



Hochschule für Angewandte Wissenschaften Hamburg
Hamburg University of Applied Sciences

Ausarbeitung - Master Seminar

Mosawer Ahmad Nurzai

Sprachsteuerung für einen ROS-basierten Assistenzroboter

*Fakultät Technik und Informatik
Studiendepartment Informatik*

*Faculty of Engineering and Computer Science
Department of Computer Science*

Mosawer Ahmad Nurzai

Sprachsteuerung für einen ROS-basierten Assistenzroboter

Ausarbeitung - Master Seminar eingereicht im Rahmen von Seminar

im Studiengang Master of Science Informatik
am Department Informatik
der Fakultät Technik und Informatik
der Hochschule für Angewandte Wissenschaften Hamburg

Gutachter: Prof. Dr. Kai von Luck

Gutachter: Prof. Dr. Bettina Buth

Betreuer: Prof. Dr. Andreas Meisel

Eingereicht am: 28. Februar 2015

Mosawer Ahmad Nurzai

Thema der Arbeit

Sprachsteuerung für einen ROS-basierten Assistenzroboter

Stichworte

Sprachsteuerung, Spracherkennung, Roboter, Assistenzroboter, bild unterstützend, ROS, Konsonant, Text mining

Mosawer Ahmad Nurzai

Title of the paper

Speech control for a ROS-based assistant robot

Keywords

Speech control, speech recognition, roboter, assistant robot, support by image, ROS, consonant, text mining

Inhaltsverzeichnis

1	Einführung	5
1.1	Aufbau der Arbeit	5
1.2	Ziel/Motivation	5
2	Eigene Arbeiten	5
2.1	Projekt 1 - Sprachsteuerungsinterface	6
2.1.1	Ziel	6
2.1.2	Vorgehen	6
2.1.3	Ausstehende Ergebnisse	8
2.2	Projekt 2 - Unterstützende Mechanismen für die Spracherkennung	9
2.2.1	Ziel	10
2.2.2	Vorgehen	10
3	Master-Thesis	11
3.1	Ziele	11
3.2	Fragestellung	11
3.3	Vorgehen	13
3.3.1	Testläufe	14
3.3.2	Erste Ideen	14
3.4	Risiken	15
4	Schluss	16
4.1	Zusammenfassung	16
	Literaturverzeichnis	17
	Abbildungsverzeichnis	18

1 Einführung

In diesem Kapitel wird der Aufbau der Ausarbeitung präsentiert, außerdem erfolgt eine Neumotivation der Arbeit. Die Ausarbeitung bzw. die praktische Umsetzung erfolgt an der HAW Hamburg in der Forschungsgruppe „Robot Vision“.

1.1 Aufbau der Arbeit

Die Arbeit ist in vier Kapitel gegliedert. Das erste (1) Kapitel ist eine Einführung zu dieser Arbeit und beschreibt die Motivation der Arbeit. Der Hauptteil dieser Arbeit stellen das zweite (2) und das dritte (3) Kapitel dar. Im zweiten (2) Kapitel werden die eigenen Arbeiten vorgestellt, die aufbauend für die Master-Thesis verwendet werden sollen. Das dritte (3) Kapitel beschreibt die Master-Thesis. In dem Kapitel sollen die Ziele, das Vorgehen und die Risiken präsentiert werden. Das vierte letzte (4) Kapitel stellt den Schluß dar. In dem Kapitel wird die Arbeit zusammengefasst.

1.2 Ziel/Motivation

Es soll eine alternative Steuerung im Robot Vision Labor implementiert werden. Zurzeit wird für die Steuerung des ROS¹-basierten Assistenzroboter (Quigley u. a. [2009]) ein Spiele-Controller verwendet. Hierbei soll eine Spracherkennung bzw. Sprachsteuerung verwendet werden. Um einen Assistenzroboter Anweisungen zu geben, denkt man häufig an die Kommunikation per Sprache. Die Sprache ist die Hauptkommunikationsart des Menschen, da diese am einfachsten und schnellsten eine Botschaft/Nachricht/Irgendetwas übermitteln kann (Goodrich und Schultz [2007]).

2 Eigene Arbeiten

In diesem Kapitel werden die eigenen Arbeiten vorgestellt, die aufbauend genutzt werden, um die Master-Thesis vorzubereiten. Hierbei werden die Erkenntnisse und Ergebnisse der Master-Veranstaltungen Projekt 1 und Projekt 2 verwendet.

