

HNO 2005 · 54:793–802
 DOI 10.1007/s00106-005-1310-6
 Online publiziert: 17. August 2005
 © Springer Medizin Verlag 2005

Redaktion
 M. Ptok, Hannover

M. Ptok¹ · C. Schwemmler¹ · C. Iven² · M. Jessen² · T. Nawka³

¹ Klinik und Poliklinik für Phoniatrie und Pädaudiologie,
 Medizinische Hochschule Hannover

² Studiengang Logopädie – Fachbereich Gesundheit,
 Europa-Fachhochschule Fresenius, Idstein

³ Abteilung für Phoniatrie und Pädaudiologie, Univ.-Klinik
 für HNO, Kopf- und Halschirurgie, Greifswald

Zur auditiven Bewertung der Stimmqualität

Nach der Definition der Union Europäischer Phoniater ist „die Heiserkeit [...] eine ausschließlich durch anatomische Veränderungen der Stimmlippen, durch Ungleichmäßigkeiten der Stimmlippenränder, durch die Irregularität der Stimmlippenschwingungen oder durch Insuffizienz des Stimmritzenschlusses zustande gekommene pathologische Klangveränderung der Stimme, die im akustischen Bild durch Aperiodizität der Grundfrequenz und der harmonischen Obertöne bzw. durch Turbulenzgeräusche charakterisiert ist“ (zit. n. [25]).

Zu dieser Definition ist anzumerken, dass anatomische Veränderungen der Stimmlippen und Ungleichmäßigkeiten der Stimmlippenränder als ursächliche, ätiologische Kriterien aufzufassen sind, und Irregularitäten der Stimmlippenschwingungen und Insuffizienzen des Stimmritzenschlusses die Pathogenese sowohl organischer als auch funktioneller Ursachen der Heiserkeit kennzeichnen.

Kriterien der Stimmqualität

Unabhängig von einer solchen Diskussion ist „Heiserkeit“ ein Merkmal für die Stimmqualität. Sie ist relativ einfach auditiv wahrnehmbar und klassifizierbar [1]. Von Isshiki u. Takeuchi [17] wurde vorgeschlagen, zwischen den auditiv unterscheidbaren Komponenten der Heiserkeit, nämlich Rauigkeit, Behauchtheit, Asthenie und Grad der Heiserkeit, zu unterscheiden. Takahashi u. Koike [36] kamen

dagegen aufgrund ihrer Untersuchungen zu dem Schluss, dass die Begriffe „rau“, „behaucht“, „asthenisch“ und „Grad der Heiserkeit“ willkürlich gewählt und irreführend seien.

In Japan und im US-amerikanischen Raum hat sich trotzdem die GRBAS-Skala [8, 12, 13, 14, 15, 16] durchgesetzt: G: „overall grade or degree“ (Grad der Heiserkeit), R: „rough“ (Rauigkeit), B: „breathy“ (Behauchtheit), A: „asthenic“ (Asthenie), S: „strained quality“ (Spannung).

In einer Untersuchung von Nawka, in der eine Hörergruppe normale und gestörte Stimmen nach der GRBAS-Skala zu beurteilen hatte, stellte sich vor allem die Qualität asthenisch (A) in der praktischen Verwendung als problematisch heraus. Sie korrelierte in hohem Maße mit dem Eindruck der Behauchtheit (B), sodass sie nicht als autonome Beurteilungsdimension angesehen werden konnte; ebenso streuten die Hörerurteile zur „Spannung“ im Gegensatz zu denen der übrigen 3 Qualitäten stark [27].

Für die auditive Einschätzung der Heiserkeit hat sich die Beschränkung auf 3 Qualitäten als praktikabel erwiesen [1]: R: Rauigkeit; B: Behauchtheit; H: Heiserkeit. Die Ausprägung dieser Merkmale wurde auf 4 Grade von 0–3 verteilt. Dabei steht der Wert 0 für normal, 1 für geringgradig, 2 für mittelgradig und 3 für hochgradig. Auf diese Weise wurden sehr gut übereinstimmende Hörerurteile gewonnen, die auch nach langen Zeiträumen gut reproduzierbar waren.

