

SPRACHTECHNOLOGIE ZUR UNTERSTÜTZUNG VON MENSCHEN MIT SINNESBEHINDERUNGEN

*Klaus Fellbaum, (ehem.) Lehrstuhl Kommunikationstechnik,
Brandenburgische Technische Universität Cottbus*

fellbaum@tu-cottbus.de

Kurzfassung: Der vorliegende Beitrag befasst sich mit dem Einsatz der elektronischen Sprachverarbeitung zur Kommunikationsunterstützung bei Sehbehinderten, Blinden sowie Schwerhörigen und Gehörlosen. Er soll zugleich auch einige Erläuterungen zu den Beiträgen der zugehörigen Sitzung liefern. Die Beiträge befassen sich mit Sprachanwendungen im Bereich intelligenter Wohnumgebungen (Stichwort: „Ambient Assisted Living“) [8], dem Einsatz von Simulationsverfahren bei Hör- und Sehbehinderungen [13], mit Intonationsfragen bei schnell gesprochener Sprache, die bei Anwendungen in Vorleseautomaten für Blinde von Bedeutung sind [14], mit Kommunikationssystemen, die in der Lage sind, Gebärdensprache in gesprochene Sprache umzusetzen [16], [17] sowie mit allgemeinen Fragen einer barrierefreien sprachbasierten Kommunikation [20].

1 Einleitung

Elektronische Sprachverarbeitungssysteme haben heute bereits ein breites Anwendungsfeld gefunden. Beispiele für Anwendungen sind der Bürobereich, in dem Diktiersysteme mit sehr leistungsfähigen Spracherkennern (Erkennung von mehr als 1 Million Wortformen in Echtzeit) eingesetzt werden [1] sowie der Kraftfahrzeugbereich mit sprachgesteuerten Navigationssystemen und Freisprecheinrichtungen [3].

Der Einsatz der Sprachverarbeitung zur Unterstützung von Personen mit motorischen Behinderungen sowie mit Sinnesbehinderungen ist dagegen noch weitgehend beschränkt auf einige wenige Anwendungen wie z.B. sprachgesteuerte Rollstühle, Vorleseautomaten, Screenreader und Sprachspeichersysteme für Blinde.

Bemerkenswert sind andererseits aber die zahlreichen internationalen Forschungsaktivitäten für die genannten Personengruppen [2]. So gewinnen im Zuge des europäischen ICT-Programmes (ICT: Information and Communication Technologies) glücklicherweise auch Technologie-Entwicklungen an Bedeutung, die dem Gesundheits- und Rehabilitationsbereich zugeordnet sind [4]. In diesen Zusammenhang gehören auch die Aktivitäten des internationalen WEB-Consortiums W3C [5], das sich mit der barrierefreien Internetseiten-Gestaltung befasst und dabei auch ausdrücklich die sprachbasierte Webseitengestaltung für Blinde propagiert.

Schließlich untersucht auch die europäische Vereinigung COST 219 (COST 219: Accessibility for All to Services and Terminals for Next Generation Networks) intensiv den Einsatz der Sprachtechnologie [6], [7].

Die nachfolgenden Ausführungen sollen einen Eindruck von aktuellen Forschungsarbeiten auf dem Gebiet der Sprachtechnologie zur Unterstützung von Menschen mit Sinnesbehinderungen vermitteln. Im Vordergrund stehen dabei Personen mit Hör- und Sehbehinderungen, vor allem Gehörlose und Blinde.

2 Sprachverarbeitung und Ambient Assisted Living

Der Terminus „Ambient Assisted Living (AAL)“ zielt ab auf die „Umsetzung und den Einsatz innovativer technischer Hilfsmittel für Menschen mit Assistenzbedarf aufgrund Behinderungen, Erkrankungen, sozialen Schwierigkeiten und Pflegebedürftigkeit“ [8].

AAL ist damit thematisch eng verwandt mit den Begriffen „Smart Home“ [9], [10] und „Ambient Intelligence“ [11]. Die gemeinsame Grundidee ist die Unterstützung von Menschen und/oder die Verbesserung der Lebensqualität in ihrer häuslichen Umgebung durch moderne Technologie.

