We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900
Open access books available

186,000
International authors and editors

200M
Downloads

154
Countries delivered to

Our authors are among the

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

CLARIVATE ANALYTICS
BOOK CITATION INDEX
INDEXED

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Volcano-Seismic Signal Detection and Classification Processing Using Hidden Markov Models - Application to San Cristóbal and Telica Volcanoes, Nicaragua

Gutiérrez, Ligdamis[1], Ramírez, Javier[1],
Ibañez, Jesús[2] and Benítez, Carmen[1]
[1]Dep. Signal Theory Networking and Communications University of Granada, Granada
[2]Instituto Andaluz de Geofísica. Campus de Cartuja s/n. University of Granada, Granada
Spain

1. Introduction

The identification and classification of seismic signals is one of primary work that a volcano observatory must do, and this work should be done in a period of time as close as possible to the real time". Each seismic event is related to a different source process, and its time and spatial distribution could be used as elements of an early warning system of volcanic eruption (see for example Chouet el el 1996 [1, 2], or Chouet 2003 [3]).

Moreover, the recognition system is based on the HMM theory, published in the late 60s by Baum et al. (1966) [4] and Baum et al. (1970) [5]. Nowadays, the Hidden Markov Models technique is the more effective one to implement voice recognition systems. Over the past years, Hidden Markov Models have been widely applied in several models like pattern [6, 7], pathologies [8] or speech recognition [9, 10], and DNA sequence analysis [11, 12]. On the other hand, previous works [13, 14, 15, 16, 16a, 16b] have probed the parallelism among speech and volcano-seismic events in terms of signal complexity and real time requirements. At the present many observatories perform this work observing on the screen or in paper the seismograms, and a technician decides the type of event under their appearance in the time domain and their experience. This work is in many cases difficulty if the rate of occurrence of seismic events per hour is high, or if weather or local conditions increase the level of seismic noise. A more detailed analysis, including for example spectral characteristics, would be too time-consuming to be carried out in real time. Furthermore, in a crisis situation, there is a need to make fast decisions that can affect the public safety. This is the reason because many researchers are focussing their efforts in the development of a robust automatic discrimination algorithm of seismic events, enabling technicians to focus their efforts in the interpretation of the situation or to analyze only a reduced number of signals. Recently Del Pezzo et al.[17] and Scarpeta et al.[18] have presented the application of neuronal networks for discrimination and classification of volcanic and artificial signals at Vesuvius Volcano and Phlegraean Fields (Italy). These methods have been successfully applied to discriminate signals for local and volcanic seismicity. However, a limitation of

these systems is that they require a manual pre-processing of the data to select a subset of recordings, each one containing just one seismic event to be classified. On the other hand, Ornhberger [19] has studied discrete hidden Markov modelling (HMM) tools for continuous seismic event classification. Just as neural networks, these methods have been successfully applied to discriminate signals for local and volcanic seismicity without the limitations of the first. Gutiérrez et al. [16], Ibañez et al [16a], Cortes et al.[16b] and Benítez et al. [13] improved the continuous HMMbased seismic event classification and monitoring system and applied the algorithm to the detection and classification of four kind of volcanic events recorded at Etna, Stromboli (Italy), Popocatepetl, Colima (México) and Deception Island (Antarctica). This system is based on the state of the art of HMM-based pattern recognition techniques, successfully applied to other disciplines such as robust automatic speech recognition (ASR) systems.

In this sense, we recordings of different seismic event types are studied at two active volcanoes; San Cristóbal and Telica in Nicaragua. The objective of the present work is to observe the validity of the method using data from active volcanoes, and to check if it is possible to exchange different databases to recognize signals belonging to different volcanoes. We use data from one single field survey carried out in February to March 2006. In this work we initially proceeded to identify the signals and to segment them to obtain a model for each class of events. Then we applied separately the model for each volcano data set and finally we mixed both data set to have a joint test of capability and accuracy of the system.

2. Regional setting

2.1 Nicaragua volcanic chain

In Nicaragua exists more than 200 volcanic structures. Some experts refer at more than 600. Counting their multiple small crater lakes (gaps) of volcanic origin, domes, maar, etc. The Nicaragua volcanic chain has produced all the mechanisms and volcanic products known in the world. The volcanic activity combined with tectonic processes taking place above the subduction zone along the Pacific Coast of Central America, formed the Volcanic Chain of Nicaragua. This chain is a part of the Central American Volcanic Front, which includes 18 volcanic centres in Nicaragua, of which eight are active. The chain is underlain by Middle Miocene – Late Pliocene volcanics and volcanogenic sediments (the widest spread Coyol and/ or Tamarindo Formations).

Young volcanoes are clustered in several morphologically well defined complexes, in some cases, separated one from another by NE trending, siniestral strike-slip faults (van Wyk de Vries, 1993 [20]), Carr and Stoiber, 1977 [21]). In Nicaragua exist ten main individual volcanic complexes, as listed below

a.   Cosigüina shield volcano
b.   San Cristóbal - Casita complex
c.   Telica complex
d.   El Hoyo - Cerro Negro complex
e.   Momotombo - Malpaisillo - La Paz Centro complex
f.   Apoyeque silicic shield volcano
g.   Miraflores - Nejapa volcano - tectonic zone
h.   Masaya volcanic complex
i.   Apoyo / Mombacho / Zapatera volcanic complex

j.   Nandaime - Granada volcano - tectonic Aligneament
k.   Ometepe island volcanoes

The respective volcanic complexes are situated in the Nicaraguan Depression, which is interpreted as an asymmetric graben. The two volcanoes of this work, San Cristóbal and Telica are the major active volcanoes in the complexes of the same name (Fig 2.1).
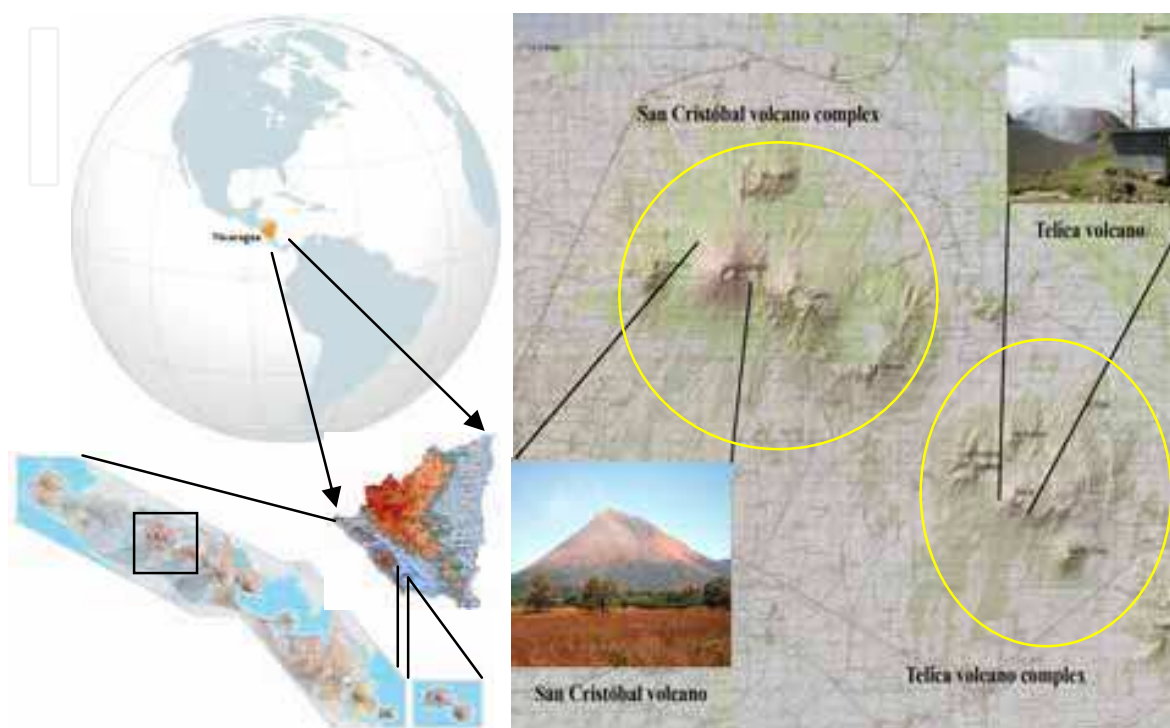

Fig. 2.1 San Cristóbal and Telica volcano complex in Nicaragua volcanic chain

The San Cristóbal volcanic complex, 100 kilometers northwest of Managua, consists of five principal volcanic edifices. The symmetrical youngest cone, San Cristóbal (also known as El Viejo) is the highest peak of the Maribios Range, and is capped by a 500 x 600 meter wide crater. The San Cristóbal is a Quaternary stratovolcano. Historical eruptions from San Cristóbal, consisting of small-to-moderate explosive activity, have been reported since the 16th century.

Telica is a Quaternary stratovolcano located in the western part of Nicaragua. It has six cones, the tallest of which is 1061 meters high. The Telica volcano, *(1,061 meters)* one of Nicaragua's most active volcanoes. Telica has erupted frequently since the Spanish Era. The Telica volcano group consists of several interlocking cones and vents with a general northwest alignment. Sixteenth-century eruptions have been reported at symmetrical Santa Clara volcano, at the southwest end of the Telica group.

