía. A diferencia del evento anterior la caída en frecuencia a lo largo del tiempo es mucho menor.

En la imagen anterior, se observa la forma alargada en el tiempo y la presencia de diversas discontinuidades energéticas. La representación en 3D de este evento se observa en las figuras E.28 y E.29.

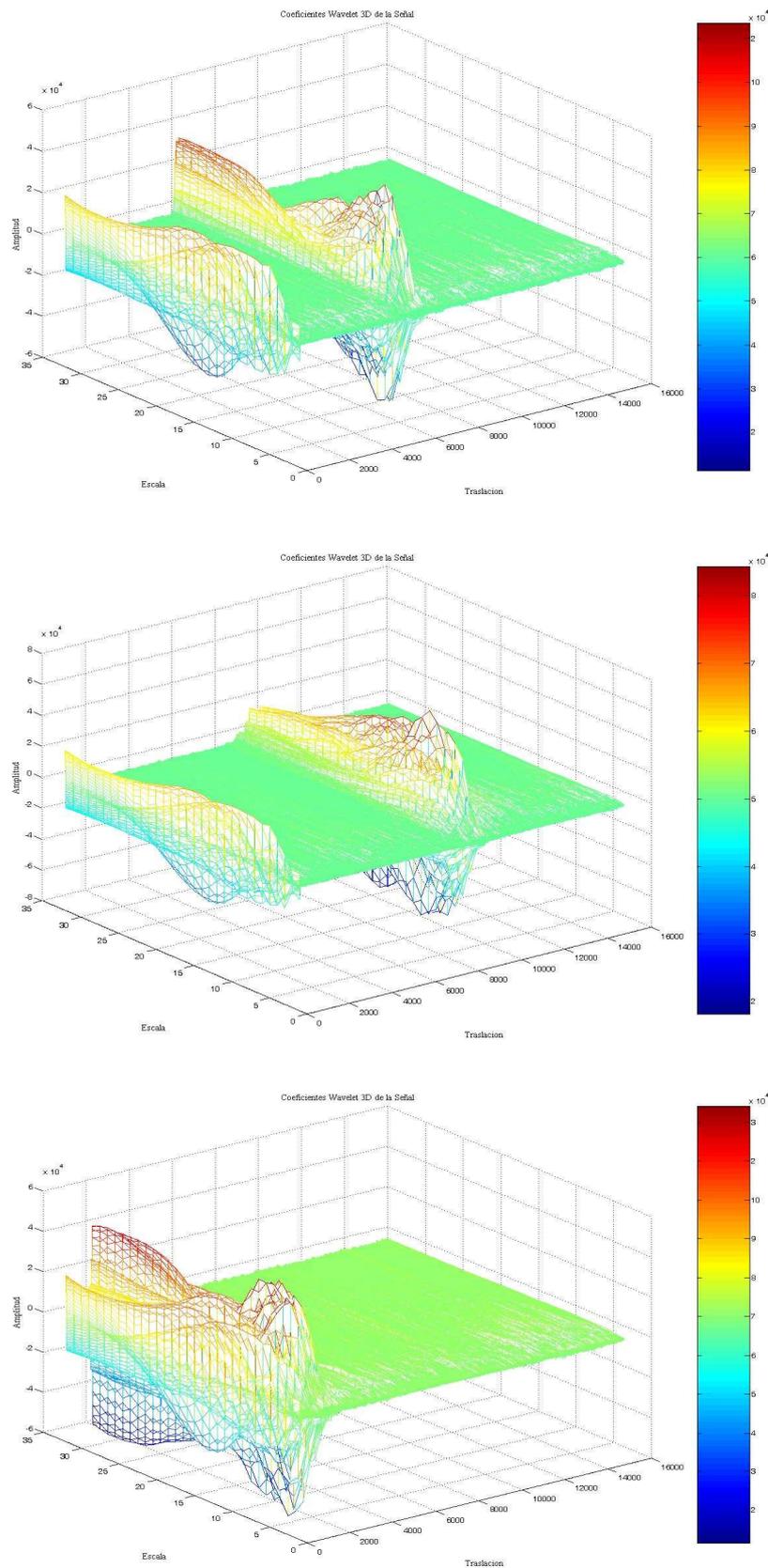


Fig. E.28 Wavelet 3D de eventos tipo S2 en tres registros del San Cristóbal. La forma de ondas en las legadas de energías de este tipo de evento es aún más irregular que del tipo S1.

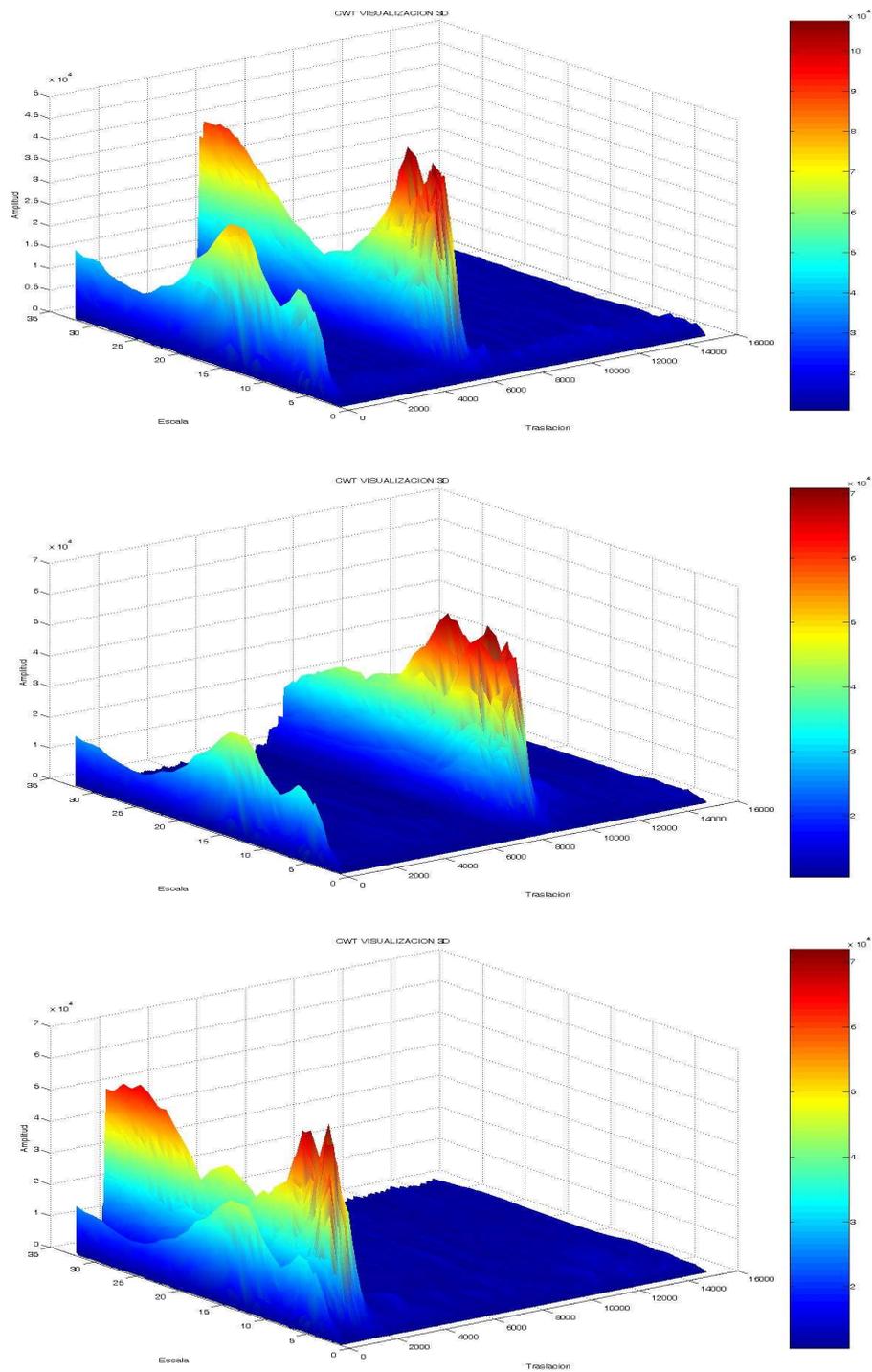


Fig. E.29 CWT 3D valor absoluto (abs.) de eventos de tipo S2 en tres registros del San Cristóbal. En este tipo de evento se pueden observar muchos más picos energéticos en llegadas. Cuanto más irregular sea la forma de onda, más picos energéticos produce.

Se puede observar en la gráfica anterior, la aparición de varios picos con máximos energéticos en este tipo de evento.

- iii) Evento tipo S3: La figura E.30 muestra la representación de la Wavelet continua (CWT), del evento S3.

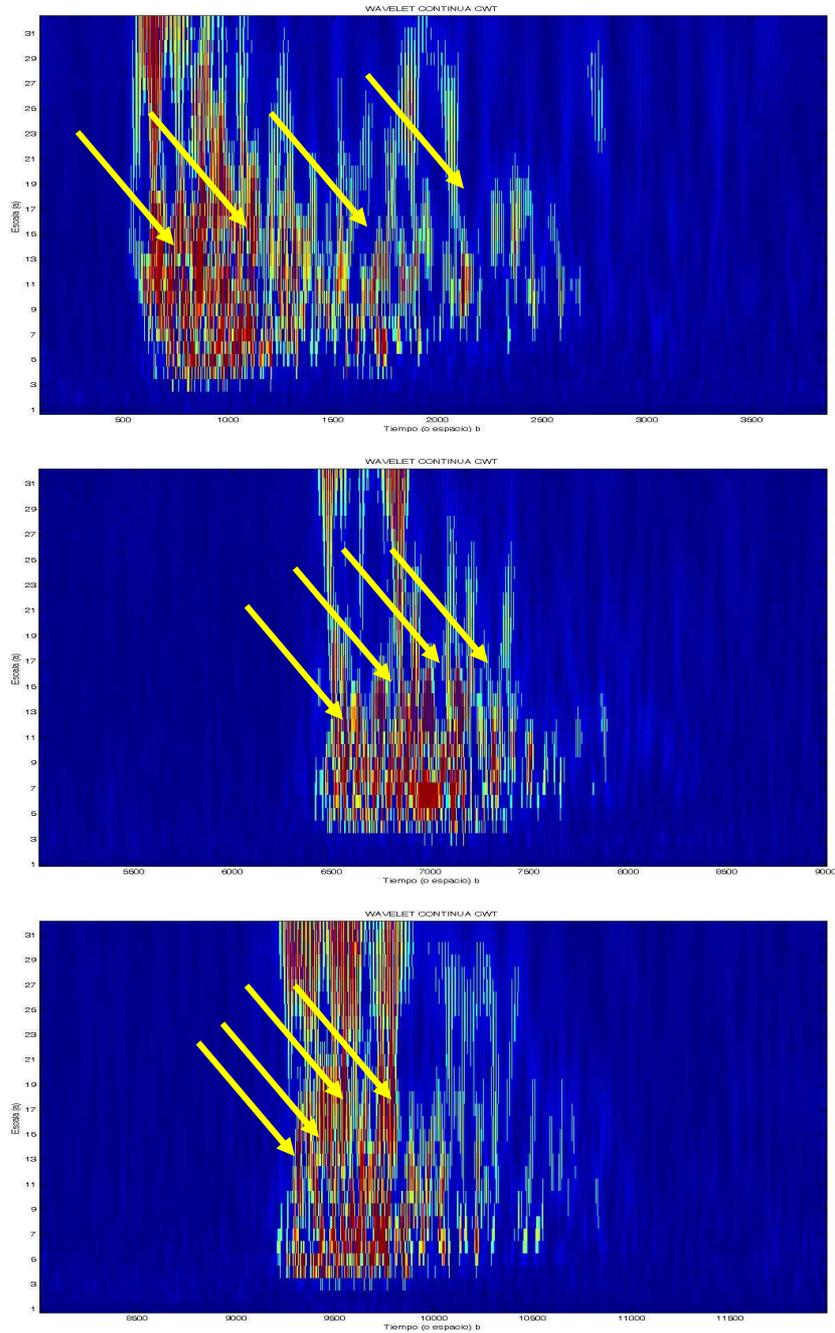


Fig. E.30 Wavelet continua (CWT) de eventos tipo S3 en tres registros del San Cristóbal. Las flechas indican que posee también múltiples llegadas en frecuencia, la caída en frecuencia es más suave que el evento anterior S2, aunque más corta que la del evento S1.

Este tipo de evento presenta una caída más brusca en energía al final, además de que se compone de diversas discontinuidades energéticas a lo largo del tiempo. La representación en 3D de este evento se observa en las figuras E.31 y E.32.

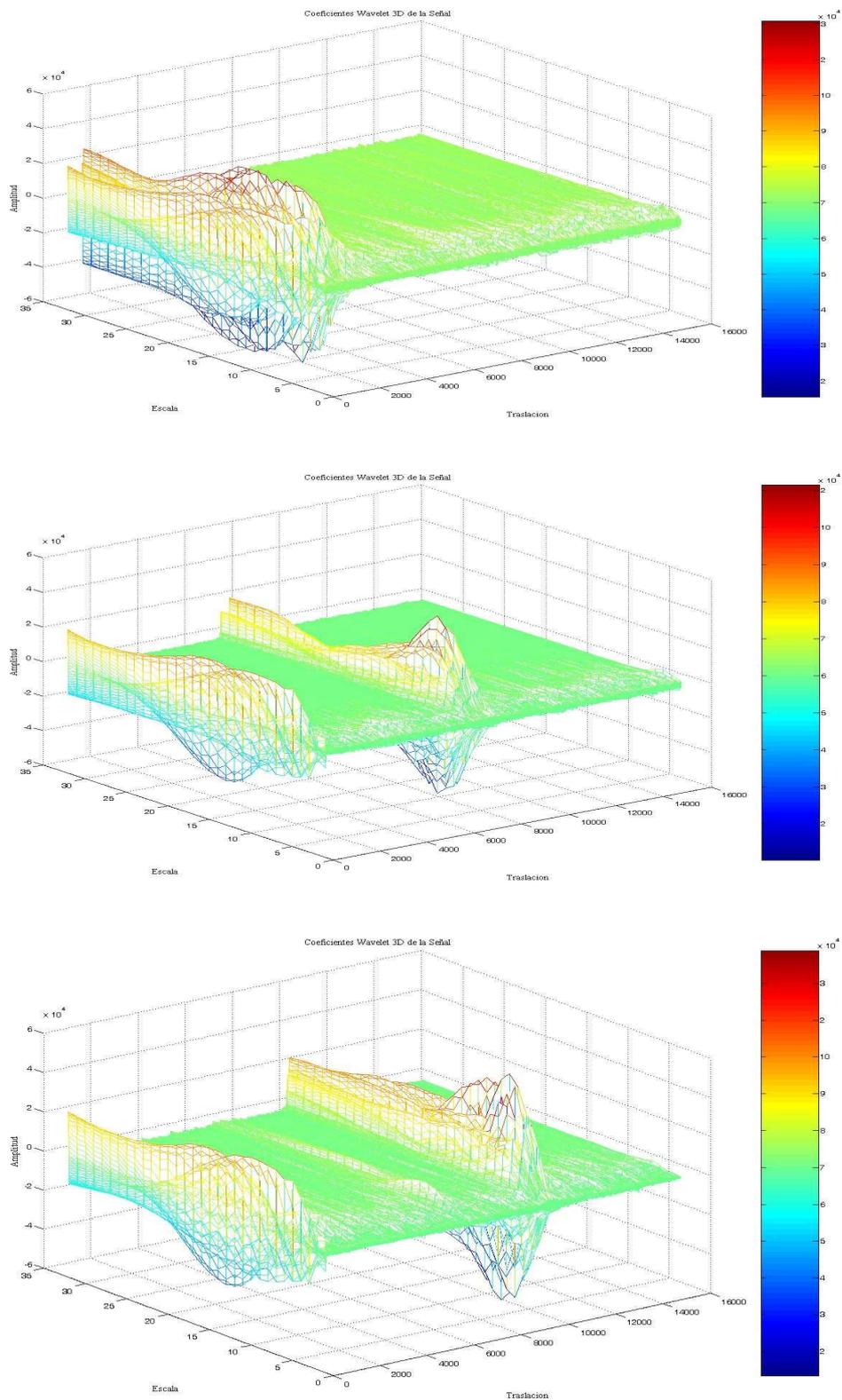


Fig. E.31 Wavelet 3D de eventos tipo S3 en tres registros del San Cristóbal. La forma de onda es similar al de tipo S1, solo que este evento es mucho mayor en el tiempo presentando una caída en energía diferente, lo que se observa en la figura.

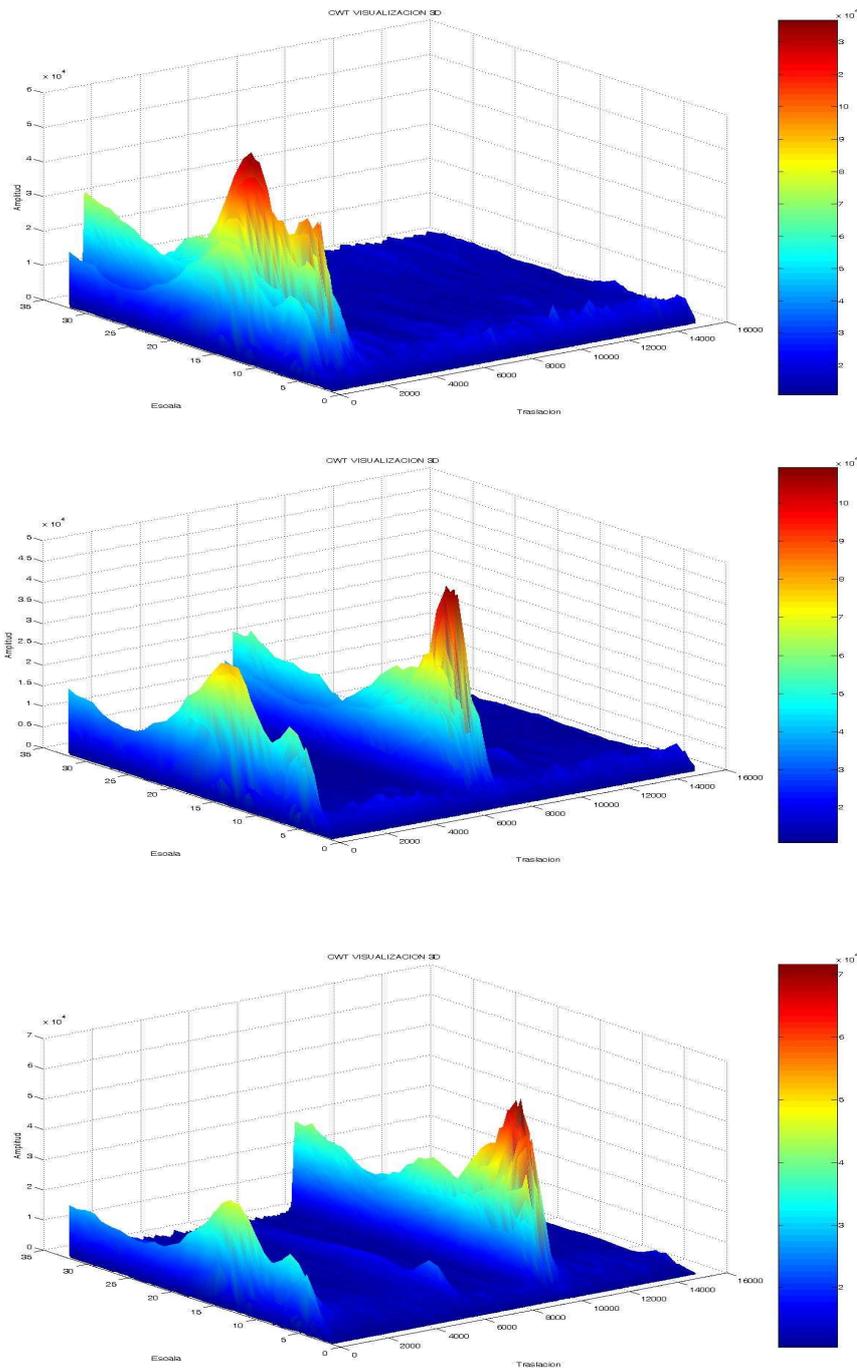


Fig. E.32 CWT 3D valor absoluto (abs.) de eventos de tipo S3 en tres registros del San Cristóbal. Este tipo de evento presenta un componente máximo en energía al que confluyen todas las llegadas, se asemeja al del tipo S1, con la diferencia de la caída en frecuencia al final del evento.

Obsérvese la semejanza en la forma de cono de los máximos energéticos de este evento. Haciendo un breve análisis de las anteriores figuras, una valoración a destacar es que se puede observar la típica figura que diferencia cada uno de los eventos en el San Cristóbal. Existe cierta similitud entre el evento S1 y el S3, sin embargo, el S3 posee menor contenido espectral y duración que el S1. El evento S2 se diferencia de los

otros dos, en que posee una llegada en baja frecuencia seguido de altas energías, no así el S2 y S3 que presentan una llegada de la señal en alta energía.

iv) Caso especial, Wavelet del tremor continuo presente en los registros del San Cristóbal

Un análisis Wavelet de este tipo de señal, muestra con mucha mayor definición las componentes y la forma de la onda de sus armónicos.

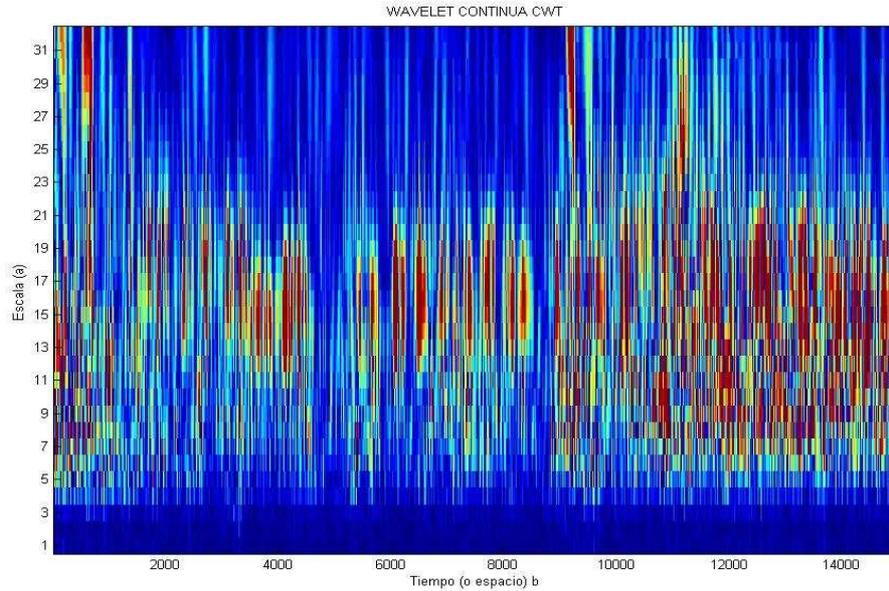


Fig. E.33 Wavelet continua CWT del tremor continuo en el volcán San Cristóbal. Se observa como el tremor está compuesto por múltiples bandas de energía.

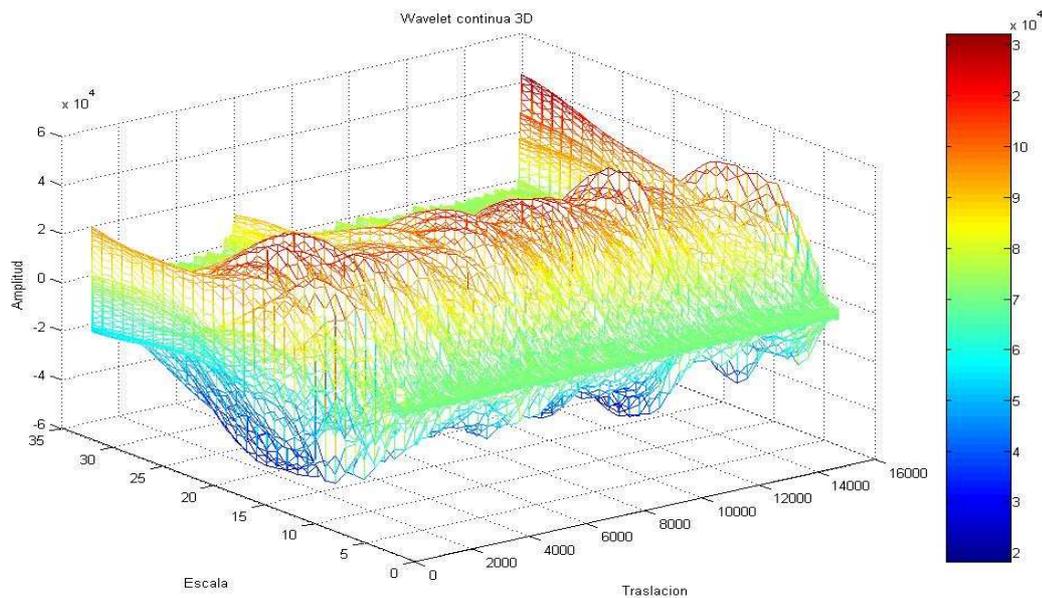


Fig. E.34 Wavelet 3D del tremor continuo presente en el volcán San Cristóbal. Se observa la multiplicidad de ondas en las diversas llegadas de energía de los armónicos presentes en el tremor.

