BARRIEREFREIES INFORMIEREN UND KOMMUNIZIEREN MITTELS SPRACHERKENNUNG

Christiane Berndt und Rainer G. Spallek

Technische Universität Dresden
Christiane.Berndt1@tu-dresden.de

Kurzfassung: Verbale Kommunikation ist, obwohl sie einen Großteil der täglichen, menschlichen Interaktion ausmacht, nicht immer und für alle barrierefrei. In dieser Arbeit soll ausgehend von den Möglichkeiten der Spracherkennung aufgezeigt werden, welche alternativen Kommunikationswege existieren und wie diese in einem neuartigen System kombiniert werden können.

1 Einleitung

Sprache bildet die Grundlage der täglichen Kommunikation zwischen Menschen. Sie ermöglicht den Austausch von Informationen und ist die Basis der sozialen Interaktion. Diese Art der Kommunikation kann aber nicht in jeder Situation oder für jeden Teilnehmer verlust- und störungsfrei erfolgen. Somit werden alternative Kommunikationsformen nötig, um einen Informationsaustausch unter Berücksichtigung der Anforderungen und Bedürfnisse der Nutzer zu ermöglichen. Daraus ergibt sich die Fragestellung, welche Möglichkeiten die Spracherkennung in Bezug auf die Vermittlung von Informationen auf alternativen Wegen bietet und wer davon profitiert.

Diese Arbeit stellt ein Konzept für ein neuartiges Assistenzsystem vor, welches existierende Lösungen integriert und erweitert, um ein ganzheitlich nutzbares System zur barrierefreien Kommunikation zu schaffen. Hierzu werden zuerst die situations- und nutzerbezogenen Barrieren in der verbalen Kommunikation auf Basis der Spracherkennung analysiert. Ausgehend von dieser Analyse werden existierende assistive Technologien bezüglich ihrer Zielgruppen betrachtet. Anschließend wird ein neues Assistenzsystem entworfen, dass sich situations- und nutzerunabhängig der Barrieren in der verbalen Vermittlung von Informationen annimmt. Zum Schluss wird darauf eingegangen wer in welchen Situationen von diesem neuen System profitieren wird.

2 Ausgangspunkt

Aus der Verarbeitung von Sprachsignalen ergeben sich nach Euler [10] mehrere mögliche Anwendungsgebiete:

- Spracherkennung - Was wurde gesprochen?
- Keyword-Spotting - Wurden bestimmte Schlüsselwörter gesprochen?
- Sprechererkennung - Wer hat gesprochen?
- Sprecherverifikation - Hat Person X gesprochen?
- Sprachenidentifikation - In welcher Sprache wurde gesprochen?

Von besonderem Interesse für das zu realisierende System ist die Spracherkennung, da sie zum Einsatz kommt, wenn andere Kommunikationskanäle belegt sind oder sich als ungeeignet erweisen. Sie ermöglicht es, Sprachsignale in eine maschinen-lesbare Form umzuwandeln und so der weiteren Verarbeitung zugänglich zu machen. Für die Schwierigkeitsstufen der Spracherkennung ergeben sich 3 Unterscheidungskriterien: Sprechweise, Wortschatz und Benutzerkreis (siehe Tabelle 1).

Tabelle 1 - Schwierigkeitsstufen der Sprachverarbeitung nach Euler [10]

	einfach	schwer
Sprechweise	einzelne Worte	Sätze
Wortschatz	klein	groß
Benutzerkreis	sprecherabhängig	sprecherunabhängig

Die größte Herausforderung liegt in der Verfügbarkeit in einer gut funktionierenden Spracherkennung, die sprecherunabhängig kontinuierlich gesprochene Sätze mit großem Wortschatz zuverlässig erkennt und als Text ausgibt. Diese Form der Spracherkennung bildet den Ausgangspunkt für die weiteren Überlegungen in diesem Projekt.

