

Nutzerspezifische Assistenz in manuellen Prozeduren mit Hilfe von Sprachinteraktion

Frieder Loch, M.Sc., Lehrstuhl für Automatisierung und Informationssysteme (AIS), Technische Universität München;
Dipl.-Inf. **Stefan Magerstedt**, KHS GmbH, Dortmund;
Prof. Dr.-Ing. **Birgit Vogel-Heuser**, Ordinaria, Lehrstuhl AIS, TUM

Kurzfassung

Maschinenbediener werden auch in Zukunft durch ihre Flexibilität und Kreativität eine wichtige Rolle in der Produktion einnehmen. Die Arbeitsbereiche von Maschinenbedienern liegen dabei neben der Maschinenbedienung im Durchführen von Wartungs- oder Umrüstprozeduren, dem Nachfüllen von Verbrauchsmaterialien, sowie dem Erkennen und Beheben von Fehlern. Diese Aufgaben, sowie die potentiell gefährliche Arbeitsumgebung, erfordern die kontinuierliche visuelle Aufmerksamkeit. Sprachbasierte Interaktion stellt somit einen vielversprechenden Ansatz dar, um Maschinenbediener zu unterstützen. Sie ermöglicht es ortsunabhängig auf Informationen zuzugreifen, ohne visuelle Aufmerksamkeit zu beanspruchen. Dieser Beitrag beschreibt ein sprachbasiertes System, das Maschinenbediener beim Durchführen von manuellen Wartungs- und Umrüsttätigkeiten unterstützt. Das System lässt sich einfach an konkrete Prozeduren und Benutzergruppen anpassen, um Benutzer mit unterschiedlichen Fähigkeiten, zum Beispiel in Bezug auf Wahrnehmung oder Berufserfahrung, gleichermaßen zu unterstützen. Die Anpassung des Systems wird über Modelle gesteuert, die das Eingabelexikon, den Benutzer und die aktuelle Prozedur beschreiben.

1. Einführung und Stand der Technik

Maschinenbediener bleiben ein wichtiger Bestandteil von heutigen und zukünftigen Produktionsumgebungen und benötigen intuitive Interaktionstechniken, um erfolgreich Ihre Aufgaben nachzukommen. Sie übernehmen dabei unterschiedliche, oft gleichzeitig auszuführende, Aufgaben. Diese Aufgaben umfassen unter anderem das Nachlegen von Verbrauchsmaterialien, das Prüfen von Qualitätsparametern sowie das Beheben von Fehlern. Zudem müssen Wartungs- und Umrüstprozeduren durchgeführt werden. Beim Bearbeiten dieser Aufgaben muss sich der Bediener kontinuierlich bewegen und den Zustand der Maschine und der Umgebung verfolgen. Konzepte die auf ortsfesten Bediengeräten beruhen, oder traditionelle Lösungen zur Assistenz, zum Beispiel papier-basierte Handbücher, stoßen in dieser Umgebung an ihre

Grenzen. Um Maschinenbediener bei ihren Aufgaben effizient zu unterstützen stellen sprachbasierte Benutzungsoberflächen eine vielversprechende und geeignete Lösung dar.

Die Möglichkeiten sprachbasierter Systeme wurden bereits in unterschiedlichen industriellen Anwendungsfällen erprobt. In der Kommissionierung werden Pick-by-Voice Systeme erfolgreich eingesetzt und führen zu einer Senkung von Kommissionierfehlern und einer Steigerung der Effizienz [1]. Zudem werden sprachbasierte Systeme für die Assistenz von Wartungsmitarbeitern [2] und die Interaktion mit Maschinen in der Forschung erprobt [3]. Existierende Lösungen zur sprachbasierten Interaktion sind aber wenig flexibel und lassen sich kaum, oder nur mit größerem Aufwand, an die Bedürfnisse eines konkreten Anwendungsszenarios oder die Fähigkeiten des Benutzers (z.B. in Hinblick auf Wissensstand oder das Vokabular) anpassen. Insbesondere die Anpassung an unterschiedliche Fähigkeiten wird im Zeitalter des demographischen Wandels und seiner Auswirkungen auf die Altersstruktur der Belegschaften an Bedeutung gewinnen.

Der vorliegende Beitrag beschreibt ein sprachbasiertes System, das Maschinenbediener bei der Durchführung von Wartungs- und Umrüsttätigkeiten unterstützt. Ein Fokus liegt auf der Anpassung dieses Systems an die Benutzer. So sollen individuelle Wissensstände und Ausdrucksweisen, sowie sensorische Einschränkungen kompensiert werden, um ein für viele Mitarbeiter gut nutzbares System zu schaffen.

