

Accepted Manuscript

Hierarchical Sparse Coding Framework for Speech Emotion Recognition

Diana Torres-Boza, Meshia Cédric Oveneke, Fengna Wang, Dongmei Jiang, Werner Verhelst, Hichem Sahli

PII: S0167-6393(17)30341-2
DOI: [10.1016/j.specom.2018.01.006](https://doi.org/10.1016/j.specom.2018.01.006)
Reference: SPECOM 2528



To appear in: *Speech Communication*

Received date: 1 September 2017

Accepted date: 29 January 2018

Please cite this article as: Diana Torres-Boza, Meshia Cédric Oveneke, Fengna Wang, Dongmei Jiang, Werner Verhelst, Hichem Sahli, Hierarchical Sparse Coding Framework for Speech Emotion Recognition, *Speech Communication* (2018), doi: [10.1016/j.specom.2018.01.006](https://doi.org/10.1016/j.specom.2018.01.006)

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Hierarchical Sparse Coding Framework for Speech Emotion Recognition

Diana Torres-Boza^a, Meshia Cédric Oveneke^{a,*}, Fengna Wang^a, Dongmei Jiang^b, Werner Verhelst^a, Hichem Sahli^{a,c}

^a VUB-NPU Joint Audio-Visual Signal Processing (AVSP) Research Lab
Vrije Universiteit Brussel (VUB)
Department of Electronics & Informatics (ETRO)
Pleinlaan 2, 1050 Brussels, Belgium

^b VUB-NPU Joint Audio-Visual Signal Processing (AVSP) Research Lab
Northwestern Polytechnical University (NPU)
Shaanxi Key Lab on Speech and Image Information Processing
Youyo Xilu 127, Xian 710072, China

^c Interuniversity Microelectronics Centre (IMEC)
Kapeldreef 75, 3001 Heverlee, Belgium

Abstract

Finding an appropriate feature representation for audio data is central to speech emotion recognition. Most existing audio features rely on hand-crafted feature encoding techniques, such as the AVEC challenge feature set. An alternative approach is to use features that are instead learned automatically. This has the advantage of generalizing well to new data, particularly if the features are learned in an unsupervised manner with less restrictions on the data itself. So in this work, we adopt sparse coding framework as a means to automatically represent features from audio named hierarchical sparse coding (HSC). Results indicates that the obtained features, in an unsupervised fashion, are able to capture useful properties of the speech that distinguish between emotions.

Keywords: Affective Computing, Speech Emotion Recognition, Sparse Coding, Support Vector Regression

*Corresponding author

Email addresses: dtboza@etrovub.be (Diana Torres-Boza), mcovenek@etrovub.be (Meshia Cédric Oveneke), fwang@etrovub.be (Fengna Wang), jiangdm@nwpu.edu.cn (Dongmei Jiang), wverhels@etrovub.be (Werner Verhelst), hsahli@etrovub.be (Hichem Sahli)

Download English Version:

<https://daneshyari.com/en/article/6960665>

Download Persian Version:

<https://daneshyari.com/article/6960665>

[Daneshyari.com](https://daneshyari.com)