

Bedienung einer Getränkemischmaschine über Sprachbefehle

STUDIENARBEIT

für die Prüfung zum

Bachelor of Science

des Studienganges Angewandte Informatik

an der

Dualen Hochschule Baden-Württemberg Karlsruhe

von

Felix Manuel Gervasi

und

Alena Sutiagina

Abgabedatum 31. März 2023

Bearbeitungszeitraum	6 Monate
Matrikelnummer	1052491
Kurs	TINF20B4
Ausbildungsfirma	

Betreuer der Studienarbeit	Prof. Dr. Jörn Eisenbiegler
----------------------------	-----------------------------

Erklärung

Ich versichere hiermit, dass ich meine Studienarbeit mit dem Thema: »Bedienung einer Getränkemischmaschine über Sprachbefehle« selbstständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe. Ich versichere zudem, dass die eingereichte elektronische Fassung mit der gedruckten Fassung übereinstimmt.

Ort Datum

Unterschrift

Zusammenfassung

TODO

Abstract

TODO

Inhaltsverzeichnis

Abbildungsverzeichnis	iii
Listingverzeichnis	v
Akürzungsverzeichnis	vi
1 Einleitung	1
2 Stand der Technik	2
2.1 Getränkemischmaschine	2
2.2 Hardware	2
2.3 Sprachverarbeitung	2
2.3.1 Spracherkennung	2
2.3.2 Vergleich verschiedener Ansätze	2
3 Konzept	3
3.1 Allgemein	3
3.2 Bewertungskriterien	3
3.3 Konzept A: Spracherkennung und -verarbeitung mittels Arduino	4
3.4 Konzept B: Spracherkennung und -verarbeitung mittels mobiler Anwendung	6
3.5 Konzept C: Spracherkennung und -verarbeitung auf Computer-Hardware	7
3.6 Konzept für die Sprachsteuerung	8
3.7 Finales Konzept	8
4 Implementierung	10
4.1 Implementierung des Sprachverarbeitungssystems	10
4.1.1 Word2Vec-Modell	10
4.1.2 Sequence-to-Sequence-Modell	10
4.2 Befehlsverarbeitung in der Mischmaschine	10
4.3 Anbindung des Sprachmodells an die Mischmaschine	10

<i>INHALTSVERZEICHNIS</i>	ii
5 Fazit und Ausblick	11
Literaturverzeichnis	vii
Liste der ToDo's	vii

Abbildungsverzeichnis

3.1	Spracherkennung und -verarbeitung mittels Arduino	5
3.2	Spracherkennung und -verarbeitung mittels mobiler Anwendung	6
3.3	Spracherkennung und -verarbeitung auf Computer-Hardware	7

Tabellenverzeichnis

3.1	Bewertung der Konzepte	8
-----	----------------------------------	---

Listingverzeichnis

Abkürzungsverzeichnis

HTTP Hypertext Transfer Protocol	4
---	---

Kapitel 1

Einleitung

TODO

Kapitel 2

Stand der Technik

2.1 Getränkemischmaschine

2.2 Hardware

2.3 Sprachverarbeitung

2.3.1 Spracherkennung

2.3.2 Vergleich verschiedener Ansätze

Kapitel 3

Konzept

In diesem Kapitel wird zunächst erläutert, wie die Steuerung der Getränkemischmaschine durch Sprachbefehle im Allgemeinen ablaufen wird. Anschließend werden mehrere Konzepte vorgestellt, die das allgemeine Konzept konkretisieren. Diese werden anhand der, in Kapitel 3.2 erläuterten Kriterien, bewertet. Zuletzt wird die Wahl des finalen Konzepts begründet.