3. Data instruments and sities

More than 600 hours of data in each volcano were analyzed and 431 seismic events in San Cristóbal and 1224 in Telica were registered at short period stations. The first step in the data analysis, after the revision of the files, is the visual inspection and segmentation of the signals. Each file is 300-s-long and could have different types of volcanic signals. Fig 3.1 and 3.2 shows the Histogram of the duration (in seconds) for each one of the seismic events. The
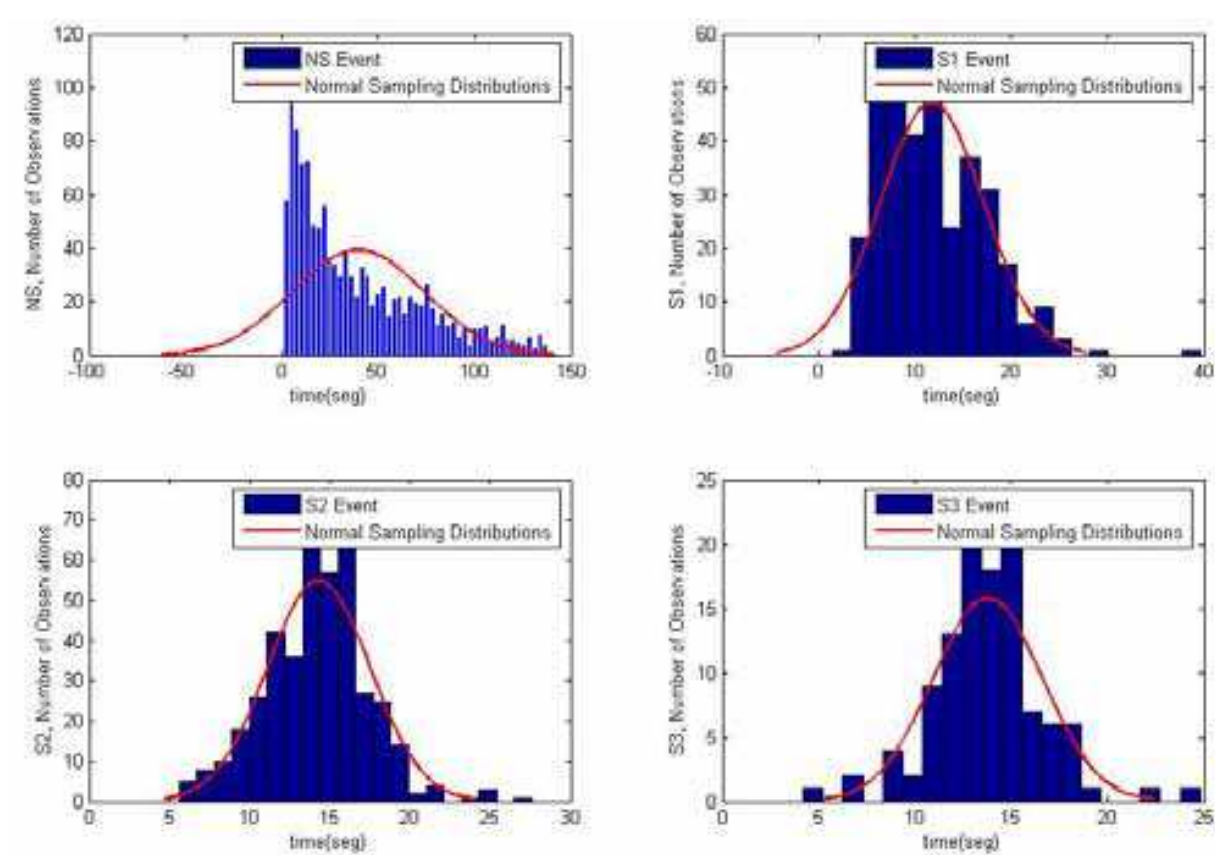
Fig. 3.1 Histogram of the duration (in seconds) for each one of the seismic events in San Cristóbal volcano.

durations of the event range between 10 and 40 seconds and tremor background noise 20-120 seconds.

For our study we proceeded to distinguish three types of signals events in the San Cristóbal volcano; San Cristóbal Long Period signal, Strombolian and San Cristóbal explosions, volcanic tremor and background seismic noise. We have defined these types as S1, S2 and S3. Additionally, we have defined the tremor background seismic noise as NS. Fig 3.3 shows and example of San Cristóbal explosion signal (S2) and Fig 3.4 shows and example of background seismic noise (NS) in which is presented a harmonic tremor.

For Telica volcano we define four types of seismic signals; Strombolian explosions, Telica explosions, Telica Long Period signal, volcanic tremor and background seismic noise. We have defined these types as T1, T2, T3 and T4. Additionally, we have defined the tremor background seismic noise as NT. Fig 3.5 shows and example of Telica signal (T1) and Fig 3.6 shows and example of Telica signal (T4).

The characteristics of the San Cristóbal explosion quakes have peaks from 3 to 8 Hz in the S1 event, 5 Hz in the S2 event to 6 Hz in the S3 event. The first 3 seconds of the signal are dominated by frequencies between 0.5 and 2 Hz, while the successive phases show a broader frequency content (see "zoom frequency spectrogram (d)" in fig. 3.3). Fig 3.4 shows and example of harmonic tremor in San Cristóbal, the signal are dominated by frequencies between 2.5 and 3.5 Hz. In the figure three bands of energy can be observed, the highest peak is about 2.8 Hz. For Telica volcano the durations of the event range between 10 and 60 seconds, and for the background noise between 20 and 140 s. The rest of the volcanic signals were not considered. In fact, In order to have an adequate model of every class of volcanic
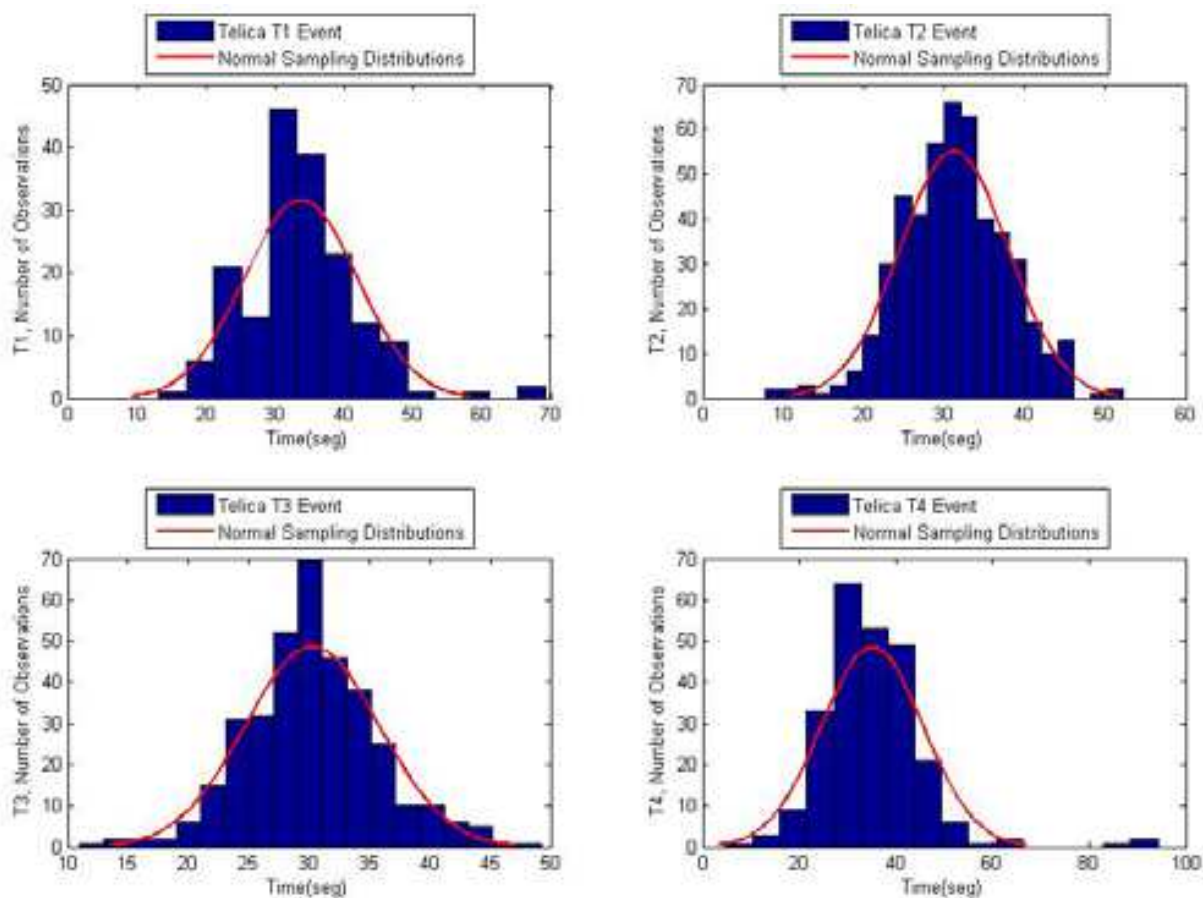
Fig. 3.2 Histogram of the duration (in seconds) for each one of the seismic events in Telica volcano

signals we need a number of events as large as possible, and only the above cited three classes in San Cristóbal and four classes in Telica could provide a number of events large enough to build a well-defined model.

