

Sprachsteuerung für einen ROS-basierten Assistenzroboter

Mosawer Ahmad Nurzai
M-INF3

- Einführung
- Arbeiten
 - Projekt 1 – Sprachsteuerungsinterface
 - Projekt 2 – Unterstützende Mechanismen für die Spracherkennung
- Masterarbeit
 - Fragestellung/Ziele
 - Vorgehen
 - Risiken

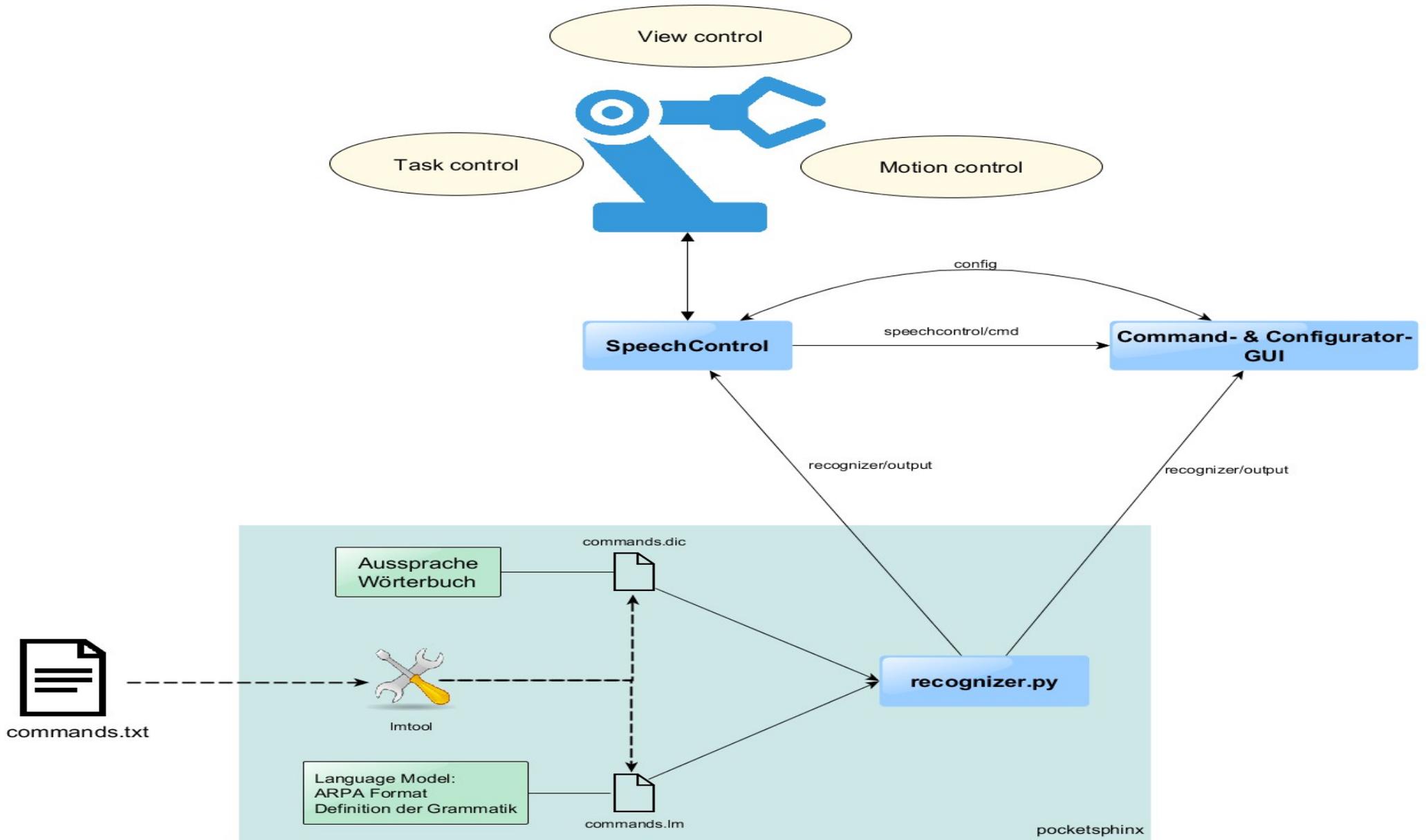
Motivation

- Sprachsteuerungsmodul für den ROS
- Alternative Steuerung im RV-Labor
 - Zurzeit nur PS3-Controller
- Spracherkennung/Sprachsteuerung in Bezug auf Assistenzroboter
- Human-Robot-Interaction
 - Kommunikation per Sprache



