Decision Support System for Controlling Home Automation Appliances with Resource Constraints

Agnieszka Bętkowska Cavalcante, Monika Grajzer, Michał Raszewski

Monika Grajzer
Gido Labs
m.grajzer@gidolabs.eu

eKNOW 2022
The Fourteenth International Conference on Information, Process, and Knowledge Management
June 26 - 30th, 2022 - Porto, Portugal

Resume of presenter

- Co-owner of Gido Labs
- M.Sc. degree in Telecommunications with honours and a Ph.D. degree (2016) from Poznań University of Technology, Poland — won the “Summa cum Laude” prize for outstanding graduates of PUT
- 10+ years of experience in designing applied research solutions

- Developed AI/ML/Deep Learning solutions for several applied research projects, including the ones realised in the international teams, with top European IT companies
- Monika’s research resulted in the provisional patent application, IETF Internet Draft, several research papers (including those in top class journals) and project deliverables
- Current research interests in IoT networking and AI for human-machine interfaces, voice biometrics, eco innovations, smart buildings

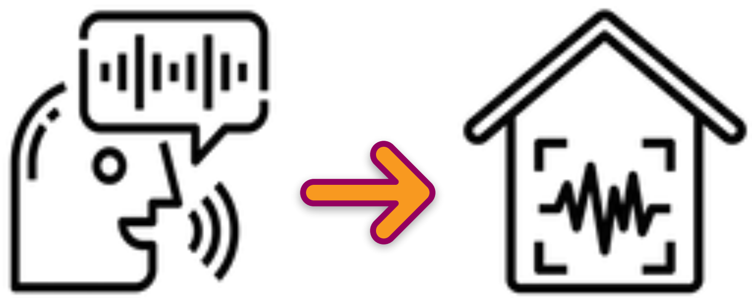
https://www.linkedin.com/in/monikagrajzer/

Outline

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

Scope of the paper

- Decision Support System (DSS) for **controlling access** to **embedded home automation** devices, working <u>fully offline</u> and with <u>limited resources</u>.



- Goal: identification of user voice commands and intentions based on which device's parameters can be set-up. This is accomplished by proposing a DSS that combines the knowledge from:
  - Keyword Spotting (KWS) — full ASR functionality is turned on after detecting a proper keyphrase
  - Speaker Recognition (SR) - voice biometrics to grant access to the system only to the known, authorized users
  - Automatic Speech Recognition (ASR) + related **Conversational Agent (CA)** — for enabling voice-based commands and short dialogs with a device (supported by speech synthesis)