The segmentation process consists in setting marks on the seismograms to define the data segments that contain a particular seismic event. Then, a label is added to declare the type of event that occurred at each segment. This procedure was done manually by a single expert, who set the beginnings and endings of the seismic events by waveform inspection and decided which class they belong to. In this way we ensure the use of the same criteria in the selection process.

For San Cristóbal volcano we segmented 224 samples of S1, 330 samples of S2, 544 samples of S3 and 1,746 samples of background noise (NS) events. For Telica volcano we segmented 360 samples of T1, 502 samples of T2, 429 samples of T3, 298 samples of T4 and 2,855 samples of background noise (NT) events.

4. Method

4.1 Introduction

We initially proceeded to identify the signals visually, and to segment the data to obtain a model for each event class. Once the recordings were manually segmented and labelled, the recognition system was carefully trained using the available Baum-Welch reestimation
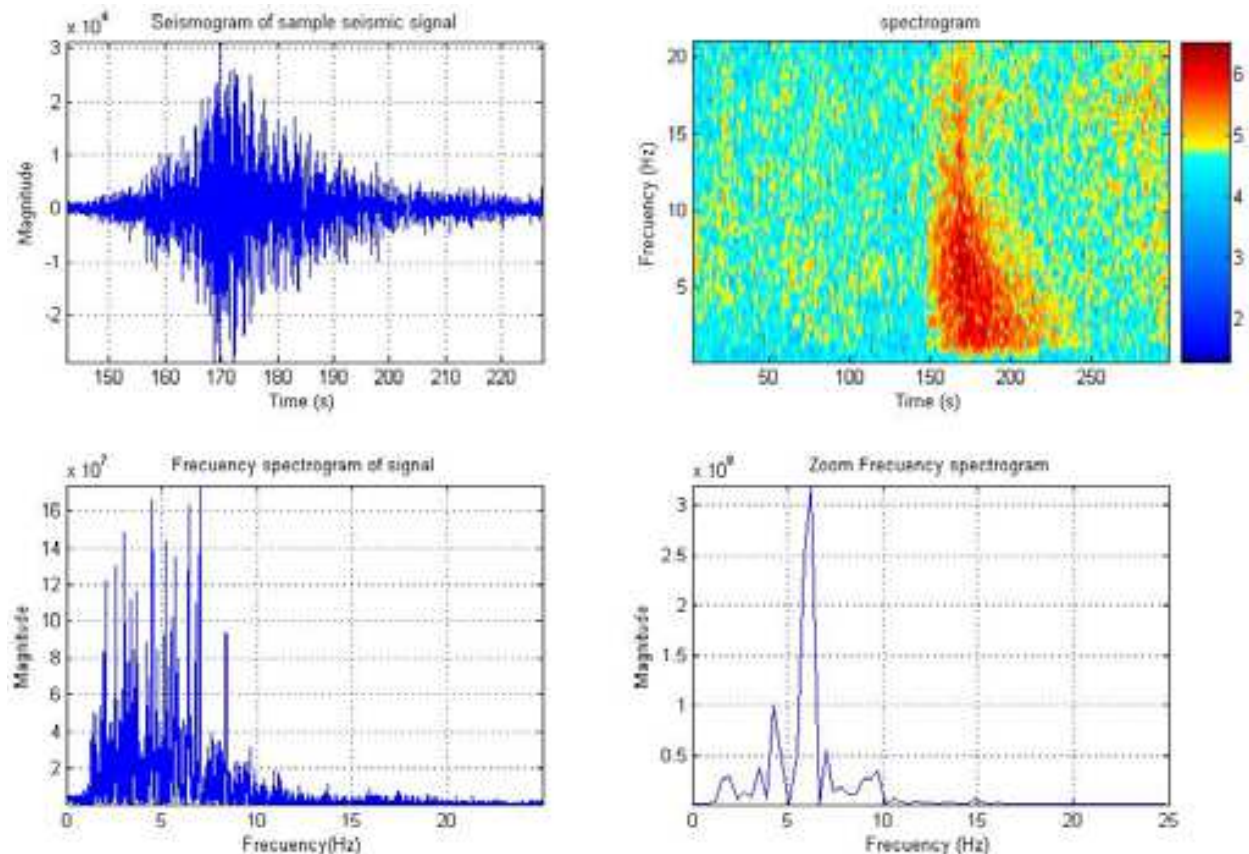
Fig. 3.3 **S2** seismic signal analysis at San Cristóbal volcano.

algorithms [22] using the Hidden Markov Model Toolkit (HTK) software [23]. Fig 4.1 shows a four-state left-to-right HMM non-emitting entry and exit states. The models were defined as left-to-right HMMs where only transitions from state i to state $i+1$ are allowed. The optimal length of the HMMs is mainly determined by the mean duration of the explosions and tremor events being modelled. Several experiments were carried out in order to determine the best choice. On the other, the HMM emission probabilities at each state were previously defined as Gaussian mixture densities and, during the training process, HMM models with increasing number of Gaussians were built in order to determine the best trade-off between recognition accuracy and performance. For our work we defined the models with 13-state HMMs with 16 Gaussians. Fig 4.2 shows the architecture of a general purpose HMM-based pattern recognition system. The training database and transcriptions are used to build the models. We applied these models separately for the volcano data set and finally mixed both data sets as a test of the portability of the system. The method analyzes the seismograms comparing the characteristics of the data to a number of event classes defined beforehand. If a signal is present, the method detects its occurrence and produces a classification. The recognition and classification system based on HMM is a powerful, effective, and successful tool [24]. 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. Also, each type of signals requires its own characterization. That is, each signal type must be represented by its own specific model, which would include the effects of source, path and sites.
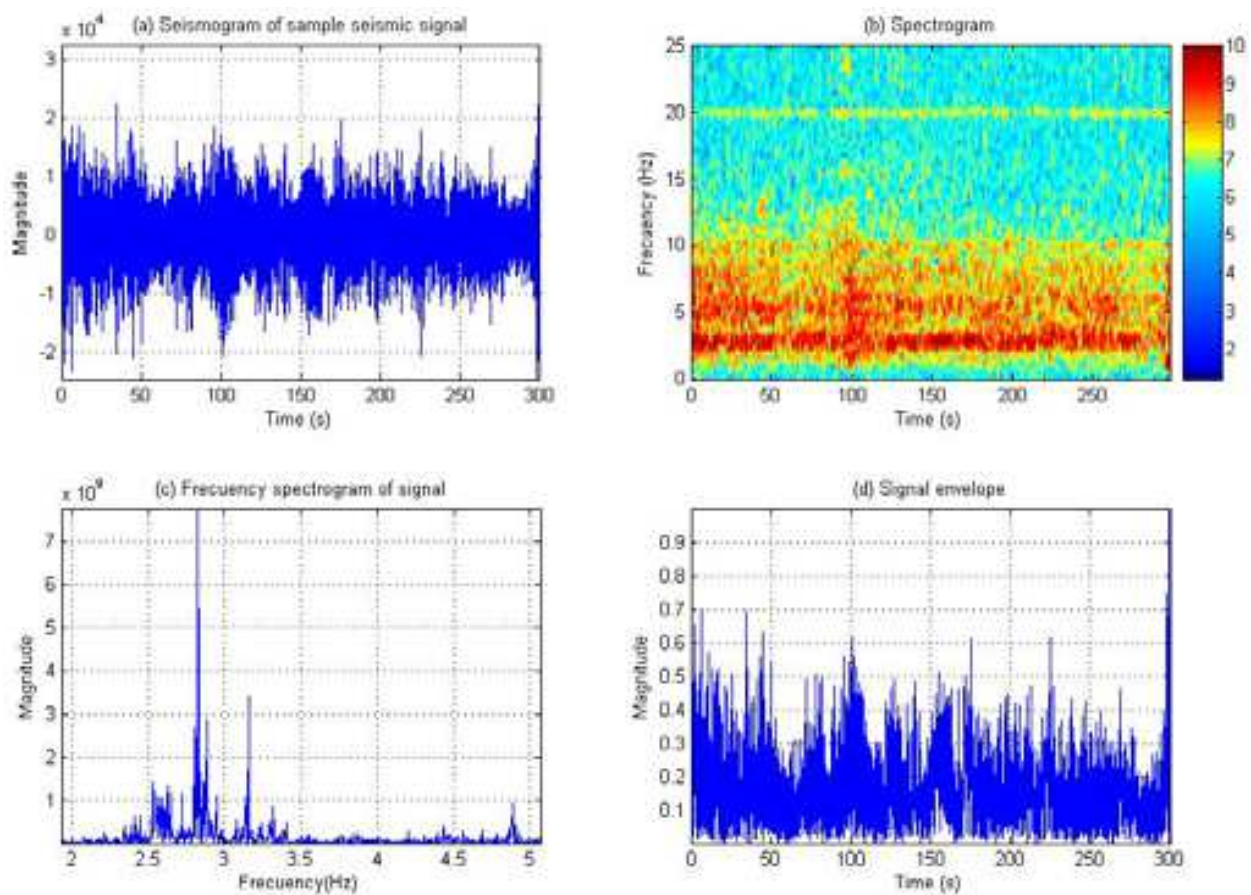
Fig. 3.4 Harmonic tremor in a **NS** background seismic noise in San Cristóbal volcano.