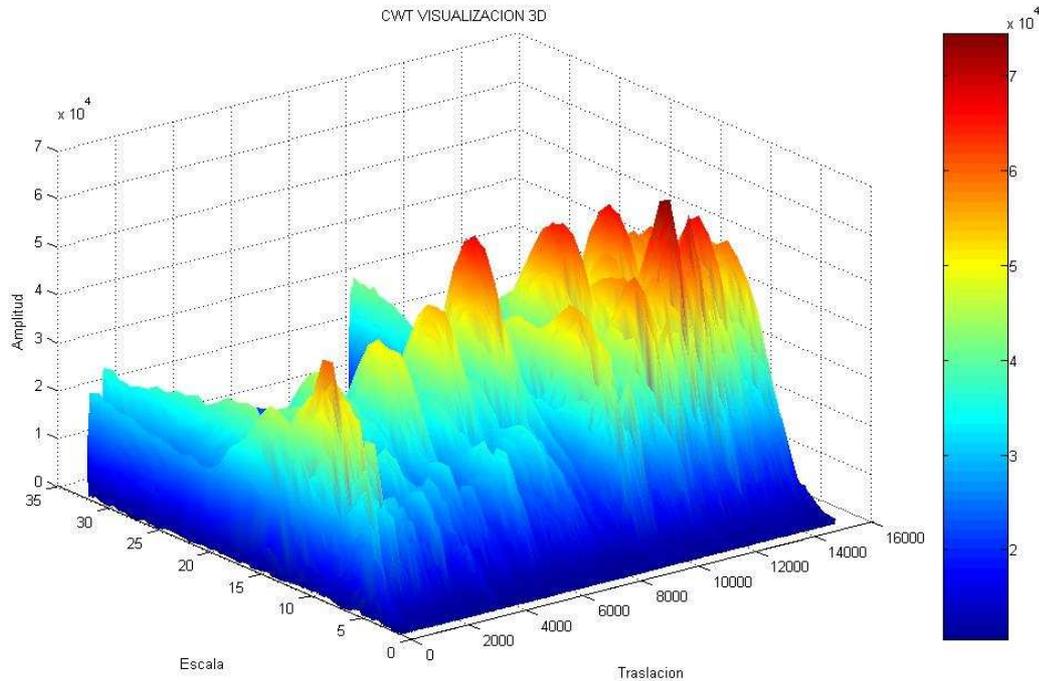


Fig. E.35 Wavelet 3D valor absoluto (abs.) del tremor continuo presente en el volcán San Cristóbal. En la gráfica se observan los diversos picos de energía que componen el tremor del San Cristóbal.

Se observa en las gráficas del evento los diversos máximos energéticos que componen los armónicos del tremor. La figura E.33 muestra una gran variedad de discontinuidades presentes en el evento. En su representación en 3D (figuras E.34 y E.35), pueden observarse dichas discontinuidades, con formas de ondas a lo largo del tiempo (fig. E.34), y presentando los diversos picos de energía que componen los armónicos del tremor (fig. E.35). En la figura se puede observar al menos cuatro bandas de armónicos bien definidos.

d) Wavelet de los eventos en el Telica

Los eventos del Telica tienen características muy particulares, que pueden ser analizadas mediante las transformadas Wavelet, como se observa en las gráficas siguientes:

- i) Evento tipo T1: La figura E.36 muestra la representación de la Wavelet continua (CWT), del evento T1

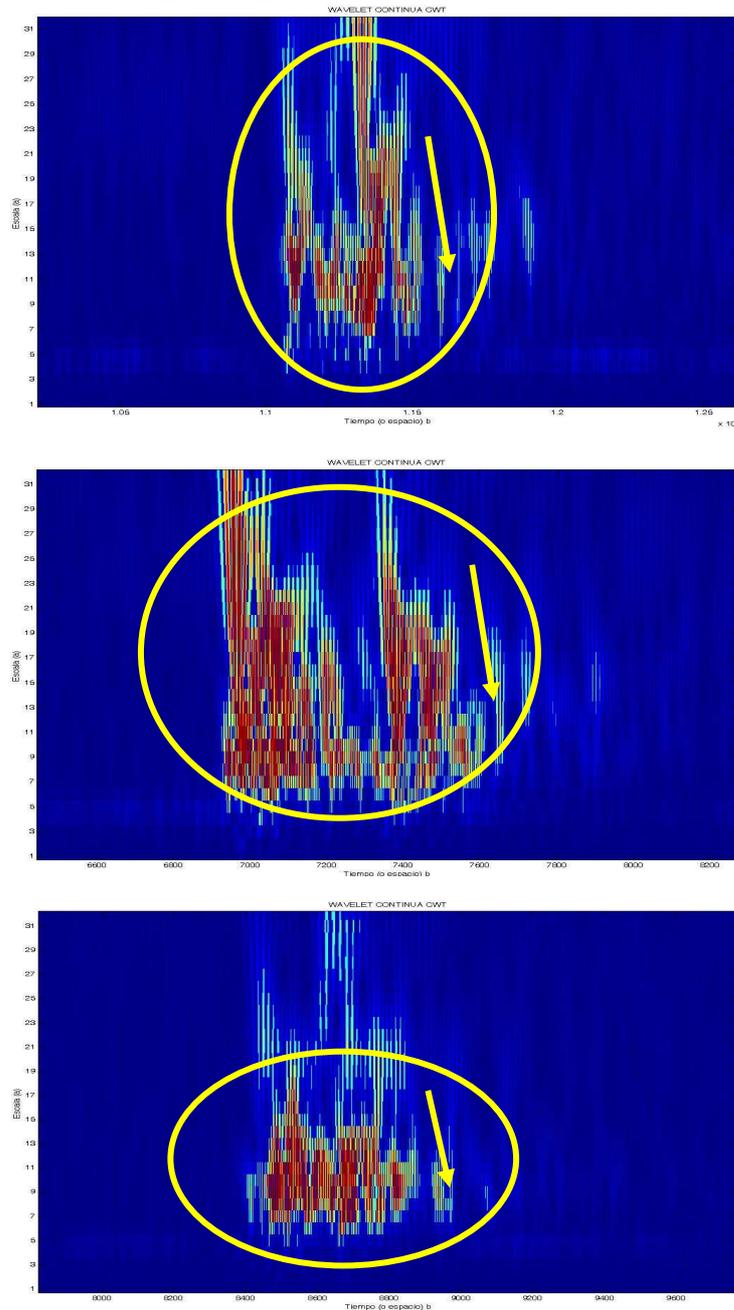


Fig. E.36 Wavelet continua (CWT) de eventos tipo T1 en tres registros del Telica. En el círculo, se pueden observar las dos llegadas energéticas que caracterizan a este tipo de evento como evento doble. Además la flecha indica la forma de la caída en frecuencia de este tipo de evento.

Se observa en la figura anterior, la forma de las componentes del evento, y los diversos picos de máximos energéticos o discontinuidades de este evento. La representación en 3D de este evento se observa en las figuras E.37 y E.38

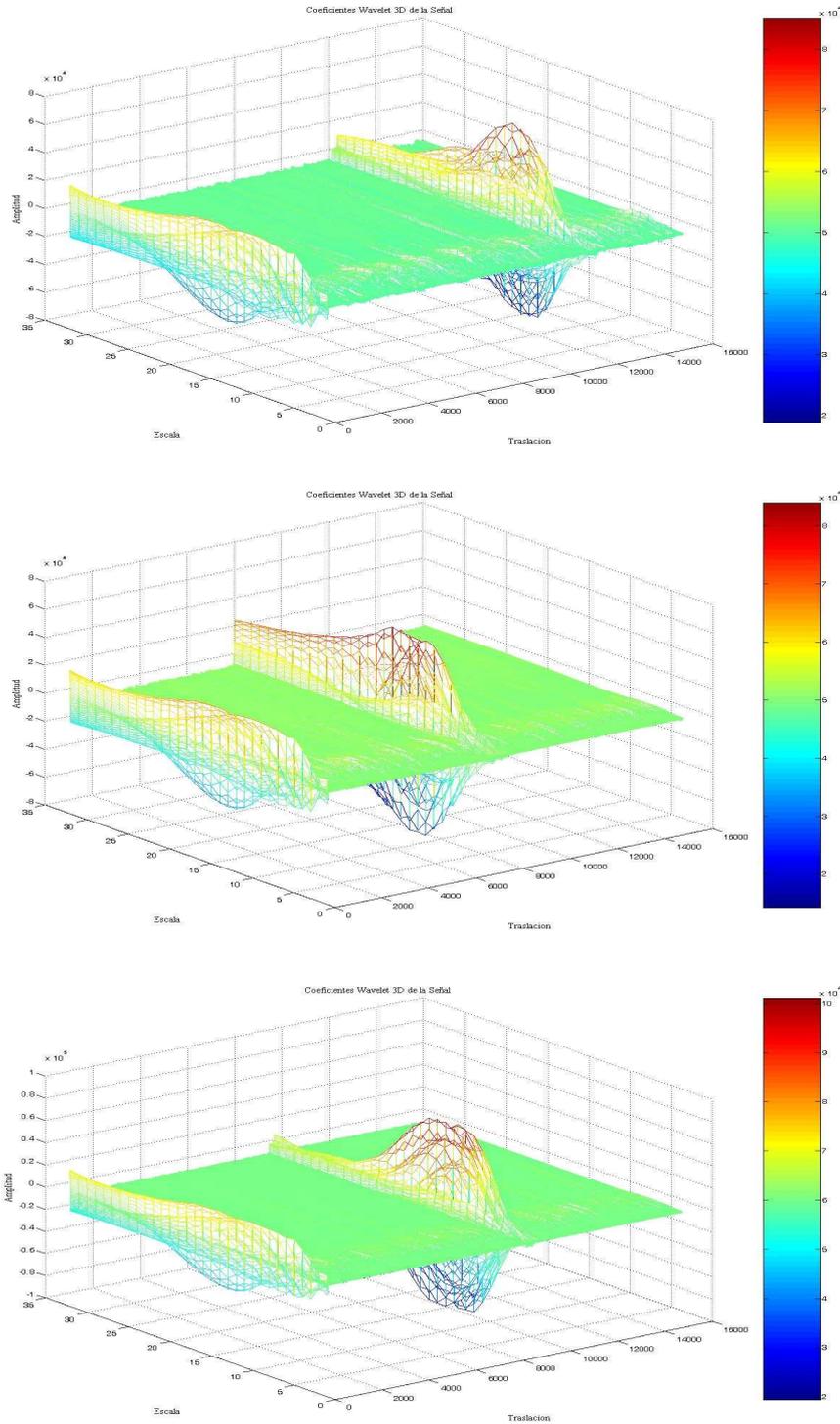


Fig. E.37 Wavelet 3D de eventos tipo T1 en tres registros del Telica. Se pueden observar las dos formas de ondas que corresponden a las dos llegadas en energía de este tipo de evento.

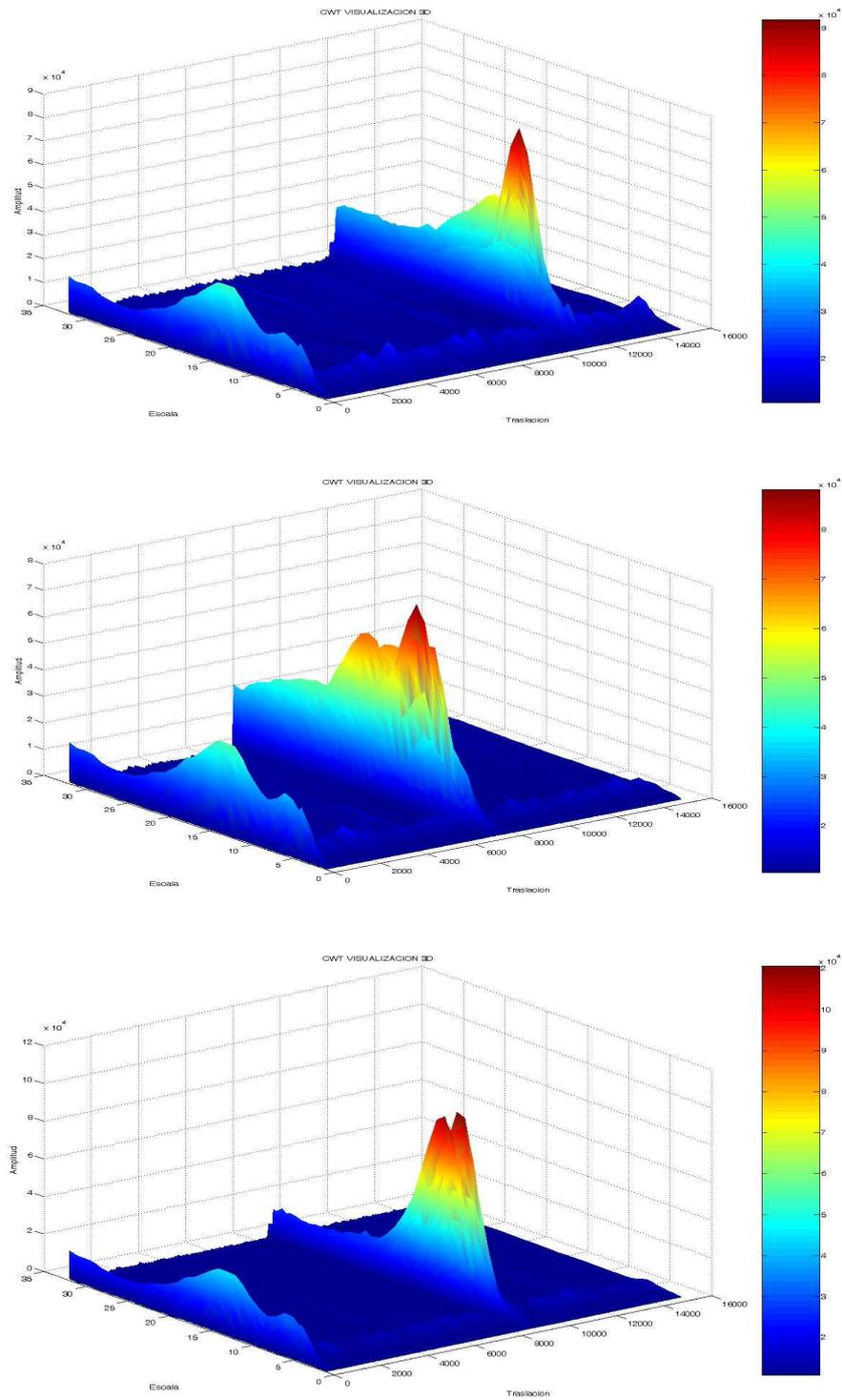


Fig. E.38 CWT 3D valor absoluto (abs.) de eventos de tipo T1 en tres registros del Telica. Aunque los tres eventos aquí mostrados difieren a lo largo del tiempo (ver estadísticas de los tipos de eventos), es en el tercero que se observa con mayor precisión, la llegada de las dos componentes en energía.

Se puede observar en las figuras E.37 y E.38, la semejanza en la forma (pico) de los máximos energéticos de este tipo de evento.

ii) Evento de tipo T2: La figura E.39 muestra la representación de la Wavelet continua (CWT), del evento T2

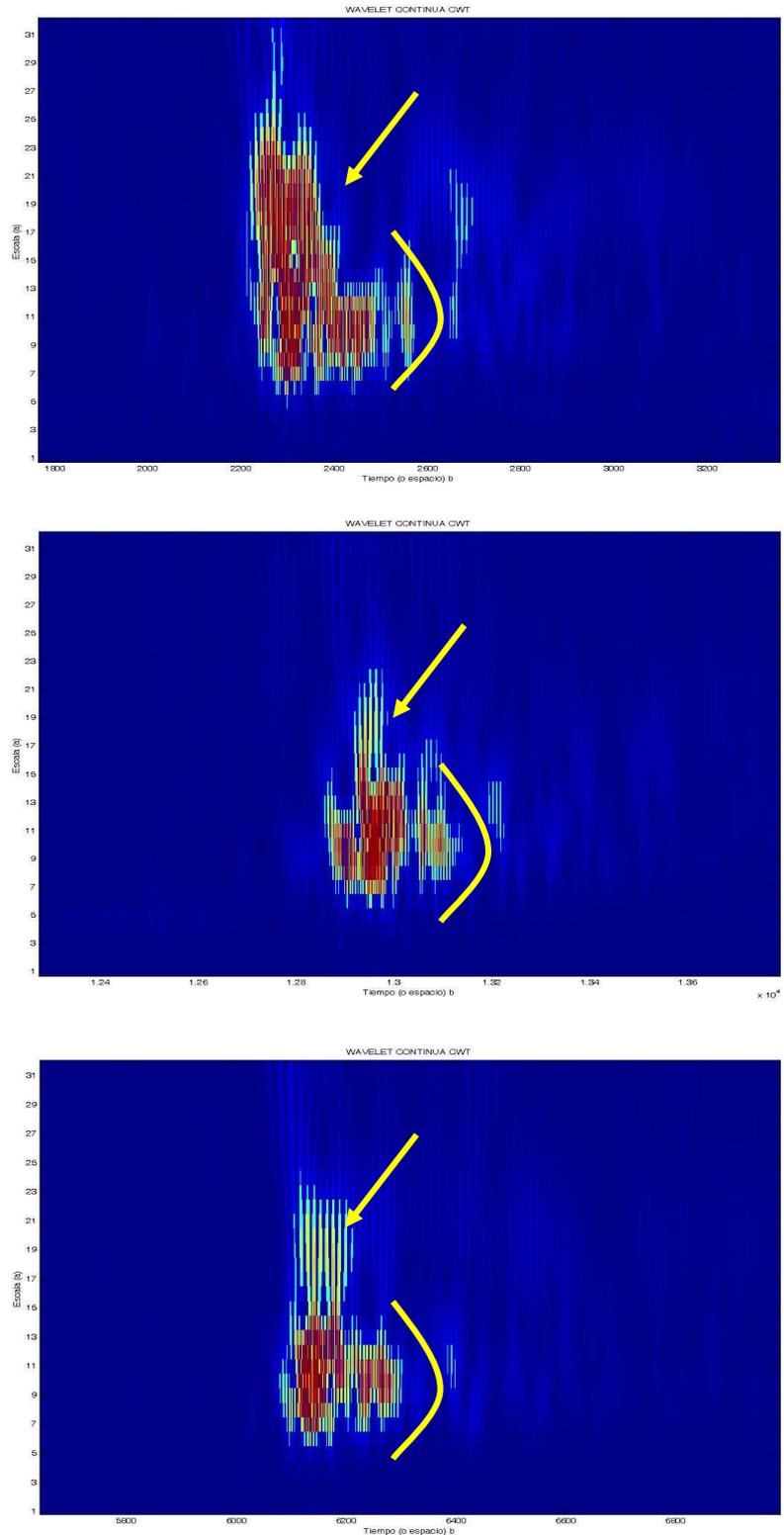


Fig. E.39 Wavelet continua (CWT) de eventos tipo T2 en tres registros del Telica. La flecha indica a diferencia del T1, una única componente mayor de llegada en energía y la curva indica la forma de caída en frecuencia de este tipo de evento.

El evento anterior posee una caída suave en energía así como una duración menor en el tiempo. La representación en 3D de este evento se observa en las figuras E.40 y E.41

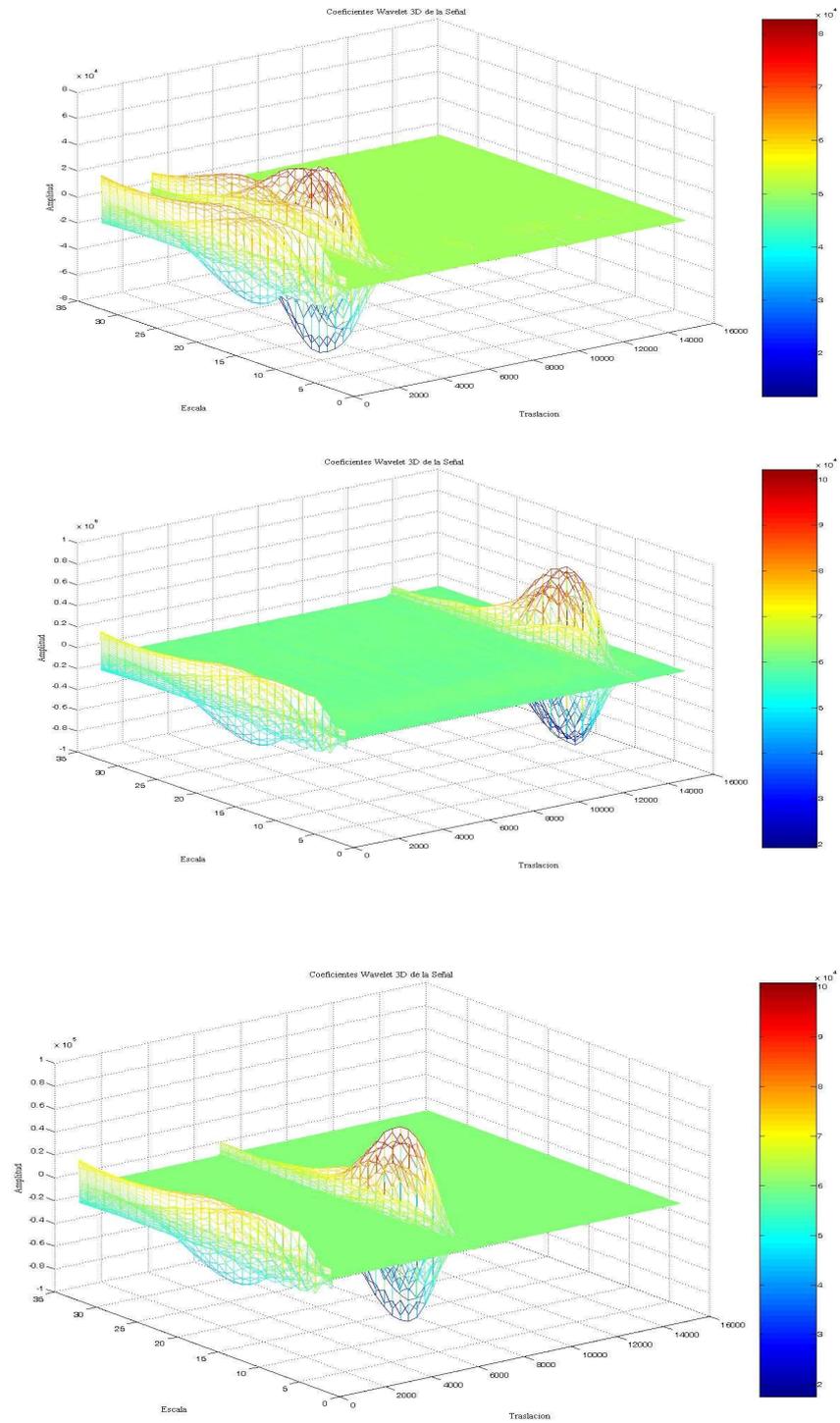


Fig. E.40 Wavelet 3D de eventos tipo T2 en tres registros del Telica. Una única onda en frecuencia muestra la llegada de la componente mayor en energía.

