

»Betrevaugeh«

Mensch-Maschine-Kommunikation am Beispiel der Sprachverarbeitung

Joe Meier

Maus und Tastatur sind Vergangenheit. Wir sagen einfach alles nur noch unserem Computer und der setzt es dann um. Eine verlockende Vorstellung. Doch bis zu einem effektiven Plausch zwischen Mensch und Maschine ist es vor allem in der Arbeitswelt noch ein weiter und steiniger Weg. Ein Erfahrungsbericht.

»Mensch-Maschine-Kommunikation« ist zu einem Schlagwort geworden und spätestens seit der Einführung von »Siri«, dem Sprachassistenten für Apples iPhone, in aller Munde.¹

Ein Sprachdialogsystem erlaubt dem Benutzer Eingaben in seinen eigenen Worten durchzuführen. Diese können je nach System auch als Text angezeigt und abgespeichert werden. Die Antwort erfolgt durch (fast) natürlich klingende Stimmen. Spracherkennung und Sprachausgabe stehen in einem engen Zusammenhang.

Allerdings erscheint mir der Begriff »Kommunikation« in diesem Zusammenhang zu hoch gegriffen. Ich würde mehr von einer »Interaktion« zwischen Mensch und programmierter Maschine sprechen, da die Einbeziehung und das Zusammenwirken aller Sinnesorgane sowie Gesten und Mimik fehlen. All dies ist aber für eine dem menschlichen Verhalten angepasste Kommunikation wichtig.²

Die Interaktion zwischen Mensch und Maschine hat in den letzten Jahrzehnten Fortschritte gemacht und führte von Geräten der Ein-/Ausgabe (Maus, Tastatur oder Bildschirm) über Handschrift- und Gesichtserkennung bis hin zur Sprachverarbeitung (Spracherkennung und Sprachausgabe, auch Sprachsynthese genannt). Eine gute Einführung in die The-

matik »Mensch-Maschine-Kommunikation« finden Sie unter:

» www.techfak.uni-bielefeld.de/~skopp/download/Kopp_MMK-Intro.pdf

Über den Entwicklungsstand der Spracherkennung als Teil der Sprachverarbeitung wurde in dieser Zeitschrift bereits vor einiger Zeit einmal ausführlich berichtet.³ Die damals gewonnenen Erkenntnisse sind auch heute noch gültig. Eine nochmalige Lektüre empfiehlt sich zur Einführung in die Thematik. Die Entwicklungen der letzten zwei Jahre gehen nunmehr eher in Richtung einer »sprecherunabhängigen« Spracherkennung, die bisher nur einen begrenzten Wortschatz verarbeiten konnte. Allerdings sind diese Lösungen schon aus Kostengründen für Privatpersonen noch (fast) unerschwinglich. Ein Besuch auf der folgenden Webseite lohnt sich aber auf jeden Fall, um sich über die Trends zu informieren:

» www.voicepro.de/demo/

Ausführliche Informationen und Demonstrationen professioneller Spracherkennung sind hier zu finden:

» www.nuance.de/index.htm

» www.linguattec.net/

» www.spracherkennung.de/demo/

» www.linguattec.net/products/stt/voice_pro

Sprachverarbeitung mit Windows 7

Zum praktischen Einstieg in die Materie empfiehlt sich die Nutzung der Windows eigenen Bordmittel. Windows 7 bietet im Menü *Zubehör* → *Erleichterte Bedienung* eine einfache Sprachverarbeitung an, die sich zum Kennenlernen eignet, aber nicht professionellen Ansprüchen genügt.



Die Sprachverarbeitungsfunktionen in Windows 7 wählen Sie unter *Zubehör* → *Erleichterte Bedienung* aus.

- 1 Mehr Informationen zu Siri unter www.apple.com/de/ios/siri/
- 2 Ausführlich dazu Cakir, Spracherkennung – gestern, heute und übermorgen, in: CuA 6/2012, 5 ff.
- 3 Siehe den Schwerpunkt »Automatische Spracherkennung vor dem Durchbruch (?)«, in: CuA 6/2012, 5 ff.



Das Sprachlernprogramm ist obligatorisch, um die Spracherkennung unter Windows 7 nutzen zu können (rechts klicken auf das Mikrofon-Symbol nach Öffnung der Spracherkennung).

Weitere Informationen und Anleitungen zur Sprachverarbeitung mit Windows finden Sie unter:

- » <http://windows.microsoft.com/de-de/windows/set-speech-recognition#1TC=windows-7>
- » http://praxistipps.chip.de/windows-7-spracherkennung-nutzen-so-gehts_24049

Sprachverarbeitung ist natürlich eine tolle Errungenschaft für all diejenigen, die insbesondere aus gesundheitlichen Gründen zu stark eingeschränkt sind, um mit herkömmlichen Eingabegeräten zu arbeiten. Wer aber noch gesund ist und flüssig mit einer Tastatur umgehen kann, wird das Tippen wohl nicht missen wollen – es ist einfach (jetzt noch) effizienter.