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

Problem statement

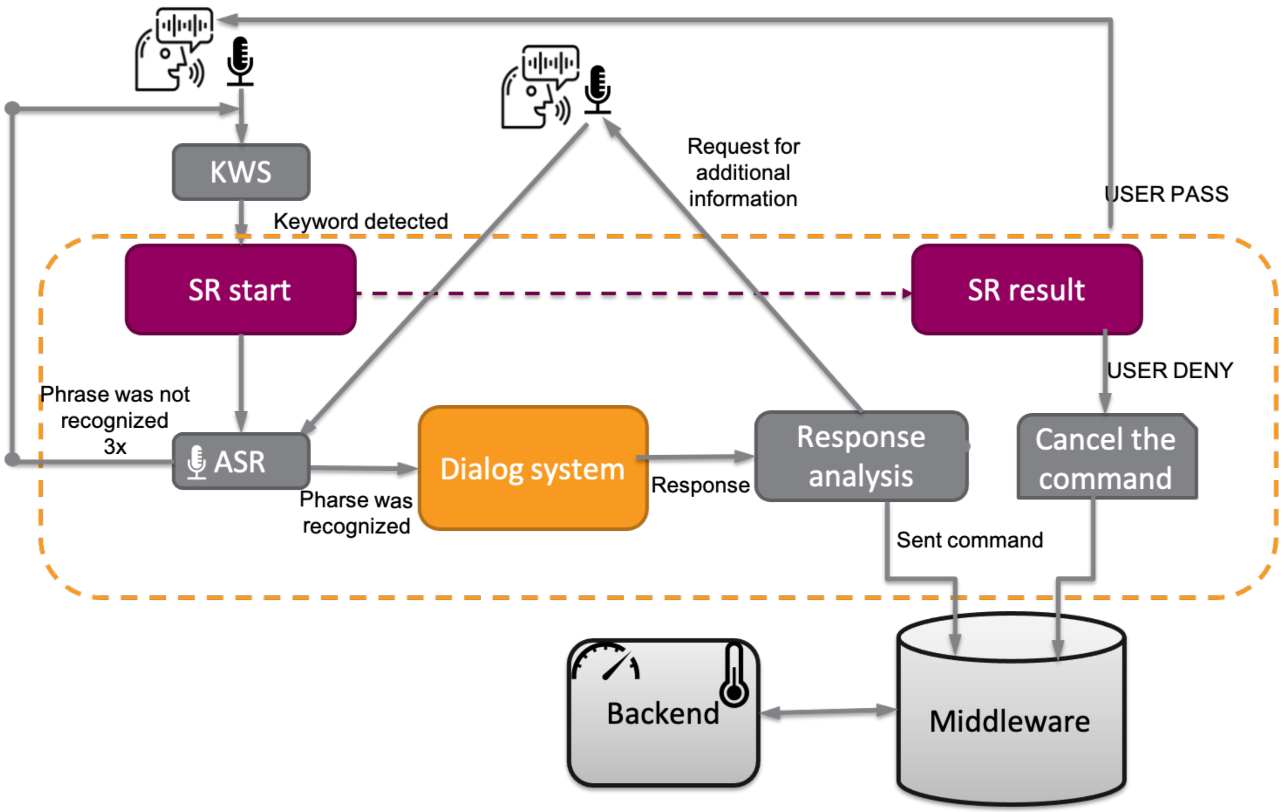
- Home automation systems with speech-based interfaces become increasingly popular.
- BUT: speech recognition is a resource-consuming task typically performed in the cloud => privacy concerns
- Offline systems working fully locally are desirable but <u>challenging on small embedded devices:</u>
    - require small resource acquisition
    - need to process audio input in real-time
    - should have low False Positive Rate (FPR) to avoid raising false alarms, granting unauthorised access
    - the number of unnecessary system activations (e.g., when someone is watching a TV) should be limited to increase performance of the ASR/CA module.
- Additional challenges:
    - support for non-English languages

Related work

- Current solutions for DSS system components are typically **cloud-based** since they require a **significant amount of resources** (like in Amazon Alexa)
- The top-performing solutions (e.g., neural networks for ASR and NLP) in an offline set-up are rare and are targeting devices with higher computational power (e.g., mobile phones) — our focus is on embedded devices (RPi and smaller)
- For access control (KWS + SR) lightweight Residual Neural Networks (ResNets) are used that can operate on embedded devices, but they require additional solutions to decrease FPR and allow for practical implementation:
  - State of the art KWS systems reach accuracy of 95% with False Positive Rate (FPR) of 2% **BUT** with this rate if a system makes prediction every second, there will be <u>~72 false alarms in an hour</u>.

Solution

- DSS system to locally control embedded devices, such as air conditioners, thermostats, and heating furnace.



