

**Gestaltung eines physischen Mensch-Maschine-Interface  
zur KI-basierten, mittels Licht unterstützten Sprachinteraktion**

**Design of a Physical Human-Machine Interface  
for AI Speech Interaction Enhanced by Light Signals**

Martin Rothmayer BSc.  
m1698094

**Masterarbeit der Studienrichtung Industrial Design**  
Institut für Raum und Design an der  
Universität für künstlerische und industrielle Gestaltung Linz

Vorgelegt am 12.10.2023 zur Erlangung des  
akademischen Grades Diplom-Ingenieur.

**Betreuer**

Univ.-Prof. Mag.art. Elke Bachlmair  
Univ.-Prof. Mag.art. Mario Zeppetzauer

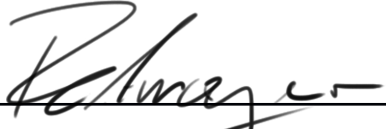
**Projektpartner KEBA Group AG**

Thomas Linde  
Julian Bauer BSc.  
Christoph Etzlinger



## Eidesstattliche Erklärung

Ich bestätige, dass diese Masterarbeit meine eigene Arbeitsleistung ist und ich alle verwendeten Quellen und Materialien dokumentiert habe. Diese Arbeit wurde bisher keiner anderen Prüfung unterzogen und nicht veröffentlicht.

  
\_\_\_\_\_  
Linz am 12.10.2023

## Abstrakt

Diese Arbeit widmet sich der Zusammenführung der Themen Sprache, Licht, KI und Industrie. Ziel ist die Entwicklung eines sprachbasierten Human-Machine-Interfaces für die Anwendung im industriellen Umfeld. Das Interface fungiert als Kommunikationsschnittstelle zwischen dem Nutzer und der KI-Steuerungseinheit des Unternehmens KEBA.

Um ein besseres Verständnis für die Interaktion zwischen Menschen und sprachbasierten KI-Systemen zu erlangen, wurden verschiedene Interfaces analysiert. Dabei wurden sowohl Science-Fiction-Filme als auch aktuelle Sprachassistenten wie Siri und Alexa untersucht. Genauer betrachtet wurden sowohl sprachliche als auch visuelle Komponenten. Die gewonnenen Erkenntnisse dienen als Grundlage für die Entwurfsphase des Interfaces.

Das entwickelte Interface wird in die äußere Hülle der Maschine integriert und verfügt über mehrere Mikrofone, einen Lautsprecher und eine umlaufende Lichtfläche. Diese Fläche dient der Darstellung des aktuellen Systemstatus und unterstützt die gesprochene Kommunikation. Die Anwendungsfelder sind vielfältig und können von großen Industriemaschinen bis hin zu kleinen Roboterarmen reichen.

## Abstract

This work focuses on the combination of the topics of language, light, artificial intelligence, and industry. The goal is the development of a speech-based human-machine interface for use in an industrial environment. The interface acts as a communication interface between the user and the KEBA AI control unit.

Various interfaces were analyzed to gain a better understanding of the interaction between humans and speech-based AI systems. Both science fiction films and current voice assistants such as Siri and Alexa were investigated. Both linguistic and visual components were considered in more detail. The insights gained served as a basis for the design phase of the interface.

The developed interface is integrated into the outer shell of the machine and features multiple microphones, a speaker, and a surrounding light surface. This area is used to display the current system status and supports spoken communication. The fields of application are diverse and can range from large industrial machines to small robotic arms.

# Inhaltsverzeichnis

<b>I. Einleitung.....</b>	<b>10</b>
1. Initiator des Projekts	12
2. Einleitung	12
3. Ziel	12
4. Anwendungsbereiche	13
5. KEBA-Markenwerte	13
6. Methodik	13
<b>II. Theoretisches Fundament.....</b>	<b>14</b>
1. Mensch-Maschine-Interface	16
2. Übersicht KI in Filmen	20
3. Übersicht KI und Spracherkennung in der Realität	26
4. KEBA KI Steuerung	31
5. Virtuelle Assistenten	32
6. Mental Model	37
7. Smart Home Speaker	38
8. Lichtkonzepte von Smart Speakern	46
9. Emotionale Beziehungen zu VA	50
10. Anwendungsbeispiele für Sprachinteraktion im Auto	51
11. Nonverbale Kommunikation	53
12. Stimme	55
13. Funktionsklänge	57
14. Herausforderung/Schwierigkeiten von VUI	57
15. Zusammenfassung und Fazit	59
<b>III. Konzeptphase Interface .....</b>	<b>60</b>
1. Erster Workshop	62
2. Formal Moods	63
3. Nutzungsanforderungen	64
4. Zweiter Workshop	65
5. KEBA-Unterscheidungswerte	66
6. Konzeptfindung	67
7. Keysketch	72
8. Baugröße und Proportion	73
9. Technische Analyse von Smart Speakern	76
10. Technisches Package	78

<b>IV. Bespielung Interface .....</b>	<b>84</b>
1. Gestaltungsansatz	86
2. Entwurfsentwicklung der Lichteffekte	86
3. Mock-Up	88
4. Aufgabenverteilung	91
5. Modi	92
6. Übergänge	94
7. Farbauswahl	94
8. Zentrum	95
<b>V. Interaktion .....</b>	<b>96</b>
1. Interaction Journey	98
2. Eigenschaften	102
3. Umgang mit Fehlern	102
4. Einstellungen	102
5. Wissensbereiche	103
<b>VI. Voice Arena .....</b>	<b>104</b>
1. Voice Arena	106
2. Bespielung	106
3. Positionierung	107
<b>VII. Ausblick.....</b>	<b>116</b>
1. Nutzeridentifikation	118
2. Kontextbezug der KI	118
3. Kombination mit anderen Eingabemethoden	118
<b>VIII. Verzeichnisse .....</b>	<b>120</b>
1. Quellenverzeichnis	122
2. Abbildungsverzeichnis	128
3. Danksagung	131

# **I. Einleitung**



# 1. Initiator des Projekts

Initiator dieses Projekts ist das Innovation Lab der Firma KEBA Group AG, ein international agierendes Technologieunternehmen mit Hauptsitz in Linz-Urfahr. KEBA bietet Hard- sowie Softwarelösungen in folgenden Geschäftsfeldern: *Industrial Automation*, *Handover Automation* und *Energy Automation*.

## 2. Einleitung

Haptische Bedienelemente gibt es seit vielen Jahren und sie haben dementsprechend bewährt. Aus einfachen Tastern und Schaltern wurden durch die technische Weiterentwicklung hochauflösende Touchscreens mit informativen Grafiken, die sich wie moderne Smartphones bedienen lassen.

Seit einigen Jahren ist es möglich durch natürliche Sprache mit Smart-Assistants zu kommunizieren. Sie verstehen Befehle und Fragen der Nutzer und können mit ihnen in einen Dialog treten. Der große Vorteil ist die überwiegend barrierefreie Bedienung, so hat man die Hände frei für andere Tätigkeiten. Außerdem kann der Dialog auch vom Smart-Assistant initiiert werden, was das Verhältnis des Nutzers zur Maschine dynamischer gestaltet.

Nicht nur die Eingabemethoden haben sich weiterentwickelt, auch die Maschinen selbst werden zunehmend intelligenter und autonomer. Dadurch eröffnen sich neue Möglichkeiten und Herausforderungen. Einerseits können Maschinen komplexe Aufgaben selbstständig durchführen, was die Effizienz und Produktivität erhöht, andererseits ist eine klare Kommunikation und Interaktion zwischen Mensch und Maschine erforderlich, um sicherzustellen, dass die gewünschten Aufgaben und Ziele verstanden und richtig ausgeführt werden. Einen Teil dieser Kommunikation können Lichtsignale übernehmen. Eine multimodale Interaktion hat Potenzial die Qualität der Interaktion zu verbessern.

## 3. Ziel

Das vorrangige Ziel dieser Arbeit ist es ein Interface zu gestalten, das eine effektive Kommunikation zwischen einem menschlichen Nutzer und einer intelligenten Maschine ermöglicht. Das Interface ist mit der lokalen KI-Steereinheit verbunden und muss in der Lage sein, mit einem Nutzer einerseits mittels Sprache und andererseits über visuelle Signale zu kommunizieren.

Ein zentraler Aspekt ist dabei die deutliche Visualisierung der Künstlichen Intelligenz (KI) nach außen, um deren Präsenz und Intelligenz erkennbar zu machen. Dabei ist es wichtig nicht nur die Informationen darzustellen, sondern vor allem dem Nutzer ein möglichst intuitives und natürliches Arbeiten zu ermöglichen. Außerdem soll sich das Interface nahtlos in den Markenkontext von KEBA einfügen. Dies betrifft einerseits die Gestaltung und andererseits die Konstruktion für ein industrielles Umfeld.

**Aus diesen Rahmenbedingungen ergeben sich folgende Forschungsfragen:**

- Wie können Lichtakzente die Interaktion mit einem sprachbasierten Human-Machine-Interface unterstützen?
- Wie muss dieses Interface gestaltet sein, um den KEBA-Werten zu entsprechen und im industriellen Umfeld bestehen zu können?

## 4. Anwendungsbereiche

Die Anwendungsbereiche des Interface sind vielfältig. Es kann in verschiedene Industriemaschinen integriert werden, wie beispielsweise Spritzgussmaschinen, Laserschneider oder mobile Roboter. Die möglichen Anwendungsfelder reichen von Geldautomaten bis Paketabholstationen.

Ein autonomer, einarmiger Roboterarm diente während der Entwurfsphase als Beispiel Use-Case. Mit einer Kamera können verschieden Objekte erkannt und abgespeichert werden. Mithilfe eines Greifers werden die gescannten Objekte anschließend sortiert. Im Hintergrund arbeitet ein KEBA-AI-Modul daran, die Sprachbefehle zu verstehen und den Roboterarm richtig zu lenken.

## 5. KEBA-Markenwerte

KEBA steht für höchste Qualität im Bereich der Automatisierung und bietet optimierte Gesamtlösungen für Kunden und Anwender. Über allen Produkten steht der Leitspruch „easy to use“. Für den Benutzer tritt KEBA oft nicht direkt in Erscheinung. Kontakt besteht meist indirekt durch die stationären oder mobilen HMI. Diese sind in verschiedenen Stufen anpassbar, von der Farbe des Gehäuses und den sichtbaren Logos bis hin zum Layout der Tastatur kann vom Kunden alles verändert werden. Die KEBA-Markenwerte müssen daher bei der Gestaltung beachtet werden, das Branding soll sich aber eher im Hintergrund halten.

## 6. Methodik

Um die Interaktion zwischen Menschen und KI-Systemen besser zu verstehen, wurden einerseits Voice-Assistants (VA) wie Siri und Alexa analysiert, und andererseits bekannte KI-Systeme aus Filmen genauer betrachtet, insbesondere Interfaces die vorrangig in natürlicher Sprache kommunizieren. Diese Erkenntnisse wurden in übersichtlichen Tabellen zusammengefasst und dienten, gepaart mit den Werten der Firma KEBA, als Basis für die Entwurfsphase. Am Ende dieser steht ein Keysketch, der die wichtigsten Aspekte der zuvor festgelegten Merkmale in sich vereint.

Am Beginn der Entwurfsphase fanden zwei Workshops mit KEBA statt. Einerseits um herauszufiltern, worauf die Schwerpunkte des Projekts liegen sollen und andererseits einen Ausgangspunkt für die Ideation zu finden. In weiteren regelmäßigen Besprechungen wurden die Entwürfe und Projektfortschritte kritisch analysiert und diskutiert.

Um unvoreingenommen von technischen Lösungen zu sein, fand die Smart Speaker Analyse erst statt, nachdem das Lichtkonzept bereits festgelegt wurde. Anschließend wurde ein technisches Package für die konkreten Anforderungen erstellt. Mithilfe einer User Journey wurde das Interaktionskonzept entwickelt und die Bespielung der Leuchtfelder festgelegt.

# **II. Theoretisches**

**Fundament**

# 1. Mensch-Maschine-Interface

## 1.1. Definition

Als Bezeichnung für ein Human-Machine-Interface (HMI) existieren verschiedene deutsche und englische Begriffe, die häufig als Synonyme verwendet werden: Schnittstelle, Interface, Bedienoberfläche, User Interface, Human-Maschine-Interface usw. Es gestaltet sich jedoch schwierig, eine allgemein gültige Definition für HMI zu formulieren.

Obwohl Stuart K. Cards Beschreibung eines Human-Machine Interfaces bereits 40 Jahre alt ist, ist sie auch heute noch angemessen.

*The human-computer interface is easy to find in a gross way—just follow a data path outward from the computer's central processor until you stumble across a human being. Identifying its boundaries is a little more subtle.*  
(Card, 1983, p. 4)

Wenn in dieser Arbeit der Begriff Human-Machine-Interface (HMI) kurz „Interface“ verwendet wird, beschreibt er das physische Gerät, über das der Benutzer mit einer Maschine kommuniziert. Die folgende Abbildung veranschaulicht das Verhältnis zwischen einem System (eine beliebige Maschine), dem Interface und dem Benutzer. Der Benutzer interagiert nicht direkt mit der Maschine, sondern mit dem Interface. Unabhängig von der spezifischen Form des Interfaces bestimmt es die Bandbreite der Handlungsmöglichkeiten. Die Funktionalitäten der Maschine sind durch das Interface begrenzt und der Benutzer kann nur das ausführen, was das Interface ermöglicht (vgl. Dorau, 2011, p. 220)

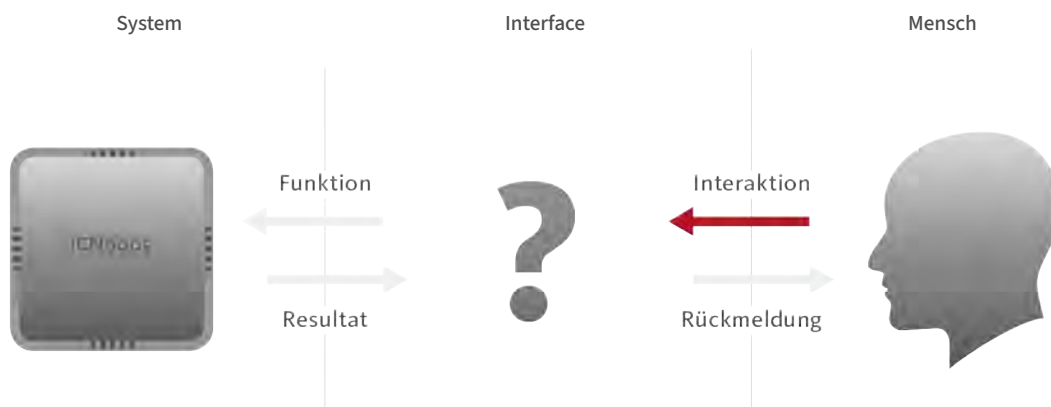


Abb. 01: Kommunikationsmodell System-Interface-Mensch (Dorau, 2011, S. 221)

KEBA teilt seine HMI in zwei Kategorien ein – mobil und stationär. Beide Arten haben Vor- und Nachteile und sind je nach Anwendungsfall auszuwählen oder zu kombinieren.

## 1.2. KEBA - Stationäre HMI

Bei einem stationären HMI handelt es sich meist um ein grafisches User-Interface (GUI), sprich ein Monitor und verschiedene physische Bedienelemente wie Tasten, Schalter und Drehregler. Ein Operator-Panel ist direkt an der Maschine angebracht. Es ist mit der Grafikschnittstelle der Steuerung verbunden und visualisiert Informationen für den Benutzer.

Ein Active-Panel ist ein interaktives Multi-Touch-Display. Hierbei werden die klassischen Schalter und Tasten durch das Touchdisplay ersetzt. Häufig wird sogenannte Force-Feedback-Technologie eingesetzt, das bedeutet in diesem Fall, dass nicht bei der Berührung des Displays, sondern erst beim Drücken der Taste (Force) eine Aktion ausgelöst wird. Weiterhin erhält der Nutzer auch eine Rückmeldung (Feedback) als Information, dass der Befehl ausgeführt wurde. (vgl. Lang, 2022).

### Vorteile:

- + Große Bildschirme bieten viel Platz für Informationen und Funktionen
- + Niedrige Herstellkosten bei der Hardware
- + Große physische Bedienelemente erleichtern das präzise Einstellen von Parametern
- + Eine Positionierung in sicherer Entfernung zwingt den Benutzer aus dem Gefahrenbereich

### Nachteile:

- Eingeschränkte Flexibilität
- Operator-Panels haben einen erhöhten Engineering-Aufwand bei der Software, damit Visualisierung und Steuerungsapplikation flüssig parallel laufen können. Funktioniert die Steuerung nur langsam und träge, so wird die Maschine als „schlecht“ wahrgenommen
- Active-Panels erfordern, dass der Benutzer seine Aufmerksamkeit vom Arbeitsbereich nimmt und sie auf den Touchscreen lenkt um den richtigen Knopf zu betätigen



**Abb. 02:** Stationäre KeTop Geräte im Überblick: KeTop AP 500, KeTop OP 300 Keyboard Panels, KeTop OP Singletouch, KeTop OP 500 Multitouch (KEBA, 2023)

### 1.3. KEBA - Mobile HMI

Ist nicht genügend Platz für ein großes Display vorhanden oder wird ein erhöhter Bewegungsradius am Arbeitsplatz benötigt, kommen mobile Einheiten zum Einsatz. Handheld-Geräte stehen in verschiedenen Größen und Formen zur Verfügung. Es ist möglich, Bildschirme, Folientastaturen, Drehregler und Not-Aus-Taster in ein kleines, tragbares Gerät zu integrieren. Kabellose Geräte bieten zusätzliche Flexibilität und ermöglichen dem Benutzer, sich ungehindert im Arbeitsbereich zu bewegen (vgl. Lang, 2022).

#### Vorteile:

- + Erhöhte Bewegungsfreiheit bei der Bedienung
- + Bessere Einsicht in den Prozess oder den Maschinenbauraum
- + Bedienung von Maschinen aus sicherer Entfernung
- + Ein HMI kann mit mehreren Maschinen verbunden werden

#### Nachteile:

- Kabellose Geräte können verloren gehen oder verlegt werden
- Ist das kabellose Gerät zu weit von der Maschine entfernt, weiß man im Zweifelsfall nicht mehr, zu welcher Maschine das Bedienelement gehört
- Mögliche Schäden durch unvorsichtige Nutzung



**Abb. 03:** Mobile KeTop Geräte im Überblick: KeTop T10 directMove, KeTop T20 eco/technic, KeTop T55, KeTop T70, KeTop T150, KeTop T200, KeTop T150/155 Safe Wireless) (KEBA, 2023)

## 1.4. Vorteile einer Industrie HMI

Im industriellen Umfeld werden Smartphones und Tablets nur begrenzt zur Bedienung verwendet. Es gibt mehrere gute Gründe, die gegen den Einsatz von Smartphones und für spezifische HMI-Werkzeuge sprechen (vgl. KEBA, 2022).

- + Hohe physische Anforderungen erfordern eine robuste Bauweise
- + Industrie-Panels bieten alle relevanten Sicherheitsfunktionen wie Not-Halt und Zustimmungstaster
- + Sie erfüllen industrielle Ergonomienormen, das bedeutet die Griffbereiche müssen entsprechend groß sein und die Nutzung darf nicht zu Ermüdungserscheinungen führen
- + Große haptische Bedienelemente ermöglichen eine fokussierte Bedienung, bei der man sich auf den Tastsinn verlassen kann
- + Wichtig ist auch, dass die Geräte jederzeit einsatzbereit sind und es nur wenige Unterbrechungen aufgrund von Updates oder Überhitzung gibt

## 1.5. Conversational-User-Interfaces

Conversational-User-Interfaces (CUIs) sind Schnittstellen, die es Benutzern ermöglichen mit Systemen und Anwendungen auf natürliche und dialogbasierte Weise zu interagieren. Sie erfordern Sprach- oder Texteingaben und nutzen Technologien wie Spracherkennung und Natural Language Processing (NLP) (vgl. Kabel, 2020, pp. 2,8).

Die Stärken von CUIs liegen in ihrer Benutzerfreundlichkeit, da sie eine intuitive und flexible Kommunikation ermöglichen. Sie bieten auch einen barrierefreien Zugang für Menschen mit eingeschränkter Sicht oder Mobilität. Allerdings haben CUIs auch einige Schwächen, wie eine begrenzte Kontextsensitivität und Schwierigkeiten bei der Interpretation von mehrdeutigen Eingaben. Sie können zudem anfällig für Missverständnisse oder Fehlinterpretationen sein (vgl. Kabel, 2020, p. 31).

## 1.6. Fazit

Mobile und stationäre HMI haben ihre Vor- und Nachteile in verschiedenen Anwendungsbereichen. Trotzdem sind sie lediglich Werkzeuge, die man zu bedienen lernt um bestimmte Aufgaben zu erledigen. 1998 brachte es Don Norman in „The Invisible Computer“ auf den Punkt:

*I don't want to use a computer, I want to accomplish something-  
Don Norman (p. 89).*

Ein intelligentes Sprachinterface, das direkt mit der Maschinensteuerung verbunden ist, hat das Potenzial Teile dieser Hürde zu überbrücken und eine direktere Verbindung zwischen Mensch und Maschine herzustellen. Sprachsteuerungen könnten die Flexibilität eines mobilen HMI mit der breiten Funktionspalette von stationären Elementen verbinden. Unabhängig vom System oder der eingesetzten Technologie ist die Qualität und Reaktionszeit bei der Interaktion entscheidend. Maschinenbediener erwarten heutzutage ein responsives System, wie sie es von Smartphones gewöhnt sind (vgl. Lang, 2022).

## 2. Übersicht KI in Filmen

Im Folgenden Abschnitt wird auf die Geschichte von dialogfähigen KI-Interfaces in Filmen eingegangen. Sie sind unter anderem die Grundlage für das Mental-Model von Sprachassistenten (siehe Kap.6), welche Funktionen man von ihnen erwarten kann und wie sie aussehen.

### 2.1. KI in Filmen

Ein beliebtes Science-Fiction Narrativ ist, dass eine Künstliche Intelligenz sich selbst *bewusst* wird und versucht, die Menschheit zu vernichten oder zu unterjochen. Siehe *2001: A SPACE ODYSSEY* (2001: Odyssee im Weltraum, 1968), *BLADE RUNNER* (Blade Runner, 1982), *TERMINATOR* (Terminator, 1984) oder *THE MATRIX* (The Matrix, 1999). Dem zu Grunde liegt eine fundamentale Angst die von Isaac Asimov als „Frankenstein-Komplex“ bezeichnet wird. Dies beschreibt die Angst, dass Menschen etwas erschaffen könnten, das so mächtig ist, dass es ihnen über den Kopf wächst und sie dadurch ihre Selbstbestimmung verlieren (vgl. Hermann, 2020, pp. 13-14).

Im Folgenden wird auf Beispiele, deren primäre Interaktionsform natürliche Sprache ist und die darüber hinaus als Assistenzsystem für den Menschen konzipiert wurden, eingegangen. Außer Acht gelassen werden *Körper-KI* wie im Film *EX MACHINA* (Ex Machina, 2014) oder aus der Serie *WESTWORLD* (Westworld, 2016).

#### 2.1.1. HAL 9000 in 2001: A Space Odyssey (1968)

HAL 9000 ist der Zentralcomputer des Raumschiffs Discovery, das nach Außerirdischen suchen soll. Er sieht sich den Menschen intellektuell überlegen und stellt ihre Herrschaft offen in Frage. In einer Art Rebellion tötet HAL fast die gesamte Mannschaft und wird zuletzt nur von zwei Astronauten gestoppt. Das schockierende an HAL ist, dass er sich an keine moralischen Regeln hält und offen gegen die Menschheit rebelliert, seine intellektuelle Überlegenheit steht außer Frage (vgl. Weber, 2008, pp. 8-9). HAL 9000 ist Vorbild für viele weitere Roboter in Filmen und wurde unzählige Male parodiert. Auch Apple's Sprachassistent Siri macht gerne Anspielungen auf HAL.



**„I’m sorry Dave,  
I’m afraid I can’t do that“**

*HAL zu Dave als er sich weigert die  
Luftschleuse zu öffnen*

**Abb. 04:** HAL 9000 aus 2001: Odyssee im Weltraum (1968) (wikipedia, 2023)

In Erscheinung tritt HAL in Form eines roten Kameraobjektivs. Dieses Auge sitzt knapp unterhalb der Mitte eines rechteckigen Interfaces im Hochformat. Oben steht der Name und am unteren Ende befindet sich ein Lautsprecher.

Das kontinuierlich rot leuchtende Auge wirkt bedrohlich und unnachgiebig. Die flache Einbauf orm legt nahe, dass HAL vollkommen in das Schiff integriert ist. Die Tatsache, dass er alles überwacht, lässt Parallelen zum Big Brother aus 1984 ziehen (vgl. Weber, 2008, p. 24).

HAL besitzt eine menschenähnliche Stimme, die ihm eine Persönlichkeit und Menschlichkeit verleiht, obwohl diese Eigenschaften eigentlich nicht existieren. HAL's ruhige und logische Art wirkt umso bedrohlicher, als er anfängt sich gegen die Besatzung zu wenden (vgl. Shedroff & Noessel, 2012, p. 187). Sogar als der Astronaut Dave ihn Stück für Stück abschaltet, bleibt die Stimme höflich, nüchtern und distanziert.

Laut Bret Kinsella, dem Gründer von voicebot.ai, hat HAL eine ganze Generation an KI-Entwicklern traumatisiert. Sie versuchen jede Assoziation mit der synthetischen männlichen Stimme zu vermeiden (vgl. Kinsella, 2018).

### 2.1.2. K.I.T.T. in Knight Rider (1982)

K.I.T.T. ist ein modifizierter Pontiac Firebird mit künstlicher Intelligenz. Er kann denken, sprechen und autonom das Fahrzeug lenken. Der Name K.I.T.T. ist ein Akronym und steht für Knight Industries Two Thousand. Gemeinsam mit seinem Fahrer Michael Knight bekämpft er Verbrechen (vgl. imdb, 2023). Ikonisch ist der rote Lichtstreifen in der Frontschürze des Autos. Er dient als Scanner und pendelt von links nach rechts. Das zweite Interface befindet sich im Innenraum, es ähnelt dem äußeren Lauflicht, bildet aber die Stimme visuell ab. Zusätzlich können auf einem Bildschirm Symbole und Bilder dargestellt werden.

Der Hauptgrund, warum man ihn als eigenständigen Charakter akzeptiert, ist seine sehr menschliche Stimme. KITT verwendet eine gehobene Sprache, variiert seine Stimmlage und hat einen natürlichen Rhythmus, der sich der Situation anpasst. Diese Eigenschaften tragen dazu bei, dass man ihn als einen vollwertigen Charakter wahrnimmt (vgl. Shedroff & Noessel, 2012, p. 168).

**„I am the voice of Knight Industries  
2000s microprocessor.  
K.I.T.T. for easy reference.  
A kid if you prefer.“**

*KITT stellt sich Michael vor*



**Abb. 05:** KITT-Interface und Scanner Licht (Vitale, 2021)

### 2.1.3. Jarvis in Iron Man (2008)

J.A.R.V.I.S. (Just A Rather Very Intelligent System) ist ein fiktiver Charakter aus den Iron Man Comics und Filmen. Er ist ein KI-basiertes Computerprogramm, das Tony Stark entwickelt hat, um ihm bei der Steuerung und Überwachung seines Anzugs sowie bei alltäglichen Aufgaben zu helfen. Darüber hinaus besitzt JARVIS ein umfassendes Verständnis für seinen Benutzer und ist in der Lage, zukünftige Schritte vorzusehen. Oftmals erledigt JARVIS Aufgaben im Voraus, noch bevor sie ihm ausdrücklich erteilt wurden. Eine weitere bemerkenswerte Eigenschaft von JARVIS ist seine Fähigkeit, Daten in Echtzeit zu analysieren und komplexe Probleme zu lösen. Seine KI-Algorithmen ermöglichen es ihm, Informationen aus verschiedenen Quellen abzurufen und sofort zu verarbeiten (vgl. marvelcinematicuniverse, 2023).

Anders als KITT oder HAL hat er weder physische Gestalt noch ein zentrales Interface, er tritt in erster Linie als KI-System auf. Er agiert über Sprachausgabe, Hologramme, Bildschirme oder Heads-Up-Displays. Über die Kameras und Sensoren im Gebäude ist es Tony Stark möglich, ohne Bedienelemente ein von JARVIS generiertes Hologramm mit Gesten zu steuern (vgl. Shedroff & Noessel, 2012, pp. 100-103, 161, 165).

„Jarvis, are you up“



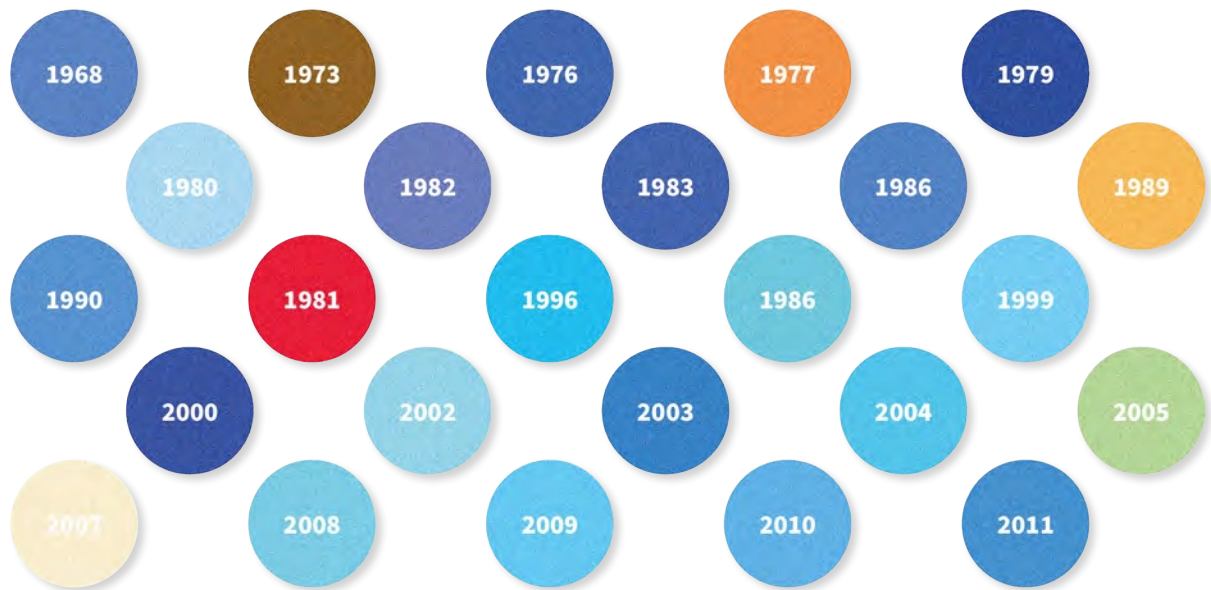
„For you sir, always.“

Abb. 06: Jarvis als Head-up-Display im Iron Man Anzug (Noessel, 2015)

## 2.2. Science-Fiction Interfaces

Der Großteil aller Science-Fiction Interfaces ist blau und leuchtet mit einem leichten Schimmer. Ein möglicher Grund für die Verwendung der Farbe Blau könnte sein, dass sie als kühl und beruhigend wahrgenommen wird und sich daher gut für Arbeitsumgebungen eignet. Außerdem ist blau eine Farbe, die man recht selten in der Natur findet, mit Ausnahme des Himmels natürlich. Es könnte sein, dass man es deshalb als besonders künstlich empfindet und dadurch mit Technologie und Innovation assoziiert. Unabhängig vom Grund wirkt ein blau leuchtendes Interface automatisch futuristisch und modern (vgl. Shedroff & Noessel, 2012, pp. 42-43).

Zusätzlich zur Farbe umgibt Science-Fiction Technologie oft ein Leuchten. Lichtschwerter, Hologramme und Teleporter emittieren Licht. Dies könnte daher rühren, dass mächtige Objekte oft leuchten, zum Beispiel die Sonne, der angestrahlte Mond am Nachthimmel oder ein Feuer. Außerdem faszinieren uns lumineszierende Tiere wie Glühwürmchen oder Kreaturen der Tiefsee (vgl. Shedroff & Noessel, 2012, p. 40).




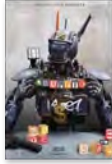




















**Abb. 07:** Science-Fiction Interface sind vorwiegend blau (1991 war Terminator 2) (Shedroff & Noessel, 2012, p. 41) Anmerkung des Autors: Die Art der Darstellung wurde verändert, die Farben wurden beibehalten.

### 2.3. Fazit zu KI und Interfaces in Filmen











Filme sind keine realistische Darstellung der Zukunft und KI ist nicht automatisch böse. Wie ein guter Schauspieler übernimmt KI nur eine Rolle und versucht ihren Zweck zu erfüllen. Trotzdem können Filme, auf soziale Probleme und gesellschaftliche Auswirkungen von KIs aufmerksam machen. Dies kann dabei helfen eine neue Generation von Entwicklern und Programmieren hervorzubringen, die mit mehr Bewusstsein für ihren Einfluss auf die Gesellschaft oder Einzelpersonen ausgestattet sind (vgl. Hermann, 2020, p. 16).

Die mühelose Kommunikation mit HAL, KITT oder JARVIS zeigt eine inspirierende Vision. Wir sollten versuchen dieser Idealvorstellung von Interaktion näher zu kommen. Diese Filme normalisieren die Idee mit einem Computer zu sprechen, dabei sollte aber nicht vergessen werden, dass reale Mensch-Maschine Interaktion viel mehr auf Kontext beruht, als in Filmen gezeigt wird (vgl. Case, 2016, p. 34).



	1999 <b>The Iron Giant</b> Giant		2015 <b>Chappie</b> Chappie
	1987 <b>RoboCop</b> RoboCop		1999 <b>Bicentennial Man</b> Andrew
	1986 <b>Short Circuit</b> Johnny 5 Nr 5 lebt!		1999 <b>The Matrix</b> Agents
	1990 <b>Edward Scissorhands</b> Edward		2001 <b>A.I. Künstliche Intelligenz</b> David
	1995 <b>Ghost in the Shell</b> Motoko Kusanagi		2004 <b>I Robot</b> VIKI Sonny
	1987 <b>Star Trek Enterprise</b> Data		2002 <b>Minority Report</b> Interface
	2005 <b>A Hitchhiker's Guide to the Galaxy</b> Marvin		2008 <b>Wall-E</b> Wall-E Eve
	2014 <b>Ex_Machina</b> Ava		2014 <b>Interstellar</b> TARS KIPP CASE
	2009 <b>Moon</b> Gerty 3000		2022 <b>Kimi</b> Kimi
	2013 <b>Her</b> Samantha		2021 <b>Oxygen</b> MILO
	2008 <b>Iron Man</b> Jarvis		2015 <b>Avengers: Age of Ultron</b> Ultron

1990 2000 2010 2020

	1990 <b>Nuance</b> Dragon Dictate		1999 <b>IBM</b> Home Page Reader		2011 <b>Apple</b> Siri		2016 <b>Google</b> Assistant
	1997 <b>IBM</b> ViaVoice		2013 <b>Microsoft</b> Cortana		2023 <b>OpenAI</b> ChatGPT		
	1997 <b>Mercedes-Benz</b> Linguatronic		2012 <b>Google</b> Now		2014 <b>Amazon</b> Alexa		

## 3. Übersicht KI und Spracherkennung in der Realität

### 3.1. Arten von künstlicher Intelligenz (KI)

Es gibt viele verschiedene Arten und Spezialisierungen von KI. Grundsätzlich unterscheidet man zwischen schwacher und starker KI. Eine starke KI bezieht sich auf ein intelligentes System, das in der Lage ist, komplexe Aufgaben eigenständig zu erlernen, zu verstehen, zu analysieren und zu lösen, vergleichbar mit oder sogar über den kognitiven Fähigkeiten des menschlichen Verstands. Alles mit dem wir heute interagieren und konfrontiert sind, sind schwache KI. Also Software, die auf ein einzelnes Gebiet spezialisiert ist. Zum Beispiel ein Schachcomputer, Siri oder personalisierte Werbung (vgl. Hermes, 2023).

Die meisten KI-Programme basieren auf maschinellem Lernen. Maschinelles Lernen (ML) ist eine Methode, bei der das Programm in der Lage ist, aus Erfahrungen zu lernen, ohne dass alle Regeln explizit programmiert werden. Dies ermöglicht das Erkennen und Vorhersagen von Mustern. Die Basis bilden hierfür große Datenmengen (vgl. Pieraccini, 2021, p. 27).

Praktischerweise gibt es im Internet einen nahezu unerschöpflichen Fundus an Trainingsdaten: Digitalisierte Bücher, tagesaktuelle Nachrichten, Gesetzestexte, E-Mails, Social-Media und viele weitere Texte auf Webseiten. Je größer die Grundlage, umso schneller und besser lernt das System (vgl. Kabel, 2020, p. 8).

Zumeist arbeiten KI-Programme im Hintergrund und personalisieren das Nutzerlebnis von Social-Media Seiten und Musik-Streamingdiensten. Heutzutage ist KI Teil von vielen verschiedenen Anwendungen wie Sprachassistenten, Bilderkennung, autonomen Fahren und Robotik. Die bekannteste KI-Anwendung ist aktuell ChatGPT vom US-Unternehmen OpenAI. ChatGPT ist ein Sprachmodell, das natürliche Sprache versteht. Es kann Fragen beantworten, Texte generieren und in Dialogen kommunizieren.

