

Eggert, Mathias; Kreuzer, Vincent

Article — Published Version

Sprachsteuerung im Gesundheitswesen – Anforderungen und Auswahl geeigneter Anbieter

HMD Praxis der Wirtschaftsinformatik

Provided in Cooperation with:

Springer Nature

Suggested Citation: Eggert, Mathias; Kreuzer, Vincent (2022) : Sprachsteuerung im Gesundheitswesen – Anforderungen und Auswahl geeigneter Anbieter, HMD Praxis der Wirtschaftsinformatik, ISSN 2198-2775, Springer Fachmedien Wiesbaden GmbH, Wiesbaden, Vol. 59, Iss. 6, pp. 1640-1652,
<https://doi.org/10.1365/s40702-022-00919-z>

This Version is available at:

<https://hdl.handle.net/10419/307625>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



<https://creativecommons.org/licenses/by/4.0/>



Sprachsteuerung im Gesundheitswesen – Anforderungen und Auswahl geeigneter Anbieter

Mathias Eggert  · Vincent Kreuzer

Eingegangen: 6. Juli 2022 / Angenommen: 8. Oktober 2022 / Online publiziert: 27. Oktober 2022
© Der/die Autor(en) 2022

Zusammenfassung Das Gesundheitswesen ist konfrontiert mit steigenden Kosten und einer immer schwieriger werdenden Personalsituation. Zeitgleich versprechen moderne Sprachsteuerungssysteme Prozesse in Arztpraxen und Krankenhäusern zu verschlanken und Vorgänge zu beschleunigen. Dennoch wird derzeit der Einsatz von Sprachsteuerungssystemen in Arztpraxen oder Krankenhäusern nur selten beobachtet, was auch an den besonders strengen Datenschutzauflagen der Datenschutzgrundverordnung (DSGVO) liegt. Darüber hinaus wirft die niedrige Nutzungsrate die Frage nach den konkreten Anforderungen und ihrer Umsetzbarkeit auf, was durch den vorliegenden Beitrag adressiert wird, indem die Ergebnisse von Interviews mit acht medizinischen Fachexperten ausgewertet werden. Ergänzend wird die technische Umsetzbarkeit einzelner Anforderungen mit unterschiedlichen Cloud-Anbietern erprobt.

Schlüsselwörter Gesundheitswesen · Sprachsteuerung · Sprachassistent · Anforderungen · Umsetzbarkeit

✉ Mathias Eggert · Vincent Kreuzer
Fachhochschule Aachen, Eupener Str. 70, 52066 Aachen, Deutschland
E-Mail: eggert@fh-aachen.de

Vincent Kreuzer
E-Mail: vincent.kreuzer@alumni.fh-aachen.de

Voice Control in Healthcare—Requirements and Provider Selection

Abstract The healthcare sector is confronted with rising costs and an increasingly difficult personnel situation. At the same time, modern intelligent voice agents promise to streamline processes in medical practices and hospitals and to speed up activities. Nevertheless, the use of intelligent voice agents in medical practices or hospitals is currently only rarely observed, which is also due to the particularly strict data protection requirements of the General Data Protection Regulation (GDPR). Furthermore, the low usage rate raises the question of concrete requirements and their feasibility, which is addressed by this paper by evaluating the results of interviews with eight medical experts. In addition, the technical feasibility of individual requirements is tested with different cloud providers.

Keywords Health sector · Voice control · Intelligent voice agent · Requirements · Feasibility

1 Einleitung

Im Zeitalter der Digitalisierung wird der Einsatz von Spracherkennung immer relevanter und populärer. Seit 2016 sind weltweit mehr als eine halbe Milliarde Sprachsteuerungssysteme wie bspw. Alexa oder Siri verkauft worden (Statista 2022). Moderne Spracherkennungssysteme ermöglichen die fast nahtlose Erkennung des gesprochenen Wortes und sind in der Lage auch den Kontext einer Aussage richtig zu erfassen. Mittlerweile können Sprachanweisungen auch ohne ein langes Trainieren hinreichend gut verstanden und ausgeführt werden. Während in der Logistik die Steuerung über Sprache bereits Einzug gehalten hat (Vollmert et al. 2022), ist im Gesundheitswesen die Anwendung noch verhalten. Dabei können Sprachassistenten auch einen Beitrag leisten, um Klinik- und Praxisprozesse zu optimieren und gesundheitliche Abläufe zu verschlanken. So kann etwa durch einen Übersetzungsservice die Sprachbarriere bei der Erstaufnahme eines ausländischen Patienten reduziert und der Aufwand zur ärztlichen Dokumentation durch Spracheingaben verringert werden. Dennoch wird derzeit der Einsatz von Sprachsteuerungssystemen in Arztpraxen oder Krankenhäusern nur selten beobachtet, was auch an den besonders strengen Datenschutzauflagen der Datenschutzgrundverordnung (DSGVO) liegt. Darüber hinaus wirft die niedrige Nutzungsrate die Frage nach den konkreten Anforderungen und ihrer Umsetzbarkeit auf, was durch den vorliegenden Beitrag adressiert wird.