2. Einsatz von Sprachinteraktion im industriellen Umfeld

Das folgende Kapitel beschreibt einen Anwendungsfall für ein sprachbasiertes Assistenzsystem. Hierfür werden zunächst die Eigenschaften beispielhafter Wartungsprozeduren, die mit einem sprachbasierten System adressiert werden sollen, beschrieben. Folgend wird der industrielle Nutzungskontext diskutiert, um Anforderungen an ein sprachbasiertes Assistenzsystem abzuleiten.

2.1 Anwendungsfall

Das Assistenzsystem soll in industriellen Wartungs- und Umrüstprozeduren zum Einsatz kommen. Solche Prozeduren werden durchgeführt, wenn die produzierte Sorte verändert werden soll oder wenn ein Fehler aufgetreten ist. Als beispielhafter Anwendungsfall wird im Folgenden eine Prozedur an einer Flaschenreinigungsmaschine der KHS GmbH beschrieben. Bei der Maschine handelt es sich um eine Maschine des Typs INNOCLEAN SE. Die Maschine wird genutzt, um Flaschen von Verunreinigungen und Etiketten zu befreien. Die Besonderheit dieser Anlage ist, dass die Flaschen auf der Seite ausgegeben werden, auf der sie auch in die

Maschine gelangt sind (SE = Single end, bzw. Einend). Bild 1 zeigt die im Anwendungsfall betrachtete Flaschenreinigungsmaschine.



Bild 1: Innoclean SE Reinigungsmaschine.

Die verunreinigten Flaschen werden von dem Transportband an die Flaschenaufnahme übergeben. Die Flaschen werden dann durch Trennwände auf Einzelspuren separiert. Von dort aus heben Finger die Flaschen in Transportkörbe, welche die Flaschen kettengeführt durch die Maschine befördern. Die Flaschen durchlaufen daraufhin mehrere Becken, in denen sie mit Lauge abgespült werden. Dabei lösen sich Etiketten und Verunreinigungen und sinken zum Grund der Becken. Abschließend werden die Flaschen mit Klarwasser von möglichen Lauge-Resten befreit und an die Flaschenabgabe ausgegeben.

Bei einer maximalen Geschwindigkeit von 170.000 Flaschen pro Stunde kann es vorkommen, dass eine Flasche zerbricht oder bei der Aufnahme verkeilt wird. Wenn dies geschieht, muss der Maschinenbediener die Maschine wieder frei fahren und die Flasche entfernen. Dazu muss der Bediener die Maschine zunächst in den Handbetrieb schalten. Im Handbetrieb können die einzelnen Antriebe durch ein Handgerät (siehe Bild 2) angesteuert werden.



Bild 2: Handgerät zur Ansteuerung der Antriebe.

Der zu verfahrenende Antrieb wird durch eine Zahl ausgewählt, die auf den Antrieben angebracht ist. Mit dem Handgerät kann der angewählte Antrieb dann frei bewegt werden. Ist der Fehler behoben, also die zerbrochene Flasche entfernt, müssen die Antriebe wieder möglichst nah an ihre Ausgangsposition gebracht werden, bevor die Maschine wieder in den normalen Betrieb übergehen kann. Dieser Schritt wird heutzutage per Augenmaß und daher nur grob durchgeführt. Die bisherige Ausgangsposition ist nicht anhand einer Skala oder Anzeige ablesbar. Insbesondere bei Schichtübergaben, längeren Ausfällen oder kognitiven Einschränkung des Bedieners kann die notwendige Position nicht mehr gezielt angefahren werden. Für eine möglichst schnelle Rückführung in den Normalbetrieb ist eine genauere Ausrichtung der Antriebe jedoch von Vorteil.

Zwar muss die Maschine aus Sicherheitsgründen weiterhin über die ursprüngliche MMS in den Handbetrieb geschaltet werden, die eigentliche Wartung wird dann aber mit der Sprachschnittstelle durchgeführt. Dem Bediener wird hierzu eine schrittweise Anleitung in Textform bereitgestellt. Der Benutzer kann mittels Sprachbefehlen („Nächster Schritt“, „Letzter Schritt“, „Wiederholen“) durch die Anleitung navigieren.

Bei der Antriebsrepositionierung wird der Anwender durch das Sprachinterface unterstützt, indem Anweisungen für den jeweiligen Antrieb (Nummer und Richtung) angegeben und vom Benutzer bestätigt werden. Durch die Handsteuerung werden die Vorgaben des Sprachinterfaces umgesetzt, hierbei kann der Bediener die jeweilige Situation permanent visuell begut-

achten und entsprechende Rückmeldungen („Antrieb bewegt sich nicht“) an das Sprachinterface übergeben. Typische Fehlerursachen („Lichtgitter unterbrochen“) werden so schnell und präzise an den Bediener übergeben und führen zu einer schnelleren Rückführung in den Normalbetrieb.