Abb. 1: Roboter im RV-Lab

- Einführung
- Arbeiten
 - **Projekt 1 – Sprachsteuerungsinterface**
 - Projekt 2 – Unterstützende Mechanismen für die Spracherkennung
- Masterarbeit
 - Fragestellung/Ziele
 - Vorgehen
 - Risiken



Ein Sprachsteuerungsinterface für einen ROS-basierten Assistenzroboter

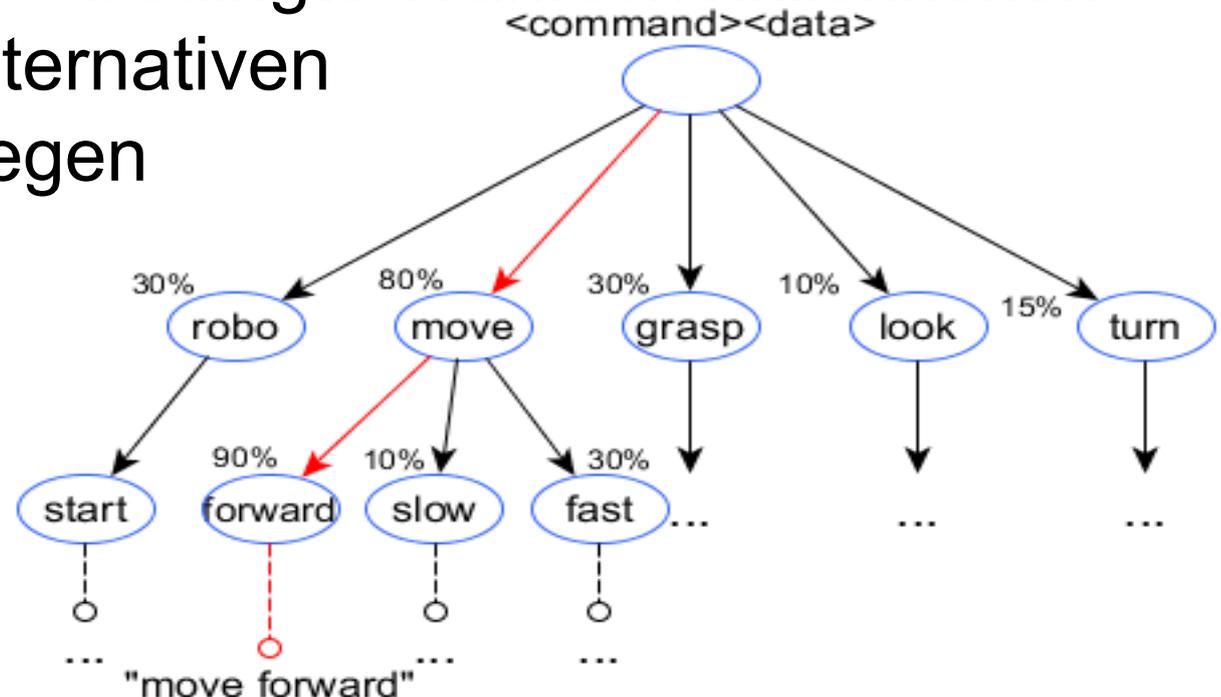
Evaluation

- Spracherkennung (pocketsphinx)
 - Erkennungsrate?
 - Schwer erkennbare Wörter?
- Sprachsteuerung
 - Befehl → Ausführung?

