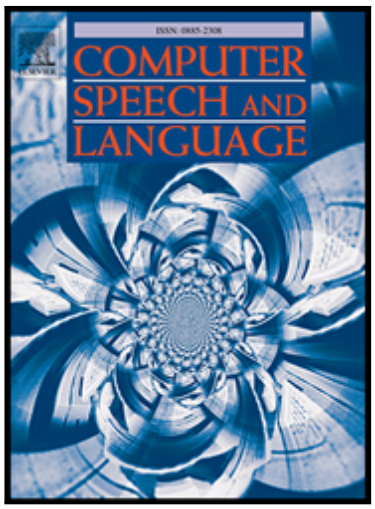

Comparing human and automatic speech recognition in simple and complex acoustic scenes

Constantin Spille, Birger Kollmeier, Bernd T. Meyer

PII: S0885-2308(17)30170-5
DOI: 10.1016/j.csl.2018.04.003
Reference: YCSLA 922

To appear in: *Computer Speech & Language*

Received date: 22 June 2017
Revised date: 1 December 2017
Accepted date: 11 April 2018

Please cite this article as: Constantin Spille, Birger Kollmeier, Bernd T. Meyer, Comparing human and automatic speech recognition in simple and complex acoustic scenes, *Computer Speech & Language* (2018), doi: 10.1016/j.csl.2018.04.003

Highlights

- Automatic speech recognition and human listeners are compared in single-channel and spatial scenes
- In single-channel scenes, ASR is on a par with normal-hearing listeners
- In spatial scenes, there is a substantial human-machine gap of 12.3 dB
- 5.3 dB of this gap can be attributed to poor localization and missing speaker-related features