3 Barrieren in verbaler Kommunikation

Kommunikation kann auf vielen Wegen erfolgen, den meist genutzten stellt dabei der verbale Informationsaustausch dar. Einige weitere Kommunikationsarten und ihre Medien sind:

- visuell: Text, Bilder, Gesten, Mimik, Gebärdensprache
- audio-visuell: Video
- taktil: Braille, Lorm.

Trotz der Vielfalt an Kommunikationsmedien erfolgt 95% der im Alltag ausgeführten Kommunikation über das Sprechen [1]. In welchen Situationen und für welche Personen die ausschließlich verbale Vermittlung von Informationen Schwierigkeiten birgt, zeigen die nächsten beiden Abschnitte.

3.1 Szenarien

Die Wissensvermittlung in der schulischen, beruflichen und universitären Aus- und Weiterbildung findet hauptsächlich über Vorlesungen und Vorträge statt, die sich meist auf die verbale Form beschränken und durch schriftliche Darstellungen bestenfalls ergänzt werden. Der Kommunikationsvorgang beschränkt sich hierbei nicht nur auf das gesprochene Wort, sondern umfasst auch die Gestik des Vortragenden und andere visuelle Darstellungen an Tafel oder durch Projektionen.

Die zu vermittelnde Information kann durch mehrere Faktoren im Umfeld gestört werden. Mögliche Störfaktoren sind zum Beispiel Umgebungsgeräusche, deren Quellen inner- oder außerhalb des Vortragsbereiches liegen können, oder ein großer Abstand zum Vortragenden, der die Verständlichkeit von Rede und die Sichtbarkeit von Gestik und visuellen Darstellungen beeinträchtigt.

Neben der direkten, unmittelbaren Kommunikation erfolgt der Informationsaustausch immer öfter durch digitale Medien, die in vielen Bereichen eine Verbesserung und Vereinfachung der Kommunikation bedeuten und neue Wege der nicht verbalen Kommunikation aufzeigen, wie

zum Beispiel E-Mail, Internetforen und Instant Messenger. Eine weitere digitale, sprachbasierte Form der Kommunikation wird durch Videokonferenzen ermöglicht. Der direkte Austausch von Bildern und Ton zeigt sich jedoch als störanfällig und nicht in vollem Umfang nutzbar, wenn entweder kein Ton vorhanden ist oder der Kommunikationspartner keinen Zugang zu Bildern und Ton hat.

Ein weit verbreitetes und stark genutztes Medium zur Informationsvermittlung ist das Internet, welches Möglichkeiten wie Webseiten, Podcasts oder Videos bietet. Doch obwohl im Bereich der Barrierefreiheit des Internets viel getan wurde und noch getan wird, sind die nicht rein visuellen Medien (Podcast und Video) ebenso störanfällig wie Videokonferenzen.

3.2 Nutzer

Neben den oben erwähnten Situationen gibt es viele Menschen, deren Zugang zu verbaler Kommunikation durch erhöhte Barrieren erschwert ist. Ein großes Hindernis beim Zugang zu sprachlich vermittelten Informationen stellt ein vermindertes Hörvermögen dar. Dies betrifft allein in Deutschland 14 Millionen Menschen, von denen 2,5 Millionen als schwerbehindert und 80.000 als gehörlos gelten [6]. Da die Kommunikation über den auditiven Kanal zu circa 60 % über optische Reize und zu circa 40 % über akustische Reize funktioniert [1], ergeben sich auch für Menschen mit vermindertem Sehvermögen Barrieren. Von einem verminderten Sehvermögen sind in Deutschland 650.000 Menschen betroffen, 150.000 davon sind blind und 500.000 sehbehindert [5]. Vor allem ältere Menschen, deren Zahl durch die demographische Entwicklung weiterhin steigt, leiden oft an einer altersbedingten Verschlechterung des Seh- und Hörvermögens. Doch auch für Menschen mit kognitiven Problemen und chronisch Kranke ergeben sich hohe Barrieren im Umgang mit einer Kommunikation, die sich auf Bild und Ton beschränkt.