Obsérvese la semejanza en la forma (onda) de los máximos energéticos de este evento

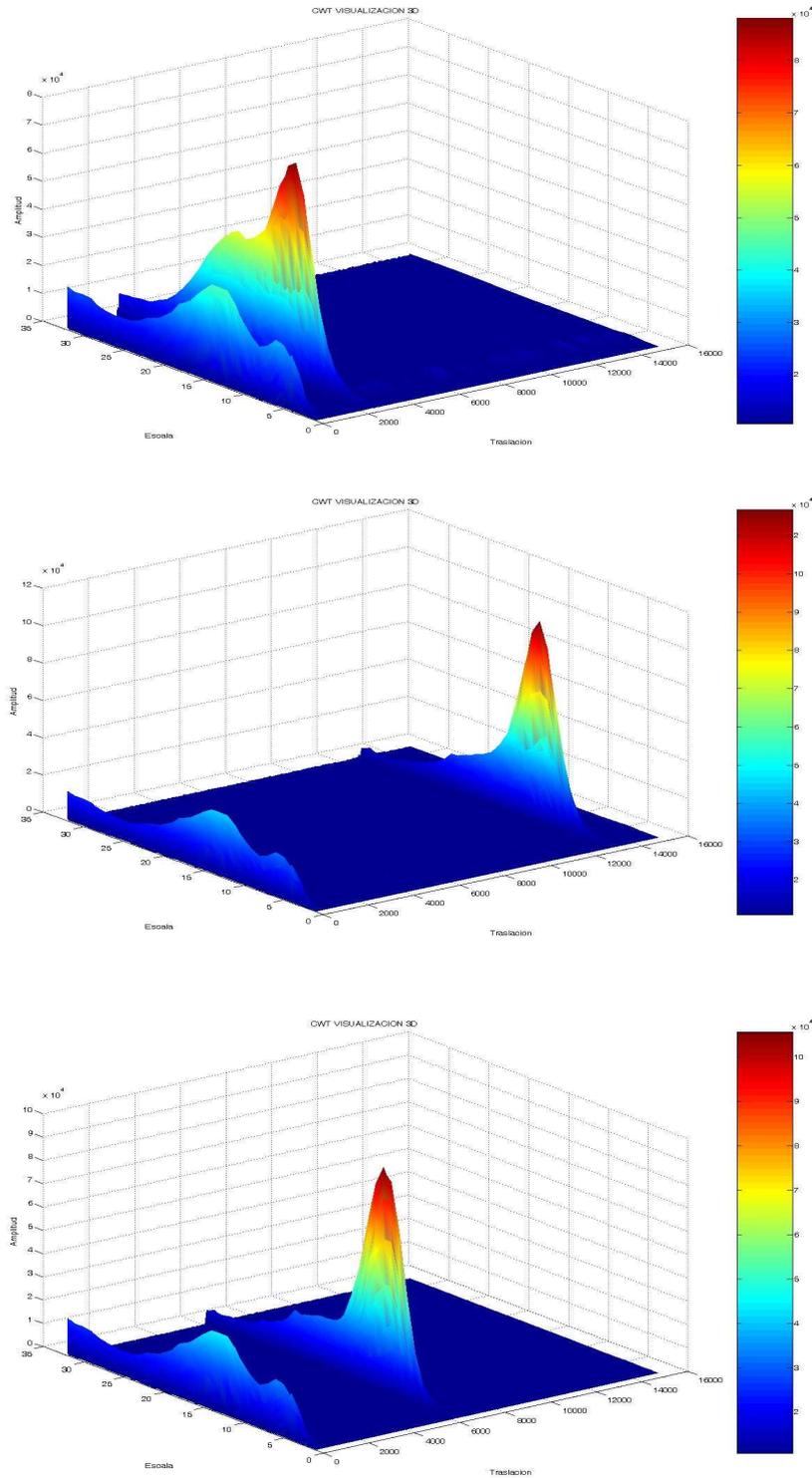


Fig. E.41 CWT 3D valor absoluto (abs.) de eventos de tipo T2 en tres registros del Telica. Se puede observar cómo se muestra el pico en frecuencia.

- iii) Evento de tipo T3: La figura E.42 muestra la representación de la Wavelet continua (CWT), del evento T3

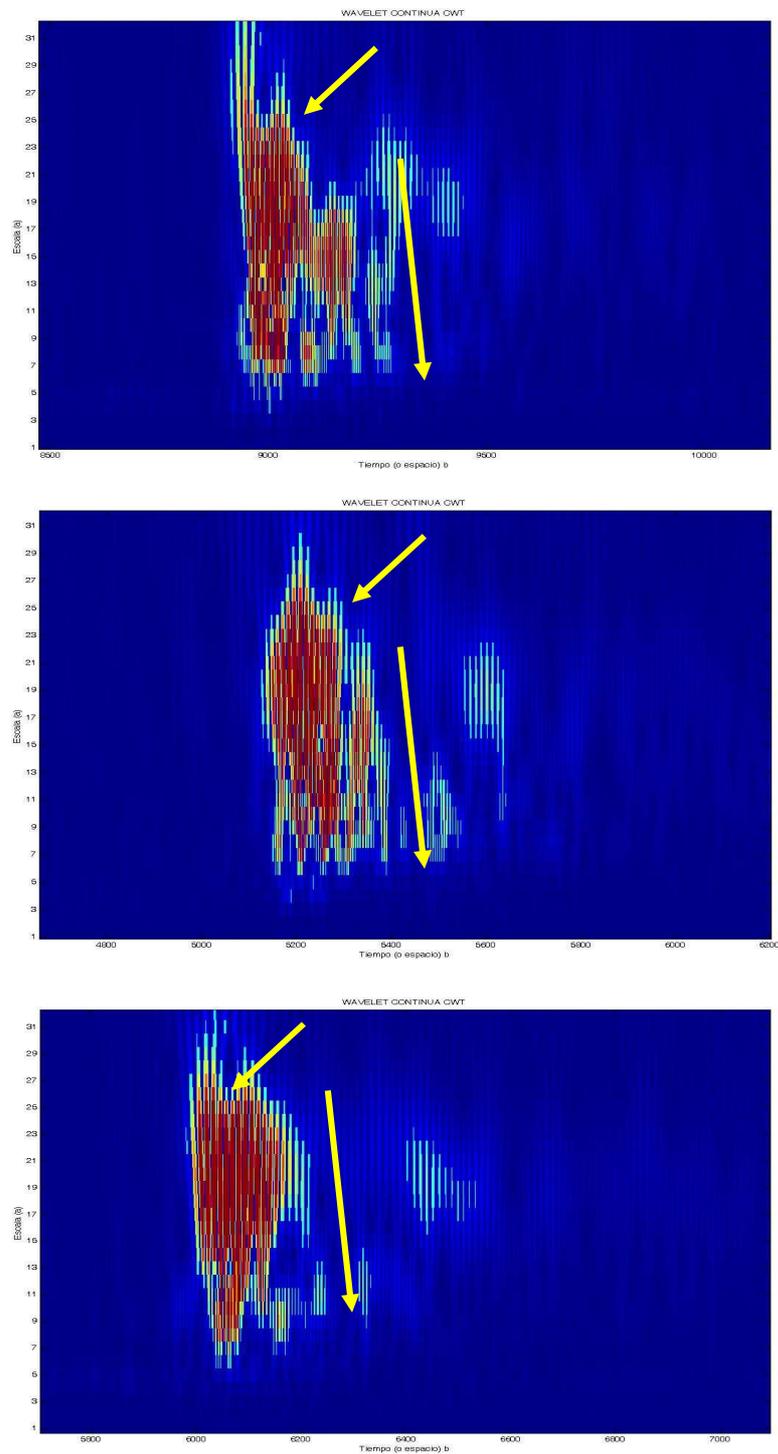


Fig. E.42 Wavelet continua (CWT) de eventos tipo T3 en tres registros del Telica. Al igual que el de tipo T2 este tipo de evento no presenta múltiples llegadas en frecuencia, la flecha superior lo indica. Sin embargo, como indica la flecha más a la derecha, la forma de caída en frecuencia es mucho más abrupta en este tipo de evento.

Obsérvese la semejanza en la forma (punta) de los máximos energéticos de este evento. Este tipo de evento posee una caída brusca a diferencia de otros al final, con un componente energético más compacto y no tan separado en discontinuidades o picos energéticos como los demás. La representación en 3D de este evento se observa en las figuras E.43 y E.44

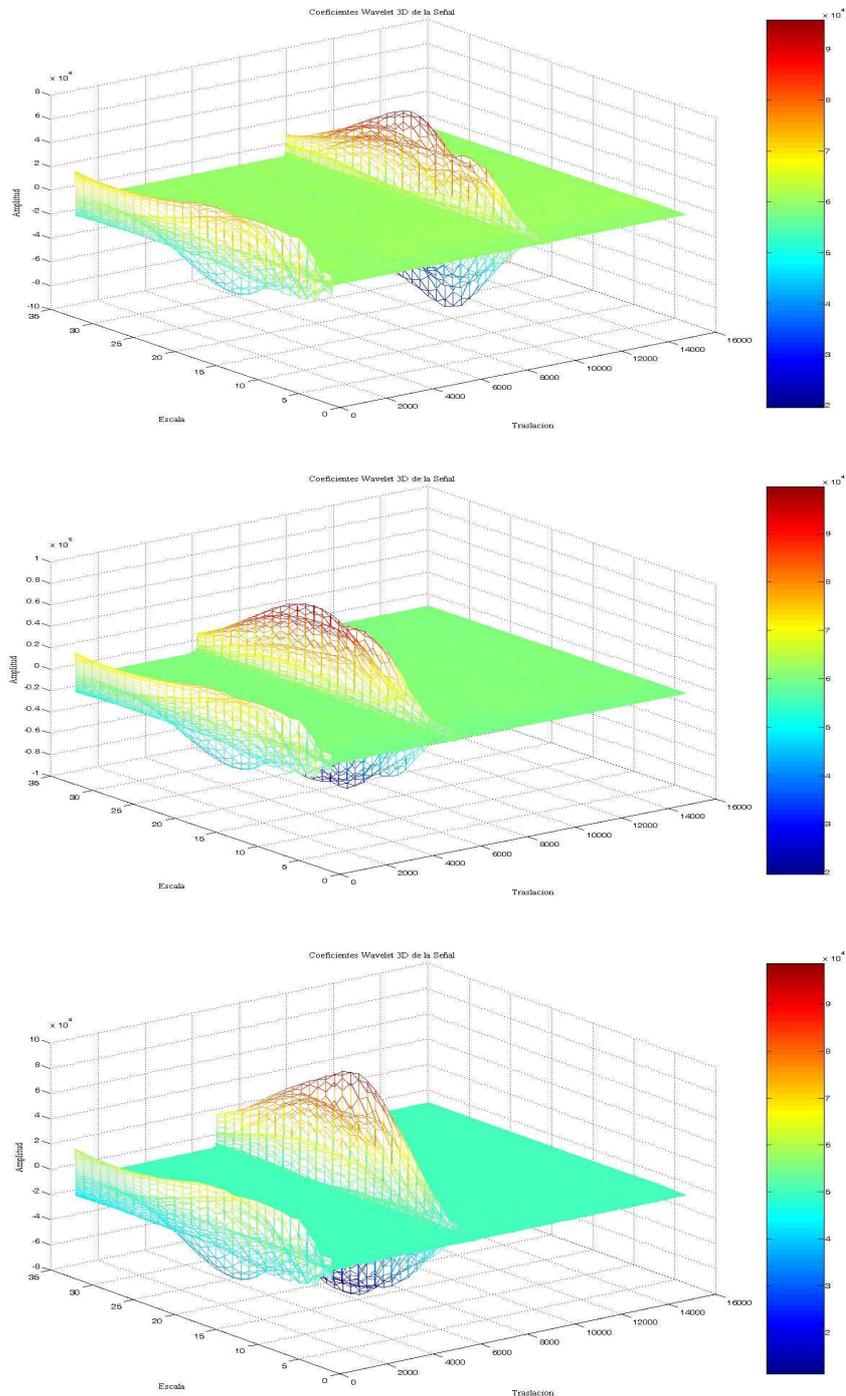


Fig. E.43 Wavelet 3D de eventos tipo T3 en tres registros del Telica. La onda energética en la forma de las componentes mayores de este tipo de evento no es tan senoidal como la del evento T2, es más aplanada.

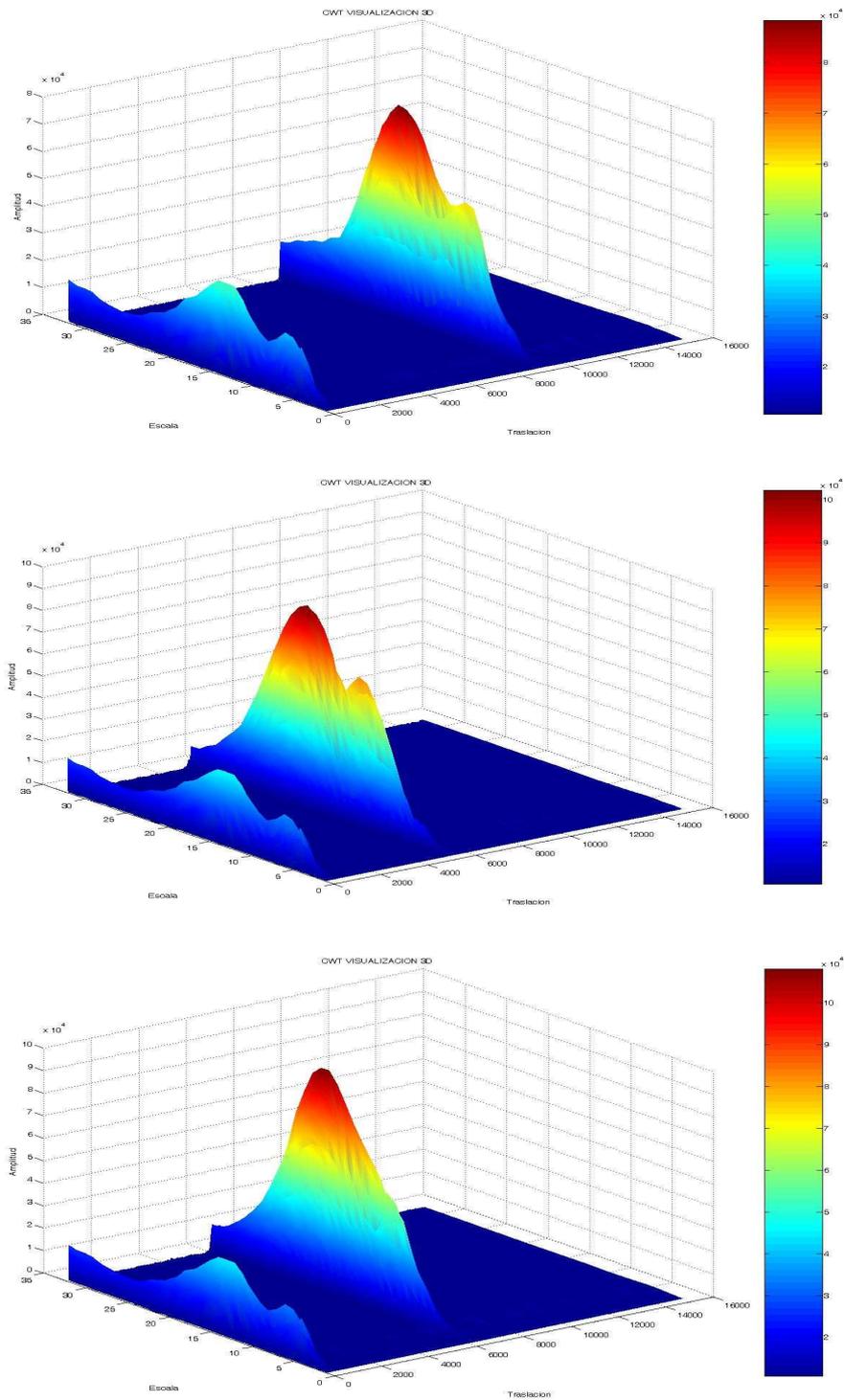


Fig. E.44 CWT 3D valor absoluto (abs.) de eventos de tipo T3 en tres registros del Telica. Se puede observar que el pico en frecuencia en este tipo de evento es más pronunciado, similar al del tipo T1.

Se observa en las figuras E.43 y E.44, la forma de cono y no de pico en los máximos energéticos de este tipo de evento.

- iv) Evento de tipo T4: La figura E.45 muestra la representación de la Wavelet continua (CWT), del evento T4

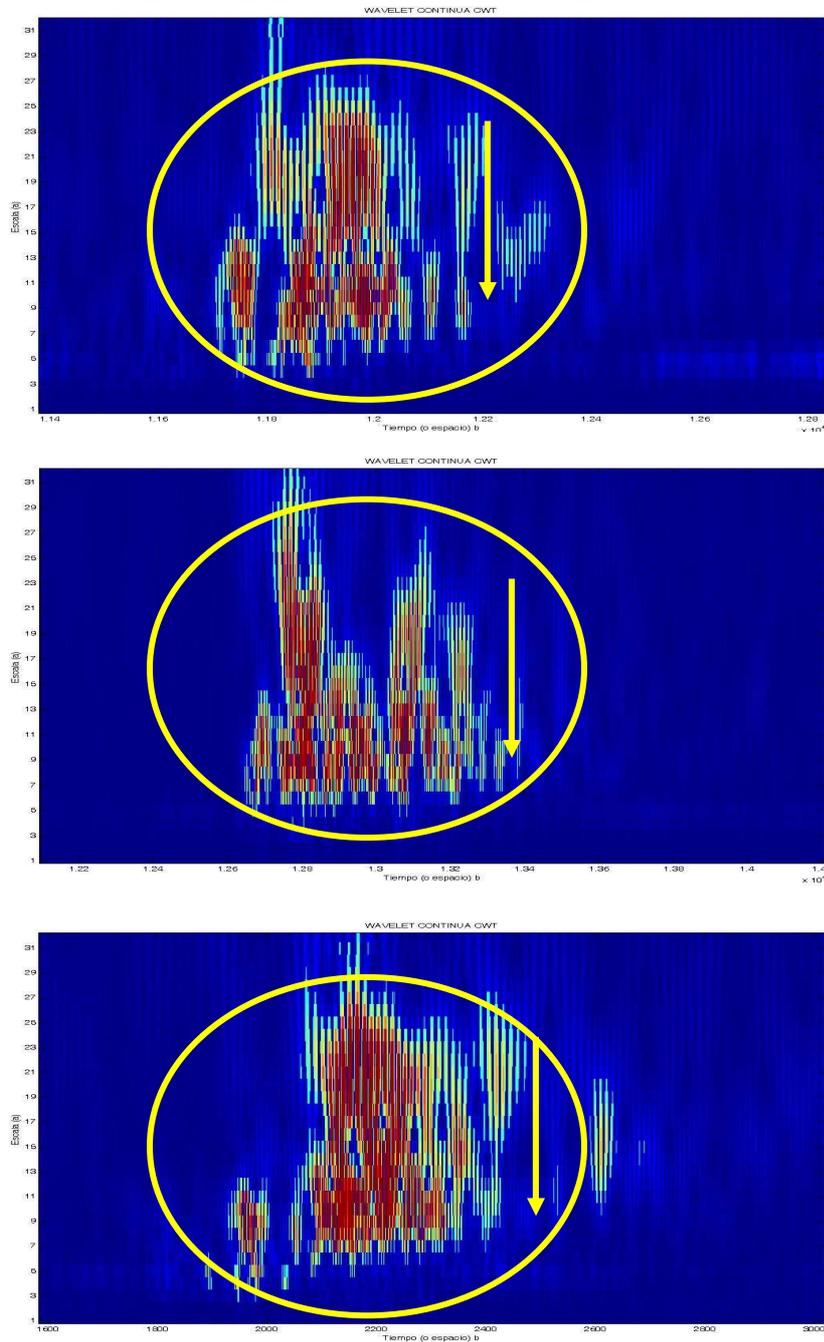


Fig. E.45 Wavelet continua (CWT) de eventos tipo T4 en tres registros del Telica. En el círculo se aprecian múltiples llegadas de energía para este tipo de evento, asimismo la flecha indica una caída en frecuencia mucho más abrupta y espaciada a lo largo del tiempo, por lo que este tipo de evento podría tratarse un evento de tipo LP.

Variados componentes energéticos o discontinuidades componen este evento así como su forma alargadas lo largo del tiempo. La representación en 3D de este evento se observa en las figuras E.46 y E.47

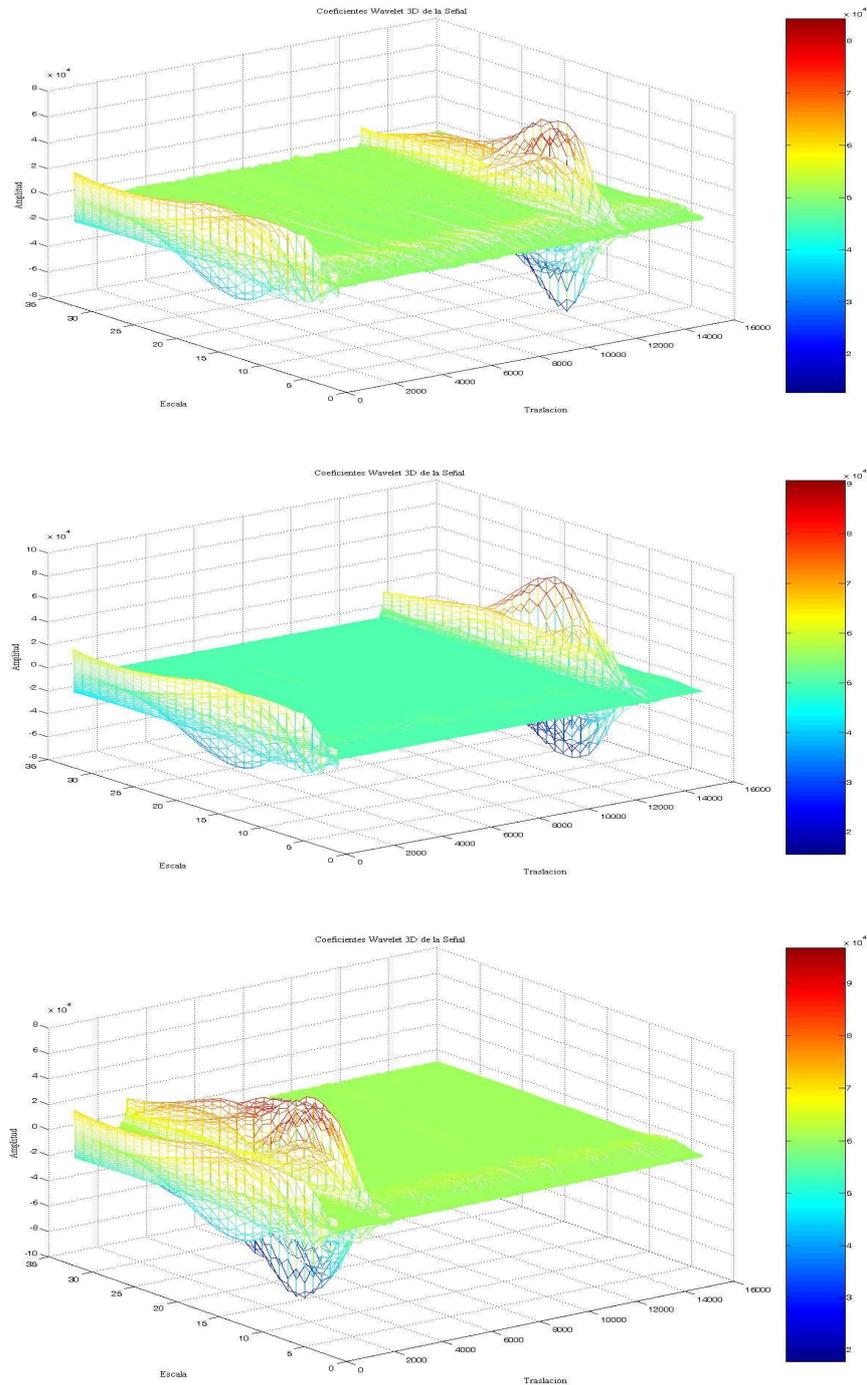


Fig. E.46 Wavelet 3D de eventos tipo T4 en tres registros del Telica. La forma de onda energética es más alargada a lo largo del tiempo.