Es werden die Projekte vorgestellt, in dem die Ziele dieser Projekte präsentiert werden. Das Vorgehen und die erhofften Ergebnisse werden ebenfalls aufgezeigt.

¹ROS (Robot Operating System) - spezielles Roboterbetriebssystem (eher Software-Framework) - <http://www.ros.org>

2.1 Projekt 1 - Sprachsteuerungsinterface

Im Projekt 1 wird ein Sprachsteuerungsinterface für einen ROS-basierten Assistenzroboter implementiert.

Das Sprachsteuerungsinterface wird implementiert für den Roboter im Robot-Vision Labor (siehe Abbildung 1).

2.1.1 Ziel

Die implementierte Anwendung wird als eine Grundlage dienen für spätere Anwendungen im Rahmen von Projekt 2 bzw. der Master-Thesis. Eines der Ziele dieses Projektes ist die Einarbeitung in das ROS-Betriebssystem und eine interne Einführung in die Spracherkennung/Sprachsteuerung.

2.1.2 Vorgehen

Das Sprachsteuerungsinterface wird Anfangs nur simple Sprachbefehle annehmen bzw. bearbeiten können, da zunächst der Fokus auf die Steuerung, ROS und die Spracherkennung gelegt wird.



Abbildung 1: Roboter im Robot Vision Labor

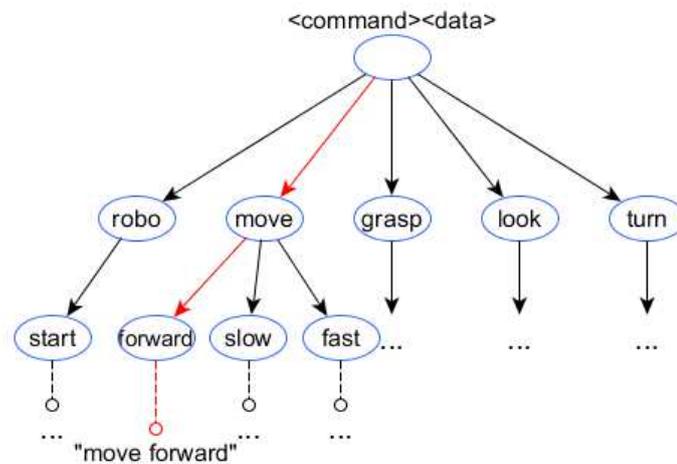


Abbildung 2: Aufbau - Simple Sprachbefehle

In Abbildung 2 ist der Aufbau der simplen Sprachbefehle zu sehen. Die Grammatik besteht aus einem „command“ und einem „data“. Dabei bilden diese beiden Wörter ein „order“. In der Abbildung ist der Pfad des Befehls „move forward“ abgebildet.

$$\langle order \rangle = \langle command \rangle \langle data \rangle \quad (1)$$

$$\langle command \rangle = \{ \langle robo \rangle, \langle move \rangle, \langle grasp \rangle, \langle look \rangle, \langle turn \rangle \} \quad (2)$$

$$\langle forward \rangle = \{ \langle forward \rangle, \langle ahead \rangle, \langle straight \rangle \} \Rightarrow 'forward' \quad (3)$$

In der obigen Formel sieht man die Grammatik der simplen Sprachbefehle. Zeile 1 besteht aus dem Aufbau der Grammatik. In Zeile 2 sind die möglichen Befehle, die der Roboter ausführen soll, aufgelistet und in Zeile 3 ist ein Datum zu sehen. In diesem Beispiel „forward“ - sind Wörter aufgelistet, die gesagt werden können und die selbe Bedeutung (nämlich „forward“) haben.

Die Anwendung, die in diesem Kapitel vorgestellt wird, benutzt für die Spracherkennung CMU Sphinx - Pocketsphinx, kurz Pocketsphinx. Pocketsphinx ist eine lightweight Spracherkennungseingine, welches speziell für mobile Geräte bzw. Geräte mit weniger Ressourcen gedacht ist. Die heavyweight Version ist das CMU Sphinx (University [2015]).

