Spracherkennung per Computer funktioniert

R. J. Lellé

Einleitung

Es steht außer Frage, dass es bei der ärztlichen Kommunikation, zum Beispiel beim zeitnahen Schreiben von Arztbriefen, Befundmitteilungen oder Operationsberichten, noch gewaltige Defizite gibt. Niedergelassene Kollegen/innen beklagen regelmäßig die schleppende Korrespondenz der Krankenhäuser.

In der Regel sind es die Assistenzärzte, die für diese Korrespondenz verantwortlich sind. Wer kennt nicht die Situation, dass der Chefarzt, alarmiert durch die Briefstapel im Arztzimmer, seine Oberärzte anhält, das Diktat der Arztbriefe nötigenfalls durch disziplinarische Mittel zu beschleunigen (klassisch: das »Operationsverbot«). Die nächste Hürde, die es zu überwinden gilt, bevor die Briefe abgeschickt werden, ist ein Mangel an Schreibkräften, sei es wegen Grippewelle, Urlaubszeit oder Personaleinsparungen. Nicht selten müssen Ärzte auch selbst – im Zweifingersystem – in die Tasten greifen.

Wie wäre es, wenn die inzwischen in praktisch jedem Krankenhaus ausreichend zur Verfügung stehenden Computer dazu benutzt werden, das diktierte Wort gleich in geschriebenen Text umzusetzen? In der Tat ist die Entwicklung von Hard- und Software inzwischen so weit gediehen, dass dies möglich ist.

Nachfolgend werden die persönlichen Erfahrungen mit einem Spracherkennungssystem dargestellt, nachdem über einen Zeitraum von mittlerweile 16 Monaten die gesamte Korrespondenz mittels Computer diktiert und geschrieben wurde. Bei der elektronischen Spracherkennung unterschätzt man leicht den sehr großen technischen Aufwand solcher Systeme. Eine Auseinandersetzung mit dem Aufbau und der Denkweise eines solchen komplizierten Systems lohnt sich.

Die erforderliche Soft- und Hardware

Bei der Software handelt es sich um Dragon NaturallySpeaking Version 7.1 Professional von Scansoft. In Zusammenarbeit mit der Firma OfficeAutomation Sander (www.oa-sa.de) wurde ein Spezialvokabular für die Gynäkologie ergänzt.

Die Software wurde auf zwei verschiedenen Laptop-Computern mit Prozessoren von 800 MHz beziehungsweise 1 GHz eingesetzt sowie auf einem Desktop-Computer mit einem 1-GHz-Prozessor. Hierbei zeigte sich, dass die Größe des Arbeitsspeichers für die Geschwindigkeit der Spracherkennung von entscheidender Bedeutung ist. Alle drei verwendeten Computer hatten hierbei einen Speicher von (nur) 256 MB. Sicherlich wäre ein höherer RAM von 512 MB oder höher wünschenswert.

Auch wenn die Geschwindigkeit des Prozessors nicht das entscheidende Kriterium ist für die Effektivität der Spracherkennung, so fand sich dennoch ein deutlicher Unterschied zwischen dem 800-MHz-Gerät und den beiden 1-GHz-Computern. Vor allem, wenn der gesprochene Text noch während des Diktats am Bildschirm mitbeobachtet und sofort korrigiert werden soll, arbeitet ein 800-MHz-Prozessor zu langsam.

Abb. 1: Die Spracherkennungssoftware arbeitet auch im Operationssaal trotz Mundschutz und Hintergrundgeräuschen. Hier wurde ein Tupfer über das Mikrofon gestülpt, um das Atemgeräusch zu dämpfen

Nach Versuchen mit unterschiedlichen Mikrofonen erwies sich ein USB-Headset (Labtec LVA-8711) als vorteilhaft. Da das Mikrofon über die USB-Schnittstelle am Computer angeschlossen wird, ist die Spracherkennung nicht von der Qualität der im Computer eingebauten Soundkarte abhängig.

Das Mikrofon nimmt hierbei nur die Geräusche in der unmittelbaren Umgebung auf. Selbst unter ungünstigen Bedingungen (Operationssaal mit Mundschutz und Hintergrundgeräuschen) lässt sich eine befriedigende Erkennungsgenauigkeit erreichen (Abb. 1).

