

Bachelorarbeit

Integration einer Sprachsteuerungsfunktion in Mobile Apps

Rubén Nuñez

Herbstsemester 2023

Bachelorarbeit an der Hochschule Luzern – Informatik

Titel: Integration einer Sprachsteuerungsfunktion in Mobile Apps

Studentin/Student: Ruben Nuñez

Studiengang: BSc Informatik

Jahr: 2023

Betreuungsperson: Dr. Florian Herzog

Expertin/Experte: xxx

Auftraggeberin/Auftraggeber: Stefan Reinhard, Bitforge AG

Codierung / Klassifizierung der Arbeit:

☒ Öffentlich (Normalfall)

☐ Vertraulich

Eidesstattliche Erklärung Ich erkläre hiermit, dass ich die vorliegende Arbeit selbständig und ohne unerlaubte fremde Hilfe angefertigt habe, alle verwendeten Quellen, Literatur und andere Hilfsmittel angegeben habe, wörtlich oder inhaltlich entnommene Stellen als solche kenntlich gemacht habe, das Vertraulichkeitsinteresse des Auftraggebers wahren und die Urheberrechtsbestimmungen der Hochschule Luzern respektieren werde.

Ort / Datum, Unterschrift _____

Abgabe der Arbeit auf der Portfolio Datenbank:

Bestätigungsvisum Studentin/Student

Ich bestätige, dass ich die Bachelorarbeit korrekt gemäss Merkblatt auf der Portfolio Datenbank abgelegt habe. Die Verantwortlichkeit sowie die Berechtigungen habe ich abgegeben, so dass ich keine Änderungen mehr vornehmen kann oder weitere Dateien hochladen kann.

Ort / Datum, Unterschrift _____

Verdankung gibt ein separiertes Kapitel dazu

Ausschliesslich bei Abgabe in gedruckter Form: Eingangsvisum durch das Sekretariat auszufüllen

Rotkreuz, den _____

Visum: _____

Abstract

Das Problem dieser Arbeit ist im wesentlichen die Erkennung von Triggerwörtern innerhalb des Kontext einer App. Grundsätzlich ist es unüblich, dass mobile Apps eine integrierte Sprachsteuerungsfunktion anbieten.

Inhaltsverzeichnis

1	Problem, Fragestellung, Vision	5
2	Stand der Forschung	6
2.1	Audio	6
2.1.1	Sampling	6
3	Ideen und Konzepte	7
4	Methoden	8
5	Realisierung	9
6	Evaluation und Validation	10
7	Ausblick	11
8	Anhang	12
8.1	Projektmanagement	12
8.2	Grobplanung	12
8.2.1	Produkt Backlog	12
8.2.2	Risikomanagement	12
	Abbildungsverzeichnis	13
	Tabellenverzeichnis	13
	Literaturverzeichnis	13

1 Problem, Fragestellung, Vision

Das Problem dieser Arbeit ist im wesentlichen die Erkennung von Triggerwörtern innerhalb des Kontext einer App. Grundsätzlich ist es unüblich, dass mobile Apps eine integrierte Sprachsteuerungsfunktion anbieten.

2 Stand der Forschung

Laut Einstein (1905) ist E gleich mc^2 .

Dieser Text wird ohne Einzug beginnen. Wie Einstein (1905) bemerkte, ist $E = mc^2$.

„Ein bekanntes Ergebnis aus der Relativitätstheorie ist $E = mc^2$ (Einstein, 1905).“

2.1 Audio

Das Verständnis von Audio ist für diese Arbeit von zentraler Bedeutung. In der digitalen Welt repräsentiert Audio Schallwellen, die durch eine Reihe von numerischen Werten dargestellt werden. Somberg et al., 2019, p.9 beschreibt Audio als: „Fundamentally, audio can be modeled as waves in an elastic medium. In our normal everyday experience, the elastic medium is air, and the waves are air pressure waves.“ Audiosignale werden durch die Funktion $\mathbf{A}(t)$ repräsentiert, wobei t die Zeit und $\mathbf{A}(t)$ die Amplitude zum Zeitpunkt t angibt. Die Amplitude ist die Stärke des Signals und die Zeit repräsentiert die Position des Signals in der Zeit. Diese Betrachtung ist vor allem in der Elektrotechnik von Bedeutung, da die Amplitude als Spannung angesehen werden kann. Grundsätzlich ist Audio ein kontinuierliches Signal. In der digitalen Welt können wir jedoch nur diskrete Werte darstellen. Daher wird das kontinuierliche Signal in diskrete Werte umgewandelt. Dieser Vorgang wird als *Sampling* bezeichnet. (Tarr, 2018, Chapter 3.1)

2.1.1 Sampling

Ein früher Ansatz zur digitalen Darstellung von analogen Signalen war die Pulse-Code-Modulation (PCM). Dieses Verfahren wurde bereits in den 1930er Jahren von Alec H. Reeves entwickelt, parallel zum Aufkommen der digitalen Telekommunikation (Deloraine und Reeves, 1965, p. 57). Im Grundsatz wird es heute noch in modernen Computersystemen nach dem gleichen Verfahren angewandt.