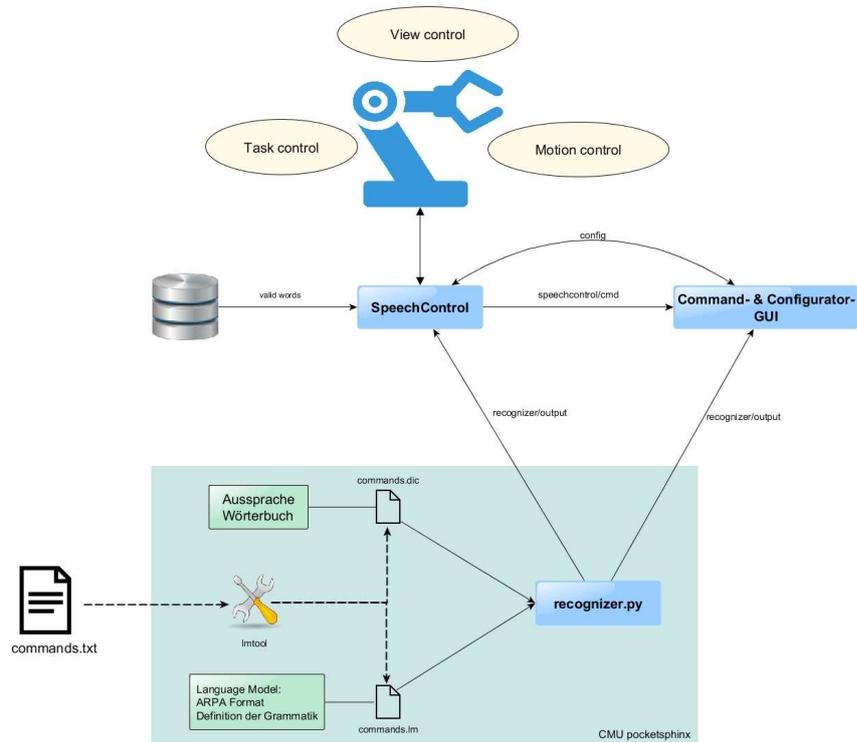


Abbildung 3: Architektur von ROS-Speech-Control-Interface

In der Abbildung 3 sieht man das Design bzw. die Architektur der Anwendung ROS-Speech-Control-Interface (RSCI). Die Anwendung besteht aus drei Hauptkomponenten bzw. Unteranwendungen. Die drei Komponenten sind „SpeechControl“, „Command- & Configurator-GUI“ und „CMU pocketsphinx“, dabei haben Teile dieser Anwendung Zugriff auf den Roboter (1), um ihn steuern zu können und einer Datenbank, um gültige Wörter zu erhalten. Die Hauptkomponenten kommunizieren untereinander, um Informationen auszutauschen und ggf. Konfigurationsmöglichkeiten anzubieten.

- **SpeechControl** Schickt die eintreffenden Befehle als Kommandos weiter an den Roboter.
- **Command- & Configurator-GUI** Eine GUI um Konfigurationsparameter einstellen zu können und aktive Kommandos anzuzeigen.
- **CMU pocketsphinx** Spracherkennung.

2.1.3 Ausstehende Ergebnisse

Es sind ausstehende Forschungen noch zu betreiben. Hierbei muss die Spracherkennung (pocketsphinx) evaluiert werden, um besser beurteilen zu können, ob die Spracherkennung

weiterhin benutzt werden kann. Es soll die Erkennungsrate gemessen werden und schwer erkennbare Wörter erkannt werden. Um die Erkennungsrate bestimmen zu können, werden in einem Versuch mehrere Wörter aufgesagt und deren bestimmte Wahrscheinlichkeiten sollen ausgelesen werden. Aus diesem Versuch wird eine Statistik aufgestellt, um ggf. neue Erkenntnisse zu erhalten.

2.2 Projekt 2 - Unterstützende Mechanismen für die Spracherkennung

In Projekt 2 soll eine Anwendung entwickelt werden, um die Spracherkennung zu unterstützen. Dies ist erforderlich, da die Spracherkennung, unter bestimmten Bedingungen, unzuverlässig sein kann. Ein undeutliches Sprechen oder eine laute Umgebung fördert beispielsweise, dass die Spracherkennung unzuverlässig wird.