Se observa en la figura E.46, la forma de onda compacta aplanada, mucho más baja que en otros tipos de eventos

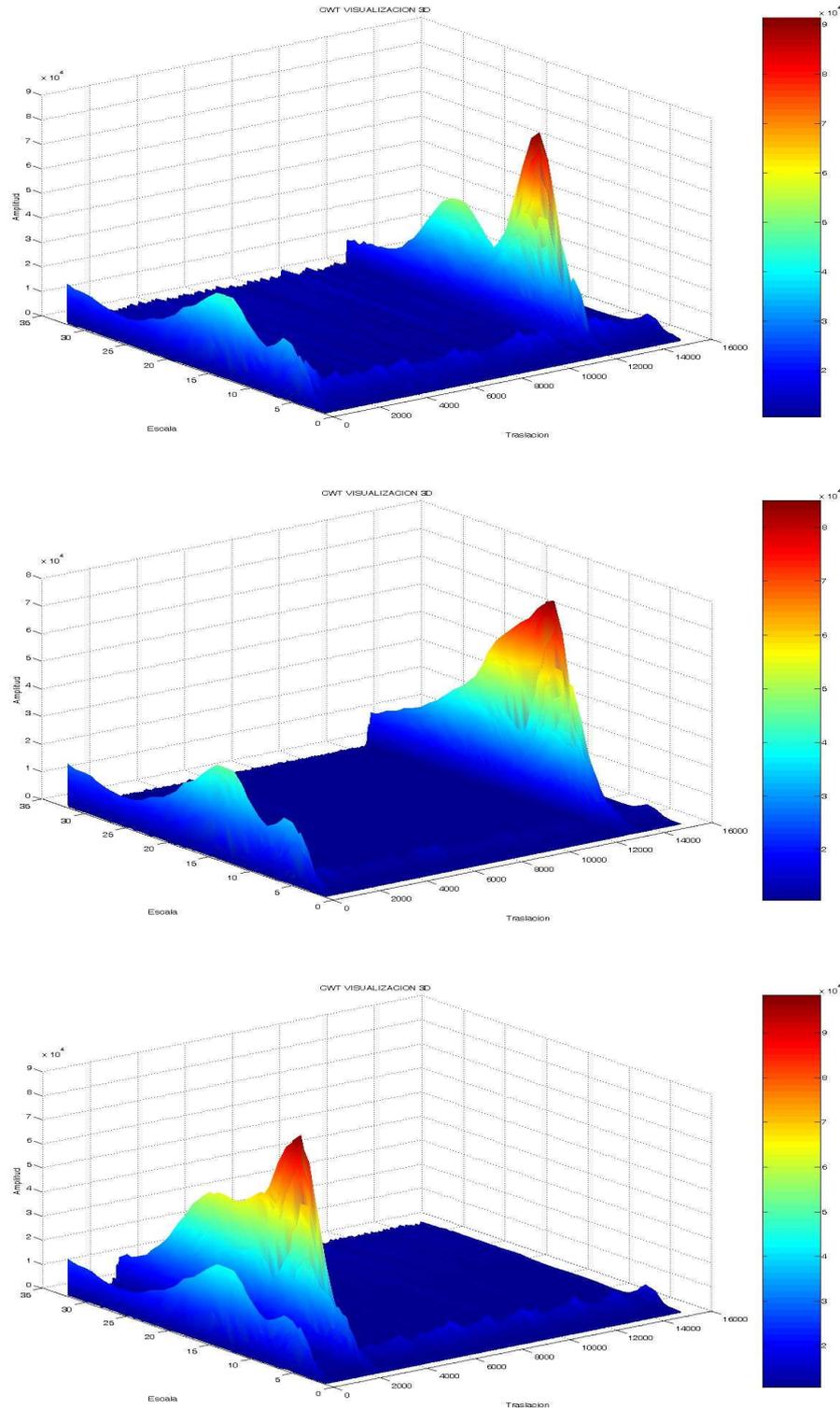


Fig. E.47 CWT 3D valor absoluto (abs.) de eventos de tipo T4 en tres registros del Telica. El pico predominante en frecuencia en este tipo de evento es similar al del T1, solo que no posee dos picos mayores, con lo que los diferentes picos menores quedan integrados en el pico predominante.

En la figura E.47 anterior, se observa la forma de pico, no tanto de cono en los máximos de este tipo de evento. En este estudio ha de destacarse que los tipos T2 y T3, podrían pertenecer a una misma familia de eventos en el volcán, diferenciándose el T3 del T2 que el primero posee una llegada inicial de alta energía y corta duración y magnitud, no así el T2 que es de mucha mayor duración y una caída en frecuencia mucho más suave que el T3. El T1 y T4 poseen mayor contenido espectral que el T2 y T3. Siendo el T1 (lo que en Nicaragua denomina eventos dobles al presentar dos llegadas sucesivas de altas energías), y el T4 los que contienen más de dos componentes espectrales en altas energías pudiendo especular a priori que se tratase de eventos de tipo LP, sin embargo, como se ha mencionado para concluir este resultado se debe de llegar a un consenso con los sismólogos locales en Nicaragua. Tanto lo visto en este apartado, como el poder tener la certeza de la existencia o no, de familias de eventos sísmicos en los volcanes, es un material de estudio bastante interesante para futuras investigaciones.

Las figuras de los registros del Telica muestran la forma típica de los eventos sísmico-volcánicos presentes en ellos. Se observa visiblemente definida, la forma típica de los cuatro eventos y sus diferencias. Cada uno de ellos podemos decir que difiere en sus componentes espectrales de los demás tipos encontrados en el volcán.

De la misma forma que se ha hecho con el espectro Wavelet, la imagen tridimensional Wavelet indica que el T1 y el T4 pueden pertenecer a una misma familia de eventos, lo mismo que el T2 y su análogo el T3, los cuales presentan algunas características similares.

El evento T1 y el evento T4, poseen más componentes energéticas que el T2 y T3. Siendo el T4 el que más componentes se observan. El T2 posee una llegada y una caída en frecuencia suavizada siendo similar a una campana de Gauss, no así el T3 que es como si se tratara de una doble campana.

Podría concluirse que el uso de herramientas complementarias como la de técnicas Wavelet, podría representar una ventaja significativa, la cual se puede observar en todas las gráficas anteriores. En cada una de las imágenes anteriores, pueden observarse las discontinuidades de las energías que conforman los eventos, definiéndose con más claridad las semejanzas y diferencias que existen. En futuras investigaciones, el empleo de esta técnica, conjuntamente con las técnicas espectrales de Fourier, los espectrogramas y la densidad de poder espectral, podrían catalogar mejor los diferentes tipos de eventos en los registros sísmico-volcánicos, para así lograr la construcción de una mejor base de datos de los volcanes. En contraposición, si solo se contara y se hiciera uso de una sola herramienta, podrían inducirse errores a la hora de definir cada tipo de evento, lo que acarrearía la mala construcción de una buena base de datos con la que puedan ser entrenados los modelos HMMs para el reconocimiento.

APENDICE F: Procedimiento de conversión de Formatos

1. Introducción.

En este apéndice, a manera de guía se presenta un detallado proceso de los pasos necesarios en la conversión de los diferentes formatos a utilizar durante el estudio. Aquí se pormenoriza la secuencia de conversión para los ficheros de Nicaragua desde el formato de datos en crudo o registros en bruto (archivos R), almacenados en las estaciones, al formato binario SEISAN [Havskov y Ottemoller (1999)], hasta llegar al formato de forma de onda y ficheros r50 que se utilizarán en las pruebas de entrenamiento y reconocimientos de los eventos sísmico-volcánicos. Este apéndice puede servir de guía a futuros investigadores que se encuentren en una situación similar, a fin de que puedan seguir todas las fases necesarias en el desarrollo de la conversión de los formatos.

Lo primero a realizar en la secuencia de conversiones es la preparación o creación de las listas de trabajo, que contienen los diversos registros en formato R, recopilados en las estaciones. Para ello se hace uso de los comandos del programa SEISAN de la siguiente forma:

Se crea la lista de todos los archivos en el directorio con el comando SEISAN siguiente:

```
dirf R*
```

Este paso, crea una lista denominada “filenr.lis”, en ella se encuentran enumerados todos los archivos R. A continuación se ejecuta el siguiente comando:

```
qnxsei
```

La ejecución del anterior comando, creará los archivos binarios con 64 canales de toda la red sísmica (todas las estaciones sísmicas instaladas en todos los volcanes). Al ejecutar el comando se pide que ingrese el nombre de la lista previa creada esto es “filenr.lis”. Hasta aquí, se tiene la información separada en los canales de cada estación, sin embargo, hay que separar dicha información para poder seleccionar el canal adecuado, que contenga los registros a analizar. Esto se realiza de la siguiente manera:

2. Separación y elección de los canales idóneos.

a) Se crea una nueva lista conteniendo los archivos creados de 64 canales. Para ello se ejecuta el comando:

```
dirf 2*
```

Esto crea la lista conteniendo el nombre de todos los archivos de 64 canales en el archivo “filenr.lis”

Por ejemplo. la lista de registros del día 16 al 23, contiene 1,049 registros de 64 canales. Una parte de esta lista con los 10 primeros registros se observa a continuación:

filenr.lis

```
1 2006-02-16-2033-38S.NIC__064
2 2006-02-16-2038-38S.NIC__064
3 2006-02-16-2043-38S.NIC__064
4 2006-02-16-2048-38S.NIC__064
5 2006-02-16-2053-38S.NIC__064 ...
. . . . .
```

b) Separación: Se ejecuta el comando “seisei” de la forma siguiente:

```
C:\prueba> seisei # Aquí se ejecuta el comando seisei
```

Al ejecutarlo se presentan 2 opciones:

- 1 – Merge (para juntar o unir)
- 2 – Split (para separar)

Se selecciona la opción número 2, esto indica separar. Ello crea por cada canal de la lista anterior un fichero. Además, el comando pregunta qué lista se va utilizar, que en nuestro caso es “filenr.lis”. A continuación, se da comienzo el proceso de la creación de los ficheros por cada uno de los 64 canales de la lista. Cada uno de estos ficheros tiene un tamaño aproximado de 3.72 Mb y al crear los ficheros se crean 60 ficheros separados por cada canal de aproximadamente 60.6 Kb cada uno. Por ejemplo, para separar un solo fichero de los 64 canales (que indicaría tomar el primero de la lista anterior de 1,049 registros de 64 canales), el procedimiento sería el siguiente:

event file

Merge (1) or split (2) files:

Al ejecutar el comando, lo primero que pregunta es si se va a fusionar (Merge) o separar (split). En cuyo caso se escoge 1 o 2. Se le indica la opción 2, pues hay que separar los canales. El comando envía el siguiente mensaje:

Filename, # or filenr.lis for all

Con lo que el sistema pregunta si se desea separar todos los ficheros contenidos en la lista “filenr.lis” o uno solo en cuyo caso solo se teclearía el nombre del fichero. De la lista anterior se va a elegir el primero.

```
2006-02-16-2033-33S.NIC__064
2006-02-16-2033-38S.NIC__064 ....
```

Por lo tanto, se le ha indicado que se desea uno solo y se le ha dado el nombre del primer fichero. A continuación, se observa a manera de ejemplo cómo se presentaría el procedimiento de la separación de los 64 canales de este fichero, cada uno de dichos canales contiene el nombre y el tiempo respectivo de proceso:

```
istat      64
TIMEL Z    0.0 299.96 XXXXS N    0.0 299.96 TEL3S Z    0.0 299.96
```

```

MGANS Z 0.0 299.96 MGANS N 0.0 299.96 MGANS E 0.0 299.96
XXXXAL Z 0.0 299.96 XXXXAL N 0.0 299.96 XXXXAL E 0.0 299.96
MGANAH Z 0.0 299.96 MGANAH N 0.0 299.96 MGANAH E 0.0 299.96
XXXXS E 0.0 299.96 PYN S Z 0.0 299.96 LEONS Z 0.0 299.96
TIME L Z106 47 2 16 20 33 38.429 50.0063 15000
2006-02-16-2033-38S.TIME__001_L__1Z

```

```

MGAN S Z106 47 2 16 20 33 38.429 50.0063 15000
2006-02-16-2033-38S.MGAN__001_S__1Z

```

.....

Filename, # or filenr.lis for all

Se observa la nueva lista resultante del fichero “2006-02-16-2033-38S.NIC__064”, que se ha tomado como ejemplo. A manera de ejemplo, a continuación se observa una parte de la lista, con la creación de estos 60 ficheros binarios resultantes.

```

1 2006-02-16-2033-38S.AGCN__001_S__1Z
2 2006-02-16-2033-38S.APON__001_S__1Z
3 2006-02-16-2033-38S.APYN__001_S__1Z
4 2006-02-16-2033-38S.BOA2__001_S__1Z ....

```

Hasta aquí, se ha finalizado la separación en los 64 canales de un solo fichero. Para poder darse una idea de todo el trabajo que hay que realizar, hay que tener en cuenta que el procedimiento anterior, se debe de realizar con todas las listas que contienen todos los ficheros de 64 canales. Es decir, con la totalidad de los registros de 64 canales. Así que la totalidad de ficheros generados de 64 canales de una sola lista, en este caso, la lista del intervalo de tiempo del 16 al 23 de un mes, contiene 1,049 registros de 64 canales. Esta cantidad se debe de multiplicar por 60, para saber cuántos ficheros se generarán. Por lo tanto, serían $1,049 \times 60 = 69,940$ ficheros resultantes que se deberán posteriormente analizar.

3. Creación y selección de los canales de interés a trabajar:

Para los volcanes de Nicaragua del presente estudio, el proceso consistirá en quedarse con los canales pertenecientes a las estaciones del San Cristóbal – “CRIN” y los del Telica – “TELN”. Por lo tanto, de la lista de todos los canales que se han obtenido previamente, se han de localizar estos dos canales, que corresponder a las estaciones sísmicas del San Cristóbal y del Telica. El procedimiento a seguir es el siguiente:

- a) Se ha de trabajar con los registros de 64 canales creados anteriormente, para ello, se debe de crear una nueva lista de ficheros con el canal seleccionado. Para el caso del “San Cristóbal” se ejecutaría el siguiente comando:

```
dirf *CRIN*S __ 1Z
```

Esto significaría, crear una lista con únicamente los datos de la estación sísmica seleccionada (CRIN), que es la del San Cristóbal. Así de 69,940 ficheros, se deben separar solo aquellos que correspondan a dicha estación. La nueva lista es designada de forma automática por el programa SEISAN, de manera similar que las anteriores listas, mediante el comando: “filenr.lis”

- b) Una vez creada dicha lista, se ejecuta el siguiente comando:

```
wavfix
```

En donde el sistema, pregunta el tiempo de corrección, o si se desea dejar con el que posee por defecto. Como se desea dejarlo con el mismo tiempo, se pulsa la tecla “enter”. A continuación, el sistema pregunta con que lista se va a trabajar, en este caso se da la opción 1, que indicaría trabajar con la lista ya creada “filenr.lis”. Una vez creados los archivos correspondientes a la estación sísmica elegida se trasladan a una carpeta aparte (creada por el usuario), para continuar con la creación de la siguiente lista de archivos del otro volcán.

De manera similar, para el Telica, se seguiría el mismo proceso. Para crear la lista a trabajar con la estación sísmica elegida, se teclea el comando siguiente:

```
dirf *TELN*S*__1Z
```

Esto creará la nueva lista de archivos “filenr.lis”, y nuevamente se ejecuta el comando “wavfix” con los mismos parámetros anteriores. Una vez que se han trasladado los archivos a sus correspondientes carpetas, el paso siguiente es la eliminación del resto de los archivos no deseados mediante el comando:

```
del *Z *N *E
```

Nota: Cabe mencionar que para la creación de las dos listas de canales, al utilizar el comando “dirf”, todo se escribe seguido y sin espacios a partir del primer carácter comodín el asterisco. Los guiones abajo van igualmente seguidos sin espacios intermedios.

Como resultado del proceso anterior, la lista creada del 16 al 23 de la estación sísmica del volcán San Cristóbal, a partir de los registros de 64 canales contendría 2,000 ficheros binarios. A continuación se observan los primeros cinco.

```
filenr.lis  
1 2006-02-16-2033-38S.CRIN__001  
2 2006-02-16-2038-38S.CRIN__001  
3 2006-02-16-2043-38S.CRIN__001  
4 2006-02-16-2048-38S.CRIN__001  
5 2006-02-16-2053-38S.CRIN__001 . . . .
```

Análogamente, este mismo procedimiento ha creado otros 2,000 ficheros del volcán Telica, los cuales en lugar de tener la extensión “CRIN__001”, tendrán la extensión “TELN__001”. Por lo que ahora, en lugar de los 69,940 ficheros de toda la red sísmica, se disponen únicamente de 4,000 ficheros binarios de las dos estaciones elegidas.

4. Creación de archivos de cabecera con la información.

Una vez creados los ficheros con los datos, se podría proceder (si se desea) a la creación de los ficheros que contendrán la cabecera con la información. Similar a los “dts” que se generan en las estaciones de las campañas de Italia. Para ello, primeramente se crea una nueva lista de los archivos que se desean, mediante el comando “dirf”. Posteriormente se ejecuta el siguiente comando:

autoreg

Al ejecutarlo, el sistema pide que tipo de eventos se deseen para todos los registros:

L – Local
R – regional
D – distante

Como se están utilizando eventos locales se le da la opción “L”

A continuación se pide el nombre de la base o dos comas “,” para base de datos local. Ya que se está trabajando en modo local se teclean estas dos comas. Por último se pide el nombre del operador (valor arbitrario). Por ejemplo, en este caso se teclea “LIG”. Con ello se comienza a crear un archivo de cabecera para cada archivo de dato.

Para la edición o manipulación de los nombres de los ficheros en las listas creadas, se utiliza un script o fichero de tratamiento por lotes (.bat), denominado “Lig.bat”. Este script tiene por objeto crear una lista en Excel para poder ser manipulada. El fichero batch contiene la siguiente información.

```
dirf *001
autoreg < C:\seismo\COM\autoreg.def
dir /b *L.S* 2006* > lista.xls
lista.xls
move lista.xls lista.bat
call lista
del lista.bat
```

El propósito de dicho fichero es que se crea y se abre la lista en Excel para renombrar los ficheros ya creados en las listas mediante las siguientes acciones:

- a) se inserta una nueva columna en A
- b) Se seleccionan las formas de onda y se mueven al inicio de la columna C, al lado de los nombres de ficheros de cabecera
- c) En la primera fila de la columna A, se teclea “ren”, esto para renombrar, se copia esta línea en toda la columna A hasta el final de la lista
- d) Se selecciona la columna C y se pulsa reemplazar en el menú editar. Aquí se reemplazan todos los puntos (.) por guiones (-) y todos los de doble guión hacia abajo 001 (_ _ 001) por un punto más “dts” (.dts) o lo que es lo mismo la extensión de fichero de cabecera (que puede ser también “txt”). Al cerrar la lista pide salvar y se da si a todo, con ello automáticamente renombra todos los archivos de las cabeceras y deja el mismo nombre que los de onda, solo que con la extensión dts, para indicar que son de datos.

Mediante el procedimiento anterior, se obtienen todas las listas renombradas con la extensión de PITSA, que es la 001. Sin embargo, se necesita tener las listas con los formatos que contienen los nombres originales, para de esta forma, poder identificar

más fácilmente, tanto la estación como el tiempo de los registros, para su posterior uso con las herramientas en HTK. Esto es porque se deben de copiar todos los registros de los ficheros en continuo. El procedimiento es el siguiente:

- a) Se crea una lista con el comando “dirf” de todos los registros Pitsa. Esto es: pitsa001.001
- b) Se renombra la lista como filenr1.lis, para no repetir con la lista a renombrar
- c) Se toma la lista con los nombres originales filenr.lis
- d) Se abren las dos listas en Excel como formato texto con separación de tabulaciones para separar las columnas.
- e) Con la lista Pitsa se copian los datos y se colocan una columna antes de la lista “filenr.lis”
- f) Se crea una nueva columna al inicio y en ella se coloca “rename” o “ren” y se copia a lo largo de toda la columna.
- g) Se reemplaza los nombres originales de las extensiones ejemplo .TELN__001 por .001 en toda la columna a renombrar.
- h) Una vez hecho esto se guarda el fichero como lista1.bat, esto es como texto con tabulaciones y además se guarda la lista como hoja de Excel para renombrar posteriormente los ficheros Wav
- i) Hecho esto en el MSDOS se coloca el path del directorio en donde se guardó la lista y solo se teclea el archivo de lotes bat. Con ello la lista se renombra automáticamente.

5. Conversión de los formatos binarios Seisan al formato ASCII PITSA

Hasta este punto, todos los ficheros de datos que se han creado anteriormente se encuentran en formato binario bajo el formato del lenguaje SEISAN. Sin embargo, para poder trabajar con el tratamiento de señales, se necesita adecuar dicho formato a un tipo que permita utilizar los script y programas que se van a utilizar, tanto para las campañas de Italia (Etna y Strómboli), como para las campañas de Nicaragua (San Cristóbal y Telica), con las nuevas listas creadas de cada estación sísmica

Esto indicaría pasar del formato SEISAN al formato ASCII tipo Pitsa que es uno de los formatos que facilitan esta labor. Con ello se procede a vectorizar los datos contenidos en los ficheros.

El lenguaje SEISAN posee una opción para realizar dicha conversión y es mediante el comando “*Seipitsa*”. La sintaxis del comando es la siguiente:

```
C:\seisan_16_23\reg_64_canales>seipitsa
```

Al ejecutar este comando se presentan las siguientes opciones:

(1) Seisan binary -> Pitsa ASCII or (2) Pitsa ASCII -> Seisan binary

Lo anterior indica que la opción 1, convierte los datos que se encuentran en binario en formato SEISAN a formato PITSA en ASCII, la segunda opción realiza el procedimiento inverso. Como se desea lo primero, la opción a elegir es 1. A

continuación, el programa pregunta si se desea convertir un fichero o la lista completa. En este caso se elige la opción de la lista y el nombre, para que convierta todos los ficheros. Esta acción crea los ficheros ASCII Pitsa. La nueva lista que se genera mediante el comando “dirf”, muestra los 2,000 ficheros ASCII en formato Pitsa ahora creados. A continuación se presenta un ejemplo de la lista con los cinco primeros.

filenr.lis

```
1 pitsa001.001
2 pitsa002.001
3 pitsa003.001
4 pitsa004.001
5 pitsa005.001 .....
```

Este mismo procedimiento se realiza con el volcán Telica y con todas las listas de los dos volcanes. Por otra parte, existe un inconveniente, que se presenta al realizar la conversión, debido a que se nombran automáticamente todos los ficheros como pitsa. Esto hace perder el orden y la localización del tiempo a que corresponde cada registro. Para volver a tener las listas con los nombres originales y en los tiempos en que fueron almacenados los registros por los instrumentos, se crea un pequeño fichero batch de procesamiento de lotes. (.bat), para volver a renombrar todas las listas. Como ejemplo de ello, a continuación se observa parte de uno de dichos ficheros, correspondiente a renombrar la lista de 2,000 registros del San Cristóbal de la lista del 16 al 23.

lista1.bat

```
rename pitsa001.001 2006-02-16-2033-38SC.001
rename pitsa002.001 2006-02-16-2038-38SC.001
rename pitsa003.001 2006-02-16-2043-38SC.001 .....
```

Con el anterior procedimiento se asegura que todos los ficheros posean el mismo orden y fecha de los registros originales, cuando fueron registrados en las estaciones.

6. Conversión de ficheros Pitsa a ficheros Wav, Raw y R50

Una vez conseguidos los ficheros de datos a formato ASCII (Pitsa), estos se deben de transformar a ficheros de audio para ser utilizados con el Transcriber y a ficheros sin cabecera Raw y r50, para poder utilizarlos en las parametrizaciones.

A partir de aquí, se estandarizan los pasos a seguir para los cuatro volcanes, tanto para los dos de las campañas de Italia, como para los de las campañas de Nicaragua, ya que los procedimientos a seguir en las siguientes conversiones y trabajos es el mismo.