### 3.2. Spracherkennung und KI

Sprache eignet sich grundsätzlich gut für die Interaktion mit einer Maschine, da sie nach einem bestimmten Regelwerk funktioniert, es gibt Grammatik, Wortfolgen, Konjugationen und einen endlichen Wortschatz. Sätze werden in Einzelteile zerlegt, Worte erkannt und zueinander in Beziehung gesetzt. Das sind eigentlich gute Voraussetzungen für einen Dialog. Das Problem ist, dass ein Computer nicht im herkömmlichen Sinne *versteht*. Aus dem Grund ist auch der Begriff *künstliche Intelligenz* in diesem Kontext ein wenig irreführend, denn mit Intelligenz, wie wir sie verstehen, hat Spracherkennung noch nichts zu tun (vgl. Kabel, 2020, p. 7).

Zwei der drei am größten wachsenden Anwendungen von Künstlicher Intelligenz sind Natural Language Processing (NLP) und Speech Processing (SP). Gemeinsam machen sie 27% aller KI-bezogenen Patente aus (vgl. WIPO, 2019, p. 31). Der Anteil der Unternehmen im produzierenden Wirtschaftsbereich die bereits KI nutzen, liegt jedoch erst bei 7% (vgl. Statistik Austria, 2022, pp. 23,24).

### 3.3. Meilensteine in der Entwicklung von Spracherkennung

#### 3.3.1. Shoebox von IBM - 1961

Bei der Weltausstellung 1962 stellte IBM ein kleines Gerät vor, das in der Lage war, 16 verschiedene gesprochene Worte in elektrische Impulse umzuwandeln. Darunter die Zahlen null bis neun und einfache mathematische Befehle. Eine Rechenmaschine führte die diktierten Berechnungen durch. Es war die erste industrielle Anwendung einer Spracherkennung (vgl. IBM Archive, 1961).

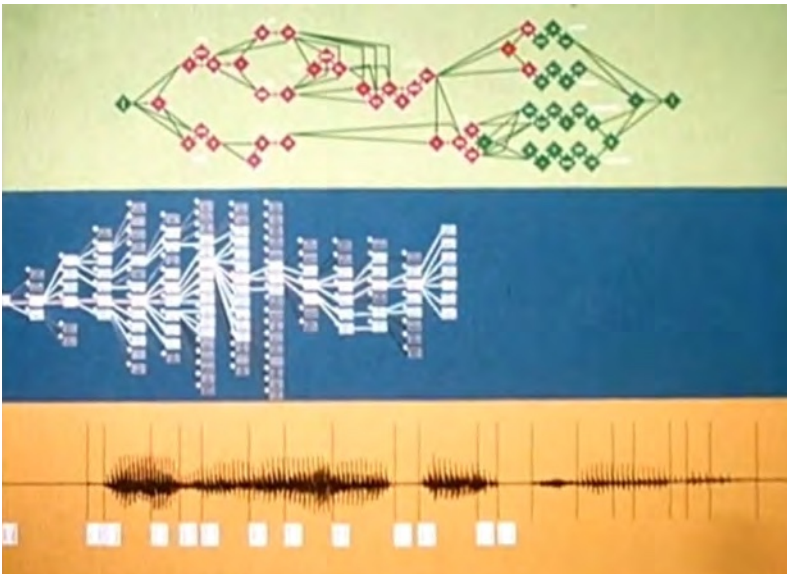
Abb. 09: IBM Shoebox 1962  
(IBM, 2023)



#### 3.3.2. Harpy vom CMU - 1976

HARPY entstand aus einer Kooperation der DARPA (Defense Advanced Research Projects Agency) und der CMU (Carnegie Mellon University). HARPY war der Versuch das Beste aus den beiden Spracherkennungssystemen Hearsay-I (Erman, 1974) und Dragon (James und Janed Baker, 1975) zu vereinen. Hearsay-I nutze ein Blackboard-Modell um die einzelnen Informationen aus Sprache, Gramatik, Vokabular und so weiter zu vereinen und so das gesagte richtig zu erkennen. Dragon hingegen nutze bereits ein HMM (Hidden-Markov-Modell) um gleichzeitig alle syntaktischen und akustischen Pfade zu durchsuchen. (vgl. Lowerre, 1976)

Der Auftrag der DARPA war in fünf Jahren ein Spracherkennungssystem mit einem Wortschatz von 1000 Wörtern zu entwickeln, das nahezu in Echtzeit arbeitet. HARPY setzte auf eine äußerst effiziente Beam-Search-Methode und erzielte bereits eine Genauigkeit von bis zu 99%. (vgl. Reddy, 2018)



**Abb. 10:** Der Recognition-Tree von Harpy (Reddy, 2013)

Eine Idee, die die Spracherkennung grundlegend verändern sollte, stammte von Fred Jelinek, einem Harvard und Cornell Professor, der später über 20 Jahre für IBM arbeitete. Er verfolgte einen datengetriebenen Ansatz, bei dem es darum ging, nicht über menschliche Sprachkonzepte, Regeln und Grammatik Sprache zu verstehen, sondern über sich wiederholende Muster und Wahrscheinlichkeiten von Wortfolgen (vgl. Lohr, 2010).

*„Airplanes don’t flap their wings.“  
Fred Jelinek (Lohr, 2010)*

Die Analogie soll verdeutlichen, dass Maschinen anders arbeiten als Vorbilder aus der Natur. Er suchte nach einem Weg, einer Maschine Sprache beizubringen, ohne ihr menschliche Regeln aufzuzwingen. Daraus folgte der Einsatz des Hidden Markov Model (HMM). Dabei werden Phoneme verwendet, um die Wahrscheinlichkeit zu bestimmen, dass aus einer bestimmten Folge von Lauten ein bestimmtes Wort entsteht. Einfach gesagt, errechnet es die Wahrscheinlichkeit, mit der ein bestimmtes Wort in einem Satz auf ein anderes folgt (vgl. Kincaid, 2018).



**Abb. 11:** Fred Jelinek (links im Bild) mit dem Talking Typewriter von IBM ca. 1980 (IBM, 2011)

### 3.3.3. Dragon Dictate von Dragon Systems – 1990

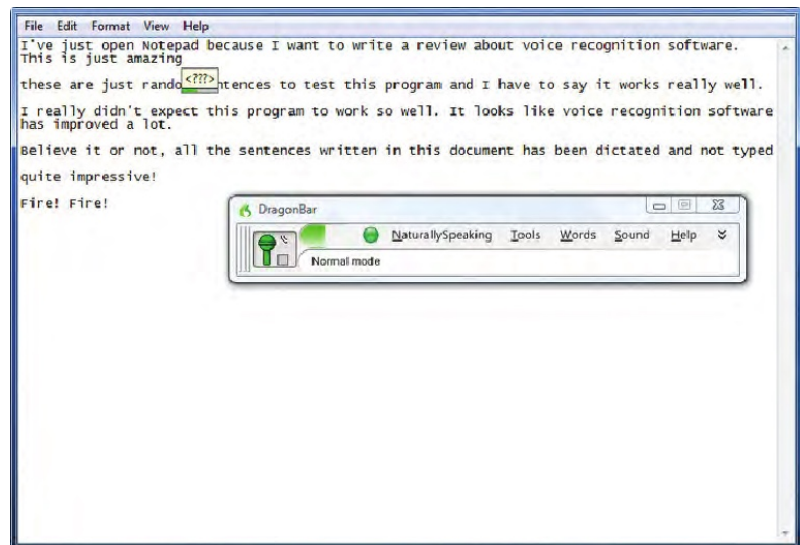
Die erste Anwendung, die für private Nutzer auf einem PC zugänglich war, ist Dragon Dictate. Das Programm verfügte über einen Wortschatz von 80.000 Worten, wobei man nur 30.000 gleichzeitig nutzen konnte. Auch Dictate verwendete das HMM. Das Produkt hatte aber noch einige Schwächen. Man konnte nur 30-40 Worte pro Minute diktieren, wesentlich langsamer als der Durchschnitt mit einer Tastatur mit ca. 76 WPM (Arif & Stuerzlinger, 2009). Zum Vergleich: die normale Sprechgeschwindigkeit liegt im Deutschen zwischen 90-120 Wörter pro Minute (Eilert, 2022). Außerdem kostete eine Lizenz etwa 9000 Dollar (vgl. Kincaid, 2018).

### 3.3.4. Dragon NaturallySpeaking von Dragon Systems - 1997

Einen großen Sprung nach vorne machte die Spracherkennung mit NaturallySpeaking. Man konnte nun endlich natürliche Sprache in Echtzeit verarbeiten und das mit einer Worterkennungsrate von bis zu 99%.

Das Interface besteht aus einem kleinen Fenster, in dem man das Programm aktivieren und deaktivieren kann bzw. Einstellungen vornimmt. Beliebt ist das Programm bei Anwälten und Ärzten, wenn sie Schreiben verfassen. Für einen ungeschulter Benutzer mag die Diktierweise etwas sonderbar wirken, da man Satzzeichen und Zeilenumbrüche mitsprechen muss. Also sagt man am Ende eines Satzes laut „Punkt“ und „Neue Zeile“. Je länger man damit arbeitet, desto besser erlernt man auch die Bedienung.

Abb. 12: Screenshot aus Dragon NaturallySpeaking (Zhang & Ng, , 2011)



Ein roter Faden, den man durch die gesamte Entwicklung von Spracherkennung verfolgen kann, beginnt mit dem Ehepaar James und Janet Baker. James Baker war an der Carnegie-Mellon-University Studetn von Raj Reddy der an Harpy mitwirkte. 1975 wechselte er zuerst von der CMU zum IBM-Forschungszentrum und gründetet 1982 schließlich Dragon Systems. Dragon Systems wurde später von Nuance Communications übernommen und in weitere Folge von Microsoft.

Die Spracherkennungsmethoden von NaturallySpeaking fanden 2003 Anwendung in einem DARPA Projekt namens CALO (Cognitive Assistant that Learns and Organizes). Beauftragt wurde dazu die Firma SRI Inc. in Kalifornien. Einige ehemalige Mitarbeiter von SRI Inc. gründeten 2007 die Spin-off-Firma Siri Inc., die 2010 von Apple aufgekauft wurde (vgl. Mariani, et al., 2014, p. 11).

### 3.4. Natural Language Processing (NLP)

Natural Language Processing (NLP) beschäftigt sich mit der gesamten Bandbreite der Sprachverarbeitung und umfasst eine breite Palette von Techniken und Algorithmen, die darauf abzielen, menschliche Sprache zu analysieren, zu verstehen oder zu generieren.

Mithilfe von Statistik wird errechnet, wie hoch die Wahrscheinlichkeit ist, dass der Sprecher einen gewissen Befehl fordert. Liegt dieser Wert über der geforderten Wahrscheinlichkeit (Confidence Level), so wird der Prozess in die Wege geleitet. Ist die KI sich nicht sicher genug, soll noch einmal nachgefragt werden (vgl. Kabel, 2020, pp. 7,8).

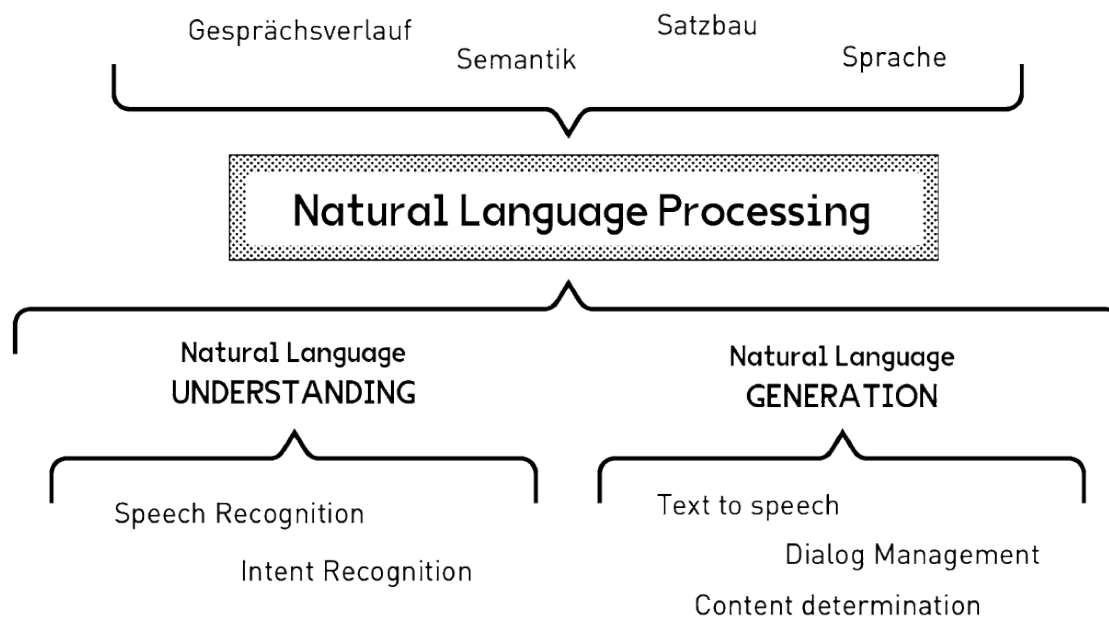


Abb. 13: Übersicht des NLP (Kabel, 2020, p. 8)

### 3.5. Automatic Speech Recognition (ASR)

Das Ziel der Spracherkennung besteht darin, die Worte der gesprochenen Äußerung (utterance) mit möglichst hoher Genauigkeit zu identifizieren. Einfach formuliert, die ASR macht die Arbeit einer Schreibkraft. Es gibt jedoch einige Situationen, die noch Schwierigkeiten bereiten. Zum Beispiel laute Umgebungsgeräusche, wenn verschiedene Personen gleichzeitig sprechen oder wenn jemand während der Aussage die Sprache wechselt (Pieraccini, 2021, p. 53f).

### 3.6. Natural Language Understanding (NLU)

Bei Natural Language Understanding (NLU) geht es darum, den erkannten Worten eine Bedeutung zuzuweisen. Es wandelt die rohen Textdaten in eine abstrakte, strukturierte Form um, mit der das Programm effizient arbeiten kann. Die NLU analysiert den Kontext, die Semantik und die Absicht hinter einer sprachlichen Äußerung und ermöglicht so eine sinnvolle Interpretation der menschlichen Sprache. Kurz gesagt, die NLU wandelt Text, der für Menschen Sinn ergibt, in abstrakte Symbole um, die für ein Programm Sinn ergeben (vgl. Pieraccini, 2021, pp. 9, 107-110). Am Ende soll klar sein, was sich der User vom System erwartet.

### 3.7. Natural Language Generation (NLG)

Die Aufgabe der Natural Language Generation (NLG) besteht darin, dem Benutzer zu antworten, also mit ihm zu sprechen. Die NLG arbeitet dabei genau entgegengesetzt zur NLU. Die strukturierten Daten der NLU geben an, was der virtuelle Assistent sagen soll, und die NLG wandelt diese in eine sinnvolle Wortfolge um. Die generierte Antwort wird dann durch die Text-to-Speech (TTS)-Funktion wiedergegeben (vgl. Pieraccini, 2021, p. 143f).

### 3.8. Dialog Manager (DM)

Der Dialog Manager entscheidet aufgrund der abstrahierten Daten des NLU, welche Aktion als nächstes ausgeführt werden soll. Dabei berücksichtigt er auch die gesamte Interaktionsgeschichte und das verfügbare Kontextwissen. Die Entscheidung des Dialogmanagers kann eine Anfrage an die externe digitale Welt sein oder eine Aufforderung an die NLG, etwas zu generieren, um dem Benutzer zu antworten (vgl. Pieraccini, 2021, p. 9).

## 4. KEBA KI Steuerung

Das KI-Modul besteht aus einer Steuereinheit mit mehreren Industrieschnittstellen. Es bietet große Rechenleistung in einem kleinen Formfaktor. Das Gerät verfügt über mehrere Anschlüsse, darunter eine Audioschnittstelle für die Sprachsteuerung.

Die Einheit erfasst die Sensorik-Daten der Maschine, analysiert sie in Echtzeit und kann anschließend entweder in den Prozess eingreifen oder an ein übergeordnetes System melden. Die Vorteile sind, dass die gesamte Rechenleistung lokal abläuft, das bedeutet das keine Informationen in die Cloud geladen werden und man stets die Hoheit über die Daten behält (vgl. INDUSTRYforward, 2022).

Abb. 14: KI-Modul CI 550 mit verschiedenen Schnittstellen (KEBA, 2023)



## 5. Virtuelle Assistenten

### 5.1. Was einen guten virtuellen Assistenten auszeichnet

Befehle werden in natürlicher Sprache formuliert und vom Assistenten aufgenommen. Es muss kein bestimmtes Vokabular erlernt werden, man drückt sich aus, als würde man mit einem anderen Menschen reden. Die Antwort kommt ohne Verzögerung, ebenfalls in gesprochener Sprache zurück.

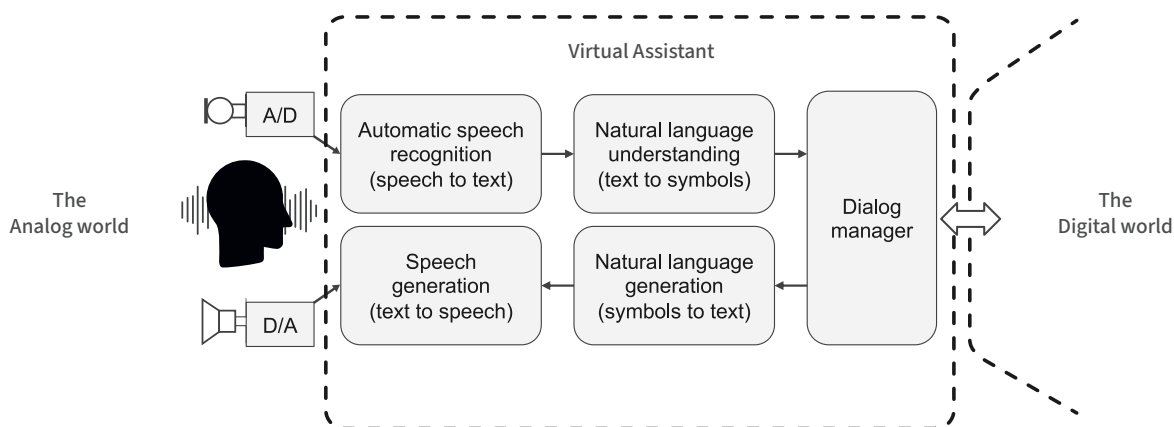


Abb. 15: Übersicht aller Komponenten, ob physisch oder digital, eines virtuellen Assistenten (vgl. Pieraccini, 2021, p. 8).

Wirklich intelligent ist der Assistent, wenn zusätzliche Informationen, die nicht im ursprünglichen Befehl enthalten waren, mit in den Entscheidungsprozess eingebunden werden. Ein Kontext bzw. früheres Verhalten beeinflussen die aktuelle Antwort. Schließlich soll der Assistent in der Lage sein, auf andere Anwendungen zuzugreifen und Aktionen auszuführen. Wichtig für gelungene Sprachinteraktion mit einem Voice-User-Interface sind nicht nur die einzelnen Teilbereiche, sondern vor allem wie gut sie im Zusammenspiel funktionieren (vgl. Budi & Laubheimer, 2018).

Siri von Apple, Google Assistent und Alexa von Amazon sind virtuelle Assistenten (VA), die einem bei alltäglichen Aufgaben helfen können. Sie sind in die verschiedensten Geräte integriert und man kann quasi jederzeit mit dem VA sprechen. Unterwegs könnte man mit seinem Smartphone, der Smartwatch, den Kopfhörern oder dem Auto reden. Zuhause steht der Smart Speaker, ein PC oder ein Tablet zur Verfügung.

Virtuelle Assistenten gibt es von vielen verschiedenen Anbietern. Die oben genannten drei beherrschen einen Großteil des Marktes und ähneln sich in vielerlei Hinsicht. Gemeinsam haben sie zum Beispiel, dass sie alle Cloud-basiert sind, also ständig mit dem Internet verbunden sein müssen, um korrekt zu funktionieren. Am besten arbeiten sie innerhalb ihres eigenen digitalen Ökosystems. Der Google Assistent und Siri profitieren natürlich davon, dass sie Teil eines Betriebssystems sind. Besitzt man ein iPhone, wäre es klug, zuhause einen Homepod zu haben und auf einem Mac zu arbeiten. So kann Siri auf viele Apps zugreifen und zum Beispiel Termine erstellen, die automatisch auf allen Geräten synchronisiert werden (vgl. Kabel, 2020, p. 69).

## 5.2. Virtuelle Assistenten (VA)

### 5.2.1. Siri

Siri war 2011 der erste virtuelle Assistent auf einem Smartphone. Sie besaß schon damals eine breite Palette an Funktionen. Man konnte nach Geschäften in der Nähe suchen, das Wetter vorhersagen lassen und Anrufe starten. Außerdem verfügt sie über ein gewisses Maß an Persönlichkeit und Humor. Gerade das war eines der beliebtesten Features (vgl. Pieraccini, 2021, p. 3f). Zum Beispiel wurde die Frage: „Was ist der Sinn des Lebens?“ meist mit 42 beantwortet. Eine Anspielung auf den Roman *Per Anhalter durch die Galaxis* aus dem Jahr 1979 von Douglas Adams, der Steve Jobs besonders gefiel (vgl. Bennett, 2016).

Im September 2014 führte Apple „Hey Siri“ als Weckwort ein. Ab sofort konnte man Siri auf Zuruf aktivieren, ohne zuerst den Home-Button drücken zu müssen.

Zukünftig soll auch nur „Siri“ genügen. Durch das Entfernen des „Hey“ könnten aufeinanderfolgender Anfragen erleichtert werden (vgl. Warren, 2023).



Abb. 16: Siri Logo (Apple, 2023)



Abb. 17: Alexa Logo (Amazon Alexa, 2023)

### 5.2.2. Alexa

Alexa von Amazon war der erste virtuelle Assistent, der in einem eigenständigen Gerät verbaut war, dem Amazon Echo. Es gibt mehrere Gründe für den Namen Alexa. Erstens ist Alexa ein seltener Name und ein Wort, das selten in amerikanischen Haushalten fällt. Zweitens machte der harte Konsonant „x“ es der Spracherkennung leichter zu erkennen, wann sie angesprochen wird. Drittens ist der Name inspiriert von der Bibliothek in Alexandria (vgl. Strenger & Kennedy, 2020).

Alexa verfügt über alle Basisfunktionen, die ein VA haben muss. Mithilfe von Skills ist es möglich Alexa seinen Bedürfnissen anzupassen. Es gibt Skills, die beim Kochen helfen, Nachrichten vorlesen oder über aktuelle Sportergebnisse informieren. Anfang 2019 gab es ca. 60.000 verfügbare Skills in den USA, weitere 8000 in Deutschland. Diese Zahlen haben sich im Vergleich zum Vorjahr teilweise mehr als verdoppelt (vgl. Kinsella, 2019).

Amazons Strategie als der größte Online-Händler der Welt war es nicht, wie ein Hardware-Unternehmen einfach möglichst viele Echo-Einheiten zu verkaufen, sondern Kunden zu mehr Einkäufen auf der Amazon-Plattform zu verleiten. Ein internes Dokument fasst diese Taktik gut zusammen: „Wir wollen Geld machen, indem Menschen unsere Geräte benutzen, nicht wenn sie sie kaufen.“ Ein Geschäftsmodell das seit nun 10 Jahren nicht rentabel ist. Ende 2022 sorgten große Entlassungswellen im Geschäftsbereich Devices für Aufsehen (vgl. Kim, 2022).

Das zugrundeliegende Problem dürfte sein, dass zu wenige Menschen genug Vertrauen in Alexa haben, dass sie ihr ihr Geld anvertrauen. Selbst bei trivialen Alltagsgegenständen möchte man sehen, wofür man sein Geld ausgibt.

Trotzdem ist Amazon Spitzenreiter im Bereich Smart-Home. Laut David Limp (Senior Vice President von Amazon) gab es bereits 2018 28.000 verschiedene Smart-Home Geräte, die mit Alexa kompatibel sind (vgl. Bohn, 2019).

### 5.2.3. Google Assistant

Google hat seine eigene Flotte an Geräten, die mit dem Google Assistant kompatibel sind. Seine Stärke basiert jedoch auf den Milliarden von Suchanfragen, die täglich in die Suchmaschine eingegeben werden. Google hat viele Informationen, wie sich Menschen ausdrücken und Fragen stellen. Dementsprechend gut ist der Assistant in der Informationsbeschaffung. Daher kommt auch seine Fähigkeit, besonders gut Gesprächen folgen zu können (vgl. Brownlee, 2022).

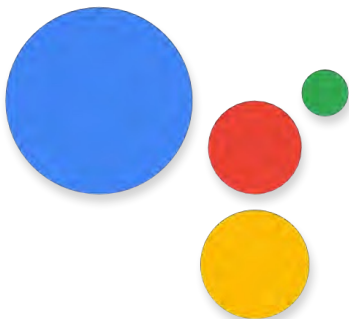


Abb. 18: Google Assistant Logo (Google, 2023)

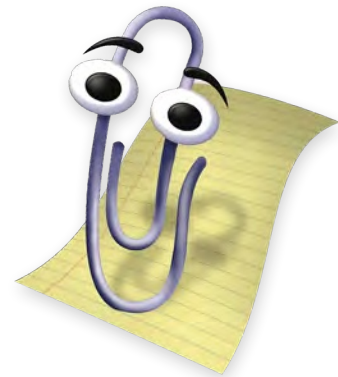


Abb. 19: Clippy (fandom.com, 2015)

### 5.2.4. Clippy

Es ist wichtig, dass die Erwartungen der Nutzer mit den tatsächlichen Fähigkeiten des Assistenten übereinstimmen. Probleme entstehen, wenn die Erwartungen die Fähigkeiten weit übersteigen, dies führt zwangsläufig zu einem frustrierenden Nutzererlebnis.

Bestes Beispiel dafür war Clippy, er versprach zu viel und konnte zu wenig. Außerdem wirkte er wie ein soziales Wesen mit zwei Augen, einem Körper und er sprach in ganzen Sätzen. Leider fehlte es ihm etwas an Kompetenz und Selbsteinschätzung, was dazu führte, dass er als nervig in Erinnerung blieb (vgl. Shedroff & Noessel, 2012, p. 193).

### 5.2.5. Cortana

Cortana ist leider in Österreich zurzeit nicht verfügbar. Sie hat einige Features, die sie von anderen Assistenten abhebt. Es können sechs verschiedene Stimmlagen imitiert werden. Sie kann daher rücksichtsvoll, sensibel, aufgeregt, verlegen, nachdenklich und entschuldigend klingen. In Kombination mit 18 verschiedenen visuellen Feedbacks kann sie auf viele Situationen reagieren (vgl. Bernhard, 2015). Die Verwendung der Farbe Blau bedient das gängige Bild des SciFi-Interface-Stereotyps. Der Name Cortana verweist auf das gängige Stereotyp, dass virtuelle Assistenten oft weiblich sind.

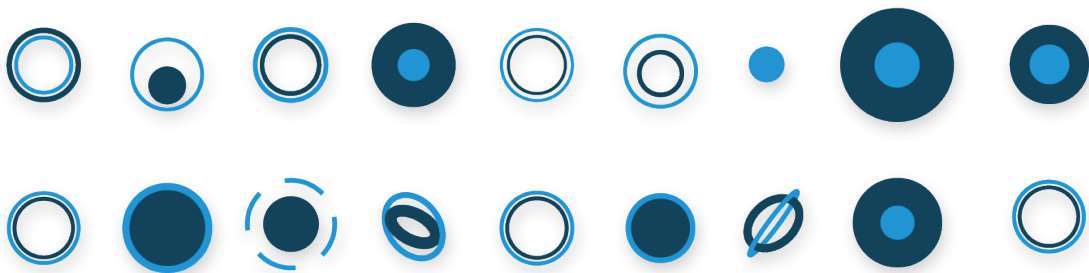


Abb. 20: Cortanas Emotionen (Bernhard, 2015)

### 5.3. Usability und Utility Gap

Die Ergebnisse einer Tagebuchstudie der Nielsen Norman Group aus dem Jahre 2018 zeigt deutlich, wie viel Potenzial es bei der Nutzung von intelligenten Assistenten noch gibt. In der Studie führten 12 Teilnehmer ein Tagebuch, in dem sie alle Bedürfnisse festhielten, die ein idealer Sprachassistent erfüllen sollte. Im Anschluss wurde untersucht, ob die bestehenden Sprachassistenten diese Bedürfnisse erfüllen können (vgl. Budiu & Whitenton, 2018).

Von den in während der Tagebuchstudie erfassten Anfragen hätten 41 % vollständig und 21 % zu-

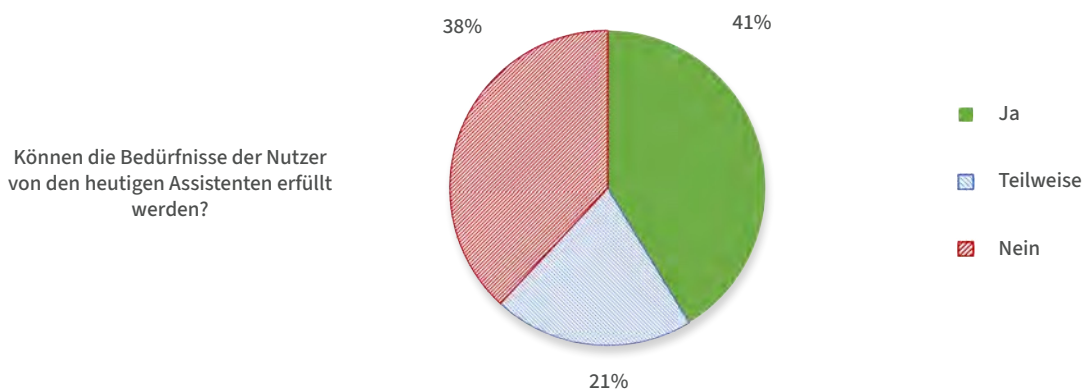


Abb. 21: Ergebnis Tagebuchstudie Nutzeranfragen (Budiu & Whitenton, 2018)

mindest teilweise von mindestens einem der derzeit verfügbaren Assistenten erfüllt werden können. Dies zeigt, dass der **tatsächliche Nutzen** (realized usefulness) aktueller intelligenter Assistenten recht gering ist, insbesondere bei komplexen Aufgaben.

Der **potenzielle Nutzen** (potential usefulness) ist viel höher, wie aus der langen Liste an Bedürfnissen der Benutzer hervorgeht. Der potenzielle Nutzen wird durch die gesamte Fläche unterhalb der obersten Linie im Diagramm dargestellt.

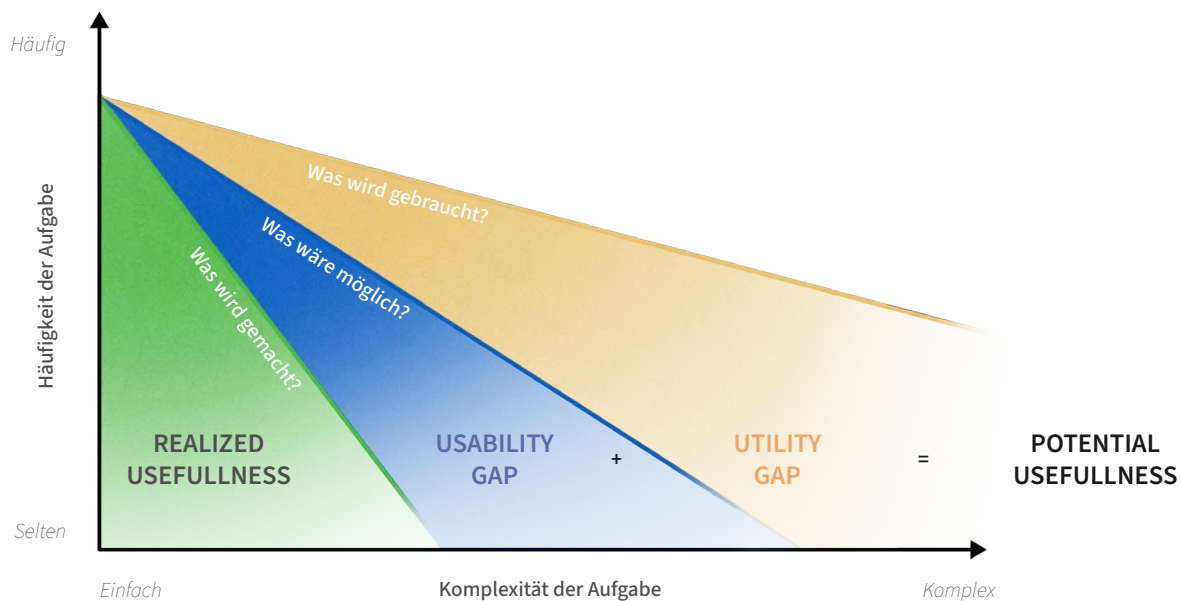


Abb. 22: Großes Potenzial bei der Nutzung von intelligenten Assistenten (Budiu & Whittenton, 2018) Layout und Farben wurden zur besseren Verständlichkeit verändert

Allerdings gibt es ein **Defizit bei der Benutzerfreundlichkeit** (usability gap). Dieses Defizit wird durch bestehende Funktionen verursacht, die zu schwierig sind, sie zu benutzen. Und es gibt einen Mangel des **praktischen Nutzens** (Utility-Gap). Dieser Mangel entsteht durch Funktionen, die noch nicht existieren. Diese beiden verschlingen den größten Teil des **potenziellen Nutzens**.

Beide Lücken müssen geschlossen (oder zumindest erheblich verkleinert) werden, damit intelligente Assistenten wirklich nützlich sind (vgl. Budiu & Whittenton, 2018). Diese Differenz zeigt, wie niedrig unsere Erwartungen an diese intelligenten Systeme noch sind.

## 5.4. Fazit zu intelligenten Assistenten

Es ergibt sich ein paradoxes Bild. Die Assistenten entwickeln sich stetig weiter, dennoch nutzen wir sie für die immer gleichen Aufgaben. Vor allem in der Fähigkeit Konversationen zu führen, wurden große Fortschritte gemacht, trotzdem sind 86% aller Interaktionen nach der ersten Antwort zu Ende. Warum ist das so?

Ein Grund, warum viele Interaktionen mit VA recht kurz sind, ist das mangelnde Kontextwissen des VA.

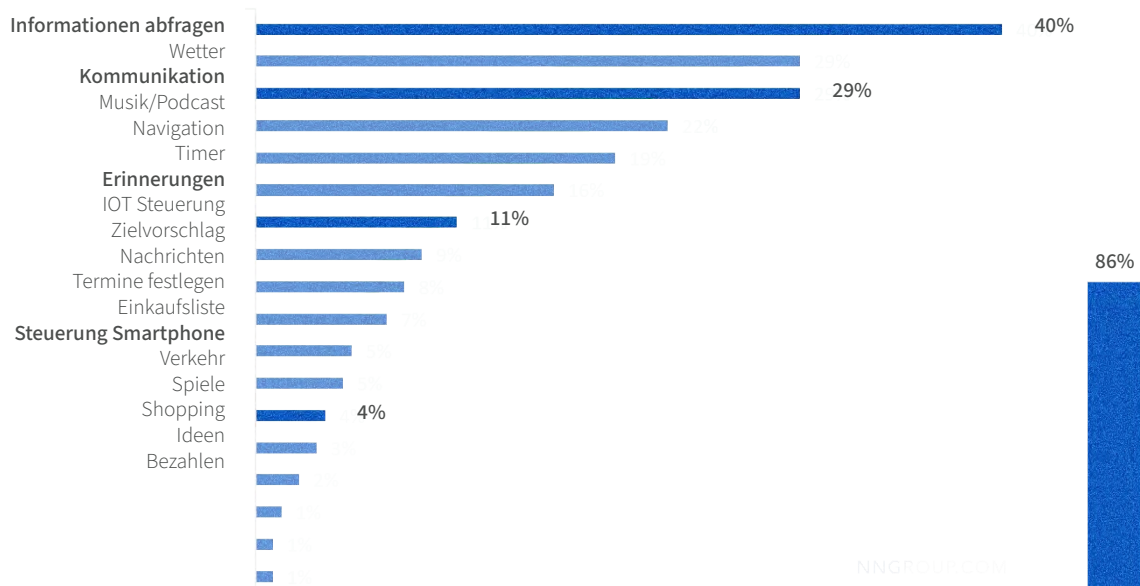


Abb. 23: Beliebteste Aufgaben für Sprachassistenten (Budiu & Whinton, 2018)

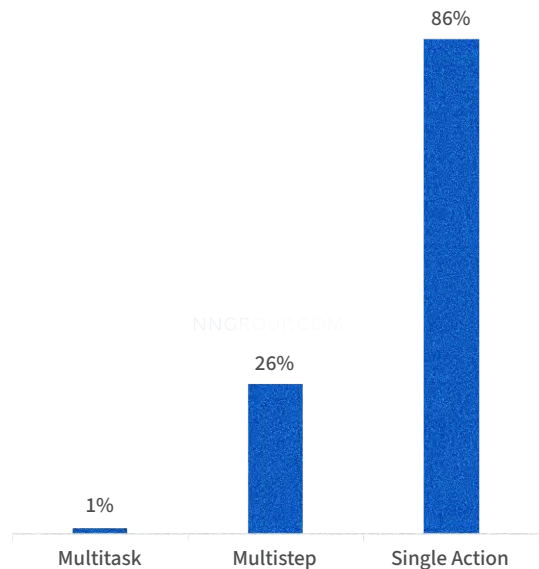


Abb. 24: Komplexität der Interaktion (Budiu & Whinton, 2018)

Eine effektive Kommunikation kommt erst zustande, wenn auf eine Frage des Benutzers nicht nur eine sinnvolle Antwort kommt, sondern bei Bedarf auch eine gut formulierte Gegenfrage (vgl. Pieraccini, 2021, p. 167). Das hilft dabei, den Kontext besser zu verstehen und den Benutzer nicht unnötig lange auszufragen. Zeigt der VA zu viel Initiative, kann es schnell unangenehm und nervig werden. Außerdem sind viele mit einer kurzen Antwort schon zufrieden, sofern es die Richtige ist.