Nun folgt eine formelle Definition von Sampling. Ein kontinuierliches Signal $\mathbf{A}(t)$ wird in bestimmten Zeitintervallen T_s gesampelt. Diese Zeitintervalle werden auch als Sampling-Periode bezeichnet. Die Sampling-Rate $F_s = \frac{1}{T_s}$ gibt die Anzahl der Samples pro Sekunde an.

3 Ideen und Konzepte

Das Problem dieser Arbeit ist im wesentlichen die Erkennung von Triggerwörtern innerhalb des Kontext einer App. Grundsätzlich ist es unüblich, dass mobile Apps eine integrierte Sprachsteuerungsfunktion anbieten.

4 Methoden

Das Problem dieser Arbeit ist im wesentlichen die Erkennung von Triggerwörtern innerhalb des Kontext einer App. Grundsätzlich ist es unüblich, dass mobile Apps eine integrierte Sprachsteuerungsfunktion anbieten.

5 Realisierung

Das Problem dieser Arbeit ist im wesentlichen die Erkennung von Triggerwörtern innerhalb des Kontext einer App. Grundsätzlich ist es unüblich, dass mobile Apps eine integrierte Sprachsteuerungsfunktion anbieten.

6 Evaluation und Validation

Das Problem dieser Arbeit ist im wesentlichen die Erkennung von Triggerwörtern innerhalb des Kontext einer App. Grundsätzlich ist es unüblich, dass mobile Apps eine integrierte Sprachsteuerungsfunktion anbieten.

7 Ausblick

Das Problem dieser Arbeit ist im wesentlichen die Erkennung von Triggerwörtern innerhalb des Kontext einer App. Grundsätzlich ist es unüblich, dass mobile Apps eine integrierte Sprachsteuerungsfunktion anbieten.

8 Anhang

8.1 Projektmanagement

Wie im vorhergehenden Kapitel «Agile Entwicklung - Automatisiertes Testen / DevOps» beschrieben, sind sämtliche gesammelten Informationen und gefällten Entscheidungen in einem Artefakt zu vermerken. Anleitungen für typische Softwareartefakte können dabei nützliche Wegleitungen sein.

8.2 Grobplanung

Die Grobplanung zeigt die wichtigsten Meilensteine des Projekts auf. Zudem werden die wichtigsten Aufgaben und deren zeitliche Abfolge dargestellt.

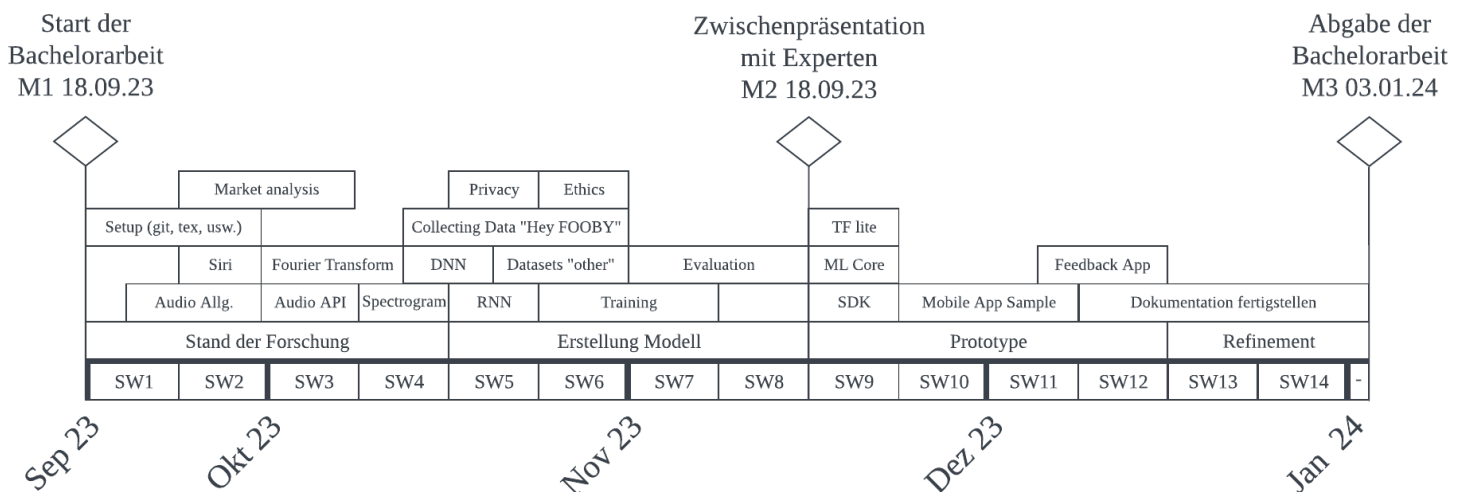


Abbildung 1: Grobplanung

8.2.1 Produkt Backlog

In der Vorbereitungsphase kann ein anfängliches Produkt Backlog als einfache Tabelle dargestellt werden. Ein Beispiel für eine solche Tabelle ist in Abbildung 5 dargestellt.