„Smart Home“, frei übersetzt „intelligentes Wohnen“ ist inzwischen sehr bekannt geworden und stellt bereits ein Qualitätskriterium für modernen Wohnkomfort dar. Im Haus sind die wichtigsten Geräte miteinander vernetzt und kommunizieren sowohl untereinander als auch mit dem Benutzer. Das sprachgesteuerte Fernseh- oder HiFi-Gerät und die sprechende Waschmaschine, die Video-Haustürüberwachung sowie eine Breitband-Vernetzung für interaktive Videodienste sind nur einige Anwendungsbeispiele für das intelligente Wohnen.

Die genannten Beispiele sind hauptsächlich „Komfort-Anwendungen“, deren Zielgruppe berufstätige Ehepaare mittleren Alters und Familien sind und daher ein lohnenswertes Käuferpotential darstellen. Es zeigt sich aber, dass Smart-Home-Systeme in zunehmendem Maße auch für Senioren und Menschen mit Behinderungen von großer Bedeutung sind; für diese Personengruppe erweisen sich oftmals Komfort-Anwendungen als wirkungsvolle Unterstützungssysteme [9].

Eine sehr ausführliche Darstellung zu diesem Thema findet man im SENTHA-Buch (SENTHA: Seniorengerechte Technik im häuslichen Alltag) [9] und in [10].

„Ambient Intelligence“ geht von der Technologie her noch weit über das Thema „Smart Home“ hinaus. Der Begriff steht für eine Umgebung, in der eine Vielzahl von elektronischen Komponenten und Systemen unsichtbar in Gegenständen und Einrichtungen der privaten Umgebung integriert sind. So kann der Mensch beispielsweise mit den Wänden seiner Wohnung, dem Mobiliar und Haushaltsgeräten kommunizieren; die genannten Dinge ihrerseits sind in der Lage, sich ad-hoc zu vernetzen, eine Eigenintelligenz zu entwickeln und dabei die Gewohnheiten, Vorlieben sowie auch Abneigungen der Bewohner zu erfassen und ähnlich wie ein aufmerksamer Butler zu reagieren [11]. In einer Ambient-Intelligence-Umgebung ist der Sprachdialog mit Alltagsgegenständen (Wänden, Bildern, Kaffeemaschine, Waschmaschine, Fernseher usw.) problemlos möglich. Da die benötigte Technologie miniaturisiert und/oder in den Gegenständen verborgen ist, „wirkt“ sie, ohne in Erscheinung zu treten. Näheres zum Thema „Ambient Intelligence“ findet man in dem Buch von *Aarts* [11].

Der Beitrag von *M. Hampicke* [12] wirft noch einmal einen Blick zurück auf die zahlreichen und wichtigen Erkenntnisse, die im Rahmen des SENTHA-Projektes – insbesondere für Senioren in ihrer häuslichen Umgebung – gewonnen werden konnten. Er gibt dann einen Überblick über aktuelle Techniken und zeigt auf, welche Anwendungen sich in der Zukunft für die Sprachtechnologie eröffnen.

3 Simulation von Hör- und Sehbehinderungen

Für Forscher und Entwickler, die Kommunikationssysteme für Menschen mit Sinnesbehinderungen entwickeln, ist es zumeist nicht leicht, sich in die Situation eines Betroffenen zu versetzen. Es besteht die Gefahr, dass die realisierten Systeme bezüglich ihrer Leistungsfähigkeit und/oder Akzeptanz deutlich hinter den Erwartungen dieser Personen zurückbleiben.

Üblicherweise geht man daher so vor, dass nach der Entwicklung der entsprechenden Geräte und Systeme umfangreiche Tests mit behinderten Versuchspersonen vorgenommen werden. Dabei besteht die Gefahr, dass entweder suboptimale Lösungen entstehen oder aber umfangreiche Adaptionen vorgenommen werden müssen.

Ein innovativer und erfolgversprechender Ansatz besteht nun darin, technische Systeme zu entwickeln, die zur Simulation von auditiven bzw. visuellen Behinderungen dienen. Damit hat man sich eine Referenz geschaffen, auf deren Basis sich angepasste Kommunikationshilfen entwickeln lassen.