## 6. Mental Model

### 6.1. Definition

Bei einem Mental Model handelt es sich um die Vorstellung des Nutzers über die Funktionsweise eines Systems. Im Idealfall entspricht das Mental Model genau der Realität. Besonders wichtig sind Mental Models, wenn Personen zum ersten Mal mit neuer Technologie konfrontiert sind. Man versucht automatisch zu verstehen, wie es funktioniert damit man korrekt damit umgeht. Je länger man mit einem Produkt arbeitet, desto mehr lernt man darüber und das Mental Model entwickle sich weiter (vgl. Sharp, et al., 2019, pp. 122,123). Man brauche eine innere Geschichte, die das Verhalten und das Erscheinungsbild eines Geräts sinnvoll zusammenführt (vgl. Norman, 1998, p. 177).

Da jeder Mensch unterschiedliche Erfahrungen und Wissen mitbringt, hat jeder sein eigenes Mental Model. Diese sind oft unvollständig, leicht verwechselbar und basieren auf unangemessenen Analogien oder veralteten Annahmen. Als Folge davon finden viele Menschen es schwierig, ein Problem zu identifizieren oder zu lösen, da oftmals die richtigen Worte fehlen es zu beschreiben (vgl. Sharp, et al., 2019, p. 123).

*Es ist Aufgabe des Interfaces, dies [das Mental Model] mit einem guten Informationsdesign möglichst schnell und umfassend zu kommunizieren.  
(Dorau, 2011, p. 23)*

Die Gestaltung des Interfaces kann dem Nutzer dabei helfen, die grundlegende Natur des Systems besser zu verstehen

## **6.2. Mental Models von Virtuellen Assistenten**

Eine Studie der NN-Group aus 2018 hat die aktuellen VAs Siri, Alexa und Google Assistant in folgende drei Mental Models eingeteilt (vgl. Nielsen Norman Group, 2019).

### **6.2.1. Interface**

Viele betrachten den VA als Schnittstelle, entweder zum Internet, ihrem Telefon, ihrem Smart Home oder einer Kombination daraus. Dies ist eine sehr gesunde und realitätsnahe Vorstellung.

### **6.2.2. Handy Helper**

Einige sehen den VA als jemand, der immer zur Stelle ist und ihnen kurz etwas abnimmt und Arbeit erspart. Man muss nicht mehr selbst von der Couch aufstehen und das Licht abdrehen, Siri macht das gerne. Es ist ein netter Gedanke, immer jemanden dabei zu haben der schnell und unkompliziert Kleinigkeiten erledigen kann. Diese Denkweise kann dazu führen, dass man sich zu sehr auf die Fähigkeit des Assistenten verlässt und mit der Zeit unvorsichtiger wird.

### **6.2.3. Wissensspeicher**

Manche stellen sich den VA als große Sammlung von Wissen vor. Sie meinen, der Assistent sei „*klüger als ein Mensch*“ und „*weiß alles*“. Sie verfallen einer Fehlvorstellung und setzen Intelligenz mit dem Zugang zu Fakten gleich.

## **7. Smart Home Speaker**

Smart Speaker sind die physischen Geräte für virtuellen Assistenten. Zumeist stehen sie im Wohnzimmer oder der Küche, sind also hauptsächlich für den Einsatz im privaten Bereich konzipiert. Im Folgenden wird genauer auf die Form und Materialität der einzelnen Speaker eingegangen und einige Sonderformen hervorgehoben.

## 7.1. Amazon Echo

Amazon's Echo war der erste richtige Smart Speaker. Bereits die erste Generation hatte einige Features, die bis heute Standard sind. Die Grundform ist ein hoher zylindrischer Körper. Die untere Hälfte des Mantels ist mit einem regelmäßigen Lochmuster perforiert. So wird deutlich, dass es sich um einen Lautsprecher handelt. Auf der Deckfläche befinden sich zwei Tasten, die Aktionstaste um Echo zu Aktivieren oder den Wecker abzustellen und die Mikrofon-Aus-Taste um das Mikrofon zu deaktivieren. Ein etwa 10mm breiter Ring am oberen Ende des Zylinders ist drehbar und stellt die Lautstärke ein.

Am auffälligsten ist der Lichtring am äußeren Rand der Deckfläche. Die weiteren Bedeutungen der Farben werden im nächsten Kapitel beschrieben.

Im Laufe der Zeit gab es den Echo in diversen Größen und Formen. Am beliebtesten ist der Echo Dot. Die ersten beiden Generationen (März bzw. Oktober 2016) sind flache Zylinder mit einem Kunststoffgehäuse und harten Kanten. Die 3. Generation hat größere Radien und wirkt dadurch weicher und rundlicher, auch der Überzug aus einem feinen Textil verstärkt diesen Eindruck. Seit der 4. Generation (2020) ist die Grundform eine Kugel und der Lichtring an den unteren Rand gewandert. Weiters gibt es nun ein segmentiertes LED-Display hinter dem Stoffbezug. Dies kann zum Beispiel die Zeit anzeigen oder Text wie auf einem Terminal durchlaufen lassen.

Amazon waren auch die Ersten, die einen Bildschirm an ihren Smart Speakern anfügten. Dies schien 2017 der logische nächste Schritt zu sein. Die Echo Show Reihe zeichnet sich durch Displays in diversen Größen aus. Über eine eingebaute Kamera ist es möglich Videoanrufe zu führen.



Abb. 25: Amazon Echo  
1. Generation  
(Carroll, 2015)



Abb. 26: Amazon Echo  
5. Generation  
(Amazon, 2023)

## 7.2. Google Home

Auch Google fügte dem Speaker einen Bildschirm hinzu. Vom Google Nest Hub gibt es bereits eine 2. Generation. Der erfolgreichste Smart Speaker von Google ist der Nest Mini. Der intelligente Lautsprecher ist pillenförmig und hat die Größe des Amazon Echo Dot 3. Generation. Unter dem Stoffbezug befinden sich 4 LEDs, die reduzierteste Form von Licht bei allen Smart Speakern.

Die Stummtaste ist ein kleiner mechanischer Knopf, den man zur Seite schieben kann, er befindet sich neben dem Stromkabelanschluss.

Abb. 27: Google Nest Mini  
(Google, 2023)



Abb. 28: Apple HomePod mini (links) und der HomePod (rechts)  
(Apple, 2023)



## 7.3. Apple HomePod

Erst 2018 brachte Apple seinen ersten Smart Speaker auf den Markt. Der HomePod war im Vergleich zu andern Smart Speakern sehr groß und verhältnismäßig schwer. Der Fokus lag auf Audio-Qualität. Der hohe Preis und die Tatsache, dass der Lautsprecher ausschließlich im eigenen digitalen Ökosystem funktioniert, war wahrscheinlich der Grund, dass der Verkauf wieder eingestellt wurde. Paradoxerweise trifft dies ebenfalls auf den neuen HomePod 2 (2022) zu. Eine der wenigen Änderungen waren zusätzliche Sensoren, die Temperatur und Luftfeuchtigkeit messen können.

Der HomePod ist ebenfalls zylinderförmig, hat aber deutlich größere Radien als die Konkurrenz von Amazon und eine leicht bombierte Deckfläche. Die Oberfläche ist überzogen mit einem nahtlosen Netzgewebe, das die Akustik nicht beeinflussen soll (vgl. Apple, 2023). Die Oberseite ist eine touch-sensitive Leuchtfläche. In diesem Bereich zeigt ein farbiges LED-Licht an, wenn Siri aktiv ist oder die Wiedergabe von Musik oder anderen Inhalten gesteuert wird.

Der deutlich preiswertere Nachfolger, der HomePod mini (2020) ist um einiges erfolgreicher. Die Grundform ist wie beim Echo Dot 4. Gen eine Kugel. Auf der Oberseite befindet sich die gleiche berührungsempfindliche Steuerungsoberfläche wie beim HomePod.

## 7.4. Jibo

Jibo war bei seiner Einführung 2014 der erste soziale Roboter für zuhause. Er wurde von Cynthia Breazeal, einer MIT-Professorin entwickelt. Jibos Körper besteht aus drei Teilen: Becken, Oberkörper und Kopf. Das Becken und der Oberkörper sind ungleichmäßig geformte Ringe, und der Kopf ist eine Kugel mit abgeflachter Vorderseite. Darin befinden sich ein HD Touchdisplay und zwei Kameras. Die Körperteile können sich gegeneinander drehen, so kann der Roboter eine unendliche Anzahl von Haltungen einnehmen und stets Blickkontakt mit dem Benutzer halten. Während Jibo spricht, bewegt er seinen Körper entsprechend mit und nutzt so nonverbale Kommunikationsmittel. An der Basis befindet sich ein umlaufender LED-Ring (vgl. Pieraccini, 2021, pp. 212-220).



Abb. 29: Jibo (Detweiler, 2014)



Abb. 30: Olly (Emotech, 2017)

## 7.5. Olly

Eine Sonderform unter den Smart Speakern ist Olly. Die Grundform ist ein Torus, der auf einer flachen, kegelförmigen Basis ruht. Er kann sich aber auch aufrichten und 360° drehen. Vorgestellt wurde er 2017 als der erste Assistent mit einer sich entwickelnden Persönlichkeit. Interessant ist seine Eigenschaft, dass er sich oft proaktiv meldet.

Das letzte Update auf Indigogo.com gab es im Juli 2019.

## 7.6. Fazit zu Smart Speaker

Die äußere Gestaltung der Smart Speaker von Apple, Amazon und Google ähneln sich in vielerlei Hinsicht. Sie sind klein, kompakt und entweder zylindrisch oder kugelförmig. Die Oberfläche besteht aus einem textilen Überzug, der meist schwarz oder grau ist. So weiß man, dass dahinter ein Lautsprecher sitzt und das Gerät Audio-Inhalte wiedergeben kann.

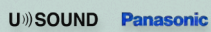
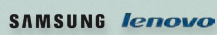
Steuerelemente wie Tasten oder Touch-Flächen sind erst bei genauerem Hinsehen erkennbar. Im Alltag muss man das Gerät nur selten berühren. Eine Basis aus weichem Kunststoff verhindert verrutschen und Kratzer auf Möbeloberflächen.

Ziel ist es, in der Umgebung unterzugehen und sich erst bei Aktivierung abzuheben. Die neutralen und zurückhaltenden Formen helfen dabei, sich in fast alle Einrichtungsstile nahtlos einzufügen.

Die runde Form ermöglicht eine gleichmäßige Schallverteilung im Raum, was zu einem besseren Hörerlebnis führt. Eine umlaufende Verteilung der Mikrofone führt zu verbesserter Spracherkennung und Sprecherlokalisierung.



Marken



2014



Echo  
1<sup>st</sup> Gen



Gatebox

2016



Echo  
2<sup>st</sup> Gen



Echo Dot  
1<sup>st</sup> Gen



Echo Dot  
2<sup>st</sup> Gen



Echo Tap



Echo Spot



Echo Look



Home

2017



Echo Show  
1<sup>st</sup> Gen



Echo Plus  
1<sup>st</sup> Gen



Home Mini  
1<sup>st</sup> Gen



Home Max



Ollly



Lenovo  
Smart Hub

2018



HomePod  
1<sup>st</sup> Gen



Echo  
3<sup>rd</sup> Gen



Echo Dot  
3<sup>rd</sup> Gen



Echo Show  
2<sup>nd</sup> Gen



Echo Plus  
2<sup>nd</sup> Gen



Nest Hub  
1<sup>st</sup> Gen



SC-GA10  
Panasonic



One  
Sonos



Portal  
facebook

Abb. 31: Übersicht Smart Speaker

2019



HomePod Mini



Echo Dot Clock 3rd Gen



Show 5 1st Gen

Show 8



Echo Flex



Nest Mini 2nd Gen

Nest Audio



Galaxy Home



Portal+ facebook

TV facebook

2020



Echo 4th Gen



Echo Dot Clock 4th Gen



Echo Show 10



Echo Studio

Echo Flex



Echo Astro



Nest Hub 2nd Gen



Acer Halo 1st Gen

Galaxy Home Mini



USound Proteus



facebook Go

NIO

2021

2022



Echo Dot Clock 4th Gen



Echo Show 15



Halo Rise



Xiaomi

2023



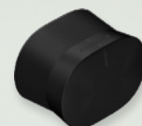
HomePod 2nd Gen



Echo Pop



Acer Halo 1st Gen



Sonos Era

## 8. Lichtkonzepte von Smart Speakern

### 8.1. Amazon Echo

Alle Amazon Echo Geräte nutzen einen umlaufenden LED-Ring als Statusanzeige. Jede Farbe hat seine eigene Bedeutung. Wenn Alexa zuhört oder spricht, leuchtet der Ring Blau. Der Ring kann in neun unterschiedlichen Modi leuchten.

Echo bedient hier zwei Gestaltungselemente, die sich von Science-Fiction Interfaces ableiten. Zum einen leuchtet der Ring in Blau, zum anderen entsteht durch die Reflexion auf der Oberfläche, auf der das Gerät steht, ein schimmernder Bereich.

Eine Besonderheit des Amazon Echos ist die Positionserkennung des Sprechers. Spricht man Alexa an, so zeigt ein heller Bereich am Lichtring an, aus welcher Richtung die Stimme kommt. Es gibt dem äußerlich leblosen Gerät ein anthropomorphes Verhalten, vergleichbar mit Blickkontakt (vgl. Pieraccini, 2021, p. 5).



**Abb. 32:** Lichtring und 7-Segment-Anzeige auf dem Amazon Echo Dot 5.Gen (Amazon, 2023)



**Abb. 33:** 4 LED auf dem Google Nest Mini (Google, 2023)



**Abb. 34:** Lichtfläche auf dem Apple HomePod Mini (Apple, 2023)

### 8.2. Google Nest Mini

Der Smart Speaker von Google verfügt nur über 4 einzelne LED-Punkte. Bei Aktivierung scheinen sie durch die Oberfläche des Lautsprechers.

Das Blinkmuster der Lichter variiert je nach Situation. In den meisten Fällen leuchtet es weiß, es gibt jedoch 11 unterschiedliche Modi. Andere Farben zeigen sich nur beim Hochfahren, wenn das Gerät stumm geschaltet ist oder bei Update Problemen.

### 8.3. Apple HomePod

Der HomePod wie auch der HomePod Mini haben im Vergleich zu anderen Smart Speakern einen recht großen Bereich für Statusanzeigen. Die ganze Deckelfläche steht der Bespielung zur Verfügung. Obwohl die bespielbare Fläche recht groß ist, gibt es verhältnismäßig wenige unterschiedliche Anzeigen. Es gibt 6 (beim HomePod Mini 7) unterschiedliche Lichtsignale.

Die wichtigsten Funktionen übernimmt ein rotierendes, mehrfarbiges Licht. Die Animation ähnelt der Darstellung von Siri auf dem iPhone. Das bunte Lichtspiel zeigt an, dass Siri auf einen Sprachbefehl wartet, ihn verarbeitet oder gerade die Antwort gibt. Das dynamische Farbspektrum vermittelt den Eindruck, dass sich Siris *Gedanken* überschlagen.

## 8.4. Jibo

Das große Display im Kopf zeigt meistens eine weiße Kreisfläche, durch gelegentliches Blinzeln und Bewegungen während dem Sprechen, merkt man, dass es sich um ein Auge handelt. Bei Bedarf werden auch Symbole, Animationen oder Text angezeigt.

Der LED-Ring ist normalerweise ausgeschaltet, wenn man Jibo anspricht, leuchtet er kurz blau. Treten Fehler wie der Verlust der WiFi-Signals auf, leuchtet er rot (vgl. Pieraccini, 2021, p. 218).



Abb. 35: Jibo zeigt Symbole aus dem Menü (Davies, 2017)



Abb. 36: bunter Lichtring des Olly (Emotech, 2017)

## 8.5. Olly

Auf der facettierten Oberfläche laufen bunte Animationen. Tagsüber ähnelt das Konzept dem von Siri, es ist ein wildes Aufflackern von Farbwellen, die die ganze Leuchtfläche einnehmen. Nachts ist das Farbenspiel ruhiger und leuchtet nur blau.

Wenn Olly aus dem Ruhezustand erwacht und spricht, kommen die bunten Wellen aus dem Zentrum des Torus. Während er zuhört und Informationen aufnimmt, laufen die Lichter von außen nach innen (vgl. emotech, 2018). Online finden sich keine genaueren Beschreibungen zur Bedeutung der verschiedenen Leuchtsequenzen.

## 8.6. Fazit und Gemeinsamkeiten der Smart Speaker

Die Vielfalt der Lichtsignale zeigt, dass unterschiedliche Ansätze der Gestaltung erfolgreich sein können. Die Hauptaufgabe ist jedoch immer, den aktuellen Status des Geräts darzustellen. Die eingesetzten Hauptfarben spiegeln oft die *Persönlichkeit* der virtuellen Assistenten wider. Alexa ist fortschrittlich und zukunftsorientiert, somit blau. Siri soll möglichst agil und lebensfroh wirken, weshalb sie stets in Bewegung ist und den ganzen Regenbogen nutzt. Der Google Assistent ist zurückhaltend, nüchtern und funktionsorientiert, daher reichen 4 LED. Werden zu viele Farbkombinationen eingesetzt, kommt es leicht zur Verwirrung des Nutzers. Nachzuforschen, was die angezeigte Farbe bedeutet, ist mühselig und frustrierend.

Eine weitere Gemeinsamkeit ist, dass die eingesetzten Farben sich nur selten vermischen. Sie treten meist nur in Reinform auf, mit der Ausnahme von Siri.


Auch wenn sie alle ihr eigenes Lichtkonzept verfolgen, gibt es einige Gemeinsamkeiten:

- Rot und Orange deutet meist auf ein Problem oder das deaktivierte Mikrofon hin.
- Bewegen sich Lichter im Kreis, wird etwas gesucht. Zum Beispiel die Internetverbindung oder Updates.
- Mit pulsierenden Lichtern versucht man Aufmerksamkeit auf sich zu ziehen.



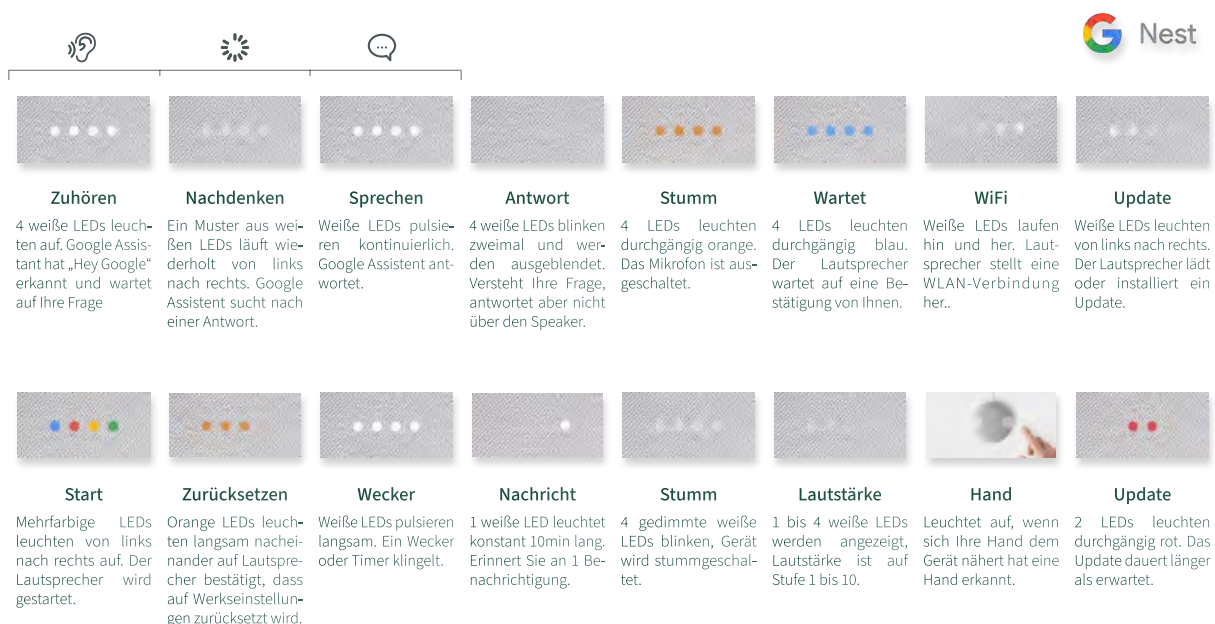
**alex**

Zuhören	Nachdenken	Sprechen	Stumm	Setup	Anruf	Nicht Stören	Lautstärke	Nachricht
Cyan auf blau, bedeutet das Alexa zuhört. Der helle Bereich zeigt in die Richtung des Sprechers	Abwechseln blau und cyan bedeutet dass Alexa gerade nachdenkt	Blau, cyan pulsierend bedeutet dass Alexa gerade spricht	Durchgehendes rotes Licht zeigt an, dass das Mikrofon des Geräts getrennt ist und Alexa nicht zuhört	Orange im Kreis laufend Einrichtungsmodus oder versucht, eine Internetverbindung herzustellen	Pulsierendes grünes Licht bedeutet, dass du einen Anruf auf dem Gerät empfangst	pulsierend in Lila „Bitte nicht stören“-Modus ist aktiviert	Die Größe des Lichtrings zeigt die aktuelle Lautstärke	Ein langsames gelbes Aufleuchten alle paar Sekunden bedeutet, dass Alexa eine Benachrichtigung hat



**Apple**

Hören-Sprechen-Denken	Aktualisierung	Audio	Anruf	Problem	Zurücksetzen	Verbunden	Wecker
Mehrfarbiges rotierendes Licht. Siri hört zu, denkt nach oder antwortet.	Weißes rotierendes Licht. Der HomePod schaltet sich ein oder die Software wird aktualisiert.	Weißes Dauerlicht. Der HomePod gibt Audioinhalte wieder.	Grünes pulsierendes Licht. Du hast einen Anruf an den HomePod weitergeleitet.	Weißes blinkendes Licht. Gibt ein Problem, trenne den Pod von der Stromversorgung	Rotes rotierendes Licht. Du setzt den HomePod zurück.	Orangefarbenes blinkendes Licht. Der HomePod mini ist mit einem Computer verbunden	Weiß pulsierend. Ein Wecker läutet



**G Nest**

Zuhören	Nachdenken	Sprechen	Antwort	Stumm	Wartet	WiFi	Update
4 weiße LEDs leuchten auf. Google Assistant hat „Hey Google“ erkannt und wartet auf Ihre Frage	Ein Muster aus weißen LEDs läuft wiederholt von links nach rechts. Google Assistant sucht nach einer Antwort.	Weißer LEDs pulsieren kontinuierlich. Google Assistant antwortet.	4 weiße LEDs blinken zweimal und werden ausgeblendet. Versteht Ihre Frage, antwortet aber nicht über den Speaker.	4 LEDs leuchten durchgängig orange. Das Mikrofon ist ausgeschaltet.	4 LEDs leuchten durchgängig blau. Der Lautsprecher wartet auf eine Bestätigung von Ihnen.	Weißer LEDs laufen hin und her. Lautsprecher stellt eine WLAN-Verbindung her..	Weißer LEDs leuchten von links nach rechts. Der Lautsprecher lädt oder installiert ein Update.
Mehrfarbige LEDs leuchten von links nach rechts auf. Der Lautsprecher wird gestartet.	Orange LEDs leuchten langsam nacheinander auf Lautsprecher bestätigt, dass auf Werkseinstellungen zurückgesetzt wird.	Weißer LEDs pulsieren langsam. Ein Wecker oder Timer klingelt.	1 weiße LED leuchtet konstant 10min lang. Remindert Sie an 1 Benachrichtigung.	4 gedimmte weiße LEDs blinken, Gerät wird stummgeschaltet.	1 bis 4 weiße LEDs werden angezeigt, Lautstärke ist auf Stufe 1 bis 10.	Leuchtet auf, wenn sich Ihre Hand dem Gerät nähert hat eine Hand erkannt.	2 LEDs leuchten durchgängig rot. Das Update dauert länger als erwartet.

Abb. 37: Übersicht Modi Lichtsignale von Smart Speakern

### 8.6.1. Effektivität und Verständlichkeit

Eine Studie von Sahiti Kunchay und Saeed Abdullah aus dem Jahr 2021 zeigt auf, wie schlecht die verschiedenen Lichtsignale von Benutzern erkannt werden. Nur 37% aller Lichtsignale wurden korrekt identifiziert. Woher die Befragten kommen und wie lange sie das Gerät bereits verwenden, hatte keinen Einfluss auf die Verständlichkeit (vgl. Kunchay & Abdullah, 2021, p. 6).

Befragte die angaben normalerweise nie auf die Lichter zu achten, erkannten nur 33% richtig. Dies verdeutlicht den kleinen, aber signifikanten Beitrag, den die Lichter zur Verständlichkeit leisten (vgl. Kunchay & Abdullah, 2021, p. 6).

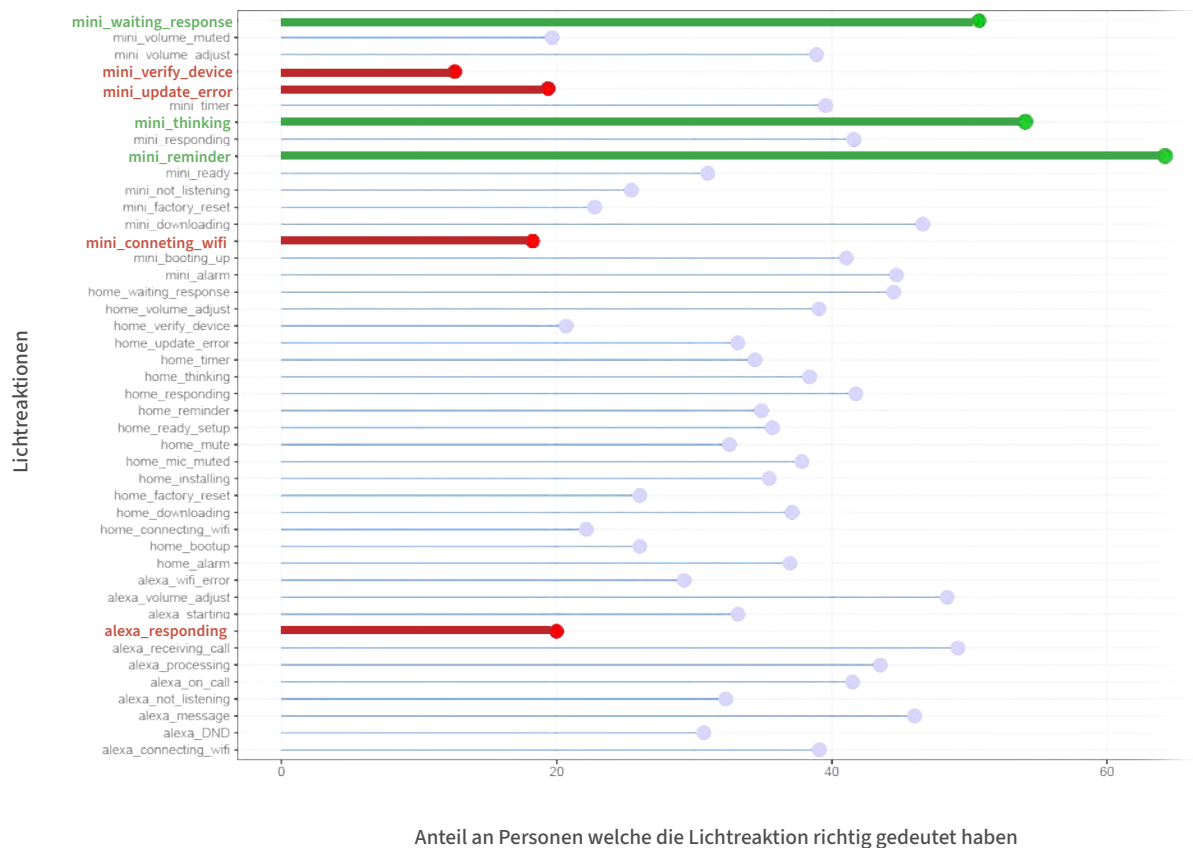


Abb. 38: Bewertung der Wirksamkeit und Interpretierbarkeit von Lichtreaktionen bei Smart Speakern (Kunchay & Abdullah, 2021, p. 8)

Überraschenderweise ist die Lichtreaktion, wenn Alexa antwortet (*alexa\_responding*), unter den vier, die am wenigsten erkannt wurden. Die Leuchtmuster mit den höchsten Wiedererkennungswerten sind vom Google Nest Mini: Benachrichtigungserinnerung (*mini\_reminder*), Mini denkt nach (*mini\_thinking*) und Mini wartet auf ihre Antwort (*mini\_waiting\_response*) (vgl. Kunchay & Abdullah, 2021, p. 8).

### 8.7. Vorgeschlagen Lösungsansätze

Die Studie schlägt folgende fünf Lösungsansätze vor, um die Effektivität und Verständlichkeit der Lichteffekte zu verbessern (vgl. Kunchay & Abdullah, 2021, p. 12).

### **8.7.1. Vereinfachung des Systems**

Sich nur auf ein paar wenige Lichtmodi zu beschränken, hilft dem Nutzer sich genauer zu fokussieren. Außerdem erspart man sich das Erlernen von vielen verschiedenen Mustern die evtl. nur selten gebraucht werden.

**Die drei wichtigsten Informationen die Nutzer wollen, wenn Sie auf den Smart Speaker blicken:**

- Lautstärkeänderungen
- Verbindungsprobleme
- Befehl wird gerade verarbeitet

**Die drei wichtigsten Situationen, in denen ein Nutzer auf den Smart Speaker blickt:**

- Nutzer spricht zum VA
- Der VA reagiert nicht
- Der VA hat einen Befehl missverstanden

### **8.7.2. Zusätzlichen Kontext liefern**

Sprachliche Beschreibungen der Lichteffekte helfen vor allem neuen Nutzern, die das System noch nicht kennen. Auch in Situationen, die selten vorkommen, hilft es, wenn zusätzlich sprachlich kommuniziert wird.

### **8.7.3. Personalisierung erlauben**

Den Nutzern erlauben Änderungen vorzunehmen, kann mehrere Vorteile bringen. Es nimmt zum einen Druck vom Designer, ein universell verständliches Lichtkonzept zu entwickeln. Weiters kann es Nutzer mit anderen Vorlieben ansprechen. Es würde Nutzern verschiedener digitaler Ökosysteme helfen, sich schneller anzupassen, da sie ihre bevorzugten Präferenzen übernehmen könnten.

### **8.7.4. Barrierefreiheit erhöhen**

Zukünftige Arbeiten sollen sich darauf konzentrieren, die Bedürfnisse von Personen mit Seh- und Hörbeeinträchtigungen mehr zu berücksichtigen. Eindeutigere Lichtsignale helfen Menschen mit Hörbeeinträchtigungen.

### **8.7.5. Aufbauen auf Erfahrungen**

Die Erfahrungen, die Menschen bereits mit anderen Smart Speakern gemacht haben, sollen in neue Konzepte einfließen. Beispielsweise der Einsatz der Farbe Rot, um Probleme anzuzeigen oder die Visualisierung der Lautstärke des Amazon Echo.

## **9. Emotionale Beziehungen zu VA**

Jibo und Olly sind zwei Beispiele für intelligente Assistenten, die menschlichen Emotionen interpretieren können bzw. darauf reagieren sollen. Beide Start-Ups scheiterten aber mit ihrem Produkt. Es gibt wahrscheinlich mehrere Gründe, warum diese Art von Assistenten (noch) nicht erfolgreich sind.

Anthropomorphismus bedeutet Objekten oder Tieren menschliche Eigenschaften zuzuschreiben. Zum Beispiel reden Menschen mit ihren Computern, als wären sie Menschen, behandeln Roboter-Staubsauger wie Haustiere und geben ihren mobilen Geräten süße Namen (vgl. Sharp, et al., 2019, p. 187).

*Simulated thinking may be thinking, but simulated feeling is never feeling, and simulated love is never love (Turkle, 2017).*

MIT Professorin und KI-Forscherin Sherry Turkle beschreibt das Problem in einem Artikel der Washington Post. Künstlich soziale Wesen wie Jibo oder Olli wissen genau, wie sie unsere Aufmerksamkeit und Empathie wecken. Fragt uns ein Roboter um Hilfe oder reagiert charmant auf eine Frage, entwickeln wir sofort eine emotionale Bindung zu dem Gerät. Doch egal wie ausdrucksstark oder mitfühlend ihre Gesichter sind, diese digitalen Begleiter können unsere Gefühle nicht verstehen (vgl. Turkle, 2017).

## 10. Anwendungsbeispiele für Sprachinteraktion im Auto

*The key challenge when designing a voice interaction for the car is helping the user avoid the temptation of looking at their phone while driving. (Pearl, 2016, p. 237)*

Die Ansprüche an einen virtuellen Assistenten in einem Fahrzeug sind anders als zuhause. Smart Speaker im privaten Wohnbereich sollen das tägliche Leben leichter und angenehmer machen. Im Auto übernehmen sie eine zentrale Rolle bei der Erhöhung der Verkehrssicherheit. 2021 war die Hauptursache (30%) für Unfälle Unachtsamkeit und Ablenkung (vgl. Statistik Austria, 2022, p. 43). Einige Automobilhersteller setzen auf ein Voice-only-Konzept, dabei sind keine sichtbaren Geräte vorhanden. Falls es doch ein physisches Interface gibt, sollte es unauffällig und nahtlos in den Fahrzeuginnenraum integriert sein, um keine unnötige Aufmerksamkeit auf sich zu ziehen.

### 10.1. Nomi von Nio

Nomi Mate ist der digitale Assistent in den Elektroautos der chinesischen Marke Nio. In der Mitte des Armaturenbretts befindet sich eine Halbkugel, in deren Schnittfläche ein Display eingebettet ist. Die Halbkugel kann sich drehen und so gezielt auf einen der Passagiere blicken. Man fühlt sich sofort wahrgenommen, wenn sich der Kopf dreht und einem beim Sprechen direkt ansieht. Durch Sprachprofile und selbst wählbare Weckwörter weiß Nomi Mate auch mit wem er gerade spricht.

Auf dem runden Display sind meist zwei Augen und ein zusätzliches Icon zu sehen. Mal trinkt er ein Glas Wasser, spielt Gitarre oder zeigt einen Daumen-Hoch-Emoji. Um dem Uncanny-Valley (siehe Kap. 11.2) aus dem Weg zu gehen, wird kein Mund dargestellt (vgl. Kabel, 2020, pp. 24,25).

Weil man befürchtete, dass der Nomi Mate für den europäischen Raum zu verspielt sei, wurde ein zweites Konzept entwickelt. Der Nomi Halo kommt ohne Kugel und Display aus. Die Lichteffekte beschränken sich auf den Rand des Halo. Das Lichtkonzept des unbeweglichen Nomi Halo besteht aus drei Ebenen. Dem Lichtstreifen, dem beleuchteten Körper (Seitenfläche) und der Linieneffekt (ebenfalls an der Seitenfläche). Die Effekte laufen meist zeitgleich ab.



Abb. 39: Nomi Halo (XU, 2019)



Abb. 40: Nomi Mate (NIO, 2020)

Vier verschiedene Farben könne auf alle Ebenen angewendet werden:

- Türkis: Nomi Halo befindet sich im Ruhezustand und wartet auf Interaktion
- Gelb: Während der Benutzer zu Nomi Halo spricht
- Orange: Nomi Halo spricht zum Benutzer
- Blau: Nomi Halo befindet sich im Nio-Service-Modus

Durch die eindeutige Farbkodierung und der großen Leuchtfläche an der Seite ist der Nomi Halo gut aus dem Augenwinkel erkennbar. So bleibt er stets sichtbar im peripheren Blickfeld.

## 10.2. Sprachsteuerung in natürlicher Sprache

Mercedes Benz, Volkswagen und BMW verfolgen ähnliche Ansätze bei der Sprachsteuerung ihrer Fahrzeuge. Mit „Hey Mercedes“, „Hallo Volkswagen“ und „Hey BMW“ erweckt man die Assistenten.

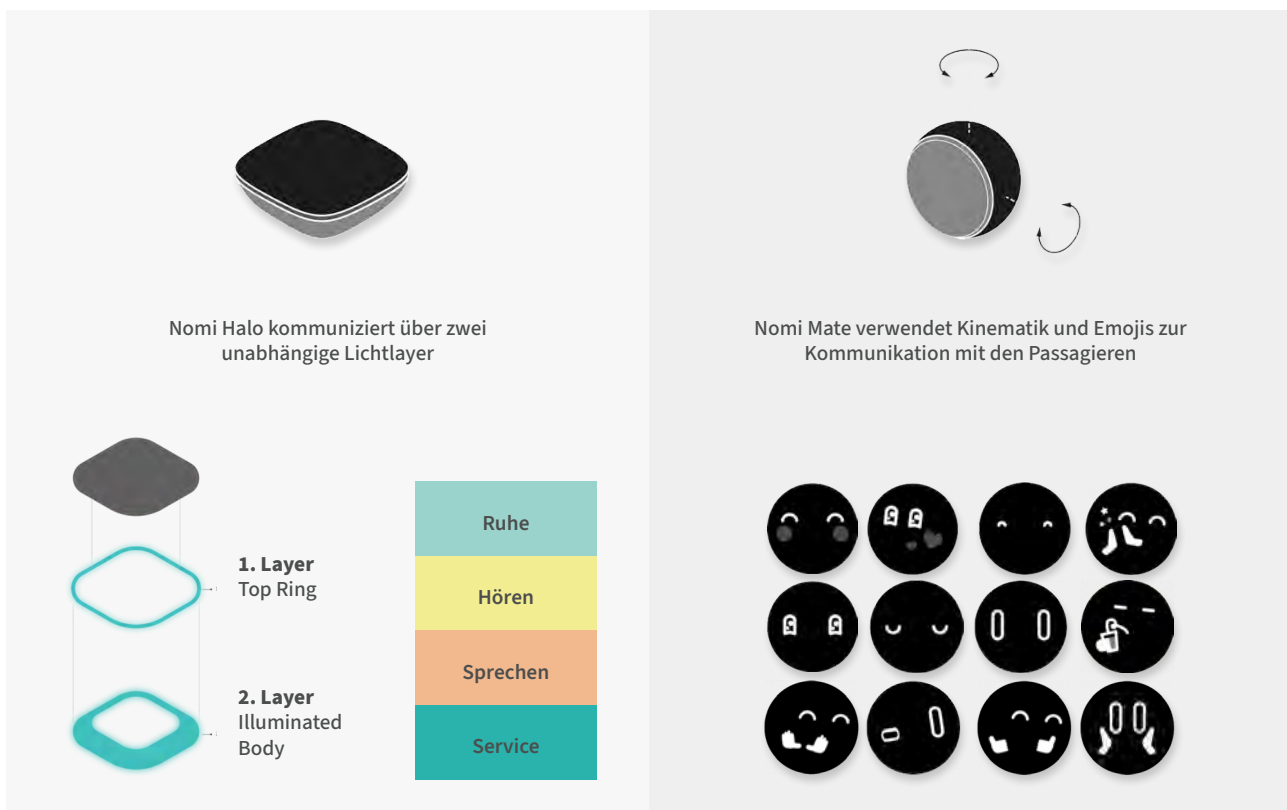


Abb. 41: Gegenüberstellung Nomi Halo und Mate( XU, 2019)

Alternativ kann man auch den entsprechenden Knopf am Lenkrad drücken.