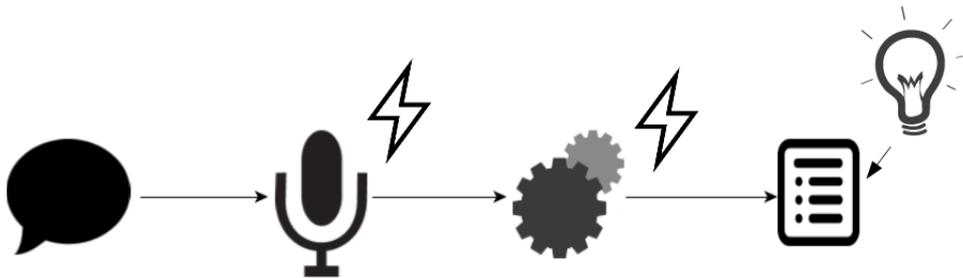


Abbildung 4: Spracherkennung Workflow - Anwendungsgebiet

Die Unterstützung kann in verschiedenen Etappen oder Prozessschritten der Spracherkennung statt finden. In Abbildung 4 sieht man abstrakt die Prozessschritte einer Spracherkennung.

- **Sprechblase:** Zunächst ist es erforderlich herauszubekommen, wie am besten ein Befehl gesagt werden soll. Dies soll in der Master-Thesis behandelt werden (3.3).
- **Mikrofon:** Das Mikrofon soll die Signalverarbeitung symbolisieren. In diesem Gebiet wird keine Unterstützung statt finden.
- **Zahnräder:** Es wird kein Algorithmus entwickelt, um Wörter aus erkannten Tönen zu bestimmen.
- **Liste:** Die Unterstützung soll in einer hohen Ebene statt finden, sprich wie sollen am besten antreffende Wörter verarbeitet werden und sollen Wörter verworfen werden?

2.2.1 Ziel

Das Ziel in diesem Projekt ist es per Bildverarbeitung die Spracherkennung zu unterstützen. Dabei geht es in diesem Projekt nur um die Bildverarbeitung und nicht um das Zusammenführen der Spracherkennung und der Bildverarbeitung.

2.2.2 Vorgehen

Im Projekt 2 wird das Lippenlesen behandelt. Hierbei wird, per Bildverarbeitung, die Bewegungen der Lippen gelesen. Dies wird in der Master-Thesis unterstützend für die Spracherkennung eingesetzt werden. Als Beispiel bzw. Vorbild für dieses Projekt wurde die Arbeit Yau u. a. [2006] genommen.

In der Arbeit werden Konsonanten, sprich Mitlaute erkannt. Es wird nicht versucht ganze Wörter zu erkennen, da die Bildverarbeitung hohe Performance benötigt. Eine weitere Problematik in diesem Feld ist, dass das Lippenlesen aus der Front- und Seitenperspektive geschehen kann. Hierbei gelten unterschiedliche Anforderungen, die zu erforschen sind.

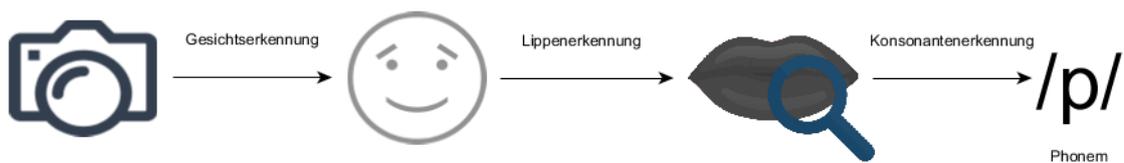


Abbildung 5: Lippenlesen Workflow

In Abbildung 5 ist exemplarisch ein Workflow bei einer Lippenlesung zu sehen. Zunächst muss das Gesicht erkannt werden. Nach der Gesichtserkennung erfolgt die Lippenerkennung. Nach dem der Bereich erkannt wurde, wo sich die Lippen befinden, erfolgt die Erkennung von Konsonanten, in dem sogenannte Phoneme erkannt werden. Ein Phonem ist die „kleinste bedeutungsunterscheidende sprachliche Einheit“ (Duden [2015]).