In dem Beitrag von *H.H. Bothe* [13] wird dieses Thema diskutiert und ein System vorgestellt, das auditive und visuelle Behinderungen simulieren kann.

4 Untersuchungen zur Verstehbarkeit von schnell gesprochener Sprache in Abhängigkeit von der Intonationsstärke

In dem Beitrag von *D. Höpfner* [14] geht es um die häufig diskutierte Frage, ob schnell gesprochene Sprache dann eine höhere Verstehbarkeit aufweist, wenn die Prosodie möglichst flach ist. Die Thematik ist vor allem für Blinde interessant, die bei einem Vorleseautomaten zumeist die maximal mögliche Sprach-Ausgabegeschwindigkeit wählen. Hierbei müsste eine monotone Sprache die höchste Verstehens-Geschwindigkeit ergeben. Andererseits ist jedoch zu berücksichtigen, dass auch die Intonation maßgebliche Informationen zum gesprochenen Inhalt liefert [14], [15]. Letztlich fallen die Verstehbarkeits-Ergebnisse keinesfalls so

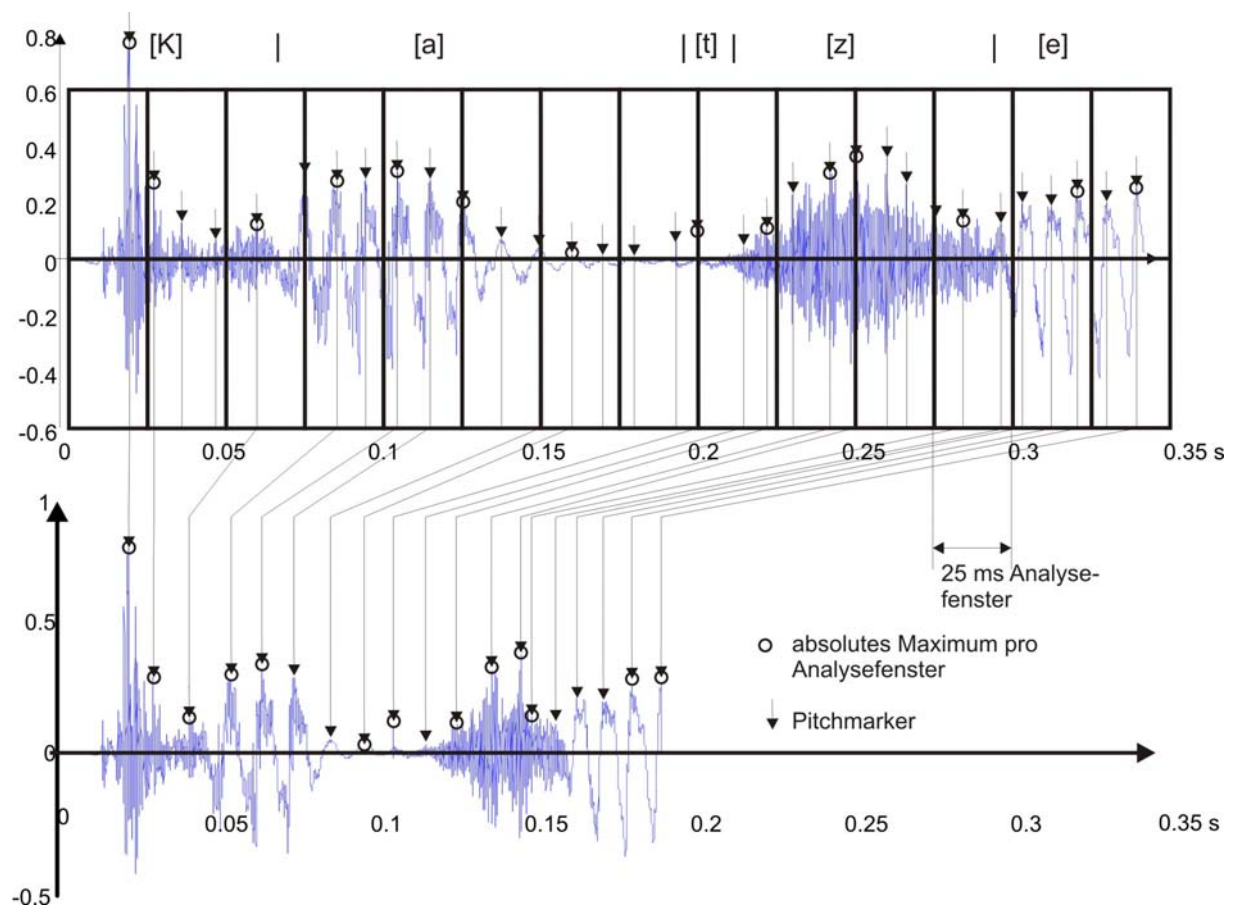


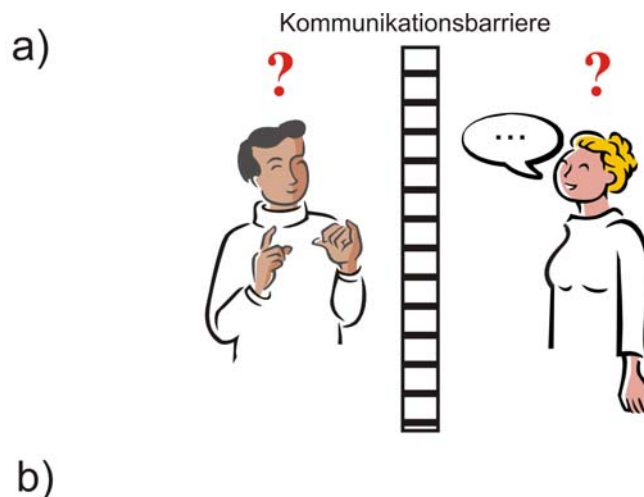
Abbildung 1 - Zeitliche Verkürzung unter weitgehender Beibehaltung der ursprünglichen Signalstruktur (nach Höpfner)

eindeutig wie vermutet aus; näheres hierzu wird ausführlich in dem Beitrag von *D. Höpfner* beschrieben. Dabei wird auch deutlich, dass eine zuverlässig arbeitende Grundfrequenz-Analyse eine besonders wichtige Rolle spielt, da sie die Basis für gezielte Intonationsveränderungen bildet.

