Automatische Bewertung der Nasalität von Kindersprache

Maier[1], Nöth[2], Wohlleben[3], Eysholdt[1], Schuster[1]

[1]Abteilung für Phoniatrie und Pädaudiologie des Klinikums der Universität Erlangen-Nürnberg, Bohlenplatz 21, 91054 Erlangen

[2]Lehrstuhl für Mustererkennung (Informatik 5), Universität Erlangen-Nürnberg, Martensstraße 3, 91058 Erlangen

[3]Logopädische Praxis Wohlleben Ulrike Dr.
Moststraße 27, 90762 Fürth

E-Mail: Andreas.Maier@informatik.uni-erlangen.de

**Einleitung**

In der aktuellen Forschung bei Sprechstörungen hat es sich gezeigt, dass automatische Klassifikationssysteme zur Bewertung der Nasalität anwendbar sind. Bisher wurde dies aber immer nur für gehaltene Vokale oder Wortteile gezeigt. Um komplette Wörter oder Sätze zu bewerten, waren bisher immer teure, apparative Methoden notwendig. Der hier beschriebene Ansatz kommt mit einem Standard-PC und einem Mikrophon aus und analysiert die akustischen Eigenschaften veränderter Lautbildung mittels Methoden der automatischen Spracherkennung.

**Patienten und Methode**

Mit der internetbasierten Bewertungsumgebung PEAKS [1] wurden Daten von insgesamt 13 Kindern (3 Mädchen und 10 Jungen) im Alter von 8,5 ± 3,0 Jahren mit Lippen-Kiefer-Gaumenspalte aufgenommen. Ihnen wurden Bilder am PC gezeigt, die Sie benennen sollten. Die Bildtafeln stellen eine digitale Version des PLAKSS-Tests [2] dar, der zur Erhebung des Lautbestandes bei Kindern dient. Dieser Test enthält alle deutschen Phoneme an verschiedenen Positionen im Wort.

Eine erfahrene Logopädin klassifizierte jedes der 99 Zielworte des Tests zur Überprüfung von phonetischen Störungen als „normal“ oder „hypernasal“. Anhand der spektralen Eigenschaften (Mel Frequenz Cepstrum Koeffizienten) der gesprochenen Wörter wurde für das Spracherkennungssystem ein Klassifikationssystem trainiert, das automatisch hypernasalierte Wörter erkennen kann. Insgesamt konnten 841 Worte in den Sprachdaten korrekt ausgeschnitten werden. 11% davon (95) waren als „hypernasal“ markiert.

Als Klassifikator für Hypernasalität wurde ein Gaussian Mixture Model (GMM) eingesetzt, welches mit einem Standard-Verfahren der Spracherkennung, dem Expectation-Maximization Algorithmus, trainiert wurde [3]. Da nicht alle Laute im Wort zwingend hypernasal sind, wird jedes Wort alle 10 ms klassifiziert. Ist ein hypernasalierter Laut im Wort vorhanden, so sollte sich eine besonders hohe Wertung für den Hypernasalitäts-Klassifikator finden. Um das Training des GMMs zu verfeinern, wurde alternativ nach dem ersten Trainingsdurchlauf die Trainingsmenge anhand des Klassifikators neu in die Klassen „normal“ und „hypernasal“ eingeteilt. Dieser Vorgang wird in der Literatur auch als „Bootstrapping“ bezeichnet [4].

Zur Evaluierung wurden die absolute Erkennungsrate (RR), die klassenweise gemittelte Erkennungsrate (CL), die Korrelationen nach Pearson und Spearman eingesetzt. Während die RR lediglich den Prozentsatz der richtig erkannten Wörter widerspiegelt, berücksichtigt die CL auch die Häufigkeit der beiden Kategorien „normal“ und „hypernasal“, um eine Verzerrung des Ergebnisses durch die Überrepräsentation einer Kategorie in der Stichprobe zu vermeiden. Daher werden für die CL zuerst die Erkennungsraten jeder einzelnen Kategorie ermittelt, aus denen dann der Durchschnitt gebildet wird. So wird jede Kategorie als ebenbürtig angesehen.

Es wurden die Anzahl der von der erfahrenen Logopädin als „hypernasal“ gekennzeichneten Wörter mit der Anzahl der automatisch detektierten Wörter pro Sprecher korreliert.