Glücklicherweise sind die Zeiten vorbei, in denen man einem Spracherkennungsprogramm stundenlang Texte vorlesen musste, damit sich das System an die individuelle Sprechweise anpassen konnte. In der Tat dauert diese Übungsphase nur noch wenige Minuten. Ebenso wie bei einem normalen Diktat wird von dem Sprecher lediglich gefordert, dass er deutlich – aber nicht

überdeutlich – spricht und dies in normaler Sprechgeschwindigkeit. Die Interpunktion muss natürlich mitdiktiert werden.

Wie arbeitet und »denkt« der Computer bei der Spracherkennung?

Der Computer versucht die natürliche Spracherkennung des Menschen nachzuahmen. Dies ist weitaus komplizierter als man zunächst glauben möchte. Das System arbeitet in Echtzeit, das heißt die Frequenzmuster müssen sofort berechnet und ausgewertet werden, sodass der gesprochene Text unmittelbar am Computer angezeigt werden kann. Es handelt sich um eine kontinuierliche Spracherkennung. Der Sprecher muss demnach keine Pausen zwischen den einzelnen Wörtern einlegen. Für die Software ist es jedenfalls eine große Herausforderung zu erkennen, wann ein Wort zu Ende ist und wann ein neues Wort beginnt.

Dem Computer steht zunächst nur ein Frequenzdiagramm zur Verfügung. Abbildung 2 zeigt ein Beispiel für ein solches Diagramm für den Ausdruck »die Fossa obturatoria«, einmal diskontinuierlich und einmal kontinuierlich gesprochen.

Vereinfacht ausgedrückt wird das Frequenzmuster in so genannte »Phoneme« zerlegt. Ein Phonem ist eine lautsprachliche Einheit und hat bei normaler Sprechgeschwindigkeit eine Dauer von 0,01 bis 0,04 Sekunden. Aus mehreren Phonemen wird ein Muster errechnet, welches mit den Referenzmustern, die in einer Datenbank in großer Zahl zur Verfügung stehen und die quasi als Schablone dienen, verglichen. Dieser Vorgang wird durch weitere mathematische Modelle sowie Methoden der künstlichen Intelligenz ergänzt.

Die rechnerisch aufwändige Analyse von Frequenzmustern und deren Umsetzung in geschriebenen Text macht es erforderlich, dass der Computer jedes diktierte Wort bereits »kennt«, das

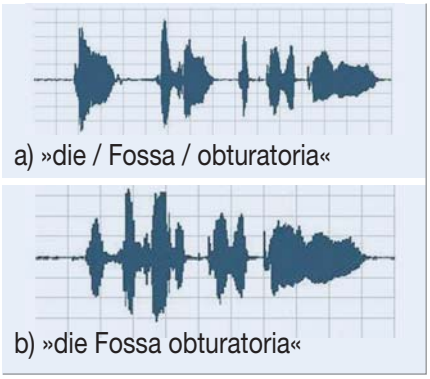

Abb. 2a und b: Frequenzmuster des Satzbausteins »die Fossa obturatoria« bei a) diskontinuierlicher und b) kontinuierlicher Sprechweise

heißt, dass das Wort im Vokabular des Programms bereits abgespeichert ist.

Dichterische Wortschöpfungen können demzufolge nicht erkannt werden. Hierzu ein Experiment mit einem Auszug aus dem Gedicht »Morgendlicher Rosenstrauß« von *Arno Holz* (1863–1929), welches von der dichterischen Freiheit der Sprache ausgiebigen Gebrauch macht:

Wie wunderbar:
Aus tiefsattem, köstlichstem,
noch taublätterigem, noch tauleuchttropfigem, noch tauglitzerigem
Dunkelglanzgrün
flimmernd, schimmernd, glimmernd,
mitten im
schattenkühlen, ebenerdigen,
weinrebenumkletterhangenen
Gartenhausraum,
Rosen!

Diktiert man das Gedicht der Spracherkennungssoftware, erhält man das folgende Ergebnis:

In 00 Mark.
Alles tiefe Saddam, köstlich dem,
noch Pterygium, wird, pflegen,
welcher Teilblitzerregern
dunkle Glanzgrünen,
flimmern, schimmernd, klingen,
Mieten eben
Schatten kühlen, eben eher
die dem, eigene Leben um Kette
handelnden
Gartenhaus waren,
Russen ...