2.2 Analyse des industriellen Nutzungskontexts

Für die Entwicklung einer gebrauchstauglichen Benutzungsoberfläche ist die Analyse und das Einbeziehen des späteren Nutzungskontexts unerlässlich. Der Nutzungskontext beinhaltet die statischen und dynamischen Eigenschaften des Benutzers (z.B. Qualifikation, Alter, physische und kognitive Fähigkeiten oder die momentane Beanspruchung). Zudem werden die technischen (z.B. eingesetzte Geräte), physikalischen (z.B. Bedingungen am Arbeitsplatz, notwendige Schutzausrüstung) und organisatorischen (z.B. rechtliche Anforderungen) Rahmenbedingungen betrachtet [4]. So wird sichergestellt, dass das entwickelte System den Anforderungen der Anwendung entspricht. Die folgenden Abschnitte beschreiben die relevanten Eigenschaften des industriellen Nutzungskontexts.

Benutzereigenschaften. Wartungsprozesse werden typischerweise von angeleiteten Maschinenbedienern durchgeführt. Diese verfügen über fundierte Erfahrungen mit der Maschine, aber oft über keine technische Ausbildung. In diesem Berufsfeld werden Mitarbeiter mit unterschiedlicher Herkunft eingesetzt, so dass unterschiedliche Sprachkenntnisse vorhanden sind und unterstützt werden müssen. Maschinen werden zunehmend auch an global verteilten Standorten eingesetzt, so dass auch das geringere Bildungsniveau der Mitarbeiter in Entwicklungsländern von einem Assistenzsystem berücksichtigt werden muss.

Physikalische Rahmenbedingungen. Der Wartungsprozess wird in einer Umgebung durchgeführt, die den Einsatz anderer Assistenzsysteme nicht zulässt oder stark erschwert. Die vorliegenden Umwelteinflüsse, wie etwa Staub, Flüssigkeit, oder der beengte Raum, schließt den Einsatz anderer Systeme zur Assistenz (z.B. Handbücher, oder touch-basierte Geräte) aus, da solche Geräte verschmutzen oder zu einer Ablenkung von der Arbeitsaufgabe und der Beobachtung der industriellen Umgebung führen würde. Zudem können diese Assistenzsysteme nicht während der Arbeitsaufgabe eingesetzt werden, da sie die Unterbrechung der manuellen Tätigkeit verlangen. In einer solchen Arbeitsumgebung, ist die Darstellung von Arbeitsschritten und Daten in auditiver Form ratsam.

Zudem ist die Umgebung in der das System eingesetzt wird, insbesondere bei Glas-Anlagen und da gewöhnlich weitere Anlagen in der Umgebung laufen, sehr laut. Hierdurch ergeben

sich weitere Anforderungen. So sollte die ausgewählte Software eine hohe Erkennungsleistung aufweisen, die auch von starken Hintergrundgeräuschen nicht beeinflusst wird. Nur so kann gewährleistet werden, dass die Befehle unabhängig von Umgebungseinflüssen sicher erkannt werden können.

Organisatorische Rahmenbedingungen. Eine reine Sprachsteuerung ist aus Gründen der Arbeitssicherheit derzeit nicht möglich. So kann beim An- und Abschalten der Maschine mit einer Sprachsteuerung nicht garantiert werden, dass sich der Bediener in einem sicheren Abstand zur Maschine befindet. Dies ist nur mit einer ortsgebundenen Lösung möglich.

Die meisten der verbreiteten Spracherkennungen senden die aufgezeichneten Sprachdaten zur Erkennung an eigene Server. Die Daten werden dort auch nach Abschluss der Erkennung gespeichert, um sie für die Verbesserung der Spracherkennung zu nutzen. Diese Datenspeicherung bedeutet eine Sicherheitslücke, da Daten über Maschinen und Prozeduren in die Hände von Unbefugten gelangen. Somit können nur Systeme verwendet werden, die nicht auf eine Internetverbindung angewiesen sind oder bei denen unternehmenseigene Server genutzt werden können.

2.3 Anforderung an ein sprachbasiertes Assistenzsystem

Auf Basis der Beschreibung des Nutzungskontexts wurden vier Anforderungen für das sprachbasierte System formuliert. Die Anforderungen werden folgend beschreiben.

