

Extension of Multimodality in Speech Assistance Systems for People with Impairments

Vortrag zur Masterarbeit von Tim Dilger

26.07.2022 – 14:00 Uhr

Zoom: <https://uni-bielefeld.zoom.us/j/98775157500?pwd=VU15NDhYU1hoYW9acnQzNytpSWJWdz09>

Über Mich

Name: Tim Dilger, B.Sc. Kognitive Informatik

Studiengang: Intelligente Systeme, MA

✉ t.dilger@uni-bielefeld.de



Projekt

Open-Source Software zur einfachen Bedienung "smarter"
Steuerungselemente in der Heimautomatisierung.

Sowohl per Sprachsteuerung als auch
über eine GUI *gleichzeitig* bedienbar

→ Bspw. Steuerung von Licht, Steckdosen, Musik, Rolläden etc.

Projekt

Für wen geeignet?

Ausgangspunkt: Das System soll auch bei leichten Defiziten, körperlichen sowie geistigen Beeinträchtigungen nutzbar sein.

→ Berücksichtigung von Barrieren

- Schriftbild
- Farbe
- Sprache
- ...

Zielsetzung

Was soll das System können?

- Vereinfachung des Alltags
im Umgang mit Smart Home Geräten
- Sprache verarbeiten und in einen Kontext bringen
- Kontextbasierte Benutzeroberfläche
zur Steuerung der Elemente
- Open-Source: System kann frei an eigene Anforderungen
angepasst werden

→ Hypothesen:

Ausführung durch Multimodalität schneller und unkomplizierter
Ausführung wird als angenehmer empfunden
im Vergleich zu unimodalen Systemen

Projekt

Motivation

- Interesse am Thema, eigene Smart-Home Projekte in Planung
- Multimodalität bietet die Möglichkeit zwischen Sprache und Touch-Steuerung zu wechseln bzw. interoperieren
- Gerade bei Personen mit Beeinträchtigungen können "smarte" Geräte im Haushalt den Alltag erleichtern

Research

Multimodalität

Sprachassistentz

Technologie

Design

Research

Multimodalität

- [END-TO-END MULTIMODAL SPEECH RECOGNITION](#), Palaskar et al., 2018
- [Multimodal Speech Recognition with Unstructured Audio Masking](#), Srinivasan et al., 2020
- [Multimodale Smartphone-Interaktion für Jung und Alt?](#), Dipl.-Ing. Radziwill, Univ.-Prof. Dr.-Ing. Schmidt, Kniewel, 2017

Research

Technologie

- "Base Cube One: A location-addressable service-oriented smart environment framework", Pohling, Leichsenring, Hermann, 2019 [BCO, 2019]
- "Evaluation of the Google Assistant as a Spoken Dialogue System for the Smart Home Framework Base Cube One", Huxohl, 2018
- Dokumentationsseiten der verwendeten Komponenten
- Suchmaschine + StackOverflow 100

Smart Home Komponenten

OpenHAB

BaseCubeOne

GraphQL



Open Source Smart Home Framework

Integration von verschiedenen
Smart-Home Technologien und Geräten

Sowohl offline als auch in der Cloud nutzbar

Unterscheidung der Komponenten in

- Things
- Channels
- Bindings
- Items
- Links

Smart Home Komponenten



Base Cube One

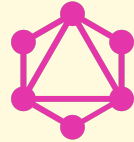
Open Source Smart Home Framework

Verwendet OpenHAB Binding
zur Zuordnung von *things/items* zu *Unit/Services*

Semantische Repräsentation von Möglichkeiten und Funktionen

Unabhängigkeit von Hard- und Software-Kompatibilität

Gute Performance für >500 *Units* und >1000 *Services*



GraphQL

```
query get_lights {  
  unitConfigs(filter: {  
    properties: {  
      unitType: COLORABLE_LIGHT  
    },  
    or: { properties: { unitType: LIGHT} }  
  }) {  
    id  
    labelString  
    placementConfig {  
      location {  
        id  
        labelString  
      }  
    }  
  }  
}
```

GraphQL API Query

Sprachassistentz

(für Menschen mit Einschränkungen)

- How Assistive Technology Has Changed the Lives of the blind and visually impaired, Hoskins & Prof. Zeeuw, 2014
- Using Automatic Speech Recognition to Assist Communication and Learning, Wald & Bain, 2005
- Assistive technology : principles and applications for communication disorders and special education, Wendt et al., 2011
- Assistive Technologies (genereller), Smeureanu & ISĂILĂ
- Assistive technology and science, Bodine, 2013

Speech Recognition Pipeline

Anforderungen

Pipeline

Sound Recording

NVIDIA NeMo Framework

Automatic Speech Recognition (ASR)

Text-to-Speech (TTS)