Dieses groteske Ergebnis kommt dadurch zustande, dass der Dichter zahlreiche Wortneuschöpfungen verwendet, die nicht im Vokabular des Computers vorhanden sein können. Das Spracherkennungsprogramm bietet immer das Wort an, welches nach der Analyse des Frequenzmusters mathematisch am wahrscheinlichsten ist. Die Tatsache, dass das Wort »taublätterigem« mit »Pterygium« assoziiert wird, gibt einen Hinweis darauf, dass die verwendete Spracherkennungssoftware aus einem medizinischen Spezialwortschatz schöpft.

Diese Eigenheiten der elektronischen Spracherkennung kann in der täglichen Praxis zu unerfreulichen Fehlern führen, die man bei flüchtiger Korrektur des Textes unter Umständen nicht erkennt. So kann aus »Jodprobe« (bei der Kolposkopie) »Idiotprobe« werden.

In dem Gedicht enthält die Zeile »flimmernd, schimmernd, glimmernd« drei ähnlich klingende Wörter, die unmittelbar aufeinander folgen ohne weitere Wortzusammenhänge. Noch problematischer für den Computer sind »Homophone«. Hierunter versteht man zwei Ausdrücke, die identisch ausgesprochen aber unterschiedlich geschrieben werden müssen. Hierzu ein Beispiel: Wie der Ausdruck »mit Hilfe« geschrieben wird, lässt sich in den folgenden Sätzen nur aus dem Zusammenhang erkennen.

»Ich wäre Ihnen für Ihre Mithilfe bei der Behandlung der Patientin dankbar.«

»Die Präparation erfolgte mit Hilfe des Skalpells.«

Um mit dieser Eigenschaft der Sprache zurechtzukommen, bedient sich das Programm der so genannten »Bigramm-Statistik«. Im vorprogrammierten Vokabular findet sich neben Einzelwörtern auch eine große Zahl von Wortkombinationen. Während des Vorgangs der kontinuierlichen Spracherkennung berücksichtigt das

Programm die Wahrscheinlichkeit, mit der bestimmte Wortkombinationen vorkommen. Gleichzeitig lernt das Programm bei jedem neu diktierten Text, welche Wortfolge der Sprecher bevorzugt.

Ohne diese aufwändige Rechenleistung, die permanent im Hintergrund abläuft, wäre es nicht möglich, mit den zahlreichen grammatikalischen Feinheiten, insbesondere den unterschiedlichen Wortendungen in der deutschen Sprache, fertig zu werden. Man betrachte zum Beispiel das Wort »schwer« in den folgenden beiden Sätzen.

»Es handelt sich um eine schwere Zervixdysplasie.«

»Bei einer schweren Zervixdysplasie wird in der Regel eine Konisation durchgeführt.«

Nur aus dem Zusammenhang des Satzes lässt sich schließen, ob es »schwere« oder »schweren« heißen muss.

Ein weiteres Problem ist die Erkennung von Groß- und Kleinschreibung wie zum Beispiel bei dem Wort »sie«, bei dem es sich um die persönliche Anrede »Sie« handeln könnte. Dass das Programm dies trotzdem meistens richtig schreibt – jedoch keinesfalls immer –, hängt ebenfalls mit der statistischen Analyse der Wortzusammenhänge zusammen.

Diktat und Korrektur

Erkennungsfehler kommen praktisch in jedem Text vor. Im medizinischen Umfeld können solche sinnentstellenden Fehler gefährliche Konsequenzen haben. So könnte an Stelle des Wortes »kein« fälschlicherweise das Wort »ein« stehen oder umgekehrt. Die nachträgliche Kontrolle und Korrektur des vom Computer geschriebenen Textes ist deshalb besonders wichtig.

Beim Diktat kann man einen Text kontinuierlich wie auf ein Tonband sprechen. Die Korrektur des Textes wird dann entweder unmittelbar nach dem Diktat vorgenommen oder aber zu einem späteren Zeitpunkt. Denn zusätzlich zu der Textinformation wird jedem einzelnen Wort das akustische Diktat unterlegt und – wenn gewünscht – abgespeichert. Demnach ist es möglich, sich später das gesamte Diktat, Auszüge hiervon oder auch nur einzelne Wörter durch Anklicken mit dem Mauszeiger nochmals vorlesen zu lassen.