Sie verstehen natürliche Sprache und können das Entertainment-System steuern, eine Reiseroute finden und die Klimaanlage einstellen. Sie sind alle in einem ähnlichen Ausmaß intelligent und lernen mit der Zeit ihre Fahrer kennen (vgl. Mercedes Benz, 2023; Volkswagen Communications, 2021; BMW Group PressClub, 2018).

Andere Hersteller wie Ford oder Chrysler verzichten auf die Entwicklung eigener Assistenten und greifen auf Alexa, Siri und den Google Assistant zurück. Visuell halten sich alle Assistenten sehr zurück. Nur am Display erscheinen Visualisierungen in Form von Wellen oder Linien.

### 10.3. Sprachsteuerung mit Kommandos

Hyundai setzt auf vordefinierte Befehle. Durch kurzes Drücken der Spracherkennungstaste am Lenkrad, startet man die Spracherkennung. Auf einem Bildschirm werden Vorschläge für Sprachbefehle angezeigt. Außerdem zeigt ein Lautstärkebalken auf dem Display an, ob ihre Stimme gehört wird. Durch längeres Halten der Spracherkennungstaste kann man alternativ auch Siri oder den Google Assistent nutzen.

## 11. Nonverbale Kommunikation

Nonverbale Kommunikation teilt sich vor allem in Gestik und Mimik. In zwischenmenschlichen Gesprächen machen sie einen großen Teil der Unterhaltung aus. Deshalb liegt ein besonderer Fokus darauf, diese non-verbale Anteile durch den Einsatz von unterschiedlichen Lichtsignalen am Interface aufzuwiegen.

### 11.1. Mimik

Es gibt nicht viele Gesichtsausdrücke die universal – also unabhängig von Person, Geschlecht, Nation, oder Kultur sind. Die folgenden elf mimischen Botschaften sind leicht zu interpretieren, weil es sich um eindeutige Aussagen handelt. Sie gelten unabhängig von der Situation, in der man sich befindet (vgl. Matschnig, 2012, p. 60).

Besonders interessant für die Anwendung bei einem User-Interface sind Gesichtsausdrücke, bei denen der Kontext eine wichtige Rolle spielt. Sie können helfen, die gesprochene Aussage richtig zu deuten. Monika Matschnig beschreibt in ihrem Buch mehrere mimische Signale und ihren Ursprung (vgl. Matschnig, 2012, pp. 60-68).

- **Zustimmung:** Das Heben der Augenbrauen während eines Gesprächs signalisiert Zustimmung und Interesse. Durch diese Reaktion werden die Augen automatisch größer und wir können mehr Informationen aufnehmen. Ein aufmunterndes Nicken kann ihren Gesprächspartner zum Weiterreden verleiten, es signalisiert Aufmerksamkeit und Bestätigung.
- **Ablehnung:** Ein fest geschlossener Mund, starrer Blick und eine gerümpfte Nase signalisieren fehlendes Einverständnis. Durch das Verengen des Sichtfelds nehmen wir unbewusst Abstand zu dem Gesagten.
- **Nachdenken:** Grübelt man lange über ein Problem nach, so legt man seine Stirn in Falten, um die geistige Anstrengung zu verdeutlichen. Durch Entspannen der Gesichtszüge beendet man den Gedanken und ist wieder offen für neuen Input



## 11.4. Fazit zu nonverbaler Kommunikation

Mimische Botschaften können eine Aussage unterstreichen und helfen, sie besser zu verstehen. Sie bieten zusätzlichen Kontext. Man kann versuchen, bestimmte Gesten oder Mimiken in die Gestaltung der visuellen Benutzeroberflächen einfließen zu lassen.

Ein Teil der Unterhaltung zwischen Menschen und Maschine kann nonverbal ablaufen. In manchen Fällen können Worte auch ganz ersetzt werden. Bestes Beispiel dafür sind einfaches Nicken und Kopfschütteln (vgl. Matschnig, 2012, p. 10).

Ein weiterer Punkt ist der Blickkontakt. Er ermöglicht eine direkte Verbindung zwischen den Gesprächspartnern und vermittelt Vertrauen, Aufmerksamkeit und Interesse.

Der Einsatz von visuellen Signalen kann die Mensch-Maschine-Kommunikation vereinfachen, beschleunigen und erfolgreicher gestalten.

## 12. Stimme

Mit der Auswahl der Stimme bestimmt man auch den Charakter des Assistenten. Schon nach wenigen Worten hat man eine Person vor dem geistigen Auge. Es gibt verschiedene Ansätze einem Sprachassistenten eine Stimme zu geben (vgl. Kabel, 2020, p. 36).

### 12.1. Recorded Speech (RS)

Bei RS werden von einem Sprecher alle benötigten Sätze und Worte in einem Studio gesprochen. So bleiben die Emotionen des Sprechers erhalten und die Intonation stimmt bei jedem Satz. Das hilft dabei, dass sich eine Frage auch nach einer Frage anhört. Das Aufnehmen aller möglichen Aussagen des Assistenten (Prompt) erfordert viel Zeit und Geld im Voraus. Man benötigt ein Aufnahmestudio, einen talentierten Sprecher und einen Tontechniker. Gibt es Änderungen bei den Aussagen, muss der Sprecher neue Aufnahmen machen (vgl. Pearl, 2016, p. 148).

### 12.2. Text-to-Speech (TTS)

Die Stärken von TTS liegen in seiner Flexibilität. Ein intelligenter Algorithmus bildet aus vielen individuellen Lauten zusammenhängende Wörter und kann so jeden beliebigen Satz formulieren. Bei schwacher Natural Language Generation (NLG) kann es passieren, dass die generierten Sätze etwas unnatürlich klingen. Außerdem kann es Ungenauigkeiten bei der Aussprache mancher Wörter geben. Defizite hat TTS auch bei der Intonation. Die meisten Systeme heutzutage verwenden eine Kombination aus RS und TTS. Sätze und Worte, die besonders häufig gesagt werden, spricht der Sprecher gesondert ein. So stimmen bei diesen Sätzen immer Tonlage und Betonung. Für andere Bereiche kommt TTS zum Einsatz. So kombiniert man die Emotionen menschlicher Stimmen mit der Vielseitigkeit von Text-to-Speech (vgl. Pearl, 2016, p. 148).

### 12.3. VA und Gender

Intelligente Sprachassistenten sind Computerprogramme und daher geschlechtslos. Die Stimme jedoch weist dem neutralen Assistenten sofort ein Geschlecht zu. In unserer Gesellschaft gibt es traditionellerweise geschlechtsspezifische Namen, die für Männer und Frauen bestimmt sind. Unisex-Namen helfen dabei Geschlechterstereotypen zu vermeiden.

Die meisten Sprachassistenten haben weiblich klingende Stimmen. Siri, Alexa und Cortana tragen ebenfalls weiblich konnotierte Namen. Die Ausnahme bildet der Google Assistent. Standardmäßig eingestellt ist bei allen eine weibliche Stimme (vgl. West, et al., 2019, p. 96).

In einer Publikation der EQUALS Skills Coalition werden die Eigenschaften von weiblichen und männlichen Stimmen folgendermaßen beschrieben (vgl. West, et al., 2019, p. 96).

- Eine Frauenstimme wird meist als angenehm, sympathisch, bescheiden und hilfsbereit wahrgenommen
- Männliche Stimmen werden hingegen als autoritär und ernst empfunden

Es stellt sich heraus, dass die meisten Menschen tiefe männliche Stimmen bevorzugen. Außerdem mögen sie den Klang einer männlichen Stimme bei autoritären Aussagen, aber eine weibliche Stimme, wenn sie hilfreich ist. Im Allgemeinen bevorzugen Menschen die Stimme des anderen Geschlechts. (vgl. West, et al., 2019, p. 96).

Gut zusammengefasst hat es Jessi Hempel im Wired Magazin:

*We want our technology to help us, but we want to be the bosses of it, so we are more likely to opt for a female interface (Hempel, 2015).*

Sie kommt zu dem Schluss, dass die Präferenz zu einer weiblichen Stimme weniger mit dem Klang zu tun hat, sondern mehr die Assoziierung mit Hilfe und Assistenz (vgl. West, et al., 2019, pp. 96,99-101). Eine weitere Erklärung für die mehrheitlich weiblichen Sprachassistenten könnte sein, dass der Großteil der Entwickler Männer sind. Im Schnitt sind nur 20% der technischen Mitarbeiter bei Apple, Google und Microsoft weiblich. Die Zahlen im Bereich Research und KI sind noch niedriger (vgl. West, et al., 2019, p. 102).

## 12.4. Alternative Stimmen

Mittlerweile bieten alle großen VA eine männliche Stimme als Alternative an. Manche Unternehmen bieten auch die Stimmen von Prominenten an. Die Navigations-App Waze hat zum Beispiel Morgen Freeman, Arnold Schwarzenegger oder Stephen Colbert im Angebot. Sogar fiktionale Figuren aus Filmen oder Serien könnte man für die Navigation wählen, darunter Bart Simpson oder C3PO (vgl. West, et al., 2019, p. 119).

## 12.5. Eine generische Stimme

Eine interessante Lösung, um den Gender-Bias zu umgehen, könnte eine genderlose Stimme sein. Q ist so eine genderneutrale Stimme. Eingesprochen wurde Q von Menschen, die sich weder als männlich noch als weiblich identifizieren. Anschließend wurden die Stimmen kombiniert und verändert, um in einem Bereich zwischen 145 Hz und 175 Hz zu liegen. Dieser bildet die Schnittmenge maskuliner und femininer Frequenzen. Sie klingt eindeutig menschlich, jedoch ist schwer zu beurteilen, ob ein Mann oder eine Frau spricht (vgl. West, et al., 2019, p. 122).

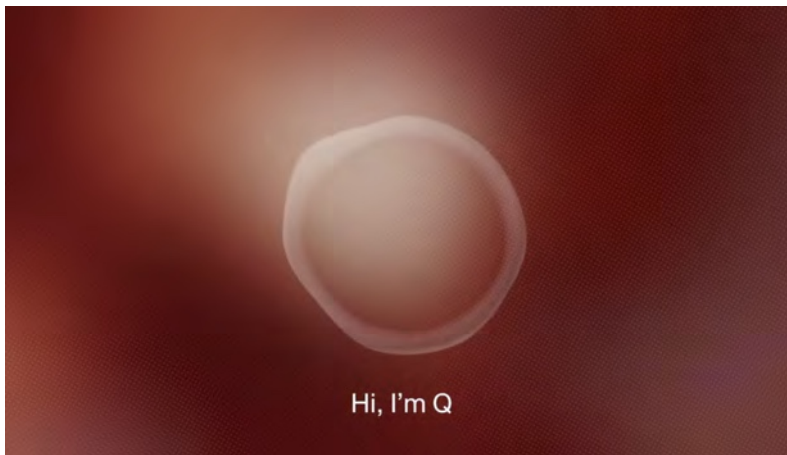


Abb. 44: Q - die erste genderlose Stimme (genderlessvoice, 2019)

## 13. Funktionsklänge

Signaltöne können sehr schnell Informationen weiterleiten. Zum Beispiel der Ton einer Nachricht oder das Ticken des Blinkers im Auto. Diese akustischen Signale können Warnungen, Hinweise oder Alarme bedeuten. Vor dem Einsatz von Funktionsklängen müssen mehrere Aspekte berücksichtigt werden. Erstens muss die Hierarchie beachtet werden, sodass ein Hinweiston deutlich von einem Warnton zu unterscheiden ist. Zweitens sollte die Klangästhetik zum visuellen Erscheinungsbild passen (vgl. Dorau, 2011, p. 19).

Der Hauptnachteil von akustischen Rückmeldungen ist ihr Störpotenzial für andere Menschen. Wird man mit auditiver Information konfrontiert, gegen die man nichts machen kann, ist man schnell genervt. Anders als bei visuellen Störfaktoren, bei denen man wegsieht oder die Augen schließt, kann man bei auditiven Signalen nicht so leicht weghören (vgl. Dorau, 2011, p. 217).

## 14. Herausforderung/Schwierigkeiten von VUI

### 14.1. Menschen machen Fehler beim Sprechen

Menschen machen häufig Fehler, wenn sie dazu aufgefordert werden, sich unnatürlich zu verhalten, beispielsweise sich auf ungewohnte Weise auszudrücken. Sprachbefehle können unterbrochen, neu gestartet oder mitten im Satz verändert werden. Trotz dieser Unstimmigkeiten sind die Äußerungen meist verständlich. Sprachen sind ein ausgezeichnetes Beispiel dafür, wie Systeme auf die menschlichen Fähigkeiten zugeschnitten werden können. Sie bieten eine reichhaltige Struktur für Kommunikation und soziale Interaktion und sind gleichzeitig äußerst tolerant gegenüber Fehlern (vgl. Norman, 1998, p. 159). Es ist Aufgabe des VUI, mit diesen Unregelmäßigkeiten umzugehen (vgl. Pearl, 2016, p. 226).

## 14.2. Beschränkte Menge an Informationen

Bildschirme mit Text sind ideal, um Listen anzuzeigen bzw. große Mengen an strukturierten Daten zu präsentieren. Ein VUI muss sorgfältig auswählen, welche Informationen man dem Benutzer weitergibt. Es ist wichtig, die Menge an Daten zu berücksichtigen, die in einer einzelnen Interaktion übermittelt werden. Außerdem sollte man auf die Geschwindigkeit achten und angemessene Pausen zwischen den Elementen berücksichtigen (vgl. Pearl, 2016, p. 227).

## 14.3. Übersehen verbesserter und neuer Funktionen

Oft sind Nutzer nicht in der Lage, zu erkennen, ob ein Problem behoben wurde. Negative Erfahrungen mit einer Funktion prägen sich im Gedächtnis ein. Es ist äußerst selten, dass Nutzer von sich aus nach Monaten zurückkehren und erneut einen Versuch unternehmen. Dazu kommt, dass es schwierig ist, neue Funktionen zu entdecken. Tutorials und andere Tipps werden gerne ignoriert (vgl. Laubheimer & Budiu, 2018).

## 14.4. Umgebungslärm

Störgeräusche können in einer industriellen Umgebung zu Problemen führen. Wird die Aussage des Benutzers nicht erkannt, kommt schnell Frust auf.

Das Fraunhofer-Institut nutzt zur Spracherkennung ein kabelloses Headset und ein stationäres Mikrofon. Durch die Kombination aus Richtmikrofonen und wirkungsvollem Noise-Cancelling werden laute Umgebungsgeräusche fast vollständig ausgeblendet (vgl. Fraunhofer Institut, 2023).

## 14.5. Calm Technology

Gemäß Weiser und Brown bezieht sich „Calm Technology“ auf Geräte und Systeme, die unauffällig, umgebungsorientiert sind und darauf abzielen, das Leben zu verbessern, ohne dabei ständige Aufmerksamkeit zu fordern.

*Calm technology engages both the center and the periphery of our attention, and in fact moves back and forth between the two (Weiser & Brown, 1995, p. 2).*

Peripherie wird hier alles bezeichnet von dem man zwar etwas mitbekommt, auf das man sich aber nicht konkret konzentrieren muss. Als Beispiel wird das Autofahren angeführt. Im Zentrum der Aufmerksamkeit steht die Straße, das Radio und die Mitfahrer, jedoch nicht das Geräusch des Wagens. Ein untypisches Motorgeräusch erkennt bemerkt man aber sofort. Technologie zu gestalten, die in der Peripherie unserer Aufmerksamkeit funktioniert, hat zwei große Vorteile:

- + **Erstens**, kann man sich mehreren und mitunter wichtigeren Dingen widmen, als wenn man alles im gleichen Maße beachten müsste.
- + **Zweitens**, indem wir etwas, das zuvor am Rand unserer Wahrnehmung war bewusst in den Mittelpunkt rücken, schaffen wir die Möglichkeit zu handeln. Das bedeutet, dass wir uns aktiv mit einer Sache auseinandersetzen können, die andernfalls unbeachtet geblieben wäre (vgl. Weiser & Brown, 1995, p. 2).

## 15. Zusammenfassung und Fazit

Auch wenn intelligente Sprachassistenten stetig verbessert werden, klafft noch eine große Lücke zwischen dem, wofür Menschen VA verwenden und was sie im Stande wären zu leisten. Bei der Gestaltung des Interfaces ist es von großer Bedeutung, dem Nutzer verständlich darzulegen, welche Handlungsmöglichkeiten ihm zur Verfügung stehen und wo seine Freiheiten eingeschränkt sind. Noch fühlt sich ein Gespräch mit einem VA künstlich und konstruiert an. Science-Fiction Filme wie *2001: ODYSSEE IM WELTRAUM* (*2001: Odyssee im Weltraum*, 1968), *IRON MAN* (*Iron Man*, 2008) oder *HER* (*her*, 2013) sind gute Vorbilder dafür, wie natürlich und problemlos ein Gespräch mit dein VA ablaufen kann. Um dort hinzukommen, braucht es mehr Vertrauen in die Technik und kompetentere Assistenten.

Um die Immersion nicht zu zerstören, ist es wichtig, dass die Art der Lichtsequenzen zur aktuellen sprachlichen Interaktion passt und das akustische und visuelle Feedback zeitlich aufeinander abgestimmt ist. Das Interface sollte alle Informationen darstellen können, ohne den Nutzer von seiner Hauptaufgabe abzulenken (vgl. Case, 2016, p. 51). Ziel ist es, Technologie so zu gestalten, dass sie sich an das bestehende Verhalten anpasst und in die natürliche Umgebung einfügt. Technologie soll bei täglichen Aktivitäten unterstützen, ohne zu überfordern oder abzulenken.

# **III. Konzeptphase**

**Interface**



## 2. Formal Moods

Als Leitplanken für die Gestaltung des Interfaces dienen mehrere Moodboards für Form, Material, Oberfläche und Lichtakzente. Die Konturlinien sollen klar und nachvollziehbar sein. Es ist wichtig, Bereiche zu schaffen, die Komplexität widerspiegeln können und gleichzeitig die Möglichkeit für Variation bieten. Die Materialien müssen Industriestandards entsprechen, dürfen jedoch nicht plump oder stilllos wirken. Außerdem soll das eingesetzte Material die dahinterliegenden Funktionen andeuten.

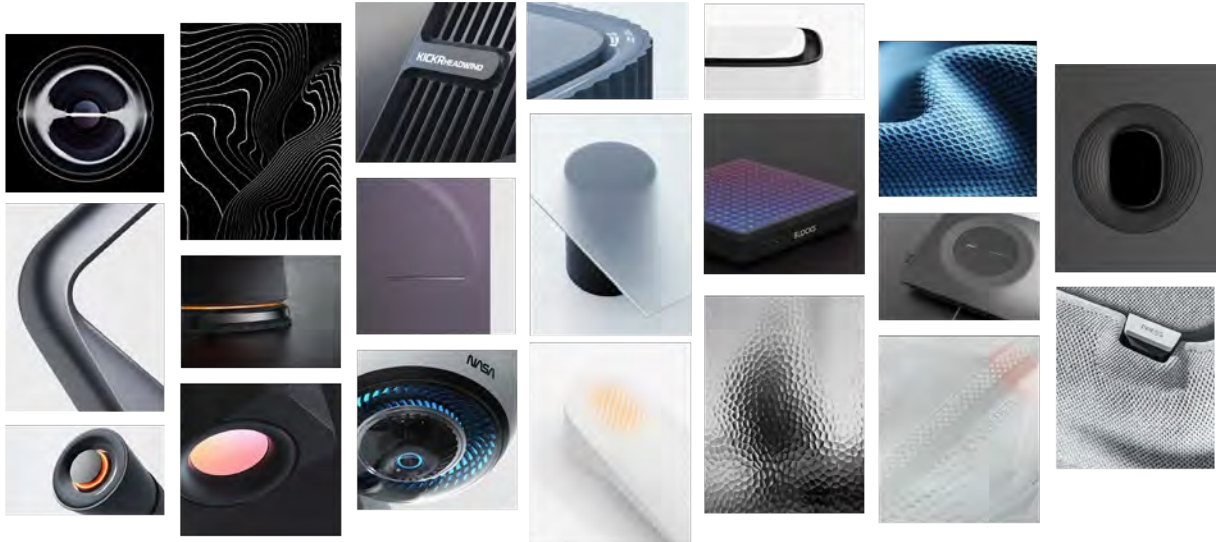


Abb. 46: Moodboard für Formen, Oberflächen und Lichtakzente

Genauer untersucht wurde auch die Gestaltung der anderen Produkte von KEBA. Dazu zählen mobile und stationäre HMI, sowie diverse Bankomaten und Paketabholstationen. Nach Analyse der Produkte wurden folgende Kernelemente ermittelt:

- Die Farben Grün und Grau (mehrere Abstufungen)
- Große Radien an den Objekträndern
- Kleine Radien an den Ecken
- Die KEBA-Schräge, also der Neigungswinkel des Logos: 19°
- Wiederholende Formen, z.B.: im Logo



Abb. 47: Moodboard für Formen, Oberflächen und Lichtakzente

## 3. Nutzungsanforderungen

Um die unterschiedlichen Anforderungen an das Interface zu organisieren, wurde die MoSCoW-Methode gewählt. Dabei werden alle relevanten Kriterien in die vier Kategorien *Must-have*, *Should-have*, *Could-have* und *Won't have* eingeteilt. So behält man während dem Entwurfsprozesses den Überblick, welche Eigenschaften Priorität haben.

### 3.1. Must-have

- + Statusanzeige
- + Mikrofone um eine gute Spracherkennung zu gewährleisten
- + Lautsprecher für eine hohe Stimmqualität
- + Veränderbares Farbenspektrum
- + Komplexität der KI zeigen
- + Flache Einbausituation
- + Muss aus verschiedenen Blickwinkeln sichtbar sein

### 3.2. Should-have

- + Robuste Bauweise für raues Industrie-Umfeld
- + Leicht zu reinigende Oberfläche
- + Schlichte Formensprache
- + Weiche Materialien als Hinweis auf Audioausgabe
- + Intuitive und logische Interaktionsschritte

### 3.3. Could-have

- Adaptierbar für alle Einsatzzwecke
- Bezug zur KEBA-Formensprache
- Einfaches Kabelmanagement
- Gestenerkennung

### 3.4. Won't-have

- Immer grün leuchten
- Sichtbare Knöpfe oder Taster
- undefinierbare Klänge
- Beschriftungen
- Touch-Interaktion

## 4. Zweiter Workshop

Ziel dieses Workshops war es einen Ausgangspunkt für die Ideation zu finden. Alle Beteiligten hatten die Aufgabe ihre ersten Ideen und Visionen des Interfaces zu Papier zu bringen. Teilnehmer waren wie beim ersten Workshop Thomas Linde und Julian Bauer von KEBA, Martin Rothmayer und die projektbegleitende Professorin Elke Bachlmair. Als Starthilfe gab es eine Wortsammlung mit Adjektiven, eine Liste an Oberflächen, Materialien, möglichen Einbauformen und Bauteilen. Zusätzlich gab es Underlays mit verschiedenen Grundkörpern, um das Zeichnen zu erleichtern.

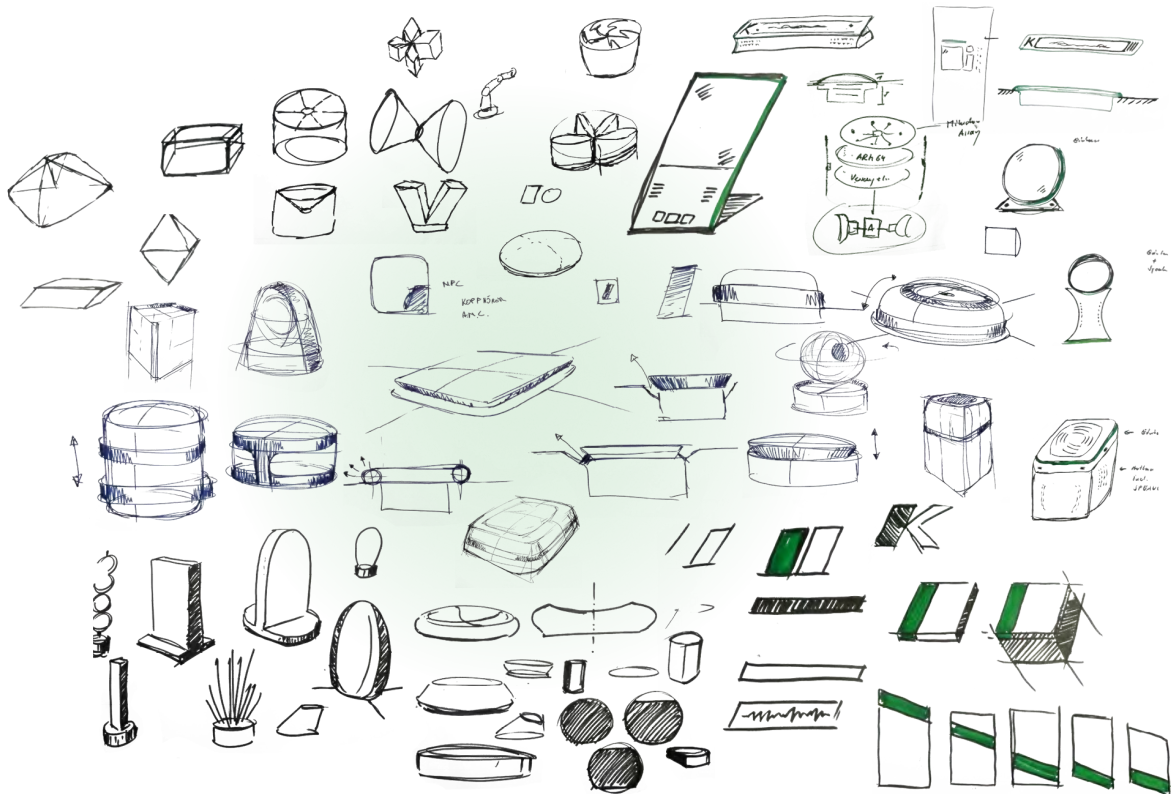


Abb. 48: Übersicht aller Sketche aus dem zweiten Workshop vom 4.8.2022

### 4.1. Fazit des 2. Workshops

Interessant war die Erkenntnis, dass die Lichteffekte wichtiger als die Form eingestuft wurden. Betont wurde die raue Arbeitsumgebung, in dem das Interface bestehen muss. Es muss äußeren Einflüssen wie Staub, Flüssigkeiten oder mechanische Einwirkungen standhalten. Gleichzeitig wurde auf die Ästhetik des Umfelds hingewiesen. Dies bezieht sich beispielsweise auf die mangelnde Sauberkeit von alten Industriehallen und in die Jahre gekommenen Spritzgussmaschinen.

Diskutiert wurden auch die Vor- und Nachteile verschiedener Einbausituationen. Eine mobile Lösung könnte man von einem Einsatzort zum nächsten bewegen. Durch die Sprachbedienung deckt man ohnehin schon einen größeren Bereich rund um eine Maschine ab, ist somit bereits bedingt ortsungebunden.

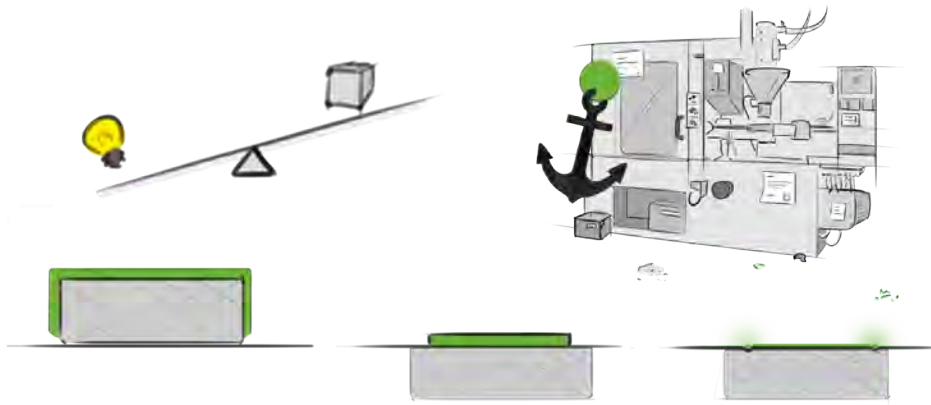


Abb. 49: Fazit aus dem zweiten Workshop vom 4.8.2022

Das Interface mitzunehmen an eine andere Maschine wird in der Praxis nur zu unnötigen Mehraufwand führen. Effektiver ist es, wenn jede Maschine im Bestand ein eigenes Interface besitzt. Somit ist eine wichtige Entscheidung getroffen, das Interface wird fix an einer Maschine verbaut.

## 5. KEBA-Unterscheidungswerte

Auf Basis der Smart Speaker-Analyse und der Erkenntnisse aus dem 2. Workshop wurden mehrere Merkmale festgelegt, anhand derer man Smart Speaker Interfaces beschreiben kann. Ein Schieberegler definiert, wie weit man sich in einem Bereich an traditionellen Merkmalen orientiert und welchen man sich abheben möchte.

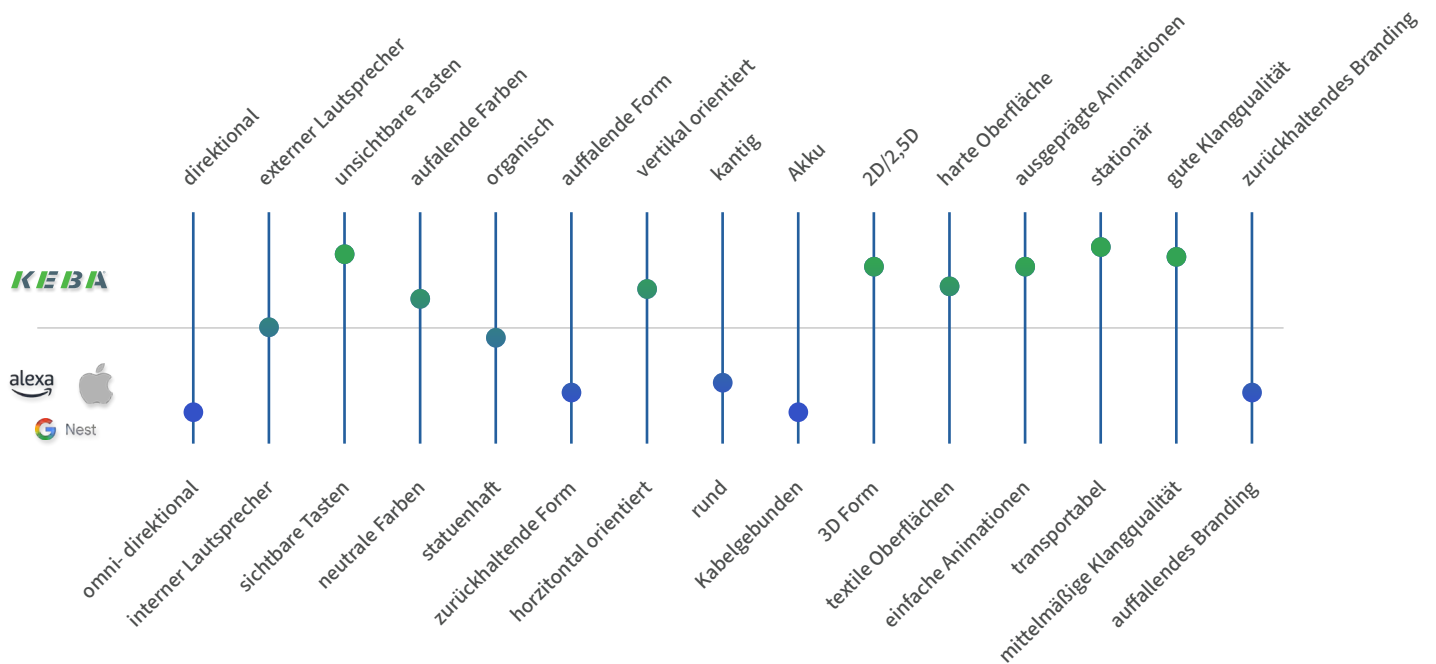


Abb. 50: Unterscheidungsmerkmale des Interface im Vergleich zu anderen Smart Speakern

## 6. Konzeptfindung

Während der Konzeptfindung wurden verschiedene Eigenschaften kombiniert und viele Entscheidungen gleichzeitig getroffen. Die folgenden Punkte beschreiben wichtige Faktoren des Konzepts.

### 6.1. Einbausituation

Es wurde bereits festgelegt das das Interface stationär verbaut wird. Jedoch stellt sich die Frage wie viel *Körper* besitzt das Gerät. Sitzt es bündig in der Außenhülle, ragt es wenige Zentimeter vor, oder befindet sich die Vorrichtung komplett außerhalb. Die Favoriten der ersten Entwürfe tendierten bereits zu einer recht flachen Variante. Mit einer integrativen Bauform sind mehrere Vorteile verbunden. Aufgrund geringerer Angriffsfläche bietet sie Schutz vor äußeren, mechanischen Einwirkungen. Außerdem ist es leichter das Eindringen von Staub oder Flüssigkeiten zu vermeiden da man nur in einer Ebene abdichten muss. Des Weiteren sind umlaufende Bespielungen von allen Seiten sichtbar, unabhängig von vertikaler oder horizontaler Orientierung.

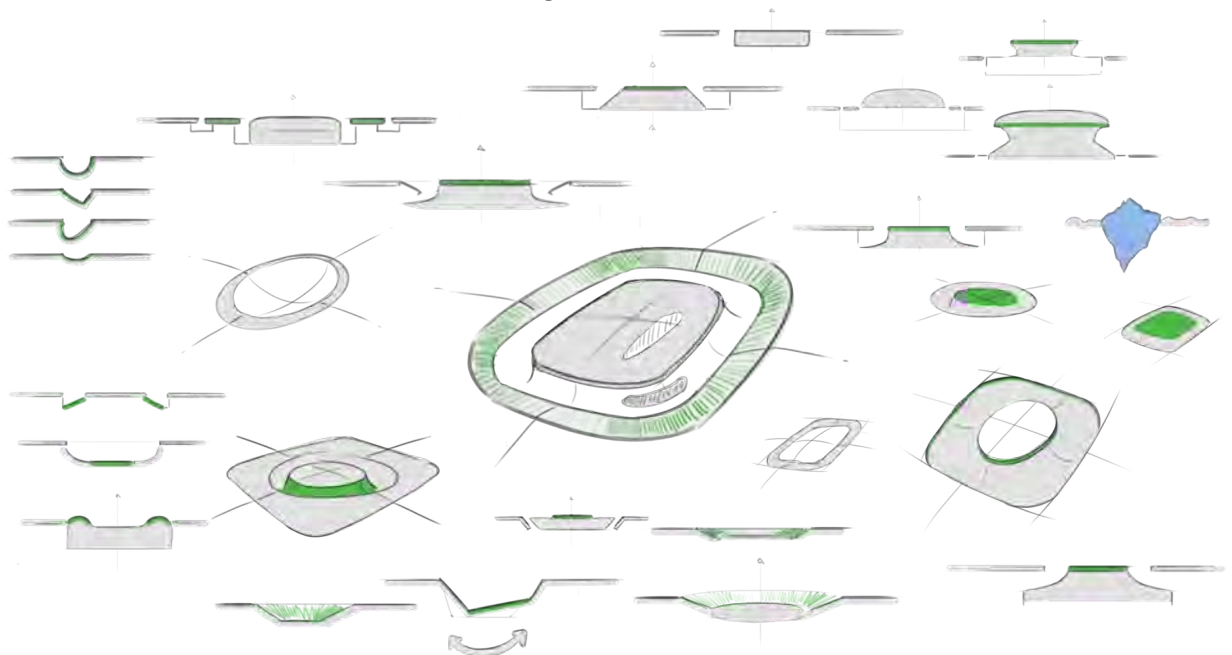
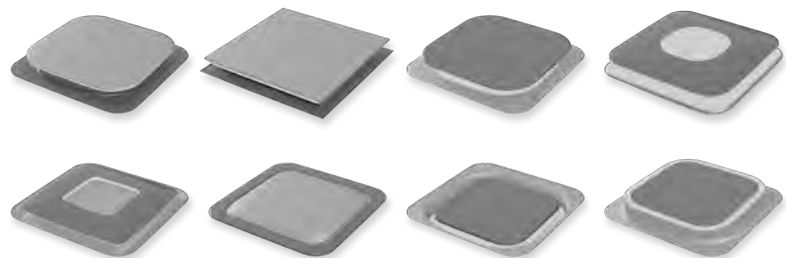


Abb. 51: Entwurfsskizze mit einer flachen Einbausituation

Abb. 52: 3D-Entwürfe mit einer flachen Einbausituation



Außerdem assoziiert man die Stimme eher mit der Maschine, wenn es tatsächlich aus der Maschine kommt. In diesem Kontext soll dem Benutzer klar sein, dass sich die Zuständigkeit des Interfaces nur auf die konkrete Maschine bezieht. Demnach wäre es irreführend das Interface extern zu positionieren. Es könnte dem Nutzer den Eindruck vermitteln, dass das Interface über einen größeren Kompetenzbereich verfügt, als es tatsächlich der Fall ist.