1. **Adaption.** Das System muss sich an Eigenschaften der adressierten Nutzergruppen anpassen lassen. So sollen unterschiedlichen Sprachkenntnisse und Bildungsniveaus unterstützt werden. Die Adaptation soll eine Flexibilität im Befehlssatz, der Komplexität der Anweisungen und der Art der Ausgabe ermöglichen.
2. **Ablenkungsarme Interaktion.** Um Störungen beim Durchführen manueller Arbeitsschritte auszuschließen, muss sich das System ausschließlich durch Sprachkommandos steuern lassen. Physische Eingaben, zum Beispiel auf einem Touchscreen, sind deshalb nicht vorgesehen.
3. **Robuste Sprachinteraktion.** Das System muss auch bei einer hohen Umgebungslautstärke Sprachkommandos robust erkennen können, um Probleme die aus Fehlerkennungen resultieren zu vermeiden. Die Sprachausgabe muss in einer industriellen Umgebung verständlich sein.

4. **Datensicherheit.** Die Spracherkennung muss als abgeschlossenes System auf dem Endgerät oder auf einem Server innerhalb des Unternehmens lauffähig sein. Dies vermeidet Risiken, die durch eine Internetverbindung des Endgeräts und die Übertragung der Sprachdateien auf den Server einer Drittfirma entstehen würden.

3. Entwicklung einer adaptiven Sprachschnittstelle

Die Adaptation an den Bediener und den Anwendungsfall ist eine wichtige Anforderung an das entwickelte System. Als Adaption wird eine automatisierte Anpassung eines interaktiven Systems verstanden [5]. So können die Bedienoberfläche oder die Funktionalität an den Bediener oder die Situation angepasst werden. Bild 3 zeigt die Komponenten des Systems und die Mechanismen zur Adaption. Die Komponenten werden im Folgenden beschrieben.

- Die *Spracherkennung* übersetzt die Eingabe des Benutzers in Text. Sie wird durch das Vokabular mit dem der Benutzer das Assistenzsystem bedienen kann definiert.
- Die *Interpretation* übersetzt die erkannten Worte in die Funktionen des Systems. Die Anpassung dieser Komponente ermöglicht ein flexibles Eingabevokabular.
- Die *Anweisungen* beschreiben die Art und den Detailgrad der Anweisungen. So lassen sich je nach Wissensgrad mehr oder weniger detaillierte Anweisungen ausgeben. Die Anweisungen werden in einem XML-basierten Format spezifiziert.
- Die *Sprachausgabe* verbalisiert die Anweisungen. Ihre Funktion lässt sich auf Basis der Parameter der Sprachsynthese adaptieren (z.B. Lautstärke oder Geschwindigkeit der Sprachausgabe) um Einschränkungen der Hörfähigkeiten zu kompensieren.

Die Adaption des Systems wird durch drei Komponenten beschreiben. Dies sind das Vokabular, das Prozessmodell und das Benutzermodell.

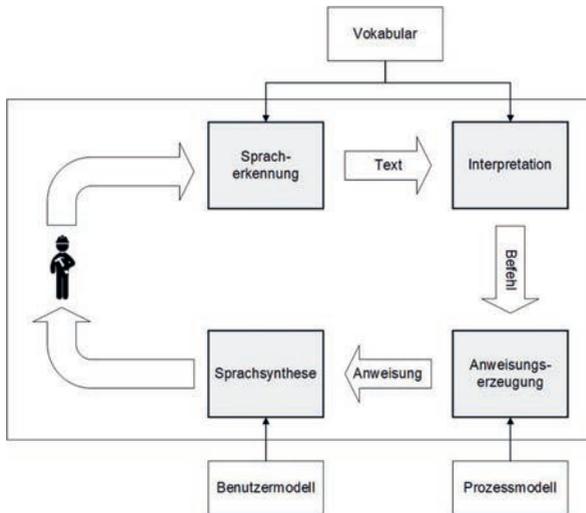


Bild 3: Architektur des sprachbasierten Assistenzsystems.

3.1 Vokabular

Das *Vokabular* beschreibt die Eingabephrasen und nimmt deren Zuordnung zu den Funktionen der Anwendung vor. Diese Komponente besteht zunächst aus einem vordefinierten Befehlssatz, der die Funktionalitäten der Anwendung beschreibt (z.B. "Nächster Schritt", "Vorheriger Schritt"). Jedem dieser Befehle können eine oder mehrere der Phrasen die vom System erkannt werden zugewiesen werden. So kann das System an das Vokabular der Benutzer, zum Beispiel in Hinblick auf regionale Dialekte, angepasst werden.