Der Vorgang des Korrigierens dient nicht ausschließlich dazu, den Text zu berichtigen. Immer dann, wenn das Korrekturfenster des Programms geöffnet wird, lernt das Programm neue Wörter und insbesondere auch Wortzusammenhänge, so genannte »Phrasen«. So muss der Eigenname einer Person nur einmal korrigiert werden und wird dann mit sehr hoher Wahrscheinlichkeit beim nächsten Mal richtig erkannt werden.

Ein ungewöhnlicher Name, zum Beispiel *»Frau Dr. Hardjolukito«,* der in der deutschen Sprache exotisch wirkt, eignet sich besonders gut für die Spracherkennungssoftware, während die verschiedenen Schreibweisen von *»Meyer, Maier, Meier, Mayer«* für den Computer ein Problem darstellen.

Wird allerdings zum Beispiel der Ausdruck *»Elisabeth Maier«* als Phrase abgespeichert und in dieser Form diktiert, so kann der Computer die Schreibweise von *»Maier«* korrekt zuordnen.

Die Korrektur falsch erkannter Wörter oder Satzpassagen ist nicht die einzige Möglichkeit, um die Erkennungsgenauigkeit im Laufe der Zeit zu verbessern. Indem ein bereits diktierter und korrigierter Text nochmals – und am besten mehrmals – in das Programm »gefüttert« wird, werden nicht nur neue Wörter in das Programmvokabular aufgenommen, sondern auch die Statistik über Wortzusammenhänge den Gewohnheiten des Sprechers angepasst.

Die Rolle der Sekretärin

In diesem Zusammenhang ein Wort zur Rolle der Sekretärin in einer Klinik, in der der Computer die Umsetzung von Sprache in Text übernommen hat. Mancher wird sich noch daran erinnern, wie in der zweiten Hälfte der achtziger Jahre des vorigen Jahrhunderts Computer und Drucker die Schreibmaschinen abgelöst haben. Damals wie heute werden diese Entwicklungen von Sekretärinnen und Schreibkräften sowie dem Personalrat mit Sorge beobachtet. Wer sich jedoch näher mit Spracherkennungssoftware beschäftigt, wird feststellen, dass er auf jeden Fall auf die Hilfe seiner Sekretärin angewiesen ist, falls er nicht alle Korrekturen selbst vornehmen möchte. So kann zum Beispiel der Abgleich zwischen akustischer Aufzeichnung und geschriebenem Text der Sekretärin übertragen werden, ebenso wie das Formatieren des Textes sowie das Drucken und Versenden der Briefe.

An Stelle des eintönigen Eintippens der Diktattexte in den Computer treten durch die gewonnene Zeit andere abwechslungsreichere und wichtigere Aufgaben. Eine Zunahme der ärztlichen Korrespondenz und vor allem eine zeitnahe Weitergabe von Informationen ist im Interesse aller, insbesondere des Patienten. Dies gilt auch für die interne Dokumentation in Klinik und Praxis. Dann würden abgegriffene handschriftliche und schwer zu entziffernde Karteikarten in der Arztpraxis endlich der Vergangenheit angehören.

Fazit

Die eigenen Erfahrungen haben gezeigt, dass Spracherkennung per Computer funktioniert. Nach kurzer Einübungszeit wird eine hohe Erkennungsgenauigkeit erreicht, vorausgesetzt dass ein Spezialvokabular in das Programm integriert ist. Bestimmte Eigenheiten der elektronischen Spracherkennung müssen berücksichtigt werden. Ganz wichtig ist eine systematische Korrektur der diktierten Texte,

um sinnentstellende Fehler zu vermeiden.

Keinesfalls verdient die elektronische Spracherkennung die Bezeichnung »elektronische Sekretärin«. Vielmehr werden die Sekretärinnen sinnvoll entlastet und die Aufgaben der Schreibkräfte interessanter und anspruchsvoller. Die Korrespondenz kann ausgeweitet werden und zeitnaher erfolgen.

Anschrift des Verfassers:
Univ.-Prof. Dr. Ralph J. Lellé, MIAC
Frauenklinik
Universitätsklinikum Münster
Albert-Schweitzer-Straße 33
48149 Münster
E-Mail info@lellenet.de