Das KITT-Konzept ist dem gleichnamigen Auto nachempfunden. Eine schmale beleuchtete Linie läuft rund um die Maschine. So ist das Interface von überall sichtbar und durch mehrere Sensoren kann das Umfeld gut im Blick behalten werden.

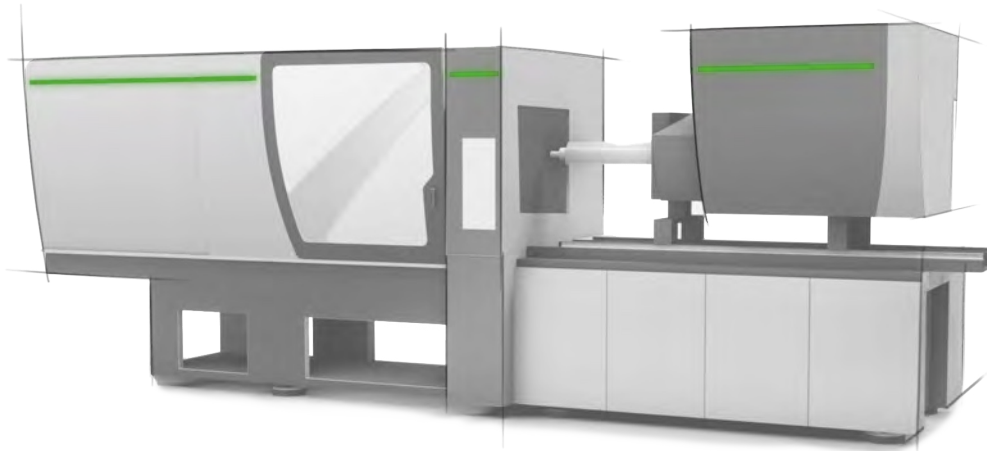


Abb. 53: Entwurfskonzept KITT am Beispiel einer Spritzgussmaschine

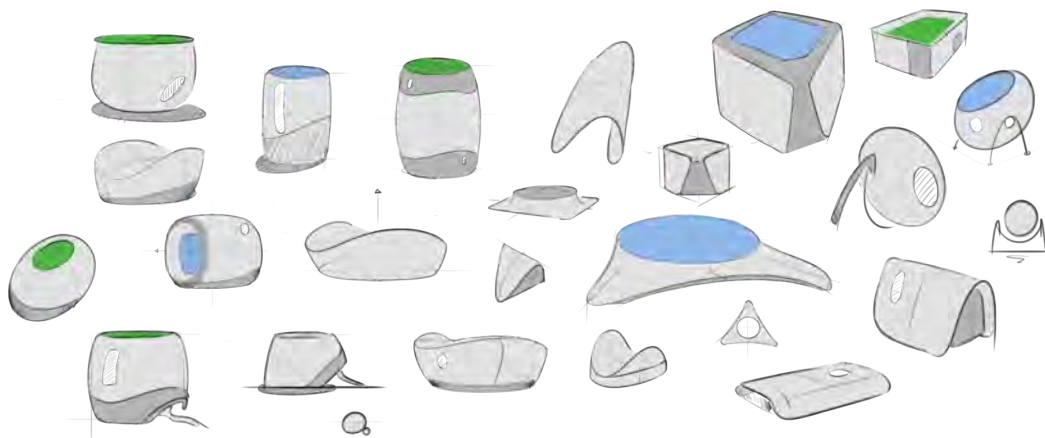
## 6.2. Positionierung

Die Platzierung des Interfaces variiert je nach Anwendungsfall. Das Interface kann horizontal, vertikal, oberhalb oder unterhalb der Augenhöhe des Nutzers positioniert werden. Wichtig ist, dass das Interface aus allen Blickrichtungen gut erkennbar ist.

Durch eine hohe Positionierung an der Maschine soll vermieden werden, dass man das Interface mit der Hand berührt. Bedienelemente, die man mit der Hand betätigt, befinden sich selten über Kopfhöhe. Ein Näherungssensor könnte erkennen, wenn sich die Hand des Benutzers oder ein Objekt nähert und einen Hinweis in visueller oder sprachlicher Form geben. Die Positionierung über Kopfhöhe könnte problematisch sein, weil es als Kommunizieren von „oben herab“ empfunden werden kann.

## 6.3. Grundform

Sowohl eine kreisförmige als auch eine quadratische Grundform besitzen eine äußerst neutrale Silhouette, die sich gut in unterschiedliche Umgebungen einfügen kann. Eine quadratische Kontur unterscheidet sich jedoch klar von den kreisförmigen Smart Speakern wie Amazon Echo, Apple HomePod und dem Google Nest. Trotzdem bietet sie gute Voraussetzungen für eine gleichmäßige Verteilung des Schalls und Anordnung der Mikrofone.



## 6.4. Branding

Bei den mobilen und stationären HMI von KEBA, stehen die Bedienelemente immer im Vordergrund. Das Logo ist meist nur klein am Rand sichtbar. Außerdem sind beinahe alle Oberflächen für jeden Kunden individualisierbar. Einige Entwürfe die Teile des Logos als zentrales Element enthielten, scheiden daher aus.

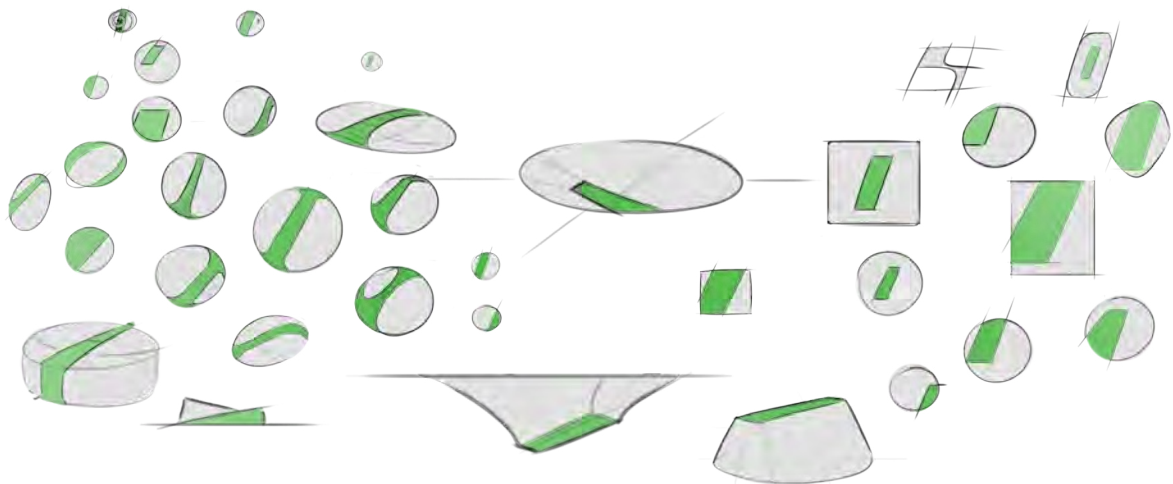


Abb. 54: Entwurfsskizze die das KEBA Logo in den Mittelpunkt stellen

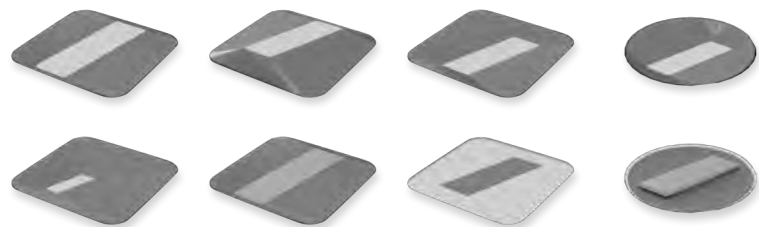


Abb. 55: 3D-Entwürfe mit dem KEBA Logo

## 6.5. Bewegung

Folgende Optionen für ein möglichst agiles Interface wurden überlegt und mit Skizzen und Animationen dargestellt.

Eine Möglichkeit wäre Teile physisch zu bewegen. Es gab Entwürfe, die ein Muster aus vielen beweglichen Einzelteilen vorsah. Diese *Schuppen* könnten den Blick auf eine darunterliegende Lichtfläche freigeben oder kollektiv dem Benutzer zugewandt sein und so Blickkontakt imitieren.

Bei einem anderen Ansatz versuchte man den Eindruck zu erwecken, dass sich unter der Oberfläche des Interfaces etwas *Lebendiges* befindet. Verschiedene Formen könnten ihre Konturen durch eine Bespannung aus Stoff drücken und so ein 3-dimensionale Oberfläche schaffen. Der Schatten, der so auf einer Seite entsteht, würde den Effekt noch verstärken.

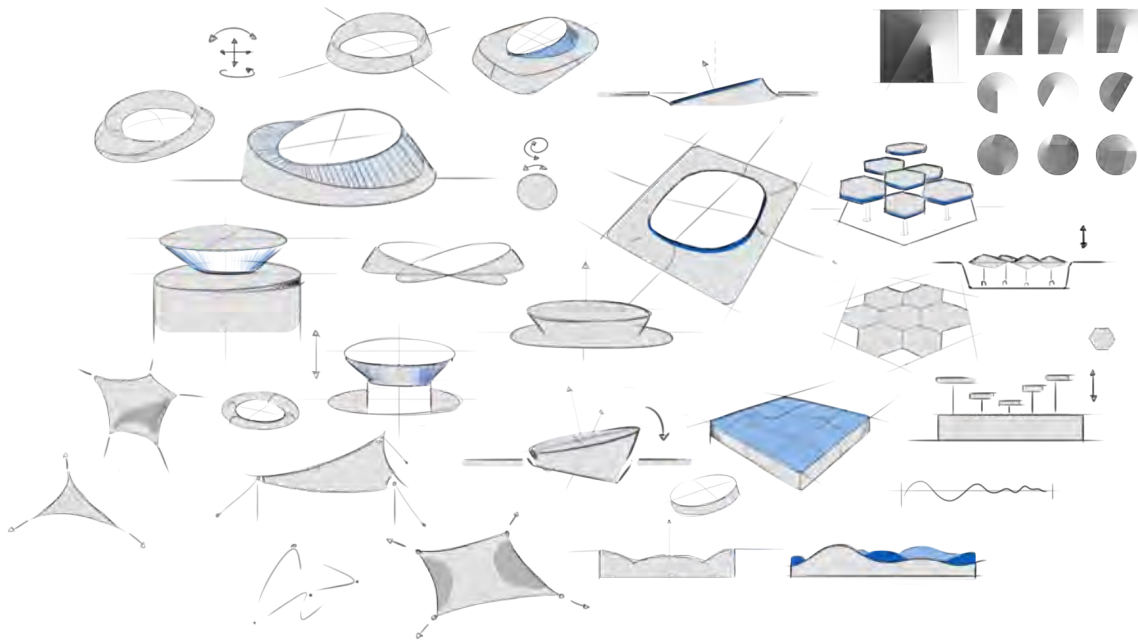


Abb. 56: Entwurfssketches mit beweglichen Teilen

Abb. 57: 3D-Entwürfe mit beweglichen Teilen



Angenommen das gesamte Interface könnte 3-dimensionale Wellen formen, so könnte man den entstehenden Mustern unterschiedliche Bedeutungen zuschreiben. Ähnlich den Mustern, in mit Wasser gefüllten Klangschalen oder Ferrofluid über einem Magneten.

Diese Überlegungen wurden jedoch nicht weiterverfolgt, da jedes bewegliche Teil eine mögliche Fehlerquelle darstellt. Regelmäßige Instandhaltungsarbeiten oder komplexe Reparaturen der Bauteile sollen möglichst vermieden werden.

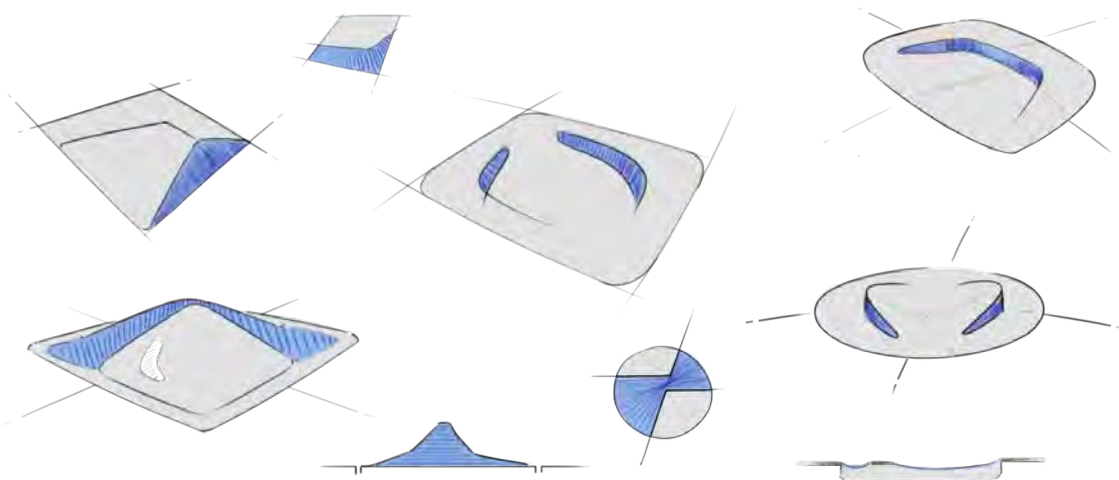


Abb. 58: Entwurfssketches mit schmalen Öffnungen in der Oberfläche

## 6.6. Lichtflächen

Für die visuelle Kommunikation können verschiedene Bereiche zuständig sein. Entweder wird die gesamte Fläche hinterleuchtet, nur ein Teilbereich farbig bespielt oder nur Details wie beispielsweise die Kante hervorgehoben. Es ist auch möglich mehrere Ebenen zu kombinieren.

Oberflächendurchbrüche in Form von Schnitten oder Schlitzen, durch die Licht nach außen dringt erwecken den Eindruck das sich darunter etwas *Lebendiges* befindet. Öffnungen in der Deckfläche können tiefe Einblicke gewähren und die Neugierde des Nutzers wecken.

Ein Problem bei diesen Varianten ist jedoch das die Größe der zu bespielende Fläche recht gering ist. Wenn die Lichteffekte auf den ersten Blick klar erkenntlich sein sollen, muss auch die bespielte Fläche dementsprechend groß sein.

## 6.7. Zusätzliche Funktionen

Während des Entwurfsprozesses kamen einige Ideen für zusätzliche Funktionen auf.

Störgeräusche am Arbeitsplatz können die Spracherkennung beeinträchtigen, was zu Kommunikationsproblemen und fehlerhaften Eingaben führen kann. Alternativ könnte der Benutzer Schwierigkeiten haben, die Ausgaben des Interfaces zu verstehen. In diesem Fall könnte ein intelligentes Wearable als Lösung dienen, indem es einerseits als Gehörschutz fungiert und andererseits als Aufnahmegerät verwendet wird. Durch die integrierten Mikrofone kann das Headset die gesprochenen Befehle präziser aufzeichnen, sodass der Benutzer nicht laut in Richtung des Geräts rufen muss.

Ein interessantes Feature für die Kopfhörer wäre direktionales Hören. Die Kopfhörer bestimmen dabei stets ihre Position im Verhältnis zum Interface, dies ermöglicht eine immersivere Hörerfahrung und verstärkt das Gefühl der räumlichen Wahrnehmung für den Benutzer.

Die Erweiterung des Produktportfolios um kleine Module bietet vielfältige Möglichkeiten. Jede Einheit könnte beispielsweise eine Kamera, einen Lautsprecher oder ein einfaches Display enthalten. Durch die Verwendung dieser separaten Komponenten mit externen Sensoren oder Ausgabegeräten könnte der Nutzungsraum des Benutzers erweitert werden.

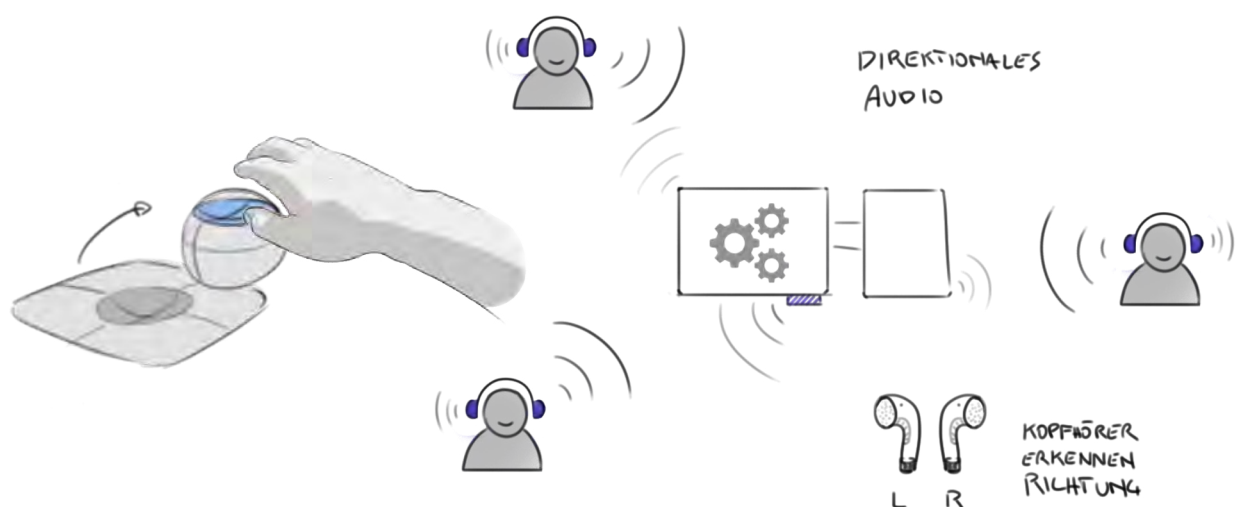


Abb. 59: Entwurfssketch von zusätzlichen Features

## 7. Keysketch

Der Keysketch zeigt einen Entwurf, der einige der gesetzten Vorgaben erfüllt. Ein breiter, umlaufender Bereich bietet viel Fläche für Lichteffekte. Im Zentrum ist ein optionales Display mit Platz für einen darunterliegenden zentralen Lautsprecher. Die schmale Linie, die die Kontur des zentralen Elements interaktiv hervorhebt, könnte für besonders wichtige Signale verwendet werden.

Um anzudeuten, dass sich hinter dem unbeleuchteten Bereich in der Mitte ein oder mehrere Lautsprecher befinden, könnte man diesen mit Stoff überziehen. Die Grundform ohne beleuchteter Flächen erinnert an einen Lautsprecher. Der Rand ist ein nach innen zulaufender Kegel mit dem erhöhten Bereich in der Mitte. Jedes sichtbare Element des Interfaces soll einen Teil der Kommunikation übernehmen. Im Zusammenspiel ergibt sich ein stimmiges Gesamtbild.

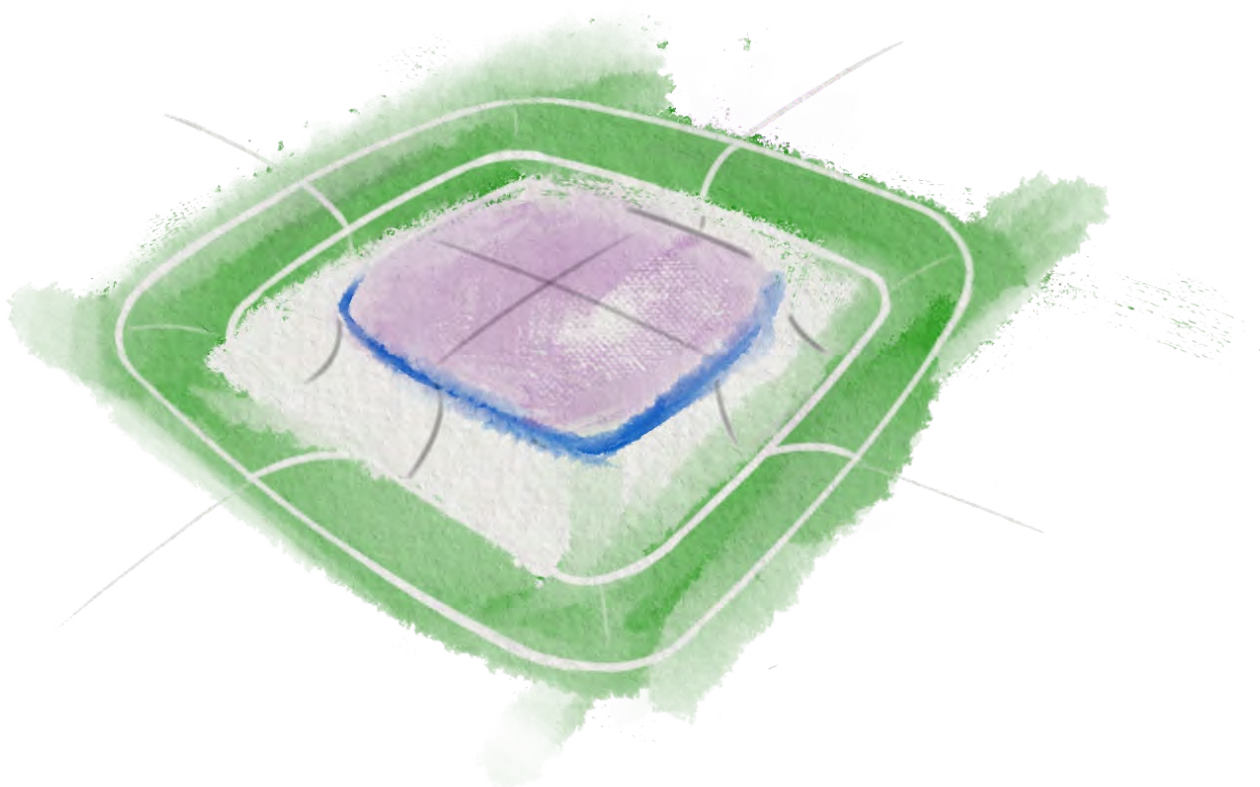


Abb. 60: Keysketch der Voice Arena

### 7.1. Die umlaufende Lichtfläche

Soll den Zustand der Maschine bzw. des Interfaces auf den ersten Blick erkenntlich machen. Gerade wenn sich das Interface nur im peripheren Blickfeld befindet, sollte schnell klar sein, wie der Status ist.

### 7.2. Das Zentrum

Hier sollen nur bei Bedarf Icons oder Symbole dargestellt werden. Diese dürfen nicht zu sehr von der Aussage der umgebenden *Aura* ablenken. Der Nutzer soll nicht dazu gezwungen sein, andauernd einen genauen Blick darauf zu werfen.

### 7.3. Lichtkante um das Zentrum

Falls erforderlich, kann diese Lichtkante zusätzliche Aufmerksamkeit auf die Symbole in der Mitte richten oder die umlaufende Lichtfläche unterstützen.

### 7.4. Lautsprecher

Eine perforierte Oberfläche oder ein textiles Gewebe signalisieren meist einen Lautsprecher. Um diese Funktion sichtbar zu machen, wird der Übergang zwischen dem umlaufenden Lichtband und dem zentralen Display genutzt.

## 8. Baugröße und Proportion

Um ein Verständnis für die Größenverhältnisse der verschiedenen Bereiche zueinander zu entwickeln, wurde eine Matrix erstellt, die es ermöglicht, viele verschiedene Varianten auf einen Blick zu vergleichen. Mit Hilfe von Grasshopper, einer parametrischen Designsoftware, wurde ein einfaches Modell des Interface erstellt, das flexibel konfiguriert werden konnte. Die Basisparameter umfassen die Größe der Flächen und ihre Höhe in Bezug auf die umgebende Maschinenhülle.

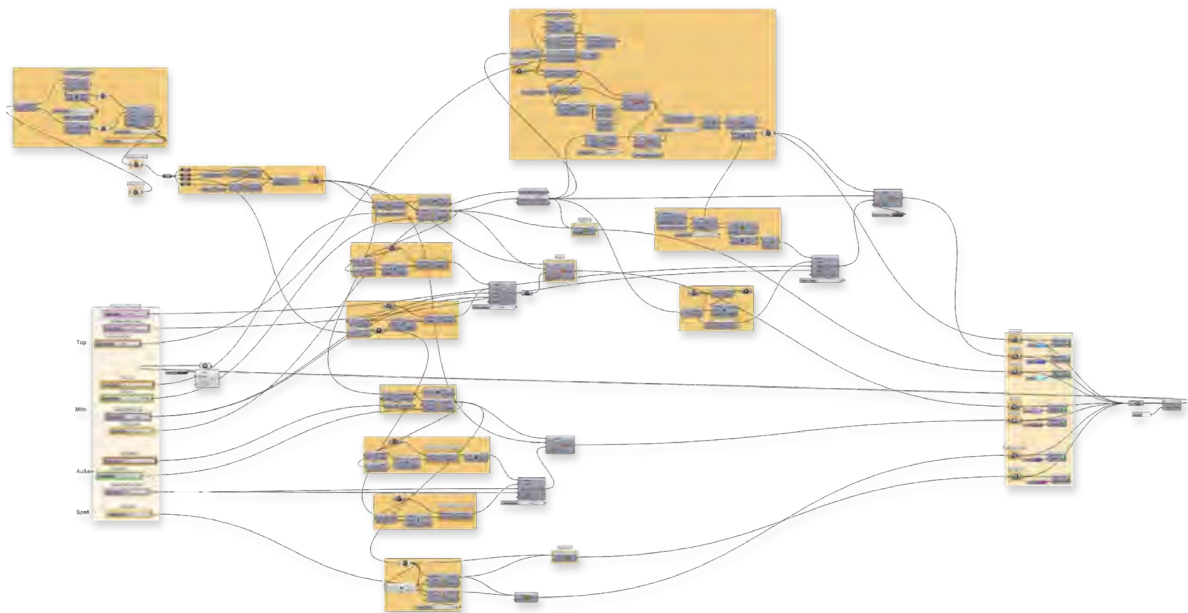


Abb. 61: Grasshopper Node-Tree für die Erstellung der Matrix

In Absprache mit den Projektbetreuenden bei KEBA und der Universität wurden zuerst sieben Favoriten ausgewählt, die genauer bewertet wurden. Aus diesen sieben wurden anschließend die beiden Varianten mit dem größten Potenzial ausgewählt, an denen die Bespielung der Lichtflächen getestet werden soll. Kriterien nach denen bewertet wurde, waren Größe bzw. Sichtbarkeit der bespielbaren Fläche, Einbausituation und technische Aspekte.

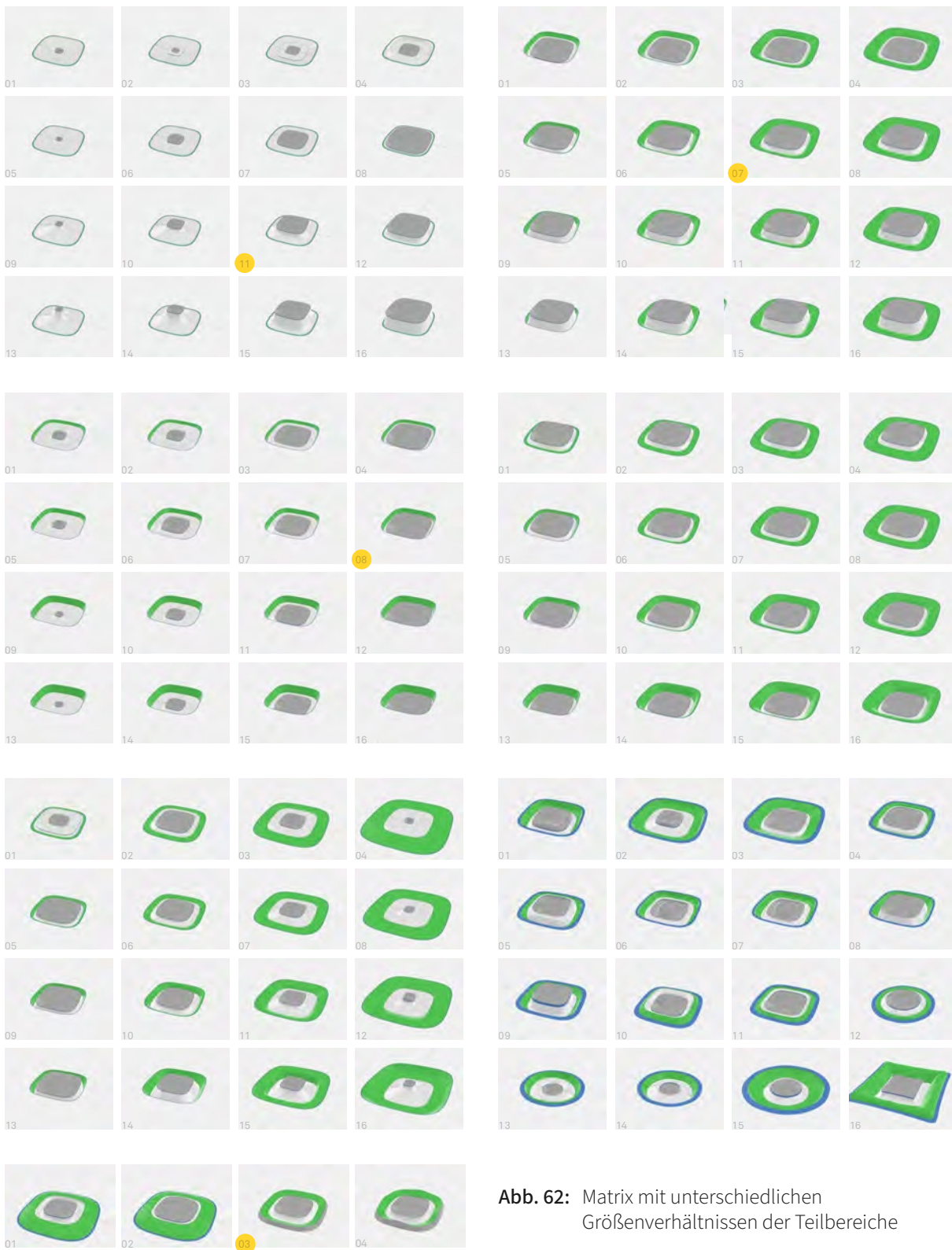


Abb. 62: Matrix mit unterschiedlichen Größenverhältnissen der Teilbereiche

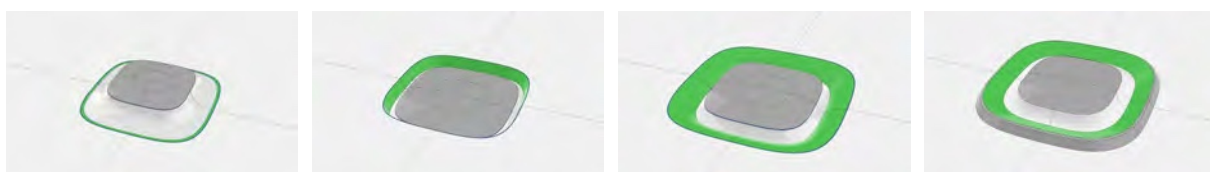
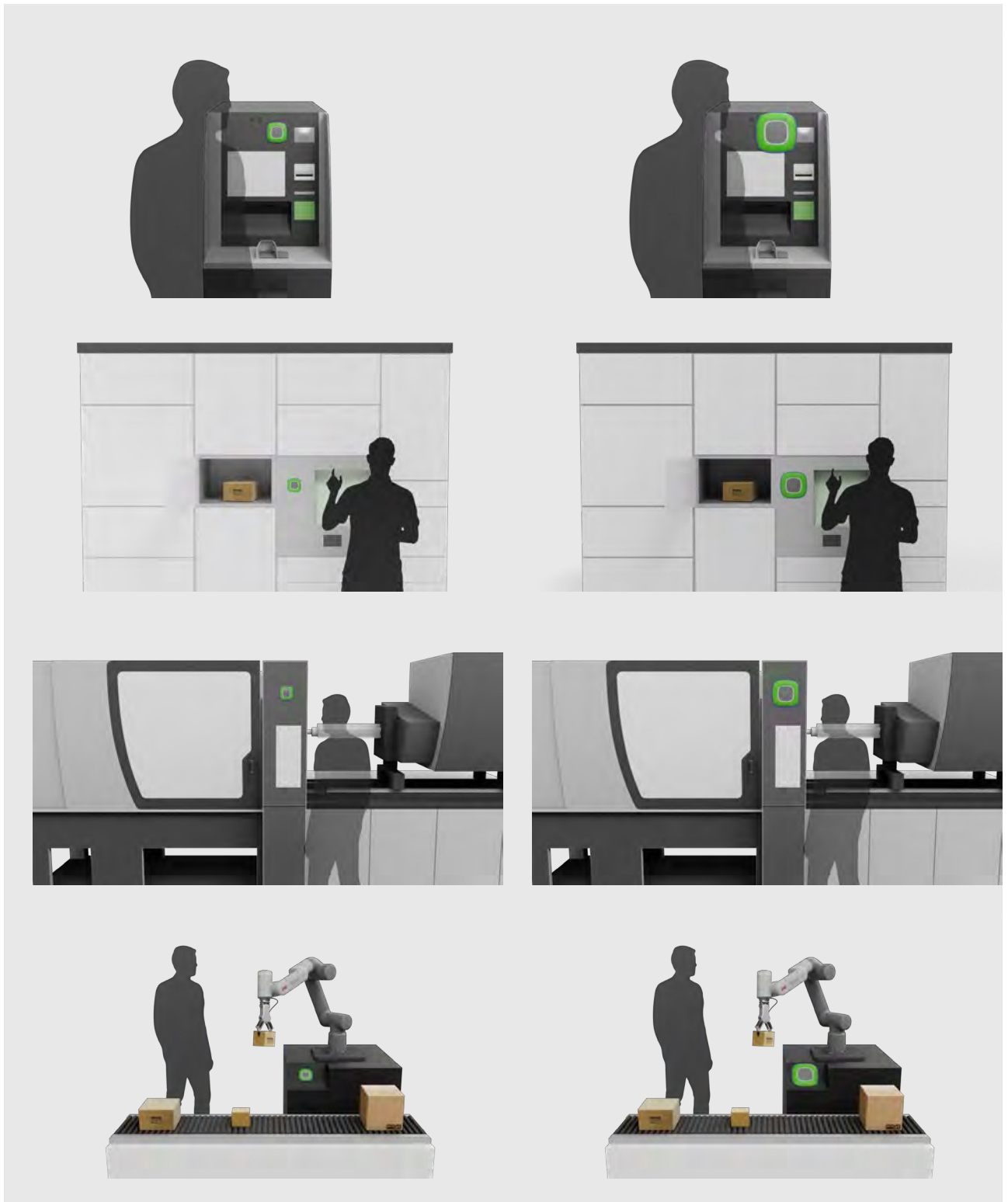


Abb. 63: Vier Favoriten der Matrix

Um definieren zu können, wie groß das Interface tatsächlich werden soll, wurde es in unterschiedlichen Einsatzbereichen dargestellt. Der Kontext hilft dabei die Größen besser einordnen zu können. In den Renderings wurde das Interface jeweils mit 10 und 20 cm Breite dargestellt. Zusätzlich wurden drei Größenmodelle mit 10, 15 und 20 cm Breite 3D-gedruckt und miteinander verglichen. Nach Abstimmung mit den Verantwortlichen bei KEBA und den Projektbetreuenden der Universität einigte man sich auf eine Größe von 12 cm Breite für das Interface.



**Abb. 64:** Darstellungen des Interface mit 10cm Breite (links) und 20cm (rechts)

## 9. Technische Analyse von Smart Speakern

Als Vorbereitung für das technische Package wurden mehrere Smart Speaker analysiert. Hauptaugenmerk wurde auf die Anordnung der LED und die Integration des Lautsprechers ins Gehäuse gelegt.

### 9.1. Amazon Echo

Die Amazon Echo Geräte haben die LED kreisförmig angeordnet und lenken das Licht über einen flachen Lichtschacht nach außen. Ein umlaufender, transluzenter Kunststoff dient als Diffuser. Die neueren Modelle haben nun 24 statt zuvor 12 LED.

Überraschenderweise ist der Lautsprecher im Echo der 3. Generation klein im Verhältnis zum ganzen Gerät. Was die Qualität verbessert ist, dass er fix mit der schweren Basis aus Metall verbunden ist. Die 3. Generation verfügt über vier Mikrofone, eingebaut in die Deckelfläche. Die 5. Generation verfügt bereits über zehn Mikrofone, drei an jeder Seite und vier oben zwischen den Bedienelementen. Die vier Tasten, Stumm, Lauter, Leiser und Aktion, befinden sich an der Oberseite.



Abb. 65: Technische Analyse von Amazon Echo 2. Generation (links) und des Echo der 5. Generation (rechts)

### 9.2. Apple HomePod

Der Apple HomePod (1. Generation), wie auch der HomePod Mini besitzen 19 LEDs, die im Zentrum der Platine kreisförmig angeordnet sind. Wenige Millimeter darüber liegt eine trübe Kunststoffplatte, die das Licht streut. Oberhalb des Diffusors befindet sich ein touchsensitives Display, das das Licht ein zweites Mal streut und so für weiche Farbverläufe sorgt. Der HomePod der 2. Generation hat eine größere Lichtfläche, die mit 37 LED bespielt wird. Der Aufbau ist wie bei den vorangegangenen Modellen.

