

Lightweight Real-Time Recurrent Models for Speech Enhancement and Automatic Speech Recognition

By Dhahbi, S (Dhahbi, Sami) ; Saleem, N (Saleem, Nasir) ; Gunawan, TS (Gunawan, Teddy Surya) ; Bourouis, S (Bourouis, Sami) ; Ali, I (Ali, Imad) ; Trigui, A (Trigui, Aymen) ; Algarni, AD (Algarni, Abeer D.)

[View Web of Science ResearcherID and ORCID](#) (provided by Clarivate)

Source [INTERNATIONAL JOURNAL OF INTERACTIVE MULTIMEDIA AND ARTIFICIAL INTELLIGENCE](#) ▾

Volume: 8 Issue: 6

DOI: 10.9781/ijimai.2024.04.003

Published JUN 2024

Indexed 2024-06-28

Document Type Article

Abstract Traditional recurrent neural networks (RNNs) encounter difficulty in capturing long-term temporal dependencies. However, lightweight recurrent models for speech enhancement are important to improve speech, while being computationally efficient and able to capture long-term temporal dependencies efficiently. This study proposes a lightweight hourglass-shaped model for speech enhancement (SE) and automatic recognition (ASR). Simple recurrent units (SRU) with skip connections are implemented where attention gates are added to the skip connections, highlighting the important features and spectral regions. The operates without relying on future information that is well-suited for real-time processing. Combined acoustic features and two training objectives are estimated. Experimental evaluations using the short time intelligibility (STOI), perceptual evaluation of speech quality (PESQ), and word error rates (WERs) indicate better intelligibility, perceptual quality, and word recognition rates. The composite measures further confirm the performance of residual noise and speech distortion. With the TIMIT database, the proposed improves the STOI and PESQ by 16.21% and 0.69 (31.1%) whereas with the LibriSpeech database, the improves STOI by

16.41% and PESQ by 0.71 (32.9%) over the noisy speech. Further, our model outperforms other deep neural networks (DNNs) in seen and unseen conditions. The ASR performance is measured the Kaldi toolkit and achieves 15.13% WERs in noisy backgrounds.

Keywords

Author Keywords: Real-Time Speech; Simple Recurrent Unit (SRU); Speech Enhancement; Speech Processing; Speech Quality

Keywords Plus: NEURAL-NETWORKS; FEEDFORWARD; MASKING; NOISE; LSTM

Addresses

¹ King Khalid Univ, Coll Sci & Art Mahayil, Dept Comp Sci, Muhayil Aseer 62529, Saudi Arabia

² Gomal Univ, Dept Elect Engn, FET, Dera Ismail Khan 29050, KPK, Pakistan

³ Int Islamic Univ Malaysia, Elect & Comp Engn Dept, Kuala Lumpur, Malaysia

⁴ Taif Univ, Coll Comp & Informat Technol, Dept Informat Technol, At Taif 21944, Saudi Arabia

⁵ Univ Swat, Dept Forens Sci, Swat, Pakistan

[...more addresses](#)

Categories/ Classification

Research Areas: Computer Science

Web of Science Categories

[Computer Science, Artificial Intelligence](#); [Computer Science, Interdisciplinary Applications](#)

Language

English

Accession Number

WOS:001239818000008

ISSN

1989-1660

IDS Number

TF3W6

[– See fewer data fields](#)

Citation Network

In Web of Science Core Collection

Use in Web of Science

1

1

0 Citations

60 Cited References

How does this document's citation performance compare to peers?

[← Open comparison metrics panel](#)

Data is from InCites Benchmarking & Analytics

Last 180 Days

Since 2013

This record is from:

Web of Science Core Collection

- Science Citation Index Expanded (SCI-EXPANDED)
-

Suggest a correction

If you would like to improve the quality of the data in this record, please

[Suggest a correction](#)

© 2024
Clarivate
Training
Portal
Product
Support

Data
Correction
Privacy
Statement
Newsletter

Copyright
Notice
Cookie
Policy
Terms of
Use

[Manage cookie preferences](#)

Follow
Us