Im Alltag entstehen auch für weitere Personenkreise Barrieren bei der Aufnahme und Verarbeitung von Informationen. So ist es Eltern mit Kind nicht immer möglich, ihre Konzentration auf eine Informationsquelle zu fokussieren und so der Informationsvermittlung ununterbrochen zu folgen, da auch das Kind ein gewisses Maß an Aufmerksamkeit fordert. Nicht-Muttersprachler erleben oft eine sprachliche Barriere bei Informationsangeboten, deren Geschwindigkeit, Deutlichkeit und Qualität sich nicht beeinflussen lassen. Somit profitieren nicht nur Menschen mit besonderen Bedürfnissen von einer Erhöhung der Erreichbarkeit und Verständlichkeit von Informationen, sondern jeder, der verschiedene Informationsquellen nutzt.

4 Assistive Technologien für Kommunikation

Für einige der im vorherigen Abschnitt angesprochen Nutzergruppen gibt es bereits Hilfsmittel, die den Zugang zu Informationen erleichtern sollen. Zuerst werden einige der etablierten Technologien vorgestellt, bevor ein neues Hilfsmittel, dass auf Spracherkennung basiert, entworfen wird.

4.1 Existierende Technologien

Bei Filmen besteht für Menschen mit Hörbehinderungen die Möglichkeit, das Gesprochene und andere Geräusche beschreibende Untertitel einzublenden. Für Menschen mit Sehbehinderungen existieren separate, das Bild beschreibende Tonspuren. Allerdings ist ersteres oft nur bei fremdsprachigen Produktionen der Fall. Die Deutsche Hörfilm GmbH hat seit 1998 circa 500 Filme mit Tonspuren für blinde oder sehbehinderte Menschen unterlegen können [3]. Dieses Angebot ist jedoch eine Spartenlösung und kann, unter anderem begründet durch den hohen Aufwand der Erstellung, nur für einen Bruchteil des existierenden Filmangebots umgesetzt werden.

Nachrichtensendungen können ihr Informationsangebot einen größeren Nutzerkreis durch die Verwendung von Laufschriften zugänglich machen. Diese sind jedoch momentan nur bei wenigen Sendern wie zum Beispiel CNN [2] und N24 [7] verfügbar. Diese Lauftexte sind nicht nur für Menschen mit besonderen Bedürfnissen interessant, sondern werden zur Übersicht über die aktuellen Informationen auch von weiteren Personengruppen ausgiebig genutzt. Desweiteren werden vereinzelt Nachrichten mit Gebärdenvideos unterlegt und somit auch tauben Menschen in vollem Umfang zugänglich gemacht, so zum Beispiel bei Phoenix [8].

Eine weitere Hilfe für hörgeschädigte Menschen in der direkten Kommunikation bilden Schriftdolmetscher, die das gesprochene Wort wortwörtlich oder in zusammengefasster Form gleichzeitig zum Sprechvorgang niederschreiben. Gebärdendolmetscher übertragen das gesprochene Wort in eine eigenständige, visuell wahrnehmbare natürliche Sprache, die für gehörlose oder stark schwerhörige Menschen zugänglich ist. Jedoch muss diese Personengruppe einen Großteil des Alltages allein und ohne fremde Hilfe bestreiten, da weder Gebärden-, noch Schriftdolmetscher zu jeder Zeit für jeden zur Verfügung stehen.

Über den Computer können weitere assistive Technologien zur Verfügung gestellt werden. Exemplarisch seien hier Braille-Zeile, Screenreader und Sprachsynthese-Programme für sehgeschädigte Menschen genannt.