Abb. 1 zeigt das Prinzip einer zeitlichen Verkürzung, bei der die ursprüngliche Signalstruktur weitgehend erhalten bleibt. Eine solche Verkürzung muss grundsätzlich lautabhängig und damit nichtlinear erfolgen; Vokale und Sprechpausen erfahren die stärksten Kürzungen, während beispielweise bei Plosivlauten keinerlei Verkürzungen vorgenommen werden.

5 Kommunikationshilfen für die Verständigung zwischen Gehörlosen und Normalhörenden

Die Kommunikation zwischen Gehörlosen und Normalhörenden erfolgt überwiegend schriftlich – z.B. über eMail oder SMS. Bei einer direkten (face-to-face-) Kommunikation ist die schriftliche Form jedoch völlig ungeeignet. Gehörlose verwenden daher zumeist die Gebärdensprache. Diese Form der Kommunikation ist in ihrer Ausdrucksfähigkeit durchaus der gesprochenen Sprache eben-bürtig. Allerdings hat die Gebärdensprache den gravierenden Nachteil, dass sie von den meisten Normalhörenden nicht „verstanden“ wird. Gehörlose bilden daher eine separate Gemeinschaft und sind in vielen Bereichen unserer hörenden Gesellschaft nicht integriert. Glücklicherweise können über eine „nonverbale“ Kommunikation (Gestik, Mimik) zumindest einige – vor allem emotionale – Informationen übermittelt werden, und viele Gehörlose sind auch darin geübt, Gesprochenes von den Lippen abzulesen.