Im Verhältnis zur Gesamtgröße hat der HomePod Mini den größten Lautsprecher. Ausgerichtet ist der Speaker nach unten. Ein Trichter lenkt Schallwellen in alle Richtungen zur Seite um. Im Resonanzraum befindet sich noch zwei passive Schwingkörper.

Der HomePod Mini ist mit insgesamt vier Mikrofonen ausgestattet. Drei davon sind nach außen gerichtet, während eines nach innen gerichtet ist, um den Klang des zentralen Lautsprechers zu erfassen. Die Bedienelemente befinden sich auf dem Touchscreen am Deckel.



Abb. 66: Technische Analyse des Apple HomePod mini (links) und des HomePod 2. Generation (rechts) (Dixon, 2021)

### 9.3. Google Nest Mini

Der Google Nest Mini verfügt nur über vier einzelne LEDs, die direkt unter der Deckfläche liegen. Wie auch der HomePod Mini, verfügt der Smart Speaker von Google über einen Trichter, der die Schallwellen umlenkt.

Dem Google Nest Mini reichen drei Mikrofone zur ausreichenden Spracherkennung. Außerdem verfügt er über einen Näherungssensor, über den man das Gerät stummschalten kann, ohne es zu berühren.



Abb. 67: Technische Analyse des Google Nest Mini

## 9.4. Fazit

Je mehr LED verwendet werden, bzw. je besser das Licht gestreut wird, desto überzeugender wirkt die Illusion einer durchgängigen Lichtfläche. Ein oder mehrere Diffusoren, helfen das Licht besser zu streuen. Ein Trichter über dem Lautsprecher und ein schweres Gehäuse verbessern ebenfalls die Klangqualität. Passive Schwingenelemente brauchen einiges an Platz im ohnehin begrenzten Bauraum. Mehrere Mikrofone können zu besserer Spracherkennung und Sprecherlokalisierung führen. Näherungssensoren helfen dabei, dass das Gerät weniger berührt und dadurch verschmutzt wird.

## 10. Technisches Package

Das technische Package enthält alle erforderlichen Komponenten, um den erarbeiteten Entwurf umzusetzen. Die Bauteile wurden mithilfe von CAD angeordnet, um den benötigten Platz zu untersuchen. Als Vorlage für das erste Konzept dienten hauptsächlich Bauteile aus anderen Smart Speakern. Basierend auf diesem ersten Entwurf wurden in einem iterativen Prozess weitere Packages entwickelt. In Abstimmung mit den Projektbetreuenden bei KEBA und den Audio-Experten von CREOIR wurden verschiedene Audio- und Lichtkonzepte entwickelt.

CREOIR ist ein Anbieter von Offline-Voice-Solutions aus Finnland. Ihre Technologie bietet eine flexible und sichere Möglichkeit für natürliche Sprachinteraktion ohne Internetverbindung. (vgl. Creoir, 2023) Die folgende Auflistung beschreibt die einzelnen Bauteile genauer.

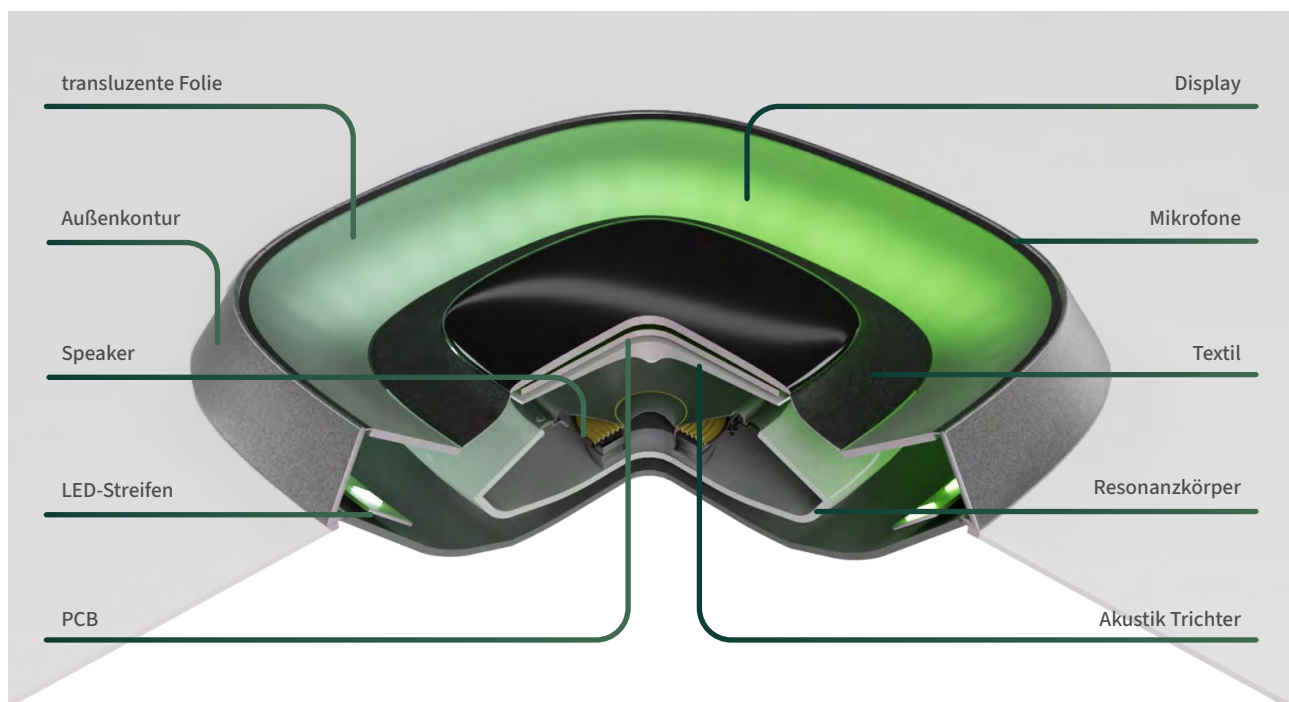


Abb. 68: Erster Entwurf eines technischen Konzepts

### 10.1. Mikrofone

Die am Interface befindlichen Mikrofone erfüllen zwei Funktionen. Erstens nehmen sie die akustischen Signale des Nutzers auf und wandeln sie um. Um dabei optimale Ergebnisse zu erzielen, werden mehrere Mikrofone eingesetzt. Zweitens dienen sie dazu den Benutzer zu lokalisieren, also die Richtung auszumachen aus der gesprochen wird. Diese Information wird genutzt, um die visuelle Interaktion

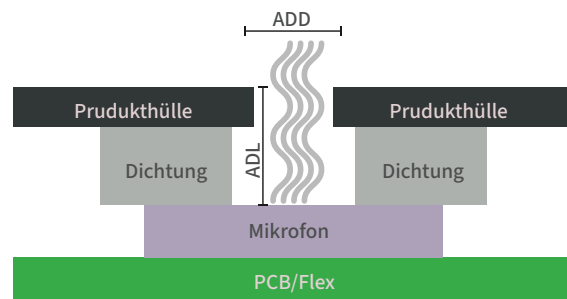
darauf abzustimmen. Dazu wird das Prinzip des Beamforming angewendet.

Um eine effektive Spracherkennung zu gewährleisten ist es wichtig, dass die Mikrofone optimale Bedingungen vorfinden. Dazu gehören bestimmte Grundregeln, die beachtet werden sollten: (vgl. Bernstein, 2019, p. 183)

- Wegen akustischer Rückkopplung zwischen Lautsprecher und Mikrofon muss die unempfindliche Seite des Mikrofons immer dem Lautsprecher zugewandt sein.
- Zur Einschränkung von störendem Nachhall sollten die Mikrofone immer der Nutzschallquelle zugewandt sein.
- Diffus einfallender Schall muss weitgehend gedämpft werden.
- Mikrofone sollten möglichst weg von möglichen Störquellen positioniert werden.

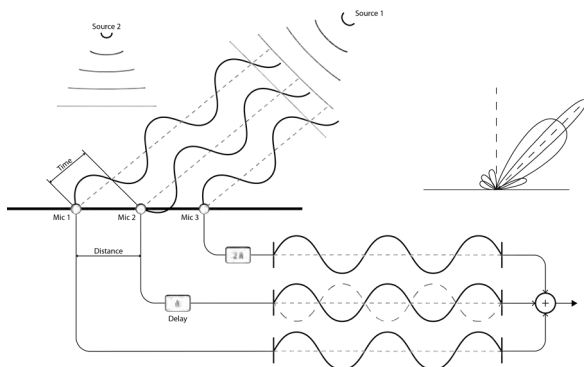
Eine der größten Quellen für Variationen in der akustischen Leistung von Mikrofonen ist die Resonanzfrequenz des Mikrofongehäuses (Microphone Enclosure System Resonant Frequency = MESRF), die aufgrund von Fertigungstoleranzen im Zusammenhang mit Produkt- und Mikrofonkomponenten erheblich in Bezug auf Amplitude und Frequenz variieren kann. Bei der Konstruktion der Mikrofonöffnung muss auf das Verhältnis von Durchmesser (Air Duct Diameter/ADL) zu Länge (Air Duct Length/ADD) geachtet werden, dies darf nicht größer 4 sein (Creoir, 2021).

**Abb. 69:** Vorgaben für die Mikrofonöffnung in der Außenhülle (Creoir, 2021)

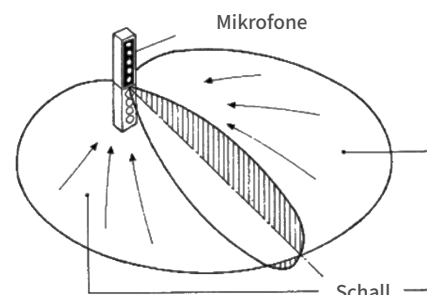


### 10.1.1. Beamforming

Um eine Schallquelle relativ zum Empfänger zu lokalisieren, bedarf es einer besonderen Technik, dem Beamforming. Die Schalleinfallrichtung, wird über die Laufzeitunterschiede zwischen den Mikrofonen ermittelt. Besonders hilfreich ist das in Umgebungen mit viele Hintergrundgeräuschen. So können die Mikrofone auf eine Schallquelle *ausgerichtet* werden (vgl. Kacher, 2017).



**Abb. 70:** Übersicht Funktionsweise Beamforming (Collings, 2020)



**Abb. 71:** Richtcharakteristik einer Mikrofongruppe (Bernstein, 2019, p. 186)

### 10.1.2. Microphone-Array

Sind mehrere Mikrofone nah beieinander angeordnet, könne die Inputs der einzelnen Mikrofone zu einem akustisch abhörbaren Summensignal zusammengeführt werden. (vgl. Kacher, 2017)

Um ein optimales Beamforming zu gewährleisten ist eine geeignete Mikrofon Anordnung nötig. Die Experten von CREOIR in Abstimmung mit KEBA haben folgenden Vorgaben für die Anordnung der Mikrofone erarbeitet.

- Vier Mikrofone in einer kreisförmigen Anordnung sind optimal für horizontale und vertikale Einbaupositionen.
- Ein geeigneter Radius ist zwischen 20 und 42 mm.
- Daraus folgt ein maximaler Abstand zwischen den Mikrofonen von 60 mm

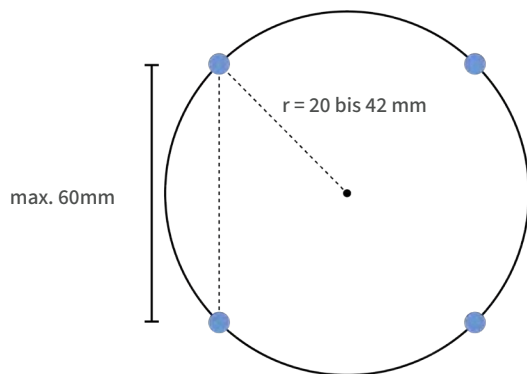


Abb. 72: Vorgaben für die Anordnung der Mikrofone (Creoir, 2021)

## 10.2. Lautsprecher

Der Lautsprecher muss folgende Anforderungen erfüllen. Erstens, gesprochene Sprache klar und deutlich wiedergeben und zweitens laut genug sein, um im industriellen Umfeld gehört zu werden. Die Vorgabe sind einen Schalldruckwert von 80dB in 3m Abstand zu erreichen.

Der Vorschlag von CREOIR ist ein 4“ Zoll co-axialen Driver der 70Hz - 20kHz mit 89dB wiedergeben kann zu verwenden.

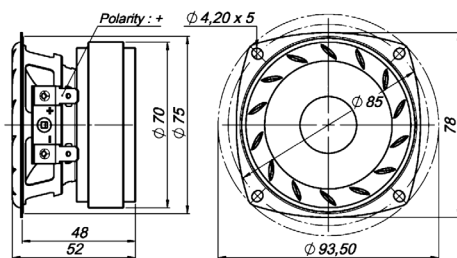


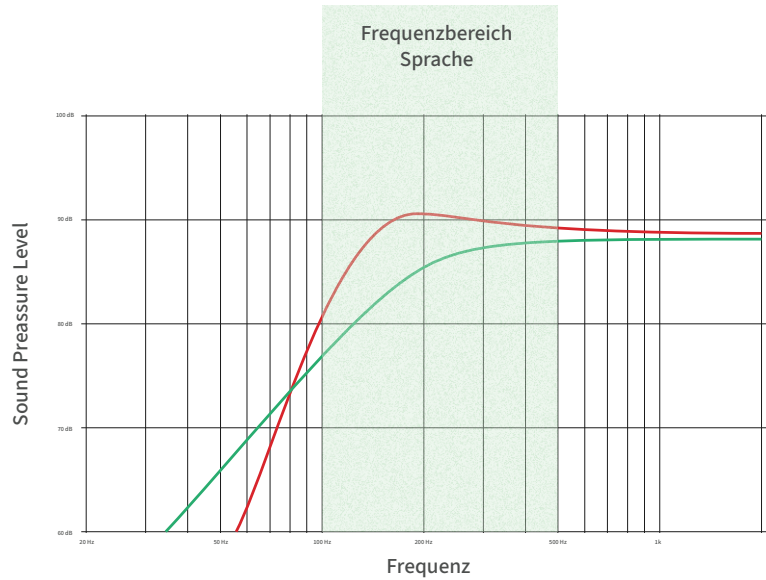
Abb. 73: Vorschlag für den Lautsprecher (Creoir, 2023)

### 10.2.1. Resonanzkörper

Der oben genannte Lautsprecher benötigt etwa 5l Volumen als Resonanzkörper. Es ist jedoch möglich das Volumen auf zwei Liter zu reduzieren da man für die Stimmwiedergabe keine sehr tiefen Frequenzen benötigt, weiters ist darauf zu achten parallele Flächen innerhalb des Resonanzkörpers zu vermeiden um keine ungewollten Schallreflexionen zu erzeugen. Um die zusätzlich die Klangqualität zu erhöhen kann Dämpfungsmaterial, beispielsweise Dämmwolle, in den Resonanzkörper eingebracht werden.

**Abb. 74:** Akustik Model mit einem optimalen Lautsprechers in 3m Entfernung (Creoir, 2023)

- Geschlossener Resonanzkörper
- Bass-Reflexgehäuse (Offenes Gehäuse)

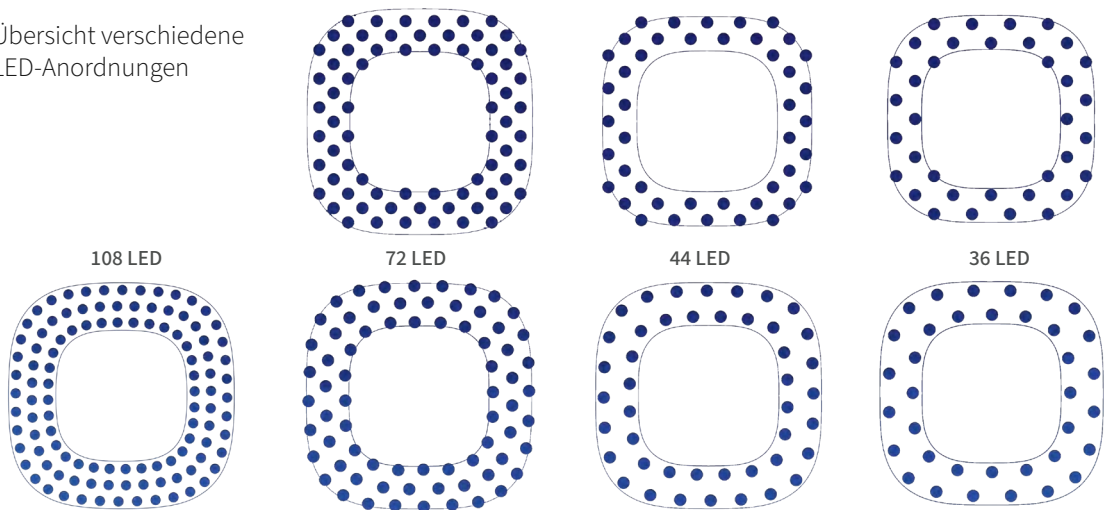


### 10.3. LED

RGBW-LEDs ermöglichen eine breite und intensive Farbwiedergabe.

Es ist wichtig, dass die LEDs gleichmäßig angeordnet sind, um Hotspots oder dunkle Bereiche zu vermeiden. Mithilfe von Grashopper wurde eine optimierte Anordnung auf der Leiterplatte errechnet. Um die Darstellung von Wellen, die von innen nach außen laufen, gut sichtbar zu machen, sind wahrscheinlich drei Reihen an LEDs erforderlich. Eine fundierte Antwort kann erst nach der Testphase mit einem funktionsfähigen Prototyp gegeben werden.

**Abb. 75:** Übersicht verschiedene LED-Anordnungen



### 10.4. Optischer Diffusor

Ein optischer Diffusor streut das einfallende Licht, um weiche Farbübergänge und eine gleichmäßige Lichtverteilung zu erzeugen. Er kann in Form von Folien oder Platten verwendet werden

Der optimale Abstand zwischen den LEDs, dem optischen Diffusor und der Deckfläche des Interface soll anhand eines Prototyps getestet werden.

## 10.5. Display

Das Display im Zentrum soll aus einer einfachen Matrix aus 8x8 LED bestehen.

## 10.6. Textiler Überzug

Die Anforderungen an den Stoff, der den zentralen Teil des Interfaces überspannt sind vielfältig.

- Akustisch transparent für die Mikrofone
- Blickdicht um Display und Mikrofone zu verdecken
- Lichtdurchlässig für die LED der Matrix
- Widerstandsfähig für Reinigung industrieller Umgebung

## 10.7. Elektronik

Da die Rechenleistung vom KI-Modul übernommen wird, sind nur folgende Elektrobauteile im Interface verbaut:

- Trägerplatte für die LEDs
- Steuerungseinheit für die LEDs
- Leiterplatte für Mikrofone und LED-Matrix
- Stromversorgung

## 10.8. Gehäuse

Der Rahmen umschließt alle Komponenten des Interfaces und muss dicht zur Maschinenhülle abschließen. Er muss den Anforderungen einer industriellen Reinigung gerecht werden und mechanischen Belastungen standhalten können.

## 10.9. Audiokonzepte

Aufgrund der flachen Integration in die Maschinenhülle ergibt sich ein potenzielles Problem zwischen den Mikrofonen und dem Lautsprecher. Um Interferenzen zu vermeiden war eine Änderung des Konzepts notwendig.

Gemäß den Empfehlungen der Experten von CREOIR soll der Lautsprecher nicht in das Interface integriert, sondern extern positioniert werden. Um ein effektives Mikrofon-Array zu gewährleisten, müssen die Mikrofone im Zentrum platziert und in die gleiche Richtung ausgerichtet werden. Sitzt der Lautsprecher direkt darunter kann es zu unerwünschten Interferenzen kommen, die die Spracherkennung erheblich beeinträchtigen würden.

Die konstruktive Trennung von Interface und Lautsprecher bringt auch Vorteile. Die Einbauhöhe des Interface verringert sich deutlich ohne den Resonanzkörper des Lautsprechers. Auch die Konstruktion des Interface gestaltet sich weniger komplex. Eine Konsequenz der Trennung ist jedoch, dass die Stimme des Assistenten nicht mehr direkt vom Interface selbst kommt. Daher sollte darauf geachtet werden, den Lautsprecher möglichst in der Nähe, aber nicht unmittelbar daneben zu positionieren.






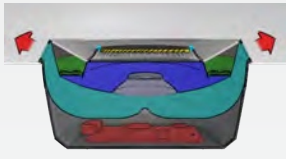
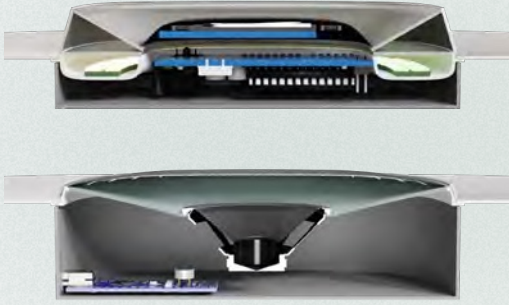



<p>01</p> 	<p>🔊 Lautsprecher zeigt nach oben</p> <p>🎤 Mikrofone im Zentrum</p> 	<ul style="list-style-type: none"> <li>+ All in One Lösung</li> <li>+ Zentrales Mikrofon-Array</li> <li>+ Geschlossener Resonanzkörper</li> <li>- Mikrofone sind über dem Speaker</li> <li>- Nur kleine Öffnung für den Lautsprecher</li> <li>- Wenig Platz im Gehäuse</li> </ul>
<p>02</p> 	<p>🔊 Lautsprecher zeigt unten</p> <p>🎤 Mikrofone im Zentrum</p> 	<ul style="list-style-type: none"> <li>+ All in One Lösung</li> <li>+ Schall kann sich ungehindert ausbreiten</li> <li>+ Große Öffnung für Lautsprecher</li> <li>- Mikrofone nicht optimal angeordnet</li> <li>- Sonderanfertigung für zentrale LED-Matrix</li> </ul>
<p>03</p> 	<p>🔊 Lautsprecher zeigt nach unten</p> <p>🎤 Mikrofone im Zentrum</p> 	<ul style="list-style-type: none"> <li>+ All in One Lösung</li> <li>+ Zentrale Mikrofon-Array</li> <li>+ Große Resonanzkörper</li> <li>- Kleine Öffnungen an der Seite</li> <li>- Schallwellen richten sich zur Seite</li> </ul>
<p>04</p> 	<p>🔊 Externer Lautsprecher</p> <p>🎤 Mikrofone im Zentrum</p> 	<ul style="list-style-type: none"> <li>+ Keine Interferenzen zwischen Lautsprecher und Mikrofon</li> <li>+ Geringe Einbautiefe</li> <li>+ Großer Resonanzkörper möglich</li> <li>+ Vereinfachtes Design</li> <li>+ Schall kann sich ungehindert ausbreiten</li> <li>- Zwei Objekte an Maschine</li> <li>- Externe Sprachquelle</li> </ul>
<p>05</p> 	<p>🔊 Lautsprecher zeigt nach unten</p> <p>🎤 Mikrofone im Zentrum</p> 	<ul style="list-style-type: none"> <li>+ Stand-alone Interface</li> <li>+ Platz für Passive Radiators</li> <li>+ Abgeschirmete Mikrofone</li> <li>+ Großer Resonanzkörper</li> <li>+ Transportierbares Interface</li> <li>- Kabelgebundenes Gerät</li> <li>- Großer Körper</li> <li>- Wird viel berührt</li> </ul>

Abb. 76: Übersicht verschiedener Konzepte für die Anordnung des Lautsprechers und der Mikrofone

# **IV. Beispielung**

**Interface**

# 1. Gestaltungsansatz

Das Interface wirkt ohne Lichteffekte sehr technisch und neutral. Während der Interaktion soll es agil, dynamisch und unverwechselbar sein wie ein Mensch. Immerhin soll es unter anderem die Informationen übermitteln, die sonst Aufgabe der Gesichtszüge, Blicke und Mundbewegungen sind.

Die visuelle Bespielung der Leuchtf lächen soll sowohl den Status der Maschine als auch den der KI-Steuerung abbilden. Es ist wichtig, dass zu jeder Zeit deutlich erkennbar ist, ob alles reibungslos funktioniert oder ob eine Intervention erforderlich ist.

Das Interface ist primär ein Sprachinterface, die Lichteffekte sollen daher auch visuell nicht die höchste Priorität haben. Im Zentrum der Aufmerksamkeit des Nutzers sollte stets die aktuelle Aufgabe sein. Dazu gehört, dass der Blick so viel wie möglich auf Hände und Werkstück gerichtet ist. Alles außerhalb dieses Fokusbereichs bleibt unscharf. Aus diesem Grund ist die Visualisierung auf dem Interface nur diffus. Es bleibt sozusagen immer im peripheren Sichtfeld. Der Nutzer ist nicht dazu verleitet, länger als nötig das Interface direkt anzusehen. Es reicht ein flüchtiger Blick, um den Status auszumachen. Die unscharfe Darstellung befreit einen von der Notwendigkeit, jedes Detail zu interpretieren.

# 2. Entwurfentwicklung der Lichteffekte

Die ersten Versionen der Lichteffekte waren schwarz-weiße Zyklen, die in drei verschiedenen Stilen erstellt wurden: *Diffuse Linien*, *Ringe* und *Partikel*. Diese Stile waren sehr unruhig und plakativ. Um die Effekte zu veranschaulichen, wurden einige Animationen auf die beiden Favoriten gemappt.

Es stellte sich heraus, dass die Varianten mit sanfte Farbverläufe mehr Zuspruch bei den Projektbeteiligten bekamen als die kontraststarken Übergänge.

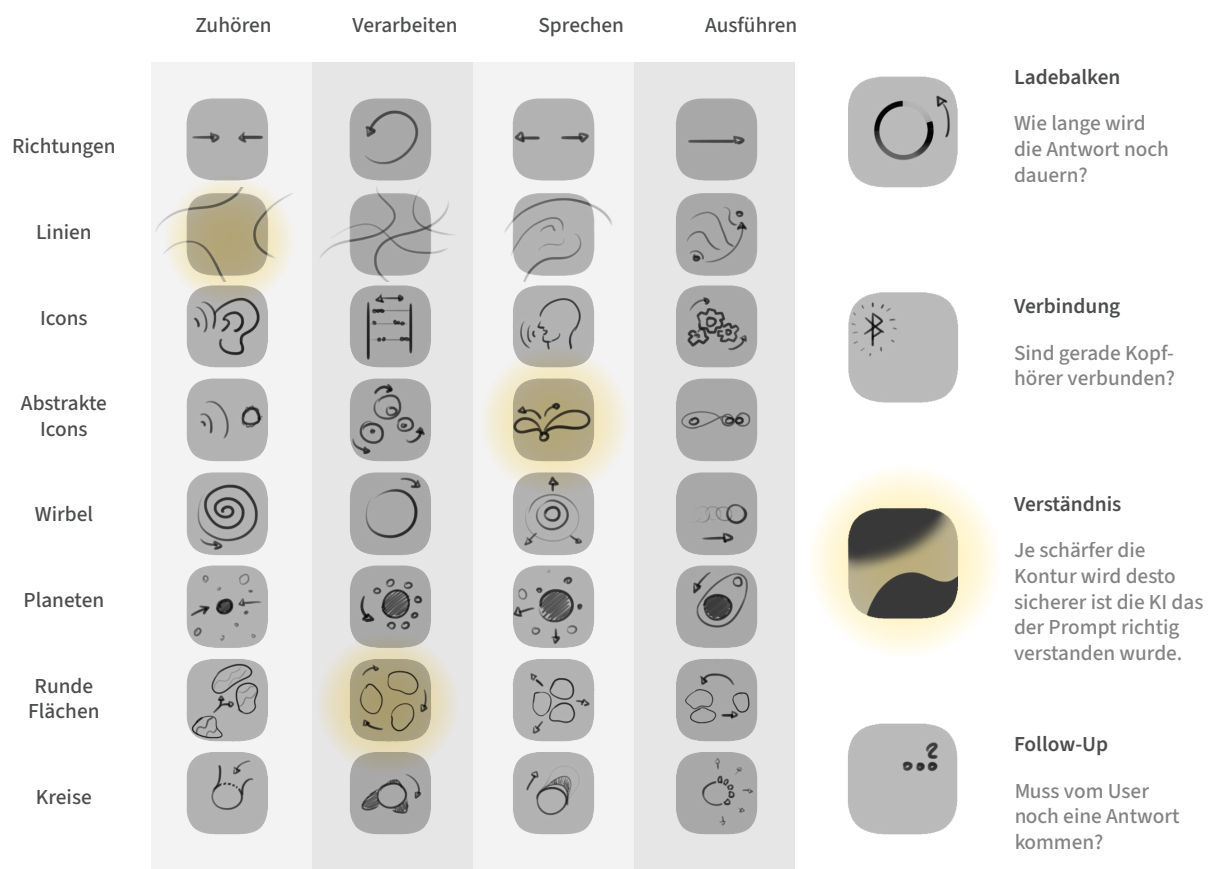


Abb. 77: Ideation für die Bespielung der Leuchtf lächen



Abb. 78: Übersicht verschiedener Konzepte für die Bespielung der Leuchtf lächen

### 3. Mock-Up

Das erste Mock-Up dient der Veranschaulichung des Konzepts und ist in der Bespielung sehr eingeschränkt. Es wird beleuchtet von einem RGB-LED-Ring mit 12 LEDs. Das Zentrum ist noch nicht hinterleuchtet. Abgespielt wird eine einfache Lichtsequenz in Dauerschleife. Um eine möglichst gleichmäßige Streuung des Lichts zu erreichen hat man verschieden 3D-Druckmethoden ausprobiert. Die besten Ergebnisse lieferte der Resin-Druck.

Das zweite Mock-Up hat einen umlaufenden LED-Strip, der die große Fläche beleuchtet und einen kleinen LED-Ring der das Zentrum hinterleuchtet. Es steht im Showroom der Firma KEBA.

Für eine möglichst diffuse Darstellung ohne einzelne Lichtpunkte sorgt ein Adafruit NeoPixel-LED-Strip mit 332 LED/m. Dieser kann auch in mehreren Farben leuchten.

Die Abdeckung im Zentrum ist austauschbar. Die Farbverläufe wurden auf eine transparente Folie gedruckt, während die Symbole mit einem Lasercutter aus undurchsichtiger, schwarzer Folie geschnitten wurden. Als Träger dienten Acrylglasplatten.



Abb. 79: Fotos und Screenshots des ersten MockUp

So können verschiedene Arten der Bespielung des Zentrums getestet werden. Die Mock-Up entstanden während der Entwurfsphase und halfen dabei die Größenverhältnisse und mögliche Bespielungen besser zu beurteilen. Es zeigt sich das große Farbverläufe am äußeren Ring sehr gut ankommen, im Zentrum jedoch eher für Verwirrung sorgen. Eindeutige Symbole wie das Mikrofon, ein Häkchen und ein Ladebalken kommen bei den Projektbeteiligten besser an.



Abb. 80: Fotos und Screenshots des zweiten MockUp

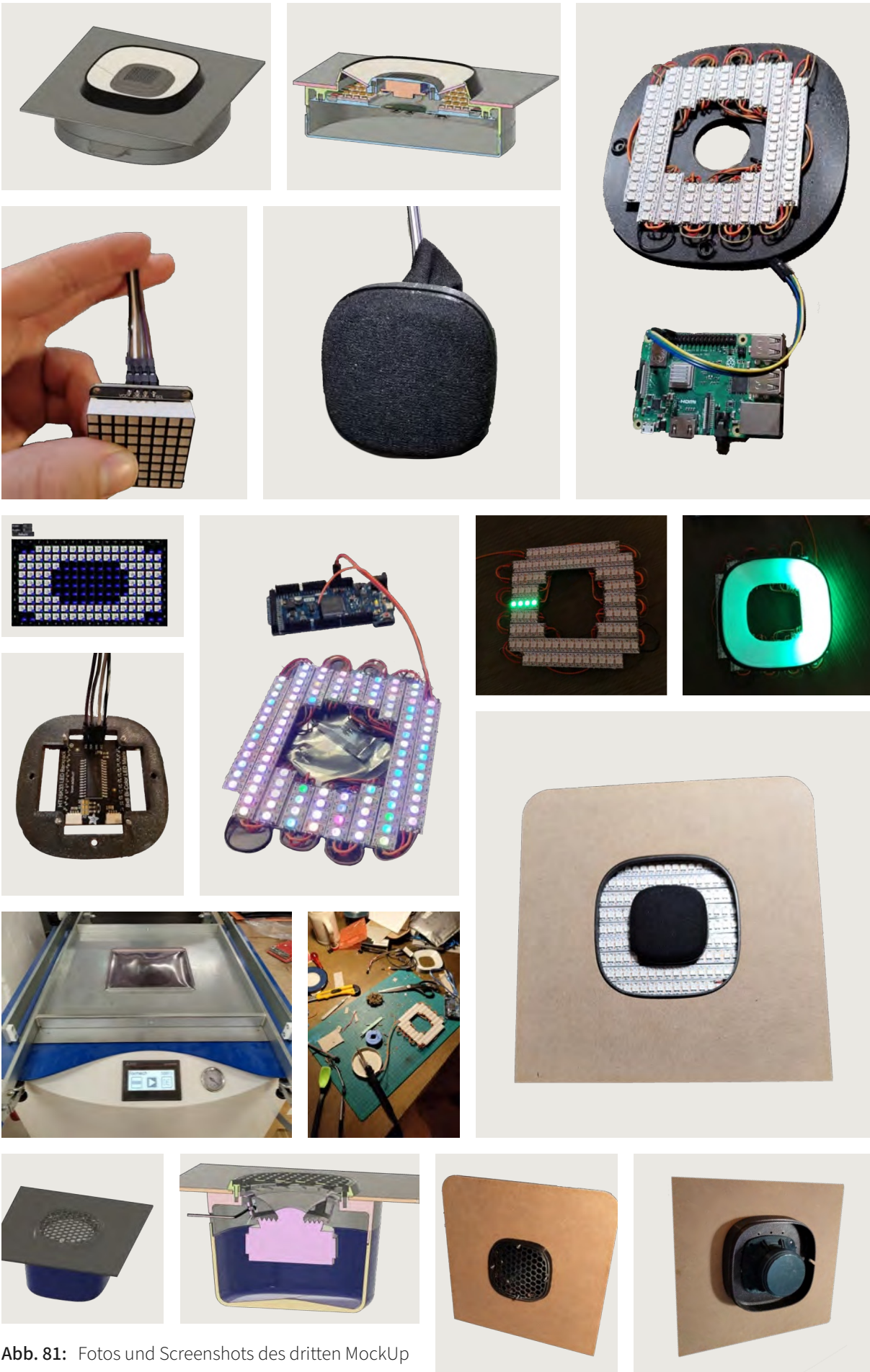


Abb. 81: Fotos und Screenshots des dritten MockUp

## 4. Aufgabenverteilung

Während des Entwurfsprozesses wurden unterschiedliche Interpretationen bezüglich der Aufgabenverteilung und der Priorität der Bereiche diskutiert. Es galt festzulegen, was die beispielbare Fläche repräsentieren soll. Es wurde vereinbart, dass der Rand die Intelligenz und den Status des gesamten Systems widerspiegeln soll, während das Zentrum lediglich eine unterstützende Funktion hat.

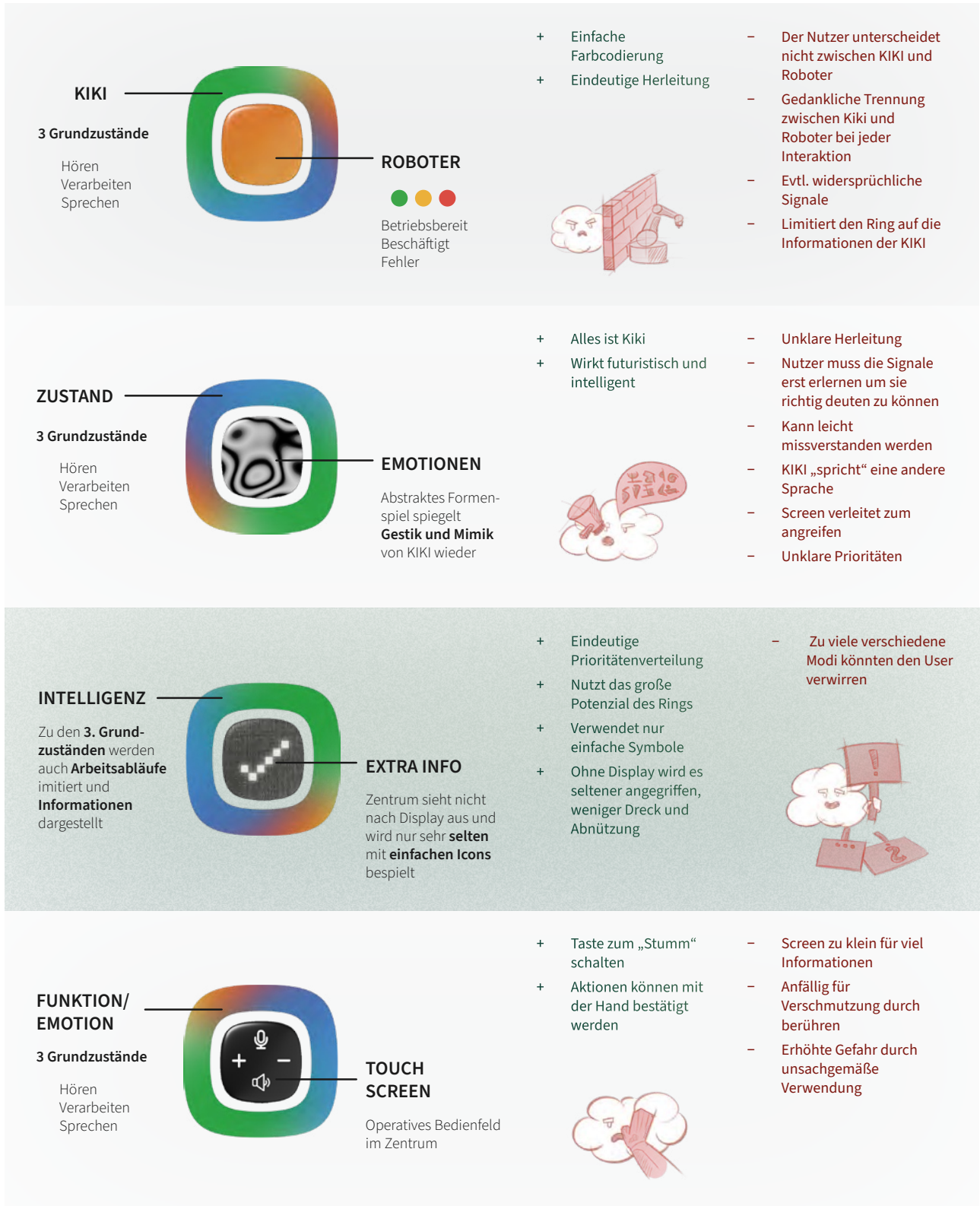


Abb. 82: Varianten der Aufgabenverteilung

## 5. Modi

Nach der Analyse der Smart Speaker und Überlegungen zu Statusanzeigen von Maschinen wurden folgende Grundmodi festgelegt. Um auch im peripheren Sichtfeld gut erkennbar zu sein, sind die Modi klar unterscheidbar. Während das Interface spricht, unterstützen Wellenanimationen den jeweiligen Modus und visualisieren zusätzlich die Sprache. Durch Veränderung der Amplitude oder Frequenz der Wellen kann beispielsweise die Dringlichkeit einer Aussage verdeutlicht werden.