La conversión de esto se realiza mediante listas, a través de tres programas de Matlab. El primer programa lee la lista en donde están los ficheros a convertir. Este programa llama a la lista “lista1.txt” y por cada uno de los ficheros ahí nombrado realiza la llamada al segundo programa, la primera línea de la lista indica el número de ficheros contenidos en dicha lista. Al final cierra la lista que ha abierto como lectura. Un fragmento de una de estas listas correspondiente a los 929 registros del 03 al 06 es el siguiente:

```
Lista1.txt
929
pitsa001
pitsa002
```

pitsa003

El primer programa de MATLAB que lee la lista es el siguiente:

```
% Programa Convierte_lista.m
% llama a las funciones para leer ficheros Pitsa
% y convertirlos en WAV y RAW
% utiliza una lista de ficheros como base de datos

% lista = input('Teclee la Lista a Convertir: ', 's');
lista = 'lista1.txt';

f1=fopen(lista,'r');
Nfrases = sscanf(fgets(f1),'%s');

for Nfrases = 1:929,
    s1 = fgets(f1);
    frase = s1(1:length(s1)-2);
    convierte_nic(frase);
end

fprintf(' Lista Convertida: %s \n', lista);
fclose(f1);
```

El segundo programa de acuerdo a las entradas de cada fichero, lee el contenido de dicho fichero ASCII, mediante el programa “lee_pitsa.m” y realiza la conversión al formato Wav, con un factor de conversión de 20000 (200 Fs) en Italia y 5000 (50 Fs) en Nicaragua, que han sido las frecuencias de muestreo con las que se almacenaron los datos en las campañas. Así, por cada fichero “Pitsa” leído de la lista, se crea un fichero, en este caso “Wav” equivalente. Es en este programa donde se realizan los cambios, dependiendo si se desea crear el fichero Wav, RAW o r50. Además, se deben de realizar algunos cambios en estos ficheros, dependiendo del tipo de formato de entrada (SAD, ASCII, SAC, etc.)

```
% Programa convierte_pitsa
% convierte ficheros ASCII de formato PITSA
% a ficheros WAV para segmentarse con el Transcriber

function convierte_nic(entrada)
fichero_asc1 = sprintf('%s',entrada);
x = lee_pitsa(fichero_asc1);
x=x-mean(x);
x=0.99*x/max(abs(x));
salida=sprintf('%s.wav',entrada);
%Creación del fichero Wav
wavwrite(x,20000,salida);
```

El tercer programa es una función que realiza la lectura de los datos ASCII del fichero de entrada.

```
function x= lee_pitsa(fichero)
fichero_asc2 = sprintf('%s.001',fichero);
fprintf('Convirtiendo el fichero: %s: \n', fichero_asc2);
fid=fopen(fichero_asc2);
linea= 0;
x= [];
while 1
    linea= linea + 1;
    tline = fgetl(fid);
```

```

        if ~ischar(tline), break, end
        if (linea>5)
            x= [ x str2num(tline) ];
        end
    end
end
fclose(fid);

```

Por lo tanto, resumiendo: el primer programa de MATLAB lee las listas, el segundo programa, de acuerdo a las entradas de cada fichero, lee el contenido de dicho fichero ASCII y realiza la conversión al formato Wav y r50, con un factor de conversión. Así, por cada fichero Pitsa leído de la lista es creado un fichero Wav equivalente, y el tercer programa es una función que realiza la lectura de los datos ASCII del fichero.

Como nota y respecto al número de ficheros, hay que advertir que al tener que trabajar con ficheros de sonidos, el proceso de segmentación lleva implícito tener el mismo número, tanto en forma de onda, como en formato Raw sin cabecera y los r50. Por ejemplo, si en el caso del Etna y el Strómboli se tienen 777 registros, se deben de tener 777 ficheros etiquetados “trs” con sus respectivos 777 ficheros de onda “Wav”, además de los 777 ficheros “r50”, y de los 777 ficheros sin cabecera “Raw”. Por lo que llegado a este punto, habrían sido analizados miles de eventos en los cuatro volcanes de forma individual, y realizado la conversión de igual número de ficheros.

7. Estructura de los Formatos

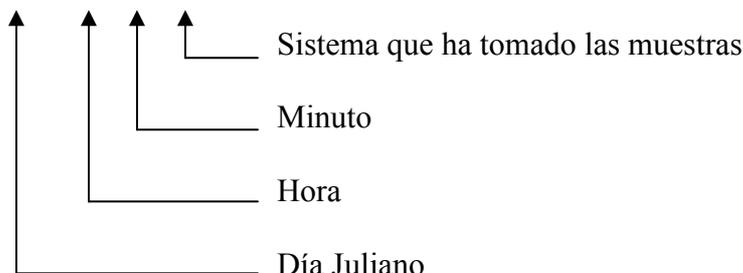
Es importante señalar, que al final las listas de los registros que se emplearán en el estudio de los cuatro volcanes, poseen un formato que en sí mismos, contienen la información básica del registro y que a continuación se detalla:

7.1.- Formatos de los registros sísmico-volcánicos

a) **Etna** -> 2540159d

En el Etna, el formato de los registros es el siguiente.

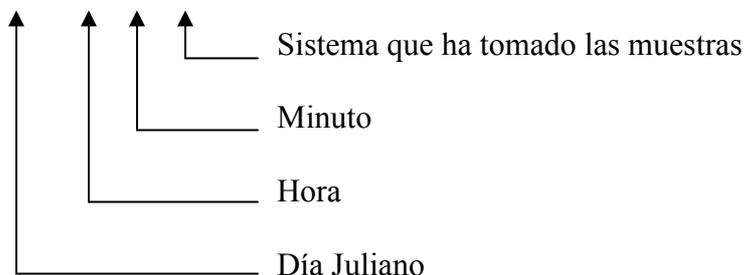
254 – 01 -59 – d



b) **Strómboli** -> 2541740G

En el Strómboli, al igual que en el Etna, el formato de los registros es el siguiente.

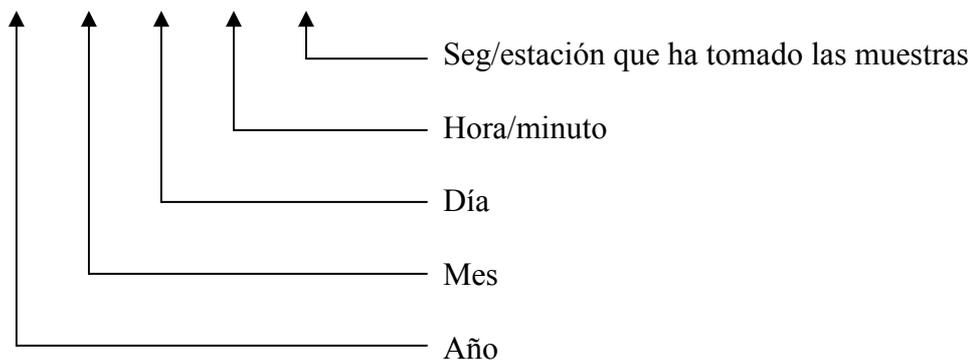
254 – 17 -40 - G



c) **San Cristóbal** -> 2006-02-18-2018-16SC

En el San Cristóbal, el formato de los registros es el siguiente

2006 – 02 – 18 – 2018 -16SC

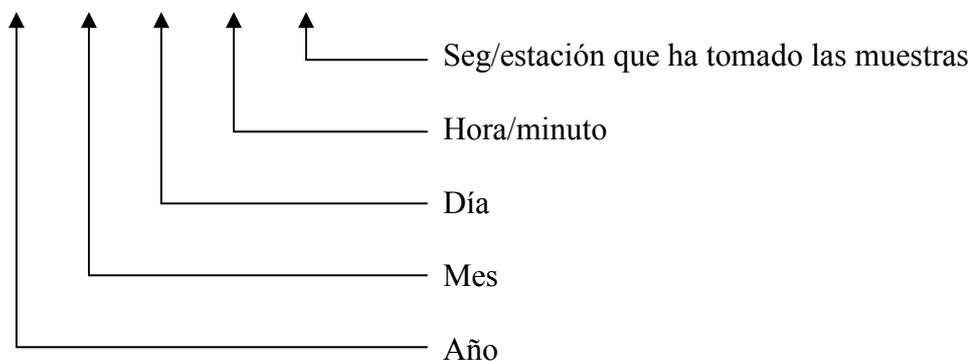


La estación para San Cristóbal es **SC**

d) **Telica** -> 2006-03-01-1411-15ST

En el Telica, al igual que en el San Cristóbal, el formato de los registros es el siguiente

2006 – 03 – 01 – 1411 -15ST



La estación para el Telica es **ST**

Los formatos anteriores brindan la información concerniente a la fecha en que se han almacenado los registros, además de que muestra la estación o el instrumento que ha tomado las muestras. Los sensores fueron programados para almacenar las fechas en las campañas de Italia, de acuerdo al calendario juliano, en cambio, en las campañas de Nicaragua en el gregoriano. Para las pruebas con los volcanes de Italia, las estaciones son las marcadas como “d” y “G”, en el Etna y Strómboli respectivamente, y para los ensayos con los volcanes de Nicaragua, las estaciones son las designadas con las siglas “SC” y “ST” en el San Cristóbal y Telica respectivamente. Con este tipo de registros se han de realizar las pruebas que se detallan en el capítulo cinco de este trabajo.

7.2.- Cabecera de información de los registros

La información o cabecera de los registros difiere en cada sistema. Para los datos de Italia, la información o cabecera del fichero, viene en un fichero diferente con extensión SAD, tanto para el Etna como para el Strómboli. En cambio para Nicaragua viene incluida dentro del mismo registro en los primeros bytes de información. En las campañas de Italia, los ficheros SAD contienen toda la información del registro sísmico. Tanto para el Etna como para el Strómboli son similares. Un ejemplo se observa en la figura F1.

```
FICHERO DTS DEL ETNA

8 trazas
30001 muestras por traza
tomadas el
99-09-11
a las
12:32:45.00
200 Hz muestreo
48 Hz 8 polos antialiasing
en las siguientes estaciones
ZD muestra mayor 2424
NSD muestra mayor 3825
EWD muestra mayor 7688
4D muestra mayor 2706
5D muestra mayor 2442
6D muestra mayor 1906
7D muestra mayor 2186
8D muestra mayor 1829
-----
Información complementaria
D Identificación Instrumento
CS5016 Serie Instrumento
Software May 95
MPS TSta TLta Pre Post Kdis Fil Dmax
200 2 16 16 32 1 1 150
Sta/Lta en incremento
```

Figura F.1 Contenido del fichero de cabecera “dts” en la campaña del Etna

La información del fichero indica cuantas trazas posee el registro, el número de muestras por traza, la fecha en que ha sido almacenado el registro, la frecuencia de muestreo a la que es tomado el registro (que es de 200 Hz), las estaciones y alguna información complementaria sobre el instrumento. El tiempo de los registros es de 600s en el Etna y 376s para el Strómboli.

La información que viene en los registros de los datos de Nicaragua se observa a continuación.

```
Registro volcán San Cristóbal
#START_TIME 2006 3 9 14 4 44.212
#SAMP_FREQ 50.01
#NDAT 15000
#STATION_CODE CRIN
#STATION_CHANNEL S Z
```

```
Registro volcán Telica
#START_TIME 2006 3 3 8 50 55.631
#SAMP_FREQ 50.01
#NDAT 15000
#STATION_CODE TELN
#STATION_CHANNEL S Z
```

La información que proporciona el registro para las estaciones de Nicaragua, es más concisa que la de Italia, proporciona el tiempo de inicio del disparo para almacenar el registro mismo que viene a ser el nombre del registro. Para el San Cristóbal, se puede observar que es “2006-03-09-1404-44” y para el Telica “2006-03-03-0850-55”. También, proporciona la frecuencia de muestreo con la que han sido registrados los eventos que en este caso es de 50 Hz. Otro aspecto es el código de la estación, que para San Cristóbal es “CRIN” y para Telica “TELN”. Además, indica el canal de la estación el cual es “Z” lo que indica que es la componente “Z” o componente vertical. Como se puede observar los registros de Italia difieren de los de Nicaragua al ser la frecuencia de muestreo de 200 Hz y la de Nicaragua 50 Hz. El tiempo de los registros para los dos volcanes es de 300s. Por lo demás, la información esencial es la misma. Al ser tomados por sismómetros de una sola componente, la vertical (Z).

APÉNDICE G: Implementación Tecnológica

1- INTRODUCCIÓN

Dentro de la Tesis doctoral, uno de los objetivos es establecer un procedimiento robusto y fiable de reconocimiento automático de señales sísmicas. Se ha demostrado que el método HMM es una herramienta potente para este reconocimiento. Puesto que esta herramienta deriva de los sistemas de reconocimiento de voz, existen algunos programas que permiten hacer este reconocimiento de una manera continua, es el caso del conjunto de herramientas HTK utilizado en el presente trabajo. Sin embargo, las señales sísmico-volcánicas requieren ser caracterizadas y preparadas para que se les pueda aplicar este proceso de reconocimiento. En la actualidad no existen muchos programas de fácil acceso y uso que permitan realizar este proceso. En este apéndice se presenta la primera versión de un programa, que permite realizar dicho proceso. Por lo que a partir de los análisis de los eventos sísmico-volcánicos, se ha logrado desarrollar un software, que provee de herramientas para desarrollar el análisis espectral de las señales. Dicho análisis, podrá permitir tener una visión mucho más clara de los tipos de eventos que se están produciendo en el volcán, ayudando así al observador humano en la comprensión de los procesos que se desarrollan en un determinado período de actividad. Lo que constituye la primera parte del desarrollo que se ha expuesto en el presente trabajo y constituye una parte esencial del posterior reconocimiento y clasificación de dichas señales mediante los Modelos Ocultos de Markov (HMMs).

La aplicación aquí desarrollada consta de varias secciones, determinadas inicialmente por dos pantallas que presentan la entrada al sistema, la primera de ellas es una interfaz de bienvenida a manera de inicio, seguida de una segunda pantalla que contiene el menú principal. A partir de dicho menú, el sistema se desarrolla en dos grupos principales. En el primer grupo se presenta una introducción y ejemplo del análisis espectral con eventos ya definidos. En esta parte, se realizan ejemplos mediante una serie de registros sísmico-volcánicos realizados mediante análisis propuestos, a partir de datos pre-establecidos de los cuatro volcanes estudiados en el presente estudio.

En un segundo grupo, constituye el núcleo del sistema, debido a que aquí se presenta la conversión de los ficheros y el análisis espectral de los ficheros ya convertidos. En la primera parte, se presenta la conversión de señales desde dos formatos básicos (ASCII y SAC), a señales de tipo onda Wav para ser posteriormente analizadas de acuerdo a los parámetros espectrales que se seleccionen. La segunda parte, se realizan diferentes análisis espectrales con los eventos previamente convertidos en la primera parte (u otros ya convertidos), como son: análisis de la señal en el dominio del tiempo, espectrogramas, envolventes de la señal, espectros de frecuencia mediante análisis de Fourier, densidad espectral de potencia de la señal, análisis wavelet de las señales e histogramas de la secuencia de datos. Además, de graficar el sistema también permite poder escuchar las señales, algo bastante útil al observador que está tratando de discernir y reconocer entre la emisión auditiva de un cierto tipo de evento y otro muy distinto. Por último, el sistema presenta una opción para poder salir de la aplicación una vez finalizadas las acciones y retornar al sistema operativo.

El sistema ha sido desarrollado utilizando Matlab V.9 para 64 bits, mediante la programación de interfaces gráficas de usuario (GUI). Las diferentes interfaces se

encuentran almacenadas en un directorio con sus correspondientes programas en código fuente (archivos .m) y los ficheros de figuras (.fig).

2.- DIAGRAMA DE FLUJO DE LAS INTERFACES DEL SISTEMA

El diagrama de flujo del sistema se presenta en la figura G1:

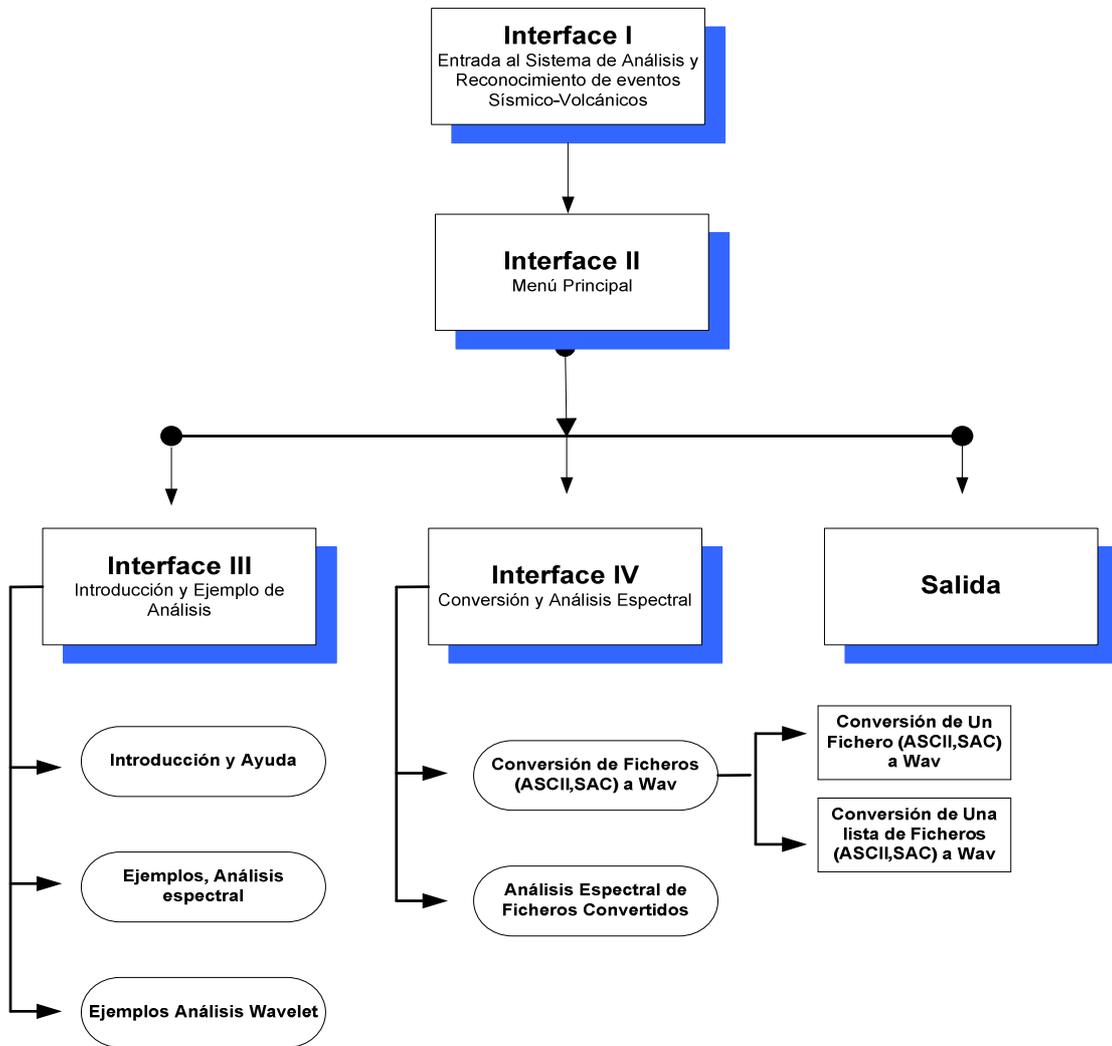


Fig. G.1 Diagrama de flujo con las interfaces del sistema

El diagrama anterior muestra los bloques en los que se divide el sistema. Las dos primeras interfaces son la de inicio y bienvenida, a través de la cual se accede al sistema y la interfaz del menú principal. La interfaz número tres representa la introducción y ejemplos de análisis espectral a través de registros de los cuatro volcanes estudiados en este trabajo. La cuarta interfaz representa el proceso de conversión de los registros desde dos tipos de formatos, ASCII y SAC a formatos de onda Wav para poder realizar los posteriores análisis espectrales. El proceso de conversión se realiza tanto de ficheros individuales como de listas. Al final de dicho menú se presenta la opción de salida del sistema. Futuras modificaciones, podrían incluir en la conversión, diferentes tipos adicionales de formatos.

3.- DESCRIPCIÓN DE LAS INTERFACES DEL SISTEMA

3.1.- Interfaz de inicio al sistema

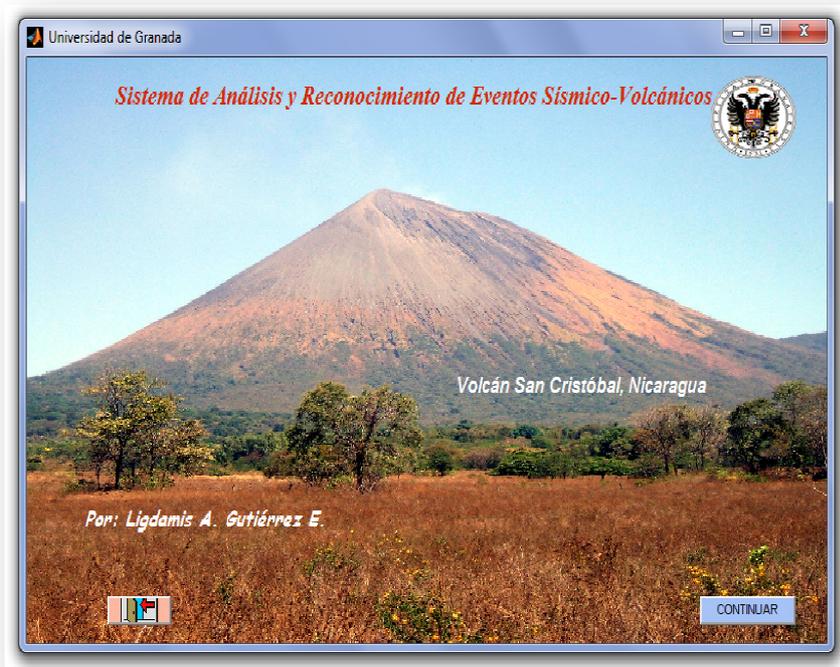


Fig. G.2 Interfaz de Inicio del Sistema de Análisis y Reconocimiento de Eventos Sísmico-volcánicos

A través de esta interfaz se da acceso al sistema. Se pueden observar en la figura anterior, dos botones de acción. Uno con la leyenda “Continuar”, que permite continuar con el proceso e introduce a la siguiente interfaz. El segundo botón de comando permite la salida del sistema si así se prefiere.

3.2.- Interfaz del menú principal

El menú principal consta de tres bloques principales. En el primer bloque se establece como primer punto una llamada para la introducción y ayuda. El segundo y tercer punto de este bloque están destinados a ejecutar acciones con eventos predefinidos. Estas acciones representan diferentes análisis espectrales de los eventos. En el segundo bloque, se establece como primer punto, una llamada para la interfaz de conversión y representación gráfica de ficheros. En el segundo punto se realizan diversos análisis espectrales con los ficheros convertidos. Para finalizar, el tercer bloque está compuesto por un botón de acción que permite salir del sistema. Esto se observa en la figura G.3.

