

## On Front-End Gain Invariant Modeling for Wake Word Spotting

*Yixin Gao, Noah D. Stein, Chieh-Chi Kao, Yunliang Cai, Ming Sun, Tao Zhang, Shiv Naga Prasad Vitaladevuni*

Wake word (WW) spotting is challenging in far-field due to the complexities and variations in acoustic conditions and the environmental interference in signal transmission. A suite of carefully designed and optimized audio front-end (AFE) algorithms help mitigate these challenges and provide better quality audio signals to the downstream modules such as WW spotter. Since the WW model is trained with the AFE-processed audio data, its performance is sensitive to AFE variations, such as gain changes. In addition, when deploying to new devices, the WW performance is not guaranteed because the AFE is unknown to the WW model. To address these issues, we propose a novel approach to use a new feature called  $\Delta$ LFBE to decouple the AFE gain variations from the WW model. We modified the neural network architectures to accommodate the delta computation, with the feature extraction module unchanged. We evaluate our WW models using data collected from real household settings and showed the models with the  $\Delta$ LFBE is robust to AFE gain changes. Specifically, when AFE gain changes up to  $\pm 12$ dB, the baseline CNN model lost up to relative 19.0% in false alarm rate or 34.3% in false reject rate, while the model with  $\Delta$ LFBE demonstrates no performance loss.





ISCA

Archive

Interspeech 2020