Gebärdensprache (Beispiel):

„Gestern bin ich vom Bahnhof zum Hotel gegangen“ lautet in Gebärdensprache:

„GESTERN ICH BAHNHOF [Index A] GEHEN HOTEL [Index B]“

Anmerkung: „Bahnhof“ und „Hotel“ werden jeweils an den Stellen A und B im Gebärdenraum positioniert und dort ausgeführt. Das „Raumverb“ „gehen“ beginnt dann bei A und endet bei B.

Abbildung 2 - Kommunikation zwischen Gehörlosen und Normalhörenden. a) Kommunikationsbarriere, b) Transkriptionsproblematik

Es verbleibt aber dennoch eine „Kommunikationsbarriere“ (Abb. 2a), die jedoch durch den Einsatz der Sprachtechnologie, kombiniert mit Videoverarbeitungstechniken, überwunden werden kann. Der/die Normalhörende spricht, die Sprache wird mittels Spracherkennung analysiert und in Text umgesetzt. Der Text kann dann für den Gehörlosen auf einem Display dargestellt werden. Alternativ besteht auch die Möglichkeit, mit Hilfe eines Avatars das Ergebnis der Spracherkennung in Gebärdensprache zu „formulieren“.

Die für diese Lösung erforderliche Technologie ist heute vorhanden – allerdings ist die benötigte Echtzeit-Spracherkennung mit erheblichem Aufwand verbunden.

Anders als bei der eben beschriebenen Kommunikationsrichtung vom Normalhörenden zum Gehörlosen ist die umgekehrte Richtung (vom Gehörlosen zum Normalhörenden) noch weitgehend Gegenstand der Forschung. Der Kommunikationsverlauf spielt sich wie folgt ab. Wenn der Gehörlose kommuniziert, benutzt er die gewohnte Gebärdensprache; die Gebärden werden von einer Videokamera aufgenommen und einer automatischen Gebärdenerkennung (Bildererkennung) unterzogen. Als Ergebnis wird ein Text produziert, den der Normalhörende lesen kann, oder der Text wird mittels Sprachsynthese in Sprache umgewandelt.

Die Schwierigkeiten liegen zum einen in der automatischen Bewegtbild-Erkennung der Gebärden [16] bis [19] und zum anderen in der Transkription der Gebärdensprache in normalen orthografischen Text. Hierbei ist zu berücksichtigen, dass die Gebärdensprache eine eigene Syntax und Semantik aufweist [17], die teilweise gravierend von derjenigen der „normalen“ Orthografie abweicht (vgl. Abb. 2b). Hinzu kommt die Problematik der Echtzeitverarbeitung. Der Normalhörende muss die Text- oder Sprachinformation in Echtzeit erhalten, damit sie synchron zu der nonverbalen Kommunikation des Gehörlosen erfolgt. Hieraus resultieren erhebliche Anforderungen an die Verarbeitungsgeschwindigkeit der automatischen Bildererkennung, der Transkription und der Sprachsynthese.

Das Thema „Bildererkennung“ wird in dem Beitrag von *P. Dreuw* [16] diskutiert; die generellen Probleme der Transkription von Gebärdensprache in gesprochene Sprache ist Gegenstand des Beitrags von *J. Lindemann* [17].

6 Barrierefreies Informieren und Kommunizieren

Die bisherigen Beiträge behandelten spezielle Kommunikationsaspekte. In dem Vortrag von *C. Berndt* [20] wird das Thema „Kommunikation“ in größerer Breite und Allgemeinheit diskutiert. Dabei wird besonderer Wert auf die „Barrierefreiheit“ gelegt. Gerade im Bereich der digitalen Medien treten in der Praxis Probleme bei der Übertragung und Darstellung von Informationen auf, die Personen mit Sinnesbehinderungen in besonderem Maße benachteiligen. In dem beschriebenen Projekt geht es um die Entwicklung eines Systems, das Informationen aufnimmt, bearbeitet und anschließend benutzergerecht und damit barrierefrei ausgibt. Eine wichtige Rolle spielen hierbei Verfahren der Spracherkennung und der Sprachausgabe.

