

Tino Haderlein<sup>1,2</sup>, Elmar Nöth<sup>2</sup>, Michael Döllinger<sup>1</sup>, Cornelia Schwemmler<sup>3</sup>, Martin Ptok<sup>3</sup>

<sup>1</sup>Phoniatische und pädaudiologische Abteilung, Klinikum der Universität Erlangen-Nürnberg

<sup>2</sup>Lehrstuhl für Mustererkennung (Informatik 5), Universität Erlangen-Nürnberg

<sup>3</sup>Klinik und Poliklinik für Phoniatrie und Pädaudiologie, Medizinische Hochschule Hannover

## **Vergleich der subjektiv-auditiven RBH-Beurteilung mit apparativer prosodischer Analyse und Irregularitätsberechnung**

### **Einleitung**

In früheren Arbeiten wurde gezeigt, dass prosodische Analyseverfahren verwendet werden können, um Stimm- und Sprechereigenschaften automatisch zu bewerten [1]. Der Fokus dieser Studie lag auf dem Vergleich einer Kombination von textbasierter prosodischer Analyse (Lehrstuhl für Mustererkennung, Universität Erlangen-Nürnberg) und Irregularitätsbeurteilung (Laryngograph Speech Studio, Laryngograph Ltd, London) [2] mit der subjektiv-auditiven RBH-Bewertung.

### **Material**

Für die vorliegende Studie wurden Stimmproben von 58 Personen (43 Frauen, 15 Männer) mit Dysphonien verschiedener Ursachen (z.B. Reinke-Ödeme, Stimmlippenparenese, chronische Laryngitiden) und ohne Dysphonien verwendet. Das Durchschnittsalter betrug  $48,7 \pm 17,8$  Jahre (min. 12,2, max. 81,9 Jahre). Als Text diente der phonetisch ausgewogene Standardtext „Der Nordwind und die Sonne“, die Aufnahme erfolgte mit einem Nahbesprechungsmikrofon (Abstand zum Mikrofon 10 cm, Abtastfrequenz 44,1 kHz, Amplitudenauflösung 16 bit). Für die prosodische Analyse wurden die Daten mit 16 kHz neu abgetastet.

### **Methode**

Als Vergleichsbasis für die automatische Evaluierung erfolgte die subjektiv-auditive Beurteilung durch 19 Logopädiestudenten nach dem RBH-Schema. Es wurde jedoch betont,

dass die Aufnahmen nur gemäß des Höreindrucks zu bewerten sind und die RBH-Konvention (Heiserkeit nicht geringer zu beurteilen als die Rauigkeit) nicht eingehalten werden muss [3].

Der erste Teil der apparativen Diagnostik bestand in der Berechnung prosodischer Merkmale. Basierend auf Wort- und Pausendauern, der Sprachgrundfrequenz  $F_0$  und der Energie im Signal [1] wurden 33 prosodische Merkmale pro Wort bzw. pro Wort-Pause-Wort-Intervall berechnet. Die größte Gruppe umfasste die  $F_0$ -Merkmale, die u.a. Mittelwert, Minimum, Maximum, den Wert bei Stimmeinsatz und -ausklang sowie ihre jeweiligen Positionen im betrachteten Intervall enthielten. 15 weitere Merkmale, auf Abschnitten von jeweils 15 Wörtern Länge berechnet, umfassten Mittelwert und Standardabweichung von Jitter und Shimmer, Anzahl, Dauer und maximale Dauer von stimmhaften und stimmlosen Abschnitten, das Verhältnis der Anzahl bzw. Dauer von stimmhaften zu stimmlosen Bereichen sowie das Verhältnis der Dauer von stimmhaften bzw. stimmlosen Abschnitten zur Gesamtdauer des Signals. Die Standardabweichung der Sprachgrundfrequenz  $F_0$  wurde hier ebenfalls textbasiert ausgewertet. Da die subjektiv-auditive Bewertung für den gesamten Text erfolgte, wurden auch für jedes prosodische Merkmal alle pro Wort bzw. Aufnahmeabschnitt berechneten Werte über die gesamte Aufnahme gemittelt.