RASA NLU

Anforderungen

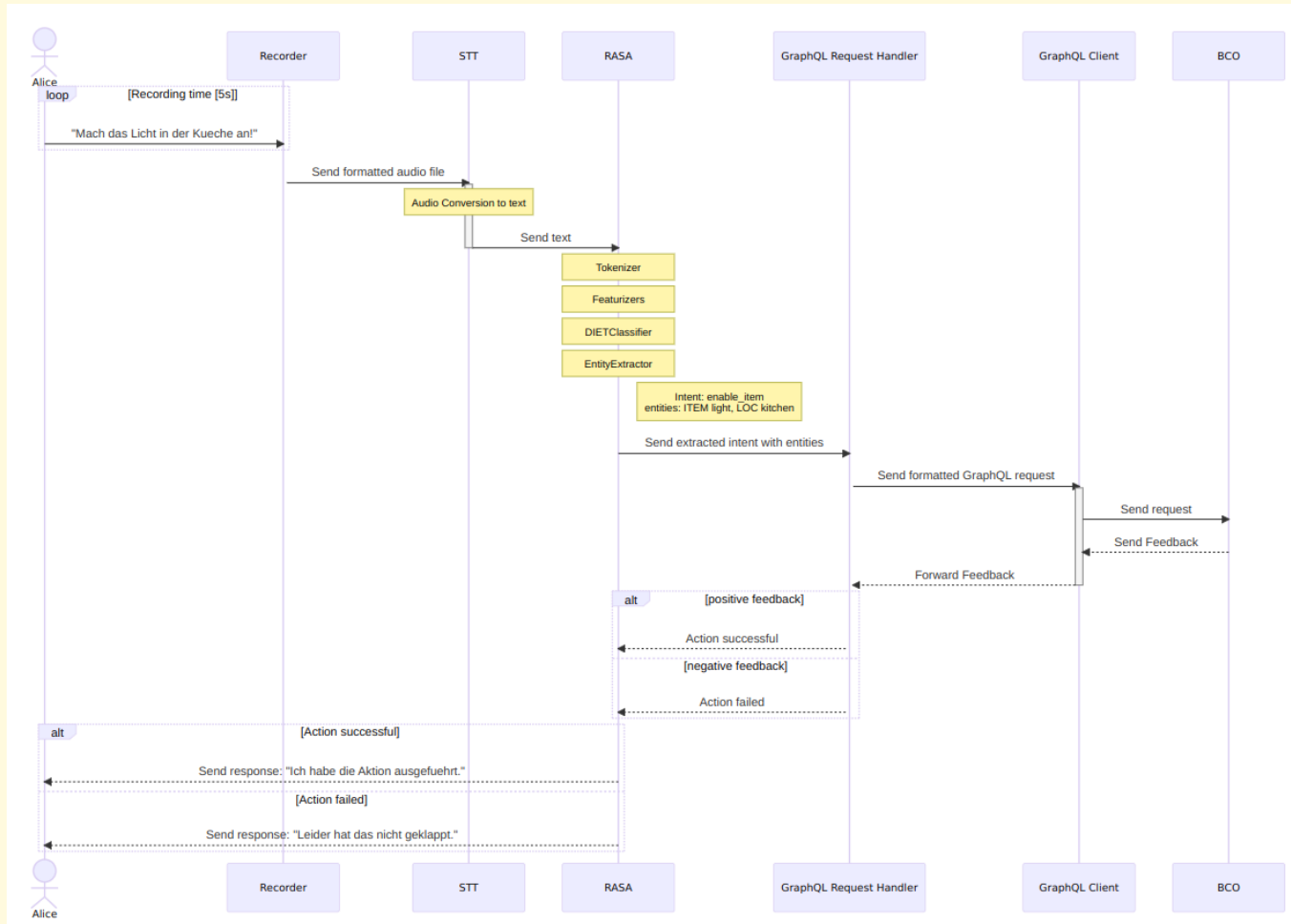
Table 1: Requirements for Speech Recognition Framework

Framework	Open Source License	Support for German language	Extensive Documentation	Up to date and ongoing maintaining	Python Language Embedding	Offline Access
MyCroft	Apache v2.0	✓	✓	✓	✓	Work-Around from private repositories
Kaldi and RASA	Apache v2.0	✓	✓	✓	✓	✓
Deepspeech	MPL-2.0 license	✓	✓	×	✓	✓
Leon	MIT	×	✓	✓	✓	✓
Python Speech Recognition	BSD	✓	✓	×	✓	Only CMU Sphinx and Snowboy
NVIDIA NeMo	Apache v2.0	✓	✓	✓	✓	✓

→ Mit der MyCroft-Komponente "**Mimic 3**" ist Ende Juni eine offline TTS-Lösung veröffentlicht worden.