- /v/ : **v**an, **m**ove, **v**oice
- /m/ : **m**an, **n**ame, **m**ake
- /g/ : **g**ame, **a**go, **g**et

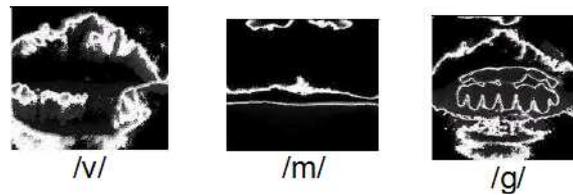


Abbildung 6: Motion History Image - Phoneme

Quelle: Yau u. a. [2006]

Um die Phoneme unterscheiden zu können, werden sogenannte „Motion History Images“ (MHI) eingesetzt. Durch die MHIs werden Abfolgen von einzelnen Phonemen abgebildet, hierbei besteht ein MHI aus zusammengesetzten Frames. Die Frames bilden dabei die Bewegungsabfolgen eines Phonems. In Abbildung 6 sind einige MHIs abgebildet. Beim Prozess (5) des Lippenlesen, können am Ende der Kette die Phoneme, beispielsweise durch ein neuronales Netzwerk, welches als Klassifikator genutzt werden kann, erkannt werden.

3 Master-Thesis

In diesem Kapitel wird das Ziel, das Vorgehen und die Risiken der Master-Thesis vorgestellt. Es erfolgt ein Bezug zum vorherigen Kapitel (2), da Projekt 1 und Projekt 2 aufbauend genutzt werden.

3.1 Ziele

In der Master-Thesis wird angestrebt eine Sprachsteuerungssoftware zu entwickeln, dabei werden die Ergebnisse und Erkenntnisse aus Projekt 1 und 2 verwendet und weiter benutzt. Konkret wird die Spracherkennung durch die Bildverarbeitung - das Lippenlesen - unterstützt. Als zusätzliches Ziel ist es, die Steuerung des Assistenzroboter per On-board-Mikrofon zu implementieren. Zurzeit wird ein Headset benutzt.

3.2 Fragestellung

Es bilden sich einige Fragestellungen in diesem Zusammenhang. Es sollen Störfaktoren, in Bezug auf das (On-board-)Mikrofon, ermittelt werden. Eventuell können die Störfaktoren (mechanische Geräusche, Reichweite) zu groß sein und es ist erforderlich, dass ein Headset benutzt werden muss.

In Projekt 1 werden Befehle entgegen genommen und an den Roboter weitergesendet. Diese Befehle sind simple zwei-Wort-Befehle. In der Master-Thesis wird zusätzlich der Forschungs-

bereich Text-Mining betrachtet, hierbei soll eine komplexere Grammatik erkannt werden. Die Fragestellung ist wie können längere Befehle, Sätze gefiltert werden?

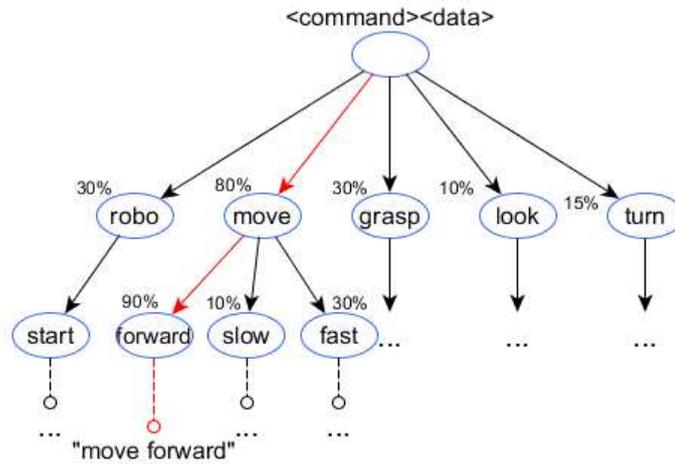


Abbildung 7: Aufbau - Simple Sprachbefehle mit bedingten Wahrscheinlichkeiten

Die Resultate aus Projekt 1 und Projekt 2 sollen miteinander kombiniert werden. Folgende Frage stellt sich: Wie können bedingte Wahrscheinlichkeiten (7) (Spracherkennung) mit den Konsonanten (Lippenlesen) kombiniert werden? In nächsten Abschnitt wird dies weiter erläutert.