3.1 Allgemein

Der Benutzer soll über Spracheingaben mit der Mischmaschine interagieren können. Dafür muss das Gesprochene zunächst durch ein Mikrofon aufgenommen werden. Anschließend können die Audiosignale weiterverarbeitet werden. Der Benutzer soll hierbei nicht auf fest vorgegebene Sprachbefehle beschränkt sein, sondern für nahezu jede Eingabe eine sinnvolle Antwort zurück erhalten. Um dies zu gewährleisten wird die Spracheingabe durch ein Sprachmodell, welches mittels maschinellen Lernverfahren trainiert wurde, verarbeitet. Ergebnisse dieser Verarbeitung sind die Antwort, die an den Benutzer zurückgegeben wird, und ein konkreter Befehl für die Mischmaschine. Ein Beispiel für einen solchen Befehl könnte etwa die Zubereitung eines bestimmten Getränks sein. Für die Ausgabe einer Antwort ist ein Lautsprecher notwendig. Denkbar wäre auch eine textbasierte Ausgabe, allerdings ginge damit der Eindruck des Benutzers verloren eine echte Konversation mit der Mischmaschine zu führen. Das Sprachmodell mit der Getränkemischmaschine zu verknüpfen stellt eine Herausforderung dieser Arbeit dar.

3.2 Bewertungskriterien

Im Folgenden sind die Bewertungskriterien für die einzelnen Konzepte aufgelistet:

- Freiheitsgrade in der Spracheingabe des Benutzers: Erhält der Benutzer passende Antworten zurück egal was er sagt oder ist er auf einige wenige Befehle beschränkt?

3.3. KONZEPT A: SPRACHERKENNUNG UND -VERARBEITUNG MITTELS ARDUINO⁴

- Hardwarekosten: Wie kostspielig ist das Konzept bezüglich der zusätzlich benötigten Hardware?
- Verfügbare Rechenleistung: Wie hoch ist die Verfügbare Rechenleistung im Vergleich zu den anderen Konzepten? Reicht diese aus um das Sprachmodell auszuführen?
- Performanz: Als wie performant wird die Lösung eingeschätzt? Ist mit Latenzen zwischen der Einagbe des Benutzers, der Ausgabe einer Antwort und Ausführung der Aktion zu rechnen?
- Overhead: Wie hoch ist im Allgemeinen der Mehraufwand einzuschätzen?

3.3 Konzept A: Spracherkennung und -verarbeitung mittels Arduino

Ein erstes Konzept sieht vor, dass das Audiosignal direkt von einem der Arduinos in der Getränkemischmaschine aufgenommen wird. Das Audiosignal wird vom Arduino interpretiert und eine passende Antwort wird ausgegeben. Außerdem sendet der Arduino die entsprechenden Signale, um die vom Benutzer gewünschte Aktion von der Getränkemischmaschine ausführen zu lassen. Ein Problem ist hierbei die Interpretation des Audiosignals durch den Arduino, da dessen Leistung nicht für das Ausführen eines Sprachmodells ausreicht. Folglich muss dieser Prozess ausgelagert werden. Das Konzept wird deshalb um ein cloudbasiertes Sprachverarbeitungssystem ergänzt, welches den Sprachbefehl des Benutzers vom Arduino entgegennimmt und einen passenden Befehl und eine passende Antwort zurückgibt (s. Abb. 3.1). Die Kommunikation zwischen Arduino und Cloudsystem kann über das Hypertext Transfer Protocol (HTTP) erfolgen.