Aufgrund des technologischen Fortschritts in der Sprachverarbeitung erscheint mir der Gedanke nicht abwegig, dass in nicht allzu langer Zeit Beschäftigte bei Arbeitsbeginn feststellen werden, dass Tastatur und Maus einfach verschwunden sind.

Auch wenn der Übergang nicht so extrem ausfallen dürfte, ist eine gesunde Portion Vorsicht gegenüber allein technisch inspirierten Neuerungen geboten. Sie wurden schon zu oft übereilt den Mitarbeitern aufgezwungen – ohne die Tragweite der Umstellungen und die zusätzlichen Belastungen am Arbeitsplatz zu bedenken. Belegschaftsvertreter sollten also auf der Hut sein, wenn der Arbeitgeber die Einführung von Spracherkennungs-Software ankündigt.

Im Folgenden befasse ich mich vor allem mit der Sprachsynthese, also der

Überführung von geschriebenem Text in das gesprochene Wort. Sie hat mehr Einsatzmöglichkeiten als die Spracherkennung und wir finden sie überall: beim Navi im Auto, bei Durchsagen im Supermarkt, auf Bahnhöfen und in öffentlichen Verkehrsmitteln – täglich kommen neue Anwendungen hinzu. Zur ersten Orientierung dient auch der folgende Link:

- » <http://de.wikipedia.org/wiki/Sprachsynthese>

Sprachausgabe mit Balabolka

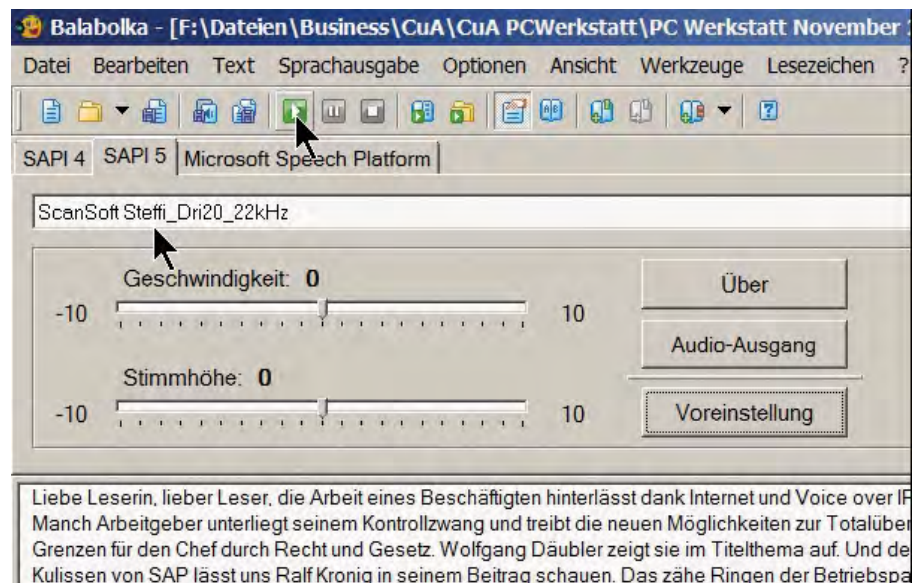
Um einen guten Einstieg in die Möglichkeiten der Sprachsynthese zu bekommen, genügt auch schon ein Freeware-Produkt wie Balabolka. Kommerzielle Produkte der Sprachausgabe – wie der Linguatéc Voice Reader Studio 15, auf dessen Möglichkeiten ich später noch im Detail eingehen werde – sind teuer (circa 500 Euro). Sie kommen also praktisch nur für Unternehmen in Frage. Der Preis wird überwiegend von der Qualität der eingesetzten

- » www.cross-plus-a.com/de/balabolka.htm

Installieren Sie bitte auch die deutsche Stimme »Steffi«, die ebenfalls auf der Balabolka-Homepage oder direkt hier zum Herunterladen zur Verfügung steht:

- » <http://dfiles.eu/files/6212458>

Die mit Balabolka erstellte Audiodatei im MP3-Format (Dateiname: Beispieltext Steffi normal.mp3) auf meiner Website klingt eigentlich ganz passabel. Vergleichen Sie dies aber mit einer professionelleren Stimme, wie Petra von Linguatéc, so werden Sie sofort den Qualitätsunterschied feststellen (Dateiname: Beispieltext Petra 95.mp3). Es besteht aber auch die Möglichkeit, professionelle Stimmen käuflich zu erwerben und diese dann mit Balabolka zu nutzen. Sie finden auf der Homepage von Balabolka die wesentlichen kommerziellen Anbieter mit Demonstrationen ihrer Produkte. Bedauerlicherweise gibt es für Balabolka (noch)



Wählen Sie – falls nicht schon als Vorgabe eingestellt – je nach Sprache eine der installierten Stimmen aus (»Steffi« ist beispielsweise für die deutsche Textwiedergabe geeignet) und klicken Sie dann auf die »Start«-Schaltfläche um sich den Text vorlesen zu lassen.