3.3 Vorgehen

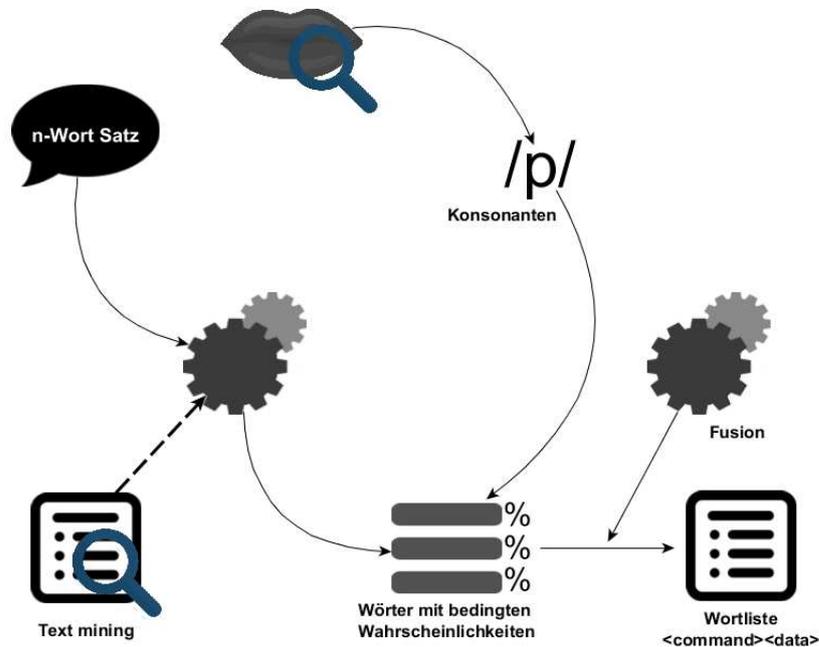


Abbildung 8: Vorgehen - Prozessflow

In der Abbildung 8 wird das gesamte Vorgehen bzw. Prozessflow der Master-Thesis gezeigt. Aus der Abbildung können einige Baustellen erkannt werden.

Zunächst ist es erforderlich, das Spracherkennungstool (Pocketsphinx) zu erweitern. Als Ausgabe liefert das Spracherkennungstool, das erkannte Wort. Für die Anwendung ist es jedoch erforderlich, als Ausgabe, eine Liste mit den erkannten Wörtern und deren bedingten Wahrscheinlichkeiten, zu erhalten (7).

Eine weitere Baustelle ist das Textmining von auftreffenden Sätzen/Befehlen. Für die Master-Thesis ist vorgesehen, dass Befehle, an den Roboter, in Form von normalen Sätzen, statt simplen zwei Wörtern, getätigt werden sollen. Die Sätze müssen gefiltert werden, hierbei sollen wichtige Wörter, die für den Befehl wichtig sind, erkannt werden und Stoppwörter müssen gelöscht werden. Dieses muss N-mal gemacht werden, da bei einem Befehl (Satz), durch die bedingten Wahrscheinlichkeiten, mehrere Sätze entstehen können. Es muss jedoch eine Grenze bzw. ein Limit geben, ab wann Wörter, für das Textmining genommen werden können, dabei dienen die bedingten Wahrscheinlichkeiten der jeweiligen Wörter als Grenze.

Im letzten Prozessschritt erfolgt eine Fusion der Wörter mit den bedingten Wahrscheinlichkei-

ten und der erkannten Konsonanten. Es erfolgt ein Abgleich der Wörter und Konsonanten. Gibt es in einem Satz mehrere Befehle, so erfolgt die Abarbeitung nach Aufkommen.

Für die Applikationen ist es erforderlich verschiedene Rahmenbedingungen festzulegen, um auf Basis dieser, Thesen und Versuche aufstellen zu können. Es muss festgelegt werden, wie der minimale und maximale Abstand zur Kamera sein muss, damit die Lippen erkannt werden können. Der Abstand zum Mikrofon muss ebenfalls festgelegt werden. Zusätzlich muss festgestellt werden, ob ein Headset oder ein On-board-Mikrofon genommen werden soll.

3.3.1 Testläufe

Um die Verbesserung der bildunterstützten Spracherkennung zu evaluieren, werden verschiedene Tests durchgeführt. Hierbei soll herausgefunden werden, ob durch die Kombination der beiden Ansätze, die Erkennungsrate gesteigert wurde. Zuerst werden Befehle ohne Lippenlesen aufgesagt und dessen Erkennungsrate bestimmt. Zum Vergleich wird der Test mit dem Lippenlesen durchgeführt. Ein weiteres Testszenario wäre, wie der Einsatz der bedingten Wahrscheinlichkeiten der Wörter, die Erkennungsrate erhöhen. Beim Test wird die höchste bedingte Wahrscheinlichkeit genommen. Hierbei soll ebenfalls überprüft werden, ob es einen gravierenden Unterschied in der Performance gibt, da sonst beim Textmining mehr Sätze überprüft/gefiltert werden müssten.