Pipeline

UML Sequenzdiagramm: Konversation Licht anschalten



Speech Recognition Pipeline

Sound Recording

Pyaudio – Python Binding für PortAudio I/O library

Vorgaben von STT

- Einteilen in Chunks
- Sampleformat 16 bits pro Sample
- 1 Channel – Mono
- Samplerate 16.000 / s
- Wie lange aufnehmen? – Default 5s

Speech Recognition Pipeline

NVIDIA NeMo Framework

Apache v2.0 Lizenz

Modulare Sammlung an Speech Recognition Komponenten

Verwendet PyTorch ML Framework

Verwendung für vortrainierte Modelle in Automatic Speech Recognition (ASR) & Text-to-Speech (TTS)

Natural Language Understanding (NLU)



RASA Open Source

config	Aufbau der NLU-Pipeline
endpoints, credentials	Anbindung, z.B. Webserver
domain	intents, entities, responses, actions
nlu	forms, intents, lookups, synonyms
rules	Gleiche festgelegte Abfolge, z.B.: Greeting
stories	Gesprächsabfolge als NLU-Training

Custom Actions



Quelle: Rasa YouTube

Research

Design

- Design of Multimodal Mobile Interfaces, Shaked, Winter et al., 2016
- Gestaltungsempfehlungen für touchscreenbasierte Benutzungsschnittstellen, Domhardt, 2018
- ColorSpace [ColorSpace]
- Farbpsychologie

GUI


Prototyp-Entwicklung: Figma 

Frontend-Tool: Flutter 

- Icons:
- Google Material Font
 - Icon8 & Flaticon mit Attribution

Farben



Barrierefreiheit 

Studien

Setup

Use-Cases

Studienaufbau

Evaluation

Setup

Lampen

- An / Aus
- Dimmen
- Farbwechsel



Steckdosen

- An / Aus



Studien

Use-Cases

"Mach die Lampe an."

"Dimme alle Lampen im Wohnzimmer etwas heller"

"Ich möchte die Leuchte im WC blau."

"Schalte die Steckdosen im Bad aus."

Studien

Studienaufbau

Hypothesen:

Der Umgang mit Assistenzsystemen ist durch Multimodalität schneller und unkomplizierter möglich.

Die Auseinandersetzung mit multimodalen Systemen wird als angenehmer empfunden im Vergleich zu unimodalen Systemen

Empirische Untersuchung der Nützlichkeit

Evaluation über Fragebögen und nachträgliche Auswertung der Sprachassistentz

Testsubjekt	Forschungsmethode	Ablauf	Voraussetzungen	Evaluation
GUI-Prototyp	Quantitative und qualitative Erhebung	Beschreibung und Durchführung der einzelnen Use-Cases	Relevante Stichprobe, Größe eher zweitrangig	Einschätzung Zufriedenheit mit einzelnen Elementen und Gesamtheit, Farbschema, Typographie, Anordnung
GUI	Quantitative und qualitative Erhebung	Beschreibung und Durchführung der einzelnen Use-Cases	Möglichst große, relevante Stichprobe	Einschätzung Zufriedenheit mit einzelnen Elementen und Gesamtheit sowie Haptik
Unimodales System (Sprache / GUI)	Quantitative Erhebung	Aufgaben erfüllt? → Falls nein, woran hat es gelegen?	Möglichst große, relevante Stichprobe	Einschätzung Zufriedenheit*
Multimodales System (Sprache & GUI kontextbasiert)	Quantitative Erhebung	Aufgaben erfüllt? → Falls nein, woran hat es gelegen?	Möglichst große, relevante Stichprobe	Einschätzung Zufriedenheit** Schwierigkeiten bei der Durchführung

*GUI bzw. Sprachassistentz **GUI, Sprachassistentz & Interoperabilität

Lizenz

Apache License 2.0

A permissive license whose main conditions require preservation of copyright and license notices. Contributors provide an express grant of patent rights. Licensed works, modifications, and larger works may be distributed under different terms and without source code.

Permissions

- Commercial use
- Distribution
- Modification
- Patent use
- Private use

Conditions

- License and copyright notice
- State changes

Limitations

- Liability
- Trademark use
- Warranty

[View full Apache License 2.0 »](#)

Quelle: choosealicense.com

Weitere Zeitplanung

Erfahrungsaustausch Bethel – Projekt *Smart im Alltag*

Fertigstellung Deployment der Speech-Recognition-Pipeline

Erstellung Prototyp GUI

Test & Evaluation – GUI Prototyp

Test & Evaluation – Interaktion mit Sprachassistentz im
Entwicklungsprozess

Weitere Zeitplanung

Entwicklung der GUI

Test & Evaluation – GUI

Test Multimodalität – GUI + Sprachassistentz

Test Vergleich mit Google Home Sprachassistentz
(Rahmenbedingungen)

Evaluation

Während des Prozesses und im Anschluss:
Schreiben der Masterarbeit ✍️

Fragen & Anregungen

Vielen Dank für die Aufmerksamkeit!