Aktuell fokussiert die Forschung im Bereich der Wirtschaftsinformatik die Entwicklung und Evaluation von Sprachsteuerungsprototypen für spezifische Probleme (Maas et al. 2020; Nitta und Murayama 2019; Sandeep Purao and Yanling Xu 2018). So wurde bspw. ein Spracherkennungssystem für die Automatisierung des medizinischen Berichtswesens entwickelt (Maas et al. 2020). Mit Ausnahme von Eggert und Stanke (2020), die Anforderungen im Kontext von Pflegeheimen erhoben und diese prototypisch implementiert haben, wurden Anforderungen an den Einsatz eines Sprachassistenten im medizinischen Umfeld bislang nicht systematisch erfasst. Darüber hinaus haben bisherige Arbeiten nicht die Besonderheiten des

deutschen Gesundheitswesens wie etwa die hohen Datenschutzanforderungen untersucht. Folglich ist es aktuell nur schwer möglich, die Entwicklung und den Einsatz von Sprachassistenten im Gesundheitswesen ganzheitlich zu betrachten.

Der vorliegende Beitrag hat das Ziel, systematisch die Anforderungen an den Einsatz von Sprachassistenten im Gesundheitswesen zu untersuchen. Dabei werden sowohl funktionale als auch nicht-funktionale Anforderungen betrachtet. Um dieses Ziel zu erreichen, wurden Interviews mit acht medizinischen Fachexperten aus Deutschland durchgeführt und systematisch mittels qualitativer Inhaltsanalyse ausgewertet. Aufbauend auf diesen Anforderungen wird ein Vergleich von Sprachsteuerungs-Anbietern durchgeführt und eine Anbieterempfehlung abgeleitet.

2 Sprachassistenten im Gesundheitswesen

Ein intelligenter Sprachassistent ist ein sprachbasierter persönlicher Agent mit künstlicher Intelligenz (KI), die so programmiert ist, dass sie Aufgaben automatisiert ausführt wie ein Mensch. Das Ziel ist dabei die Herstellung einer natürlichen Dialogsituation (Otoo und Salam 2018). Ein Sprachassistent stellt kontextbezogene Informationen bereit, die durch sprachliche Abfrage des Nutzers generiert werden (Guy 2016). Die Dialogsteuerung von Sprachassistenten erfolgt über ein Spracherkennungsmodul, welches die automatische Erkennung der Sprachinhalte ermöglicht (Tsuji et al. 2013). Dies erlaubt eine direkte Mensch-Maschinen-Kommunikation und damit die direkte Ausführung von Aktionen durch einen Befehl, der dem Sprachassistenten via Sprache übergeben wird (Helmke et al. 2015). Der Aufgabenumfang gängiger Sprachassistenten ist nahezu unbegrenzt und reicht von rein administrativen Aufgaben, wie beispielsweise das Ausfüllen einer elektronischen Patientenakte mithilfe von Sprachbefehlen, bis zur Informationsbeschaffung, wie die Suche nach Nebenwirkungen von Medikamenten. Administrative Aufgaben erfordern die Integration von bestehenden Systemen wie der elektronischen Patientenakte oder einen digitalen Kalender (Jiang et al. 2015). Informative Aufgaben beinhalten Websuchen sowie die Suche in Datenbanken (Kepuska und Bohouta 2018). Für Konsumenten werden etwa Sprachassistenten von Apple (Siri), Amazon (Alexa), Samsung (S Voice) und Google angeboten (Kepuska und Bohouta 2018).

Nach Jiang et al. (2015) weisen Sprachassistenten drei signifikante Funktionen auf: *Gerätedialoge* ermöglichen die Übermittlung und Erkennung von Sprachbefehlen und damit den Aufbau einer Interaktion mit dem Sprachassistenten. So kann beispielsweise ein Dialog über die Bestellung des Abendessens im Pflegeheim geführt werden. Die *Web-Suche* umfasst die sprachgestützte Internetsuche, welche automatisch nicht zu verarbeitende Sprachbefehle in eine Internetsuchanfrage umwandelt und entsprechend weiterleitet. Mit Hilfe dieser Funktion kann zum Beispiel eine Wetter- oder Pollenflugauskunft eingeholt werden. Die *Chat-Funktion* ermöglicht es Nutzern, eine Konversation mit dem Sprachassistenten aufzubauen, was etwa zur Bearbeitung von medizinischen Fragebögen genutzt werden kann. Sprachassistenten verfügen über vordefinierte Interaktionsmodelle, welche durch den Dienstanbieter erstellt werden (Jiang et al. 2015). Technologisch arbeiten alle Ansätze mit Deep-

Learn-Verfahren (vertiefend: Schmidhuber 2015), um die gesprochene Sprache zu erkennen (Kepuska und Bohouta 2018).

Das Gesundheitswesen als Anwendungsfeld für Sprachassistenten ist nicht neu. So wurde bereits ein erster Prototyp zur Erforschung von Sprachassistenten in der Altenpflege entwickelt und evaluiert (Eggert und Stanke 2020). Adelmeyer et al. (2019) verglichen Cloud-Dienste mit privaten Rechenzentren im Hinblick auf den Umgang mit Gesundheitsdaten und arbeiteten heraus, dass Cloud-Dienste für die Verarbeitung von Gesundheitsdaten sorgfältig ausgewählt werden müssen. Maas et al. (2020) entwickelten eine Software, die einen Sprachassistenten nutzt, um medizinische Dokumentationen zu erfassen. Des Weiteren wurde ein Prototyp zur sprachgesteuerten Erkennung und Behandlung von emotionsgestörten Menschen vorgestellt (Catania et al. 2019).

