

Grafische Darstellung von Sprechstörungen mithilfe der Sammon-Transformation

Tino Haderlein¹, Dominik Zorn², Elmar Nöth², Hikmet Toy¹, Ulrich Eysholdt¹, Frank Rosanowski¹

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

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

E-Mail: Tino.Haderlein@informatik.uni-erlangen.de

Einleitung

Automatische, objektive Methoden zur Stimmevaluation basieren derzeit meist lediglich auf einer Aufnahme eines gehaltenen Vokals. Mithilfe der automatischen Spracherkennung ist es jedoch möglich, verschiedene Laute in die Analyse einzubeziehen. Die dabei auftretenden großen Datenmengen müssen in eine übersichtliche Form gebracht werden, um dem medizinischen Personal im klinischen Alltag nützlich zu sein. Zu diesem Zweck wird die sog. Sammon-Transformation verwendet: Sie bildet die automatisch berechneten Messwerte in einer zweidimensionalen Grafik ab.

Material und Methoden

Vier Sprechergruppen wurden in dieser Studie evaluiert. 18 laryngektomierte Männer mit einer seit mehr als einem Jahr funktionierenden tracheoösophagealen Ersatzstimme und einem Durchschnittsalter von 64,2 Jahren [1] bildeten zusammen mit jeweils 9 chronisch heiseren Männern und Frauen (Durchschnittsalter 47,6 Jahre) die Menge der untersuchten pathologischen Sprecher. Dem gegenüber standen zwei Gruppen von Normalsprechern. 18 Männer, im Durchschnitt 65,4 Jahre alt, stellten die erste Kontrollgruppe. Die zweite bestand aus 9 Männern und 7 Frauen, deren

Durchschnittsalter ca. 25 Jahre betrug. Eine genaue Altersangabe war nicht möglich, da diese Aufnahmen einem Sprachkorpus entnommen wurden, in welchem diese Daten nicht erhoben worden waren (BAS Strange Corpus 1, www.phonetik.uni-muenchen.de/Bas/).

Alle Sprecher lasen jeweils den „Nordwind und Sonne“-Text vor. Die Aufnahmen erfolgten mit einem „dnt Call 4U Comfort“-Headset (Abtastfrequenz 16 kHz, Amplitudenauflösung 16 bit). Die jungen Normalsprecher wurden auf DAT-Band aufgenommen (Abtastfrequenz 48 kHz, danach unterabgetastet mit 16 kHz; Amplitudenauflösung 16 bit).

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. Kommerziellen Erfolg beim Vertrieb des Systems mit Telefondialogsystemen erzielt eine Ausgründung des Lehrstuhls (www.sympalog.de).

Zum Zwecke der Stimmevaluation wurden die internen Parameter des Spracherkennungssystems jeweils mithilfe der Textaufnahme an den aktuellen Sprecher adaptiert [2]. Die gewünschten Messgrößen ergaben sich aus den Unterschieden zu den ursprünglichen Parameterwerten. Mit der Sammon-Transformation [3] wurde deren große Zahl dann in eine zweidimensionale Darstellung umgerechnet. Der große Vorteil der Sammon-Transformation gegenüber eingeführten Verfahren, wie z.B. der Principle Component Analysis (PCA), besteht darin, dass der Begriff des „Abstandes“ zweier Sprecher dabei keinerlei mathematischen Einschränkungen (z.B. lineare Transformation) unterliegt.

Ergebnisse

Abb. 1 zeigt eine Grafik, die aus allen vorhandenen Sprachaufnahmen erstellt wurde. Sie zeigt, dass nicht nur die einzelnen Sprechergruppen, die sich in Stimmqualität und Alter

unterscheiden, fast völlig voneinander getrennt wurden, sondern innerhalb dieser Gruppen auch Männer und Frauen.

Diskussion

In der gezeigten Grafik nimmt der Grad der Pathologie von links nach rechts zu, die Grundfrequenz der Stimme von oben nach unten. Welche Stimmeigenschaften bzw. Messgrößen entlang welcher Achse innerhalb der Grafik dargestellt werden, ist von den eingespielten Daten abhängig. Bei Verwendung einer wohl dokumentierten Sprechermenge als Grundlage für die Berechnung ist es jedoch möglich, einen neuen, unbekanntem Sprecher in die Grafik hineinzuprojizieren. Seine Lage gibt dann Auskunft über die Stimmeigenschaften im Vergleich zu den bekannten Sprechern. Die Koordinaten im Bild dienen folglich als objektive Bewertungsparameter.

Danksagung

Diese Arbeit wird von der Deutschen Krebshilfe (Fördernr. 106266) gefördert.