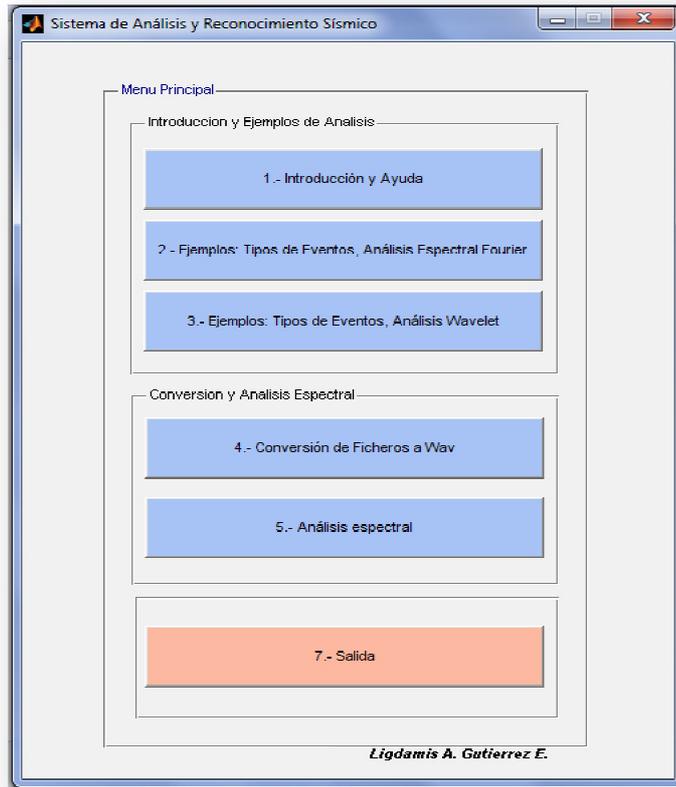


Fig.G.3 Interfaz de Menú Principal de la Aplicación

3.3.- Interfaz de ayuda

La pantalla de interfaz de ayuda presenta los créditos del sistema. A futuro, podrá presentar una serie de comandos de ayuda a medida que vayan creciendo las posibilidades del sistema. Está contemplada para poder auxiliar al usuario, sobre la documentación general de la aplicación. Dicha interfaz es una página Web, que ha sido creada mediante un script de Matlab y se accede a ella a través del botón de comando (punto primero), del menú principal. La primera parte de dicha interfaz, lleva los contenidos que son enlaces a los puntos principales de ayuda. Dichos puntos enmarcan los diferentes aspectos que contiene la aplicación. Y son los siguientes:

- a) Introducción
- b) Diagrama del Sistema
- c) Entrada a la aplicación (Interfaz de inicio)
- d) Menú Principal
- e) Interfaz de ejemplos de eventos
- f) Interfaz de Conversión de Ficheros
- g) Interfaz de análisis espectral

Al igual que una página Web, dicha interfaz puede navegarse a través de las barras de desplazamiento situadas en los extremos de la pantalla o con las flechas de dirección del teclado. La ubicación de dicha interfaz, se encuentra en el directorio:

“C:\Nicaragua\html\Docum.html”. Una perspectiva de dicha interfaz se observa en la figura G.4.

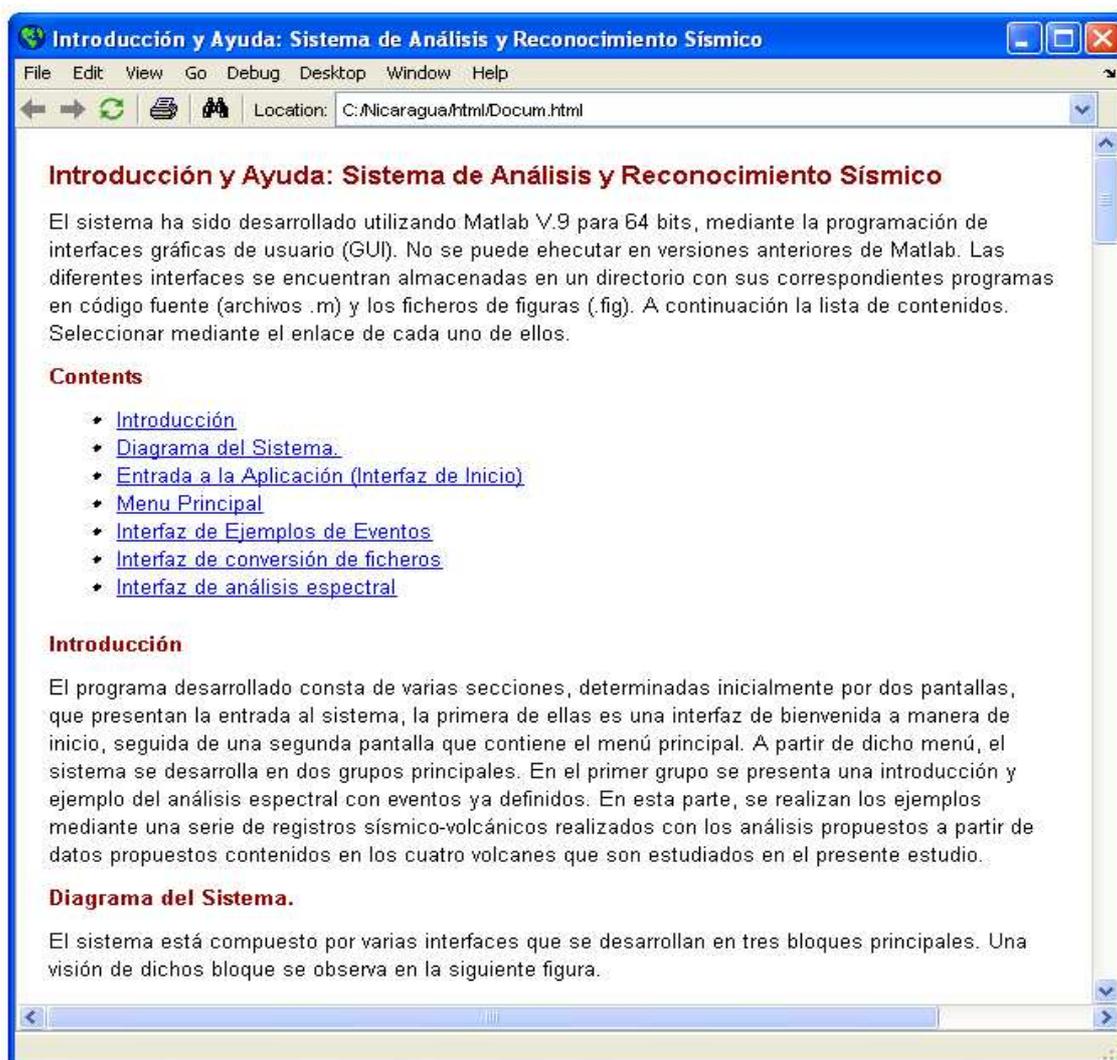
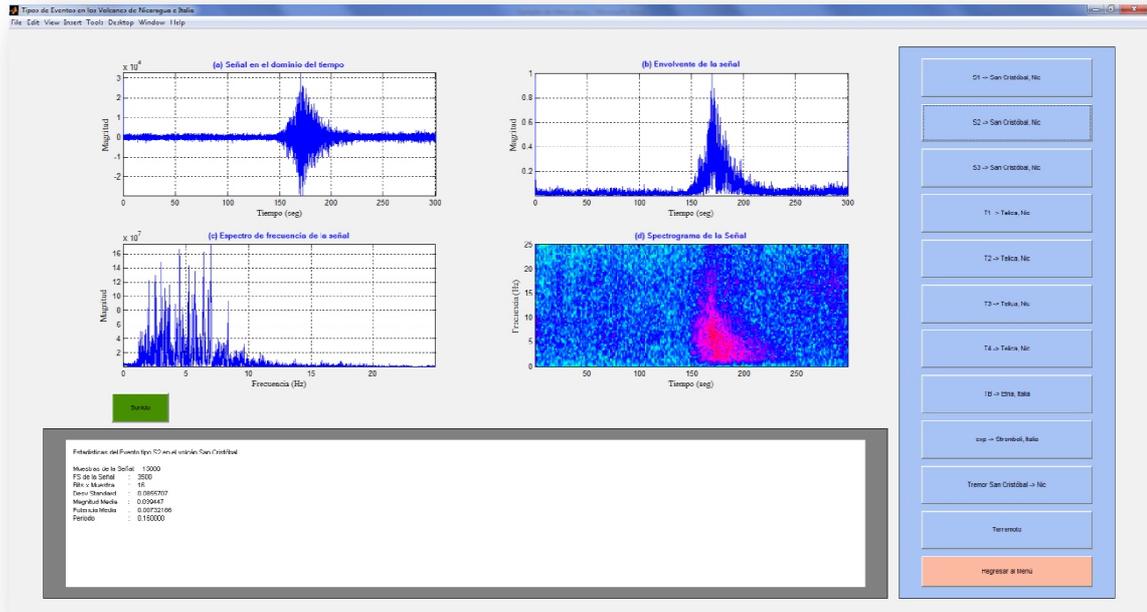


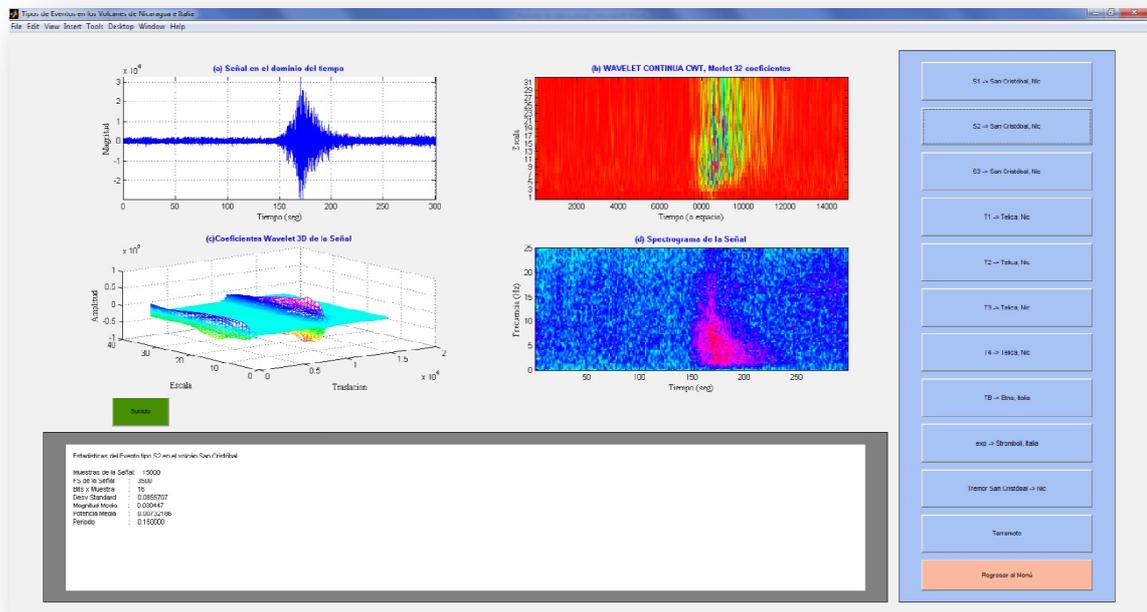
Fig.G.4 Interfaz de Ayuda

3.4.- Interfaz de ejemplos de eventos

Esta interfaz está diseñada para representar a manera de ejemplo lo que puede desarrollar el sistema con los eventos. Previamente está ya definida para ser ejecutada con eventos que han sido definidos en los cuatro volcanes. Consta de dos pantallas principales, que corresponden a los puntos 2 y 3 del menú principal (Fig. G.3), en cada una de dichas opciones son seleccionados once diferentes tipos de eventos. La primera pantalla (Fig.G.5), representa la señal en el dominio del tiempo, mediante un análisis de Fourier, la envolvente de la señal, el espectro de frecuencia de la señal y la señal en el dominio de la frecuencia o el espectrograma. La segunda pantalla (Fig.G.6), vuelve a incluir la representación de la señal en el dominio del tiempo y su espectrograma. Esto se hace para poder comparar las gráficas, con la aplicación del análisis wavelet. De esta forma, se pueden ver los distintos tipos de variación en las energías de los componentes de la señal presentes en los eventos sísmico-volcánicos.



La siguiente figura muestra ejemplos de los eventos con Análisis Wavelet



Se puede observar en la figura, la señal en el dominio del tiempo, además de tres gráficas adicionales, una de ellas en la parte inferior derecha muestra el espectrograma del evento, las otras dos indican un análisis Wavelet en dos y tres dimensiones.

3.5.- Interfaz de conversión de ficheros

Esta interfaz provee las herramientas para poder convertir los tipos de ficheros básicos manejados en el observatorio de Nicaragua (ASCCI y SAC), a un formato de tipo onda Wav para posteriormente, ser analizados mediante sus características espectrales. Se presentan diferentes bloques de acciones en la pantalla:

a) Datos a convertir. Aquí se selecciona el tipo de evento original a convertir, ya sea de tipo ASCII o SAC. También delimita si se trata de un fichero simple, o una lista a convertir. Además se le indica al sistema la extensión del fichero original a convertir (.001 si es de tipo ASCII o de .SAC si es SAC). Nota: Para la construcción de las listas, la primera línea debe de indicar el número de registros a convertir. Los ficheros SAC se encuentran en *Little Endian*.

b) En esta sección se definen los valores de entrada y salida. Se designa el directorio donde se encuentran los datos o ficheros originales a convertir (directorio de entrada de datos). Se designa el directorio donde se van a almacenar los ficheros ya convertidos (directorio de salida de datos). Por último, en este bloque se establece el valor de la frecuencia de muestreo en Hz., con la que convertir los ficheros.

c) Este tercer bloque es el que establece los comandos de acción una vez definidos todos los anteriores parámetros. Consta de tres botones de acción, el primero establece la conversión, la cual si es llevada con éxito aparecerá una ventana emergente que así lo denota. El segundo botón permite graficar dicho fichero convertido. En caso de que alguno de los parámetros no sea correcto, diferentes ventanas de error se presentarán de acuerdo a la acción a tomar. Existen secciones (Dato de Entrada y Fichero de Salida), en donde se pueden visualizar las rutas definidas por el usuario. De manera similar, se presentan dos comandos o botones de acción que permiten limpiar las entradas y eliminar el fichero en caso que se desee, por ejemplo si se tratase de un registro que contenga únicamente ruido y ningún evento reconocible.

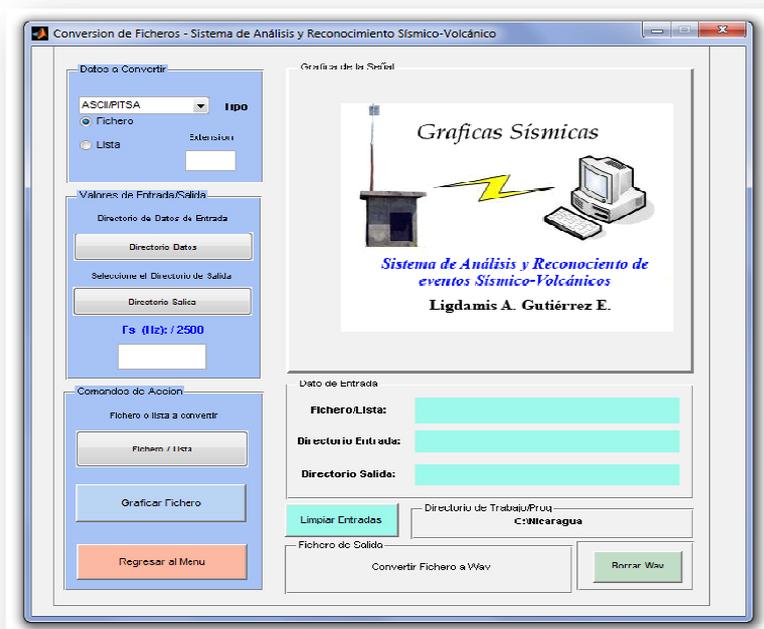


Fig. G.7 Interfaz del bloque II, Conversión de Ficheros.

De acuerdo a la interfaz de “Datos a Convertir”, el primer paso es elegir qué tipo de dato se desea convertir, si es de tipo ASCII o de tipo SAC. Seguidamente se debe de elegir si se trata de un fichero o si se desea convertir una lista de ficheros. Se debe de establecer qué tipo de extensión poseen el o los ficheros a convertir. La ruta y el nombre del fichero o de la lista se presenta en la parte denominada “Dato de Entrada” en el apartado “Fichero/Lista”. En la interfaz de la figura G.6, en el bloque de comandos de la parte izquierda, se designa el directorio en donde están ubicados los datos o registros a convertir y el directorio de salida. Esto se realiza en la parte denominada “Valores de Entrada/Salida”. La información de dichas entradas se presenta en el mismo bloque de “Datos de Entrada”, bajo los apartados de “Directorio de Entrada” y “Directorio de Salida”. Adicionalmente se debe designar el valor de la frecuencia de muestreo, dado en Hz. Si no se asigna ningún valor se presentará una pantalla de error indicando que se debe establecer un valor válido en dicha casilla. Cuando ya todos los datos están seleccionados se presiona el botón de “Fichero/Lista”, que se encuentra en el bloque de comando de la parte izquierda de la interfaz bajo el título de “Comandos de Acción”, con lo que se despliega una pantalla (Fig. G.8), mostrando la lista de ficheros que se desean convertir.

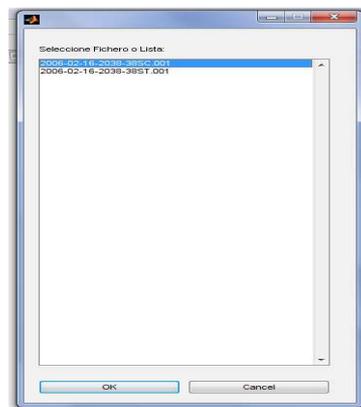


Fig. G.8 Pantalla de selección de un fichero

Al realizar lo anterior se procederá a la conversión con los parámetros establecidos. El nombre del fichero o lista resultante se presentará en la parte denominada “Fichero de Salida”. Si el proceso ha tenido éxito se presentará una pantalla que confirma esto, similar a la que se presenta a continuación (Fig.G.9).



Fig. G.9 Pantalla de éxito en la creación de un fichero

Una vez que el fichero ha sido creado, para visualizar la señal se le da clic al botón de comando “Graficar Fichero”, con lo que se presenta la señal en el dominio del

tiempo de dicho registro. En la interfaz quedan indicada toda la información de la configuración anterior. Esto se observa en la figura G.10.

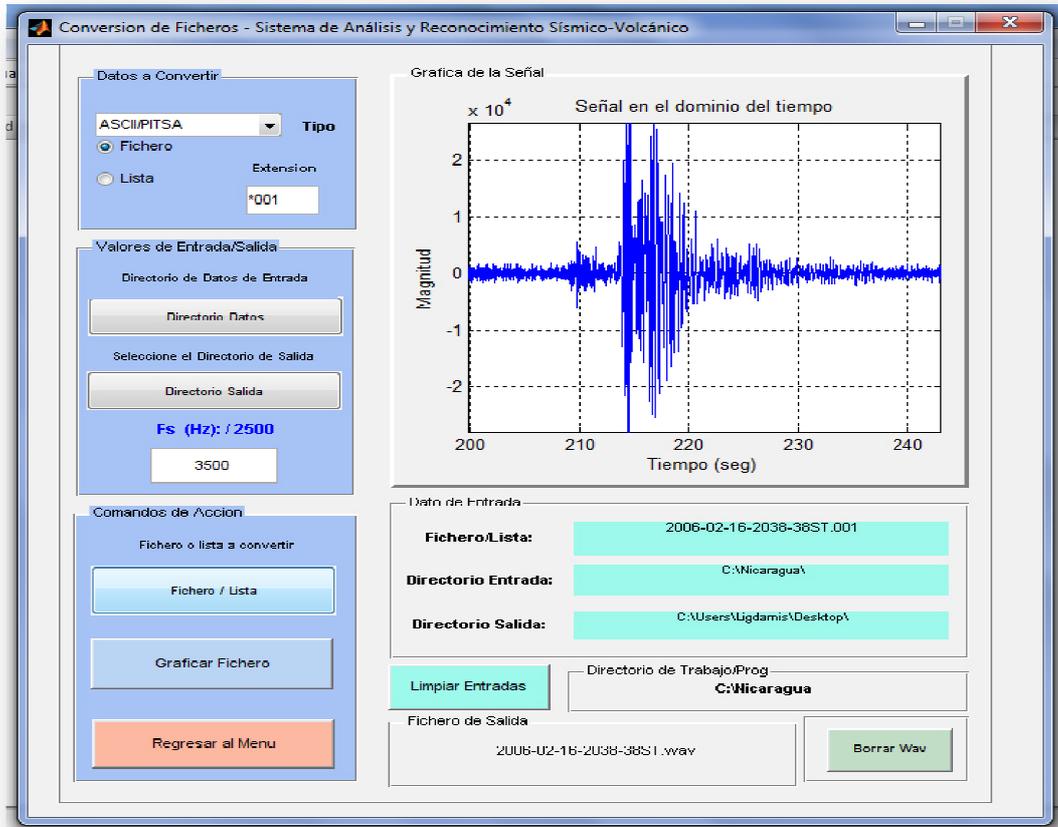


Fig. G.10 Interfaz del bloque II, Ejemplo de Conversión de Ficheros

Como se observa en la figura anterior, la señal se presenta en la parte denominada “Grafica de la señal”, pudiéndose realizar un acercamiento o “zoom” a dicha gráfica para poder observar con más detenimiento las características del evento. En la pantalla de la interfaz presentada en la figura anterior también se observa la información general del proceso de carga y conversión del registro. Es decir: la extensión del registro original, así como su nombre y ubicación. El valor de la frecuencia utilizada en la conversión, el nombre y la ubicación del directorio en donde se almacenará el fichero.

Es importante destacar que si el fichero que se ha convertido no es útil o presenta por ejemplo solo ruido, se puede eliminar, mediante el botón de comando “Borrar Wav”, esto es útil, ya que desde esta aplicación se descartan los registros que no contengan eventos significativos o que son determinantes de acuerdo al observador. En caso de que se necesite procesar otro registro se puede hacer uso de una nueva elección o proceder a la limpieza de los datos mediante el botón de comando “Limpiar Entradas”. Sin embargo si se aplica dicho comando, se tendrá que introducir de nuevo toda la información.

3.6.- Interfaz de análisis espectral

Una vez realizada la conversión de un fichero o lista mediante la interfaz anterior, el siguiente paso es realizar un análisis espectral del fichero convertido. Este proceso se realiza mediante la interfaz de la figura G.11. En el bloque de entrada de datos se selecciona el fichero y el directorio donde se encuentra almacenado dicho fichero. Además se establece el valor de la frecuencia de muestreo que puede ser diferente a la que se utilizó para convertir el fichero. Esto es útil para poder escuchar el sonido del evento a diferentes frecuencias, ya que existe un botón de comando que permite escucharlo. Una vez que los datos han sido cargados al sistema se procede a seleccionar el tipo de análisis que se desea mediante un menú desplegable. Una vez escogido dicho análisis se selecciona el botón de comando para graficar dicho análisis. Con estas acciones se tiene además de la representación gráfica, el sonido característico de dicho evento. Existe una sección en donde se presenta la información estadística del evento. También el usuario puede observar la información tanto de fichero que se va a analizar como su ubicación en la sección de información de Salida. Adicionalmente, existe un comando de acción para la limpieza de entradas ante cualquier error.



Fig. G.11 Pantalla de Interfaz para el análisis espectral de los ficheros

La imagen anterior muestra todos los elementos ya mencionados. Para elegir un registro que ya ha sido convertido en el proceso anterior, se hace uso del bloque denominado “Entrada de Datos”, situado en la parte izquierda de la interfaz. En dicho bloque se establece el sitio o el directorio en donde están almacenados los ficheros a analizar, el nombre del fichero y la frecuencia de muestreo designada para representar el

sonido de la señal sísmica-volcánica o evento elegido. Una vez hecho esto se elige el tipo de análisis que se le va a asignar a la señal. Esto hace que automáticamente se presenta en la parte denominada “Gráficas”, la señal con dicho tipo de análisis, así como la información estadística de dicho evento. La información de la ubicación y el nombre del registro se presenta en el bloque denominado “Información de Salida de los Ficheros”. Un ejemplo de este proceso se muestra a continuación en las siguientes figuras. Una vez que se le da clic al botón de “Fichero”, se presenta la pantalla siguiente en donde se procede a la selección del fichero a analizar

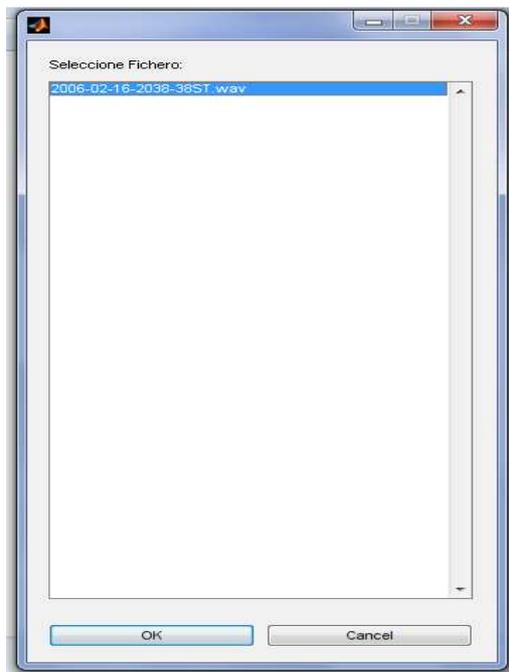


Fig. G.12 Pantalla de selección de los ficheros a analizar

Cuando se ha designado el fichero y elegido el tipo de análisis se muestra una pantalla similar al de la figura G.13, en donde se despliega una lista con los diversos tipos, así como los resultados y una información estadística básica de dichos análisis. Una vez seleccionado el tipo deseado se le da clic al botón de “Graficar” para observar y analizar la gráfica resultante. Adicionalmente, se dispone de un botón de comando que permite limpiar las entradas y otro botón de comando que regresa al menú principal.