Zur Zeit gibt es jedoch keine Technologie, die sich umfassend der Barrieren im Umgang mit verbaler Kommunikation annimmt und versucht, diese zu reduzieren. Die existierenden Insellösungen beschränken sich auf einzelne Teilgebiete und es erfolgt keine Zusammenführung und Erweiterung der vorhandenen Technologien zu einem sinnvollen und anwendbaren System.

4.2 Entwurf eines neuen Hilfssystem

Die in Abschnitt 4.1 vorgestellten Assistenzsysteme beschränken sich auf einzelne Situationen oder Nutzergruppen und bietet keine übergreifende Lösung für die Barrieren in der verbalen Kommunikation. Im Folgenden wird auf Grundlage der Spracherkennung ein neues System konzipiert und vorgestellt, welches einen umfassenden Zugang zu verbalen Informationen bietet.

4.2.1 Grundlage

Ausgangspunkt für das neue System ist eine verbal vermittelte Information, die ein nicht verbales Medium umgewandelt werden soll. Um die gesprochene Information in eine textuelle Form umzuwandeln, sind in der Spracherkennung mehrere Schritte notwendig (siehe Abbildung 1).

Geräuschreduktion, Quellentrennung
Merkmalsextraktion
Akustische Modelle
Wortlexikon
Statistisches Sprachmodell
Syntax
Semantik
Pragmatik

Abbildung 1 - Verarbeitungsstufen in einem sprachverstehenden System nach Euler [10]

Diese Abbildung erhebt keinen Anspruch auf Vollständigkeit und universelle Gültigkeit, sondern dient nur der Veranschaulichung der Komplexität der Spracherkennung. All diese Schritte

führen am Ende zu einem Wort, einer Aussage oder einem Satz, der die höchste Wahrscheinlichkeit bezüglich des Gesprochenen aufweist und der in textueller Form ausgegeben werden kann. Erfolgt die Spracherkennung sprecherunabhängig, für große Wortschätze und für ganz Sätze, steigt die Komplexität und Schwierigkeit des Vorhabens.

4.2.2 Prinzipielles Vorgehen

Um eine verbal kommunizierte Information weiterbearbeiten zu können, wird sie per Video aufgezeichnet, welches einer manuellen Audio- und Video-Bearbeitung unterzogen werden kann. Anschließend erfolgt mittels der in Abbildung 1 aufgeführten Verarbeitungsschritte die Spracherkennung und Umwandlung in eine textuelle Form. Dieser Text bildet die Basis für das weitere Vorgehen.

Das Medium Text bietet viele Möglichkeiten zur Umwandlung und Weiterverarbeitung. So kann der Text als Untertitel in die Aufzeichnung der Information eingebettet werden oder separat, parallel zur akustischen Wiedergabe der Information dargestellt werden. Mittels farblicher Unterlegung kann dann im Volltext das gerade Gesprochene hervorgehoben werden.

Neben einer textuellen Ausgabe der verbalen Informationen bietet die Ausgabe in Gebärdensprache eine interessante Perspektive. Hierzu muss der Text mit Hilfe einer Gestennotation in Bewegungsdaten überführt werden, welche dann durch einen Avatar visualisiert werden können. Für die Generierung der Gestennotation wird ein Lexikon der Gebärdensprach-Grammatik benötigt, da diese anderen Regeln folgt als die gesprochene Sprache [9].

Eine weitere Möglichkeit der Ausgabe ist die Sprachsynthese, die den schon vorverarbeiteten Text nach einer Analyse der Wort- und Satzstruktur und der Verkettung und Generierung der Sprachsignalparameter als synthetisiertes Sprachsignal ausgibt [11]. Hierbei wird zukünftig zu prüfen sein, in wie weit eine Sprachsynthese die Verständlichkeit des Gesprochenen erhöhen kann.

Zusätzlich zu der Ausgabe über verschiedene Medien kann durch das Verfahren des Keyword-Spotting eine Indizierung des Textes mit Schlagworten erfolgen, die so eine Suche innerhalb des Textes und des Videos nach bestimmten Stichworten ermöglicht.