8.2.2 Risikomanagement

Risikomanagement dient dem Zweck, mögliche Probleme vorwegzunehmen. Die Verwendung von Checklisten, Brainstorming mit den Anspruchsgruppen und die von Erfahrungen aus früheren Projekten sind mögliche Strategien zur Identifikation möglicher Risiken.

Tabelle 1: Beispiel-Tabelle für Risikomanagement

Kopf 1	Kopf 2	Kopf 3
Wert 1	Wert 2	Wert 3
Wert 4	Wert 5	Wert 6

Tabelle 2: Eine einfache Tabelle



Abbildung 2: Tabelle für das anfängliche Product Backlog

Abbildungsverzeichnis

1	Grobplanung	12
2	Tabelle für das anfängliche Product Backlog	13

Tabellenverzeichnis

1	Beispiel-Tabelle für Risikomanagement	12
2	Eine einfache Tabelle	12

Literaturverzeichnis

- Deloraine, E. M., & Reeves, A. H. (1965). The 25th anniversary of pulse code modulation. *IEEE Spectrum*, 2(5), 56–63. <https://doi.org/10.1109/MSPEC.1965.5212943>
- Einstein, A. (1905). Zur Elektrodynamik bewegter Körper. *Annalen der physik*, 322(10), 891–921.
- Somberg, G., Davidson, G., & Doumler, T. (2019). A Standard Audio API for C++: Motivation, Scope, and Basic Design [“C++ is there to deal with hardware at a low level, and to abstract away from it with zero overhead.” – Bjarne Stroustrup, Cpp.chat Episode #44]. *Programming Language C++*.
- Tarr, E. (2018). *Hack audio : : an introduction to computer programming and digital signal processing in MATLAB* (1st edition). Routledge.

Aufgabenstellung

Integration von Sprachsteuerungstechnologien in Mobile Apps, insbesondere zur Erkennung von Triggerwörtern.

Projektteam

- Student:in: Rubén Nuñez
- Betreuer:in: Herzog
- Firma: Bitforge AG

Auftraggeber

- Firma: Bitforge AG
- Ansprechperson: Stefan Reinhard
- Funktion: Head of Mobile
- Adresse: Zeughausstrasse 39, 8004 Zürich
- Telefon: +41 55 211 02 41
- E-Mail: stefan.reinhard@bitforge.ch
- Website: www.bitforge.ch

Ausgangslage und Problemstellung

Sprachsteuerungstechnologien haben ein grosses Potenzial und werden bisher vor allem als Sprachsteuerungsassistenten genutzt. Während es etablierte Sprachassistenten wie Siri gibt, fehlt es an Lösungen für eine integrierte Sprachsteuerung in Mobile Apps, insbesondere in Bezug auf das Erkennen von Triggerwörtern.

Ziel der Arbeit und erwartete Resultate

Ziel der Arbeit ist es zum einen, eine Grundlage zu schaffen, um ein Triggerwort oder eine Sequenz von Triggerwörtern in der akustischen Sprache erkennen zu können. Dabei werden Methoden und Werkzeuge aus dem Bereich des Machine Learnings verwendet. Zum anderen soll diese Erkenntnis in eine mobile Plattform wie iOS oder Android integriert werden. Für den Rahmen dieser Arbeit genügt die Integration in eine der genannten Plattformen. Weiterhin werden das Thema Datenschutz und die ethischen Aspekte berücksichtigt.

Gewünschte Methoden, Vorgehen

Das Projekt kann beispielsweise in drei Phasen durchgeführt werden: Technische Abklärungen, Datensammlung und Modelltraining, sowie die Erarbeitung eines Prototypen. Agile Vorgehensweisen sind wünschenswert.

Kreativität, Methoden, Innovation

Bisher sind Sprachsteuerungsfunktionen fast ausschliesslich grossen Akteuren wie Siri vorbehalten. Der innovative Ansatz dieser Arbeit zielt darauf ab, einen Anreiz zu setzen, um diese Funktionen auch in herkömmlichen Apps einzusetzen. Die handfreie Bedienung durch Sprachsteuerung hat das Potenzial, das Benutzererlebnis erheblich zu verbessern.

Sonstige Bemerkungen

Grundkenntnisse in Machine Learning, speziell im Bereich der Spracherkennung, sowie Erfahrung mit entsprechenden APIs sind erforderlich.