Alle Experimente wurden im Leave-One-Speaker-Out-Verfahren überprüft. So ist sichergestellt, dass jeder Sprecher einmal aus der Trainingsmenge entfernt wurde und als Testsprecher diente. Die Klassifikations- und Korrelationsergebnisse wurden daher immer mit einer disjunkten Trainings- und Testmenge durchgeführt.

**Ergebnisse**

Für das GMM wurde die Zahl der Komponenten der Mischverteilung variiert. Dabei ergaben sich ein Maximum der RR bei 10 Dichten mit 75,7 % und ein Maximum der CL bei 5 Dichten mit 64,9 %. Für die Korrelationen zwischen der subjektiven und automatisch ermittelten Bewertung der Hypernasalität wurden hierbei keine signifikanten Ergebnisse festgestellt (siehe Tabelle 1).

In Tabelle 2 sind die Ergebnisse mit Bootstrapping dargestellt. Hier ergeben sich etwas schlechtere Erkennungsraten als für den Fall ohne Bootstrapping. Jedoch konnten für 5 Dichten eine hohe Übereinstimmung zwischen der subjektiven und

automatisch ermittelten Bewertung der Hypernasalität mit Korrelationen von 0,79 nach Pearson und 0,75 nach Spearman erreicht werden. Beide Ergebnisse sind mit $p < 0,01$ signifikant.

**Diskussion**

Die gezeigten Ergebnisse sind sehr viel versprechend. Es konnte mit der genannten Methode bereits eine gute automatische Klassifikation der Hypernasalität erreicht werden, was sich in der hohen Korrelation zwischen der Zahl der detektierten und der von der Erfahrenen als „hypernasal“ gekennzeichneten Wörter zeigt. Während sich das Bootstrapping positiv auf die Korrelation zur subjektiven Bewertung durch eine Erfahrene auswirkte, konnten damit nur etwas niedrigere Erkennungsraten erreicht werden. Wir führen dies auf die wortweise Annotation zurück, die ja ein ganzes Wort als „nasal“ markiert, auch wenn nur ein Laut im Wort betroffen ist. Eine Überprüfung und Verbesserung der Methode ist an einer größeren Stichprobe geplant. Damit ließe sich dann eine automatische, nicht-invasive, und wenig aufwändige Quantifizierung der Hypernasalität von Kindersprache ermöglichen. Eine Erweiterung der Methode auf andere Lautbildungsstörungen wird derzeit untersucht.

| # Dichten | Gesamterkennungsrate (RR)/<br>klassenweise gemittelte Erkennungsrate (CL) |
|---|---|
| 2 | 77,5 / 64,8 |
| 5 | 74,4 / **64,9** |
| 10 | **75,7** / 63,8 |
| 15 | 73,7 / 63,1 |

Tabelle 1: Erkennungsraten für Hypernasalität auf Wort-Ebene.

| # Dichten | RR / CL | Korrelation<br>Pearson / Spearman |
|---|---|---|
| 2 | 82,0 / 47,2 | -0,14 / 0,24 |
| 5 | **74,7** / 52,7 | **0,79 / 0,75*** |
| 10 | 55,6 / **59,4** | 0,68 / 0,63* |
| 15 | 34,6 / 52,7 | 0,35 / 0,53 |

Tabelle 2: Erkennungsraten in % auf Wort-Ebene und Korrelation der Anzahl der detektieren hypernasalen Wörter zu tatsächlich Vorhandenen mit der „Bootstrapping“ Methode. Die signifikanten Korrelationen ($p<0.01$) wurden mit einem * gekennzeichnet.

[1] A. Maier, T. Haderlein, C. Hacker, E. Nöth, F. Rosanowski, U. Eysholdt, S. Schuster: Automatische internetbasierte Evaluation der Verständlichkeit . In: Gross, Manfred ; Kruse, Friedrich E. (Eds.) : Aktuelle phoniatrisch-pädaudiologische Aspekte 2006 (23. Wissenschaftliche Jahrestagung der Deutschen Gesellschaft für Phoniatrie und Pädaudiologie Heidelberg 15. - 17. September 2006). Vol. 14. Norderstedt : Books On Demand GmbH Norderstedt, 2006, pp. 87-90.
[2] 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
[3] E. G. Schukat-Talamazzini. Automatische Spracherkennung, 1995, Vieweg Verlag, Braunschweig.
[4] S. Abney: Bootstrapping. In 40th Annual Meeting of the Association for Computational Linguistics: Proceedings of the Conference. 2002.