Aktuelle Arbeiten zur Anwendung von Sprachassistenten im Gesundheitswesen legen ihren Fokus verstärkt auf die Entwicklung von Prototypen für einen spezifischen Anwendungskontext und weniger auf generelle Anforderungen. Dieser Beitrag adressiert diese Lücke indem generische Anforderungen durch Interviews mit medizinischem Fachpersonal abgeleitet werden. Die Vorgehensweise dazu wird im folgenden Kapitel erläutert.

3 Methodisches Vorgehen

Die Vorgehensweise wurde in vier Phasen gegliedert. In der Phase der *Interviewvorbereitung* wurde der Leitfaden zu den semi-strukturiert durchgeführten Interviews entwickelt und die Zielgruppe definiert. Dabei nahmen primär Ärzte, Pfleger und Therapeuten, die aktiv im Gesundheitswesen tätig sind und mindestens drei Jahre Berufserfahrung aufweisen, an der Befragung teil, um eine ausreichend große Fachexpertise sicherzustellen. Um die Befragten demographisch einordnen zu können, wurden zu Beginn des Interviews Fragen zum demographischen Hintergrund der Interviewteilnehmer sowie zu deren Erfahrung im Umgang mit Sprachassistenten gestellt. Im Anschluss folgten offene Fragen zu funktionalen und nicht-funktionalen Anforderungen, deren Verständlichkeit und Nachvollziehbarkeit in einem Pretest mit einem Arzt getestet wurden:

- Was sind Ihre Erfahrungen im Umgang mit Spracherkennung?
- Wie oft nutzen Sie ein Spracherkennungssystem in der Woche?
- Wie ist Ihre Einstellung zum Datenschutz im Kontext Gesundheitsanwendungen?
- Gibt es Bereiche im Gesundheitswesen, in denen Sie es als sinnvoll empfinden, eine Spracherkennung einzusetzen, um bspw. Zeit zu sparen?
- Wie könnte Ihrer Meinung nach eine Spracherkennung einen medizinischen Prozess (Dokumentierung, Kommunikation) effizienter gestalten?
- Haben Sie ähnliche direkte Vorschläge für den Einsatz von Sprachassistenten im Gesundheitswesen?

In der Phase *Interviewdurchführung* wurden insgesamt acht Fachexperten aus dem Gesundheitswesen befragt (Tab. 1). Bei allen acht handelt es sich um Berufstätige des deutschen Gesundheitswesens. Jede Person wurde einzeln und getrennt interviewt.

Tab. 1 Demographie der Interviewteilnehmer

Interview ID	Geschlecht	Alter	Berufliche Tätigkeit	Arbeitsplatz	Erfahrungen mit Spracherkennungssysteme	Nutzungsfrequenz
I1	m	45	Facharzt	Praxis, Krankenhaus	Alexa, Dragon Naturally Speaking	5/Woche
I2	m	23	Gesundheits- & Krankenpfleger	Krankenhaus	Alexa	Unregelmäßig
I3	w	22	Ergotherapeutin	Praxis	Siri, Navigationsgerät Spracherkennung	Täglich
I4	w	22	Pflegefachfrau	Krankenhaus	Alexa	1/Woche
I5	w	48	Logopädin	Praxis	Siri, Messenger Spracherkennung	Täglich
I6	w	22	Gesundheits- & Krankenpflegerin	Krankenhaus	Google Spracherkennung	1–2/Woche
I7	w	57	Zahnärztin, Oralchirurgin	Praxis	–	–
I8	w	51	Krankenschwester	Krankenhaus, Klinik	Siri, Messenger & Auto Spracherkennung	Täglich

Zwei der Teilnehmer lehnten ein mündliches Interview per Videokonferenz ab. Diese erhielten einen ausführlichen schriftlichen Fragebogen mit den Interviewfragen. Alle anderen sechs Befragten willigten zu einem Online-Interview ein. Die Interviews dauerten zwischen 18 und 25 min. Alle Online-Interviews wurden aufgezeichnet und anschließend transkribiert. Die Interviews wurden von Oktober bis November 2021 geführt.

In der *Ergebnisanalyse* wurden alle Transkripte systematisch kodiert und kategorisiert, wobei die qualitative Inhaltsanalyse nach Mayring (2015) zur Anwendung kam. Dafür wurden alle Kernaussagen markiert und extrahiert. Daraufhin wurde jede Kernaussage einer Kategorie zugeordnet. Der Vorgang wurde wiederholt bis alle Transkripte analysiert sind und keine neuen Kategorien mehr entwickelt werden können. Alle Aussagen, die sich funktionalen und nicht-funktionalen Anforderungen zuordnen lassen, wurden markiert und isoliert.

Die letzte Phase umfasst einen *Anbietervergleich*. Um einen Überblick über die Fähigkeiten aktueller Spracherkennungsdienste zu geben, wurde abschließend ein Vergleich von Sprachdiensten mittels der Kriterien *Speech-to-Text Performance*, *Preisgestaltung* und *Datenschutz* vorgenommen. Die Speech-To-Text Performance ist relevant für die Fähigkeit auch komplexere medizinische Begriffe schnell und korrekt zu erkennen. Die Preisgestaltung soll zumindest einen ersten Eindruck über die Kosten der Dienste vermitteln und die Betrachtung des Datenschutzes ist für personenbezogene Gesundheitsdaten obligatorisch.