3.3.2 Erste Ideen

Für die Master-Thesis werden drei Implementierungen ausgeführt: Output bedingte Wahrscheinlichkeiten, Text mining der Befehle und Fusion Lippenlesen und Spracherkennung.

- **Output bedingte Wahrscheinlichkeiten:** Die Pocketshpinx API muss weiter studiert werden und das ROS package muss erweitert werden, damit zu den jeweiligen erkannten Wörtern, die bedingten Wahrscheinlichkeiten als Output mitgeliefert werden.
- **Text mining der Befehle:** Die antreffenden Sätze müssen gefiltert werden. Es werden Stopwörter (häufig auftretende Wörter, keine Relevanz) gelöscht und es findet ein „Matching“ mit bekannten Wörtern statt.

$$\text{'Hello Robo can you please move forward'} \Rightarrow \text{'Robo move forward'} \quad (4)$$

Aus dem erkannten Satz wird der Identifikator (Robo), der Befehl (move) und ein Set von Daten (forward) gefiltert.

- **Fusion Lippenlesen und Spracherkennung:** Die Spracherkennung liefert die bedingten Wahrscheinlichkeiten der Wörter und das Lippenlesen die Konsonanten. Beim „Matching“ wird das korrekte Wort, aus der Liste, der erkannten Wörter der Spracherkennung, genommen.

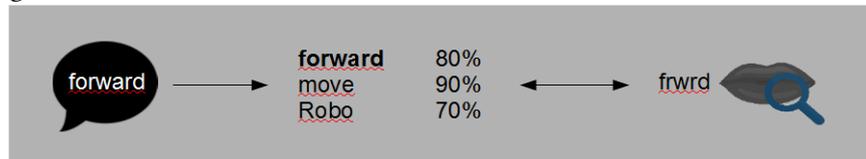


Abbildung 9: Matching - Spracherkennung und Lippenlesen

In Abbildung 9 wird beispielsweise „forward“ gesagt. „move“ erhält, in der Erkennung, die höchste Wahrscheinlichkeit. Beim Lippenlesen wurden jedoch die Konsonanten „frwrđ“ erkannt. Nun kann beim „Matching“ das richtige Wort ausgewählt werden.

3.4 Risiken

In diesem Abschnitt werden die Risiken erläutert, die vor und während der Entwicklung der Anwendung entstehen können. Die Machbarkeit spielt in der Risikoanalyse eine entscheidende Rolle. Die Machbarkeit der Anwendung bezieht sich zunächst auf das Mikrofon und die Kamera. Ein passendes Mikrofon zu finden, welches im Assistenzroboter einsetzbar ist, kann sich als schwer erweisen. Primäres Ziel, in diesem Zusammenhang, ist es ein On-board-Mikrofon zu nutzen. In einem typischen Assistenzroboterszenario werden Befehle ohne zusätzlich elektrische Hilfsmittel erteilt. Ist es aus technischen Gründen nicht möglich ein On-board-Mikrofon zu nutzen, muss ein passendes Headset verwendet werden. Eine passende Kamera muss ebenfalls ausgesucht werden. Das Lippenlesen sollte nicht unter erschwerten Bedingungen funktionieren. Erschwerte Bedingungen wären beispielsweise, dass Befehle nur sehr dicht an die Kamera gesagt werden können, da die Bildverarbeitung es nicht leisten kann aus verschiedenen und weiten Abständen, die Lippen zu erkennen.

Ein geringes Risiko stellt die zu verwendete Spracherkennung dar, da gegebenenfalls die Spracherkennung, die erkannten Wörter, mit ihre bedingten Wahrscheinlichkeiten als Ausgabe, nicht liefern kann. In diesem Fall muss nach einer Alternative geschaut werden.