Eine Studie, in der 11 Gruppen aus 3 Ländern mit unterschiedlicher professioneller Erfahrung in der Stimmbeurteilung den Grad der Heiserkeit (G, im Deutschen mit H für Heiserkeitsgrad oder Gesamtheiserkeitsgrad benannt) zu bewerten hatten, zeigte sich eine weitgehende Übereinstimmung der Urteile von professionell geschulten Hörern und Laien sowie Beurteilern unterschiedlicher Nationalität [40].

Zur Problematik von Stimmaufnahmen

Typischerweise wird die Heiserkeit einer Stimme während fortlaufender Sprache bestimmt. Darin sind die Vokale erheblich kürzer (50–300 ms) und durch Transienten eingefasst. Die auditive Beurteilung ist bei einer Unterteilung in 4 Grade ausreichend differenziert, um Unterschiede zu erfassen, die durch externe Kriterien (z. B. vor und nach Therapie) definiert werden. Untersuchungen der Sprechstimme vor und nach stimmverbessernden Operationen haben gezeigt, dass statistisch eine hochsignifikante Verbesserung des Stimmklangs postoperativ erzielt wurde, wenn als Bewertung die auditive Heiserkeitsklassifizierung zugrunde lag. Die akustischen Analysen von Jitter und Shimmer der Grundfrequenz der Sprechstimme wiesen dagegen eine weit weniger deutliche Verringerung als erwartet auf [28].

Die Stimmproben der gehaltenen Phonation von Vokalen eignen sich bei

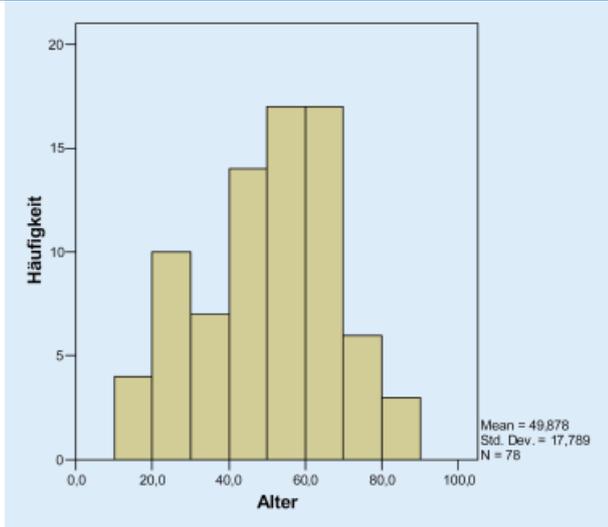


Abb. 1 ◀ Altersverteilung der untersuchten Patienten

ausreichender Dauer (mindestens 2–3 s) sehr gut für Periodizitätsanalysen, die Bewertung der Heiserkeit ist aber besonders vom Anlaut geprägt. Die auditive Einschätzung der Heiserkeit von Sprechstimme und Vokal wird deshalb nur vergleichbar, wenn der gesamte Vokal mit Anlaut bewertet wird [33, 34]. Die fortlaufende Sprache eignet sich somit besser zur Heiserkeitsbeurteilung [11].

In der phoniatischen Routine werden typischerweise zur auditiven Stimmeinschätzung phonetisch ausbalancierte Standardtexte verwendet, die laut gelesen und aufgenommen werden.

Der hier verwendete Text „Der Nordwind und die Sonne“ (Äsop¹-Fabel) ist als Standardtext – auf der Grundlage von Empfehlungen der Arbeitsgruppe Akustik in der Union der Europäischen Phoniater (UEP) – in der phoniatischen Praxis allgemein bekannt und akzeptiert [4]. Es handelt sich um eine Kurzgeschichte mit 109 Wörtern. Sie erfordert eine durchschnittliche Lesezeit von etwa 45 s.

Ziele der Arbeit/Hypothesen

Ziel der Arbeit war es, die Zuverlässigkeit im Sinne der Interrater-Reliabilität der auditiven Einschätzung der Stimmqualität durch die 3 Kategorien Rauigkeit, Behauchtheit und Gesamtheiserkeit (RBH-Skala) an Stimmproben mit fortlaufender Sprache zu überprüfen. Nicht überprüft

werden sollte die Intrarater-Reliabilität oder die Validität.