3.2 Prozessmodell

Das *Prozessmodell* beschreibt die Prozedur durch die das Assistenzsystem führt. Es ist hierarchisch angelegt, um eine Segmentierung in zusammenhängende Gruppen zu ermöglichen und unterschiedlich detaillierte Anweisungen erzeugen zu können. Ein Schritt (z.B. "Komponente A entfernen") kann so zur Unterstützung eines weniger erfahrenen Benutzers durch Teilschritte ausgedrückt werden (z.B. "Schraubverbindungen von Komponente A lösen", "Schrauben entfernen und Komponente A entnehmen").

3.3 Benutzermodell

Das *Benutzermodell* beschreibt die relevanten Eigenschaften des Benutzers. Es beschreibt Faktoren, wie die individuelle Konstitution, die Qualifikation und Kompetenz, sowie die momentane Beanspruchung. Die Konstitution betrifft Attribute wie Geschlecht oder Kultur. Die Disposition ist variabel und beschreibt Faktoren wie das Alter. Der momentane Zustand des Benutzers hängt von der Situation ab [6]. Die im Benutzermodell beschriebenen Eigenschaften ermöglichen es Regeln für die Sprachausgabe zu formulieren. So kann zum Beispiel die Lautstärke oder Wiedergabegeschwindigkeit angepasst werden oder die Komplexität der Anweisungen abhängig von der Erfahrung der Bediener variiert werden.

4. Technische Umsetzung

Das beschriebene System wurde auf eine Weise umgesetzt, die die Verwendung des Ansatzes in industriellen Szenarien erleichtert. Aus diesem Grund wurden weitestgehend freie Softwarekomponenten eingesetzt, die auf unterschiedlichen technologischen Plattformen lauffähig sind. Als Betriebssystem wurde Windows ausgewählt. Diese Entscheidung wurde auf Grund der großen Verbreitung dieses Systems in industriellen Umgebungen, sowie der Unterstützung einer Vielzahl von Geräteklassen (z.B. Desktop-, Mobil- oder Einplatinencomputer) getroffen. Spracherkennung und Sprachsynthese wurden mit dem C# .NET-Framework implementiert. Die Modelle, mit denen das Vokabular, die Benutzereigenschaften und die Prozeduren beschrieben werden, basieren auf XML. Ein Editor für die Erstellung und Pflege der Prozeduren und Modelldateien wurde entwickelt, um die Anpassung des Systems an verschiedene Anwendungsfälle zu erleichtern.

5. Zusammenfassung und Ausblick

Im vorliegenden Beitrag wurde ein System zur adaptiven Assistenz von Maschinenbedienern beschrieben. Ausgehend von einer Beschreibung eines Anwendungsfalls wurde die Verwendung eines sprachbasierten Assistenzsystems motiviert und Anforderungen abgeleitet. Auf dieser Basis wurde ein System beschrieben, das sich an konkrete Anwendungsfälle und Benutzergruppen anpassen lässt. Ein solches Assistenzsystem erlaubt es den Anforderungen heutiger und zukünftiger industrieller Umgebungen zu entsprechen und die Effizienz und Zufriedenheit unterschiedlicher Nutzergruppen in industriellen Umgebungen zu erhöhen.

Danksagung

Die beschriebenen Ergebnisse wurden im Projekt "Smart and adaptive interfaces for INCLUSIVE work environment" erarbeitet, das im Rahmen des Grant Agreements Nr. 723373 von der Europäischen Union im Programm "Horizon 2020" gefördert wird.

Literaturangaben

- [1] Lolling, A.: Fehlerraten verschiedener Kommissionersysteme in der Praxis. Zeitschrift für Arbeitswissenschaft 56 (2002), S. 112–115
- [2] Fischer, J., Pantförder, D. u. Vogel-Heuser, B.: Improvement of maintenance through speech interaction in cyber-physical production systems. 2017 IEEE 15th International Conference on Industrial Informatics (INDIN). IEEE 2017
- [3] Overmeyer, L., Podszus, F. u. Dohrmann, L.: Multimodal speech and gesture control of AGVs, including EEG-based measurements of cognitive workload. CIRP Annals - Manufacturing Technology 65 (2016) 1, S. 425–428
- [4] Maguire, M.: Context of Use within usability activities. International Journal of Human-Computer Studies 55 (2001) 4, S. 453–483
- [5] Dix, A., Finlay, J., Abowd, G. D. u. Beale, R.: Human-computer interaction. Pearson/Prentice Hall 2004
- [6] Luczak, H.: Wesen menschlicher Leistung. In: Arbeitsgestaltung in Produktion und Verwaltung. Köln: Wirtschaftsverlag Bachem 1989, S. 39–64