7 Schlussbemerkungen

Die präsentierten Beiträge zeigen sehr eindrucksvoll, dass die elektronische Sprachverarbeitung gerade für behinderte und ältere Menschen eine wertvolle Unterstützung bieten kann. Sie ist insbesondere in der Lage, Kommunikationsbarrieren abzubauen und die Betroffenen in die Gesellschaft zu integrieren. Schlüsselbegriffe sind „Barrierefreiheit“ und „Design für alle“. Spezial-Lösungen sind nur selten eine brauchbare Alternative, da sie zumeist mit

großem Aufwand und entsprechenden Kosten verbunden sind. Gleiches gilt auch für spezielle Dienste und Standards.

Die europäische Vereinigung COST219 [6], das Projekt SENTHA [9] sowie zahlreiche weitere Institutionen und Aktivitäten haben gezeigt, dass technische Entwicklungen und Standards, die von Beginn an auch die Bedürfnisse von behinderten und älteren Menschen mit berücksichtigen, für alle von Nutzen und letztlich – allein schon wegen der größeren Benutzerzahl – auch besonders kostengünstig sind.

Literatur

- [1]: Spracherkennungs- und Synthesysteme, Fa. Nuance, <http://www.nuance.de/>
- [2]: Fellbaum, K. (Guest editor). Electronic speech processing for persons with disabilities. Special Issue. *Technology and Disability*. IOS Press. August 2008.
- [3]: Jeschke, B.: Sprachausgaben von Sprachdialogsystemen im KFZ. ITG-Fachtagung „Sprachkommunikation“, 8.-10. Oktober 2008 Aachen.
- [4]: 7. EU-Rahmenprogramm, ftp://ftp.cordis.europa.eu/pub/fp7/ict/docs/ict-wp-2009-10_en.pdf
- [5]: W3C-Konsortium, <http://www.w3.org/>
- [6]: COST 219: siehe www.cost219.org
- [7]: Roe, P.R. (Ed.). Towards an inclusive future. COST 219ter publication Brussels 2007.
- [8]: Ambient Assisted Living. <http://www.ambient-assisted-living.org/index.php?id=188>
- [9]: Friesdorf, W.; Heine, A.(Hrsg.): SENTHA-seniorengerechte Technik im häuslichen Alltag. Springer-Verlag Berlin, Heidelberg 2007
- [10]: Willems, C.G.: Special Issue: Smart Homes. *Technology and Disability*, IOS Press, Vol. 15, Nr. 3 2003
- [11]: Aarts, E.; Marzano, S.: The New Everyday. Views on Ambient Intelligence. Koninklijke Philips Electronic N.V. 010 Publishers, Rotterdam, The Netherlands (2003)
- [12]: Hampicke, M.: Einsatz von Sprachanwendungen im AAL-Bereich-Rückblick, aktueller Stand und Zukunftsperspektiven. Beitrag in dieser Sitzung.
- [13]: Bothe, H.H.: Simulation audio-visueller Schädigungen für Studien an technischen Kommunikationshilfen. Beitrag in dieser Sitzung.
- [14]: Höpfner, D.: Verstehbarkeits-Untersuchung synthetisch geglätteter Intonationsstrukturen in zeitskalierter Sprache. Beitrag in dieser Sitzung.
- [15]: Höpfner, D.: Nichtlinearer Zeitskalierungsalgorithmus für gespeicherte natürliche Sprache. ITG Fachtagung „Sprachverarbeitung“ Aachen, 2008.
- [16]: Dreuw, P.: SignSpeak-Bridging the Gap Between Signers and Speakers. Beitrag in dieser Sitzung.
- [17]: Lindemann, J.: Zur Kommunikation zwischen Gehörlosen und Normalhörenden. Beitrag in dieser Sitzung.
- [18]: Brasher, H.; Starner, T.; Lukowicz, P. and Junker, H.: Using Multiple Sensors for Mobile Sign Language Recognition, *Proceedings of the Seventh IEEE International Symposium on Wearable Computers*, 2003, 45–52.
- [19]: Kraiss, K.F. (Ed.): *Advanced Man-Machine Interaction*. Springer-Verlag Berlin, Heidelberg, New York 2006.
- [20]: Berndt, C.: Barrierefreies Informieren und Kommunizieren. Beitrag in dieser Sitzung.