Es wurde, im Gegensatz zu anderen Studien, bei denen gehaltene Vokale untersucht wurden (z. B. [19, 22]), bewusst eine Leseprobe gewählt, um eine hohe Realitätsnähe zu den alltagsrelevanten Stimmansforderungen zu gewährleisten.

In dieser Studie wurde anhand der Textaufnahmen von 78 Patienten die Reliabilität der RBH-Bewertung durch 19 Beurteiler berechnet.

Es wurde angenommen, dass trotz fehlender quantitativer Definitionen die Heiserkeit ein Merkmal der Stimme ist, das auditiv übereinstimmend wahrgenommen und ordinal skalierbar eingeschätzt werden kann.

Es sollte überprüft werden, ob es Einzelbeurteilungen gibt, die vom gemittelten Wert aller anderen Bewertungen so abweichen, dass sie als „Ausreißer“ gewertet werden müssen.

Material und Methoden

Studientyp

Experimentelle Reliabilitätsstudie; verblindete Bewertung von aufgezeichneten Stimmproben; Berechnung des gemittelten Gruppenurteils; Berechnung des α -Koeffizienten, Berechnung der Reliabilitäts-Gesamt-Item-Statistik.

Stichprobe

Für die Untersuchung wurden Stimmproben von Patienten, die sich in der Zeit von

Januar 2004–März 2004 in der Abteilung Phoniatrie/Pädaudiologie vorstellten, verwendet. Die Stimmen wurden im Rahmen von Routineuntersuchungen für die Diagnostik aufgezeichnet. Kam ein Patient mehrfach während dieses Zeitraums zur Untersuchung, wurde jeweils nur die erste Stimmprobe verwendet.

Es handelt sich um 21 männliche und 57 weibliche Patienten im Alter von 12,2–89,2 Jahren (Altersverteilung s. **Abb. 1**).

Die Auswahl der Stimmproben sollte möglichst realitätsnah sein, deshalb wurde bewusst eine querschnittliche Auswahl über einen definierten Zeitraum gewählt und keine weitere Selektion vorgenommen. Somit wurden einerseits gestörte Stimmen, andererseits auch „normale“ Stimmen (z. B. Nachuntersuchungen nach erfolgreich abgeschlossener Therapie) inkludiert.

Folgende klinische Diagnosen wurden gestellt:

Ödeme:

Reinkeödem bds. 3×
Randkantenödem 1×

Paresen:

Rekurrensparese links 8×
Rekurrensparese rechts 15×
Rekurrensparese beidseits 2×

Gutartige Tumoren,

Pseudotumoren:

Raumforderung der Stimmlippe rechts 1×
Stimmlippenpolyp 6×
Stimmlippenzyste 1×
Stimmlippenknötchen 3×
Stimmlippengranulom 1×
Papillomatose 1×

Entzündungen:

Laryngitis 7×

Zentrale Bewegungsstörungen:

Spasmodische Dysphonie 2×
Balbuties 3×
Sonstige zentrale laryngeale Bewegungsstörung 1×

Funktionelle Dysphonie:

Dysphonie ohne erkennbares morphologisches Korrelat 23×

¹ griechischer Fabeldichter um 550 v. Chr.

Datenmaterial/Aufzeichnung

Für die Off-line-Untersuchung wurde nur die aufgezeichnete Leseprobe „Der Nordwind und die Sonne“ verwendet.

Die Aufzeichnungen erfolgten mit dem Laryngograph-System bzw. dessen Systemkomponenten. Insbesondere wurde das mit dem Laryngograph-System mitgelieferte Headset-Mikrofon verwendet und generell auf einen Mund-Mikrofon-Abstand von 10 cm eingestellt. Auf die Verwendung eines höherwertigen Mikrophons wurde bewusst verzichtet, um bei den Auswertungen möglichst praxisnahe Bedingungen zu haben.