Stimmen (TTS = »Text to Speech«) bestimmt. Hier gibt es große Unterschiede wie Sie auch den Hörproben auf meiner Website entnehmen können:

- » www.joemeier.net/sprachverarbeitung

Die Software Balabolka (russisch für »Schwätzer«) gibt es kostenlos zum Download:

kein Benutzerhandbuch – die Hilfefunktionen im Programm sind etwas dürftig ausgefallen. Hier hilft ein Erkunden der Menüstruktur weiter. Sie finden unter anderem Funktionen wie das Speichern des gesprochenen Textes in gängigen Audioformaten wie MP3 und das Einfügen fremdsprachlicher Wörter. Auch Benutzerwörterbücher lassen sich anlegen.

Bund.2

Fremdsprachliche Wörter werden meist richtig wiedergegeben, wenn sie sich auch in der deutschen Sprache schon eingebürgert haben, wie etwa »Features«. Abkürzungen machen aber immer wieder zu schaffen. Ein gutes Beispiel hierfür ist IT (Informationstechnologie). Wenn Sie diese Abkürzung nicht als fremdsprachliches Wort definiert haben, wird nur »it« oder »i-t« anstatt »ei-tii« ausgegeben. Es ist aber in Balabolka möglich, fremdsprachlichen Text (auch mit einer anderen der vorinstallierten Sprachen) durch ein sogenanntes Markup einzufügen.

Für Geschäftskunden mag es allerdings notwendig sein sich schon heute auf die zukünftigen Trends in ihrer `<voice required= Name=ScanSoft Emily_Dn40_16kHz>IT (Information Technology)</voice>` Landschaft einzustellen und sich dementsprechend vorzubereiten.

Auch innerhalb eines deutschen Textes können fremdsprachliche Begriffe korrekt ausgesprochen werden.

Balabolka eignet sich sehr gut, um sich mit den Grundlagen der Sprachsynthese vertraut zu machen. Vorteil ist die kostenlose Bereitstellung mehrerer Weltsprachen, die auch kombiniert angewendet werden können. Zusätzlich professionelle Sprachen und fachspezifische Wörterbücher können käuflich erworben werden.

Sprachausgabe mit Linguattec Voice Reader Studio 15

Bei der kommerziell vertriebenen Software »Voice Reader Studio 15« von Linguattec handelt es sich um ein Produkt für den professionellen Einsatz im Unternehmen:

» www.linguattec.net/products/tts/voice_reader/vrs15

Dort finden Sie auch Online-Demos der Sprachsynthese mit dem Linguattec Voice

Reader Studio 15, die sich anzusehen lohnen:

» www.linguattec.net/products/tts/usecases/vrs

» www.linguattec.net/products/tts/voice_reader/vrs15onlinedemo

» www.linguattec.net/products/tts/voice_reader/vrs15demo

Bitte berücksichtigen Sie, dass es sich um kurze Online-Demos handelt und wegen des Datentransfers die Qualität nicht unbedingt professionellen Ansprüchen genügt, die Sie beim käuflich erworbenen

Produkt erwarten können. Ich verweise auch auf die mit dem Produkt von mir selbst erstellten Beispieltex-te auf meiner Website.

Das ausführliche Handbuch für die Software finden Sie dort ebenfalls (Dateiname: Handbuch Voice Reader.pdf), damit man sich mit den gebotenen Funktionen vertraut machen kann (siehe auch die Abbildung unten).

Die Menüführung von Voice Reader orientiert sich wie die neueren Microsoft Office-Programme am sogenannten Menüband, das 2007 eingeführt wurde.

Das Programm Voice Reader ist für viele Sprachen erhältlich. Fremdsprachliche Begriffe können beim Verwenden einer ML (Multi-Lingual) Sprache wie »Anna-ML, Weiblich« durch das Einfügen von sogenannten Voice Markups korrekt wiedergegeben werden (siehe dazu die Abbildung auf Seite 13 oben).



Voice Reader Studio 15 ist für den professionellen Einsatz gedacht und verfügt über zahlreiche Funktionen, die sich in einem Freeware Produkt bisher nicht realisieren lassen.