Tab. 2 Anforderungen an Sprachassistenten im Gesundheitswesen

Anforderung	Interviewfrequenz
<i>Funktional</i>	
Sprachübersetzung	8
Therapie-/Behandlungsdokumentation	8
Zugriff und Verwaltung der Patientenakte	7
OP-Berichte diktieren	4
Arztbriefe diktieren	3
Sprachaktivierte Geräte steuern	2
Fotodokumentation aufrufen	1
Termine vereinbaren	1
Neuaufnahme	1
<i>Nicht-funktional</i>	
Zugriffsschutz	5
Mobile Verfügbarkeit	4
Verschlüsselte Datenübertragung	3
Leicht bedienbare Oberfläche	2
Altersgerechte Bedienung	1

4 Anforderung aus der medizinischen Praxis

Aus den acht Interviews, die mit medizinischen Fachexperten geführt wurden, konnten insgesamt neun funktionale und fünf nicht-funktionale Anforderungen an Sprachassistenten im Gesundheitswesen abgeleitet werden. Tab. 2 fasst die funktionalen und nicht-funktionalen Anforderungen, die aus den Interviews erhoben werden konnten sowie deren Interviewfrequenz, zusammen. Im Folgenden wird auf einzelne Aussagen der Interviewteilnehmer mittels der Bezeichnung ‚I‘ gefolgt von der Nummer der Teilnehmer referenziert.

Die *Sprachübersetzung* während der Behandlung von Patienten ist die wichtigste funktionale Anforderung an den Einsatz von Sprachassistenten im Gesundheitswesen. Im Kern geht es um die Verbesserung der Verständigung zwischen Patienten, die ausschließlich Sprachen beherrschen, die vom medizinischen Personal nicht gesprochen oder verstanden werden. Das größte Problem bei der Kommunikation mit dem Patienten ist laut I8, „wenn die Patienten kein Deutsch [...] sprechen können“ und „eine Verständigung nur mit Händen und Füßen“ möglich ist. I2 weist darauf hin, „dass ab und an ein Dolmetscher nötig [ist], der dann aber natürlich seinen Arbeitsplatz für den Moment verlassen muss“. Ist niemand mit einer passenden Sprachqualifikation verfügbar, erfolgt die Kommunikation heute über Gestik und Mimik (I3). In seltenen Fällen wird ein Online-Übersetzer hinzugezogen (I3), wobei man sich bei der Verwendung bewusst ist, dass schnell Fehler in der Übersetzung entstehen können (I5). Darüber hinaus „kann man auch nicht alles ‚unmedizinisch‘ erklären“ (I3), d. h. man kommt nicht um eine Nennung medizinischer Fachbegriffe herum.

Mit ebenfalls acht Nennungen ist die Anforderung einer *Therapie-/Behandlungsdokumentation* ähnlich relevant wie die Sprachübersetzungsfunktion. Im Gesundheitswesen spielt die Dokumentation eine wesentliche Rolle. Dabei ist es egal, ob es

sich um einen OP-Bericht oder ein Patientengespräch handelt. „Im Gesundheitswesen musst du im Endeffekt alles, was du machst, dokumentieren“ (I2). I1 berichtet in diesem Zusammenhang vom Einsatz eines Spracherkennungssystems in der Form einer Diktiersoftware. Alle anderen Interviewteilnehmer wünschen sich ebenfalls eine Art Diktiersoftware, die in der Lage ist, medizinische Dokumentationsaufgaben durchzuführen. So sieht I3 eine Diktierfunktion als effizientes Hilfsmittel an, da es „irgendwo schon einfacher [ist], wenn du das einfach irgendwie in [das] Handy reinsprichst, das speichert das dann und du kannst das dann ausdrucken“. Patienten und deren Stellvertreter müssen häufig Dokumente ausfüllen. Bei allen befragten Interviewteilnehmern erfolgt dies aktuell noch schriftlich, so dass es auch dazu kommen kann, dass ein Angestellter den ursprünglichen Arbeitsplatz verlassen muss, um beim Ausfüllen eines Formulars zu helfen (I3). Als Formulare, die durch einen Speech-to-Text-Service digitalisiert werden können, wurden Datenschutzerklärung, Anamnesebogen und Neuaufnahmen genannt (I1, I3, I7, I8).

Zugriff und Verwaltung der Patientenakte sind für die Interviewteilnehmer weitere wichtige Anforderungen an den medizinischen Einsatz von Sprachassistenten. I3 wünscht sich ein Anlegen einer Patientenakte per Spracherkennung: „[...] könnte man nicht auch für jeden seiner Patienten eine Akte dadurch anlegen“? Allerdings ist dann die Speicherdauer zu berücksichtigen: „Aufgrund von rechtlichen Vorgaben, sind wir dazu verpflichtet, diese Daten mindestens 19 Jahre nach Abschluss der Behandlung [aufzubewahren]“ (I7). Eine Sprachassistentenlösung muss folglich in der Lage sein, Daten in das Patientenverwaltungssystem der medizinischen Einrichtung einfließen zu lassen. Des Weiteren befürwortet I1 ein Abspeichern der via Sprache erfassten Daten auf der elektronischen Gesundheitskarte (eGK).

Vier Interviewteilnehmer erwarten, dass eine Sprachsteuerung in der Lage ist, konkrete Dokumente zu erstellen. Die Anforderung „*OP-Bericht erstellen*“ wurde dabei dreimal genannt. I8 spricht über „eine Spracherkennung direkt im OP, um einen sofortigen und genauen OP-Bericht vor Ort zu [verfassen]“. Im Zentrum stehen die Aufzeichnung und das Verstehen von medizinischen Fachbegriffen. So wird unter anderem von I5 die besondere Herausforderung der Spracherkennung gesehen, „alle Fachbegriffe“ richtig zu erkennen. Zudem wurde die Anforderung *Arztbriefe diktieren* dreimal erwähnt.

