HAL open science

Babble noise augmentation for phone recognition applied to children reading aloud in a classroom environment

Lucile Gelin, Morgane Daniel, Thomas Pellegrini, Julien Pinquier

To cite this version:

Lucile Gelin, Morgane Daniel, Thomas Pellegrini, Julien Pinquier. Babble noise augmentation for phone recognition applied to children reading aloud in a classroom environment. Speech in Noise Workshop (SPiN 2020), Jan 2020, Toulouse, France. 10.13140/RG.2.2.32155.21287 . hal-03372743

HAL Id: hal-03372743

https://hal.science/hal-03372743v1

Submitted on 11 Oct 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

BABBLE NOISE AUGMENTATION FOR PHONE RECOGNITION APPLIED TO CHILDREN READING ALOUD IN A CLASSROOM ENVIRONMENT

Lucile Gelin[1,2], Morgane Daniel[1], Thomas Pellegrini[2], Julien Pinquier[2]
[1]Lalilo, [2]Institut de Recherche en Informatique de Toulouse (IRIT)

OBJECTIVES

Reading is a key step in a child's learning, but teachers do not have time to spend individually with each student to monitor progress. We are developing a tool that aims at automatically evaluating children's fluency and pronunciation as they read aloud.

CHALLENGES

- Lack of children speech data in French
- Characteristic non-reader children's prosody
- **Babble noise present in classroom environment**

CORPORA

- Lalilo French children corpus, composed of 5-7 year-old children reading aloud in a classroom environment
- Commonvoice French adult corpus
- DEMAND babble noise corpus [1]

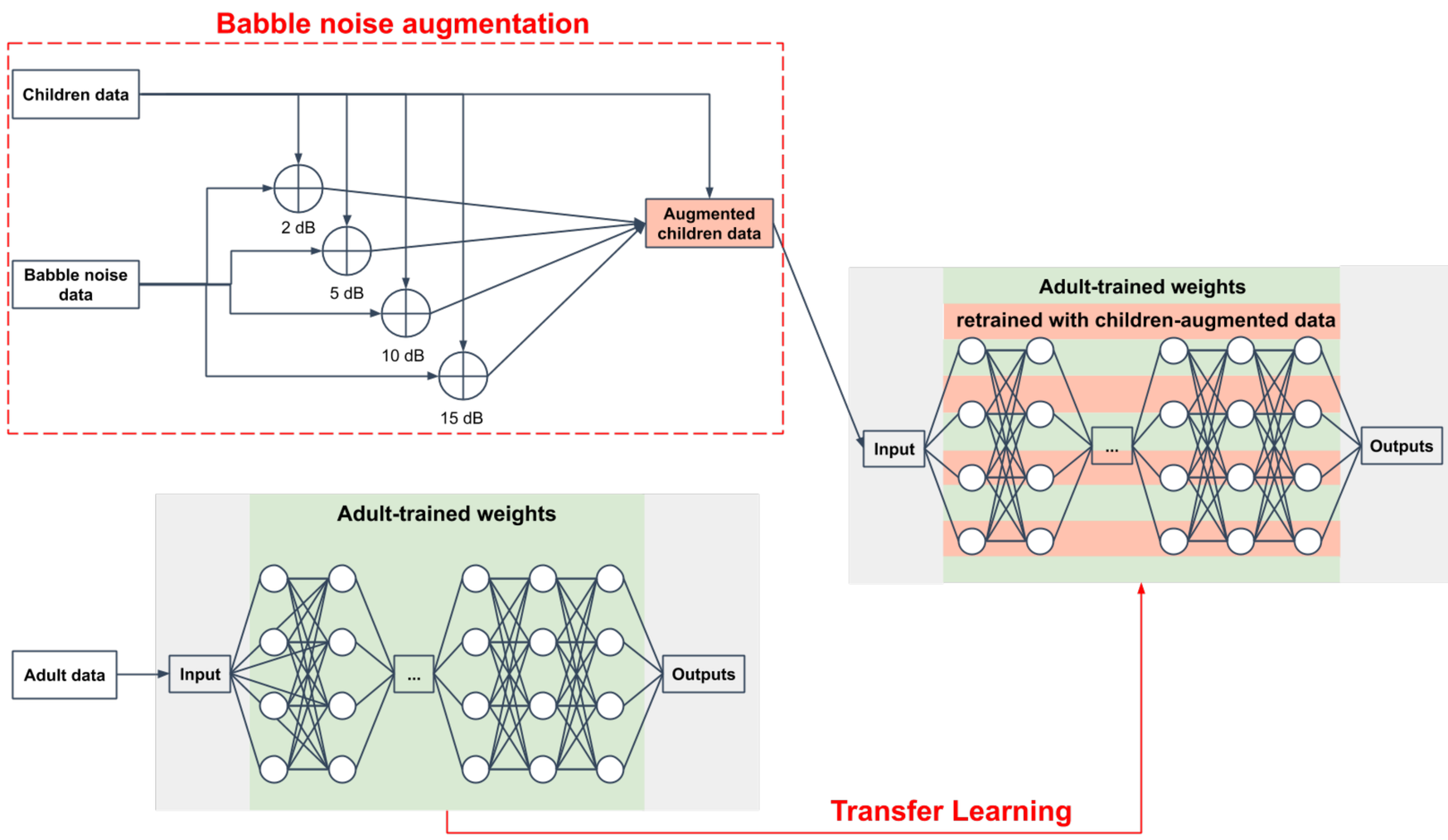
Corpus	**Lalilo**		**Commonvoice**	**DEMAND**
Set	Train	Test	Train	Babble
Duration (h)	13.8	0.8	148.9	1.3
SNR (dB)	24 ±13	22 ±11	35 ±16	–