Fig. G.13 Pantalla con un ejemplo de la Interfaz para el análisis espectral. Se observa el menú de análisis

A continuación se observan a manera de ejemplo, tres tipos de análisis espectrales elegidos de la lista (Fig.G.13), mediante el mismo fichero (a través de la interfaz de análisis), lo que se representa en las tres figuras (Fig. G.14, G.15 y G.16) que a continuación se muestran. En la primera pantalla (Fig. G.14) se observa un análisis del espectro de frecuencia de la señal. En la segunda (Fig. G.15) se observa la señal mediante la transformada Wavelet continua. La tercera pantalla (Fig. 16), presenta el mismo análisis en 3D.

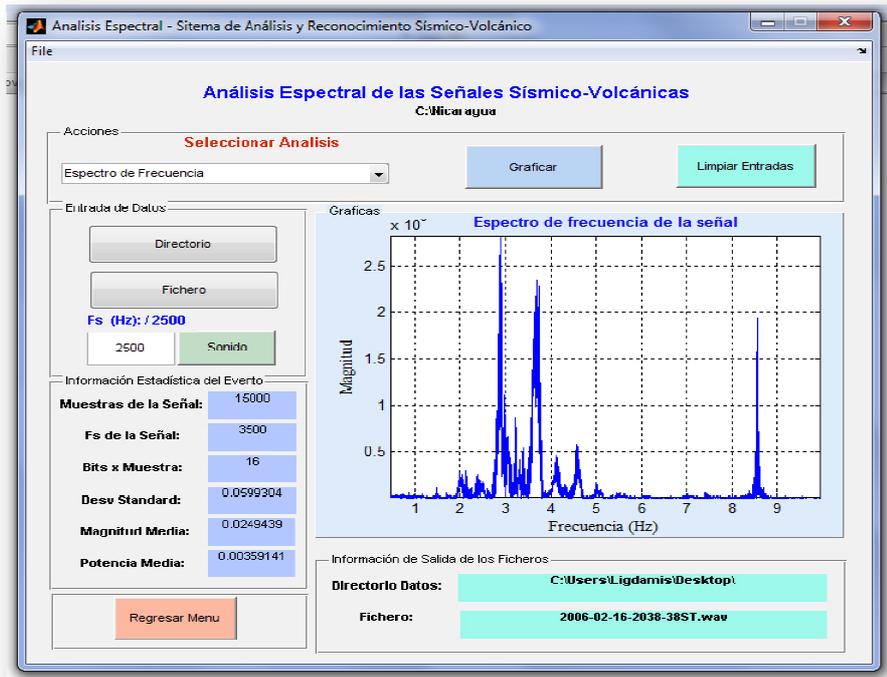


Fig. G.14 Pantalla de análisis espectral. Selección de análisis por espectro de frecuencia

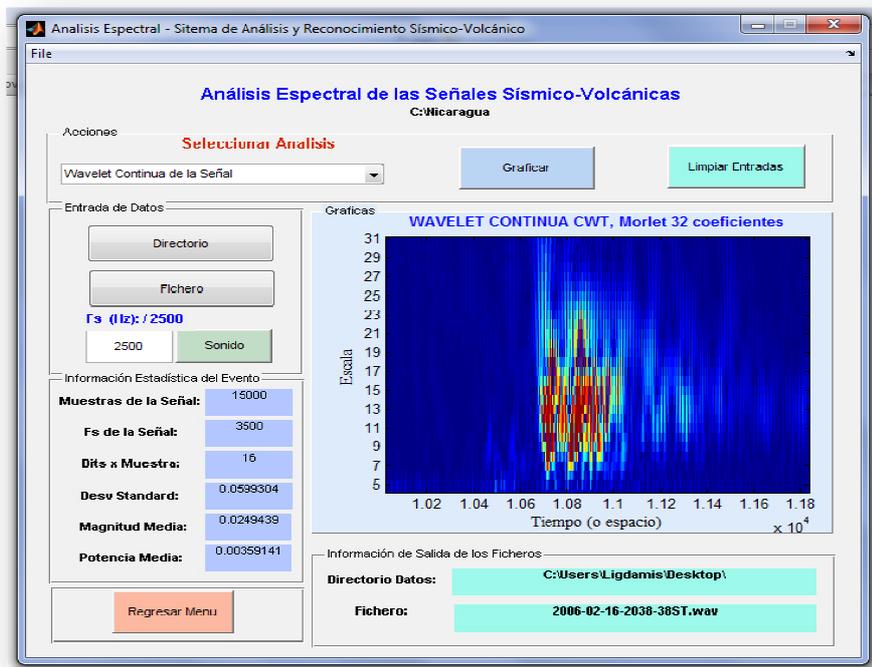


Fig. G.15 Pantalla de análisis espectral. Selección de análisis Wavelet

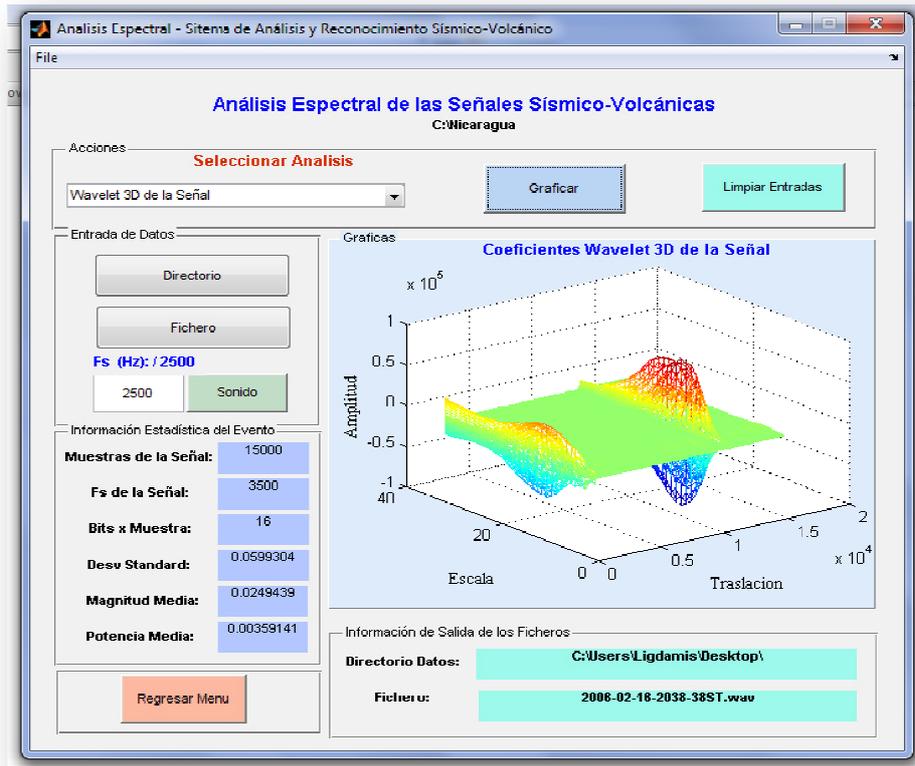


Fig. G.16 Pantalla de análisis espectral. Selección de análisis Wavelet 3D

4.- PANTALLAS DE ERROR

4.1.- Pantallas de error en la conversión de ficheros

El sistema presenta diversas pantallas de error en caso de que algún dato o entrada no esté correcto. De esta forma, se pueden recuperar las entradas y seguir ejecutando la aplicación sin ningún problema crítico de salida. Son las siguientes:

- a) Pantalla de error en la conversión de ficheros



Fig. G.17 Pantalla de error en la Conversión de Ficheros

b) Pantalla de error en la lectura de fichero de datos



Fig. G.18 Pantalla de error al seleccionar un fichero

c) Pantalla de error al graficar fichero

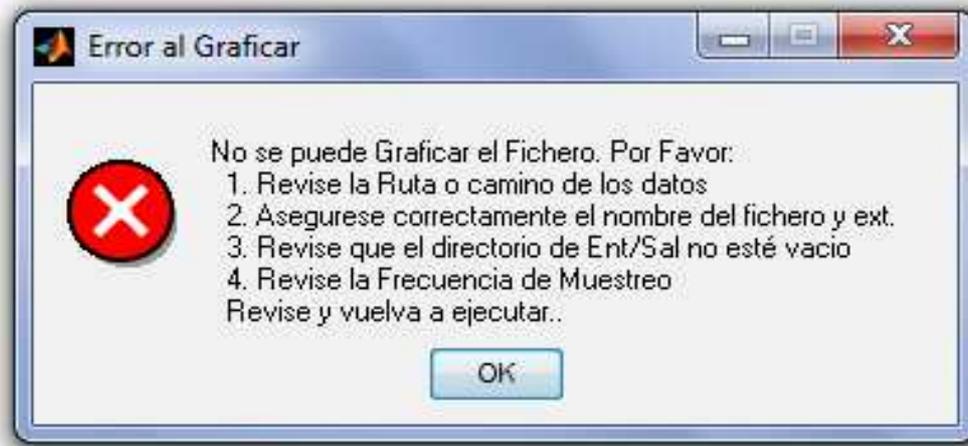


Fig. G.19 Pantalla de error al graficar un fichero

d) Pantalla de error al reproducir el sonido del fichero



Fig. G.20 Pantalla de error al reproducir un fichero

4.2 Pantallas para la eliminación de un fichero creado

a) Pantalla para solicitar permisos previos a eliminar un fichero Wav



Fig. G.21 Pantalla de confirmación para eliminar un fichero

b) Pantalla de confirmación de la eliminación del fichero Wav



Fig. G.22 Pantalla que confirma que el fichero ha sido eliminado

4.3.- Pantallas de error en selección de ficheros

a) Pantalla de error al seleccionar la ruta de los ficheros



Fig. G.23 Pantalla de error al seleccionar la ruta de los ficheros

b) Pantalla de error en la entrada de datos



Fig. G.24 Pantalla de error en la entrada de los ficheros

Todas las pantallas anteriores de error y advertencia se presentan cuando el usuario realiza un procedimiento no válido. Estas pantallas a su vez, permiten restablecer el sistema sin provocar un fallo generalizado en la ejecución del mismo. De esta forma, el proceso de ejecución retorna el valor al procedimiento anterior que lo ha llamado.

De acuerdo a como está estructurada, la aplicación a futuro puede enriquecerse mediante la adición de módulos como pueden ser: aumento de más tipos de formatos para la conversión de ficheros, módulo para la segmentación de las señales, módulo para el reconocimiento mediante modelos HMMs. Cualquier otro módulo que pueda ser útil para un observador sismológico en un observatorio. Los códigos de los programas (incluidos en la aplicación), pueden modificarse de acuerdo a las necesidades que se estimen convenientes.



APÉNDICE H: Programas

En este apéndice se describen los principales script, que son necesarios para la generación de las gramáticas, los diccionarios, la lista de transcripciones con el fichero “mlf”, las estadísticas de los eventos y el fichero de entrenamiento y reconocimiento utilizados en las pruebas realizadas durante el desarrollo de la tesis.

1. Creación de Gramáticas

La construcción de las gramáticas necesarias para el entrenamiento de los modelos, puede realizarse mediante el uso de un editor de texto, y así crear un fichero de texto denominado:”gram.gram”, el cual contiene una lista de los eventos que deben de contener la gramática de red.

gram.gram

```
$evento = NT | T1 | T2 | T3 | T4 ;
( < $evento > )
```

Seguidamente se hace uso del script denominado “crea_gram.csh” para realizar la creación de la gramática de red. El contenido del script es el siguiente:

crea_gram.csh

```
#!/bin/csh
# Sitio en donde se ubica el fichero gram

set GRAM = /Nicaragua/datos/gram.gram

# llamada a la herramienta HParse de HTK que construye
# la gramática de red.
# El fichero resultante se denomina gram.wnet

HParse $GRAM gram.wnet
```

En la segunda línea de este script se establece la ubicación del fichero “gram.gram” que contiene los eventos a representar en la gramática. El fichero invoca a la herramienta de HTK “HParse” para crear el fichero de salida con la gramática el cual se denominará gram.wnet. El contenido de este fichero de salida ya generado es el siguiente:

gram.wnet

```
VERSION=1.0
N=9 L=19
I=0 W=T5
I=1 W=!NULL
I=2 W=T4
I=3 W=T3
I=4 W=T2
I=5 W=T1
I=6 W=NT
I=7 W=!NULL
I=8 W=!NULL
J=0 S=0 E=1
J=1 S=1 E=7
J=2 S=1 E=6
J=3 S=1 E=5
J=4 S=1 E=4
J=5 S=1 E=3
```

J=6	S=1	E=2
J=7	S=1	E=0
J=8	S=2	E=1
J=9	S=3	E=1
J=10	S=4	E=1
J=11	S=5	E=1
J=12	S=6	E=1
J=13	S=8	E=6
J=14	S=8	E=5
J=15	S=8	E=4
J=16	S=8	E=3
J=17	S=8	E=2
J=18	S=8	E=0

Se denomina “Red” a la secuencia de palabras que pueden reconocerse, mientras que el diccionario describe los alófonos que componen cada una de las palabras. Pueden crearse distintos tipos de redes dependiendo del modelo de lenguaje utilizado, (todas estarán descritas mediante el formato: Estándar Lattice Format). La sintaxis divide la red en tres partes: cabecera (es opcional; contiene información general sobre el lattice), nodos y enlaces. Los enlaces pueden ser: unigrama (probabilidad de una palabra) o bigrama (transiciones posibles entre unas palabras hacia otras).

2. Diccionarios de eventos

El contenido de los ficheros de diccionarios de eventos que se utilizan en los entrenamientos y reconocimiento de los modelos son simplemente dos ficheros de texto con la extensión “.txt” y uno con la extensión “.dic”, que contiene la etiqueta de cada evento. Un ejemplo de dichos diccionarios se observa a continuación con los eventos del Telica:

eventos.txt

NT
T1
T2
T3
T4

eventos_def.txt

NT
T1
T2
T3
T4

eventos_def1.dic

NT NT
T1 T1
T2 T2
T3 T3
T4 T4

El orden en que se encuentran los eventos, obedece simplemente al orden de aparición de los eventos en los ficheros de transcripción.

3.- Creación de ficheros cepstrales parametrizados mfc

Un paso previo a realizar es crear una lista que contenga los ficheros de origen, (ficheros r.50) y los ficheros de destino (mfc). Conjuntamente, se crea el script o fichero de configuración general con los parámetros cepstrales que contendrán los ficheros mfc. Este script se nombra “Hcopy.conf”. El fichero de configuración se muestra a continuación.

hcopy.conf

```

# Los valores son sobre un factor 100 en la Frecuencia de muestreo
# Esto es, a 20KHz en lugar de 200Hz
SOURCEFORMAT      = NOHEAD
# Coeficientes cepstrales comprimidos en el Rango de frecuencias
# Más los parámetros de energía, Delta y aceleración
TARGETKIND        = MFCC_E_D_A
# Frecuencia de muestreo de entrada
SOURCERATE        = 500.0
# Frecuencia de 'features' (desplaza 100 muestras = 50000/500)
# Indica el rango de salida de la señal muestreada
TARGETRATE        = 50000.0
# Ventana de análisis = (800 muestras = 400000/500)
WINDOWSIZE        = 400000.0
# Se estrecha la ventana de muestreo de tipo Hamming
USEHAMMING        = T
# Energía logarítmica normalizada
ENORMALISE        = T
# Parámetro de configuración booleana,
# para mejorar el rendimiento del análisis de la señal
ZMEANSOURCE       = T
# Coeficientes de pre-énfasis
PREEMCOEF         = 0.0
# Numero de coeficientes cepstrales
NUMCEPS           = 12
# Numero de canales
NUMCHANS          = 23
# Mínima frecuencia del banco de filtros; 0Hz
LOFREQ            = 0
# Máxima frecuencia del banco de filtros; 20Hz
HIFREQ            = 2000
# Se leerán y escribirán los datos binarios en orden natural
# Es decir, sin realizar intercambio de bytes
NATURALWREADORDER = TRUE
NATURALWRITEORDER = TRUE

```

Además, para trabajar con los ficheros es necesario crear una lista de los registros que se van a utilizar. Una vez creadas las listas, que contendrían los ficheros tanto de origen como de destino y el fichero de configuración anterior, se ejecuta la librería de HTK “Hcopy”, mediante el siguiente comando:

HCopy -C hcopy.conf -S hcopy.tmp > & hcopy1.log &

En donde “-C”, hace alusión al fichero de configuración a utilizar y “-S”, al fichero que contiene las listas de origen y destino de los ficheros cepstrales a crear. Por cada fichero de origen se creara un fichero con los parámetros cepstrales “mfc”.

4. Lista de Transcripciones y Estadística de los eventos

Para obtener una mejor y mayor fiabilidad en el reconocimiento, se necesitan obtener al menor 200 eventos de cada tipo.

Este procedimiento se realiza a partir de la segmentación y etiquetación manual de los ficheros, lo que da un fichero resultante con extensión “trs”, muy similar a la siguiente figura:

```

<?xml version="1.0" encoding="ISO-8859-1"?>
<!DOCTYPE Trans SYSTEM "trans-13.dtd">
<Trans scribe="(unknown)" audio_filename="2541132d" version="1"
version_date="050115">
<Episode>
<Section type="report" startTime="0" endTime="1.5">
<Turn startTime="0" endTime="1.5">
<Sync time="0"/>
NE
<Sync time="0.084"/>
TB
<Sync time="0.268"/>
NE
<Sync time="0.629"/>
TB
<Sync time="0.74"/>
NE
<Sync time="0.82"/>
TB
<Sync time="0.978"/>
NE
</Turn>
</Section>
</Episode>
</Trans>

```

Se observa cómo cada evento etiquetado posee un tiempo de inicio y final. El procedimiento a seguir para obtener una lista de las transcripciones a partir de los ficheros etiquetados es el siguiente:

- a) Es necesario construir ficheros con la duración de cada uno de los registros. Para elaborar estos ficheros, se debe obtener el fichero maestro de los eventos “mlf”, mediante un grupo de tres scripts. Los cuales son los siguientes:

Script No. 1

La salida de este fichero es el fichero denominado “data_segml.mlf”, que contiene las transcripciones de los registros segmentados. Para ello, dicho fichero llama a una lista con el registro de los nombres de ficheros, la cual se encuentra en el fichero “data_segml.scp”. A cada registro de esta lista, se le asigna la extensión de los ficheros del Transcriber “trs”. Los elementos de esta lista se procesan con el segundo script trs2mlf.csh.

crea_mlf_data.csh

```

#!/bin/csh

# Nombre del fichero de salida asignado
set SALIDA = data_segml.mlf

# Cabecera de Salida
echo '#\!MLF\!#' > $SALIDA

# Por cada línea del fichero
foreach f ( `cat data_segml.scp` )
    echo $f.trs
    trs2mlf.csh $f.trs >> $SALIDA
end

```

Script No. 2

En este script se realiza la llamada a cada uno de los elementos de la lista, para poder construir los elementos del fichero de transcripción. La línea de comando que realiza la creación es “gawk”, el parámetro “-f”, indica el fichero que contiene la

subrutina para la creación, y se utiliza una llamada a una variable en “-v”, en donde está contenido cada uno de los elementos del fichero de transcripción.

trs2mlf.csh

```
#!/bin/csh

if($#argv != 1) then
    echo "Formato: trs2mlf.csh <frase.tr>"
    exit
endif

set frase = $argv[1]
set nombre = `basename $frase .trs`
cat $argv[1] | tr -s '
"<>/' ' -' | gawk -f trs2mlf.awk -v FICHERO=$nombre
```

NOTA: Este script llama al fichero “trs2mlf.awk” que con la ejecución del comando “gawk” realiza la creación de los ficheros de transcripción “mlf”.

Script No. 3

trs2mlf.awk

```
BEGIN {
    inicio = 0;
    final = 0;
    mini = -1;
    mfin = -1;
    ts = 1e7;
    etiqueta = "UNKNOWK";
    printf("\n\*\*/%s.lab\n\n", FICHERO);
}

{
    if($1=="Turn") {
        inicio = $3;
        final = $5;
    }

    if($1=="Sync") {
        mini = mfin;
        mfin = $3;
        if((mini>=0) && (mfin>=0)) printf("%d %d
%s\n", mini*ts, mfin*ts, etiqueta);
    }

    if($1=="-Turn") {
        mini = mfin;
        mfin = final;
        printf("%d %d %s\n", mini*ts, mfin*ts, etiqueta);
    }

    if(NF==1) etiqueta = $1;
}

END {
    printf(".\n");
}
```

La salida final de estos tres scripts, es el fichero de transcripciones o “MASTER LABEL FILE (MLF)”, denominado “data_segml.mlf”, que por cada uno de los registros indica su correspondiente fichero de transcripción “lab”. Un ejemplo del formato de salida, en donde se observan los dos primeros registros obtenidos es el siguiente:

```
#!MLF!#
"/2006-03-02-2106-00ST.lab"
0 1500000 NT
1500000 2060000 T4
2060000 7500000 NT
.
"/2006-03-02-2116-00ST.lab"
0 5390000 NT
5390000 6060000 T2
6060000 7500000 NT
.
```

Nota: Se coloca un punto al final indicando el fin de la lista

La anterior lista de ficheros se obtiene a partir de los ficheros resultantes de etiquetar los registros con el Transcriber. Es decir se utilizan los ficheros “trs”. Esta misma lista se puede obtener a partir de los ficheros creados con la herramienta “HSLab”, mediante el siguiente script:

Script No. 4

```
#!/bin/csh

# script para generar el MASTER LABEL FILE (MLF)
# a partir de los ficheros lab etiquetados con el
# HSLab de HTK

# coloca la línea inicial
echo "#\!MLF\!#" > $1

# lista todos los ficheros con extensión lab
# y los re-direcciona a una variable con un fichero
# de extensión scp llamado l.scf
ls *.lab > $1.scf

# por cada uno de los ficheros lab
# escribe los contenidos y al final
# de cada fichero coloca un punto "."
foreach lab_file (`cat $1.scf`)
    echo "$lab_file" >> $1
    more $lab_file >> $1
    echo '.' >> $1
end
```

El fichero MLF resultante debe ser similar al siguiente:

```
#!MLF!#
"/2006-03-01-1416-15SC.lab"
1622500 2489000 S1
3333000 4309000 S2
```

```

4824500 5657500 S2
.
"/2006-03-01-1426-15SC.lab"
1370500 2236500 S2
3497500 4155500 S3
4462500 5032500 S3
5493000 5921000 S2
.

```

Como se puede observar, al comparar los dos ficheros resultantes MLF la estructura es la misma solo que se obtienen a través de dos procedimientos distintos. Uno lleva incluido el ruido ya que el Transcriber debe etiquetar todo el registro. En cambio por medio del otro procedimiento se le puede excluir dejando exclusivamente los eventos a entrenar.

- b) Una vez obtenidos estos ficheros de transcripción MLF. Se ejecuta un script para obtener a partir de ellos la primera estadística de la cantidad obtenida de eventos en cada tipo, representada en la Tabla 1 anterior. El script es el siguiente:

Script No. 5

Este script utiliza o se vale para la identificación, de una lista creada con los tipos de eventos “eventos_def.txt”. Posteriormente llama al fichero de transcripción y mediante comando como “HLEd”, realiza una sustitución de los eventos malos. Un ciclo foreach (por cada uno), realiza el barrido de cada uno de los elementos del fichero de transcripción “mlf”, llamado “data_seg_def1.mlf”, anteriormente obtenido y separa los elementos, logrando al final un total de todos los registros encontrados.

refina_mlf.csh

```

#!/bin/csh

# Creación de la lista de eventos originales
HLEd -i /dev/null -n eventos.txt null.hed data_segml.mlf
echo "Eventos originales"
echo "====="
cat eventos.txt
echo "====="

# Sustitución de los MALOS
HLEd -i data_segml.mlf -n eventos_def.txt sustituye.hed data_segml.mlf
sed -e 's/^\/*\/*/' data_segml.mlf > data_seg_def1.mlf
echo ""
echo "Eventos restantes"
echo "====="
set allev = (`cat data_seg_def1.mlf`)
set eventos = (`cat eventos_def.txt`)
foreach e ( $eventos )
    set ne = `grep $e data_seg_def1.mlf | wc`
    set ne = $ne[1]
    printf "%-7s %d\n" $e $ne
end
echo "-----"
echo "Total " $#allev
echo "====="

```

El comando “HLEd”, utiliza un script en blanco llamado “null.hed”, para eliminar las transcripciones erradas de la lista de transcripciones “data_segml.mlf”. Además, crea los diccionarios de eventos. “eventos.txt” y “eventos_def.txt”. Por otra parte, realiza una sustitución mediante el script “sustituye.head”, de todos aquellos eventos dudosos de escritura para sustituirlos mediante la nomenclatura válida con la

que se trabaja. Así, envía un formato de salida del número de ocurrencias que corresponden a cada evento.

