Using deep learning to improve the intelligibility of a target speaker in noisy multi-talker environments for people with normal hearing and hearing loss

Iordanis Thoidis[1], Tobias Goehring[2]

[1]Aristotle University of Thessaloniki, Thessaloniki Greece, [2]MRC Cognition and Brain Sciences Unit, University of Cambridge, UK

Introduction

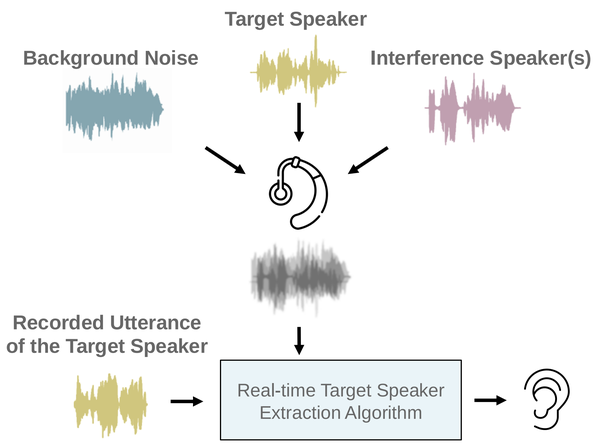
Speech perception in noisy environments with multiple talkers is a challenging task for listeners with hearing loss and for normal-hearing listeners.

Speech enhancement based on deep learning has shown great potential to improve speech intelligibility in the presence of background noise.

Realistic acoustic scenes are inherently dynamic, with alternations and overlap between speakers that vary both in level and in timing.

Aim: Restore ability of a listener to understand a target speaker and ignore others in realistic noise.

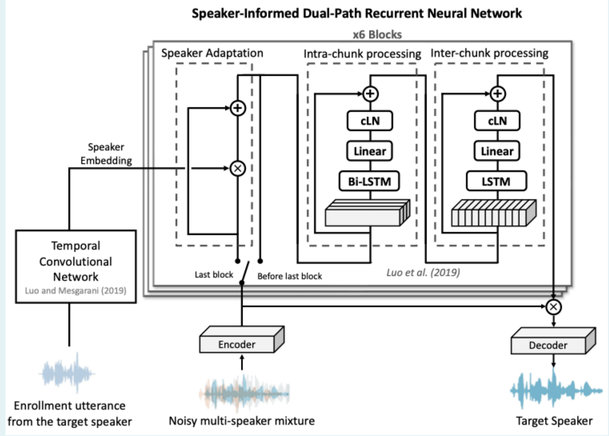
Approach: Real-time Target Speaker Extraction
Extract the voice of a target speaker from a mixture, given a recorded utterance of the target speaker. Evaluate the intelligibility of the target speaker.



Check out the full paper: https://doi.org/10.1121/10.0028007
Thoidis, I., & Goehring, T. (2024). Using deep learning to improve the intelligibility of a target speaker in noisy multi-talker environments for people with normal hearing and hearing loss. *JASA*, *156*(1), 706-724.

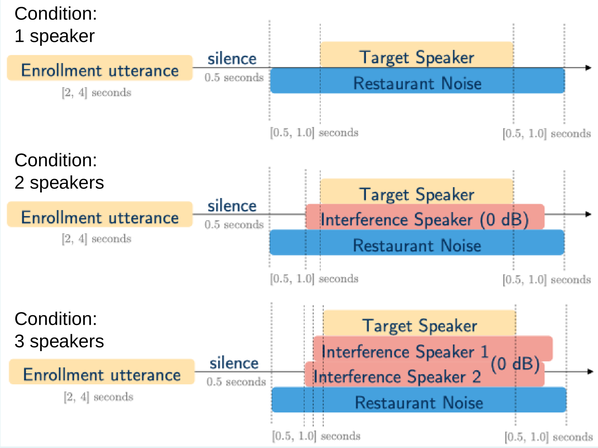
Deep Learning Model

- Deep Neural Network: 11.8M parameters
- Real-time processing: Quasi-causal (5 ms)