Solution - operational details

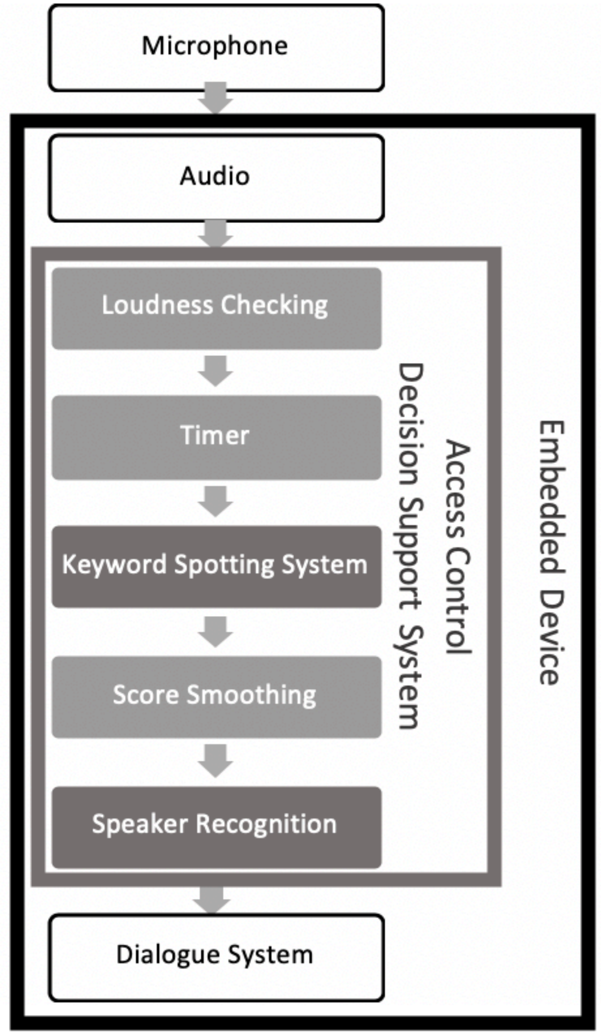
- Access control DSS constantly analyses the signal from the microphone and searches for a specific keyword
- Once that keyword is spotted — ASR starts listening. The command spoken by the user (e.g., "set the temperature in the living room to 5 degrees") is converted to text.
- The transcribed utterance is processed by the CA, which tries to understand the user's intent ("set the temperature") and assesses whether the input contains enough information.
  - If so, the DSS decides the type of command, it's parameters, and recipient device.
  - If not, the Dialog System will continue the conversation and ask the user for the missing information.
- The constructed technical command is sent to backend via a dedicated middleware.

Solution - operational details

- In parallel, the system authorizes a user with the SR module and the stored voice biometric patterns (the commands are executed only if the user's voice is recognized)
- SR is performed in the background, since the process can take a few seconds because of the limited computation capabilities of the targeted devices.
- All decision-making subsystems form a DSS, where at each step decisions are made based on the knowledge collected on the previous steps.

Solution - components

- Access control - KWS + supporting modules:
  - small ResNet (110k parameters) using 40 MFCCs as input, transfer learning with 698 positive examples of 36 people
  - System performance: accuracy of 90.77%, FPR of 4.87%
  - for long audio recording with no keywords present — false activations reduced from approx. 72/h to 0



Solution - components

- Speaker Recognition:
  - ResNet re-trained with a transfer learning on the dataset of 100 polish speakers - to increase accuracy for Polish language (can recognize both Polish and ENglish speakers)
  - enrollment: 10 repetitions of a custom phrase, approx. 1s each; recognition: custom phrase (approx. 1s long)
  - the new model + the proposed text-dependent system design, allowed to improve the Equal Error Rate for the identification of a single speaker from 9.83% to 1.7%

Solution - components

- Automatic speech recognition:
  - Based on CMU open source vocabulary, speaker- independent continuous speech recognition engine (HMMs)
    - For English, an acoustic model and phonetic dictionary were provided with pocketsphinx library + custom- prepared grammar to enable voice control tasks for home automation systems.
    - For the Polish language, a dedicated acoustic model was trained with recordings of 100 people, 49 males and 51 females + custom grammar
  - In field trial experiments on a Raspberry Pi platform, in office and home environments the word accuracy of the system was 97.98% for Polish and 94.77% for English.
- Dialog system/ conversational agent:
  - based on Bayesian-networks implementation from the openDial system enhanced with custom dialog models
  - In addition, the dialog management module was designed to make decisions regarding dialog states and flow –- asking user to repeat the sentence, ask about missing information, finish the dialog, etc.

Solution - components

- Middleware:
  - the commands identified by the logic of the DSS system are aquired and stored in a dedicated database and communicated to the actuators with a proper timing and order
- Hardware:
  - STM32MP157 microprocessor (based on an ARM-A7 architecture), 2 cores at 650 MHz, 1GB RAM, a 868 MHz radio to communicate with peripheral devices
  - OS: Embedded Linux based on OpenSTLinux and customised with the Yocto framework
  - 5 digital MEMS microphones placed on the PCB in a semicircle

Figure 2. Voice control embedded device – exterior view.