Die Performance stellt ebenfalls ein Risikofaktor dar. Bildverarbeitungsmethoden benötigen viele Ressourcen und durch die Methodik, die in Projekt 2 (2.2) vorgestellt wurde, kann das Erkennen der vollständigen Lippen langsam sein. In diesem Fall müssen weitere Forschungen getätigt werden, um neue Techniken zu finden, um Lippen zu erkennen. Ein weiterer Punkt ist, dass durch die Kombination der Spracherkennung, dem Text mining und der Bildverarbei-

tung, die Performance darunter leiden kann. Der Ausführungsfluss kann gestört werden, da unabhängige Berechnungen stattfinden.

4 Schluss

In diesem Kapitel soll diese Arbeit kurz zusammengefasst werden.

4.1 Zusammenfassung



Abbildung 10: Phasen der Arbeit

In dieser Arbeit wurde das Vorgehen zur Master-Thesis präsentiert. In Abbildung 10 sind die Phasen der Arbeit abgebildet. Diese Phasen wurden konkret in dieser Arbeit vorgestellt. In der ersten Phase wird ein Sprachsteuerungsinterface implementiert. Ziele in der Phase sind, sich mit ROS vertraut zu machen, eine alternative Steuerung für das ROS-Labor zur Verfügung zu stellen und ebenfalls die erste Berührung mit der Spracherkennung zu haben. Die Anwendung wird als Software für die spätere Weiterentwicklung benutzt. In der zweiten Phase wird versucht die Spracherkennung zu unterstützen. Hierbei wird Bildverarbeitung benutzt, um die Lippen zu erkennen. Gründe für den Einsatz der Bildverarbeitung und die Verbesserung der Spracherkennung werden in 2.2 gezeigt. Die dritte Phase stellt die Master-Thesis dar, in der alle Kenntnisse aus den beiden vorherigen Phasen genutzt werden. Es erfolgt eine Fusion der Sprachsteuerung mit dem Lippenlesen, um eine durch Bildverarbeitung unterstützende Spracherkennung zu implementieren. Text mining wird in der Master Thesis behandelt, da ganze Sätze als Kommandos interpretiert werden sollen. Die Evaluation der unterstützenden Spracherkennung wird in der Master Thesis durchgeführt. Es wird untersucht, ob sich die Erkennungsrate der Spracherkennung, durch die Bildverarbeitung verbessert hat.

Literatur

- [Duden 2015] DUDEN: *Phonem, Fonem, das*. 2015. – URL <http://www.duden.de/rechtschreibung/Phonem>
- [Goodrich und Schultz 2007] GOODRICH, Michael A. ; SCHULTZ, Alan C.: Human-robot Interaction: A Survey. In: *Found. Trends Hum.-Comput. Interact.* 1 (2007), Januar, Nr. 3, S. 203–275. – URL <http://dx.doi.org/10.1561/11000000005>. – ISSN 1551-3955
- [Quigley u. a. 2009] QUIGLEY, Morgan ; CONLEY, Ken ; GERKEY, Brian P. ; FAUST, Josh ; FOOTE, Tully ; LEIBS, Jeremy ; WHEELER, Rob ; NG, Andrew Y.: ROS: an open-source Robot Operating System. In: *ICRA Workshop on Open Source Software, 2009*
- [University 2015] UNIVERSITY, Carnegie M.: *CMU Sphinx*. 2015. – URL <http://cmusphinx.sourceforge.net>
- [Yau u. a. 2006] YAU, Wai C. ; KUMAR, Dinesh K. ; ARJUNAN, Sridhar P.: Voiceless Speech Recognition Using Dynamic Visual Speech Features. In: *Proceedings of the HCSNet Workshop on Use of Vision in Human-computer Interaction - Volume 56*. Darlinghurst, Australia, Australia : Australian Computer Society, Inc., 2006 (VisHCI '06), S. 93–101. – URL <http://dl.acm.org/citation.cfm?id=1273385.1273401>

Abbildungsverzeichnis

1	Roboter im Robot Vision Labor	6
2	Aufbau - Simple Sprachbefehle	7
3	Architektur von ROS-Speech-Control-Interface	8
4	Spracherkennung Workflow - Anwendungsgebiet	9
5	Lippenlesen Workflow	10
6	Motion History Image - Phoneme	11
7	Aufbau - Simple Sprachbefehle mit bedingten Wahrscheinlichkeiten	12
8	Vorgehen - Prozessflow	13
9	Matching - Spracherkennung und Lippenlesen	15
10	Phasen der Arbeit	16