4.2 Description of the recognition system

An HMM-based seismic event recognition system [13,16,16a,16b] assumes that the signal is a realization of a sequence of one or more symbols. In order to perform the recognition process the recognizer decomposes the incoming signal as a sequence of feature vectors. This sequence is assumed to be a precise representation of the seismic signal while the length of the analysis window is such that the seismic waveform can be considered as stationary. The goal of a seismic event recognition system is to perform a mapping between a sequence of feature vectors and the corresponding sequence of seismic events.

4.3 Seismic event recognition

Let a sequence of seismic events $\mathbf{w} = \{w_1, w_2, ..., w_l\}$ be represented as a sequence of feature vectors $\mathbf{o}_t$ or observations $\mathbf{O}$, defined as

$$\mathbf{O} = \mathbf{o}_1, \mathbf{o}_2, ..., \mathbf{o}_t, ..., \mathbf{o}_T \tag{1}$$

where $\mathbf{o}_t$ is the feature vector observed at time t. The solution to the problem of continuous event recognition is to select the sequence of events $\mathbf{w}$ with the maximum probability $P(\mathbf{w} \mid \mathbf{O})$, that is:

$$\underset{\mathbf{w}}{\arg\max} \quad P(\mathbf{w} \mid \mathbf{O}) \tag{2}$$
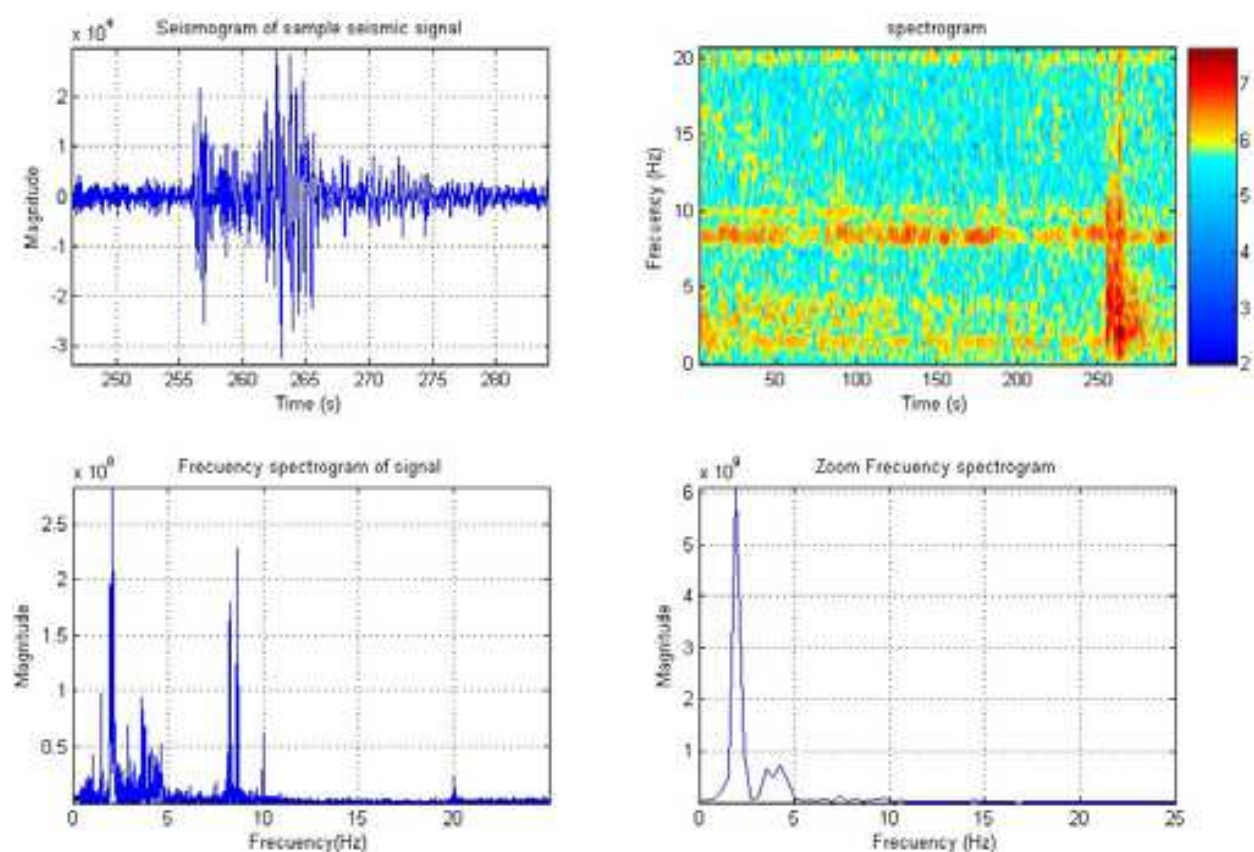
Fig. 3.5 **T1** seismic signal analysis at Telica volcano

4.4 HMM-based seismic event classification

HMM-based pattern recognition systems normally assume that the sequence of observed feature vectors corresponding to each event is generated by a Markov model. A Markov model is essentially a finite state machine with several states. Figure 4.1 shows a four-state left-to-right HMM with non-emitting entry and exit states. A change of state takes place every time unit and a feature vector $\mathbf{o}_t$ is generated from a probability density $b_j(\mathbf{o}_t)$ determined during the training process. Moreover, transition from state i to state j is governed by the transition probabilities a_{ij} which are used to model the delay in each of the states and the transitions through the entire model.

Figure 4.2 shows the architecture of a general purpose HMM-based pattern recognition system. The training database and transcriptions are used to build the models. Once the models are initiated, the recognition system performs feature extraction and decoding based on the Viterbi algorithm (Fig. 4.3). The output is the sequence of recognized events, confidence measures and global accuracy scores.

4.5 Signal processing and feature extraction

The first step of the recognition process is the signal processing feature extraction which converts the volcano seismic waveform in a parametric representation, with less redundant information, for further analysis and processing. As the short-time spectral envelop representation of the signal has been widely used, with good results, in speech recognition systems (Rabiner and Juang,1993) [25], a similar representation for our volcano seismic recognition system is used in this work. The feature extraction process is described as
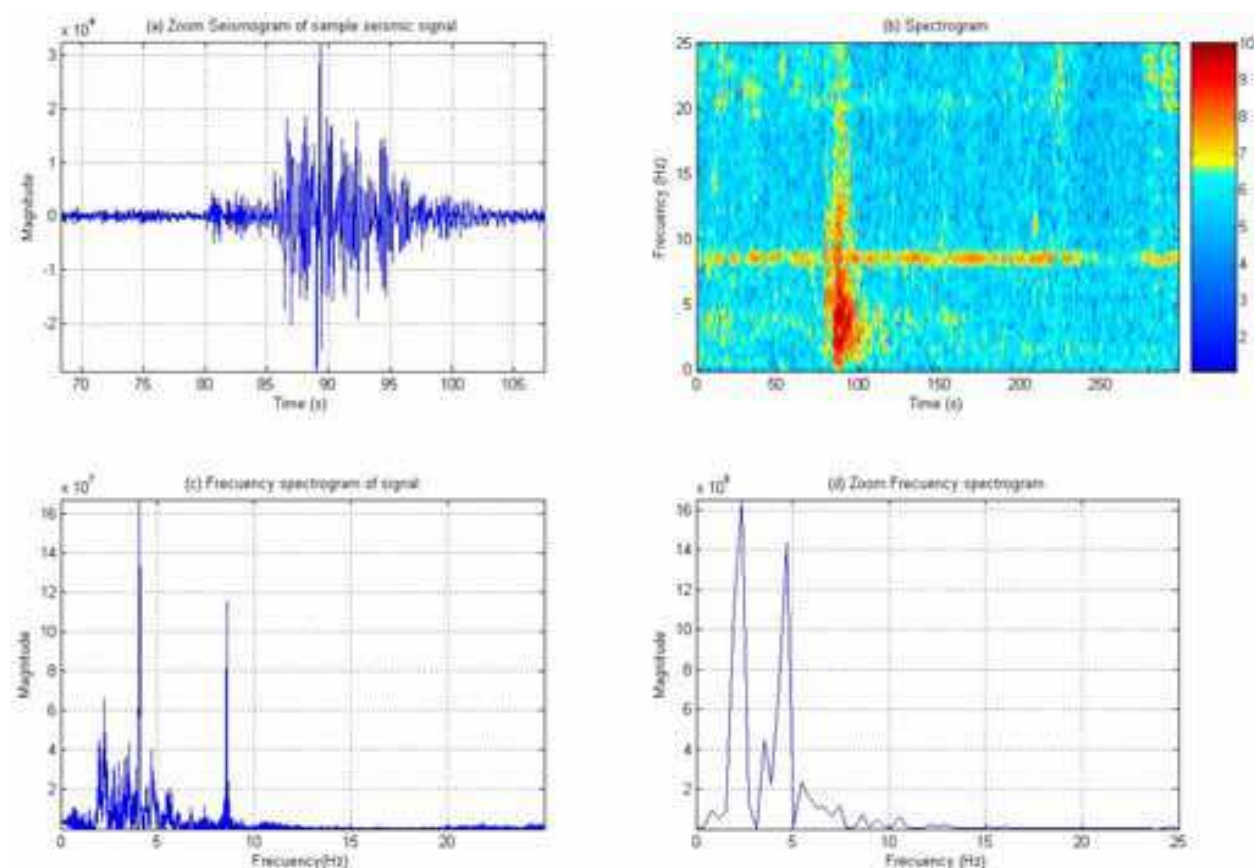
Fig. 3.6 **T4** seismic signal analysis at Telica volcano