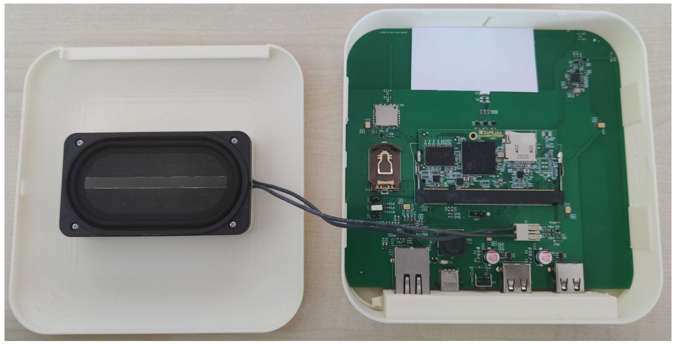


Figure 3. Voice control hardware platform – interior.

Evaluation

- Real-life scenarios
- Testers had to perform 27 assignments using the targeted device: tasks to set a desired configuration to the chosen home automation system or to collect data from it — the <u>creation of a final command was left to the user</u>
- Accuracy evaluated based on the number of positively completed tasks: in the first, 2nd or 3rd and none of the attempts
- A survey of the level of user satisfaction has been also performed: evaluating intuitiveness of the system and its subjective effectiveness
- 2 languages: Polish and English

Results

- Polish: 8 users, male and female
  - The average accuracy of task completion in the first attempt was 82.2% — that includes keyword spotting, the successful understanding of the dialog with the user, and correct user's voice verification
  - The percentage of correctly performed tasks in at most three attempts increased to 97.1%
  - intuitiveness: 8.8 out of 10, effectiveness: 8.4 out of 10

TABLE I. TASK COMPLETION ACCURACY FOR ACCESS CONTROL DEVICE – POLISH NATIVE SPEAKERS

User	Acc. of successfully completed tasks		Acc. of [%]	
	1st attempt	*2nd/3rd attempt*	*SR verification*	*failure*
user1	77.8%	14.8%	100%	7.4%
user2	88.9%	7.4%	90.5%	3.7%
user3	70.4%	29.6%	100%	0.0%
user4	85.2%	14.8%	81.8%	0.0%
user5	85.2%	14.8%	68.2%	0.0%
user6	96.3%	3.7%	100%	0.0%
user7	74.1%	22.2%	95.2%	3.7%
user8	80.0%	12.0%	84.2%	8.0%

TABLE II. SURVEY OF THE LEVEL OF USER SATISFACTION WITH ACCESS CONTROL DSS

user	Effectiveness	Intuitiveness
user1	9.5	9.0
user2	8.0	9.0
user3	8.0	10
user4	9.0	9.0
user5	7.0	6.0
user6	10	10
user7	9.0	8.0
user8	7.0	9.0

Results

- English: 6 users, male and female
  - due to COVID the users were not native speakers
  - average accuracy of task completion in a first attempt: 78.7%
  - performance of 94.3% was in at most 3 attempts.
  - On average, the testers evaluated intuitiveness of a DSS system as 8.5, and the effectiveness as 7.8 out of 10.
- Observations: some types of errors lowered the prototype performance:
  - the testers were using grammatically incorrect commands (as it happens in colloquial speech), or they were making mistakes and correcting themselves
  - the sequence of words spoken by the testers was very unique and did not fit into rules of the dialog system
  - the ASR system had problems with correctly recognising numbers when they were not spoken clearly (this is related to how grammar is being constructed)
  - the command was understood correctly by the DSS, but the user verification was not successful

Conclusion

- We have described the DSS system for voice-controlled home automation devices running on embedded platforms with limited resources
- Field trials were conducted with a device prototype — the testers freely used natural language to convey their commands
- We have shown that with a tailored design, the voice-controlled interface can achieve performance levels, which are sufficient to properly control the home automation device.

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