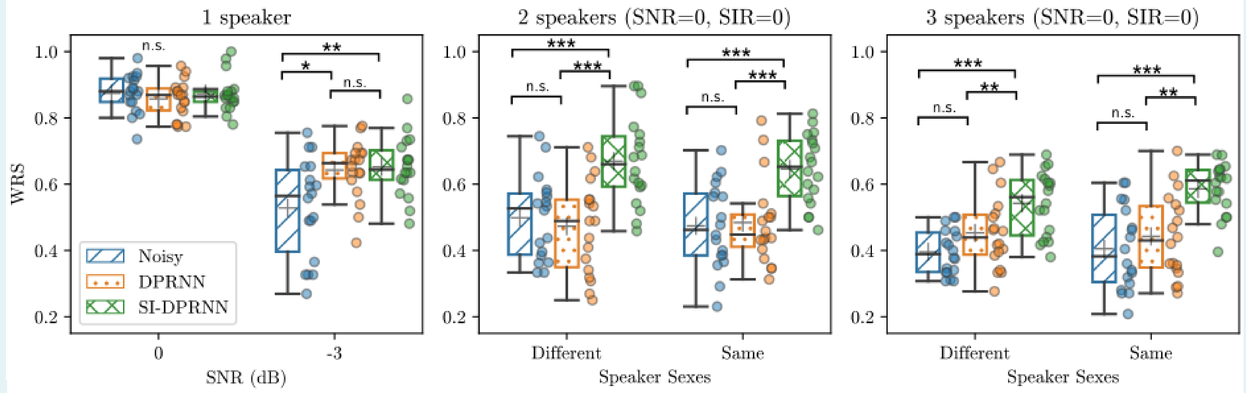
Stimuli

Target Speaker: Greek Harvard speech corpus
Interference Speakers: Clarity British English

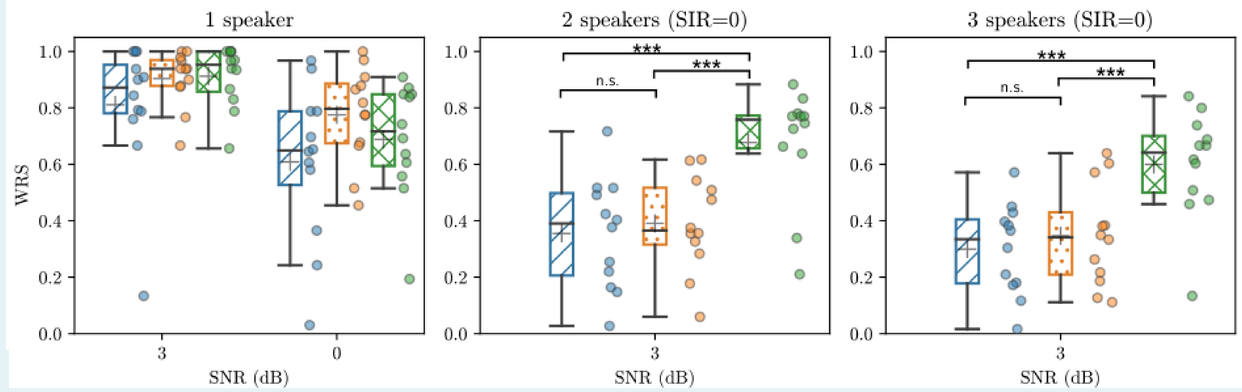


Results

Group with Normal Hearing ($N = 18$)



Group with Hearing Loss ($N = 12$, Bilateral, PTA $\in$ *[36, 61]*)



Discussion

+**10.1 dB** ΔSI-SDR in noisy multi-talker conditions

A **single algorithm** that generalizes to different speakers, noises, and number of speakers.

Significant speech intelligibility improvements of **17% for people with normal hearing** and **31% for people with hearing loss**.

Acknowledgements

Author TG was supported by Career Development Award MR/T03095X/1 from the Medical Research Council UK.

Thanks to Prof. Konstantinos Markou, Anastasia Kypriotou, and Iriana Chrysikou for their help in the recruitment process.