follows. The signal is arranged into 4 seconds overlapping frames with a 0.5 seconds frame shit using a Hamming window. A 512-point FFT is used to compute the magnitude spectrum which serves as the input of an emulated filter-bank consisting of 16 triangular weighting functions uniformly spaced between 0 Hz and 20 Hz. The overlap between adjacent filters is 50%. The purpose of the filter bank analyzer is to give a measurement of the energy of the signal in a given frequency band. Then, the natural logarithm of the output filter-bank energies is calculated resulting a 16-parameter feature vector. Since the log-filter bank energies are highly correlated and the recognition system uses continuous observation HMMs with diagonal covariance matrices, it is necessary to apply a decorrelation transformation. Thus, the Discrete Cosine Transform (DCT) is used to decorrelate the features and reduce the number of components of the feature vector from 16 to 13 coefficients. Finally, the feature vector is augmented with linear regressions of the features (derivatives and accelerations) obtaining a total of 39 parameters.

4.6 Training

The recognition system [13,16,16a,16b] is based on continuous hidden Markov models (CHMM). CHMM are trained for each seismic event to be recognized and a noise model is used to represent sequences with no events. Both, training and recognition processes are performed using HMM Tool Kit (HTK) software (Young et al., 1997 [26]). In a CHMM the emission probabilities for a feature vector $\mathbf{o}_t$ in state $x(t)$, $b_{x(t)}(\mathbf{o}_t)$ are given by:

$$b_{x(t)}(\mathbf{o}_t) = \prod_{s=1}^{S} \sum_{k=1}^{K} c_{ik} N(\mu_k, \sigma_k, \mathbf{o}_t) \tag{3}$$
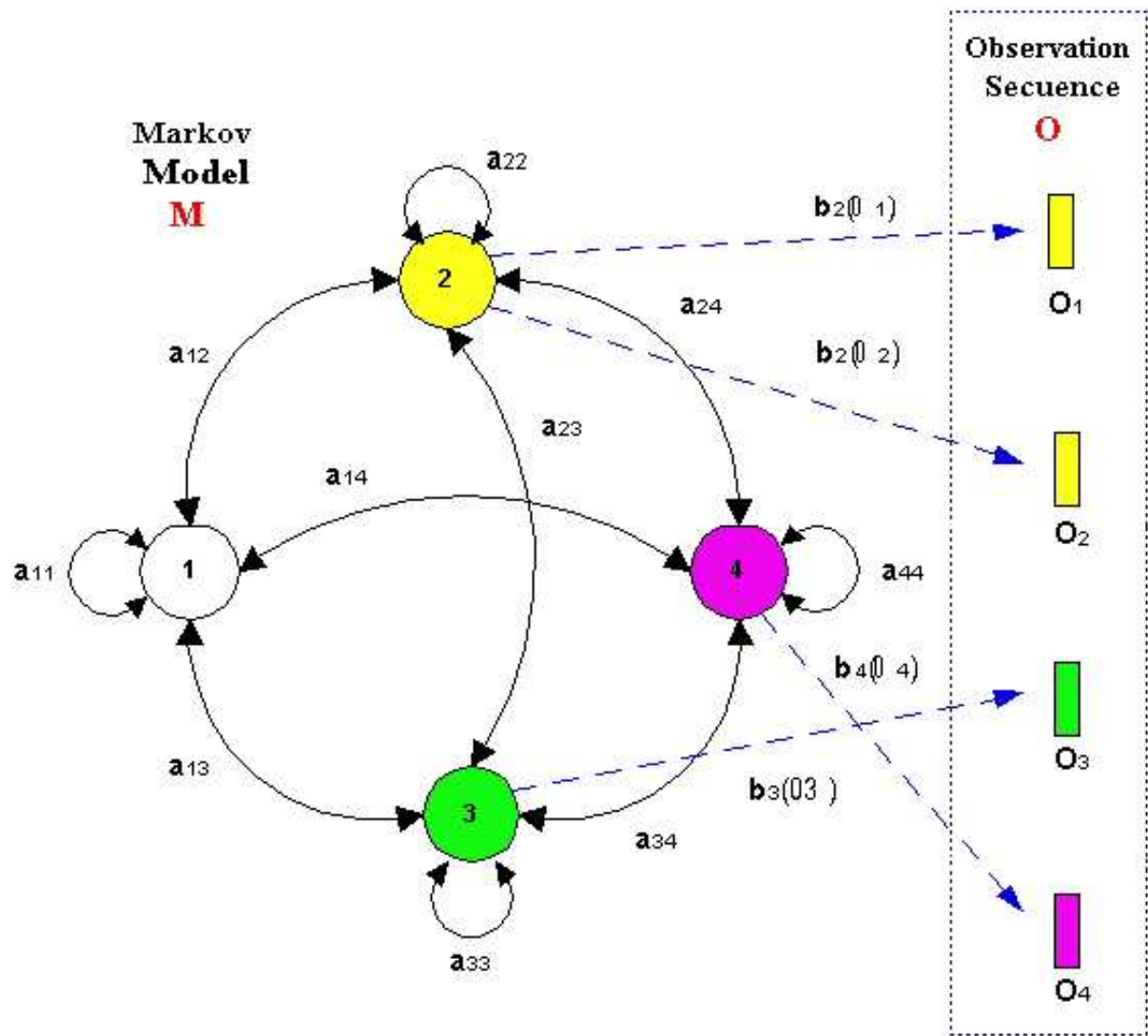
Fig. 4.1 Four state left-right HMM with non-emitting entry and exit states.

where S is the number of parameters in the feature vector, and K is the number of probability density functions (PDFs) considered.

The training algorithm for the HMM consists on finding the parameters of the model (i.e.: the weights for each state of the HMM, c_{ik}, and the transition probabilities between states a_{ij} of the model) from a previously labelled training database. Usually, the maximum likelihood criterion is chosen as the estimation function to adjust the model parameters that is, the maximization of the $P(\mathbf{O}|M)$ over M, where M defines an HMM. However, there is no known way to obtain a solution in a closed form. The Baum-Welch algorithm (Bahl et al., 1983 [27] and Dempster and Rubin, 1977 [28]) is an iterative procedure which provides a locally optimum solution to solve this problem (Fig. 4.4).

