Access Control Method for the Offline Home Automation System

Mikołaj Pabiszczak, Monika Grajzer, Łukasz Sawicki

Mikołaj Pabiszczak
Gido Labs
m.pabiszczak@gidolabs.eu

eKNOW 2020
The Twelfth International Conference on Information, Process, and Knowledge Management
21-25 November 2020 - Valencia, Spain

Resume of presenter

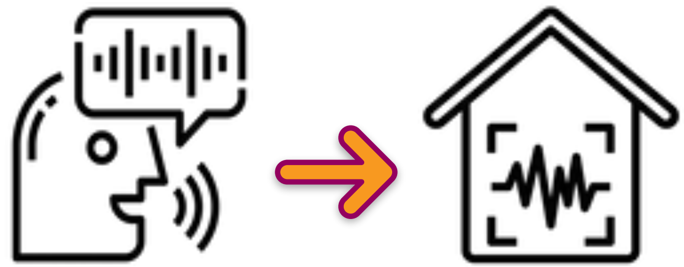
- BEng in Materials' Engineering obtained at Poznań University of Technology, Poland.
- MA in Theoretical Mathematics obtained at Adam Mickiewicz University in Poznań, Poland.
- Machine Learning Engineer at Gido Labs.

Outline

- Scope of the paper
- Problem statement
- Related work
- The proposed solution
- Experiments and results
- Conclusion

Scope of the paper

- Home Automation Systems using small, embedded control devices



- Voice-based interface: access to the main speech recognition system (ASR) is controlled by an Access Control Decision Support System:
  - Voice Activity Detection (VAD) system for recognizing audio only when voice is detected AND
  - **Keyword Spotting Module (KWS)** — full ASR functionality is turned on after detecting proper keyphrase
  - Speaker Recognition - voice biometrics to grant access to the system only to authorized users

The graphics on this slide: House by Komkrit Noenpoempisut from the Noun Project, voice control by monkik from the Noun Project

Problem statement

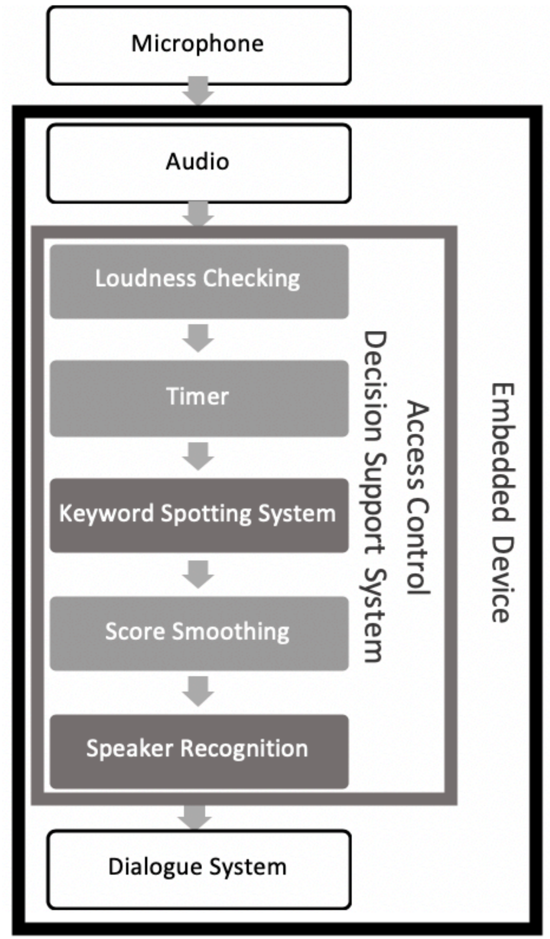
- Home automation systems with speech-based interfaces become increasingly popular.
- BUT: speech recognition is a resource-consuming task typically performed in the cloud => privacy concerns
- Offline systems working fully locally are desirable but <u>challenging on small embedded devices</u>
- Additional challenges:
  - support for non-English languages
  - relatively small dataset of examples with the recordings of a selected keyword

Related work

- KWS is a core part of an Access Control DSS
- Convolutional Neural Networks and Residual Neural Networks (ResNets) used for KWS:
  - State of the art KWS systems reach accuracy of 95% with False Positive Rate (FPR) of 2%.
  - BUT: this makes those solutions inapplicable „as is" in commercial set-ups - if a system makes prediction every second, with FPR=2% there will be ~72 false alarms in an hour.
- ResNets for Speaker Recognition