Ein Sprachsteuerungsinterface für einen ROS-basierten Assistenzroboter

Weitere Schritte

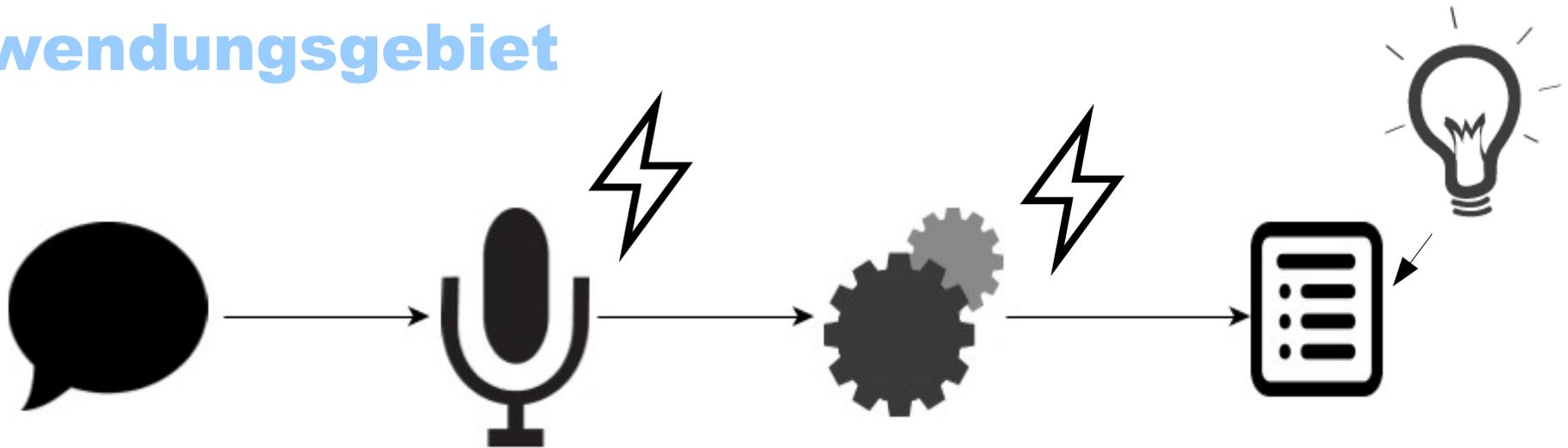
- Pocketsphinx output → bedingte Wahrscheinlichkeiten
 - Nicht machbar → Alternativen
- Randbedingung festlegen
 - Welches Mikrofon?
 - Abstand Mikrofon?



- Einführung
- Arbeiten
 - Projekt 1 – Sprachsteuerungsinterface
 - **Projekt 2 – Unterstützende Mechanismen für die Spracherkennung**
- Masterarbeit
 - Fragestellung/Ziele
 - Vorgehen
 - Risiken

Unterstützende Mechanismen für die Spracherkennung

Anwendungsgebiet



Warum?

- Unzuverlässige Spracherkennung
 - Undeutliches Sprechen
 - Laute Umgebung

Unterstützende Mechanismen für die Spracherkennung

Was gibt es?

- Bildverarbeitung
 - Lippenbewegung
Erfolgsrate 84.7% [Yka06]
 - Gesten [Dmrd11]
- Sensorik
 - Elektroglottograph (EGG) [DiSch95]
Verbesserung 94.37% → 99.37%
 - Gesichts-Elektromyographie (EMG) [LoLa10]
Erfolgsrate 70%

Unterstützende Mechanismen für die Spracherkennung

Lippenbewegung

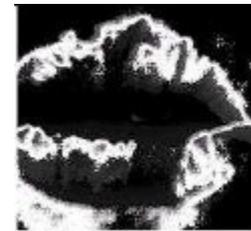
- Konsonanten [Yka06]
- Nicht Erkennung von ganzen Wörtern
- Problematik
 - Frontsicht oder seitlich
 - Performance Bildverarbeitung

Unterstützende Mechanismen für die Spracherkennung

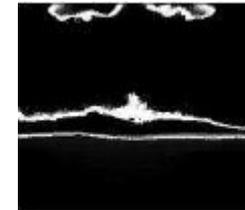


■ Phonem

- /v/ → van, move, voice
- /m/ → man, name, make
- /g/ → game, ago, get



/v/



/m/



/g/ [Yka06]

- Motion history image → Neuronales Netzwerk als Klassifikator

- Einführung
- Arbeiten
 - Projekt 1 – Sprachsteuerungsinterface
 - Projekt 2 – Unterstützende Mechanismen für die Spracherkennung
- **Masterarbeit**
 - Fragestellung/Ziele
 - Vorgehen
 - Risiken