I1 und I2 verweisen auf die Anwendung von *sprachaktivierten Geräten* im Gesundheitswesen. I1 beschreibt eine Idee, in der ein Röntgengerät mit einer Sprachfunktion ausgestattet ist. Die Spracherkennung soll es ermöglichen, auch ohne einen weiteren Assistenten im Raum das Röntgengerät auszulösen. Als Voraussetzung für dieses Einsatzgebiet nennt I2 eine absolute Präzision der Spracherkennung. Der Spracherkennungsdienst müsste „eine KI haben, die das 100 % präzise erkennen kann“ (I2). Ferner wird erwartet, dass deutlich geringere Komplikationen auftreten, weil die Gefahr, dass es „viele Kabel gibt, über die man stolpern kann oder viele Knöpfe, die man aus Versehen falsch drücken kann“ durch den Einsatz von Sprachassistenten sinkt (I2).

Die folgenden drei Anforderungen wurden jeweils nur einmal genannt und werden als eher unbedeutende Anforderungen betrachtet. I6 erwähnt, dass als erweiterte funktionale Anforderung der Einsatz einer Spracherkennung zum *Aufruf einer Fotodokumentation* genutzt werden kann. Ebenfalls könnte via Sprachassistent eine Inte-

gration mit dem Kalender der medizinischen Einrichtung hergestellt werden, so dass eine effiziente *Terminvereinbarung* ermöglicht wird. Abschließend wurde einmal auf die Digitalisierung der Neuaufnahme von Patienten durch einen Sprachassistenten hingewiesen.

Ergänzend zu den funktionalen Anforderungen an den Einsatz von Sprachassistenten im Gesundheitswesen wurden auch nicht-funktionale Anforderungen von den Teilnehmern vorgetragen. Dabei war mit fünf Nennungen der „*Zugriffsschutz*“ die wichtigste Anforderung. Einige der medizinischen Experten nennen Vertrauensprobleme als größte Einstiegsbarriere in der Verwendung eines Spracherkennungssystems (I1, I3, I5, I7, I8). I3 und I5 drücken ihre Bedenken aus, da beide das Gefühl haben, nach Benutzung einer Spracherkennungssoftware „Werbung für bestimmte Sachen“ vorgeschlagen zu bekommen, von denen sie kurz vorher noch gesprochen haben. I5 sagt zudem, dass er deswegen „keine sensiblen Daten darüber“ bespricht. Der Einsatz eines Sprachassistenten ist ein potenzielles Einfallstor für unberechtigte Zugriffe, so dass der Zugriff intelligent abgesichert sein muss und nur berechtigtes medizinisches Personal Zugriff haben darf. I3 ist der Auffassung, dass für eine ausreichende Sicherheit „[...] diese Face-ID oder Code eingeben eigentlich schon fast gar nicht mehr [ausreichen]“. I4 und I7 erwarten ein Sperren des Zugriffs, sobald man nicht mehr auf den Sprachassistenten zugreifen möchte. I4 merkt an, dass es öfter vorkommen kann, dass jemand seinen Arbeitsplatz verlässt, um kurz auf Toilette zu gehen und dabei „vergisst den Computer zu sperren“. Folglich sollte ein Sperren des Sprachassistenten ermöglicht werden.

Die zweitwichtigste nicht-funktionale Anforderung stellt die „*mobile Verfügbarkeit*“ dar. Die interviewten Mediziner empfinden es als einfacher und angenehmer eine mobile Smartphone-App für die Spracherkennung zu nutzen, anstatt eine installierte Software auf einem Computer (I3–I6). „Es wäre bestimmt gut, wenn das auf Mobilgeräten laufen würde, weil die hat man immer dabei“ meint etwa I5. Die dadurch entstehende Parallelisierung von Arbeitsschritten führt zu Zeitersparnissen (I3, I4). Zu den wichtigsten Eigenschaften einer solchen App gehört laut I5 die ständige Einsetzbarkeit: Der Sprachassistent „[...] sollte natürlich funktionieren“ (I5).

Die *Verschlüsselung der Datenübertragung* ist den Beteiligten ebenfalls wichtig. Es wäre ein großer Vorteil, wenn der Sprachassistent alle Sprachdaten und die Interaktion mit Bestandssystemen verschlüsselt, sodass ein Zugriff von außerhalb nur beschränkt möglich ist (I5). „Die Programme sind hier alle verschlüsselt“ erläutert I5. I1 weist darauf hin, dass bei einem Austausch von Patientendaten mit medizinischen Einrichtungen die Kommunikationsschnittstelle „ziemlich aufwändig verschlüsselt“ ist. Auch I5 ist der Ansicht, dass man „nicht fahrlässig mit diesen Daten“ (I5) umgehen sollte und verweist dabei auf das Verschicken von Patientendaten. I7 betont ebenfalls, dass im Falle einer Übertragung von Patientendaten eine Verschlüsselung unumgänglich ist.

