Automatische internetbasierte Evaluation der Verständlichkeit

Maier, Haderlein, Hacker, Nöth, Rosanowski, Eysholdt, Schuster

Einleitung: Bisher existierten kaum objektive, untersucherunabhängige Bewertungsinstrumente für die Verständlichkeit von gesprochener Sprache. In der aktuellen Forschung bei Stimmstörungen und Sprechstörungen hat es sich gezeigt, dass in der Medizin für diesen Zweck automatische Spracherkennungssysteme anwendbar sind. Als Maß für die Verständlichkeit dient dabei die sogenannte Worterkennungsrate (WR) bzw. die Wortakkuratheit (WA), welche den prozentualen Anteil richtig erkannter Wörter angibt.

Für eine Anwendung an verschiedenen Untersuchungsplätzen wurde hierzu nun eine zentralisierte Auswertung durch eine Client-Server-basierte Software realisiert, um die schnelle Auswertung am PC über das Internet zu ermöglichen.

Material: Diese neue Methodik wurde bisher an insgesamt 35 Kindern und Jugendlichen mit Lippen-Kiefer-Gaumenspalte im Alter zwischen 3,3 und 18,5 Jahren (Mittelwert 8,3 ± 3,6 Jahre, Median 7,7 Jahre), davon 13 Mädchen und 22 Jungen, überprüft. Ihnen wurden Bilder am PC gezeigt, die Sie benennen sollten. Die Bildtafeln stellen eine digitale Version des PLAKSS-Tests [1] dar, der zur Erhebung des Lautbestandes bei Kindern dient. Dieser Test enthält alle deutschen Phoneme an verschiedenen Positionen im Wort (Anfang, Mitte und Ende). Die Untersuchung wurde im Rahmen der interdisziplinären Spaltsprechstunde des Universitätsklinikums Erlangen mit einem Head-Set aufgezeichnet und digital gespeichert.

Methode: Als Maß für die Verständlichkeit wurde bisher die Wortakkuratheit herangezogen. Eigentlich dient diese als Gütekriterium für Spracherkennungssysteme. Wie jedoch in [2] und in [3] dargestellt wurde, kann sie auch als Maß für die Verständlichkeit verwendet werden. Dabei wird die Fehlerrate der erkannten Wortfolge auf Grund der tatsächlich gesprochenen Wortfolge ermittelt. In den vorangegangenen Arbeiten musste daher auch immer die komplette Sprachstichprobe vor der Analyse transkribiert werden. Da dieser Prozess sehr arbeitsintensiv ist, wird hier ein alternativer vollautomatischer Weg aufgezeigt:

Da aus dem Aufnahmeprozess die exakten Zeitpunkte der Wechsel der Bildtafeln bekannt sind, kann statt der tatsächlich gesprochenen Wortfolge auch die gezeigte Tafel zur Approximation herangezogen werden. Dabei werden die Namen der gezeigten Bilder als Referenz angenommen. Mittels dieser Referenz kann dann die Wortakkuratheit nach folgender Formel ermittelt werden, wobei C die Zahl der korrekt erkannten Worte, I die Zahl der fälschlich eingefügten Worte und R die Zahl der Worte der Referenz sind:

$$WA = \frac{C - I}{R} * 100\%$$

Jedoch stellen fälschliche Einfügungen ein Problem dar. Oft verwenden die Kinder „Trägersätze“ (z.B. „Da ist ein Mond, ein Eimer und ein Baum“), welche nicht in der geschätzten Referenz enthalten sind, die ja nur aus den Namen der gezeigten Bilder (z.B. „Mond, Eimer, Baum“) besteht. Folglich werden weitere Wörter als Fehler ($I = 6$) betrachtet, obwohl die Sprache in unserem Beispiel perfekt erkannt wurde ($C = R = 3$). Dies kann zu sehr niedriger Wortakkuratheit führen ($WA = -100\%$). Daher kann auch ein anderes Maß – die Worterkennungsrate (WR) – herangezogen werden:

$$WR = \frac{C}{R} * 100\%$$

So ergibt sich ein Maß, das trotz Trägersätzen zu guten Ergebnissen führt ($WR = 100\%$). Nach der Aufnahme werden beide Maßzahlen innerhalb von ein bis zwei Minuten – je nach der Aufnahmelänge – berechnet und sind danach per Mausklick verfügbar.

Zur Prüfung des Spracherkennungssystems wurden dessen automatische Bewertungen mit denen von zwei Experten verglichen. Diese vergaben für jede Äußerung Noten auf einer Likert-Skala zwischen 1 (sehr gut) und 5 (mangelhaft). Für die Gesamtnote wurde dann der Mittelwert der Noten berechnet. Als Maß für die Übereinstimmung zwischen Experten und Bewertungssystem wurde die offene Korrelation zwischen dem Mittel der Expertenbewertung und der Wortakkuratheit bzw. der Worterkennungsrate herangezogen.

Ergebnisse: Die Bewertungen der beiden Experten zeigten eine hohe Übereinstimmung (0,91; $p<0,01$). Wie Tabelle 1 entnommen werden kann, ist auch eine hohe Übereinstimmung zwischen den Bewertungen der Experten und des Spracherkennungssystems vorhanden. Dabei liefert bei dieser Anwendung die Worterkennungsrate eine bessere Übereinstimmung als die Wortakkuratheit. Die

Korrelationen zwischen Experten und Spracherkennungssystem sind negativ, da eine hohe Erkennungsrate bei der automatischen Bewertung einem hohen prozentualen Wert, bei den Experten jedoch einer niedrigen Note entspricht (siehe Abbildung 1).