3.3. KONZEPT A: SPRACHERKENNUNG UND -VERARBEITUNG MITTELS ARDUINO5

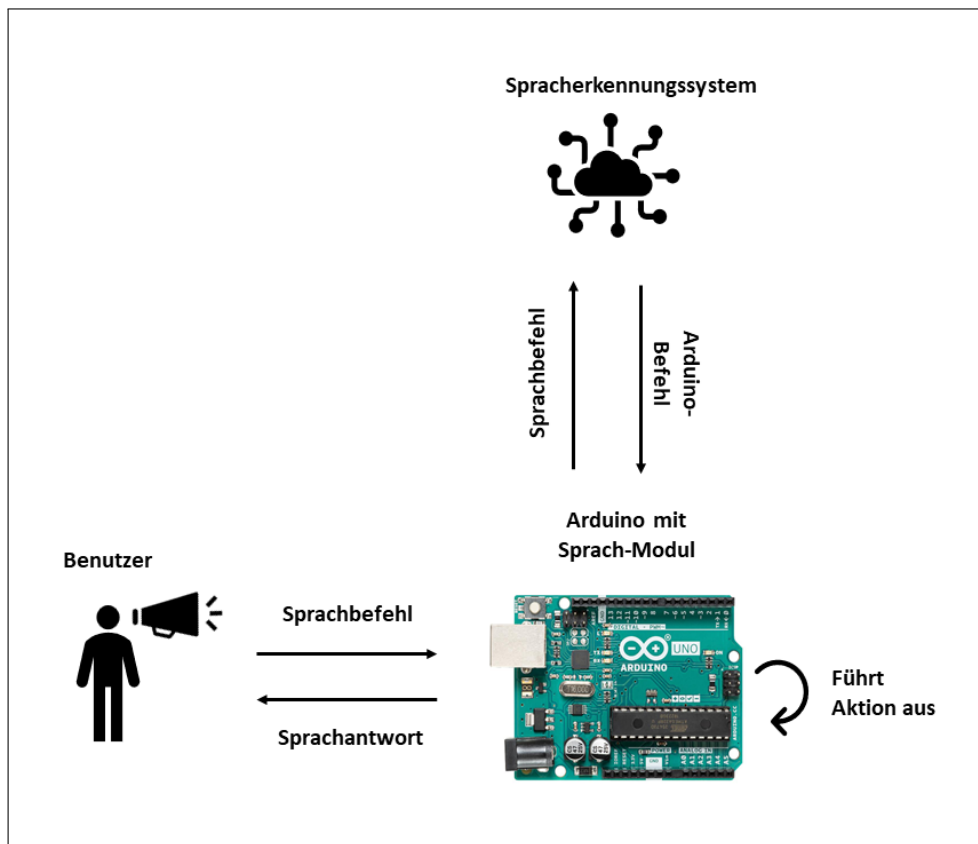


Abbildung 3.1: Spracherkennung und -verarbeitung mittels Arduino

Es muss ein geeignetes Format zum Versenden des Sprachbefehls über HTTP gefunden werden. Eine Möglichkeit besteht darin, das eingehende Audiosignal im Arduino in textform umzuwandeln und diesen String zu versenden. Die auf dem Markt verfügbaren Arduino-Sprach-Module sind jedoch nicht in der Lage beliebige Spracheingaben in Text umzuwandeln, sondern bieten diese Funktionalität nur für vordefinierte Werte an. Dies würde das Ziel dieser Arbeit verfehlen, dem Benutzer eine Konversation mit der Mischmaschine zu ermöglichen. Ein weiteres Problem dieser Lösung besteht darin, dass beispielsweise bei wechselnder Getränkeauswahl die zur Verfügung stehenden Sprachbefehle wie „Ich hätte gerne Getränk xy“ jedes Mal aufs neue manuell angepasst werden müssten. Dies hat zur Folge, dass auch die reine Spracherkennung aus der Mischmaschine ausgelagert werden muss. Ein denkbare Format sind die rohen Audiosignale, die vom Arduino aufgenommen werden.

3.4 Konzept B: Spracherkennung und -verarbeitung mittels mobiler Anwendung

Die Audiosignale über ein Mikrofon in der Mischmaschine aufzunehmen und eine Antwort über einen Lautsprecher auszugeben, so wie es in Konzept A der Fall ist, kann ein Problem darstellen. Zum Einen wird dadurch zusätzliche Hardware benötigt und zum Anderen muss diese korrekt verbaut werden. Das Tonsignal muss vom Mikrofon in einer guten Qualität aufgenommen werden können und die Antwort aus dem Lautsprecher für den Benutzer verständlich sein. Konzept B umgeht dieses Problem durch den Einsatz einer mobilen Anwendung, die durch den Benutzer installiert wird. Über diese Anwendung können anschließend die Aufnahme der Audiosignale, die Spracherkennung und die Kommunikation mit dem Sprachverarbeitungsservice und der Mischmaschine abgewickelt werden, wie in Abbildung 3.2 zu sehen ist.

