

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 | ii |
| Listingverzeichnis | iii |
| Akürzungsverzeichnis | iv |
| 1 Einleitung | 1 |
| 2 Stand der Technik | 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 . . . | 5 |
| 3.5 Konzept C: Spracherkennung und -verarbeitung auf Computer-Hardware | 6 |
| 3.6 Konzept D: | 7 |
| 3.7 Finales Konzept | 7 |
| 4 Implementierung | 8 |
| 5 Fazit und Ausblick | 9 |
| Literaturverzeichnis | v |
| Liste der ToDo's | v |

Abbildungsverzeichnis

| | | |
|-----|---|---|
| 3.1 | Spracherkennung und -verarbeitung mittels Arduino | 4 |
| 3.2 | Spracherkennung und -verarbeitung mittels mobiler Anwendung | 5 |
| 3.3 | Spracherkennung und -verarbeitung auf Computer-Hardware | 6 |

Listingverzeichnis

Abkürzungsverzeichnis

| | |
|---|---|
| HTTP Hypertext Transfer Protocol | 4 |
|---|---|

Kapitel 1

Einleitung

TODO

Kapitel 2

Stand der Technik

TODO

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
- Hardwarekosten
- Verfügbare Rechenleistung

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.

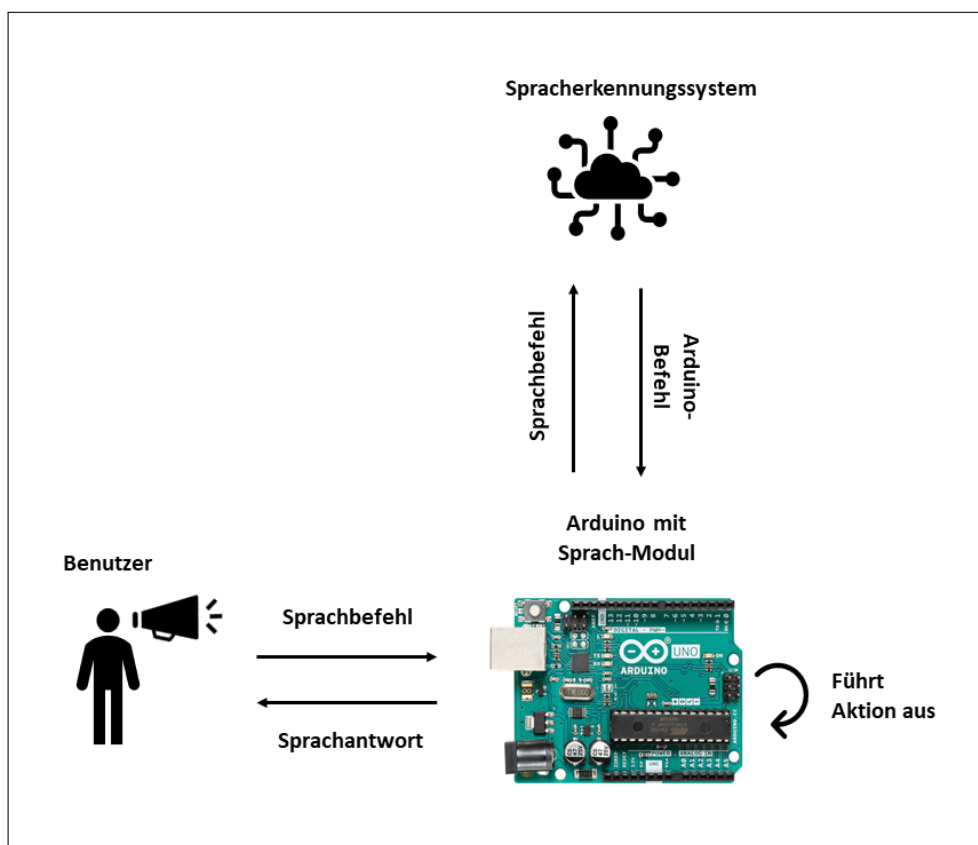


Abbildung 3.1: Spracherkennung und -verarbeitung mittels Arduino

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

3.4. KONZEPT B: SPRACHERKENNUNG UND -VERARBEITUNG MITTELS MOBILER ANWENDUNG

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.