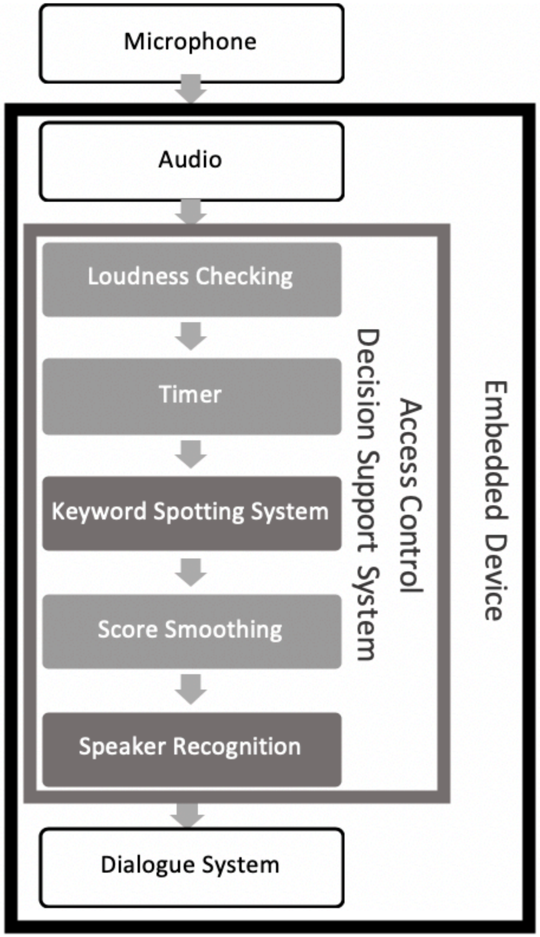
Solution

- In typical reference system only KWS and SR modules would be present
- Our solution: several other modules proposed (light gray)
- GOAL: reduce FPR of the AccessControl DSS



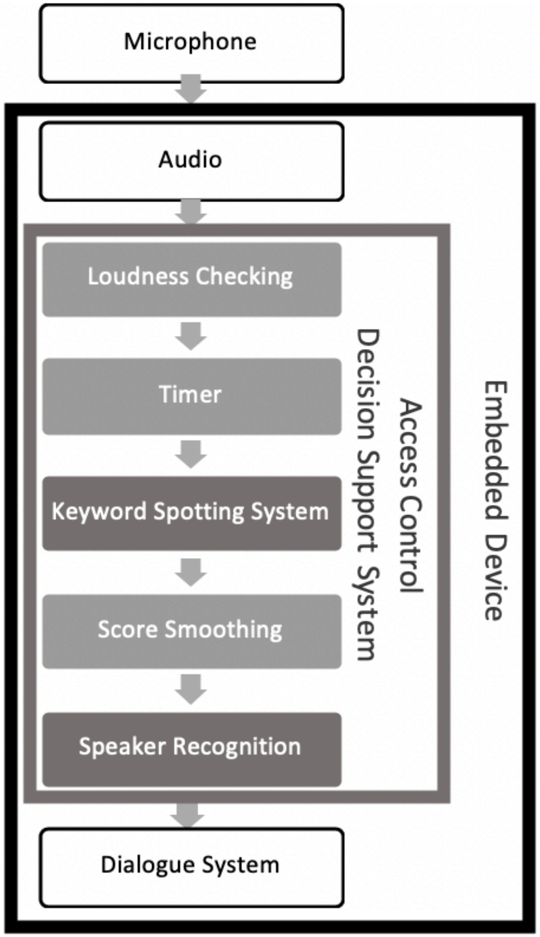
Solution - details

- Loudness Checking - allow only audio with the loudness above certain threshold to be processed further.
- Timer - after minimal loudness was reached, process only first 1.2s of audio (keyword is approx. 1s long)



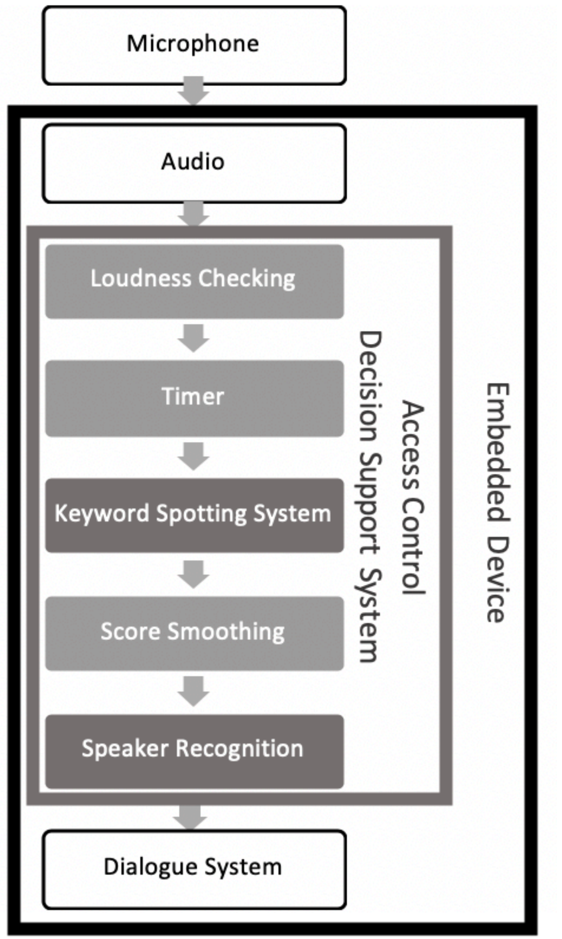
Solution - details

- KWS module: a small ResNet (110k parameters) using 40 MFCCs as input
- Transfer learning:
  - model trained for recognizing 10 English keywords as a basis
  - adapted for recognizing single keyword from Polish language (by using smaller dataset of examples)