Vergleichen Sie dazu auch die Sprachausgabe der Abkürzung »IT« in den folgenden Beispieldateien auf meiner Website:

» **Beispieltext Petra IT 1.mp3**

(IT wie »it« ausgesprochen)

» **Beispieltext Anna ML IT.mp3**

(IT korrekt wie »ei-tii« ausgesprochen)

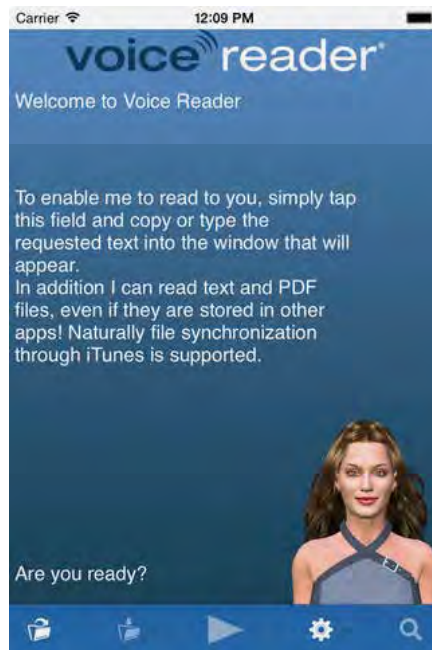
Abkürzungen stellen nach meinen Erfahrungen immer eine Herausforderung dar. Es empfiehlt sich, möglichst darauf bei den zu vertonenden Texten möglichst zu verzichten. »IT« kann auch mit »Informationstechnologie« oder unter Verwendung eines Markups in englischer Sprache mit »Information Technology« umschrieben werden. Ein weiteres Beispiel wäre die gängige Abkürzung »BetrVG« für das Betriebsverfassungsgesetz, die in so etwas wie »Betrevauegeh« ausartet. Es gibt allerdings für diesen Einsatz auch auf die juristischen Belange zugeschnittene Wörterbücher und Stimmen.⁴

Textnormierungen können ebenfalls vom Menü *Vorlesen* → *Voice Markup* eingefügt werden, etwa ein korrekt gesprochenes Datum, die Uhrzeit oder eine Adresse in der richtigen Formatierung. Die Redewendungen in der Abbildung unten auf dieser Seite finden Sie als MP3-Datei auf meiner Site (Dateiname: Petra Textnormierung.mp3). Die heutigen Möglichkeiten der Sprachsynthese haben aber noch ihre Grenzen, da sie nicht ganz dem natürlichen menschlichen Verhalten entsprechen.

Korrekturaufwand nicht zu unterschätzen

Betrachten Sie zum Beispiel einmal den folgenden Satz: »Renn, renn – der Zug fährt gleich ab!«

Der erste Teil des Satzes »Renn, renn« wird vom Menschen bewusst schneller



Und so präsentiert sich VoiceReader auf dem Smartphone

gesprochen, um den Gesprächspartner zur Eile zu ermahnen, der zweite Teil dann im normalen Tonfall. Dies lässt sich noch nicht zufriedenstellend mit den

heute zur Verfügung stehenden Mitteln der Sprachsynthese realisieren.

Wichtig ist beim Einsatz von Spracherkennung und Sprachausgabe auch die Korrektur möglicher Eingabe-/Ausgabefehler. Bei der Spracherkennung kann der Text am Bildschirm vor der Druckausgabe mit einem Textverarbeitungsprogramm verbessert werden. Bei der Sprachsynthese lässt sich das Resultat – sollte es nicht den Anforderungen genügen – nur durch ein Überarbeiten der zugrunde liegenden Texte und einer Neuerstellung der Audio-Datei verbessern. Der (Zeit-) Aufwand hierfür ist meinen Erfahrungen nach sehr hoch.

Von einem sinnvollen und vollwertigen Sprachdialog im Sinne menschlicher Kommunikation sind wir aber noch weit entfernt. Das Video: »Computer of the future – Apple Knowledge Navigator« ist ein bisschen Zukunftsmusik. Ich empfehle, es sich zum tieferen Verständnis aber einmal anzusehen – Sie finden es hier:

» www.youtube.com/watch?v=9bjve67p33E

Autor

Joe Meier ist IKT-Journalist und PC-Sachverständiger. Für Rückfragen zu den besprochenen Programmen steht er CuA-Lesern gerne zur Verfügung:

» joemeier@consultant.com

» www.joemeier.net

Wir haben versucht Sie am <Tn=date/>20.11.2014 telefonisch um <Tn=time/>14 Uhr<Tn=normal/> zu erreichen. Hat sich Ihre nachstehende Adresse geändert?
<Tn=address/>
Hans Müller
Bahnstrasse 12
65824 Schwalbach am Taunus
<Tn=normal/>
Falls ja, geben Sie uns bitte Bescheid. Vielen Dank.

Mit Textnormierungen können Sie eine korrekte Wiedergabe von Datum, Uhrzeit und anderen Einfügungen bei der Sprachausgabe erreichen.