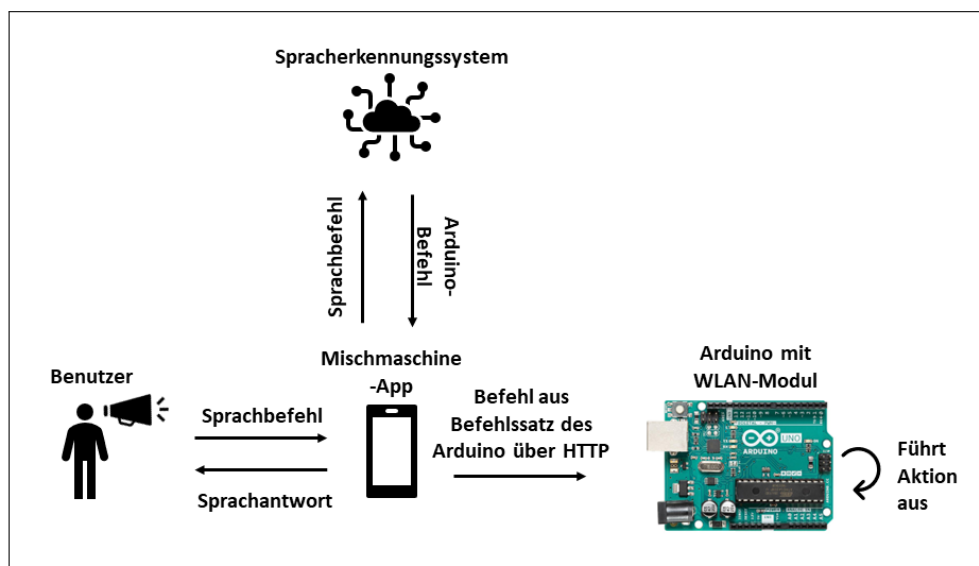


Abbildung 3.2: Spracherkennung und -verarbeitung mittels mobiler Anwendung

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.

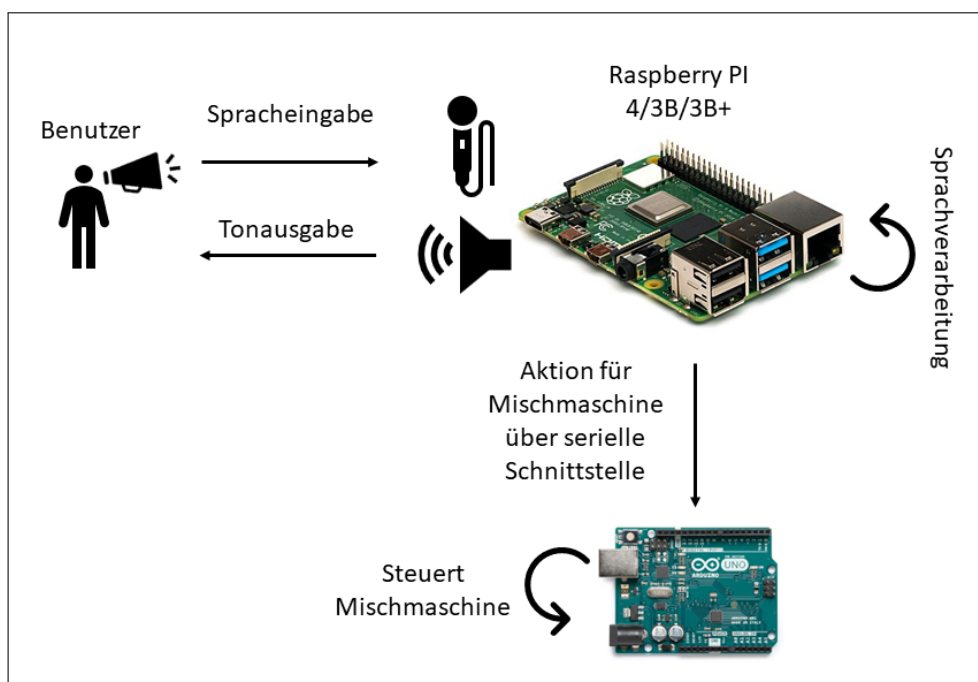


Abbildung 3.3: Spracherkennung und -verarbeitung auf Computer-Hardware

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 D:

3.7 Finales Konzept

| | Konzept A | Konzept B | Konzept C | Konzept D |
|---|-----------|-----------|-----------|-----------|
| Freiheitsgrade in der Spracheingabe des Benutzers | | | | |
| Hardwarekosten | hoch | niedrig | sehr hoch | |
| Verfügbare Rechenleistung | niedrig | hoch | hoch | |

// TODO: Konzepte anders aufteilen damit die Bewertungsmatrix einen Sinn ergibt; Konzept A weiter unterteilen in alles auf Arduino/in Maschine und Äufteilen zwischen Maschine und Cloud"

Kapitel 4

Implementierung

TODO

Kapitel 5

Fazit und Ausblick

TODO

Liste der ToDo's