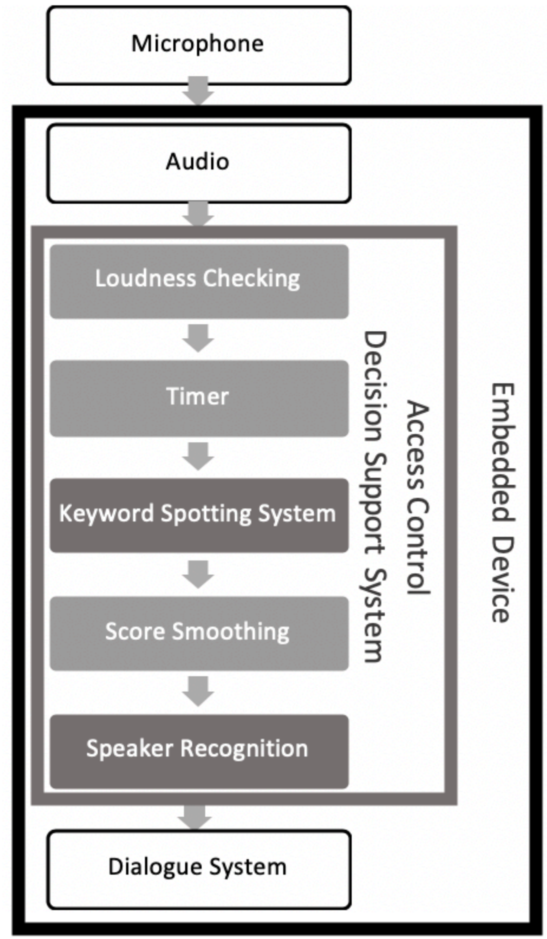
Solution - details

- Score Smoothing - calculates mean score of the last *n* predictions and checks if it is above certain threshold:
  - comparison between averaging over 3 or 4 predictions



Solution - details

- Speaker Recognition - takes a single frame with the strongest trigger.
- Utilises different ResNet as a classifier.



Evaluation

- Evaluation of the 3 main components proposed for Access Control DSS and their combinations:
  - Loudness Checking
  - Timer
  - Score smoothing with averaging over last 3 or 4 predictions
- AIM: estimate their influence on both False Positive Rate (FPR) and True Positive Rate (TPR)
- Reference system: only KWS block present
- The DSS system was implemented on a RaspberryPi 3B (CPU: 1,2 GHz quad-core; 1 GB RAM) with a custom-made microphone matrix (5 independent microphones)

Results

- FPR measurements:
  - analysing long audio recording with no keyword present
  - counting the number of falsely positive system activations (i.e. when the system has wrongfully detected the keyword)
  - FPR = ratio of the number of false alarms to the number of all analysed audio frames

Design	FPR [%]
Reference system (only KWS)	**2.23**
Loudness checking [LC]	0.90
LC + Timer [T]	0.64
Score smoothing (avg. last 3) [SS3]	1.43
Score smoothing (avg. last 4) [SS4]	1.30
LC + T + SS3	0
LC + T + SS4	0

Results

- Performance of the best set-ups — analysing audio samples recorded live from 13 users:
  - 30 repetitions of the keyword
  - 10 other words 3 times each (both phonetically similar and vey different from the keyword)

Design	TPR [%]	FPR [%]	Acc. [%]
Reference system - KWS only	90.77	5.90	92.44
KWS + SS3	**86.41**	**4.87**	**90.77**
KWS + SS4	84.10	4.87	89.62

Conclusion

- Proposed Access Control Decision Support System allows to decrease FPR to an acceptable level while retaining high TPR:
  - overall accuracy above 90%
  - the proposed solution allowed to entirely suppress false alarms caused by background radio voices, while the reference set-up generated approx. 122 unwanted activations per 5471 analysed frames
- The proposed design is computationally lightweight — works on an embedded device => <u>commercially applicable Access Control DSS</u>

GIDO LABS

ELARS

IARIA

Thank you for your attention!

The presented research has been supported by the National Centre for Research and Development in Poland under the grant no. POIR.01.01.01-00-0044/17

The National Centre for Research and Development