

PEAKS: Ein Client-Server-Internetportal zur Bewertung der Aussprache

A. Maier, T. Haderlein, E. Nöth, M. Schuster

Zielsetzung

Für Sprechstörungen existierten bisher keine objektiven, untersucherunabhängigen Bewertungsinstrumente. In der aktuellen Forschung hat es sich gezeigt, dass die Anwendung von automatischen Spracherkennungssystemen auch in der Medizin denkbar und nützlich ist [1]. Dabei kann eine zentralisierte Auswertung hilfreich sein. Hierfür wurde eine Client-Server-basierte Software realisiert, um die Auswertung über das Internet zu ermöglichen (<http://peaks.informatik.uni-erlangen.de>) [2]. Unser Programm arbeitet unabhängig vom jeweiligen Betriebssystem (Windows, Linux, Mac OS) des Clients. So kann man von jedem PC aus ohne lange Installationsprozeduren, wie es nötig wäre, wenn die komplette Spracherkennungssoftware installiert werden müsste, mit einem einfachen Internet-Browser die Datenanalyse starten. Mit dem System soll dann eine internetbasierte Beobachter-unabhängige Auswertung der Sprachdaten möglich werden. Diese Methode könnte beispielsweise für ein Sprach-Screening eingesetzt werden.

Methodik und Daten

Mittels des Clients, der frei auf unserer Portalseite verfügbar ist, kann gelesene Sprache von einer Testperson aufgezeichnet werden. Kann die Testperson noch nicht lesen, können zu benennende Piktogramme verwendet werden. Die Sicherheit der Daten ist gewährleistet, da alle Übertragungen ausschließlich verschlüsselt und

pseudonymisiert durchgeführt werden. Nur die aufnehmende Kontrollperson (Arzt, Logopäde) hat Zugang zu Daten ihrer Testpersonen. Die Daten werden dann zum Server übertragen und dort mittels Sprachverarbeitungsmethoden [3,4] ausgewertet. Ein Spracherkennungssystem, das schon seit mehreren Jahren an unserem Lehrstuhl ständig weiter entwickelt wird, ermittelt die Anzahl der richtig erkannten Wörter im Verhältnis zu der Textreferenz, die von der Testperson gelesen wurde. Diese Zahl wird auch als Worterkennungsrate bezeichnet und gibt den Verständlichkeitsgrad an.

Nach der Aufnahme wird diese innerhalb von ein bis zwei Minuten – je nach der Aufnahmelänge – berechnet. Danach ist das Ergebnis per Mausklick abrufbar.

Auf teure Spezial-Komponenten am Client-Rechner kann verzichtet werden, da diese nur beim Server gebraucht werden [5]. Am Client ist lediglich ein Mikrofon oder Head-Set und eine Soundkarte notwendig. Zur Standardisierung werden USB-Head-Sets verwendet, da diese eine einheitliche Digitalisierung bieten und damit Audiodaten in vergleichbarer Qualität unabhängig von der verwendeten Hard- und Software ermöglichen.

Insgesamt wurden sechs Kontrollgruppen in ganz Deutschland aufgezeichnet. Diese enthalten Schulkinder der ersten bis vierten Klasse. In Erlangen wurden in den Jahren 2006 bis 2008 insgesamt 322 Kinder aufgezeichnet. Weiterhin wurden in Hannover 126, in Karlsruhe 131 und in Leipzig 61 Kinder aufgenommen.

Diese Aufnahmen stehen im Vergleich zu einem Patientenkollektiv, ebenfalls mit Schulkindern der ersten bis vierten Klasse. Alle 59 Kinder der Patientenkollektivs wurden aufgrund einer Lippen-Kiefer-Gaumenspalte (LKG) schon im Kindesalter operiert. Die Aufnahmen der Patientengruppe (LKG-Schule) wurden in der interdisziplinären Spaltsprechstunde des Universitätsklinikums Erlangen während der regulären Nachsorgeuntersuchung erstellt.

Das Einverständnis der Eltern wurde vor allen Aufnahmen eingeholt. Tabelle 1 enthält eine Übersicht über alle Daten.

Bei 35 der Kinder mit LKG wurde weiterhin eine perzeptive Bewertung der Verständlichkeit von vier Experten durchgeführt. Die Bewertung wurde pro Kind auf einer Skala von 1 (sehr gut) bis 5 (sehr schlecht) ermittelt. Durch Bilden des Durchschnitts über die Experten wurde eine Verständlichkeitsbewertung pro Kind errechnet.

