Low-Resource Keyword Spotting for Hearing Assistive Devices

Iván López-Espejo[1], Zheng-Hua Tan[1] and Jesper Jensen[1,2]

[1]Department of Electronic Systems, Aalborg University, Denmark
[2]Oticon A/S, Denmark
{ivl,zt,jje}@es.aau.dk, jesj@oticon.com

AALBORG UNIVERSITY
DENMARK

CASPR
Centre for Acoustic Signal Processing Research

Motivation

- Manual operation of **hearing assistive devices** (**HADs**) is cumbersome in a number of situations.
- To assist in addressing this issue, **voice interfaces** are envisioned as a means **for handling and operating HADs** in a practical manner.

Objectives

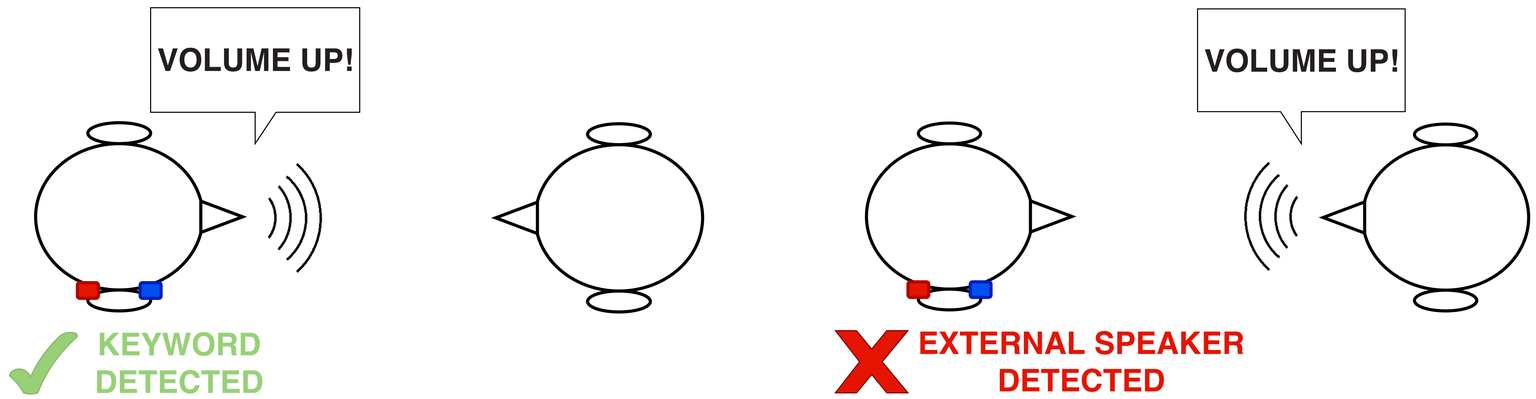
- Research and development of **keyword spotting** (**KWS**) systems **for HADs**:
  - Personalization.
  - Robustness against noise.
  - Low memory and low computational complexity.

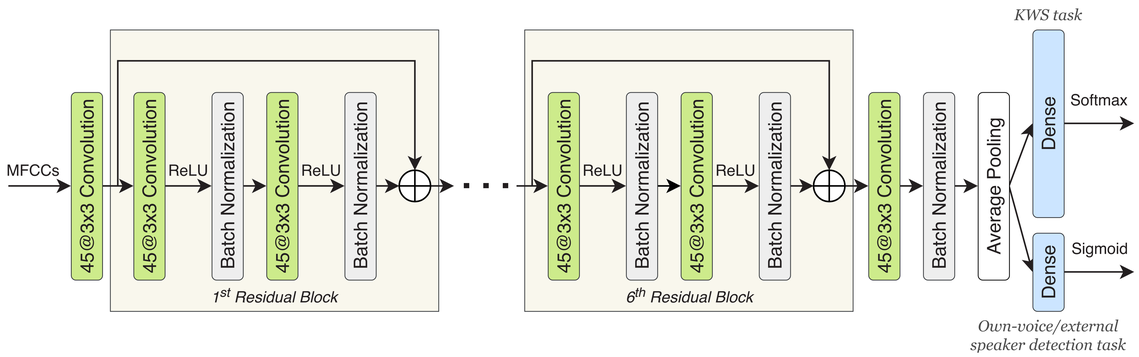
- To accomplish these objectives, **we explore**...
  - ...the combined use of **multi-microphone signals** from HADs along with signal processing **and** the latest **deep learning** techniques.
  - ...the utilization of **user-specific aspects**, e.g., voice characteristics or head-related acoustics of the specific user.

- We expect to contribute to enhance the life quality of hearing-impaired people.

Example: KWS for HADs Robust to External Speakers

- **KWS** systems **for HADs must be robust against external speakers**, that is, the user must be the only one allowed to trigger actions on her/his HAD.
- A first attempt on **personalized**, i.e., user-dependent, **KWS for HADs**:
  - Iván López-Espejo, Zheng-Hua Tan and Jesper Jensen, "Keyword Spotting for Hearing Assistive Devices Robust to External Speakers", *in* Proc. of Interspeech 2019, pp. 3223–3227, Graz (Austria), 2019.





- We proposed **HAD user-dependent KWS** drawing from a KWS system **based on deep residual learning** and dilated convolutions [Tang18].

- **Experiments on a** (simulated) **two-microphone hearing aid speech database** showed that our proposed approach is robust against external speakers.

	Keyword spotting accuracy (%)	
	Own-voice subset	*Overall set*
Baseline [Tang18]	94.21 ± 0.39	71.87 ± 0.30
Proposal	94.59 ± 0.32	94.86 ± 0.39

[Tang18] Raphael Tang and Jimmy Lin, "Deep residual learning for small-footprint keyword spotting", *in* Proc. of ICASSP 2018, pp. 5484–5488, Calgary (Canada), 2018.