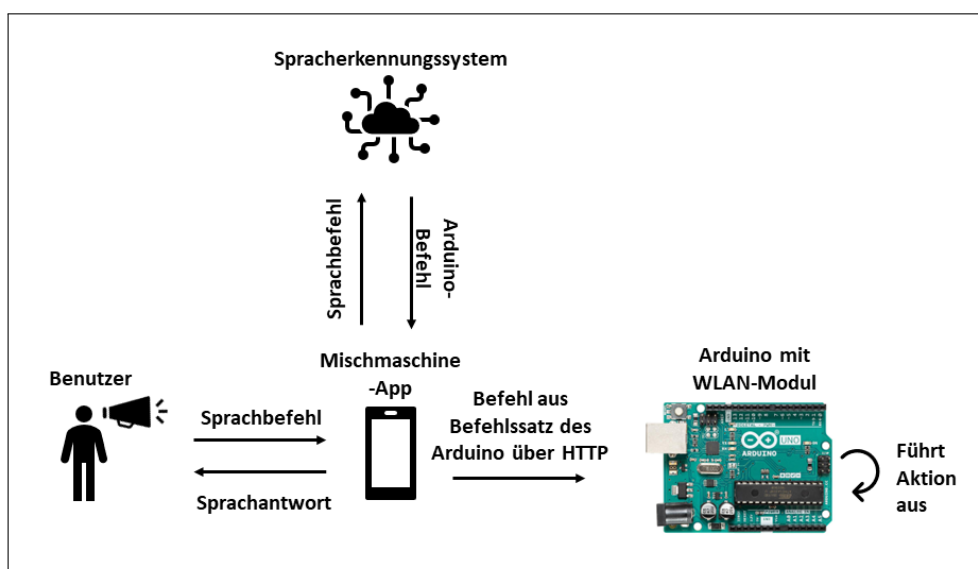


Abbildung 3.2: Spracherkennung und -verarbeitung mittels mobiler Anwendung

Ein Problem dieser Lösung ist der offensichtliche Mehraufwand durch die Entwicklung einer eigenen Anwendung für Mobiltelefone. Auch der Anwender hat zusätzlichen Aufwand durch die Installation. Außerdem ist die Spracheingabe und -ausgabe über das Mobiltelefon nicht intuitiv, da der Anwender eigentlich mit der Maschine kommunizieren sollte. Dieser Effekt kann dadurch abgeschwächt werden, dass wenigstens die Antwort durch einen Lautsprecher in der Mischmaschine an den Benutzer zurückgegeben wird.

3.5 Konzept C: Spracherkennung und -verarbeitung auf Computer-Hardware

Ein weiteres Konzept stützt sich auf die Verwendung eines Computers in der Mischmaschine anstelle eines Mikrocontrollers wie dem Arduino. Motivation ist hierbei der Leistungsgewinn gegenüber eines Mikrocontrollers, um die Spracherkennung und -verarbeitung mittels Sprachmodell zu gewährleisten. Ein Beispiel für einen solchen Miniaturcomputer ist der Raspberry-Pi. Dieser bietet genügend Schnittstellen, wie etwa USB-Hubs, zum Verbinden von Mikrofon als auch Lautsprecher. Nimmt der Computer das Audiosignal auf verarbeitet er dieses und generiert daraus die Antwort, die durch den Lautsprecher ausgegeben wird, zusammen mit der Aktion für die Getränkemischmaschine. Diese muss an den Arduino, welcher die Mischmaschine steuert, übermittelt werden. Um dies zu ermöglichen können der Computer und der Arduino über eine serielle Schnittstelle, wie etwa einem USB-Kabel, miteinander verbunden werden. Abbildung 3.3 stellt den konzeptionellen Aufbau graphisch dar.

