

Spracherkennung im Field Service

Carolin Edler-Mende

Wie sehen klassische Aufgabenstellungen im Field Service aus?

- **Terminvorbereitung**
 - **Protokollieren eines Termins / Besuchsbericht**
 - **Auslösung von Folgehandlungen, wie bspw. Bestellungen**
 - **Ausfüllen von Formularen (Neukundendaten, Angebotserstellungen)**
 - **Ableitung von To-Do's und Dokumentation**
- x **All diese Tätigkeiten müssen **neben Termin und Fahrtzeit** umgesetzt werden, z.B. Direkt vor Ort mühsam auf Mobile Device oder später am stationären Endgerät.**

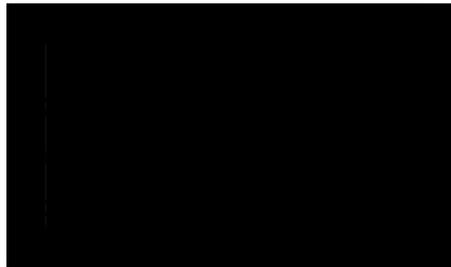
Wie können diese Tätigkeiten durch Sprachtechnologie optimiert werden?

- **Terminvorbereitung**
 - Vorlesen des letzten Berichts während Anfahrt durch TTS-Stimme
 - **Protokollieren des Termins**
 - Protokoll per Spracherkennung erfassen, im System ablegen oder als Email versenden
 - **Folgehandlungen auslösen**
 - Per Spracherkennung/Sprachdialog mit Backendanbindung erfassen und auslösen
 - **Ausfüllen von Formularen**
 - Per Sprachdialog Felder abfragen und ausfüllen
 - **Ableitung von To-Do's und Dokumentation**
 - Per Spracherkennung erfassen
- ✓ Umsetzung **während der Fahrt ohne visuelle Interaktion** möglich.
- ✓ Keine mühselige Texteingabe auf Mobilgerät (Scrollen durch Produktnamen, Tippen, etc.)

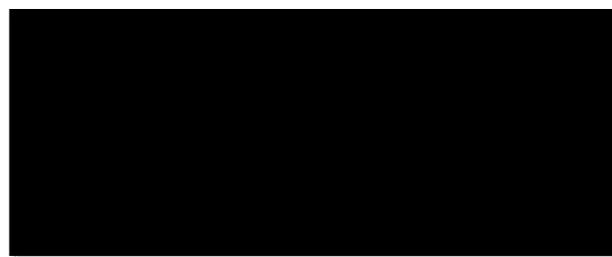
Wie funktionieren moderne Spracherkennung?

- Definition

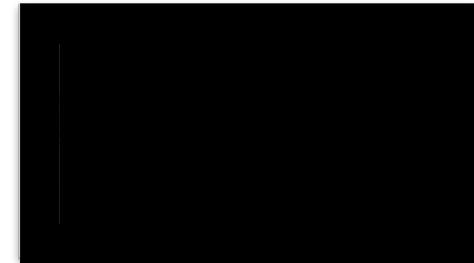
- **Software, die auf Basis von Text, Audio und Aussprachelexikon trainiert wird, um gesprochene Sprache in den entsprechenden Text zu transkribieren**



Aussprachelexikon



Language model



Acoustic model

Wie funktionieren moderne Spracherkennung?

- **Erkennungsmodi**
 - Freie Erkennung, Grammatikerkennung, Stichworterkennung
- **Betrieb**
 - Von Hochlastserver in der Cloud über on-premise-Lösung bis embedded Anwendung ist alles umsetzbar
- **Künstliche Intelligenz**
 - Verbesserung der Erkennung durch Machine Learning Prozesse
- **Qualität abhängig von**
 - Trainingsmethoden
 - Trainingsdaten
 - Audioqualität der Eingabe
 - Sprechweise/Sprecherverhalten

Achtung: Unterschied zwischen **Spracherkennung und **semantischer Auswertung****

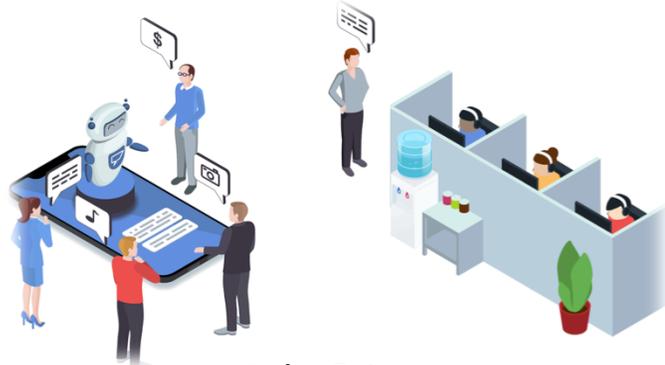
Wo werden Spracherkennung heute eingesetzt?