Auditive Stimmbewertung

Die auditive Beurteilung der nicht verwürfelten Stimmproben erfolgte durch 19 Studenten des Studiengangs Logopädie der Europa-Fachhochschule Fresenius in Idstein. Die Studenten wurden im Rahmen des regulären Unterrichts im Fach Stimme mit den theoretischen Hintergründen und dem veröffentlichten Audiomaterial des RBH-Systems [26] vertraut gemacht. An einem separaten, späteren Termin wurden die zu beurteilenden Stimmproben im Unterrichtsraum durch eine hochwertige Audioanlage dargeboten. Die Studenten hatten ihre Beurteilungen in einem Antwortbogen einzutragen. Sie wurden instruiert, jeweils alle 3 Stimmqualitäten gemäß ihrem auditiven Eindruck auf einer Skala von 0–3 zu beurteilen.

Obwohl den Studenten die RBH-Konvention, nach der die Heiserkeit nicht geringer bewertet werden soll als die Rauigkeit oder Behauchtheit, aus dem Unterricht bekannt war, wurde diese Konvention zum Zeitpunkt der Untersuchung nicht wiederholt, sondern stattdessen betont, dass die Beurteilung ohne weiteres Nachdenken rein auf Grundlage des Wahrnehmungseindrucks zu erfolgen hatte. Ein vergleichbares Verfahren, bei dem die rein auditive Beurteilung ohne bewusstes Beachten der RBH-Konvention zugrunde lag, wurde bereits in einer Reihe von Stimmuntersuchungen am Institut für Phonetik und Phonologie der Universität des Saarlandes verwendet

HNO 2005 · 54:793–802 DOI 10.1007/s00106-005-1310-6
© Springer Medizin Verlag 2005

M. Ptok · C. Schwemmler · C. Iven · M. Jessen · T. Nawka
Zur auditiven Bewertung der Stimmqualität

Zusammenfassung

Hintergrund. In der phoniatischen Praxis wird bei Dysphoniepatienten routinemäßig auditiv die Heiserkeit bewertet. In Deutschland hat sich hierfür die RBH-Skala etabliert, nach der in einem vierstufigen System die Rauigkeit R, die Behauchtheit B und die Heiserkeit H beurteilt wird. Obwohl diese Skala verbreitet ist, gibt es wenige Daten zur Reliabilität der auditiven Bewertung der Sprechstimme beim Lesen des Textes „Der Nordwind und die Sonne“, der mittlerweile als Standard in der phoniatischen Praxis gilt.

Material und Methoden. 78 Patienten lasen den phonetisch ausbalancierten Text „Der Nordwind und die Sonne“. Die Stimmqualität der Sprechstimme wurde in Anlehnung an die RBH-Methode von 19 Studenten des Studienganges Logopädie beurteilt. Für diese Beurteilungen wurde die Interrater-Reliabilität bestimmt.

Ergebnisse. Die Berechnungen ergaben eine sehr hohe Interrater-Reliabilität mit Cronbach- α von 0,94, die Gesamt-Item-Statistik zeigte, dass es bei den Bewertungen keine „Ausreißer“ gab.

Diskussion. Die hier erarbeiteten Daten lassen keine Aussage zur Validität bzw. der Interrater-Reliabilität der RBH-Methode zu, da lediglich die Interrater-Reliabilität untersucht wurde. Diese Ergebnisse bestätigen aber die wenigen bisher vorhandenen Angaben in der Literatur, dass die RBH-Skala gut für die auditive Einschätzung der Stimmqualität geeignet ist. Deshalb sollte, auch zur Qualitätssicherung, diese Bewertung bei allen Stimmpatienten vorgenommen werden.

Schlüsselwörter

Dysphonie · Rauigkeit · Heiserkeit · Behauchtheit · RBH-Skala

On the auditory evaluation of voice quality

Abstract

Background. For routine clinical purposes, dysphonic voices are assessed perceptually using the GRBAS scale or analogues. For clinical application, the crucial question is the interrater reliability (IRR) of the auditory perceptual assessment of voice quality. Therefore, the IRR of the four point RBH (roughness, breathiness, hoarseness vs overall grade) scale was studied. Other parameters, e.g. validity and intrarater reliability were not considered.