Selten wurden Anforderungen in Bezug auf das Design der Benutzeroberfläche des Sprachassistenten genannt. Selbst wenn die Kommunikation hauptsächlich via Sprache stattfindet, so wird eine mobile App wahrscheinlich auch eine Oberfläche zur schnelleren Administration besitzen. Unter „*leicht bedienbare Oberfläche*“ verstehen die Interviewteilnehmer eine intuitive Bedienung des Prototyps, die keiner zusätzlichen Erklärung bedarf (I5, I7). In Bezug auf die Übersetzungsfunktion er-

klärt I5, dass diese „[...] eine sehr einfach zu bedienende Oberfläche haben [muss]“. I7 lehnt die Nutzung von Spracherkennungssystemen aufgrund einer komplizierten Darstellung ab. Spracherkennungssysteme werden von I7 als „kompliziert, zeitaufwändig und die Daten [sind] nicht geschützt“ dargestellt. I7 bräuchte auch diese Systeme nicht, „da ich mit den Fremdsprachen, die ich beherrsche, die Kommunikation mit meinen Patienten gewährleisten [kann]“. I5 denkt bei einer *altersgerechten Bedienung* an eine große Benutzeroberfläche mit Buttons, die groß genug sind, um sie sehen und drücken zu können (I5). Dabei wird vor allem auf Barrierefreiheit abgezielt. I5 ist der Meinung, dass durch große Felder „die Bedienung [der App] sehr intuitiv sein“ sollte, um es auch älteren Menschen zu ermöglichen die App zu bedienen.

Die erhobenen Anforderungen sind eine erste Grundlage zur Entwicklung von Sprachassistenten im Gesundheitswesen. Sie sind jedoch kontextabhängig und verlangen eine Reflektion in jedem medizinischen Anwendungsgebiet. Die Steuerung von Geräten ist etwa bei der Neuaufnahme eines Patienten im Krankenhaus weniger sinnvoll als während einer Behandlung.

5 Anbietervergleich

Um die funktionalen und nicht-funktionalen Anforderungen umzusetzen, ist im Kern ein Spracherkennungsdienst zur Transformation von Sprache in Text notwendig. In diesem Beitrag wird die Auswahl von Spracherkennungsservices auf Cloud-Anbieter fokussiert, da der Funktionsumfang bereits sehr ausgereift und die Zugriffsbarriere niedrig ist. Für die Auswahl eines Speech-to-Text-Services kommen folgende Anbieter in Frage: Amazon Web Services (AWS), IBM Cloud, Google Cloud (GCloud), Microsoft Azure. Die Anbieter werden im Hinblick auf die Kategorien Speech-to-Text Performance, Preisgestaltung und Datenschutz bewertet und verglichen.

Speech-To-Text Performance Damit die *Speech-to-Text Performance* messbar gemacht werden kann, wurden die vier Anbieter einem Test unterzogen. Dafür wurde eine 30-sekündige Sprachdatei erstellt, in der ein Abschnitt aus einem medizinischen Fachbuch vorgelesen wurde. Diese Datei wurde den Spracherkennungsdiensten der Cloud-Anbieter zur Verfügung gestellt. Das Ergebnis wurde daraufhin sowohl auf Rechtschreibung (hauptsächlich Groß- und Kleinschreibung) als auch auf falsch verstandene Wörter und Syntaxzeichen kontrolliert. Darüber hinaus wurde die Zeit von

Tab. 3 Performance-Vergleich Spracherkennungsdienste

Cloud-Dienstleister	Zeitbedarf (in Sek.)	Falsches Wort (absolute Häufigkeit)	Korrekte Satzzeichen (relative Häufigkeit) (%)	Rechtschreibfehler (absolute Häufigkeit)	Preis (\$/Std.)
AWS	28,28	11	90	2	1,44
IBM	35,24	16	20	15	1,20
GCloud	22,08	7	30	3	1,44
Azure	17,17	9	100	2	1,00

der Eingabe bis zum fertigen Schlussergebnis gemessen, was einen Vergleich ermöglicht. In Tab. 3 sind die Ergebnisse des Performancevergleichs dargestellt. Der Zeitbedarf für die Sprachumwandlung liegt zwischen 17 und 28 s, wobei Azure mit 17 s die schnellste Umwandlung durchführen kann. Falsch erkannte Wörter wurden zwischen neun und 16-mal wiedergegeben. Bei Azure war während der Spracherkennung zu beobachten, dass der erkannte Text automatisch mit einer grammatikalischen KI überprüft wird. Lediglich bei der Erkennung des Wortes „Aortenaneurysma“ hatte Azure Schwierigkeiten. Die Erkennung von Satzzeichen ist sehr unterschiedlich und reicht von 20 % korrekt gesetzter Satzzeichen bis hin zu 100 %. Ähnlich wie bei der IBM Cloud ist bei der GCloud aufgrund fehlender Satzzeichen nicht zu erkennen, wann ein Satz logisch aufhört oder beginnt. Bei der Setzung von Satzzeichen schneidet AWS ähnlich gut wie Azure ab, hatte aber nicht nur Probleme beim Verstehen des Wortes „Aorta“, sondern auch bei „Gefäßaussackung“ und „Ruptur“. Die 15 Rechtschreibfehler bei IBM setzen sich ausschließlich aus Groß- und Kleinschreibfehlern zusammen. Darüber hinaus scheint es so, als würde IBM lediglich die Worte verstehen und den grammatikalischen oder logischen Sachzusammenhang nicht überprüfen. Außerdem erscheint der Text auf den ersten Blick als ein sinnloses Aneinanderreihen von Worten. Übersetzer versuchen in der Regel innerhalb eines Satzes anhand der Grammatik und Syntax einen Sachzusammenhang zu erkennen und sinngemäß zu übersetzen.