Literatur

[1] 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

[2] Steidl S, Stemmer G, Hacker C, Nöth E (2004)

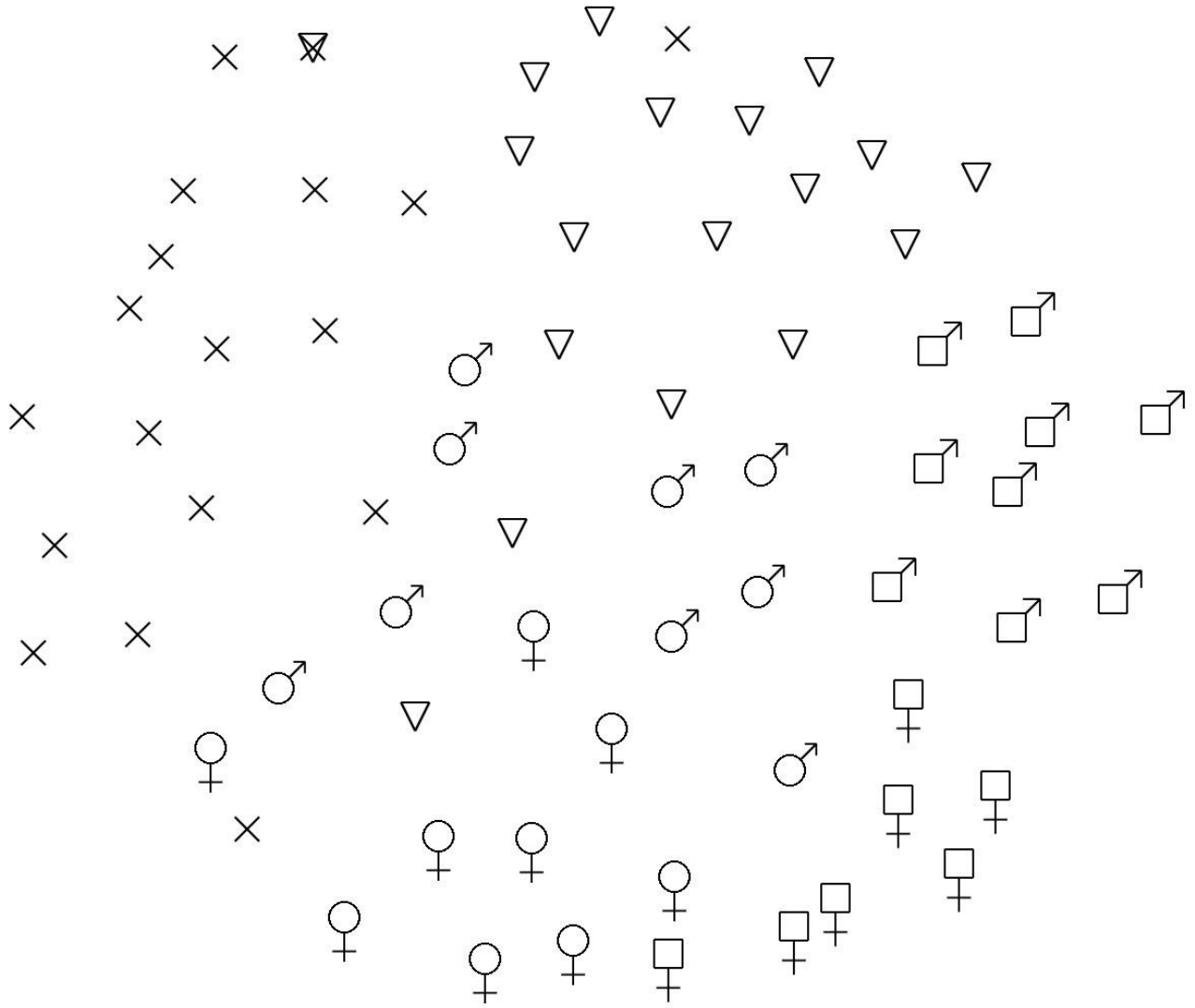
Adaptation in the Pronunciation Space for Non-Native Speech Recognition. In: Proc. Int. Conf. on Spoken Language Processing (ICSLP), Jeju Island, Korea vol. 1, pp. 318-321

[3] Sammon J (1969)

A nonlinear mapping for data structure analysis. IEEE Trans. Computers C-18:401-409

Abbildung

Abb. 1: Sammon-Karte, erstellt aus Aufnahmen von pathologischen und Normalstimmen



×	Laryngectomized	▽	Old reference speakers
○	Hoarse	□	Young reference speakers