Ergebnisse

Mit der vorgestellten Software kann die Messung der Worterkennungsrate durchgeführt werden. Die Worterkennungsrate ist ein Maß für die Verständlichkeit [6]. Dies wird am Beispiel von Patienten mit Lippen-Kiefer-Gaumenspalte dargestellt. Wie in Abbildung 1 dargestellt, ist die Worterkennungsrate hochgradig mit der subjektiven Verständlichkeit korreliert ($r = 0.9$; $p < 0.001$). Die Korrelation liegt in der gleichen Größenordnung wie die eines erfahrenen Experten zu den übrigen Experten.

Weiterhin ist es möglich, zwischen Kindern mit Sprechpathologie und Kindern aus einem Kontrollkollektiv zu unterscheiden. Untersuchung von Gruppen aus verschiedenen Teilen Deutschlands (Karlsruhe, Hannover, Erlangen und Leipzig) belegen dies. Dabei wurden bei den Kindern mit Lippen-Kiefer-Gaumenspalte signifikant niedrigere Worterkennungsraten gemessen als bei den vier Gruppen von Kontrollkindern ($p < 0.003$). Damit ist das Verfahren auch für ein Screening von Kindern geeignet.

Relevanz der Ergebnisse für die Versorgung

Mit PEAKS ist es erstmals möglich, die Sprachverständlichkeit objektiv zu messen. Das Verfahren ist telemedizinisch an Kliniken und Arztpraxen einsetzbar; die Technik kann aber auch für Kindergärten und Schulen als Screening eingesetzt werden. Die Registrierung und Nutzung ist für Ärzte kostenlos.

Literatur

- [1] T. Haderlein. *Automatic Evaluation of Tracheoesophageal Substitute Voices*. Logos Verlag, Berlin, Germany, 2007.
- [2] A. Maier, E. Nöth, A. Batliner, E. Nkenke, and M. Schuster. “Fully Automatic Assessment of Speech of Children with Cleft Lip and Palate”. *Informatica*, Vol. 30, No. 4, pp. 477–482, 2006.
- [3] A. Maier. *Parallel Robust Speech Recognition*. VDM Verlag Dr. Müller, Saarbrücken, Germany, 2008.
- [4] A. Maier. *Speech Recognizer Adaptation*. VDM Verlag Dr. Müller, Saarbrücken, Germany, 2008.
- [5] E. Nöth, A. Maier, T. Haderlein, K. Riedhammer, F. Rosanowski, and M. Schuster. “Automatic Evaluation of Pathologic Speech — from Research to Routine Clinical Use”. In: V. Matoušek and P. Mautner, Eds., *10th International Conf. on Text, Speech and Dialogue (TSD)*, pp. 294–301, Springer, Berlin, Heidelberg, New York, 2007.
- [6] M. Schuster, A. Maier, B. Vogt, E. Nöth, E. Nkenke, A. Marchis, U. Eysholdt, and F. Rosanowski. “Objektive und automatische Ermittlung der Verständlichkeit von Kindern und Jugendlichen mit Lippen- Kiefer-Gaumenspalten”. In: M. Gross and F. Kruse, Eds., *Aktuelle phoniatisch-pädaudiologische Aspekte 2006*, pp. 43–46, Books On Demand GmbH, Norderstedt, Germany, 2006.

Tabelle 1: Kontroll- und Patientengruppen im Vergleich. Die Kontrollgruppen liegen alle im gleichen Alter im Vergleich zu der Patientengruppe (LKG-Schule), d.h. es gibt keinen signifikanten Unterschied in ihren Altersverteilungen. Die Erkennungsrate des Spracherkennungssystems – also die WR – hingegen zeigt signifikante Unterschiede ($p < 0.003$)

Gruppe	Ort der Aufnahme	#	Alter	WR
Erlangen-06	Erlangen	89	8.8 ± 1.3	60.5 ± 10.7
Erlangen-07	Erlangen	76	8.5 ± 1.4	62.7 ± 10.4
Erlangen-08	Erlangen	157	8.4 ± 1.2	61.1 ± 9.0
Hannover-07	Hannover	126	8.6 ± 1.1	63.7 ± 10.0
Karlsruhe-07	Karlsruhe	131	8.3 ± 1.1	64.1 ± 8.6
Leipzig-07	Leipzig	61	7.9 ± 2.1	59.0 ± 9.8
LKG-Schule	Erlangen	59	8.5 ± 1.6	52.2 ± 15.2

Abbildung 1: Vergleich der subjektiven Bewertung eines Expertengremiums mit der Worterkennungsrate: Es liegt eine hohe Übereinstimmung zwischen beiden vor (Korrelation nach Pearson: $r = 0.9$; $p < 0.001$)

