Sprachenunabhängige Verständlichkeitsanalyse bei Kindern mit orofazialen Spaltfehlbildungen auf Deutsch und Italienisch mittels akustischer Modellierung

Einleitung

Zur objektiv-apparativen Bewertung der Verständlichkeit von Kindern mit Lippen-Kiefer-Gaumenspalten (LKG) werden von der berichtenden Arbeitsgruppe für mehrere Sprachen erfolgreich Spracherkennungssysteme eingesetzt [1]. Das Training eines Spracherkenners für verschiedene Sprachen ist dabei sehr aufwendig. In der vorliegenden Studie wurde untersucht, ob für eine automatische Verständlichkeitsbewertung ein sprachenabhängiger Spracherkenner zwingend erforderlich ist, oder ob die Verwendung eines rein akustisch getriebenen Ansatzes ausreicht, um sprachenunabhängig eine Bewertung abzugeben.

Material und Methoden

Es wurden zwei unterschiedliche Datensätze verwendet: 14 italienische und 35 deutsche Kinder mit LKG. Die Kinder wurden beim Sprechen eines Standardtextes aufgenommen. Im Fall der deutschen Kinder wurde der PLAKSS-Test verwendet, die italienischen Kinder führten einen standardisierten Test durch, der im Nachsprechen von 19 italienischen Sätzen bestand. Ausführlichere Informationen zu den italienischen Daten sind in [2] zu finden.

Referenzwerte für das automatische System waren für beide Gruppen auditive Verständlichkeitsbewertungen. Bei den deutschen Kindern lagen Bewertungen von 5 Logopädinnen auf einer Skala von 1 ("sehr gut verständlich") bis 5 ("sehr schlecht verständlich") vor. Die italienischen Kinder wurden von einer Logopädin auf einer

Skala von 1 bis 4 bewertet. Auch hier gilt: Je niedriger der Wert, desto besser die Verständlichkeit.

Die Grundidee des Systems ist es, die gesprochenen Äußerungen eines Sprechers durch Gaußsche Mischverteilungsmodelle (GMMs) zu modellieren. Dazu werden zuerst aus den Sprachaufnahmen automatische Messwerte (Mel-Frequency-Cepstrum-Coefficients, MFCCs) gewonnen. Zur Erzeugung dieser Messwerte wird das Sprachsignal in 16 ms dauernde Abschnitte unterteilt. Für jeden Abschnitt werden anschließend die MFCCs berechnet. Im nächsten Schritt werden alle Messwerte aller Sprecher zusammengefasst, um ein sogenanntes "Hintergrundmodell" (Universal Background Model, UBM) zu trainieren. Auf der Grundlage dieses allgemeinen Modells werden anschließend mittels Maximum-A-Posteriori-Adaption (MAP) die akustischen Modelle für die einzelnen Sprecher abgeleitet. Aus den einzelnen Sprechermodellen werden nun die Mittelwerte und die Kovarianzen extrahiert und zu sprechercharakteristischen Merkmalen konkateniert. Durch eine Regression wird eine Funktion geschätzt, die aus diesen Merkmalen einen Vorhersagewert für die menschliche, auditive Bewertung des jeweiligen Kindes liefert. Eine detaillierte Beschreibung des Systems ist in [3] zu finden. Unterschiedliche Werte für die Anzahl der Gaußdichten der einzelnen Sprechermodelle wurden getestet.

Ergebnisse

Die Korrelation zwischen den auditiven Befunden und den Vorhersagewerten des automatischen Systems lag bei maximal r=0,81 für die deutschen und r=0,83 für die italienischen Kinder. Ausführliche Ergebnisse auf dem deutschen Datensatz sind in Tabelle 1 zu finden. In Tabelle 2 befinden sich die Ergebnisse des italienischen Datensatzes.

Tabelle 1: Korrelation zwischen der automatischen Bewertung und der menschlichen Bewertung in Abhängigkeit von der Anzahl der Gaußdichten und der verwendeten Merkmale (dt. Datensatz)

Anzahl der Gaußdichten	Mittelwert	Kovarianz	Kombination aus Mittelw. + Kova.
32	0,67	0,72	0,79
64	0,75	0,72	0,80

128	0,77	0,72	0,80
256	0,79	0,72	**0,81**

Tabelle 2: Korrelation zwischen der automatischen Bewertung und der menschlichen Bewertung in Abhängigkeit von der Anzahl der Gaußdichten und der verwendeten Merkmale (ital. Datensatz)

Anzahl der Gaußdichten	Mittelwert	Kovarianz	Kombination aus Mittelw. + Kova.
32	0,69	**0,83**	0,76
64	0,63	0,78	0,71
128	0,53	0,51	0,61
256	0,49	0,47	0,42

Diskussion

Die hohe Korrelation zwischen menschlicher und automatischer Bewertung zeigt, dass das verwendete System zur objektiven Stimmevaluierung geeignet ist. Die Korrelationen bei spracherkennungsbasierten Systemen [3,4] sind in derselben Größenordnung wie das in dieser Studie vorgestellte Verfahren. Somit ist eine automatische Verständlichkeitsbewertung von Kindern mit LKG nicht auf eine detaillierte Analyse der gesprochenen Wörter angewiesen, sondern ist auch durch rein akustische Sprechermodellierung möglich. Diese Tatsache erlaubt es, auch das vorgestellte System ohne aufwendiges Training in unterschiedlichen Sprachen anzuwenden, wie in dieser Arbeit am Beispiel von Deutsch und Italienisch gezeigt wurde. Die Verlässlichkeit der Methode ist in der Zukunft an größeren Stichproben zu untersuchen und auf andere Landessprachen auszudehnen.

Literatur

[1] 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 Oct;70(10):1741-7

[2] Scipioni M, Gerosa M, Giuliani D, Nöth E, Maier A.

Intelligibility Assessment in Children with Cleft Lip and Palate in Italian and German.

Proc. Interspeech 2009, 967-970, 2009

[3] Bocklet T, Haderlein T, Hönig F, Rosanowski F, Nöth E.

Evaluation and Assessment of Speech Intelligibility on Pathologic Voices Based upon Acoustic Speaker Models.

Proc. AVFA 2009, 89-92, 2009

[4] Maier A, Haderlein T, Eysholdt U, Rosanowski F, Batliner A, Schuster M, Nöth E.

PEAKS - A System for the Automatic Evaluation of Voice and Speech Disorders.

Speech Communication, 51(5):425-437, 2009.