EVALUATING ROBUSTNESS OF A HMM-BASED CLASSIFICATION SYSTEM OF VOLCANO-SEISMIC EVENTS AT COLIMA AND POPOCATEPETL VOLCANOES

Guillermo Cortés (1) {gcortes@correo.ugr.es}, Raúl Arámbula (2), Ligdamis A. Gutiérrez (1), Carmen Benítez (1), Jesús Ibáñez (3), Philippe Lesage (2), Isaac Álvarez (1) and Luz García (1)

(1) Departamento de Teoría de la Señal, Telemática y Comunicaciones. Universidad de Granada. Granada, Spain.
(2) Instituto de Geofísica, Universidad Nacional Autónoma de México. D.F. México. Mexico.
(3) Instituto Andaluz de Geofísica. Granada, Spain.

1. INTRODUCTION

Classifying events at active volcanoes is usually a hard and time-consuming task carried out by expert technicians in a non-stop monitoring process. Nowadays, automatic detection and classification methods are being developed in order to help with this task and to build early warning robust systems. In this area, the Hidden Markov Models (HMM) are arising as the most promising solution like it was in Automatic Speech Recognition (ASR) [1,2,3,4] versus other approaches [5].

In the present paper we advance one step forward following the ideas already stated in ASR. Previous works [1,2,4] have probed the parallelism between speech and volcano-seismic events in terms of signal complexity and real time requirements. Multi-speaker ASR databases lead to independent classification models able to transcribe the speech of an unknown speaker. Our aim is the design of a general recognition system integrating several volcano-seismic classes from different types of volcanoes to be effortless portable to any other volcano. At first approach, only two Mexican volcanoes of the same type have been mixed for building a complex database used to test robustness of a continuous HMM-based recognition system. For increasing system reliability, a method to assign class confidence scores to each recognized event has been implemented facilitating result interpretation by experts in a critical alert situation or simply in a signal cataloging or monitoring work. We also made a comparative study between the system of the mixed corpus and the ones built with standalone databases analyzing independent results for each event class.

2. BUILDING AND EVALUATING THE CLASSIFICATION SYSTEM

Automatic continuous classification systems based in HMM state of the art have been built by Colima and Popocatepetl corpus. A speech-based Mel Frequency Cepstral Coefficients (MFCC) parametrization of seismic data is achieved to extract relevant signal features [1,2] instead of classic seismic properties used in previous works [4,5]. Multivariate Gaussian probability density functions will capture the underlaying statistics of the feature space. A flexible left to right HMM topology has been adopted named as multi-state HMM which allows to have different number of states for each model class selected in function of their time and feature space variability. Multi-state HMM slightly improve fixed models allowing a class independent compensation of inserted and deleted events.

More than 195 hours of data in Colima (4687 events) and 138 (2101 events) in Popocatepetl were registered at short period and wide band stations and manually labeled by 3 expert technicians into 7 classes: explosions, long period and volcano-tectonic events, collapses, lahars, regional earthquakes, noise and 3 subclasses of volcanic tremor: spasmodic, with repetitive pulses and harmonic. Unlabeled segments were previously removed from continuous records. Independent recognition accuracy scores for each class will help to choose the adequate number of states for the HMM and the components number of the Gaussian mixture which achieves the best recognition results averaging class scores instead of event scores.

2.1. System integrity tests

Tests using the same data corpus for building and evaluating the models were done to check system function. 80%, 89% and 82% of recognition accuracy were obtained respectively for Popocatepetl, Colima and the mixed database. Isolated blind event recognition was also performed to analyze the system response without event insertions and deletion related issues launching 91%, 90% and 88% for the above mentioned corpus. All the tests have been averaged to improve result reliability by 3 equally data partitions, taking two of them for building models and the other for system evaluation, repeating the process three times alternating the train/evaluation partitions.

2.2. System evaluation

Finally, blind tests were carried out achieving more than 73%, 81% and 77% accuracy results for Popocatepetl, Colima and Popocatepetl+Colima. In order to evaluate the robustness of the result reliability assignation method a re-scoring selecting the maximal class confidence score to label each recognized event was made obtaining similar blind test rates.

3. CONCLUSIONS

The result obtained is a very good start point for an automatic continuous classification system built by a database of 8 classes and 3 subclasses collected from 2 volcanoes. A score of 77 % accuracy in blind test is fair enough given the complexity of the recognition task compared to previous studies. The HMM approach is also valid when SNR is too low for the classic detection algorithms to properly work [4]. It is worthwhile to note that the supervised event labeling process is the key to achieve precise models, and human technicians sometimes suffer the lack of an unified criteria which hardly can overcome a 80% of agreement in manual classification. The scoring method assigns a %percent class-belonging confidence value for a recognized event, becoming useful and reliable for cataloging purposes and even for early warning systems. The time to replace human operators in monitoring tasks maybe has come, allowing them to paying their attention towards to result interpretation and alert situations.

4. REFERENCES

[1] C. Benítez, J. Ramírez, J.C. Segura, J.M. Ibáñez, J. Almendros, 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.

[2] J.M. Ibañez, 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.

[3] M. Ohrnberger, “Continuous automatic classification of seismic signals of volcanic origin at Mt. Merapi, Java, Indonesia”, *Ph.D. dissertation*, Math.- Naturwissen. Facultat der Univ. Potsdam, Potsdam, Germany, 2001.

[4] 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.

[5] E. Del Pezzo, A. Esposito, F. Giudicepietro, M. Marinaro, M. Martini, and S. Scarpetta, “Discrimination of earthquakes and underwater explosions using neural networks”, *Bulletin of the Seismological Society of America*, vol. 93, pp. 215–223, February 2003.

[6] Chouet, B., *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, 1996.

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

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