4.7 Recognition

The operation of the recognition system is described in equation 3. Note that $P(\mathbf{O}|\mathbf{w})$ is the conditional probability of the feature vector sequence $\mathbf{O}$ given the sequence of events $\mathbf{w}$ which can be computed by the HMMs while $P(\mathbf{w})$ is the probability of the sequence of events $\mathbf{w}$. As there is no statistical knowledge of possible event sequences, we assume that
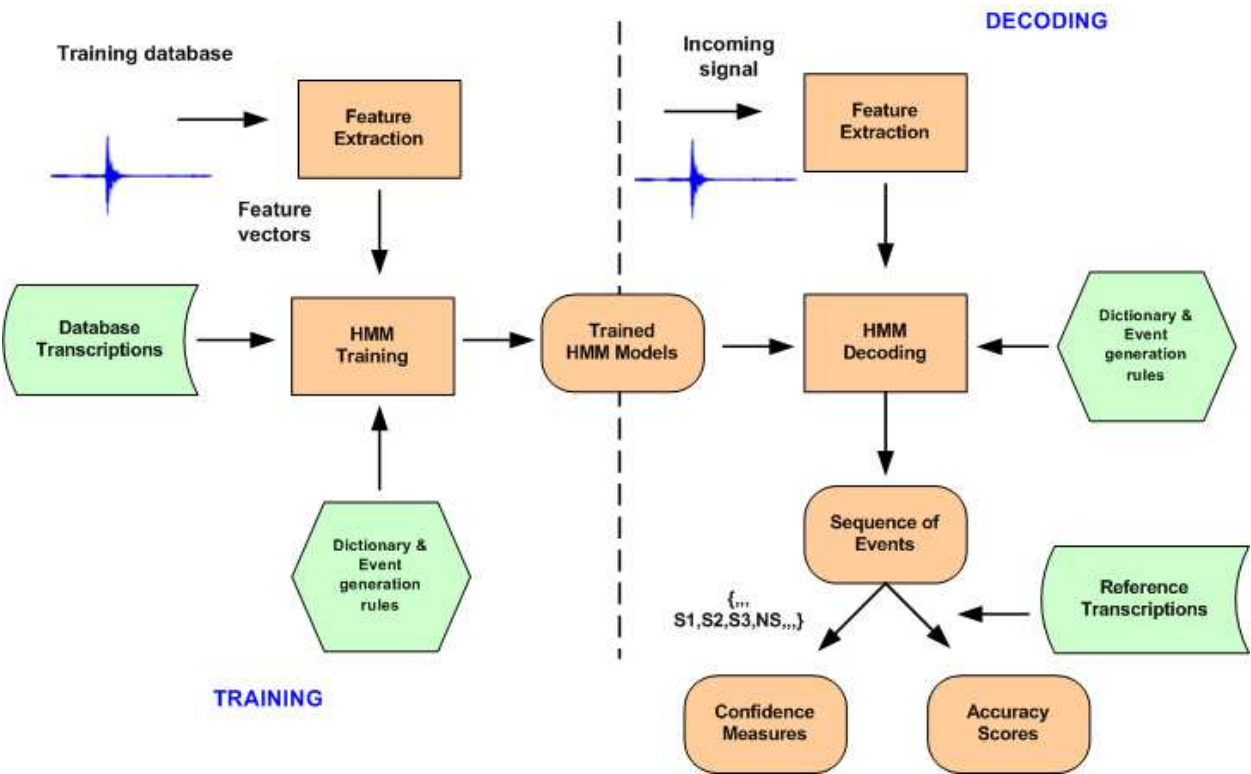
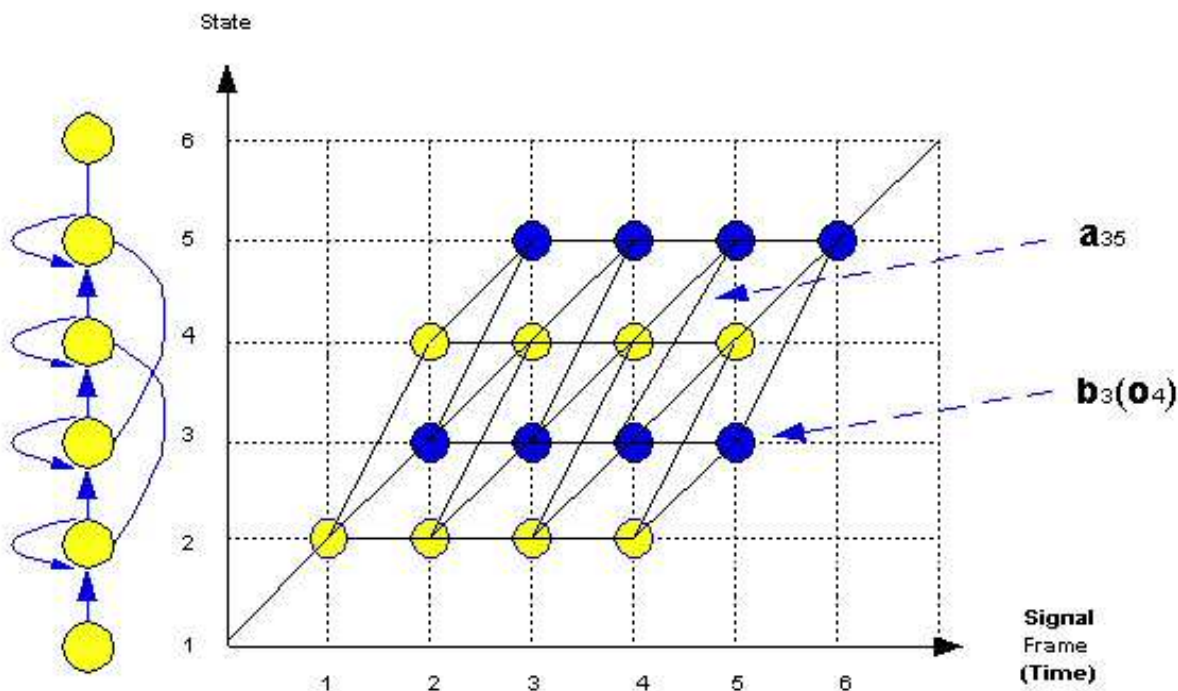Fig. 4.2 Architecture of an HMM-based seismic event classification system.
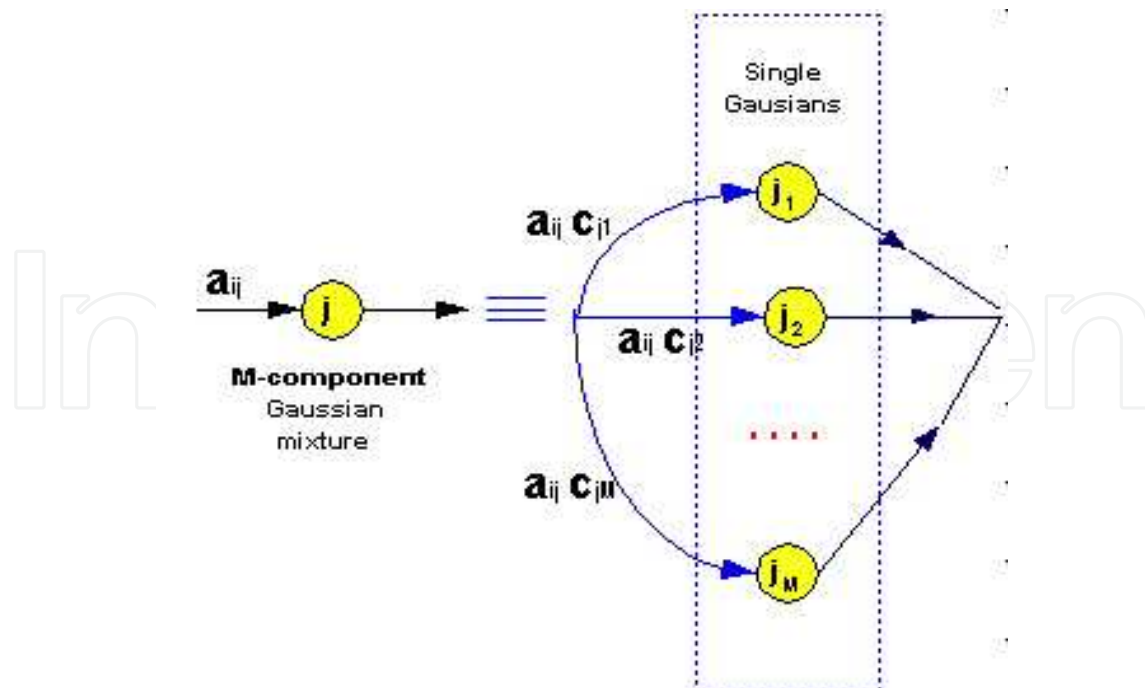

Fig. 4.3 Viterbi algorithm

Fig. 4.4 Baum-Welch algorithm

after a particular event, any other one or noise could appear with the same probability. The recognition system combines the probabilities generated by the models and the probabilities obtained by the allowed transition for the seismic events. Equation 3 indicates that it is necessary to generate all the sequences of events and to evaluate all of them, thus selecting the one with maximum probability. Our recognition system solves this problem by means of the Viterbi decoding algorithm (Rabiner et al. 1993 [25], Furui and Soundhi 1992 [29], Young et al. 1997 [26]).

5. Results

5.1 Preparation of the database

The first step in the data analysis, after the revision of the files, is the visual inspection and segmentation of the signals. As commented before, each file is 300-s-long and could have different types of volcanic signals. For our study we proceeded to distinguish four types of events for Telica volcano and three types of event for San Cristóbal volcano, according with the volcanic activity of each one. For Telica volcano we define four types of seismic signals; Strombolian explosions, Telica explosions, Telica Long Period signal, volcanic tremor and background seismic noise. We have defined these types as T1, T2, T3 and T4. Additionally, we have defined the tremor background seismic noise as NT. In San Cristóbal volcano we identified three types of signals; Strombolian explosions, San Cristóbal explosions, volcanic tremor and background seismic noise. We have defined these types as S1, S2 and S3. Additionally, we have defined the tremor background seismic noise as NS. The rest of the volcanic signals were not considered. In fact, in order to have an adequate model of every class of volcanic signals we need as large as possible number of data, and only the above cited four class could provided a number of events to built a well defined model. The segmentation process consists in to mark over the seismogram the segment that corresponds to every class of events.(Fig. 5.1) This procedure was done by eye and by an unique expert

who has decided the start and end of every signal. In this way we are confident of the use of the same criteria in the selection process. For San Cristóbal volcano we segmented 224 samples of S1, 330 samples of S2, 544 samples of S3 and 1,746 samples of background noise (NS) events. For Telica volcano we segmented 360 samples of T1, 502 samples of T2, 429 samples of T3, 298 samples of T4 and 2,855 samples of background noise (NT) events.
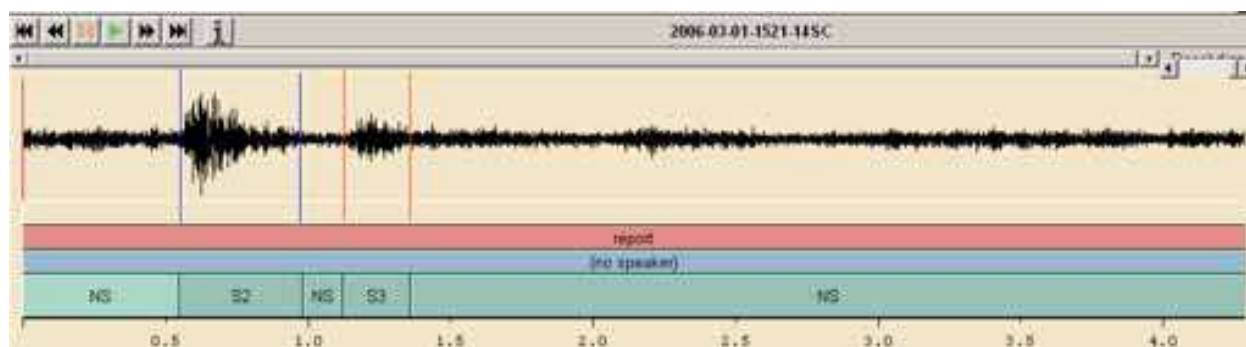