Preisgestaltung Ein weiteres wichtiges Auswahlkriterium ist die *Preisgestaltung* der Dienste. Der Einfachheit halber wurden alle Modelle auf Kosten pro Stunde hochgerechnet und verglichen (siehe Tab. 3). GCloud bietet zwei Preismodelle an. Das billigere Kostenmodell mit 0,004\$/15 Sek. beinhaltet ein sogenanntes Daten-Logging. Mit einem Aufpreis von 0,002\$/15 Sek. erhält man die gleiche Leistung, jedoch ohne den Mitschnitt und die Auswertung der Daten. Um direkt das Thema Datenschutz einzubeziehen, wird für GCloud das teurere Preis-Modell in Betracht gezogen. Azure verlangt mit einem Dollar pro Stunde den niedrigsten Preis.

Datenschutz AWS Kunden können selbst entscheiden, wie mit ihren Daten umgegangen wird. So können sie bspw. entscheiden, dass alle Daten anonymisiert oder verschlüsselt gespeichert werden. AWS folgt ferner den Standardvertragsklauseln, die von der Europäischen Kommission definiert und genehmigt wurden, um personenbezogene Daten von einem Datenverantwortlichen an einen Datenverarbeiter außerhalb des Europäischen Wirtschaftsraums zu übertragen, ohne gegen geltendes Recht zu verstoßen. IBM überlässt es dem Kunden selbst, die DSGVO-Richtlinien einzuhalten. Dabei bietet IBM verschiedene Dienste an (z. B. IBM Security Guardium-Software, IBM Data Risk Manager), die bei der Einhaltung der DSGVO unterstützen. Außerdem werden sogenannte Datenschutzservices angeboten (u. a. Datenschutzrichtlinienservices, Datensicherheitsservices), mit denen der Kunde festlegen kann, wie die Daten geschützt werden sollen. IBM verwendet ebenfalls die EU-Standardvertragsklauseln. GCloud hat einen kundenorientierten Ansatz in Bezug auf Datenschutz, Kontrolle und Compliance. GCloud verpflichtet sich zur Einhaltung der DSGVO in Bezug auf die Verarbeitung personenbezogener Daten in allen angebotenen Diensten. Darüber hinaus werden zusätzliche Sicherheitsfeatures an-

geboten, die stets aktualisiert werden¹. Der Kunde muss allerdings selbst die Methoden und den Zweck zur Verarbeitung personenbezogener Daten festlegen. Auch GCloud setzt EU-Standardvertragsklauseln ein. Azure versichert, dass die Daten nicht zum Data Mining genutzt werden oder an Werbedienste weitergegeben werden. Der Speicherort der Daten kann vom Kunden gewählt und überprüft werden. Es werden sowohl ruhende Daten als auch übertragende Daten mittels Verschlüsselung und Transportprotokoll (IPsec, TLS) geschützt. Allen Anbietern ist jedoch gemein, dass sie – unabhängig von Standort des Rechenzentrums – Daten aufgrund staatlicher Anordnung, wie bspw. dem U.S. Patriot Act, an Dritte weitergeben müssen, was im Falle von Gesundheitsdaten besonders kritisch ist.

Betrachtet man alle Kriterien und die Bewertung der Anbieter, so ist derzeit die Entwicklung von Sprachsystemen für die Anwendung im Gesundheitswesen auf Basis von Microsoft Azure vielversprechend. Die Entwicklung und Anwendung eines Prototyps kann hier weitere Klarheit bringen, was zukünftige Arbeiten motiviert.

6 Diskussion

Der vorliegende Beitrag betrachtet funktionale und nicht-funktionale Anforderungen an Sprachassistenten im Gesundheitswesen. Auf Basis von Interviews konnten in Summe neun funktionale und fünf nicht-funktionale Anforderungen erhoben werden. Die ermittelten Anforderungen sind nicht gleichsam in allen medizinischen Bereichen anwendbar. So ist etwa eine Übersetzungsfunktion sehr sinnvoll bei der medizinischen Erstaufnahme oder während eines Arztgesprächs mit dem Patienten. Bei einer Operation im OP-Saal wird diese Funktion eher weniger zur Anwendung kommen. Besondere Bedeutung bei den nicht-funktionalen Anforderungen hat der Zugriffsschutz, also der Schutz personenbezogener Daten. Die hohe Bedeutung des Datenschutzes lässt sich zum Teil damit erklären, dass die Studie in Deutschland durchgeführt wurde. Die Sensibilisierung für Datenschutz ist dort besonders hoch und zudem gesetzlich durch die DSGVO reguliert.

Die Ergebnisse dieses Beitrags müssen differenziert betrachtet werden und deren Aussagekraft ist begrenzt. Zur Datenerhebung wurden acht medizinische Fachexperten aus Deutschland befragt. Sie sind Endbenutzer und haben aus ihrer Perspektive Anforderungen benannt. Nicht befragt wurden IT-Entscheider, die neben den Endnutzern eine wichtige Rolle bei der Einführung von Informationssystemen spielen. Die Ergebnisse der Interviews wurden nicht weiter evaluiert. Bei der Auswahl der Cloud-Dienstleister wurde die Anzahl auf vier große US-amerikanische Anbieter beschränkt. Kleinere Start-ups und Sprach-Dienste aus anderen Ländern wurden nicht betrachtet.