Script No. 6 y 7

null.head (vacío),

sustituye.hed (Si existe alguna duda de algún evento, aquí se coloca)

Por ejemplo el fichero utilizado en el Etna, respondió a la siguiente nomenclatura.

```
RE sil noise moise niose noise??
RE eq eq? eq?? eq-regional
RE NODEF ? ?? ??? Episode
RE lp lp3
RE hy hibr hibrido
RE NE noise Etna
RE TB tremor Etna
```

Al final, se obtiene la estadística de los eventos, que no es más que una relación de los eventos presentes en las segmentaciones similar a la siguiente:

```
Eventos restantes
=====
NT      1737
T4      200
T2      281
T1      234
T3      209
-----
Total   9753
=====
```

En donde se observa la cantidad de eventos existentes por cada tipo.

5.- Creación del modelo prototipo

La creación del modelo prototipo se realiza con el script llamado “**crea_prototipo.csh**”.

Script crea_prototipo.csh

```
#!/bin/csh

# Creación de modelos prototipo HMM
# Ligdamis A. Gutiérrez E.
# Abajo indica la nomenclatura del comando
# crea_prototipo.csh 6 (para 8) 39 1 > & proto.39.8 &

if ( $#argv < 3 ) then
    echo "Formato: crea_proto_hmm.csh <Nestados> <Nfeatures> <tipo>"
    exit
endif

set NE = $argv[1]
set NF = $argv[2]
set TY = $argv[3]
```

```

@ NEE = $NE + 2

# Cabecera
echo "<BeginHMM>"
echo "<NumStates> $NEE <VecSize> $NF <$TY> <nullD> <diagC>"
echo "<StreamInfo> 1 $NF"

# Definiciones para cada estado (excepto el 1ro y ultimo)
set ne = 2
set n = 1
while ( $n <= $NE )
  echo "<State> $ne <NumMixes> 1"
  echo "<Stream> 1"
  echo "<Mixture> 1 1.0"
  echo "<Mean> $NF"
  set j = 1
  while ( $j <= $NF )
    echo -n " 0.0"
    @ j = $j + 1
  end
  echo ""
  echo "<Variance> $NF"
  set j = 1
  while ( $j <= $NF )
    echo -n " 1.0"
    @ j = $j + 1
  end
  echo ""
  @ ne = $ne + 1
  @ n = $n + 1
end

# Matrix de Transicion

echo "<TransP> $NEE"

set ne = 1
echo -n " 0.0 1.0"
set ie = 3
while ( $ie <= $NEE )
  echo -n " 0.0"
  @ ie = $ie + 1
end
echo ""

set ne = 2
while ( $ne < $NEE )
  @ ne1 = $ne + 1
  set ie = 1
  while ( $ie <= $NEE )
    if ( $ie == $ne ) then
      echo -n " 0.6"
    else if ( $ie == $ne1 ) then
      echo -n " 0.4"
    else
      echo -n " 0.0"
    endif
    @ ie = $ie + 1
  end
  echo ""
  @ ne = $ne + 1
end

set ne = $NEE
set ie = 1
while ( $ie <= $NEE )
  echo -n " 0.0"

```

```

@ ie = $ie + 1
end
echo ""
echo "<EndHMM>"

```

Al comando de acción se le deben indicar tres parámetros; el primero es el número de estados, el segundo parámetro, representa el número de coeficientes a utilizar por el modelo, en este caso 39, y el último es el tipo, en este caso de tipo 1, seguidamente se da el nombre del fichero resultante de salida que será proto.39.8: Por lo tanto, la ejecución completa del comando es la siguiente:

crea_prototipo.csh 6 39 1 > & proto.39.8 &

El anterior script, realiza la creación del prototipo o modelo matemático inicial (11.39), que contiene 11 estados y con tamaño de vectores en 39 (coeficientes a utilizar) un fragmento de dicho modelo se muestra a continuación.

```

<BeginHMM>
<NumStates> 11 <VecSize> 39 <MFCC_E_D_A> <nullD> <diagC>
<StreamInfo> 1 39
<State> 2 <NumMixes> 1
  <Stream> 1
    <Mixture> 1 1.0
      <Mean> 39
        0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0
0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0
      <Variance> 39
    . . .
  <EndHMM>

```

6. Fichero de entrenamiento y reconocimiento

Una vez que se han obtenido todos los parámetros de configuración, la creación del prototipo y los ficheros necesarios, se ejecuta ahora el script denominado “trabajo2.csh”. Dicho script realiza la creación de los modelos de entrenamiento y la ejecución de los modelos de Markov, mediante el uso de las iteraciones con los estados y Gausianas configurados. El fichero se utiliza en las pruebas cerradas para el entrenamiento y el reconocimiento. Utiliza las diversas herramientas de HTK como HVite, HCompV, HRest, HERest y HResults. La línea de comandos de ejecución del Script es la siguiente:

trabajo2.csh > & resultados.log &

Con la ejecución de la línea anterior se creará un fichero (.log), en donde se almacena la información de salida, para revisarlo en caso de que suceda algún tipo de error en la ejecución. Por cada Gausiana, el script crea dos ficheros de resultados, uno conteniendo las transcripciones del reconocimiento por parte del sistema (rec.13.32.gram.mlf), y el otro las estadísticas (rec.13.32.gram.res). El contenido del script trabajo2.csh es el siguiente:

trabajo2.csh

```

#!/bin/csh
# Tesis Doctoral

```

```

# Ejemplo del fichero trabajo con parámetros para
# el volcán Telica de Nicaragua
# Ligdamis A. Gutiérrez Espinoza
# Segmentos de entrenamiento y reconocimiento
# test_train y test_test

# Declaración de directorio de datos
# Declaración de variable FEADIR para directorio de trabajo
set FEADIR = /home/Nicaragua/Telica/Telica/Datos

# Declaración del fichero SCP con las listas de registros
set SCP = /home/Nicaragua/Telica/Train/lista_r50.scp
set SCPT = /home/Nicaragua/Telica/Train/lista_r50.scp

# Declaración del número de iteraciones a utilizar
set NITER      = 3
#Declaración del número de gaussianas a utilizar
set NGAUS      = 16

# Declaraciones Generales
set HCOPLY_CONF = /home/Nicaragua/Telica/lib/hcopy.conf
set CONF        = /home/Nicaragua/Telica/lib/htrain.conf
set HMM_DIR     = ./HMM-fe
set EVFILE      = /home/Nicaragua/Telica/lib/eventos.txt
set NE          = 13 # Número de estados del modelo
set NUM_COEF    = 39 # Número de coeficientes
set PAR_TYPE    = MFCC_E_D_A
set PROTO       = /home/Nicaragua/Telica/lib/proto.$NUM_COEF.$NE
set BINDIR      = /home/Nicaragua/bin
set EPS         = 0.1
set MAXITER     = 35 # Número máximo de iteraciones a realizar
set MLF         = /home/Nicaragua/Telica/Train/data_segml.mlf
set GRAM        = gram
set WNET        = /home/Nicaragua/Telica/lib/gram.wnet
set DICT        = /home/Nicaragua/Telica/lib/eventos_defl.dic

#####
# Entrenamiento #
#####

entrena: # Entrenamiento del modelo

rm -rf ${HMM_DIR}
mkdir ${HMM_DIR}

# Modelos iniciales FLAT
mkdir $HMM_DIR/hmm.0.0
HCompV -C $CONF -o hmmdef -f 0.01 -m -S train.tmp -M $HMM_DIR/hmm.0.0 $PROTO
$BINDIR/macro $NUM_COEF $PAR_TYPE {$HMM_DIR}/hmm.0.0/vFloors
{$HMM_DIR}/hmm.0.0/macros

set eventos = `cat $EVFILE`
foreach ev ( $eventos )
  echo "~h " "'"$sev"'" >> $HMM_DIR/hmm.0.0/models
  gawk '(NR>4) {print $0 ;}' $HMM_DIR/hmm.0.0/hmmdef >>
  $HMM_DIR/hmm.0.0/models
  sed -e "s/hmmdef/$sev/" $HMM_DIR/hmm.0.0/hmmdef >>
  $HMM_DIR/hmm.0.0/$sev
end

echo "Fin de la creación de los prototipos"

```

```

# Se realiza el Entrenamiento a partir de las segmentaciones

mkdir $HMM_DIR/hmm.0.1
foreach ev ( $eventos )
  HRest -T 1 -C $CONF -e $EPS -i $MAXITER -I $MLF -S train.tmp \
    -H $HMM_DIR/hmm.0.0/macros -l $ev \
    -M $HMM_DIR/hmm.0.1 $HMM_DIR/hmm.0.0/$ev
end

# Creación del fichero global de modelos con 1 gaussiana
set hori = $HMM_DIR/hmm.0.0
set hdest = $HMM_DIR/hmm.1.0
mkdir $hdest
cp $hori/macros $hdest
HHEd -w $hori/models -M $hdest \
  -d $hori/home/Nicaragua/Telica/lib/null.hed ${EVFILE}

# Re-estimacion con 1 gaussiana

set ngaus = 1
set j = 1
while ( $j <= $NITER )
  @ i = $j - 1
  set hori = ${HMM_DIR}/hmm.${ngaus}.${i}
  set hdest = ${HMM_DIR}/hmm.${ngaus}.${j}
  mkdir $hdest
  echo "Iteración $j con $ngaus gaussianas"
  HERest -C $CONF -I $MLF -t 250.0 150.0 1000.0 -S train.tmp \
    -H $hori/macros -H $hori/models -M $hdest $EVFILE
  @ j = $j + 1
end

#####
# Decodificación (reconocimiento) #
#####

# Decodificación con 1 gaussiana

decol:

set HMM_REC = $HMM_DIR/hmm.${ngaus}.${NITER}
set out = rec.$NE.$ngaus.$GRAM
HVite -C $CONF -p -s 50.0 -t 250.0 -H $HMM_REC/macros -H
$HMM_REC/models -w $WNET \
  -S train.tmp -l '*' -i $out.mlf $DICT $EVFILE
HResults -e "???" sil -p -I $MLF $EVFILE $out.mlf > $out.res
cat $out.res

# Incremento de las gaussianas

mixup:

set ngaus = 2
while ( $ngaus <= $NGAUS )

  @ n = $ngaus - 1

  set hori = $HMM_DIR/hmm.${n}.${NITER}
  set hdest = $HMM_DIR/hmm.${ngaus}.0
  mkdir $hdest

```

```

HHed -T 2 -H $hori/macros \
      -H $hori/models \
      -M $hdest \
      /home/Nicaragua/Telica/lib/mixup.hed $EVFILE

set j = 1
while ( $j <= $NITER )
  @ i = $j - 1
  set hori = ${HMM_DIR}/hmm.${ngaus}.${i}
  set hdest = ${HMM_DIR}/hmm.${ngaus}.${j}
  mkdir $hdest
  echo "Iteración $j con $ngaus gaussianas"
  HERest -C $CONF -I $MLF -t 250.0 150.0 1000.0 -S train.tmp \
        -H $hori/macros -H $hori/models -M $hdest $EVFILE
  @ j = $j + 1
end

# Decodificación

set HMM_REC = $HMM_DIR/hmm.${ngaus}.${NITER}
set out = rec.$NE.$ngaus.$GRAM
HVite -C $CONF -p -s 50.0 -t 250.0 -H $HMM_REC/macros -H
$HMM_REC/models -w $WNET \
      -S train.tmp -l '*' -i $out.mlf $DICT $EVFILE
HResults -e "???" sil -p -I $MLF $EVFILE $out.mlf > $out.res
cat $out.res

@ ngaus = $ngaus + 1
end

```

El fichero “resultados.log”, que contiene el registro de los mensajes de la prueba. Un fragmento presentando el contenido de dicho fichero se muestra a continuación.

resultados.log

```

Fin de la creación de los prototipos
WARNING [-7032] OWarn: change HMM Set covKind in HRest
Reestimating HMM ./HMM-fe/hmm.0.0/NS . . .
States : 2 3 4 5 6 7 8 9 10 11 12 (width)
Mixes s1: 1 1 1 1 1 1 1 1 1 1 1 ( 39 )
Num Using: 0 0 0 0 0 0 0 0 0 0 0
Parm Kind: MFCC_E_D_A
Number of owners = 1
SegLab : NS
MaxIter : 35
Epsilon : 0.100000
Updating : Transitions Means Variances

- system is PLAIN
258 Examples loaded, Max length = 143, Min length = 11
Ave LogProb at iter 1 = -6298.54102 using 258 examples
Ave LogProb at iter 2 = -6228.62646 using 258 examples change =
69.91456
. . .
. . .
Ave LogProb at iter 22 = -5917.01562 using 258 examples change =
0.09485
Estimation converged at iteration 22

```

7.- Reconocimiento

Para el reconocimiento (pruebas abiertas, ciego-ciego) se utiliza un script que se denomina “decode.csh”. El fichero utiliza los modelos previamente entrenados HMMs con el anterior fichero, utilizando para ello los modelos almacenados en los “mlf” iniciales y presenta como resultados una nueva lista de ficheros de transcripción “mlf”.

Archivo decode2.csh

```
#Modificación fichero decode para test ciego (pruebas abiertas)
#Proyecto Etna

#Fichero decode2.csh Ligdamis para el volcán Telica
#!/bin/csh
# Modificación del fichero decode con los datos del Etna
# Fichero decode2.csh

set NITER      = 6      # Numero de iteraciones a utilizar
set NGAUS      = 32     # Numero de gaussianas a utilizar
set ngaus      = 32
set HCOPY_CONF = hcopy.conf
set CONF       = htrain.conf
set HMM_DIR    = HMM-fe
set EVFILE     = eventos_defl.txt
set NE         = 11     # Numero de estados de los modelos
set NUM_COEF   = 39     # Numero de coeficientes
set PAR_TYPE   = MFCC_E_D_A
#set PROTO     = proto.$NE
set PROTO      = proto.$NE
set BINDIR     = bin
set EPS        = 0.1
set MAXITER    = 35
set MLF        = data_segm_defl.mlf
set GRAM       = gram
set WNET       = gram.wnet
set DICT       = eventos_defl.dic

#####
# Codificación #
#####

codifica:

HCOPY -C ${HCOPY_CONF} -S listal.scp

# Decodificación llama a los ficheros listal con extensión r.50
# y test1.tmp que contiene la extensión mfc

set HMM_REC = $HMM_DIR/hmm.${ngaus}.${NITER}
set out     = rec.$NE.$ngaus.$GRAM
HVite -C $CONF -H $HMM_REC/macros -H $HMM_REC/models -w $WNET \
      -S test2.tmp -l '*' -i $out.mlf $DICT $EVFILE
HResults -e "???" sil -p -I $MLF $EVFILE $out.mlf > $out.res
cat $out.res
```

APÉNDICE I: Formulario

En este apéndice se describen las principales ecuaciones matemáticas utilizadas en la solución de los tres problemas básicos de los HMMs. De acuerdo con la nomenclatura expuesta en [Young et. al. (2002)].

1. Solución a los tres problemas básicos de un HMM

a) Solución al problema de Evaluación. **Algoritmo de adelante-atrás (forward-backward)**

Este algoritmo define dos variables $\alpha_t(i)$, que se ocupa hacia adelante, y $\beta_t(i)$, que se ocupa hacia atrás, determinados por las ecuaciones 1 y 2

$$\alpha_t(i) = P(O_1, O_2, \dots, O_t, q_t = i | \lambda) \quad (1)$$

Donde la probabilidad de la secuencia parcial de observación O_1, O_2, \dots, O_t , hasta el tiempo t , dado un estado i , en un tiempo t , y el modelo λ

$$\beta_t(i) = P(O_{t+1}, O_{t+2}, \dots, O_T | q_t = i, \lambda) \quad (2)$$

Donde la probabilidad de la secuencia parcial de observación desde $t + 1$, hasta el final, dado un estado i , en un tiempo t , y el modelo λ

Para encontrar estas dos probabilidades parciales se utiliza un bucle inductivo. El bucle para resolver $\alpha_t(i)$ es el siguiente:

i) Inicialización

$$\alpha_1(i) = \pi_i b_i(O_1), \quad 1 \leq i \leq N \quad (3)$$

ii) Inducción

$$\alpha_{t+1}(j) = \left[\sum_{i=1}^N \alpha_t(i) a_{ij} \right] b_j(O_{t+1}) \quad 1 \leq t \leq T - 1, 1 \leq j \leq N \quad (4)$$

iii) Término

$$P(O|\lambda) = \sum_{i=1}^N \alpha_T(i) \quad (5)$$

De manera similar, el bucle para resolver $\beta_t(i)$, es el siguiente:

i) Inicialización

$$\beta_t(i) = 1, \quad 1 \leq i \leq N \quad (6)$$

ii) Inducción

$$\beta_t(i) = \sum_{j=1}^N \alpha_{ij} b_j(O_{t+1}) \beta_{t+1}(j) \quad (7)$$

iii) Término

$$t = T - 1, T - 2, \dots, 1 \quad 1 \leq i \leq N \quad (8)$$

b) Solución al problema de decodificación: **Algoritmo de Viterbi**

La solución al problema de decodificación es buscar un algoritmo para encontrar un modelo λ que produzca el máximo valor de $P(O|\lambda_i)$, y de ahí que pueda ser utilizado para el reconocimiento.

En la práctica, sin embargo, es preferible basar el reconocimiento, en la secuencia de estados de máxima probabilidad ya que esto se generaliza fácilmente para el caso de la señal continua, mientras que, el uso de la probabilidad total no se efectúa. Esta probabilidad es computada como el cálculo de la probabilidad hacia adelante, utilizando esencialmente el mismo algoritmo exceptuando que la sumatoria es reemplazada por la máxima operación. Para un modelo dado λ , $\phi_j(t)$ representa la máxima probabilidad de observar un vector de señal o_1 a o_t en un estado j en un tiempo t . Esta probabilidad parcial puede computarse eficientemente utilizando la recursión dada en la ecuación 9

$$\phi_j(t) = \max_i \{ \phi_i(t-1) a_{ij} \} b_j(o_t) \quad (9)$$

En donde,

$$\phi_1(1) = 1 \quad (10)$$

$$\phi_j(1) = a_{1j} b_j(o_1) \quad (11)$$

Para $1 < j < N$. La máxima probabilidad $\hat{P}(O|\lambda)$ está dada por

$$\phi_N(T) = \max_i \{ \phi_i(T) a_{iN} \} \quad (12)$$

Como el caso de la re-estimación, la computación directa de las múltiples probabilidades, se realiza utilizando probabilidades logarítmicas. Así, la fórmula de recursión 12 queda de la siguiente forma:

$$\psi_j(t) = \max_i \{\psi_i(t-1) + \log(a_{ij})\} + \log(b_j(o_t)) \quad (13)$$

La anterior ecuación de recursión es el llamado *Algoritmo de Viterbi*

c) Solución al problema de aprendizaje: **Algoritmo de Baum-Welch**

El problema esencial es estimar los recursos y variaciones de un HMM, en el que la salida de distribución de cada estado es un simple componente Gaussiano y viene dada por la ecuación 14.

$$b_j(O_t) = \frac{1}{\sqrt{(2\pi)^n |\Sigma_j|}} e^{-\frac{1}{2}(o_t - \mu_j)' \Sigma_j^{-1} (o_t - \mu_j)} \quad (14)$$

Si hubiera un simple estado j en el HMM, la estimación del parámetro sería relativamente fácil. El estimado máximo posible de μ_j y Σ_j se reduciría a los promedios simples dados por las ecuaciones 15 y 16.

$$\bar{\mu}_j = \frac{1}{T} \sum_{t=1}^T o_t \quad (15)$$

y

$$\bar{\Sigma}_j = \frac{1}{T} \sum_{t=1}^T (o_t - \mu_j)(o_t - \mu_j)' \quad (16)$$

En la práctica, hay múltiples estados y no hay ninguna asignación directa de la observación de vectores a los estados individuales, porque la sucesión de estados subyacente es desconocida. Sin embargo, alguna aproximación de la asignación de vectores a los estados se puede realizar con las ecuaciones 2.22 y 2.23, para dar los valores iniciales requeridos por los parámetros.

Ya que la probabilidad total de cada sucesión de la observación se basa en la sumatoria de todas las posibles secuencias de estados, cada vector de observación o_t , contribuye a la computación de la máxima probabilidad de los valores paramétricos de cada estado j . En otros términos, en lugar de asignar cada vector de la observación a un estado específico (como en la aproximación previa), cada observación es asignada a cada estado en proporción a la

probabilidad de que el modelo esté en ese estado cuando el vector fue observado.

Así, la probabilidad en el inicio de un estado j en un tiempo t , convierte las ecuaciones 2.22 y 2.23 a los valores representados en las ecuaciones 17 y 18

$$\bar{\mu}_j = \frac{\sum_{t=1}^T L_j(t) o_t}{\sum_{t=1}^T L_j(t)} \quad (17)$$

y

$$\bar{\Sigma}_j = \frac{\sum_{t=1}^T L_j(t) (o_t - \mu_j)(o_t - \mu_j)'}{\sum_{t=1}^T L_j(t)} \quad (18)$$

En donde, las sumatorias en los denominadores son incluidas para obtener la normalización requerida. Las ecuaciones 17 y 18, son las formulas de **re-estimación Baum-Welch** para las medias y covarianzas de un HMM.

Claro está, que para aplicar las ecuaciones 17 y 18 se debe calcular la probabilidad de ocupación de un estado $L_j(t)$. Este resultado se obtiene de forma eficiente, aplicando el llamado algoritmo “Forward-Backward” (algoritmo de delante-atrás), antes mencionado. El cual permite que la probabilidad delantera $\alpha_j(t)$ para un modelo λ con N estados.

Hasta este punto, toda la información que necesita un *HMM* para ejecutar un parámetro de re-estimación utilizando el algoritmo *Baum-Welch* ha sido expuesta. Los pasos de ejecución de este algoritmo pueden definirse de la siguiente forma:

- a. Para cada vector/matriz parametrizado que requiere una re-estimación se asigna espacio de almacenamiento para el numerador y la sumatoria del denominador de acuerdo a las ecuaciones expresadas en las fórmulas 17 y 18
- b. Se calcula las probabilidades delante y trasera (forward and backward) para todos los estados j y los tiempos t .
- c. Para cada estado j y tiempo t , se utiliza la probabilidad $L_j(t)$ y el correspondiente vector de observación o_t para actualizar los acumuladores para ese estado.
- d. Se utiliza los valores acumulados finales para calcular nuevos valores de parámetros.
- e. Si el valor de $P = P(O | \lambda)$ para esta iteración no es más alto que el valor de la iteración previa entonces se detiene, en otro caso se repiten los pasos anteriores utilizando nuevos valores para los parámetros de re-estimación.

Un punto final a mencionar es que la computación de la probabilidad del algoritmo delantero-trasero, involucra la toma de productos de un gran número de

probabilidades. Esto significa, que en la práctica los números reales involucrados sean muy pequeños. Aquí, para evitar problemas numéricos, la computación del algoritmo delantero-trasero en *HTK* se realiza utilizando aritmética logarítmica.

Bien, se ha terminado « Todo largo camino, comienza con el primer paso, pero que reconfortante es haber dado el último y llegar al fin a la meta ».

Solo no lo hubiera logrado, Gracias a todos aquellos por su apoyo y oraciones, además de haber creído en mí. Lo he conseguido.

Ligdamis A. Gutiérrez E.