Fig. 5.1 Example of segmentation S2 and S3 signals in San Cristóbal volcano

5.2 Recognition system setup

Once the recordings were manually labelled during the data preparation stage, the recognition system was carefully trained using the available Baum-Welch reestimation algorithms [23] present in the Hidden Markov Model Toolkit (HTK) software [27]. The first step in the training process is the feature extraction process on a frame by frame basis as described above. The 39 parameter feature vector including static coefficients as well as dynamic features is used for training and testing. The training process consists of the initialization of an HMM for each of the events based on a training labelled data set. In this way, different HMMs were built for each of the events analyzed at the different volcanoes: that is, Telica and San Cristóbal event. Moreover, the different background noises observed at both volcanoes were also modelled using different HMMs. The models were defined as left-to-right HMMs where only transitions from state i to state $i+1$ are allowed. The optimal length of the HMMs is mainly determined by the mean duration of the types of events being modelled. Different experiments were carried out in order to determine the best choice. On the other, the HMM emission probabilities at each state were previously defined as Gaussian mixture densities and, during the training process, HMM models with increasing number of Gaussians were built in order to determine the best trade-off between recognition accuracy and performance.

To check the confidence of the method, two types of exercises have been done: closed and blind test. Closed test are done when the same data are used for train and test. For blind test the volcano database is divides in three subset and, rotary two subset are used to train the models and the other one for test; the final result is the average of the three different experiments carried out. Closed test could help to obtain an initial configuration of the recognition system, an initial valuation of the difficulty of the task and, even an idea of the quality of the supervised segmentation of the database. The blind test show objective results of the automatic classification system.

The direct application of the system, specifically trained for the San Cristóbal and Telica volcanoes, provided the following results: We obtained an 88.92% and 74.78% of accuracy in blind test (see table III and VI). Tables I, II, IV and V shows the confusion matrices for the

experiments conducted on the San Cristóbal and Telica volcanoes with 13-state HMMs defined with 16 Gaussians. Tables I and IV shows a set of closed test experiments for the different events for San Cristóbal and Telica database and Tables II and V shows a set of blind test experiments for the different events for San Cristóbal and Telica database.

These matrices indicate the different errors that appear in a automatic recognition system; substitution errors, insertion errors and deletion errors. A substitution error is when the system assign a wrong label to a well know event; a insertion error is when the system identify an event inside of a non labelled segment, for example, an explosion in the middle of the noise, and a deletion error is when the system ignore a labelled event. In all the cases, the confusion matrices are almost diagonal. Reading across the rows of the confusion matrix, each column represents the number of times that the event was automatically labelled by the system as such event. For example Table I shows that 102 San Cristóbal Long Period signal (S1) events were recognized as S1 event and 6 times was not identified by the recognition system ("del" row); the "ins" row indicates the number of time that each one of the events was incorrectly detected when just noise is present in the signal. Tables III and VI shows the percentage correct (%Corr) and the accuracy (%Acc) in closed and blind test.

	NS	S1	S2	S3
NS	617	0	0	13
S1	0	102	0	0
S2	0	0	130	0
S3	0	0	1	193
Ins	0	1	0	1
Del	72	6	21	27

Table I. San Cristóbal training result (closed test)

	NS	S1	S2	S3
NS	422	0	0	0
S1	1	10	0	2
S2	2	1	46	3
S3	2	2	5	157
Ins	3	0	11	11
Del	13	3	5	12

Table II. San Cristóbal test result (blind test)

	%Corr	%Acc
Closed test	88.16	87.99
Blind test	92.57	88.92

Table III. San Cristóbal % result (closed and blind test)

	NT	T1	T2	T3	T4
NT	2193	21	23	6	5
T1	1	210	6	3	6
T2	1	8	360	3	1
T3	1	2	3	330	3
T4	1	0	2	0	264
Ins	65	39	64	32	6
Del	164	28	59	38	30

Table IV. Telica training result (closed test)

	NT	T1	T2	T3	T4
NT	145	0	2	0	0
T1	0	9	2	1	9
T2	0	0	12	0	0
T3	0	1	3	13	3
T4	0	4	3	5	8
Ins	0	1	5	2	1
Del	8	3	2	3	4

Table V. Telica test result (blind test)

With these preliminary set of experiments we have set up the HMM-based recognition system and, in order to validate it, we have carried out experiments with a database containing recordings from both San Cristóbal and Telica volcanoes. With these tests we want to determine if the seismic signals as well as the background noises observed at San Cristóbal and Telica are clearly separable on a mixed data set. Table VII shows the confusion matrix for this test with 16-Gaussian HMMs. It is shown that it is almost diagonal and that San Cristóbal events, Telica events and the different background noises observed at San Cristóbal and Telica are effectively recognized with an accuracy of about 91.41%. In this

way, no confusion occurred between the events at San Cristóbal and Telica so that no San Cristóbal event was classified as an Telica event and viceversa.

6. Discussion and conclusions

In the present work we have developed a volcanic seismic event discrimination system by using an automatic classification system based on Hidden Markov Models (HMM). The present system uses the statistical information compiled from a training set of data, when the event classification was supervised by an expert. From these data sets we have established a model of every event type. The present work used data from two active

	%Corr	%Acc
Closed test	89.00	83.54
Blind test	78.24	74.78

Table VI. Telica % result (closed and blind test)

======================= HTK Results Analysis =======================
-- Overall Results --

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]

	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

Table VII. Confusion matrix for mixed San Cristóbal/ Telica recognition experiments.

MODELS	RECORDS		
	TELICA	**SAN CRISTOBAL**	**MIXED RECORDS**
TRAINING	**83.29%**	**87.99%**	**--**
TEST	**74.48%**	**88.92%**	**--**
TELICA DB TRAINING	**SUCCESS**	**NO SUCCESS**	**--**
SAN CRISTÓBAL DB TRAINING	**NO SUCCESS**	**SUCCESS**	**--**
TELICA AND SAN CRISTÓBAL MIXED DB TRAINING	**--**	**--**	**SUCCESS TRA 91.41% But only with individual training**

Table VIII. Telica and San Cristóbal mixed data bases training.?
DB = DATA BASE, TRA = TRAINING

volcanoes of Nicaragua, San Cristóbal and Telica, with different eruptive processes. We use data from two field surveys carried out in February to March 2006. At every volcano we have identified different types of events, three types and noise for San Cristóbal (S1,S2 and S3), and four types for Telica (T1, T2, T3 and T4) and background noise. The direct application of the system, specifically trained for every volcano, provided the following results:

6.1 San Cristóbal

We obtained, see table III, a 88.92% of accuracy in blind test, with a 11.1% of error. In the error estimate we include two types of results, the insertion of new events (20% of the errors), not labelled previously by the expert, and the deletion of signals. In this situation, the insertion of new events in the data base only could be considered as an error if these new events are not true events. In our case, more than the 90% of the insertions produced by the system correspond to true event that were no previously labelled in the training process. In figures 3.3 and 3.4 we show the seismogram, spectrogram and power spectra of samples of explosions (S2) and background tremor (NS) for San Cristóbal volcano. It is clear the difference between both types of signals.

6.2 Telica

Observing table VI, the application of the system for the data set of Telica has an accuracy of 74.78%, with an error of 25.22%. In this case, the insertions reach up to 69.2% of the total errors, and more than 90% of these insertions were a posteriori identified visually as a valid event. Thus the real accuracy of the system is higher than 80%. Comparing Figures 3.5 and 3.6, we find that the spectral and time-domain characteristics of T1 and T4 event are similar. In spite of these similarities, the system is able to discriminate between both of them.

We have to underscore that in the training process we selected those signals with good signal to noise ratio. We did not consider signals with low quality or with uncertainties in the classification process. However, the system seems to be able to recognize these dubious signals. Therefore, the system helps us to correct omissions performed in the initial training process of the data base. In this sense, the insertions could be considered as successes rather than errors. This ability of the system to detect events that were not previously recognized

by the expert is a consequence of the modelling of the temporal evolution of the power spectra of the seismic events, while the training process is made exclusively attending to the time-domain waveforms. The fact that our overall success rate exceeds 80% reflects the robustness and effectiveness of the proposed methodology.