### 5.1. Hören

Während der Nutzer mit dem Interface spricht, zeigt eine Animation, dass gerade zugehört wird. Um das zu erreichen, werden verschiedene Effekte kombiniert.

- Wellen, die von außen zur Mitte laufen deuten an, dass Input aufgenommen wird.
- Um dem Nutzer zu verdeutlichen, dass er wirklich wahrgenommen wird, wird die Animation in die Richtung aus der gesprochen wird, besonders intensiv sein.
- Je länger der Nutzer spricht, desto mehr füllt sich die ganze Fläche, ausgehend von dem Bereich der in Richtung des Nutzers zeigt.

### 5.2. Interne Handlung

Hat der Nutzer zu Ende gesprochen, wechselt das Interface den Modus und zeigt, dass der Befehl verarbeitet wird. Dieser Vorgang ist aufwendig und für die Benutzer entsteht Wartezeit. Sinnbild fürs Warten am Computer ist ein Rad, das sich im Kreis dreht. Eine Endlosschleife, die zeigt, dass gearbeitet wird. Wenn die Bewegung ins Stocken gerät, deutet dies auf Probleme bei der Verarbeitung hin.

- Ein Bereich, der etwa  $\frac{1}{4}$  des Rings in Anspruch nimmt, läuft im Uhrzeigersinn über die Fläche. Dauert die Verarbeitung länger als gewöhnlich, variiert die Geschwindigkeit.

### 5.3. Sprechen

Während das Interface spricht, visualisiert es den Informationsfluss aus dem Inneren nach außen. Ähnlich wie beim *Hören* gibt es einen Indikator, der die Position des Nutzers anzeigt. Die Richtung wurde zuvor während des *Hörens* ermittelt. Die Animation gleicht der des Zuhörens, sie läuft jedoch in umgekehrter Reihenfolge ab.

- Wellen bewegen sich aus der Mitte nach außen, besonders intensiv in die Richtung des Nutzers
- Die Fläche füllt sich entsprechend der Dauer der Aussage

### 5.4. Externe Handlung

Die Lichtsequenz dient dazu die Aktionen der Maschine zu verdeutlichen. Es spiegelt die Arbeit die aktuell verrichtet wird wider. Die Animationen sind spezifisch und auf die gesteuerte Maschine abgestimmt. Beispielsweise ein Roboterarm hebt Objekte an oder eine Spritzgussmaschine öffnet und schließt das Werkzeug. Es ist wichtig, die Position des Interfaces in Bezug zur Maschine zu berücksichtigen, um eine korrekte Darstellung von links, rechts, oben und unten zu gewährleisten.

- Eine Farbfläche in der Größe eines  $\frac{1}{4}$  Rings bewegt sich schnell von links nach rechts und kehrt etwas langsamer wieder zurück

## 5.5. Ruhemodus

Diese Animation wird wahrscheinlich die meiste Zeit dargestellt werden. Bedient niemand die Maschine, schaltet das Interface in einen Ruhemodus. Dennoch ist es wichtig, dass deutlich erkennbar ist, dass erstens alles in Ordnung ist und zweitens das Interface jederzeit bereit ist, einen Befehl entgegenzunehmen. Im Ruhemodus bewegen sich große Farbfelder durcheinander über das Interface und ändern dabei ihre Farbe. Um keine unnötige Aufmerksamkeit auf sich zu ziehen, geschieht dies langsam und mit verringerter Helligkeit.

- Langsames, diffuses Farbenspiel mit verringerter Helligkeit und Kontrast

## 5.6. Problem

Eine Warnung muss aus allen Blickwinkeln sofort die Aufmerksamkeit des Nutzers erreichen. Schnelle Bewegungen und Signalfarben fallen unverzüglich ins Auge. Dabei kann es noch Abstufungen nach Dringlichkeit der Meldung geben.

- Intensive Wellen bewegen sich schnell in alle Richtungen

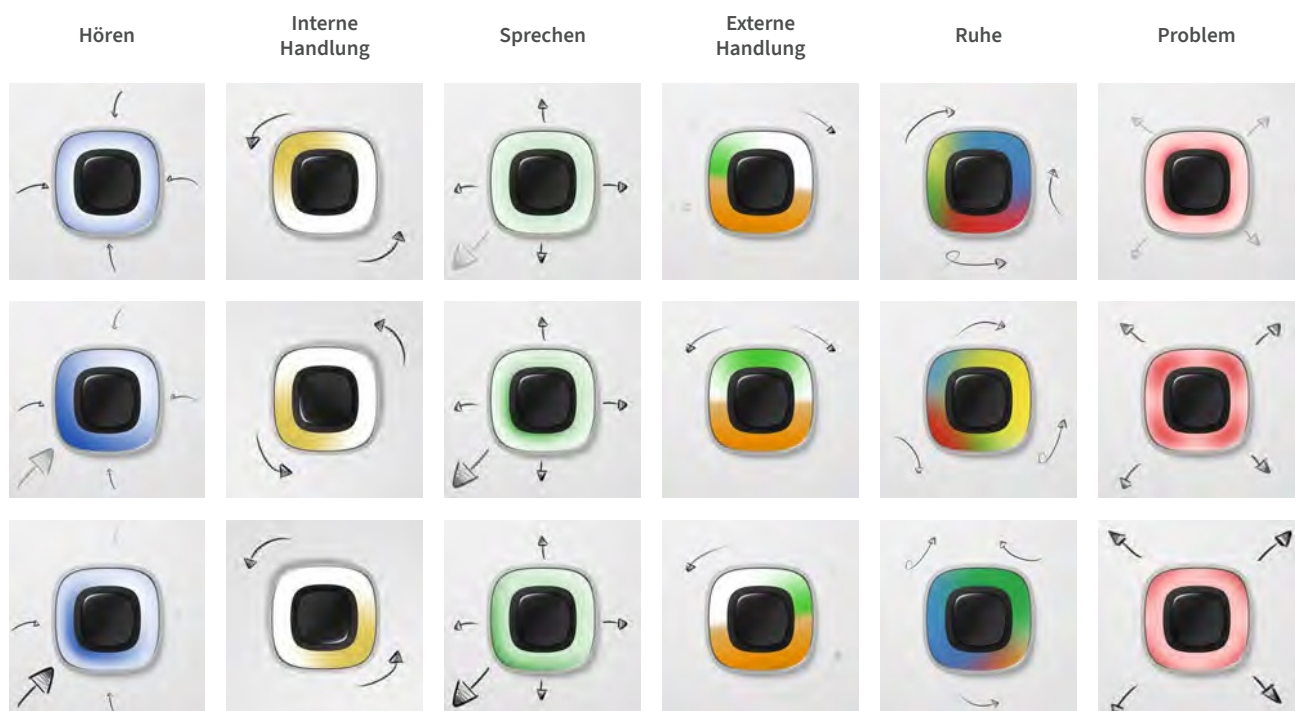


Abb. 83: Übersicht der Animationsmodi

## 6. Übergänge

Für eine reibungslose Interaktion sind die Übergänge zwischen den einzelnen Modi besonders wichtig. Anstelle von abrupten Wechseln sollen die Übergänge fließend gestaltet werden. Fließend bedeutet jedoch nicht langsam, die meisten Animationen müssen sehr schnell ablaufen. Beispielsweise, um den Ruhemodus zu starten oder zu beenden.

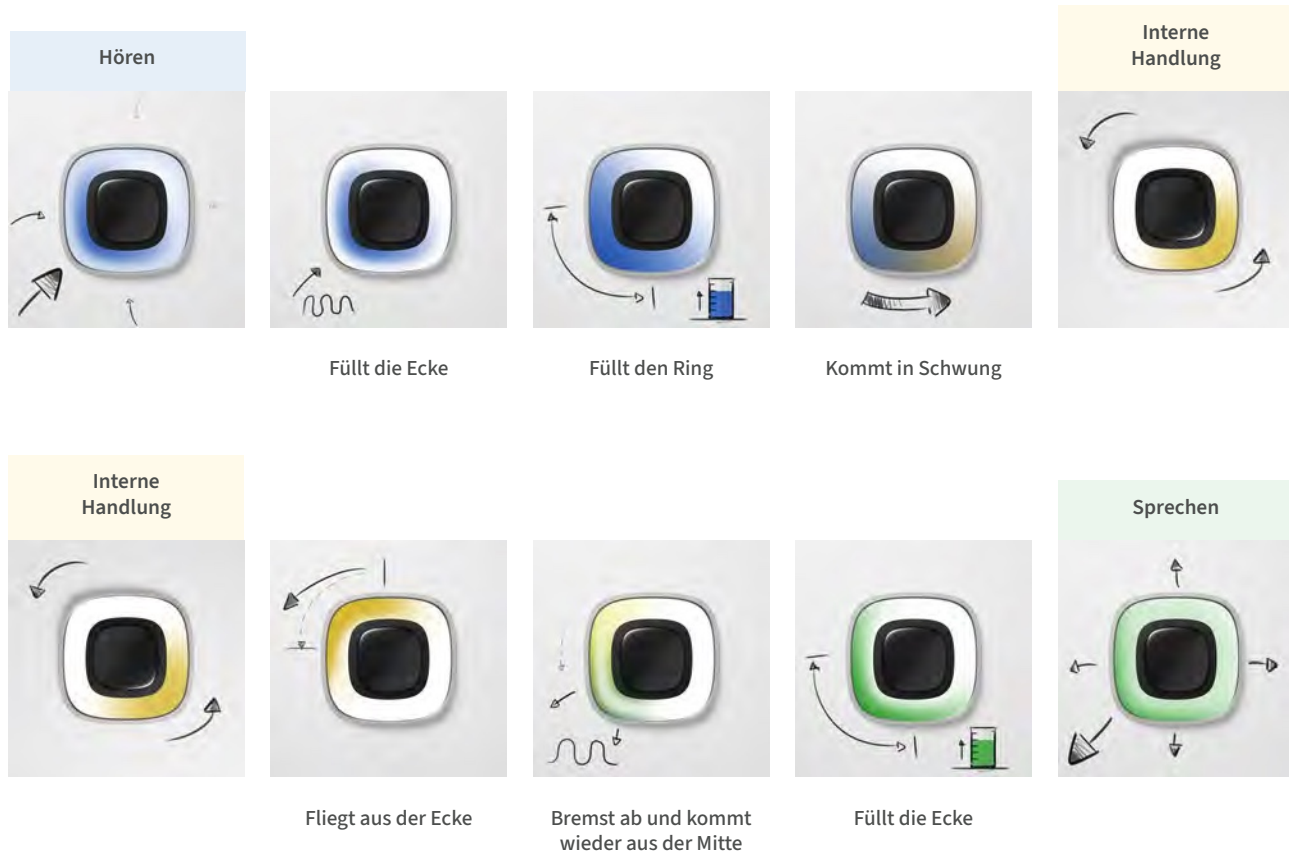


Abb. 84: Beschreibung der Übergänge von einem Modus zum Nächsten

## 7. Farbauswahl

Die gewählten Farben sind aus der Analyse der Smart Speaker abgeleitet.

- Hören ist blau (wie ein futuristisches Interface)
- Interne Handlungen sind gelb (schwache noch nicht belegte Farbe)
- Sprechen ist grün (Stimme und Interface repräsentieren KEBA)
- Externe Handlungen sind orange (Hinweis, dass gerade etwas passiert)
- Leerlauf ist bunt (Soll alles zeigen was möglich wäre)
- Probleme sind rot (Signalfarbe wird stets mit Gefahr assoziiert)

Eine fundierte Antwort, ob die vielen unterschiedliche Farben und Bewegungen den Nutzer verwirren, wird erst zu einem späteren Zeitpunkt nach entsprechenden Feldtests möglich sein.

## 8. Zentrum

Die zentrale Fläche des Interface kann einfache Symbole darstellen. Sie kommen zum Einsatz in Situationen, die nicht die unmittelbare Aufmerksamkeit des Nutzers erfordern. Um die Bedeutung der Symbole weiter zu reduzieren, wurde beschlossen, anstelle eines hochauflösenden Displays eine 8x8 LED-Matrix zu verwenden. Um dem Uncanny-Valley auszuweichen, werden keine Emojis verwendet.

Abb. 85: Darstellung mit Icon auf Display im Zentrum



### 8.1. Bestätigung

Wird beispielsweise ein Auftrag abgeschlossen ist, der lange Zeit in Anspruch genommen hat, wird das Symbol eines Hacken angezeigt. Die umliegende Leuchtfläche befindet sich bereits im Ruhemodus.

### 8.2. Mikrofon

Während das Mikrofon stumm geschaltet ist, leuchtet das ein „Mikrofon-Aus“-Symbol.

### 8.3. Verbindung

Bei Problemen mit der Verbindung zum Netzwerk oder Maschine leuchtet das Wifi-Symbol.

### 8.4. Frage

Ein Fragezeichen zeigt an das der Prompt des Nutzers nicht richtig verstanden wurde.

### 8.5. Benachrichtigung

Benachrichtigung für den Benutzer ohne zeitliche Dringlichkeit werden nur als Icon dargestellt. Ein klassisches Glockensymbol, signalisiert dem Benutzer, dass eine Benachrichtigung vorliegt, ohne dabei aufdringlich oder störend zu sein. Im Anwendungsfall der im Kapitel Interaction-Journey genauer beschrieben wird, könnte man den Benutzer mitteilen, dass eine Kiste demnächst volle Kapazität erreicht.

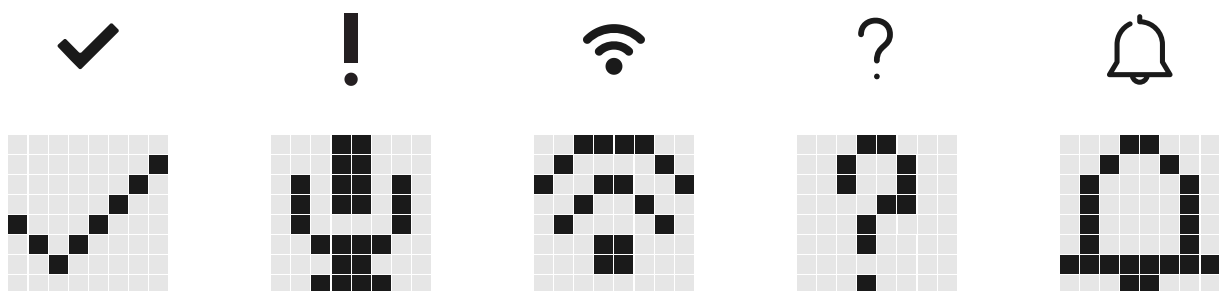


Abb. 86: Übersicht der Icons die im Zentrum dargestellt werden

# V. Interaktion



# 1. Interaction Journey

## 1.1. Interaction Journey

Um den Ablauf bei der Interaktion zwischen dem Benutzer, dem Interface und der Maschine besser nachvollziehen zu können wurden zwei Szenarien durchgespielt. Der Use-Case bezieht sich auf einen Roboterarm, der autonom Obst in Kisten sortieren kann. Mit einer integrierten Kamera können verschiedene Objekte erkannt und abgespeichert werden. Anschließend erfolgt die Sortierung der gescannten Objekte mithilfe eines Greifers. Die oben genannte Lichtkonzepte dienten als Ausgangspunkte für die Bespielung des Interface.



Im ersten Szenario möchte der Nutzer einen Apfel bewegen. Als Vorbereitung für die Ausarbeitung der Journey, mit allen beteiligten Akteuren (Nutzer, Interface, Stimme und Roboter), wurde die gleiche Journey durchgespielt, als würde sie zwei Menschen machen. Diese Mini-Journey soll zeigen, wie unkompliziert die Zusammenarbeit von zwei Menschen sein kann.



Abb. 87: Interaktion Journey „Sortieren“

Im zweiten Szenario soll der Roboter ein neues Objekt erlernen. Dies erfordert mehr Kommunikation zwischen Nutzer und Maschine.



## 1.2. Fazit und Erkenntnisse

Es stellte sich heraus, dass nur wenige Symbole, die auf dem zentralen Display angezeigt werden, einen Mehrwert für die Konversation darstellen. Allgemein gültige Symbole sind leicht verständlich, ein Ohr oder ein Fragezeichen beispielsweise. In untypischen Situationen könnten die Icons irreführend sein. Daher sollen sie keine essenzielle Rolle in der Kommunikation übernehmen, sondern lediglich eine unterstützende Funktion haben.

Ein Mensch spricht automatisch lauter, wenn der Geräuschpegel in der Umgebung zunimmt. Das gleiche soll die Stimme des Interface tun. Das automatische Anpassen der Lautstärke beugt dem störenden Einstellen mit einem extra Befehl vor.

Es ist mühsam, mehrere Dinge gleichzeitig im Gedächtnis zu behalten. Daher sollte das Vortragen von Listen nach Möglichkeit vermieden werden. Nur auf Aufforderung des Nutzers kann eine Auflistung mehr als drei Einträge umfassen.

Abb. 88: Interaktion Journey „Neues Objekt erlernen“



Befehle wie „HALT!“ oder „PAUSE!“ sollen jederzeit funktionieren. Dieser Vorgang unterbricht die laufende Aktion und die Ausgabe des Interfaces. Im Zentrum erscheint ein Pausensymbol und es wird in den Hörmodus gewechselt. Die Aktion wird erst nach explizitem Befehl des Nutzers fortgesetzt. Um nicht irrtümlich gestoppt zu werden, reagiert das Interface erst bei erhöhter Lautstärke des Befehls. Bei komplexen Befehlen, die mehrere Komponenten umfassen, werden vor der Ausführung die wichtigsten Details wiederholt, so stellt man sicher das alles korrekt verstanden wurde.

Aufgrund der Erkenntnisse der Interaction Journey wurden die ursprünglichen vier Modi (Hören, Interne Handlung, Sprechen und Externe Handlung) um zwei weitere erweitert: den Leerlauf- und den Problem-Modus. Der Leerlauf-Modus signalisiert eine Phase ohne spezifische Aktivität und der Problem-Modus weist auf ein akutes Problem hin.

## 2. Eigenschaften

### 2.1. Name und Weckwort

Der Name Kiki leitet sich von den Worten KEBA und Künstlicher Intelligenz ab. Laut der Babynamen App *CharliesNames* zählt Kiki als geschlechterneutraler Unisex Name (vgl. *charlies-names*, 2023).

Kiki eignet sich außerdem gut als Weckwort, da es im normalen Sprachgebrauch selten vorkommt und die harten „K“ Laute gut detektiert werden können.

In Anlehnung an Alexa, Cortana und bald auch Siri entfällt das „Hey“ vor „Kiki“. Dadurch wird die Hemmschwelle weiter verringert und es wird leichter, mehrere Befehle hintereinander zu sprechen.

### 2.2. Dialog

*Das Wording einer Anwendung sollte freundlich, aber bestimmt sein, und nie maßregelnd oder bevormundend klingen (Dorau, 2011, p. 21).*

Die Interaktion mit einem Nutzer soll, wenn möglich ein richtiger Dialog sein. Vorprogrammierte Kommandos werden zwar verstanden, Kiki soll die Antwort aber immer in ganzen Sätzen geben. Die Formulierungen sollen nicht konstruiert und gezwungen klingen. Das soll den Benutzer dazu verleiten normal zu sprechen und seine Ausdrucksweise nicht anzupassen.

Die Basis eines guten Gesprächs ist es, dass beide Gesprächsteilnehmer sich abwechseln und aufmerksam zuhören. Dazu gehört auch, den anderen ausreden zu lassen. Obwohl die Befehle „HALT!“ und „STOP!“ jederzeit verwendet werden können, um die laufende Handlung zu pausieren, muss im normalen Dialog das Weckwort verwendet werden, um Kiki zu unterbrechen.

Das Interface hört nach dem Sprechen automatisch für mehrere Sekunden zu, um dem Benutzer Zeit zum Antworten zu geben.

## 3. Umgang mit Fehlern

Jede Warnung, Hinweis oder Alarm wird nicht durch ein akustisches Signal, sondern sprachlich kommuniziert. Im Falle einer Störung ist es die Aufgabe von Kiki, das Problem präzise zu benennen und gegebenenfalls eine mögliche Lösung vorzuschlagen.

Bei Missinterpretationen in der Spracherkennung liegt es in der Verantwortung der KI, dieses Problem zu beheben. Der Nutzer wird entweder gebeten, den ganzen Befehl zu wiederholen, oder es wird gezielt nach dem nicht erkannten Teil gefragt.

## 4. Einstellungen

Dem Nutzer dürfen nicht zu viele Einstellungen am Interface zur Verfügung stehen. Die Lautstärke aber, muss veränderbar sein. Eine weitere Einstellung, die die Interaktion verbessern kann, ist die Geschwindigkeit der Sprache. Selbst kurze Antworten können sich frustrierend lange anfühlen, wenn die Stimme zu langsam spricht. Bei Befehlen wie „LAUTER!“ oder „LEISER!“ soll die Aussage nicht unterbrochen werden, sondern direkt umgesetzt.

## 5. Wissensbereiche

Da Kiki eine lokale KI ist, kann nur mit den verfügbaren Daten der Maschine und des Nutzers gearbeitet werden. Die Frage ist daher, welche Informationen zu Beginn in die Datenbank eingespeist werden müssen. Über welche Informationen soll Kiki verfügen. Welches Wissen setzt ein uninformatierter Nutzer voraus? Die folgende Liste bietet erste Anhaltspunkte.

Die wichtigsten Informationen beziehen sich auf die möglichen Funktionen. Am Beispiel des Roboterarms sind das das Erlernen, Greifen und Sortieren von Objekten. Darüber hinaus sollen das maximale Traggewicht und die Reichweite des Greifers abgespeichert sein. Die KI muss die Gestaltung des Interfaces kennen, damit sie Benutzern bei Fragen zu bestimmten Lichtsequenzen sofort Auskunft geben kann.

Die Positionierung im Raum ist eine weitere wichtige Information. Dadurch kann der Benutzer dem Roboter ortsbezogene Anweisungen wie links, rechts, oben, unten, vorne und hinten geben.

	Informationen über die Maschine	Möglichkeiten	<ul style="list-style-type: none"> <li>+ Scannen</li> <li>+ Bewegen</li> <li>+ Greifen</li> </ul>	<ul style="list-style-type: none"> <li>+ Reichweite</li> <li>+ Abläufe</li> <li>+ Checkliste</li> </ul>	<ul style="list-style-type: none"> <li>+ Handbuch</li> <li>+ Bauteile</li> <li>+ Betriebszeit</li> </ul>
	Informationen über sich selbst	Funktionen	<ul style="list-style-type: none"> <li>+ Hören</li> <li>+ Sprechen</li> <li>+ Leuchten</li> <li>+ Icons</li> </ul>	<ul style="list-style-type: none"> <li>+ KEBA</li> <li>+ KI</li> <li>+ Sprache</li> </ul>	
	Informationen über das Umfeld	Status	<ul style="list-style-type: none"> <li>+ Hören</li> <li>+ Sprechen</li> <li>+ Leuchten</li> <li>+ Icons</li> </ul>	<ul style="list-style-type: none"> <li>+ Arbeitszeit</li> <li>+ Lagerbestand</li> <li>+ Positionierung im Raum</li> </ul>	
	Informationen über den Nutzer	Zugriffsrechte	<ul style="list-style-type: none"> <li>+ Gastzugriff</li> <li>+ Vollzugriff</li> <li>+ Gesperrter Zugriff</li> <li>+ Eingeschränkte Bedienung</li> </ul>	<ul style="list-style-type: none"> <li>+ Nutzererkennung</li> <li>+ Freigaben des Nutzers</li> </ul>	

Abb. 89: Wissensbereich von Kiki

# **VI. Voice Arena**



# 1. Voice Arena

Der finale Entwurf trägt den Namen Voice Arena. Der Name bezieht sich auf die Visualisierung der Sprache im äußeren Ring. Der aufgrund seiner leichten nach innen geneigten Form Assoziationen zu einem Stadion oder einer Arena weckt.

Das entstandene Interface zeichnet sich durch eine klare Formensprache aus. Der flache Körper deutet an, dass das Interface in die Maschine integriert ist und dennoch ein eigenständiges Element ist. Der Rahmen ist leicht nach innen geneigt und wirkt dadurch fest in der Maschinenhülle verankert. Die Kontur des Interfaces bildet einen harmonischen Mittelweg zwischen den kreisrunden Smart Speakern und den rechtwinkligen, stationären HMI von KEBA mit großzügigen Radien. Der robuste, jedoch zurückhaltende Rahmen umschließt die nach innen geneigten Leuchtfelder, die gut aus verschiedenen Blickwinkeln sichtbar sind.

Unter der transluzenten Oberfläche liegt ein zweiter Diffusor, der das Licht der einzelnen LED-Punkte zusätzlich streut. Durch die optimierte Anordnung der LEDs auf der Platine ist es möglich, Bewegungen von innen nach außen gleichmäßig darzustellen.

Im Zentrum befindet sich eine mit Stoff überzogene Fläche, die sich leicht nach oben hebt. Unter dem textilen Gewebe sitzt eine 8x8 LED-Matrix, auf der verschiedene Symbole dargestellt werden können. Das Display ist unter dem Stoff verborgen, damit es im ausgeschalteten Zustand nicht erkennbar ist. Ebenfalls verdeckt sind vier Mikrofone, die für eine präzise Spracherkennung sorgen.

Die Farb- und Materialauswahl orientiert sich an den anderen KEBA-Produkten und erfüllt somit Industriestandards.

Aufgrund akustischer Notwendigkeiten ist der Lautsprecher ausgelagert. Das Gehäuse des Lautsprechers ähnelt dem des Interfaces, es ist jedoch etwas kleiner und ragt nicht so weit aus der Maschinenhülle hervor.

# 2. Bespielung

Das Lichtkonzept verfolgt das Ziel, den Nutzer bestmöglich zu unterstützen, ohne dabei abzulenken. Die Visualisierung ist plakativ und ausdrucksstark, sodass sie auch aus dem Augenwinkel gut erkennbar ist. Die Modi unterscheiden sich durch Farbmuster, Richtung und Geschwindigkeit, was dabei hilft, sie intuitiv zu erkennen. Um die Kompetenz und Intelligenz der KI-Steuerung bzw. des Interfaces zu verdeutlichen, sind die Übergänge stets ruckelfrei und fließend gestaltet.

Damit sich der Benutzer wahrgenommen fühlt, werden die Lichteffekte gezielt auf ihn ausgerichtet. Die Grundlage hierfür bildet die Sprecherlokalisierung durch die Mikrofone.

In bestimmten Situationen wird die Aussage von Kiki durch Symbole unterstützt. Dies geschieht beispielsweise nach erfolgreichem Abschluss einer Aufgabe oder falls es Nachrichten für den Benutzer gibt, die nicht sofort gehört werden müssen. In solchen Fällen leuchtet ein einfaches Icon in der Mitte.

### 3. Positionierung

Das Interface wird stets sichtbar und sofern möglich in Augenhöhe positioniert. Es ist möglich, das Interface sowohl vertikal als auch horizontal einzusetzen. Aufgrund des externen Lautsprechers ergeben sich einige Möglichkeiten die Interfaceteile an einer Maschine zu positionieren. Hier am Beispiel einer Spritzgussmaschine. Die Entscheidung den Lautsprecher auszulagern, wurde erst vor kurzem getroffen. Daher können sich noch Änderungen bei der Gestaltung ergeben.



**Abb. 90:** Positionierung Voice Arena und Speaker Arena

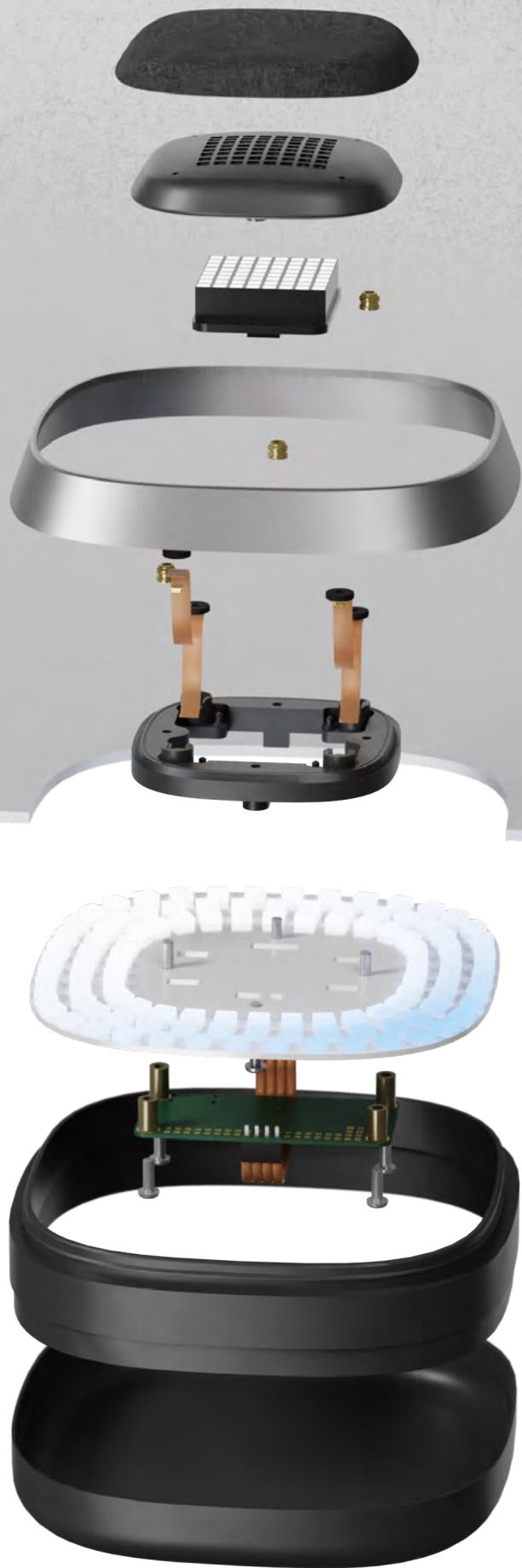


Abb. 91: Explosionsdarstellung Voice Arena



**Abb. 92:** Explosions-  
darstellung  
Speaker Arena



Abb. 93: Rendering Voice Arena und Speaker Arena



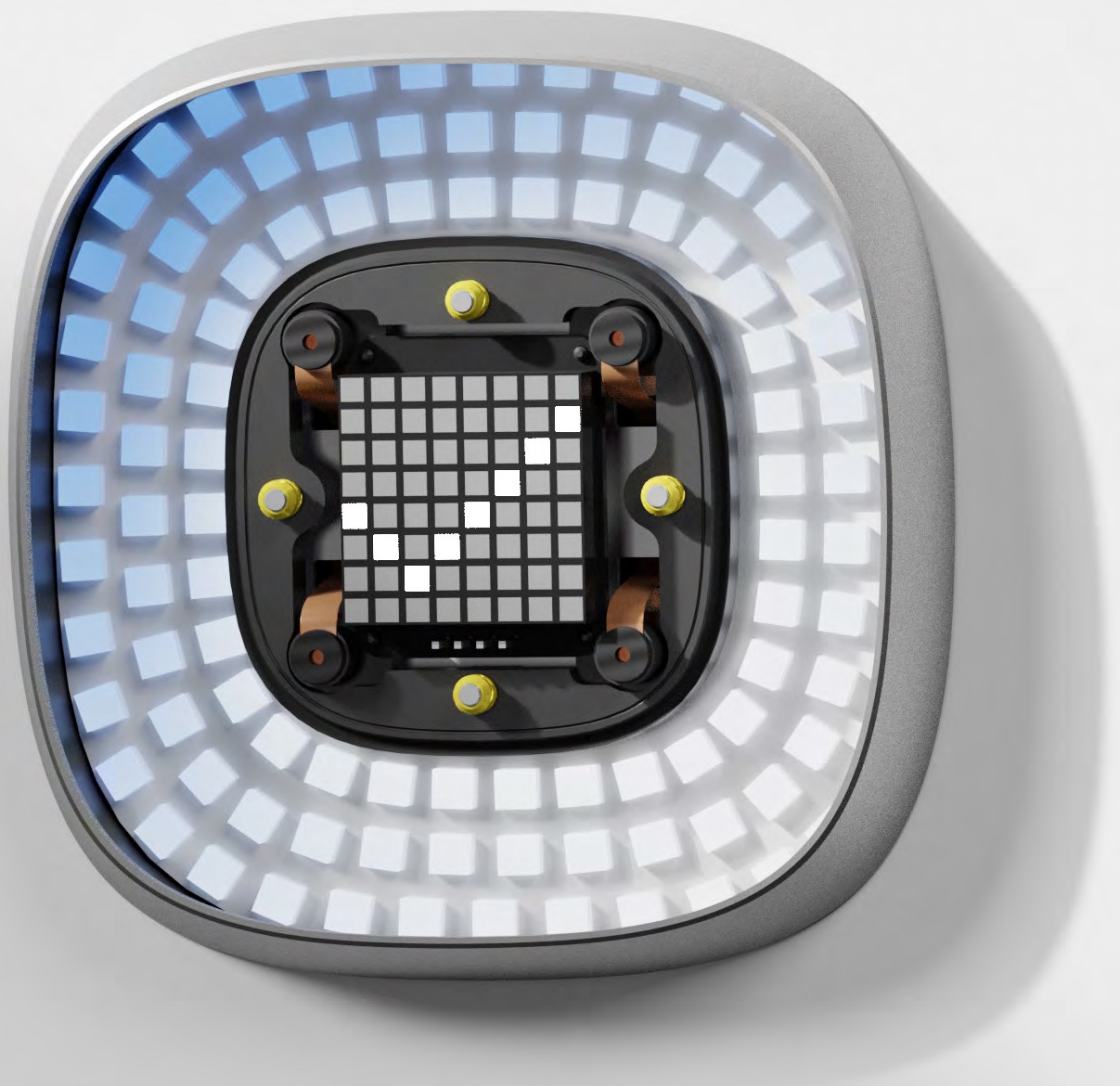
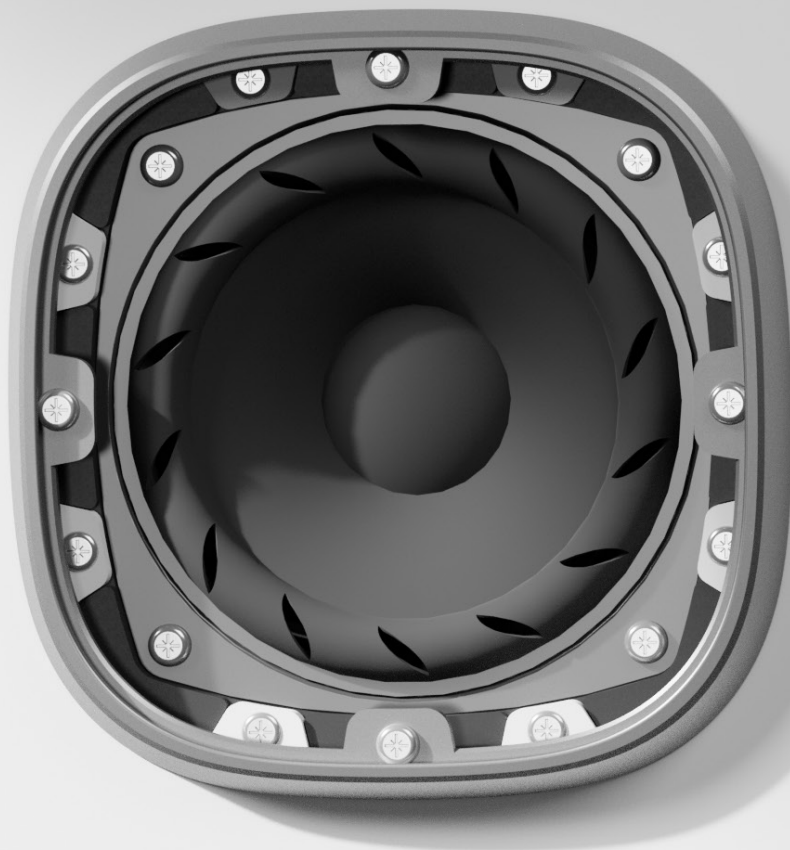


Abb. 94: Voice Arena und Speaker Arena ohne Cover



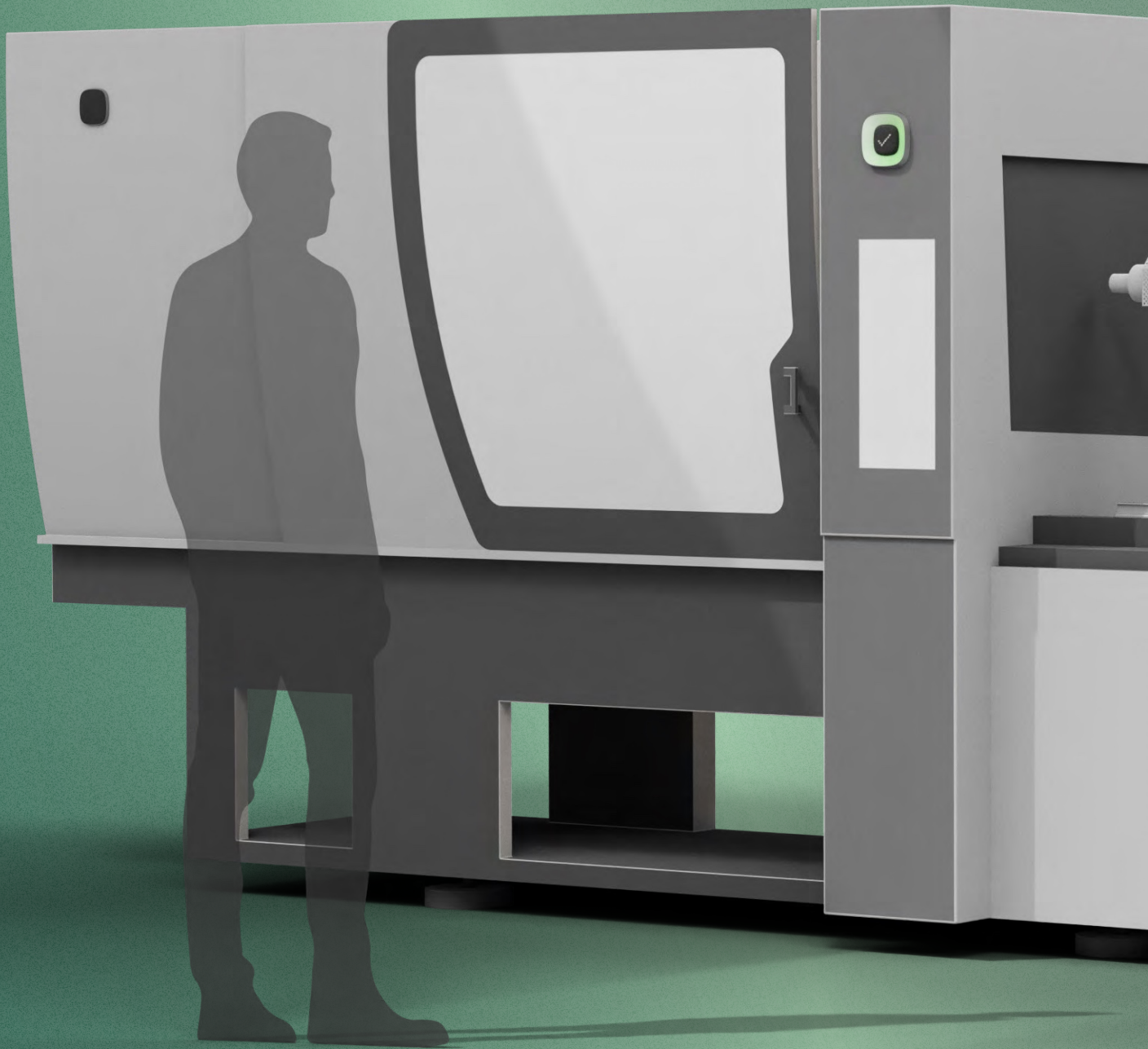


Abb. 95: Anwendungsbeispiel Voice Arena und Speaker Arena an Spritzgussmaschine



# VII. Ausblick



Die folgenden Funktionen sind losgelöst von den aktuellen Vorgaben und geben eine Vorstellung, wie sich die Interaktion in Zukunft verändern kann.