Zukünftige Arbeiten sollten die Ergebnisse dieser Studie validieren in dem die Befragung in anderen Ländern und mit IT-Entscheidern medizinischer Einrichtungen wiederholt wird. Ferner kann die Durchführung von Fallstudien mit Sprachassistenten in Krankenhäusern und Arztpraxen dabei helfen, die Akzeptanz bzw. Barrieren

¹ siehe hierzu: <https://cloud.google.com/docs/security/encryption/customer-supplied-encryption-keys?hl=de>.

von Sprachassistenten bei medizinischem Personal zu untersuchen und Handlungsempfehlungen für den Einsatz im Gesundheitswesen zu erarbeiten.

Funding Open Access funding enabled and organized by Projekt DEAL.

Open Access Dieser Artikel wird unter der Creative Commons Namensnennung 4.0 International Lizenz veröffentlicht, welche die Nutzung, Vervielfältigung, Bearbeitung, Verbreitung und Wiedergabe in jeglichem Medium und Format erlaubt, sofern Sie den/die ursprünglichen Autor(en) und die Quelle ordnungsgemäß nennen, einen Link zur Creative Commons Lizenz beifügen und angeben, ob Änderungen vorgenommen wurden.

Die in diesem Artikel enthaltenen Bilder und sonstiges Drittmaterial unterliegen ebenfalls der genannten Creative Commons Lizenz, sofern sich aus der Abbildungslegende nichts anderes ergibt. Sofern das betreffende Material nicht unter der genannten Creative Commons Lizenz steht und die betreffende Handlung nicht nach gesetzlichen Vorschriften erlaubt ist, ist für die oben aufgeführten Weiterverwendungen des Materials die Einwilligung des jeweiligen Rechteinhabers einzuholen.

Weitere Details zur Lizenz entnehmen Sie bitte der Lizenzinformation auf <http://creativecommons.org/licenses/by/4.0/deed.de>.

Literatur

- Adelmeyer M, Meier P, Teuteberg F (2019) Security and privacy of personal health records in cloud computing. In: 14. Internationale Tagung Wirtschaftsinformatik
- Catania F, Di Nardo N, Garzotto F, Occhiuto D (2019) Emoty: An emotionally sensitive conversational agent for people with neurodevelopmental disorders. In: 52th Hawaii International Conference on System Sciences (HICSS). Maui, Hawaii
- Eggert M, Stanke M-A (2020) Adoption of integrated voice assistants in health care—Requirements and design guidelines. In: 15th International Conference on Wirtschaftsinformatik. Potsdam
- Guy I (2016) Searching by talking. Analysis of voice queries on mobile web search. In: Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval. Pisa, S 35–44
- Helmke H, Rataj J, Mühlhausen T, Ohneiser O, Ehr H, Kleinert M et al (2015) Assistant-based speech recognition for ATM applications. In: Eleventh USA/Europe Air Traffic Management Research and Development Seminar (ATM2015)
- Jiang J, Awadallah HA, Jones R, Ozertem U, Zitouni I, Gurunath Kulkarni R, Khan OZ (2015) Automatic online evaluation of intelligent assistants. In: Proceedings of the 24th International Conference on World Wide Web—WWW '15. Florence, Italy, S 506–516
- Kepuska V, Bohouta G (2018) Next-Generation of virtual personal assistants (Microsoft Cortana, Apple Siri, Amazon Alexa and Google Home). In: 2018 IEEE 8th Annual Computing and Communication Workshop and Conference (CCWC). IEEE, Las Vegas, S 99–103
- Maas L, Geurtsen M, Nouwt F, Schouten S, van de Water R, van Dulmen S et al (2020) The Care2Report system: Automated medical reporting as an integrated solution to reduce administrative burden in healthcare. In: Proceedings of the 53th Hawaii International Conference on System Sciences (HICSS)
- Mayring P (2015) Qualitative Inhaltsanalyse: Grundlagen und Techniken, 11. Aufl. Beltz, Weinheim, Basel
- Nitta Y, Murayama Y (2019) Privacy-aware remote monitoring system by skeleton recognition. In: Proceedings of the 52nd Hawaii International Conference on System Sciences
- Otoo BAA, Salam AF (2018) Mediating effect of intelligent voice assistant (IVA), user experience and effective use on service quality and service satisfaction and loyalty. In: International Conference on Information Systems (ICIS)
- Purao S, Xu Y (2018) So talk to me: Designing voice-enabled tools for health-related information. In: Proceedings of the Americas' Conference on Information Systems (AMCIS)
- Schmidhuber J (2015) Deep learning in neural networks: an overview. *Neural Netw* 61:85–117. <https://doi.org/10.1016/j.neunet.2014.09.003>

- Statista (2022) Absatz von intelligenten Lautsprechern weltweit bis Q1 2022. <https://de.statista.com/statistik/daten/studie/818982/umfrage/absatz-von-intelligenten-lautsprechern-weltweit-pro-quartal/>. Zugegriffen: 26. Sept. 2022
- Tsujino K, Iizuka S, Nakashima Y, Isoda Y (2013) Speech recognition and spoken language understanding for mobile personal assistants: a case study of “Shabette Concier”. In: IEEE 14th International Conference on Mobile Data Management (MDM). IEEE, Milan, S 225–228
- Vollmert M, Hoogen E, Calchera CA (2022) Voice assistant technology in logistics: An opportunity for optimization. <https://dhl-consulting.com/insights/blog/voice-assistant-technology-logistics/>. Zugegriffen: 26. Sept. 2022