We wondered if the recognition data base designed using data from a particular volcano and eruptive period could be applied to other volcanoes or different eruptive episodes. In the present work, we have used separately two data base from San Cristóbal and Telica volcanoes. To answer this question, we performed an experiment consisting in the mixing of both data bases, and its application to the discrimination of seismic signals from both volcanoes. There are nine event types in this new data base: S1, S2, S3 and San Cristóbal noise (NS), T1, T2, T3, T4 and Telica Noise (NT). The results are show in Tables VII and VIII. From these results, we conclude that: (a) We do not observe confusion or mixing in the event classification, that is, no event of San Cristóbal was identified as Telica event and viceversa; (b) The events of San Cristóbal was identified separately from the events of Telica; (c) We maintained the same high accuracy rate than in the previous experiments. It is remarkable that when we have mixed both data base we do not have confusion between all the signals. Also we point out that it is not possible to use the data base trained for San Cristóbal to recognize events of Telica, and viceversa. Thus each volcano and signal require an specific training process.

Finally, we conclude that the recognition and classification system based on HMM is a powerful, effective and successful tool. From the application performed over data belonging to two active volcanoes (San Cristóbal and Telica), we have demonstrated that in order to have a reliable result, it is necessary a careful and adequate segmentation process. Also, we proved that each type of signals requires its own characterization. That is, each signal type must be represented by its own specific model, which would include the effects of source, path and sites. Once we have built this model, the success level of the system is high.

7. References

[1] Chouet, B., 1996 *Monitoring and mitigation of volcano hazards*, chap. New methods and future trends in seismological volcano monitoring, pp. 23–97, R Scarpa and Tilling R ed.,Springer-Verlag.

[2] Chouet, B., 1996. Long-period volcano seismicity; its source and use in eruption forecasting, *Nature, 380* (6572), 309–316.

[3] Chouet, B., 2003. Volcano seismology, *Pure Appl. Geophys., 160* (3-4), 739–788.

[4] Baum, L. and Petrie, T. 1966 Statistical inference for probabilistic functions of finite state Markov chains. *Annals of Mathematical Statistics*, 37:1554-1563.

[5] Baum, L., Petrie, T., Soules, G., and Weiss, N., 1970 A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains, Ann. *Math. Statist.*, vol. 41, no. 1, pp. 164—171.

[6] J A. Sánchez, C. M. Travieso, I. G. Alonso, M. A. Ferrer, Handwritten recognizer by its envelope and strokes layout using HMM's, 35rd Annual 2001 IEEE Internacional Carnahan Conference on Security Technology, (IEEE ICCST'01), London, UK, 2001, 267-271.

[7] M. A. Ferrer, J L. Camino, C. M. Travieso, C. Morales, Signatura Classification by Hidden Markov Model, 33rd Anual 1999 IEEE Internacional Carnahan Conference on Security Technology, (IEEE ICCST'99), Comisaría General de Policía Científica,

Ministerio del Interior, IEEE Spain Section, COIT, SSR-UPM, Seguritas Seguridad España S.A, Madrid, Spain, Oct. 1999, 481-484.

[8] J. B. Alonso, C.Carmona, J. de León y M. A. Ferrer, Combining Neural Networks and Hidden Markov Models for Automatic Detection of Pathologies, 16_th Biennial International Eurasip Conference Biosignal 2002, Brno, Check Republic, June 2002.

[9] Renals, S., Morgan, N., Bourlard, H., Cohen, M. & Franco, H. (1994), Connectionist probability estimators in HMM speech recognition, IEEE Transactions on Speech and Audio Processing 2(1), 1994, 161-174.

[10] L.R. Bahl, P.F. Brown, P.V. de Souza, and R.L. Mercer, Maximum mutual information estimation of HMM parameters for speech recognition,. In Proc. IEEE Intl. Conf. on Acoustics, Speech, and Signal Processing, , Tokyo, Japan, December 1986, 49-52

[11] Yin, M.M., Wang, J.T.L., Application of hidden Markov models to gene prediction in DNA, Information Intelligence and Systems, 1999] Proceedings. International Conference on, 1999, 40 – 47.

[12] Cohen, A., Hidden Markov models in biomedical signal processing, Engineering in Medicine and Biology Society, 1998. Proceedings of the 20th Annual International Conf. of the IEEE, Vol. 3, 1998, 1145 – 1150.

[13] C. Benítez, J. Ramírez, J.C. Segura, J.M. Ibáñez, J. Almedros, A. García-Yeguas, and G. Cortés, "Continuous hmm-based seismic event classification at Deception island," *IEEE Trans. Geoscience and Remote Sensing*, vol. 45, pp. 138–147, January 2007.

[14] J.M. Ibáñez, C. Benítez, L.A. Gutiérrez, G. Cortés, A. García-Yeguas, and G. Alguacil, "Classification of seismo-volcanic signals using hidden Markov models: an application to Stromboli and Etna volcanoes," *Submitted for publication to Volcanology and Geothermal Research*, 2009.

[15] M. Beyreuther, R. Carniel, and J.Wassermann, "Continuous hidden Markov models: Applications to automatic earthquake detection and classification at Las Cañadas caldera, Tenerife," *Volcanology and Geothermal Research*, vol. 176, pp. 513–518, 2008

[16] Gutiérrez L., Ibañez J, Benítez C., Ramírez J, Almendros J, García-Yeguas A. "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.

[16a] Ibáñez Jesús M., Benítez Carmen, Gutiérrez Ligdamis A., Cortés Guillermo, Yeguas-García Araceli and Alguacil Gerardo "The classification of seismo-volcanic signals using Hidden Markov Models as applied to the Stromboli and Etna volcanoes" Journal of Volcanology and Geothermal Research Volume 187, Issues 3-4, 10 November 2009, Pages 218-226

[16b] Cortes, G.; Arambula, R.; Gutierrez, L.A.; Benitez, C.; Ibaez, J; Lesage, P.; Alvarez, I.; Garcia, L.; Evaluating robustness of a HMM-based classification system of volcano-seismic events at colima and popocatepetl volcanoes. Geoscience and Remote Sensing Symposium,2009 IEEE International, IGARSS 2009

[17] E. Del Pezzo, A. Esposito, F. Giudicepietro, M. Marinaro, M. Martini, and S. Scarpetta, "Discrimination of earthquakes and underwater explosions using neural networks", *Bull. Seism. Soc. Am..*, vol. 93, no. 1, pp. 215–223, 2003.

[18] S. Scarpetta, F. Giudicepietro, E. Ezin, S. Petrosino, E. Del Pezzo, M. Martini and M. Marinaro, "Automatic classification of seismic signals at Mt. Vesuvius volcano, Italy, using neural networks", *Bull. Seism. Soc. Am.*, vol. 95, no. 1, pp. 185–196, 2005

[19] M. Ohrnberger. *Continuous automatic classification of seismic signals of volcanic origin at Mt. Merapi, Java, Indonesia*, Ph.D. dissertation, Mathematisch-Naturwissenschaftlichen Facultat der Universitat Potsdam, 2001.

[20] Van Wyk de Vries B., 1993. Tectonic and magma evolution of Nicaraguan volcanic systems. Ph.D. Thesis. – Open University, Milton Keynes, 328 p.

[21] M. J Carr and R. E. Stoiber Geologic setting of some destructive earthquakes in Central America *Geological Society of America Bulletin* (January 1977), 88(1):151-156

[22] L.R. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," *Proc. of the IEEE*, vol. 77, no. 2, February 1989.

[23] S. Young, G. Everman, M. Gales, T. Hain, D. Kershaw, X. Liu, G. Moore, J Odell, D. Ollason, D. Povey, V. Valtchev, and P. Woodland, ," *The HTK Book (Version 3.4)*. Cambridge University Engineering Department, 2006.

[24] M. Karnjanadecha, S. Zahorian. "Signal modeling for High-Performance Robus Isolated. Word Recognitions". IEEE Transactions On speech and Audio Processing. Vol 9 Nº 6. September 2001.

[25] L. Rabiner and B. Juang, *Fundamentals of Speech Recognition*. Prentice Hall, 1993.

[26] S. Young, J Odell, D. Ollason, V. Valtchev, and P. Woodland, *The HTK Book*. Cambridge University, 1997.

[27] L. Bahl, F. Jelinek, and R. Mercer, "A maximum likelihood approach to continuous speech recognition," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-5, pp. 179–190, 1983.

[28] A. Dempster and N. L. and D.B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J Roy. Stat. Soc.*, vol. 39, no. 1, pp. 1–38, 1977.

[29] S. Furui and M. M. Sondhi, *Advances in Speech Signal Processing*. Marcel Dekker, Inc., 1992.
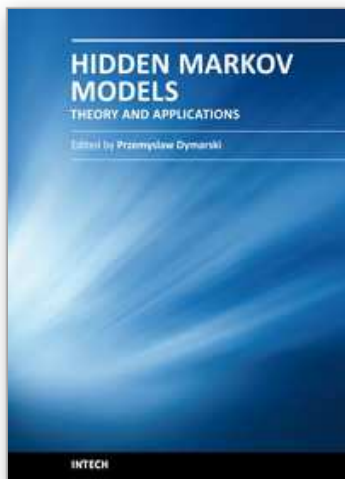
Hidden Markov Models, Theory and Applications

Edited by Dr. Przemyslaw Dymarski

Hidden Markov Models (HMMs), although known for decades, have made a big career nowadays and are still in state of development. This book presents theoretical issues and a variety of HMMs applications in speech recognition and synthesis, medicine, neurosciences, computational biology, bioinformatics, seismology, environment protection and engineering. I hope that the reader will find this book useful and helpful for their own research.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

INTECH
open science | open minds