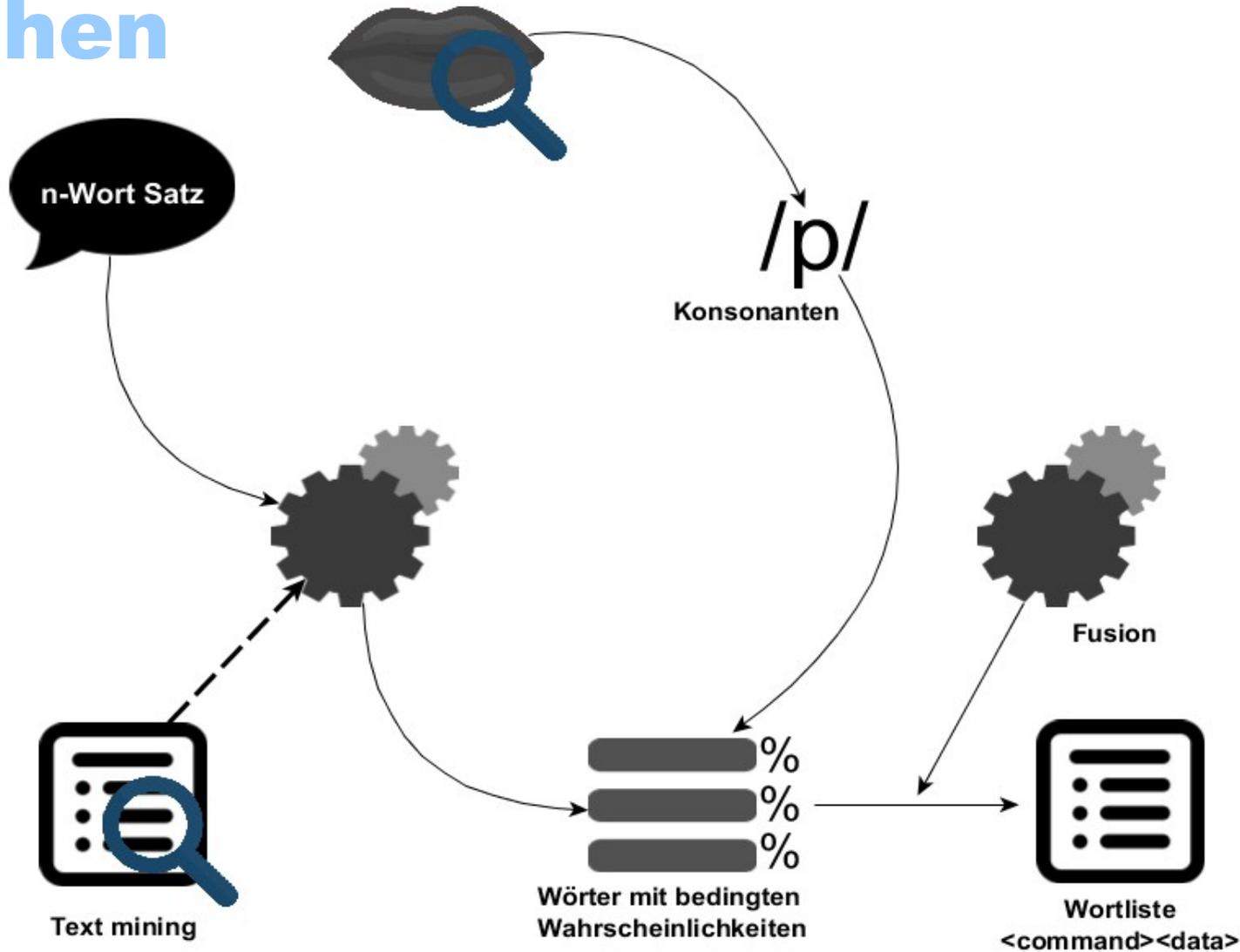
- Weitere Vorgehensweise?
 - n-Wort Sätze
 - Rahmenbedingung festlegen
 - PJ1 + PJ2

Fragestellung / Ziele

- Ziele
 - Steuerung des Assistenzroboter per On-board-Mikrofon
 - Sprachsteuerungssoftware: Lippenlesen + Spracherkennung
- Fragestellung
 - Welche Störfaktoren gibt es? Bezug: Mikrofon
 - Wie n-Wort Sätze filtern?
 - Wie bedingte Wahrscheinlichkeiten (Spracherkennung) mit Konsonanten (Lippenlesen) kombinieren?

