

## Abstract

**title:** Spoken Emotion Recognition using Restricted Boltzmann Machines and Deep Learning

**author:** John Goddard

**affiliation:** Depto. de Ingeniería Eléctrica, Universidad Autónoma Metropolitana, Iztapalapa, Mexico City, Mexico, uamijohn@gmail.com.

### text

The purpose of this talk is to share/present some of the research conducted by members of the Universidad Autónoma Metropolitana (UAM), Mexico in collaboration with researchers from the Universidad Nacional del Litoral (UNL), Argentina on Emotional Speech Recognition. The main focus has centered on the classification problem [1-3], principally applying and comparing restricted boltzmann machines (rbm) and deep belief networks (dbn) to other, commonly used classifiers on some well-known emotional speech databases. This followed previous work using hierarchical classifiers by researchers from the UNL [4] and jointly [5].

The reason for this approach is that in the last few years rbm and dbn have become major topics of research in pattern recognition in general, and speech recognition, speech synthesis, and emotional speech recognition in particular ([6-8] are just three references). This is due to the introduction of new algorithms (e.g. contrastive divergence), which makes the training of these techniques more feasible, as well as a variety of papers showing improvements achieved by applying them. There is also the idea that “there is theoretical evidence which suggests that in order to learn complicated functions that can represent high-level abstractions (e.g. in vision, language, and other AI-level tasks), one needs deep architectures.” [c.f. 9]. Time will tell how important these techniques turn out to be, especially in more complicated problems of spontaneous speech recognition, but initial results in that direction seem promising [10].

### main references

1. Spoken Emotion Recognition Using Deep Learning, E. M. Albornoz, M. Sánchez-Gutiérrez, F. Martínez-Licona, H. L. Rufiner, J. Goddard, Lecture Notes in Computer Science, Vol. 8827, pp. 104-111, 2014.
2. Deep Learning for Emotional Speech Recognition, M. Sánchez-Gutiérrez, E.M. Albornoz, F. Martínez-Licona, H.L. Rufiner, and J. Goddard, Lecture Notes in Computer Science Vol. 8495, pp. 311-320, 2014.
3. Reconocimiento de emociones utilizando técnicas de aprendizaje maquina, Máximo E. Sánchez Gutiérrez, Masters Thesis (*in Spanish*), U.A.M.-I, 2013.
4. Spoken emotion recognition using hierarchical classifiers, Albornoz, E., Milone, D., Rufiner, H., Computer Speech and Language 25, 556–570, 2011.
5. Hierarchical Clustering and Classification of Emotions in Human Speech Using Confusion Matrices, M. Reyes-Vargas, M. Sánchez-Gutiérrez, L. Rufiner, M. Albornoz, L. Vignolo, F. Martínez-Licona, and J. Goddard-Close, Lecture Notes in Artificial Intelligence, Vol. 8113, pp. 162-169, 2013.
6. G. Hinton, L. Deng, D. Yu, G. Dahl, A. Mohamed, N. Jaitly, A. Senior, V. Vanhoucke, P. Nguyen, T. Sainath, and B. Kingsbury, “Deep Neural Networks for Acoustic Modeling in Speech Recognition”, IEEE Signal Processing Magazine, vol. 29, no. 6, pp. 82-97 Nov. 2012.
7. Stuhlsatz, A., Meyer, C., Eyben, F., Zielke, T., Meier, G., Schuller, B.: Deep neural networks for acoustic emotion recognition: Raising the benchmarks. In: ICASSP 2011, pp. 5688–5691 (2011).
8. Deep Learning in Speech Synthesis, H. Zen, 8th ISCA Speech Synthesis Workshop, Barcelona, Spain (2013).
9. Bengio, Y.: Learning Deep Architectures for AI. Foundations and Trends, in Machine Learning 2(1), 1–127 (2009).
10. Le, D.; Provost, E.M., "Emotion recognition from spontaneous speech using Hidden Markov models with deep belief networks," 2013 IEEE Workshop on Automatic Speech Recognition and Understanding (ASRU), pp.216-221, 2013.