Information on corpora

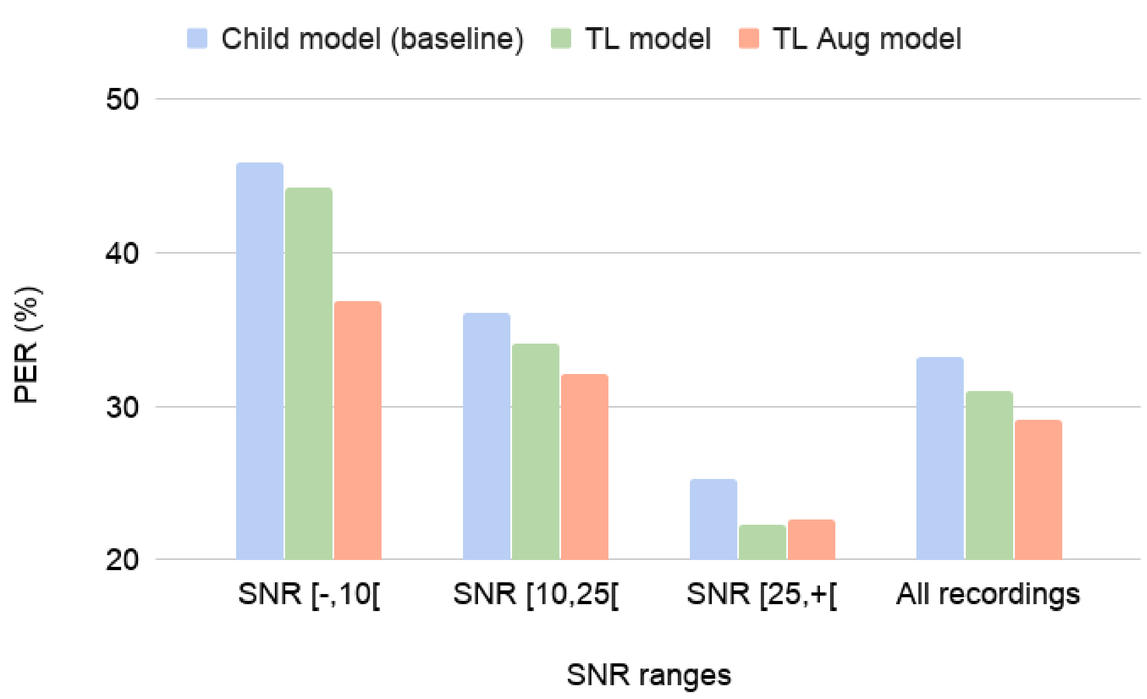
METHOD

The baseline model is a children-trained factorized time-delay neural network [2]. We combine two methods to improve our system's performance:

- **Babble noise augmentation (Aug)**: concatenating children data with babble noise data at different signal to noise ratios (SNR) to train the acoustic model → makes the model more robust to classroom babble noise
- **Transfer Learning (TL, [3])**: transferring knowledge learned from a big corpus of adult data by re-training it with children data → overcomes lack of children data



RESULTS



Phoneme Error Rate (PER,%) obtained with the different acoustic models, global (All files) and for different SNR ranges

CONCLUSION

- Transfer learning brings overall relative improvement of 7%, with a greater efficiency on clean recordings.
- Babble noise augmentation brings further decrease of PER, especially on low and medium SNR ranges (respectively 17% and 6% relative improvement over TL model). While the improvement over TL model is not significant for clean recordings, the PER is not degraded. While the improvement over TL model is not significant for clean recordings, the global relative PER improvements brought by the TL-Aug model are 6% over the TL model and 12% over the baseline model.

References

[1] Thiemann J. et al. (2013). *DEMAND: a collection of multi-channel recordings of Acoustic Noise in Diverse Environments*
[2] Povey D. et al. (2018). *Semi-Orthogonal Low-Rank Matrix Factorization for Deep Neural Networks*
[3] Shivakumar P.G. and Georgiou P. (2018). *Transfer Learning from Adult to Children for Speech Recognition: Evaluation, Analysis and Recommendations.*