Durch geeignete Darstellungsformen soll es nun möglich sein, die aufgezeichnete Information als Video und die gewünschten Zusatzoptionen wie Text, Untertitel oder Gebärdenvideo synchron abzurufen. Potential hierfür bietet DAISY (Digital Accessible Information System, [4]), ein weltweiter Standard für navigierbare, zugängliche Multimedia-Dokumente, der sich momentan noch auf Hörbücher beschränkt, in naher Zukunft jedoch auf Video-Inhalte erweitert werden soll.

Eine geeignete Archivierung kann nun die auf unterschiedliche Weisen dargestellte Information sichern und auch zukünftig zugänglich und adressierbar machen.

Das gesamte Vorgehen wird in Abbildung 2 noch einmal kurz, schematisch verdeutlicht. Eine interessante Fragestellung ist hierbei noch, ob ein solches System auch live, das heißt ohne wahrnehmbare Verzögerung, betrieben werden kann.

5 Schlussfolgerungen

Die Entwicklung eines Systems, dass verbal kommunizierte Informationen in andere Medien umwandelt, stellt eine große Herausforderung dar. Hierbei werden verschiedenste Arbeits- und Fachgebiete kombiniert, wie zum Beispiel Video- und Audiobearbeitung, Spracherkennung und -synthese, Textbearbeitung, Darstellung von Text und Informationen, Datenbanken, sowie Gebärdensprache.

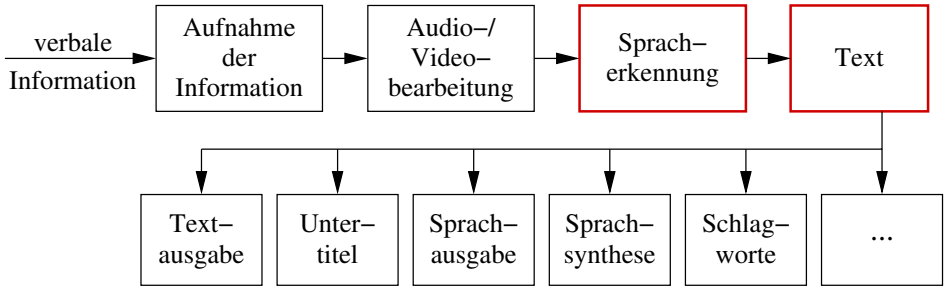

Abbildung 2 - System zur Transformation verbaler Information in andere Medien

Abschließend wird noch einmal kurz auf die Personengruppen eingegangen, denen das System einen auf ihre Bedürfnisse angepassten Zugang zu Informationen bietet, und auf die Situationen, in denen es angewandt werden kann.

Menschen mit Hörbehinderungen An Vorlesungen und Vorträge, die sonst durch Störfaktoren nur unzureichend zugänglich waren, kann nun aktiv und direkt teilgenommen werden. Gebärdenvideos ermöglichen auch tauben Menschen eine Zugang zu vorher nicht erreichbaren Informationsangeboten.

Menschen mit Sehbehinderungen Für sehgeschädigte Menschen können nun durch eine vergrößerte Wiedergabe der Aufnahme auch die visuellen Darstellungsformen eines Vortrages, einer Nachrichtensendung oder einer Verwaltungsinformation zugänglich gemacht werden und somit die akustischen Informationen ergänzen. Desweiteren können die vorher möglicherweise nicht gut verständlichen akustischen Informationen nun in schriftlicher Form per Braille oder Sprachsynthese ausgegeben werden.

Vorlesungen und Vorträge Abgesehen von der kurzfristigen Erhöhung der Zugänglichkeit durch die Präsentation des Inhaltes durch andere Medien wird der Vortrag durch die Archivierung auch dauerhaft zugänglich, nutzbar und bietet neue Möglichkeiten durch die Stichwortsuche. Desweiteren können Vorträge und Vorlesungen in einem störungsfreien Umfeld wiederholt werden.