Diskussion: Die automatische Spracherkennung wurde zur Bewertung der Verständlichkeit bisher bei Stimm- und Sprechstörungen angewandt. Dabei wurden bereits die Stimmen von Laryngektomierten [2] und das Sprechen von Kindern mit Lippen-Kiefer-Gaumenspalte [3] untersucht.

Mit der nun entwickelten Software, die eine angepasste Spracherkennungstechnik mit standardisierten Tests zur Erhebung des Lautbefundes kombiniert, ist es erstmals möglich, die Beurteilung der Verständlichkeit von Sprache vollautomatisch über das Internet zu generieren. Dabei arbeitet die Software unabhängig vom jeweiligen Betriebssystem (Windows, Linux, Mac OS) des Clients, da wir Java™ 2 verwenden. So kann man von jedem PC aus mit einem einfachen Internet-Browser die Datenanalyse starten, ohne lange Installationsprozeduren durchgehen zu müssen, wie es nötig wäre, wenn die komplette Spracherkennungssoftware installiert werden müsste.

Ein weiterer Vorteil dieser Architektur ist es, dass Software-Updates nicht bei jedem Client einzeln installiert werden müssen. Stattdessen wird beim Starten stets die neueste Software vom Server geholt. Auch auf teure Spezial-Komponenten kann verzichtet werden, da diese ja nur beim Server gebraucht werden. Am Client ist ein einfaches Mikrophon oder Head-Set und eine Soundkarte ausreichend. Weiterhin ist auch die Sicherheit der Daten gewährleistet, da alle Übertragungen ausschließlich verschlüsselt durchgeführt werden und die Patientendaten pseudonymisiert eingegeben werden können. Nur der behandelnde Arzt hat Zugang zu seinen jeweiligen Patientendaten.

Fazit: Die dargestellte Methodik eignet sich für eine verlässliche und objektive Bewertung der Verständlichkeit von Sprache und kann mit einfachen Mitteln an einem Arbeitsplatz mit Computer und Internet-Anschluss angewendet werden.

Danksagung

Diese Arbeit wird von der Johannes-und-Frieda-Marohn-Stiftung des Universitätsklinikums Erlangen gefördert.

Literatur

[1] Fox AV: PLAKSS - Psycholinguistische Analyse kindlicher Sprechstörungen, Swets & Zeitlinger, Frankfurt a.M., 2002, now available from Harcourt Test Services GmbH, Frankfurt a.M. http://www.harcourt.de

[2] Schuster M, Haderlein T, Nöth E, Lohscheller J, Eysholdt U, Rosanowski F (2006) Intelligibility of laryngectomees' substitute speech: automatic speech recognition and subjective rating. Eur Arch Otorhinolaryngol 263(2):188-193

[3] Schuster M, Maier A, Haderlein T, Nkenke E, Wohlleben U, Rosanowski F, Eysholdt U, Nöth E. Evaluation of Speech Intelligibility for Children with Cleft Lip and Palate by Automatic Speech Recognition. Int J Pediatr Otorhinolaryngol (*im Druck*)

Tabelle 1: Korrelationen zwischen Spracherkennungssystem und Experten

	Bewerter M	Bewerter S	Mittel der Bewerter
WA	-0,83	-0,77	-0,82
WR	-0,88	-0,85	-0,89

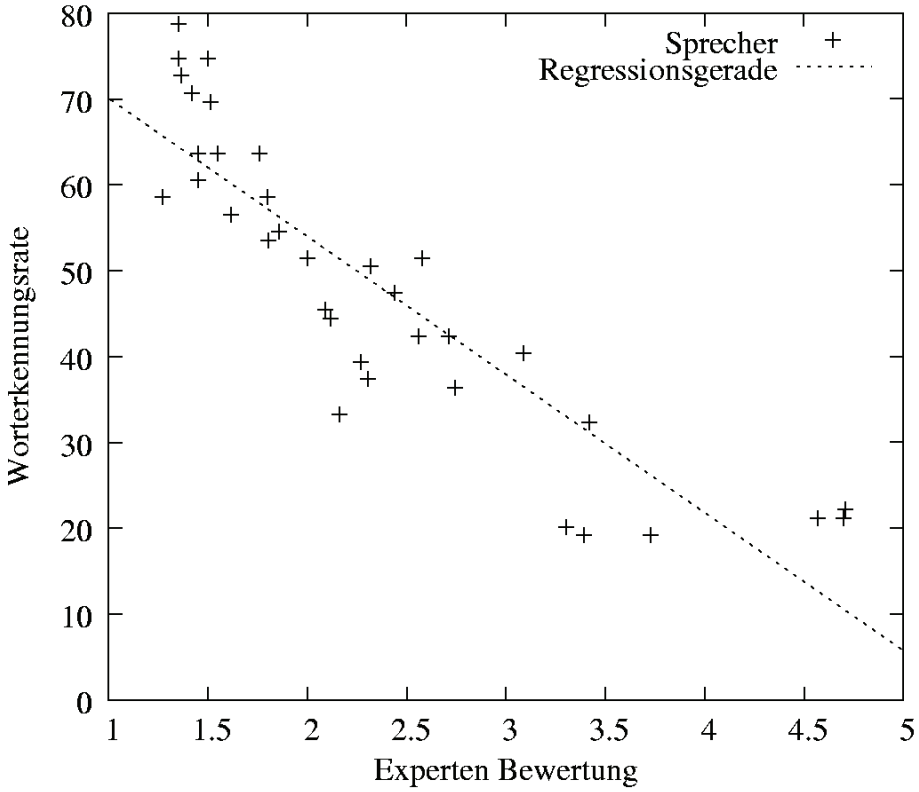

Abbildung 1: Graphische Darstellung der Korrelation zwischen Spracherkennungssystem und Experten