Die vom Laryngograph-System gemessenen Frequenzunterschiede von jeweils zwei aufeinanderfolgenden Schwingungszyklen werden durch den  $CF_x$ -Wert repräsentiert [2]. Der  $CQ_x$ -Wert bezeichnet das prozentuale Verhältnis zwischen der Dauer des Stimmlippenschlusses (Quasigeschlossenzeit) und der Gesamtdauer eines Zyklus bei aufeinanderfolgenden Schwingungszyklen. Beide Irregularitätsparameter wurden in der vorliegenden Studie verwendet.

Mithilfe der Support-Vektor-Regression (SVR) wurde schließlich aus allen gemessenen Werten die aussagekräftigste Kombination bestimmt und ein Vorhersagewert für die durchschnittliche menschliche Bewertung des jeweiligen Patienten berechnet.

## Ergebnisse

Die durchschnittlichen „Noten“ der 19 Bewerter für die 58 Stimmproben lagen bei 0,88 (R), 0,59 (B) und 0,81 (H). Die berechneten Korrelationswerte  $r$  lauten:

	R	B	H
Inter-Rater-Korrel. (ein Bewerter gegen Mittelwert der übrigen)	0,65	0,58	0,59
Mensch-Maschine-Korrelation (prosodische Merkmale)	0,66	0,36	0,36
Mensch-Maschine-Korrelation (pros. Merkm. + Laryngograph)	0,71	(0,36)	0,53

Die beste Merkmalsmenge war für alle Bewertungskriterien ähnlich. Sie enthielt u.a. die Länge eines Wort-Pause-Wortintervalls und damit einen Hinweis auf Sprechanstrengung bzw. -geschwindigkeit, sowie den Mittelwert des Jitter. Für die Rauigkeit war der Mittelwert der  $F_0$ , für die Behauchtheit der  $F_0$ -Wert beim Stimmeinsatz enthalten. Die Position des Stimmausklangs war Teil der besten Menge für das Heiserkeitsmodell. Für Behauchtheit und Heiserkeit spielte außerdem die normierte Energie (in Bezug auf Normalsprecher) eine wesentliche Rolle. Von den Werten des Laryngograph-Geräts war  $CF_x$  in der Merkmalsmenge für die Rauigkeit und  $CQ_x$  in der Menge für Heiserkeit enthalten. Für die Behauchtheitsmodellierung brachten weder  $CF_x$  noch  $CQ_x$  einen Vorteil.

## Diskussion und Fazit

Die Korrelation zwischen der automatischen Bewertung und dem Durchschnitt der 19 Stimmbewerter liegt in dieser Studie für Rauigkeit und Heiserkeit im selben Bereich wie die Korrelation zwischen einem zufällig ausgewählten Hörer und dem Rest der Gruppe. Folglich ist die maschinelle Bewertung gleichwertig zur menschlichen und stellt eine wertvolle Ergänzung in der Stimmdiagnostik dar. Die Behauchtheitsbewertung konnte mit den vorhandenen Messwerten nicht zufriedenstellend modelliert werden. Die Einbeziehung der Werte  $CF_x$  und  $CQ_x$  des Laryngograph-Systems können die Mensch-Maschine-Korrelationen wesentlich verbessern, jedoch nicht für die Behauchtheit. Die berechneten Korrelationswerte sind für den klinischen Einsatz noch zu gering, jedoch bleibt zu bedenken, dass auch die Bewerter untereinander nicht besser übereinstimmen.

## **Danksagung**

Die Arbeit wurde von der Else Kröner-Fresenius-Stiftung (Fördernr. 2011\_A167) gefördert.

## **Literatur**

[1] Bocklet T, Toy H, Nöth E, Schuster M, Eysholdt U, Rosanowski F, Gottwald F, Haderlein T. Automatic Evaluation of Tracheoesophageal Substitute Voice: Sustained Vowel versus Standard Text. Folia Phoniatr Logop 2009;61:112-6.

[2] Fourcin A. Aspects of Voice Irregularity Measurement in Connected Speech. Folia Phoniatr Logop 2009;61:126-36.

[3] Ptok M, Schwemmle C, Iven C, Jessen M, Nawka T. Zur auditiven Bewertung der Stimmqualität. HNO 2005; 54:793-802.