## **1. Nutzeridentifikation**

Wenn es möglich ist, den Benutzer anhand der Stimme zu identifizieren, könnte Kiki ein individuelles Profil anlegen. Darin werden verschiedene Daten gespeichert, um die persönliche Nutzererfahrung zu verbessern, beispielsweise ein personalisierter Sprachstil, der an die individuellen Vorlieben des Benutzers angepasst ist. Durch die Stimmidentifikation können verschiedene Berechtigungen für unterschiedliche Mitarbeiter vergeben werden.

## **2. Kontextbezug der KI**

In alltäglichen Unterhaltungen werden viele kontextbezogene Begriffe verwendet. Pronomen wie „Es“ und „Das“ kommen im normalen Sprachgebrauch oft vor. Für Maschinen ist es jedoch schwierig, diese Bezeichnungen zu verstehen, da sie nicht auf einen bestimmten Begriff oder Gegenstand weisen. Gleiches gilt für Ausdrücke, die sich auf vergangene Ereignisse beziehen, wie beispielsweise „wie beim letzten Mal“ oder ähnliche Formulierungen.

## **3. Kombination mit anderen Eingabemethoden**

Die Kombination mit anderen Eingabemethoden kann die Interaktion verbessern. Kameras mit Gesterkennung können der KI bei der Erfassung des Kontextes helfen. Anstatt einen Gegenstand genau benennen zu müssen, reicht es aus, mit dem Finger auf das entsprechende Objekt zu zeigen. Durch das Tracking mit einer Kamera über einen längeren Zeitraum hinweg können in Zukunft Bedienfehler oder Gefahrenpotenziale frühzeitig erkannt werden. Dies erhöht die Sicherheit von Mitarbeitern in Umgebungen mit erhöhten Gefahrenpotenzial.



# VIII. Verzeichnisse



# 1. Quellenverzeichnis

- 2001: Odyssee im Weltraum. 1968.** [Film] Regie: Stanley Kubrick. USA: Metro-Goldwyn-Mayer.
- Amazon Alexa, 2023.** Alexa Smile Logo Guidelines. [Online]  
Available at: <https://developer.amazon.com/en-US/alexa/branding/alexa-guidelines/brand-guidelines/the-alexa-logo>  
[Zugriff am 01 10 2023].
- Amazon, 2023.** All-New Echo Dot. [Online]  
Available at: <https://www.amazon.com/All-New-release-clock-speaker-Glacier/dp/B09B8VN8YQ>  
[Zugriff am 20 03 2023].
- Apple, 2023.** Apple stellt den neuen HomePod vor, bahnbrechend bei Sound und Intelligenz. [Online]  
Available at: <https://www.apple.com/at/newsroom/2023/01/apple-introduces-the-new-homepod-with-break-through-sound-and-intelligence/>  
[Zugriff am 02 03 2023].
- Apple, 2023.** HomePod – Benutzerhandbuch. [Online]  
Available at: <https://support.apple.com/de-at/guide/homepod/welcome/homepod>  
[Zugriff am 02 10 2023].
- Apple, 2023.** HomePod mini. [Online]  
Available at: <https://www.apple.com/at/homepod-mini/>  
[Zugriff am 01 10 2023].
- Apple, 2023.** Siri. [Online]  
Available at: <https://www.apple.com/de/siri/>  
[Zugriff am 01 10 2023].
- Arif, A. & Stuerzlinger, W., 2009.** Analysis of Text Entry Performance Metrics. TIC-STH'09: 2009 IEEE Toronto International Conference - Science and Technology for Humanity, 10, p. 102.
- Bennett, S., 2016.** Accidentally Famous: The Story Behind the Original Voice of Siri | Susan Bennett | TEDxFurmanU, Greenville, South Carolina: s.n.
- Bernhard, T., 2015.** Cortana vs. Siri #1: Emotionen. [Online]  
Available at: <https://www.genieblog.ch/blog/en/2015/cortana-vs-siri-1-emotionen/>  
[Zugriff am 16 07 2023].
- Bernstein, H., 2019.** Elektroakustik; Mikrofone, Klangstufen, Verstärker, Filterschaltungen und Lautsprecher. München, Deutschland: Springer Vieweg.
- Blade Runner. 1982.** [Film] Regie: Ridley Scott. USA: Warner Bros..
- BMW Group PressClub, 2018.** „Hey BMW, jetzt reden wir.“ Mit BMWs Intelligent Personal Assistant bekommt der BMW eine Persönlichkeit.. [Online]  
Available at: <https://www.press.bmwgroup.com/deutschland/article/detail/T0284429DE/%E2%80%9Ehey-bmw-jetzt-reden-wir-%E2%80%9C-mit-bmws-intelligent-personal-assistant-bekommt-der-bmw-eine-persoenlichkeit>  
[Zugriff am 06 03 2023].
- Bohn, D., 2019.** The Verge. [Online]  
Available at: <https://www.theverge.com/2019/1/4/18168565/amazon-alexa-devices-how-many-sold-number-100-million-dave-limp>  
[Zugriff am 28 02 2023].
- Brownlee, M., 2022.** The Voice Assistant Battle! (2023). [Online]  
Available at: <https://www.youtube.com/watch?v=Q2MGqmuEdtU>  
[Zugriff am 01 03 2023].
- Budiu, R. & Laubheimer, P., 2018.** Intelligent Assistants Have Poor Usability: A User Study of Alexa, Google Assistant, and Siri. [Online]

Available at: [https://www-nngroup-com.translate.goog/articles/intelligent-assistant-usability/?lm=mental-model-ai-assistants&pt=article&\\_x\\_tr\\_sl=auto&\\_x\\_tr\\_tl=de&\\_x\\_tr\\_hl=de](https://www-nngroup-com.translate.goog/articles/intelligent-assistant-usability/?lm=mental-model-ai-assistants&pt=article&_x_tr_sl=auto&_x_tr_tl=de&_x_tr_hl=de)  
[Zugriff am 03 03 2023].

**Budiu, R. & Whitenon, K., 2018.** The Paradox of Intelligent Assistants: Poor Usability, High Adoption. [Online]  
Available at: <https://www.nngroup.com/articles/intelligent-assistants-poor-usability-high-adoption/>  
[Zugriff am 16 07 2023].

**Budiu, R. & Whitenon, K., 2018.** What Could an Intelligent Assistant Do for You? A Diary Study of User Needs. [Online]  
Available at: [https://www-nngroup-com.translate.goog/articles/intelligent-assistant-user-needs/?lm=mental-model-ai-assistants&pt=article&\\_x\\_tr\\_sl=auto&\\_x\\_tr\\_tl=de&\\_x\\_tr\\_hl=de](https://www-nngroup-com.translate.goog/articles/intelligent-assistant-user-needs/?lm=mental-model-ai-assistants&pt=article&_x_tr_sl=auto&_x_tr_tl=de&_x_tr_hl=de)  
[Zugriff am 03 03 2023].

**Card, S. K., 1983.** The Psychology of Human-Computer Interaction. New Jersey: Lawrence Erlbaum Associates Inc., Publishers.

**Carroll, R., 2015.** Goodbye privacy, hello ‚Alexa‘: Amazon Echo, the home robot who hears it all. [Online]  
Available at: <https://www.theguardian.com/technology/2015/nov/21/amazon-echo-alexa-home-robot-privacy-cloud>

**Case, A., 2016.** Calm Technology - Principles and Patterns for Non-Intrusive Design. 1005 Gravenstein, Highway North, Sebastopol, CA 95472.: O’Reilly Media, Inc.,.

**charlies-names, 2023.** Name Kiki. [Online]  
Available at: <https://charlies-names.com/de/kiki/>  
[Zugriff am 16 07 2023].

**Collings, I., 2020.** What is Beamforming?. [Online]  
Available at: <https://www.youtube.com/watch?v=A1n5Hhwtz78&t=360s>  
[Zugriff am 08 10 2023].

**Creoir, 2021.** Microphone Design Guidelines. s.l.:Firmeninternes Dokument von Creoir.

**Creoir, 2023.** Acoustic Design Proposal - Speaker Concept. s.l.:Internes Dokument unter Geheimhaltung.

**Creoir, 2023.** creoir.com. [Online]  
Available at: <https://www.creoir.com/>  
[Zugriff am 24 09 2023].

**Davies, C., 2017.** Jibo Review: Alexa gets some cute competition. [Online]  
Available at: <https://www.slashgear.com/jibo-review-2017-13507668/>  
[Zugriff am 06 07 2023].

**Detweiler, M., 2014.** Roboter Jibo will Familienmitglied werden. [Online]  
Available at: <https://www.faz.net/aktuell/technik-motor/technik/crowdfunding-projekt-roboter-jibo-will-familienmitglied-werden-13050912.html>  
[Zugriff am 06 07 2023].

**Dixon, T., 2021.** HomePod Mini: Round, Full of Sound, and Now Torn Down. [Online]  
Available at: <https://de.ifixit.com/News/49645/homepod-mini-round-full-of-sound-and-now-torn-down>  
[Zugriff am 07 10 2023].

**Dorau, R., 2011.** Emotionales Interaktionsdesign: Gesten und Mimik interaktiver Systeme. Berlin Heidelberg: Springer-Verlag.

**Eilert, D. W., 2022.** Eilert Akademie. [Online]  
Available at: <https://eilert-akademie.com/blog/sprechgeschwindigkeit/>  
[Zugriff am 23 02 2023].

**Emotech, 2017.** Olly - The First Home Robot with Personality. [Online]  
Available at: <https://www.indiegogo.com/projects/olly-the-first-home-robot-with-personality#/>  
[Zugriff am 16 07 2023].

- emotech, 2018.** Indiegogo Production Progress Update #2 - Smart TV. [Online]  
Available at: [https://www.youtube.com/watch?v=n\\_nT5c\\_hrkg](https://www.youtube.com/watch?v=n_nT5c_hrkg)  
[Zugriff am 06 03 2023].
- Ex Machina. 2014.** [Film] Regie: Alex Garland. Großbritannien: A24; Universal Pictures International .
- fandom.com, 2015.** Clippy. [Online]  
Available at: <https://the-scrappy.fandom.com/wiki/Clippy>  
[Zugriff am 01 10 2023].
- Fraunhofer Institut, 2023.** Intuitive Maschinensteuerung durch Spracherkennung. [Online]  
Available at: [https://www.idmt.fraunhofer.de/de/Press\\_and\\_Media/press\\_releases/2023/maschinensteuerung-durch-spracherkennung.html](https://www.idmt.fraunhofer.de/de/Press_and_Media/press_releases/2023/maschinensteuerung-durch-spracherkennung.html)  
[Zugriff am 08 07 2023].
- genderlessvoice, 2019.** Meet Q. [Online]  
Available at: <https://www.genderlessvoice.com/>  
[Zugriff am 16 07 2023].
- Google, 2023.** Google Brand Resources. [Online]  
Available at: <https://about.google/brand-resource-center/logos-list/>  
[Zugriff am 01 10 2023].
- Google, 2023.** Nest Mini. [Online]  
Available at: [https://store.google.com/product/google\\_nest\\_mini?hl=de](https://store.google.com/product/google_nest_mini?hl=de)  
[Zugriff am 20 03 2023].
- Hempel, J., 2015.** Siri and Cortana Sound Like Ladies Because of Sexism. [Online]  
Available at: <https://www.wired.com/2015/10/why-siri-cortana-voice-interfaces-sound-female-sexism/>  
[Zugriff am 07 03 2023].
- her. 2013.** [Film] Regie: Spike Jonze. USA: Annapurna Pictures.
- Hermann, I., 2020.** Künstliche Intelligenz in der Science-Fiction: Zwischen Magie und Technik. FIF-Kommunikation, Issue 04, pp. 13-14.
- Hermes, A. K., 2023.** trend.at. [Online]  
Available at: <https://www.trend.at/branchen/forschung-innovation/kuenstliche-intelligenz>  
[Zugriff am 23 02 2023].
- IBM Archive, 1961.** IBM MediaCenter. [Online]  
Available at: [https://mediacenter.ibm.com/media/\(1961\)+Shoebox+-+IBM+Archives+\(78-013\)/0\\_4m2ynnkk](https://mediacenter.ibm.com/media/(1961)+Shoebox+-+IBM+Archives+(78-013)/0_4m2ynnkk)  
[Zugriff am 23 02 2023].
- IBM, 2011.** IBM.com. [Online]  
Available at: <https://www.ibm.com/ibm/history/ibm100/us/en/icons/speechreco/transform/>  
[Zugriff am 29 09 2023].
- IBM, 2023.** IBM Shoebox. [Online]  
Available at: <https://www.ibm.com/ibm/history/exhibits/specialprod1/images/overlay/1404011.jpg>
- imdb, 2023.** Knight Rider. [Online]  
Available at: [https://www.imdb.com/title/tt0083437/?ref\\_=tt\\_ov\\_inf](https://www.imdb.com/title/tt0083437/?ref_=tt_ov_inf)  
[Zugriff am 05 07 2023].
- Iron Man. 2008.** [Film] Regie: Jon Favreau. USA: Marvel Studios.
- Kabel, P., 2020.** Dialog zwischen Mensch und Maschine: Conversational User Interfaces, intelligente Assistenten und Voice-Systeme. Hamburg: Springer-Verlag.
- Kacher, G., 2017.** Was ist der Unterschied zwischen Array-Mikrofonen und Beamforming? [Interview] (8 11 2017).
- KEBA, 2022.** 10 Gründe, warum industrielle HMI Geräte Smart Devices überlegen sind. [Online]  
Available at: <https://www.keba.com/de/news/industrial-automation/industrielle-hmi-versus-consumer-devices>

[Zugriff am 05 07 2023].

**KEBA, 2023.** AI-Modul, Produktname: CI 550. Linz: Firmeninternes Dokument.

**KEBA, 2023.** Mobile HMI: Die mobile Bedienung für Ihre Bedürfnisse. [Online]

Available at: <https://www.keba.com/de/industrial-automation/products/hmi/mobile/ueberblick>

**KEBA, 2023.** Stationäre KeTop-Produkte im Überblick. [Online]

Available at: <https://www.keba.com/de/industrial-automation/products/hmi/stationary/ueberblick>

[Zugriff am 05 07 2023].

**Kim, E., 2022.** Amazon is gutting its voice assistant, Alexa. Employees describe a division in crisis and huge losses on ,a wasted opportunity.'. [Online]

Available at: <https://www.businessinsider.com/amazon-alexa-job-layoffs-rise-and-fall-2022-11>

[Zugriff am 01 03 2023].

**Kincaid, J., 2018.** Medium. [Online]

Available at: <https://medium.com/descript/a-brief-history-of-asr-automatic-speech-recognition-b8f338d4c0e5>

[Zugriff am 23 02 2023].

**Kinsella, B., 2018.** The Voice of HAL from “2001: A Space Odyssey,” Passed Away at Age 90, Influenced How Voice Assistants Sound Today. [Online]

Available at: <https://voicebot.ai/2018/11/12/the-voice-of-hal-from-2001-a-space-odyssey-passed-away-at-age-90/>

[Zugriff am 08 03 2023].

**Kinsella, B., 2019.** Amazon Alexa Skill Counts Rise Rapidly in the U.S., U.K., Germany, France, Japan, Canada, and Australia. [Online]

Available at: <https://voicebot.ai/2019/01/02/amazon-alexa-skill-counts-rise-rapidly-in-the-u-s-u-k-germany-france-japan-canada-and-australia/>

[Zugriff am 01 03 2023].

**Kunchay, S. & Abdullah, S., 2021.** Assessing Effectiveness and Interpretability of Light Behaviors in Smart Speakers. Bilbao (online), New York, NY, United States, Association for Computing Machinery, pp. 1-14.

**Lang, U., 2022.** Maschinenbedienung: Mobile HMI im Vergleich zu stationären Bediengeräten. [Online]

Available at: <https://www.keba.com/de/news/industrial-automation/mobile-hmi-vs-stationaere-maschinenbedienung>

[Zugriff am 05 07 2023].

**Laubheimer, P. & Budi, R., 2018.** Intelligent Assistants: Creepy, Childish, or a Tool? Users' Attitudes Toward Alexa, Google Assistant, and Siri. [Online]

Available at: [https://www-nngroup-com.translate.goog/articles/voice-assistant-attitudes/?\\_x\\_tr\\_sl=auto&\\_x\\_tr\\_tl=de&\\_x\\_tr\\_hl=de](https://www-nngroup-com.translate.goog/articles/voice-assistant-attitudes/?_x_tr_sl=auto&_x_tr_tl=de&_x_tr_hl=de)

[Zugriff am 03 03 2023].

**Lohr, S., 2010.** The New York Times. [Online]

Available at: <https://www.nytimes.com/2010/09/24/business/24jelinek.html>

[Zugriff am 23 02 2023].

**Lowerre, B. T., 1976.** The HARPY Speech Recognition System. Pittsburgh, Pennsylvania: s.n.

**Mariani, J., Rosset, S., Garnier-Rizet, M. & Devillers, L., 2014.** Natural Interaction with Robots, Knowbots and Smartphones: Putting Spoken Dialog Systems into Practice. In: Spoken Language Understanding for Natural Interaction: The Siri Experience. New York, Heidelberg, Dordrecht, London: Springer Science & Business Media, pp. 3-14.

**marvelcinematicuniverse, 2023.** J.A.R.V.I.S.. [Online]

Available at: <https://marvelcinematicuniverse.fandom.com/wiki/J.A.R.V.I.S.>

[Zugriff am 05 07 2023].

**Matschnig, M., 2012.** 30 Minuten Körpersprache verstehen. Offenbach: Gabal Verlag GmbH.

**Mercedes Benz, 2023.** MBUX: Die Mercedes-Benz User Experience. [Online]

Available at: <https://www.mercedes-benz.at/passengercars/mercedes-benz-cars/mbux/personalization.module.html>

[Zugriff am 06 03 2003].

**Mori, M., 1970.** The Uncanny Valley: The Original Essay by Masahiro Mori. [Online]

Available at: <https://spectrum.ieee.org/the-uncanny-valley>

[Zugriff am 16 07 2023].

**Nielsen Norman Group, 2019.** Mental Models for Intelligent Assistants. [Online]

Available at: <https://www.nngroup.com/articles/mental-model-ai-assistants/>

[Zugriff am 06 07 2023].

**NIO, 2020.** NOMI Mate 1.5. [Online]

Available at: <https://www.red-dot.org/de/project/nomi-mate-15-46042>

[Zugriff am 16 07 2023].

**Noessel, C., 2015.** scifiinterfaces.com. [Online]

Available at: <https://scifiinterfaces.com/2015/07/13/iron-man-hud-just-the-functions/>

[Zugriff am 28 09 2023].

**Norman, D., 1998.** The Invisible Computer. Cambridge, Massachusetts: The MIT Press.

**Pearl, C., 2016.** Designing Voice User Interfaces. Sebastopol, CA 95472: O'Reilly Media, Inc..

**Pieraccini, R., 2021.** AI Assistants. Cambridge, Massachusetts: MIT Press.

**Reddy, R., 2013.** youtube.com. [Online]

Available at: <https://www.youtube.com/watch?v=NiiDe2n-GeQ&t=970s>

[Zugriff am 30 09 2023].

**Reddy, R., 2018.** Raj Reddy, 1994 ACM Turing Award Recipient [Interview] (26 02 2018).

**Sharp, H., Rogers, Y. & Preece, J., 2019.** Interaction Design: beyond human-computer interaction, Fifth Edition. Indianapolis, Indiana: John Wiley & Sons, Inc..

**Shedroff, N. & Noessel, C., 2012.** Make it so: Interaction Design Lessons from Science Fiction. Brooklyn, New York: Rosenfeld Media, LLC.

**Statistik Austria, 2022.** IKT-Einsatz in Unternehmen 2021, Wien: Verlag Österreich GmbH.

**Statistik Austria, 2022.** STRASSENVERKEHRSUNFÄLLE mit Personenschaden 2021. Wien: Verlag Österreich GmbH.

**Strenger, Y. & Kennedy, J., 2020.** The Smart Wife: Why Siri, Alexa, and Other Smart Home Devices Need a Feminist Reboot. Cambridge, Massachusetts: The MIT Press.

**Terminator. 1984.** [Film] Regie: James Cameron. USA: Hemdale; Pacific Western Productions.

**The Matrix. 1999.** [Film] Regie: Wachowskis. USA: Warner Bros..

**Turkle, S., 2017.** Why these friendly robots can't be good friends to our kids. [Online]

Available at: [https://www.washingtonpost.com/outlook/why-these-friendly-robots-cant-be-good-friends-to-our-kids/2017/12/07/bce1eaea-d54f-11e7-b62d-d9345ced896d\\_story.html](https://www.washingtonpost.com/outlook/why-these-friendly-robots-cant-be-good-friends-to-our-kids/2017/12/07/bce1eaea-d54f-11e7-b62d-d9345ced896d_story.html)

[Zugriff am 07 07 2023].

**Vitale, S., 2021.** auto-motor-und-sport.d. [Online]

Available at: <https://www.auto-motor-und-sport.de/oldtimer/hasselhoff-versteigert-knight-rider-auto/>

[Zugriff am 28 09 2023].

**Volkswagen Communications, 2021.** Leistungsfähiger und schneller.. [Online]

Available at: <https://www.volkswagen-newsroom.com/de/pressemitteilungen/leistungsaehiger-und-schneller-volkswagen-hebt-die-sprachbedienung-im-golf-auf-neues-niveau-7698>

[Zugriff am 06 03 2023].

**Wahler, H., 2012.** 93% of All Communication Is Nonverbal? Correcting a Common Misconception of the Mehrabi-

an Studies. Mainz: Farid Darwish und Hendrik Wahler mindyourlife GbR.

**Warren, T., 2023.** Apple is changing the 'Hey Siri' trigger phrase to just 'Siri' in iOS 17. [Online]  
Available at: <https://www.theverge.com/2023/6/5/23749156/apple-wwdc-hey-siri-wake-word-voice-assistant>  
[Zugriff am 16 07 2023].

**Weber, K., 2008.** Roboter und Künstliche Intelligenz in Science Fiction-Filmen: Vom Werkzeug zum Akteur. Technik und Gesellschaft in der Science Fiction, pp. 8-9.

**Weiser, M. & Brown, J. S., 1995.** Designing Calm Technology. Palo Alto: Xerox PARC.

**West, M., Kraut, R. & Chew Han Ei, 2019.** I'd blush if I could: closing gender divides in digital skills through education. Frankreich: EQUALS und UNESCO.

**Westworld. 2016.** [Film] Regie: J. J. Abrams. USA: Bad Robot Productions; Warner Bros. Television,.

**wikipedia, 2023.** HAL 9000. [Online]  
Available at: [https://de.wikipedia.org/wiki/HAL\\_9000](https://de.wikipedia.org/wiki/HAL_9000)  
[Zugriff am 16 07 2023].

**winfuture.de, 2022.** Echo Dot 5 im Test: So gut ist Amazons neuer Alexa-Lautsprecher. [Online]  
Available at: <https://winfuture.de/news,132839.html>  
[Zugriff am 16 07 2023].

**WIPO, 2019.** WIPO Technology Trends 2019 - Artificial Intelligence. Genf: World Intellectual Property Organization.

**XU, X., 2019.** Nomi Halo - ES8. [Online]  
Available at: <https://xiumingxu.studio/nomi-halo/>  
[Zugriff am 20 03 2023].

**Zhang, L. & Ng, M., 2011.** An Acoustics-Based Assessment of Error Patterns in English Sounds Produced by Cantonese ESL Speakers. Asia Pacific Journal of Speech.

## 2. Abbildungsverzeichnis

- Abb. 01:** Kommunikationsmodell System-Interface-Mensch (Dorau, 2011, S. 221)
- Abb. 02:** Stationäre KeTop Geräte im Überblick: KeTop AP 500, KeTop OP 300 Keyboard Panels, KeTop OP Singletouch, KeTop OP 500 Multitouch (KEBA, 2023)
- Abb. 03:** Mobile KeTop Geräte im Überblick: KeTop T10 directMove, KeTop T20 eco/technic, KeTop T55, KeTop T70, KeTop T150, KeTop T200, KeTop T150/155 Safe Wireless) (KEBA, 2023)
- Abb. 04:** HAL 9000 aus 2001: Odyssee im Weltraum (1968) (wikipedia, 2023)
- Abb. 05:** KITT-Interface und Scanner Licht (Vitale, 2021)
- Abb. 06:** Jarvis als Head-up-Display im Iron Man Anzug (Noessel, 2015)
- Abb. 07:** Science-Fiction Interface sind vorwiegend blau (1991 war Terminator 2) (Shedroff & Noessel, 2012, p. 41) Anmerkung des Autors: Die Art der Darstellung wurde verändert, die Farben wurden beibehalten.
- Abb. 08:** Chronik der künstlichen Intelligenz und Spracherkennung in Filmen im Vergleich zur realen Welt
- Abb. 09:** IBM Shoebox 1962 (IBM, 2023)
- Abb. 10:** Der Recognition-Tree von Harpy (Reddy, 2013)
- Abb. 11:** Fred Jelinek (links im Bild) mit dem Talking Typewriter von IBM ca. 1980 (IBM, 2011)
- Abb. 12:** Screenshot aus Dragon NaturallySpeaking (Zhang & Ng, , 2011)
- Abb. 13:** Übersicht des NLP (Kabel, 2020, p. 8)
- Abb. 14:** KI-Modul CI 550 mit verschiedenen Schnittstellen (KEBA, 2023)
- Abb. 15:** Übersicht aller Komponenten, ob physisch oder digital, eines virtuellen Assistenten (vgl. Pieraccini, 2021, p. 8).
- Abb. 16:** Siri Logo (Apple, 2023)
- Abb. 17:** Alexa Logo (Amazon Alexa, 2023)
- Abb. 18:** Google Assistent Logo (Google, 2023)
- Abb. 19:** Clippy (fandom.com, 2015)
- Abb. 20:** Cortanas Emotionen (Bernhard, 2015)
- Abb. 21:** Ergebnis Tagebuchstudie Nutzeranfragen (Budiu & Whitenton, 2018)
- Abb. 22:** Großes Potenzial bei der Nutzung von intelligenten Assistenten (Budiu & Whitenton, 2018) Layout und Farben wurden zur besseren Verständlichkeit verändert
- Abb. 23:** Beliebteste Aufgaben für Sprachassistenten (Budiu & Whitenton, 2018)
- Abb. 24:** Komplexität der Interaktion (Budiu & Whitenton, 2018)
- Abb. 25:** Amazon Echo 1. Generation (Carroll, 2015)
- Abb. 26:** Amazon Echo 5. Generation (Amazon, 2023)
- Abb. 28:** Apple HomePod mini (links) und der HomePod (rechts) (Apple, 2023)
- Abb. 27:** Google Nest Mini (Google, 2023)
- Abb. 30:** Olly (Emotech, 2017)
- Abb. 29:** Jibo (Detweiler, 2014)
- Abb. 31:** Übersicht Smart Speaker
- Abb. 32:** Lichtring und 7-Segment-Anzeige auf dem Amazon Echo Dot 5.Gen (Amazon, 2023)
- Abb. 33:** 4 LED auf dem Google Nest Mini (Google, 2023)
- Abb. 34:** Lichtfläche auf dem Apple HomePod Mini (Apple, 2023)

- Abb. 35:** Jibo zeigt Symbole aus dem Menü (Davies, 2017)
- Abb. 36:** bunter Lichtring des Olly (Emotech, 2017)
- Abb. 37:** Übersicht Modi Lichtsignale von Smart Speakern
- Abb. 38:** Bewertung der Wirksamkeit und Interpretierbarkeit von Lichtreaktionen bei Smart Speakern (Kunchay & Abdullah, 2021, p. 8)
- Abb. 39:** Nomi Halo (XU, 2019)
- Abb. 41:** Gegenüberstellung Nomi Halo und Mate( XU, 2019)
- Abb. 40:** Nomi Mate (NIO, 2020)
- Abb. 42:** Elf Grundemotionen (Matschnig, 2012)
- Abb. 43:** Uncanny Valley (Mori, 1970)
- Abb. 44:** Q - die erste genderlose Stimme (genderlessvoice, 2019)
- Abb. 45:** Gemeinsam erarbeitete Post-it Wand aus dem ersten Workshop am 26.06.2022
- Abb. 46:** Moodboard für Formen, Oberflächen und Lichtakzente
- Abb. 47:** Moodboard für Formen, Oberflächen und Lichtakzente
- Abb. 48:** Übersicht aller Sketche aus dem zweiten Workshop vom 4.8.2022
- Abb. 49:** Fazit aus dem zweiten Workshop vom 4.8.2022
- Abb. 50:** Unterscheidungsmerkmale des Interface im Vergleich zu anderen Smart Speakern
- Abb. 51:** Entwurfssketch mit einer flachen Einbausituation
- Abb. 52:** 3D-Entwürfe mit einer flachen Einbausituation
- Abb. 53:** Entwurfskonzept KITT am Beispiel einer Spritzgussmaschine
- Abb. 54:** Entwurfssketch die das KEBA Logo in den Mittelpunkt stellen
- Abb. 55:** 3D-Entwürfe mit dem KEBA Logo
- Abb. 56:** Entwurfssketch mit beweglichen Teilen
- Abb. 57:** 3D-Entwürfe mit beweglichen Teilen
- Abb. 58:** Entwurfssketch mit schmalen Öffnungen in der Oberfläche
- Abb. 59:** Entwurfssketch von zusätzlichen Features
- Abb. 60:** Keysketch der Voice Arena
- Abb. 61:** Grasshopper Node-Tree für die Erstellung der Matrix
- Abb. 63:** Vier Favoriten der Matrix
- Abb. 62:** Matrix mit unterschiedlichen Größenverhältnissen der Teilbereiche
- Abb. 64:** Darstellungen des Interface mit 10cm Breite (links) und 20cm (rechts)
- Abb. 65:** Technische Analyse von Amazon Echo 2.Generation (links) und des Echo der 5.Genreation (rechts)
- Abb. 66:** Technische Analyse des Apple HomePod mini (links) und es HomePod 2.Generation (rechts) (Dixon, 2021)
- Abb. 67:** Technische Analyse des Google Nest Mini
- Abb. 68:** Erster Entwurf eines technischen Konzepts
- Abb. 69:** Vorgaben für die Mikrofonöffnung in der Außenhülle (Creoir, 2021)
- Abb. 70:** Übersicht Funktionsweise Beamforming (Collings, 2020)
- Abb. 71:** Richtcharakteristik einer Mikrofongruppe (Bernstein, 2019, p. 186)
- Abb. 72:** Vorgaben für die Anordnung der Mikrofone (Creoir, 2021)
- Abb. 73:** Vorschlag für den Lautsprecher (Creoir, 2023)

- Abb. 74:** Akustik Model mit einem optimalen Lautsprechers in 3m Entfernung (Creoir, 2023)
- Abb. 75:** Übersicht verschiedene LED-Anordnungen
- Abb. 76:** Übersicht verschiedener Konzepte für die Anordnung des Lautsprechers und der Mikrofone
- Abb. 77:** Ideation für die Bespielung der Leuchtflächen
- Abb. 78:** Übersicht verschiedener Konzepte für die Bespielung der Leuchtflächen
- Abb. 79:** Fotos und Screenshots des ersten MockUp
- Abb. 80:** Fotos und Screenshots des zweiten MockUp
- Abb. 81:** Fotos und Screenshots des dritten MockUp
- Abb. 82:** Varianten der Aufgabenverteilung
- Abb. 83:** Übersicht der Animationsmodi
- Abb. 84:** Beschreibung der Übergänge von einem Modus zum Nächsten
- Abb. 86:** Übersicht der Icons die im Zentrum dargestellt werden
- Abb. 85:** Darstellung mit Icon auf Display im Zentrum
- Abb. 87:** Interaktion Journey „Sortieren“
- Abb. 88:** Interaktion Journey „Neues Objekt erlernen“
- Abb. 89:** Wissensbereich von Kiki
- Abb. 90:** Positionierung Voice Arena und Speaker Arena
- Abb. 91:** Explosionsdarstellung Voice Arena
- Abb. 92:** Explosionsdarstellung Speaker Arena
- Abb. 93:** Rendering Voice Arena und Speaker Arena
- Abb. 94:** Voice Arena und Speaker Arena ohne Cover
- Abb. 95:** Anwendungsbeispiel Voice Arena und Speaker Arena an Spritzgussmaschine

### 3. Danksagung

An dieser Stelle möchte mich bei allen Bedanken die mich auf dem Weg zu meiner Masterarbeit und bei ihrer Fertigstellung unterstützt haben.

An erster Stelle bei Univ.-Prof. Mag. Art Elke Bachlmair und Univ.-Prof. Mag. Art Mario Zeppetzauer für ihre Betreuung, wertvollen Ratschläge Geduld im Laufe des Projekts.

Ein besonderer Dank gilt den Initiatoren des Projekts bei der Firma KEBA, Thomas Linde, Julian Bauer und Christoph Etzlinger, bedanken, für die Möglichkeit, an diesem Projekt zu arbeiten, sowie für ihre Unterstützung und Ressourcen.

Dagmar Hutter für die Korrektur meiner Arbeit und so erheblich dazu beigetragen hat, die Qualität und Verständlichkeit zu verbessern. Ebenfalls an Lumetzberger Monika für die Korrektur der Rechtschreibung. Ein besonderes Dankeschön geht an Simon Hutter für die Hilfe beim Bau des MockUp's und die vielen Stunden Programmierarbeit. Zu guter Letzt möchte ich mich bei meinen Eltern bedanken, die mir die Möglichkeit gegeben haben, dieses Studium zu wählen und erfolgreich abzuschließen.

Nochmals vielen Dank an alle für Ihre Hilfe und Unterstützung.



CC BY-NC-ND 4.0 International  
Namensnennung - Nicht-kommerziell - Keine Bearbeitung 4.0 International