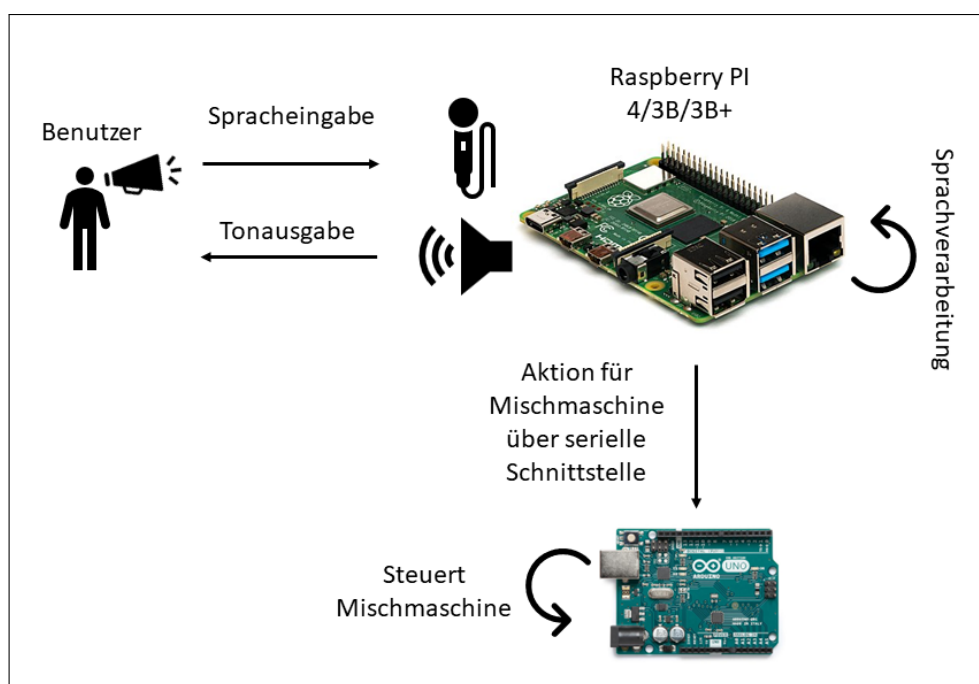


Abbildung 3.3: Spracherkennung und -verarbeitung auf Computer-Hardware

Obwohl ein Computer wie der Raspberry-Pi im Allgemeinen eine höhere Leistung als ein Mikrocontroller hat ist damit nicht sichergestellt, dass diese zur Ausführung des Sprachmodells ausreicht. Beispielsweise ist das vierte Modell der Raspberry-Pi-Serie mit nur maximal acht Gigabyte Arbeitsspeicher erhältlich. Das Sprachmodell könnte allerdings noch weitaus mehr Daten im Arbeitsspeicher benötigen. Des Weiteren ist zu beachten, dass die Miniaturcomputer von Raspberry-Pi im Speziellen zum Zeitpunkt dieser Arbeit kaum zu vertretbaren Preisen

verfügbar sind.

3.6 Konzept für die Sprachsteuerung

3.7 Finales Konzept

Die folgende Tabelle zeigt eine Übersicht bezüglich der Bewertung der einzelnen Konzepte.

