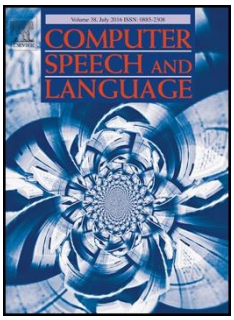
Accepted Manuscript



Title: On the use of deep feedforward neural networks for automatic language identification

Author: Ignacio Lopez-Moreno, Javier Gonzalez-Dominguez, David Martinez, Oldřich Plchot, Joaquin Gonzalez-Rodriguez, Pedro J. Moreno

PII: S0885-2308(15)30036-X
DOI: http://dx.doi.org/doi: 10.1016/j.csl.2016.03.001
Reference: YCSLA 768

To appear in: *Computer Speech & Language*

Received date: 30-10-2015
Accepted date: 18-3-2016

Please cite this article as: Ignacio Lopez-Moreno, Javier Gonzalez-Dominguez, David Martinez, Oldřich Plchot, Joaquin Gonzalez-Rodriguez, Pedro J. Moreno, On the use of deep feedforward neural networks for automatic language identification, *Computer Speech & Language* (2016), http://dx.doi.org/doi: 10.1016/j.csl.2016.03.001.

On the use of Deep Feedforward Neural Networks for Automatic Language Identification

Ignacio Lopez-Moreno[a,*], Javier Gonzalez-Dominguez[b], David Martinez[c], Old ř ich Plchot,[d], Joaquin Gonzalez-Rodriguez[b], Pedro J. Moreno[a]

[a]*Google Inc., New York, USA*

[b]*ATVS-Biometric Recognition Group, Universidad Autonoma de Madrid, Madrid, Spain*

[c]*I3A, Zaragoza, Spain*

[d]*Brno University of Technology, Brno, Czech Republic*

*Corresponding author

Email address: E-mail address:elnota@google.com (Ignacio Lopez-Moreno)

HIGHLIGHTS

- This work presents a novel comprehensive study on the use of deep neural networks (DNNs) for automatic language identification (LID).
- It includes a detailed performance analysis for different data selection strategies, DNN architectures and test segments of different durations.
- Presented results are computed on the public and largely benchmarked NIST Language Recognition Evaluation 2009 (LRE'09), and tested the proposed systems against a highly competitive state-of-the-art i-vector baseline system,
- It also presents novel an approach that combines DNN and i-vector systems by using Bottleneck Features.
- Best system performance is achieved by combining the results from the proposed Bottleneck System and the baseline i-vector system. Compared to an state-of-the-art i-vector system, the combined system achieves a 45% of relative improvement in both EER and Cavg, on 3s and 10s.

Abstract

In this work, we present a comprehensive study on the use of deep neural networks (DNNs) for automatic language identification (LID). Motivated by the recent success of using DNNs in acoustic modeling for speech recognition, we adapt DNNs to the problem of identifying the language in a given utterance from its short-term acoustic features. We propose two different DNN-based approaches. In the first one, the DNN acts as an end-to-end LID classifier, receiving as input the speech features and providing as output the estimated probabilities of the target languages. In the second approach, the DNN is used to extract bottleneck features that are then