Tino Haderlein, Maria Schuster, Elmar Nöth, Frank Rosanowski

Einfluss von Lesefehlern auf die textbasierte automatische Verständlichkeitsanalyse

Einleitung

Objektiv-apparative Stimmbewertungen werden derzeit meist auf der Basis gehaltener Vokale durchgeführt. Jedoch reflektiert ein isolierter Vokal keine reale Kommunikationssituation. In früheren Arbeiten wurde gezeigt, dass automatische Spracherkennungsverfahren verwendet werden können, um die Verständlichkeit von pathologischen Sprechern automatisch zu bewerten [1,2,3]. Grundlage der Methode war die Annahme, dass das Spracherkennungssystem umso weniger Wörter eines vorgegebenen Standardtextes „versteht“, je schlechter die Stimmqualität des Sprechers ist. Das Programm kann jedoch nur diejenigen Wörter erkennen, die in seiner Vokabularliste gespeichert sind. Weicht der Patient aufgrund von Lesefehlern, wie z.B. „Der Mor- Nordwind“, oder Äußerungen wie „Ich habe meine Brille nicht auf.“ von dem Standardtext ab, kann dies trotz guter Stimmqualität zu einer niedrigen Erkennungsrate und damit zu einer falschen Bewertung der Stimme führen. Der Fokus dieser Studie lag deshalb auf dem Einfluss von Lesefehlern auf die Korrelation zwischen der automatischen Evaluierung und der Referenzbewertung durch Experten.

Material

Als Testsprecher dienten 85 Personen mit Krebserkrankungen des Kehlkopfes, davon 65 nach einer Larynxteilresektion. Das Durchschnittsalter innerhalb der Gruppe betrug 60,7±9,2 Jahre (min. 34,0, max. 83,0 Jahre), zehn der Patienten waren weiblich. Jede Testperson las den “Nordwind und Sonne”-Text vor und wurde dabei mit einer Abtastfrequenz von 16 kHz und einer Amplitudenauflösung von 16 bit aufgenommen.

Als Vergleichsbasis für die automatische Evaluierung bewerteten fünf Experten das Kriterium „Gesamtverständlichkeit“ bei jedem Sprecher mit Noten von 1 („sehr gut verständlich“) bis 5 („extrem schlecht verständlich“). Aus den fünf Bewertungen für jede Aufnahme wurde jeweils eine Durchschnittsnote gebildet.

Methode

Aus den Originalaufnahmen des Textes wurde jeweils eine zweite Variante erstellt, aus der Lesefehler und Äußerungen, die nicht zum Text gehörten, herausgeschnitten wurden. Insgesamt wurden 368 (3,9%) der 9519 Wörter und Wortfragmente in den Aufnahmen auf diese Weise entfernt (vgl. Abb. 1).

Das auf Hidden-Markov-Modellen basierende Spracherkennungssystem war unabhängig vom gegenwärtigen Projekt am Lehrstuhl für Mustererkennung der Universität Erlangen-Nürnberg entwickelt und bereits in zahlreichen Forschungsprojekten erfolgreich eingesetzt worden. Von einer Ausgründung des Lehrstuhls (www.sympalog.de) wird es mit Erfolg zum Einsatz in Telefondialogsystemen vertrieben.

Zielkriterium der automatischen Analyse waren die Wortakkuratheit WA und die Worterkennungsrate WR, die mit der Verständlichkeitsbewertung durch die Experten korreliert wurden. Die Wortakkuratheit errechnet sich aus der Formel

$WA\ [\%] = 100 * [1-(N_{sub}+N_{del}+N_{ins})/N_{ges}]$, wobei

N_{sub}: Anzahl der vom Erkennungssystem durch andere Wörter ersetzten, d.h. „verwechselten“, Wörter (Substitutionen)

N_{del}: Anzahl der nicht erkannten Wörter (Deletionen)

N_{ins}: Anzahl der fälschlicherweise eingefügten Wörter (Insertionen)

N_{ges}: Anzahl aller gesprochenen Wörter

Die Wortkorrektheit oder Worterkennungsrate (engl. „word recognition rate“, WR), wird genau wie die Wortakkuratheit berechnet, allerdings ohne Berücksichtigung der fälschlicherweise eingefügten Wörter N_{ins}. Der Maximalwert von Wortakkuratheit und Wortkorrektheit beträgt 100%. Der mögliche Minimalwert der Wortkorrektheit ist 0%, während die Wortakkuratheit bei großem N_{ins} auch negativ werden kann.

Ergebnisse

Die Korrelation zwischen den menschlichen Verständlichkeitsbewertungen und der Maschine lagen bei r=-0,61 für die WA und r=-0,55 für die WR, sowohl für die Originalaufnahmen als auch für die Aufnahmen, aus denen die Lesefehler eliminiert

worden waren. In der folgenden Tabelle sind jeweils die Messgröße, ihr Mittelwert μ und ihre Standardabweichung σ, der Minimal- und Maximalwert sowie die Korrelation r zur Expertenbewertung angegeben:

	Messgröße	μ	σ	min.	max.	r
mit Lesefehlern	WA	48,0	17,2	3,4	81,3	-0,61
ohne Lesefehler	WA	49,3	17,0	10,1	81,3	-0,61
mit Lesefehlern	WR	53,2	15,3	9,1	82,2	-0,56
ohne Lesefehler	WR	54,1	15,4	9,1	82,2	-0,55

Diskussion

Im Hinblick auf die breite klinische Anwendung der Messmethode kann folgendes geschlossen werden: Lesefehler müssen nicht gesondert eliminiert werden. Sie haben keinen entscheidenden Einfluss auf das Auswertungsergebnis. Es bleibt zu prüfen, ob sich die Gesamtkorrelation noch verbessern lässt, wenn häufig auftretende Fehler oder zusätzliche Phrasen, die nicht zum Text gehören, in das Erkennungsvokabular aufgenommen werden.

Danksagung

Diese Arbeit wird von der Deutschen Krebshilfe (Fördernr. 107873) gefördert.

Literatur

[1] Schuster M, Haderlein T, Nöth E, Lohscheller J, Eysholdt U, Rosanowski F. Intelligibility of laryngectomees' substitute speech: automatic speech recognition and subjective rating. Eur Arch Otorhinolaryngol 2006;263(2):188-93.

[2] 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 means of automatic speech recognition. Int J Pediatr Otorhinolaryngol 2006;70(10):1741-7.

[3] Windrich M, Maier A, Kohler R, Nöth E, Nkenke E, Eysholdt U, Schuster M. Automatic Quantification of Speech Intelligibility of Adults with Oral Squamous Cell Carcinoma. Folia Phoniatr Logop 2008;60(3):151-156.

Abbildung

Abb.1: Anzahl von Lesefehlern in den 85 Aufnahmen des „Nordwind und Sonne“-Textes