Bewertungsmatrix	Konzept A (nur Arduino)	Konzept A	Konzept B	Konzept C
Freiheitsgrade in der Spracheingabe des Benutzers	niedrig	sehr hoch	sehr hoch	hoch
Hardwarekosten	niedrig	hoch	niedrig	sehr hoch
Verfügbare Rechenleistung	niedrig	sehr hoch	sehr hoch	hoch
Performanz	sehr hoch	mittel	mittel	sehr hoch
Overhead	niedrig	hoch	sehr hoch	niedrig

Tabelle 3.1: Bewertung der Konzepte

Ein erster Ansatz wurde im Kapitel 3.3 zu Konzept A erläutert. Hierbei war die Idee, die Spracherkennung und -verarbeitung nur auf dem Arduino auszuführen. Aufgrund der mangelnden Leistung eines Arduinos können mit Hilfe von Sprachmodulen jedoch nur vordefinierte Sätze oder Wörter erkannt werden. Damit ist der Freiheitsgrad in der Spracheingabe des Benutzers äußerst eingeschränkt. Dafür sind die aufzuwendenden Hardwarekosten minimal. Lediglich das Sprachmodul sowie ein Lautsprecher müssten besorgt werden. Die Performanz wird als sehr hoch eingeschätzt, da die Spracherkennung direkt auf dem Arduino erfolgen kann, der auch die Mischmaschine steuert. Niedrig ist hingegen der benötigte Mehraufwand, da kaum zusätzliche Anwendungen, Services oder Hardwarekomponenten benötigt werden.

Das Konzept wurde schließlich durch einen Sprachverarbeitungsservice in der Cloud ergänzt (Konzept A). Der Benutzer hat bei diesem Ansatz große Freiheiten in seinen Formulierungen, da es kein Problem darstellt ein großes Sprachmodell in der Cloud auszuführen. Die Hardwarekosten könnten allerdings hoch sein, je nach dem, ob der Server selbst bereitgestellt oder von einem externen Anbieter bezogen wird. Auch die letzten Endes tatsächlich benötigte Rechenleistung spielt dabei eine Rolle. Die verfügbare Rechenleistung ist theoretisch unbegrenzt, wobei die Performanz des Gesamtsystems nur als mittelmäßig eingestuft werden kann. Nach der Aufnahme und eventuell einer Vorverarbeitung des Aduiosignals durch den Arduino müssen HTTP-Nachrichten gesendet und Empfangen werden. Je nach Last auf dem Netzwerk kann es dadurch zu Latenzen oder sogar Verbindungsabbrüchen kommen. Außerdem ist der Overhead durch den Einsatz einer

Cloud recht hoch.

Konzept B unterscheidet sich in Sachen Freiheitsgrad, Rechenleistung und Performanz nicht von Konzept A, da auch hier eine Cloud zum Einsatz kommt. Die Hardwarekosten sind jedoch niedriger, da immerhin kein Sprachmodul, Mikrofon und Lautsprecher benötigt werden. Die Aufgaben dieser Komponenten kann das Mobiltelefon des Anwenders übernehmen. Der Overhead ist deutlich größer, das das Konzept die Entwicklung einer eigenen Mobilanwendung voraussetzt.

Der Freiheitsgrad wird bei Konzept C als hoch, jedoch nicht als sehr hoch, bewertet. Grund hierfür ist die, im Vergleich zur Cloud, etwas beschränkte Leistung, welche die Leistung/Größe des Sprachmodells beeinträchtigen könnte. Hardwarekosten können jedoch sehr hoch werden. Bei dem Einsatz einer Cloud kann ein günstiger Anbieter gefunden werden, sodass die Anschaffung eigener Hardware entfällt. Dies ist hier nicht der Fall. Die Performanz des Gesamtsystems kann, wie bei Konzept A (nur mittels Arduino), sehr hoch eingeschätzt werden, da Spracherkennung und -verarbeitung direkt in der Mischmaschine von der Hardware übernommen wird. Der Mehraufwand ist gering, da weder ein Cloudservice noch eine externe Anwendung entwickelt werden müssen.

Kapitel 4

Implementierung

4.1 Implementierung des Sprachverarbeitungssystems

4.1.1 Word2Vec-Modell

4.1.2 Sequence-to-Sequence-Modell

4.2 Befehlsverarbeitung in der Mischmaschine

4.3 Anbindung des Sprachmodells an die Mischmaschine

Kapitel 5

Fazit und Ausblick

TODO

Liste der ToDo's