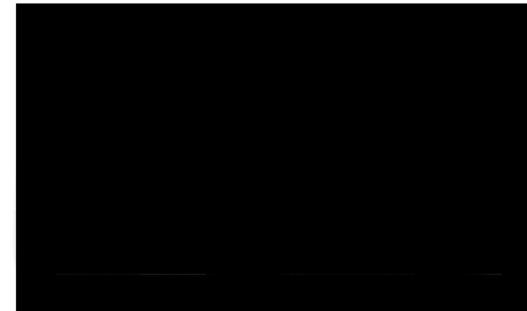
Smart devices



Diktiergeräte



Voice-Bots



Digitale Assistenten

Wie kann eine Integration aussehen?

- Cloud Lösung im **Rechenzentrum** (Achtung Datenschutz)
- Cloud Lösung **On-Premise**
- Umsetzung **On-Device** (Achtung evtl. Hardwarelimitierung -> Performanceverlust)

Fragestellungen:

- Welche Datenquellen können zur Erhöhung der Erkennungsqualität verwendet werden?
- Wie komplex ist die Backend-Anbindung zur direkten Übertragung erfasster Daten, bzw. Zum Auslösen eines Prozesses?
- Ist bereits ein System vorhanden, welches eine Integration ermöglicht? (z.B. Mobile App)
 - Voice-Enabling

Wo stößt die Technologie auf ihre Grenzen?

- **Kurze, phonetisch naheliegende Wörter**
 - Zwei oder Drei? F oder S?
- **Nicht oder schwer vorhersehbare Wortketten**
 - Namen, Adressen, Email-Adressen
- **Fachvokabular**
 - Produktbezeichnungen, Softwarenamen, Projektnamen, Anglizismen
- **Schlechte Bedingungen**
 - Während der Autofahrt durch Motorgeräusche, Widerhall im Auto
- **Unkooperative User**
 - „Ähh, ja... also A, ich meine A wie Anton, und dann F...“
- **Netzabdeckung bei Cloud-Lösung**

Wie kann man diese Grenzen überwinden?

- ✓ **Vorbereiten des Spracherkenners auf Anwendungskontext:**
 - Training des Language Models mit Fachvokabular, bzw. Fachsprache
 - Training des Akustischen Models mit Aufnahmen aus dem realen Anwendungsszenario, bspw. Im Auto
 - Erweitern des Aussprachelexikons mit Fachvokabular
- ✓ **Nutzen bestehender Datenquellen zur Identifikation der richtigen Erkennungshypothese über Nähenanalyse**
- ✓ **Embedded Betrieb ist möglich und garantiert Funktion auch in Gebieten mit schlechter Infrastruktur**
- ✓ **Im Sprachdialog: gezielte Formulierung der richtigen Fragestellung, um User zur gewünschten Sprechweise zu bewegen**

Wie sieht ein ideales Projektvorgehen aus?

1. Auftaktworkshop

- a) Betrachten der **Ausgangslage**
- b) Identifikation von Prozessen, die sich leicht Sprach-automatisieren lassen
- c) Identifikation von Prozessen, deren Automatisierung den meisten Nutzen bringt
- d) Ableiten sinnvoller Schritte und Priorisierung

2. Entwicklung

- a) Agile, anwendungsnahe Entwicklung
- b) Frühes Testing durch friendly Usergruppen
- c) Optimierungs- und Erweiterungsiterationen auf Basis der Erkenntnisse aus den Testrunden

3. Deployment

- a) Ausrollen auf alle Usergruppen

Welche Vorteile kann der Einsatz von Sprachtechnologie bringen?

- ✓ **Steigerung der Produktivität**
- ✓ **Mitarbeiterzufriedenheit**
- ✓ **Höhere Datengenauigkeit durch zeitliche Nähe des Erfassens zum Termin**
- ✓ **Business Case: Kostenersparnis durch mehr Termine pro Tag, weniger Nacharbeit**

- **Zeitersparnis durch effektive Nutzung von Downtime während der Reise**

Vielen Dank für Ihre
Aufmerksamkeit