# Evaluation tracheoösophagealer Ersatzstimmen durch naive Hörer, Experten und automatische Spracherkennung

Martina Bellanova[1], Maria Schuster[1], Tino Haderlein[1], Elmar Nöth[2], Ulrich Eysholdt[1], Frank Rosanowski[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
E-Mail: Tino.Haderlein@informatik.uni-erlangen.de

## Einleitung

Die Stimmrehabilitation laryngektomierter Patienten mit Stimmventilprothesen (tracheoösophageale Ersatzstimme, TE-Stimme) ist heute „state of the art“. Die objektive Bewertung solcher Ersatzstimmen steht an der Schwelle zur Klinikreife: So ist die maschinelle Verständlichkeitsbewertung eines vorgelesenen Textes durch ein automatisches Spracherkennungssystem möglich; in einer früheren Pilotstudie mit 18 Patienten ergab ein Vergleich der automatischen Evaluation mit der durchschnittlichen Expertenbewertung eine Korrelation von r=-0,84 [1]. Im Hinblick auf die methodische Optimierung dieses automatischen Spracherkennungssystems wurde nun die Stichprobe auf 33 Patienten und die Bewerter um eine Gruppe naiver Hörer erweitert. Es wurde untersucht, inwieweit sich die Auswertungen der Experten und naiven Hörer sowie des Spracherkennungssystems decken.

## Material und Methode

33 männliche Patienten mit einem Durchschnittsalter von 61,8±7,7 Jahren, die nach der Laryngektomie mit einer Provox®-Stimmventilprothese versorgt worden waren, lasen den „Nordwind-und-Sonne“-Text vor und wurden dabei mit einem „dnt Call 4U Comfort“-Headset (Abtastfrequenz 16 kHz, Amplitudenauflösung 16 bit) aufgenommen. Anschließend wurden diese Nahbesprechungsaufnahmen von einer Gruppe von fünf Experten sowie einer Gruppe von elf naiven Hörern hinsichtlich der Verständlichkeit und

Gesamtqualität bewertet. Die Verständlichkeit wurde dabei auf einer Likertskala von 1 („sehr gut“) bis 5 („extrem schlecht“), die Gesamtqualität auf einer visuellen Analogskala mit Werten zwischen 0,0 („sehr gut“) und 10,0 („sehr schlecht“) markiert. Um die Schwankungen in der Bewertung der naiven Hörer zu ermitteln, erfolgte eine zweite Evaluation der Aufnahmen durch diese Gruppe im Abstand von ca. sechs Wochen. Darüber hinaus lagen ca. zwei Jahre alte Ergebnisse des Expertengremiums für 18 der 33 Patienten zum Vergleich mit den aktuellen Bewertungen vor.

Sowohl bei der Bewertung durch die Experten, als auch durch die naiven Hörer erfolgte das Abspielen der Aufnahmen in zufälliger Reihenfolge. Die Aufnahmen wurden nur einmal abgespielt, die Bewertungen erfolgten zur selben Zeit. Die automatische Analyse der digitalisierten Aufnahmen erfolgte durch ein Spracherkennungssystem des Lehrstuhls für Mustererkennung der Universität Erlangen-Nürnberg, das bereits für Marktzwecke professionalisiert wurde (www.sympalog.de). Zielkriterium der automatischen Analyse war zunächst die Wortakkuratheit (WA), die dem Kriterium der Gesamtverständlichkeit der menschlichen Bewerter entspricht.

**Ergebnisse**

Die Korrelation in den Bewertungen durch die naiven und erfahrenen Bewerter sowie des Spracherkennungssystems sind in Tabelle 1 zusammengefasst. Insgesamt sind die Ergebnisse für Verständlichkeit und Gesamtqualität sehr ähnlich. Die Einzelbewertungen für die beiden Kriterien korrelieren untereinander sowohl bei naiven Hörern (r=0,98 im ersten und 0,97 im zweiten Durchlauf) als auch bei den Experten (r=0,98 und 0,96) äußerst stark. Die Intra-Rater-Korrelation zwischen erster und zweiter Bewertungssitzung ist in Tabelle 2 dargestellt.

| | Gesamtqualität | | Verständlichkeit | |
|---|---|---|---|---|
| | 1. Bew. | 2. Bew. | 1. Bew. | 2. Bew. |
| Experten vs. autom. Spracherkennung | -0.86* | -0.81 | -0.84* | -0.86 |
| Naive vs. autom. Spracherkennung | -0.72 | -0.74 | -0.71 | -0.73 |

| | | | | |
|---|---|---|---|---|
| Experten vs. naive Bewerter | -0.91 | -0.91 | -0.88 | -0.90 |

Tabelle 1: Korrelationskoeffizient r zwischen Evaluationen verschiedener Bewerter (*=Auswertung auf 18 Sprechern)

| | Gesamtqualität | Verständlichkeit |
|---|---|---|
| Experten | 0,98* | 0,96* |
| Naive Hörer | 0,95 | 0,94 |

Tabelle 2: Intra-Rater-Korrelation r zwischen erster und zweiter Bewertungssitzung (*=Auswertung auf 18 Sprechern)

**Diskussion**

Die Ergebnisse bestätigen, dass eine zuverlässige automatische Bewertung der Ersatzstimmen möglich ist, wobei die automatische Evaluation näher am Urteil der Experten als an dem der naiven Hörer liegt. Trotzdem ähneln sich die Bewertungen von Experten und naiven Hörern stark, wobei naive Hörer tendenziell schlechtere Noten vergeben als Experten (siehe Abbildung 1). Der Grund hierfür liegt möglicherweise darin, dass naive Bewerter TE-Stimmen eher unwillkürlich mit einer gesunden Stimme vergleichen, wohingegen Experten aufgrund des größeren Erfahrungsschatzes die jeweilige Ersatzstimme in Relation zu anderen Ersatzstimmen bewerten. Die Ergebnisse solcher Bewertungen sind sowohl bei naiven Hörern als auch bei Experten reproduzierbar. Weiterhin lassen die Ergebnisse darauf schließen, dass die Verständlichkeit bei menschlichen Hörern eine sehr große Rolle bei der Beurteilung der Stimmqualität spielt. Hierfür sprechen die sich fast perfekt deckenden Bewertungen für die beiden Kriterien.

Die weitere Optimierung des automatischen Systems, auch zur maschinellen Beurteilung von TE-Sprechern über ein Telefon, ist Gegenstand aktueller Arbeiten.

**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

**Abbildung**

Abb.1: Vergleich der Durchschnittsnote jeder Aufnahme für das Kriterium „Gesamtverständlichkeit“, bewertet von Experten und naiven Hörern (erster Durchgang) mit Noten von 1 („sehr gut verständlich“) bis 5 („extrem schlecht verständlich“)