- Einführung
- Arbeiten
 - Projekt 1 – Sprachsteuerungsinterface
 - Projekt 2 – Unterstützende Mechanismen für die Spracherkennung
- Masterarbeit
 - Fragestellung/Ziele
 - **Vorgehen**
 - Risiken

Vorgehen



Vorgehen

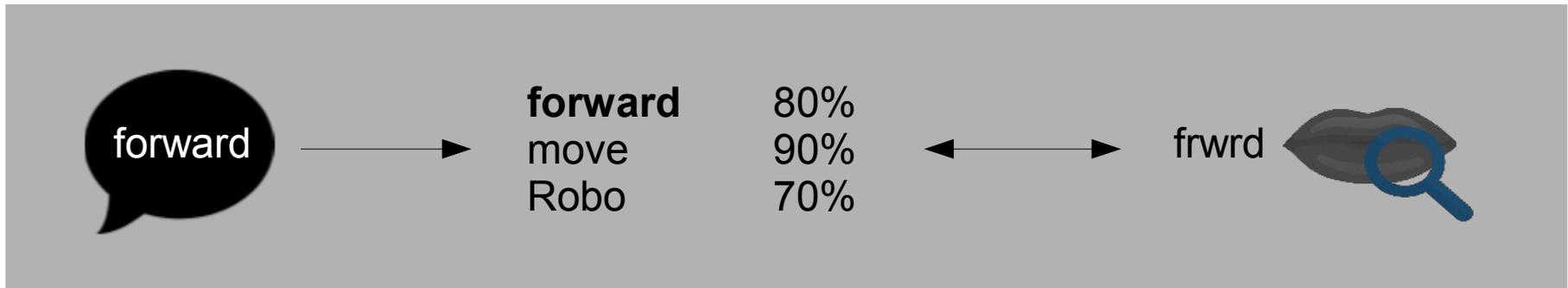
- Implementierungen
 - 1. Output bedingte Wahrscheinlichkeiten
 - 2. Text mining der Befehle
 - 3. Fusion Lippenlesen und Spracherkennung
- Rahmenbedingung festlegen
 - Abstand zur Kamera
 - Abstand zum Mikrofon
 - Mikrofon?
- Testläufe
 - Erkennungsrate? Steigerung?
 - Performance Text mining
 - Noch Zeit → Grenzen des Mikrofons

Vorgehen – Erste Ideen

- Implementierung 1 Output bedingte Wahrscheinlichkeiten
 - Pocketsphinx API studieren
 - ROS package erweitern
- Implementierung 2 Text mining der Befehle

„Hello Robo can you please move forward“ → „Robo move forward“

- Implementierung 3 Fusion Lippenlesen und Spracherkennung



Risiken

- Machbarkeit? → Mikrofon, Kamera
- Kein output von bedingten Wahrscheinlichkeiten
- Performance
 - Bildverarbeitungsmethode
 - Kombination Spracherkennung + Text mining + Bildverarbeitung

Plan

Phase 1

Projekt 1

Sprachsteuerungsinterface

- ROS
- Alternative Steuerung
- Erste Berührung Spracherkennung
- Software für spätere Weiterentwicklung

Phase 2

Projekt 2

Spracherkennung unterstützen

- Bildverarbeitung
- Lippenlesen
- Performanter Output

Phase 3

Master-Thesis

Durch Bildverarbeitung unterstützende Spracherkennung

- Fusion PJ1 und PJ2
- Text mining
- Evaluation Spracherkennung/steuerung und Performance

Vielen Dank für eure Aufmerksamkeit!

Paper

- **[Yka06]** Voiceless Speech Recognition Using Dynamic Visual Speech Features – Yau et al. 2007
- **[Dmrd11]** An Exploration of Gesture-speech Multimodal Patterns for Touch Interfaces – Dey et al. 2011
- **[DiSch95]** Electroglottograph as an additional source of information in isolated word recognition – Dikshit and Schubert 1995
- **[LoLa10]** Syllable-based speech recognition using EMG – Lopez-Larraz et al. 2010

Internetseiten

- CMU Sphinx – <http://cmusphinx.sourceforge.net/>
- Pocketsphinx (ROS) – <http://wiki.ros.org/pocketsphinx>
- ROS – <http://www.ros.org/>