Nachrichtensendungen, Videos und Podcasts Durch das Versehen mit Untertiteln oder separaten Text werden Nachrichtensendungen direkt für hörgeschädigte Menschen zugänglich. Auch Menschen mit kognitiven Problemen können durch das Nachlesen im Text den sonst manchmal schwer verständlichen Nachrichten folgen. Die Nachbearbeitung von existierenden Videos oder Podcasts im Internet öffnet das Benutzerfeld hin zu Menschen, denen der Zugang zu akustischen Formaten bis dahin verwehrt blieb.

Videokonferenzen Diese können jetzt von Nutzern mit unterschiedlichen Anforderungen genutzt werden. So sind Konferenzen zwischen Hörenden und Hörgeschädigten, sowie Sehgeschädigten und Sehenden möglich.

6 Zusammenfassung

Ausgehend von den Anwendungsgebieten der Spracherkennung stellte sich die Frage, welche Möglichkeiten es gibt, verbal vermittelte Informationen durch alternative Medien darzustellen und wer in welchen Situationen davon profitiert. In Abschnitt 3 wurde gezeigt, dass in

vielen Alltagssituationen die verbale Kommunikation eingeschränkt oder gestört ist und das Verständnis der Informationen nicht immer gewährleistet werden kann. Dies liegt nicht nur in den Bedürfnissen und Anforderungen der Menschen, sondern auch in den Situationen und ihrer Störanfälligkeit selbst begründet. Bei Betrachtung der existierenden assistiven Technologien (Abschnitt 4.1) fällt auf, dass viele dieser Hilfsmittel sich nur für bestimmte Nutzergruppen und Situationen eignen und derzeit kein System existiert, dass sich umfassend den Barrieren in der verbalen Kommunikation annimmt und sie benutzergruppen- und situationsübergreifend beseitigt. Um diesen Mangel zu beseitigen, wurde ein neues Hilfssystem auf Basis der Spracherkennung konzipiert (Abschnitt 4.2). Dieses neue System ermöglicht den einzelnen Benutzergruppen einen Zugriff auf verbal vermittelte Informationen unter Berücksichtigung ihrer Anforderungen und Bedürfnisse, unabhängig von der gegebenen Situation. Zudem können die aus der verbalen Kommunikation erstellten nicht verbalen Medien auch mit zusätzlichen Informationen versehen und gespeichert werden. Somit wird ein späterer Zugriff auf einmal vermitteltes Wissen möglich.

Literatur

[1] *Vorlesung Technische Sprachkommunikation*. Technische Universität Dresden, Institut für Akustik und Sprachkommunikation, AG Systemtheorie und Sprachkommunikation, April 2004.

[2] *CNN*. www.cnn.com, Februar 2009.

[3] *Deutsche Hörfilm GmbH*. www.hoerfilm.de, Februar 2009.

[4] *Deutsche Zentralbücherei für Blinde zu Leipzig, DAISY*. www.dzb.de, April 2009.

[5] *Deutscher Blinden- und Sehbehindertenverband (DBSV)*. www.dbsv.org, Februar 2009.

[6] *Gebärdenwerk*. www.dgs-filme.de//GWHomepage/, März 2009.

[7] *N24*. www.n24.de, Februar 2009.

[8] *Phoenix*. www.phoenix.de/content/phoenix/start, Juli 2009.

[9] *Virtual Human Signing at UEA*. www.visicast.cmp.uea.ac.uk, März 2009.

[10] EULER, S.: *Grundkurs Spracherkennung - Vom Dialog zum Sprachsignal - Grundlagen und Anwendungen verstehen - Mit praktischen Übungen*. Vieweg Verlag.

[11] GEISER, G.: *Mensch-Maschine-Kommunikation*. Oldenbourg Verlag.