Methods. A total of 78 patients read a standard text “Der Nordwind und die Sonne”. These samples were evaluated by 19 speech and voice therapy students according to the degree of roughness, breathiness and hoarse-

ness. Data were subjected to reliability analysis.

Results. Our data indicate a high IRR with a Cronbach's α of 0.94. No single rating of the 19 raters could be omitted without decreasing the IRR.

Discussion. The data indicate that the perceptual assessment of hoarseness for running speech is highly reliable. The application of the RBH scale is suitable for clinical purposes. It should be considered as an outcome measure.

Keywords

Dysphonia · Roughness · Hoarseness · Breathiness · RBH-Scale

Tab. 1 Korrelationen zwischen den gemittelten Werten der auditiven Bewertung von Rauigkeit R, Behauchtheit B und Heiserkeit H

		B_Mittel	H_Mittel
R_Mittel	ρ	0,252(*)	0,466(**)
	Signifikanz (2-seitig)	0,026	0,000
	n	78	78
B_Mittel	ρ	1,000	0,708(**)
	Signifikanz (2-seitig)		0,000
	n	78	78

** Die Korrelation ist auf dem 0,01 Niveau signifikant (zweiseitig).

* Die Korrelation ist auf dem 0,05 Niveau signifikant (zweiseitig).

[18, 30, 31]. Die Studenten wurden nicht über die Diagnosen bzw. den klinischen Status informiert.

Statistik

Für die auditiven Bewertungen wurde die Interrater-Reliabilität nach Cronbach- α geprüft. Dann wurde die Korrelation (Rangkorrelationskoeffizient nach Spearman) zwischen den Subskalen ermittelt und die Gesamt-Item-Statistik berechnet.

Die Auswertung erfolgte durch das Statistikprogramm SPSS (Statistical Package for Social Sciences, Version 12.0).

Datenschutz

Die durchgeführte Untersuchung orientiert sich hinsichtlich des Datenschutzes an den rechtlichen Grundsätzen und Empfehlungen, die vom Berliner Datenschutzbeauftragten für den Bereich „Bildung und Wissen“, insbesondere empi-

risch-medizinische Sozialforschung und epidemiologische Studien im universitären Bereich, ausgearbeitet wurden (vgl. zur aktuellen Fassung: DatenschutzWeb-Index, über: <http://www.datenschutz-berlin.de>).

Ergebnisse

Interrater-Reliabilität

Die Interrater-Reliabilität für den Rauigkeits-, den Behauchtheits- sowie für den Gesamtheiserkeitsgrad ergab ein Cronbach- α von 0,938 für die R-Werte, von 0,961 für die B-Werte und von 0,924 für die H-Werte.

Die Gesamt-Item-Statistik (Tab. 1) zeigte eine hohe Übereinstimmung der verschiedenen Urteile. Lediglich bei der Heiserkeitsbewertung ergaben sich 2 „Ausreißer“, H6 und H10. Würde man

diese Urteile nicht berücksichtigen, hätte sich Cronbach- α noch verbessert.

Korrelationsanalysen

Bei der Berechnung des Korrelationskoeffizienten Spearman- ρ ergaben sich signifikante Zusammenhänge zwischen allen gemessenen Variablen (gemittelte Werte, s. Tab. 1). Hierbei lagen die Korrelationskoeffizienten für den gemittelten B-Wert mit dem gemittelten H-Wert relativ hoch ($\rho=0,708$). Rauigkeit und Behauchtheit korrelierten nur schwach ($\rho=0,252$).

RBH-Verteilung

Die Beurteiler haben die Urteile in den 3 Dimensionen nicht homogen abgegeben. Diesen Sachverhalt demonstrieren Abb. 2, Abb. 3 und Abb. 4. Je „höher“ die einzelnen Balken sind, desto „uneinheitlicher“ fielen die Urteile aus: So waren sich die Beurteiler für Patient Nr. 19 z. B. hinsichtlich der Rauigkeit sehr einig, während für Patient Nr. 17 offensichtlich die Meinungen zur Rauigkeit sehr unterschiedlich waren.

Diskussion

Die auditive Stimmbewertung hat sich in der phoniatriischen Praxis als subjektives

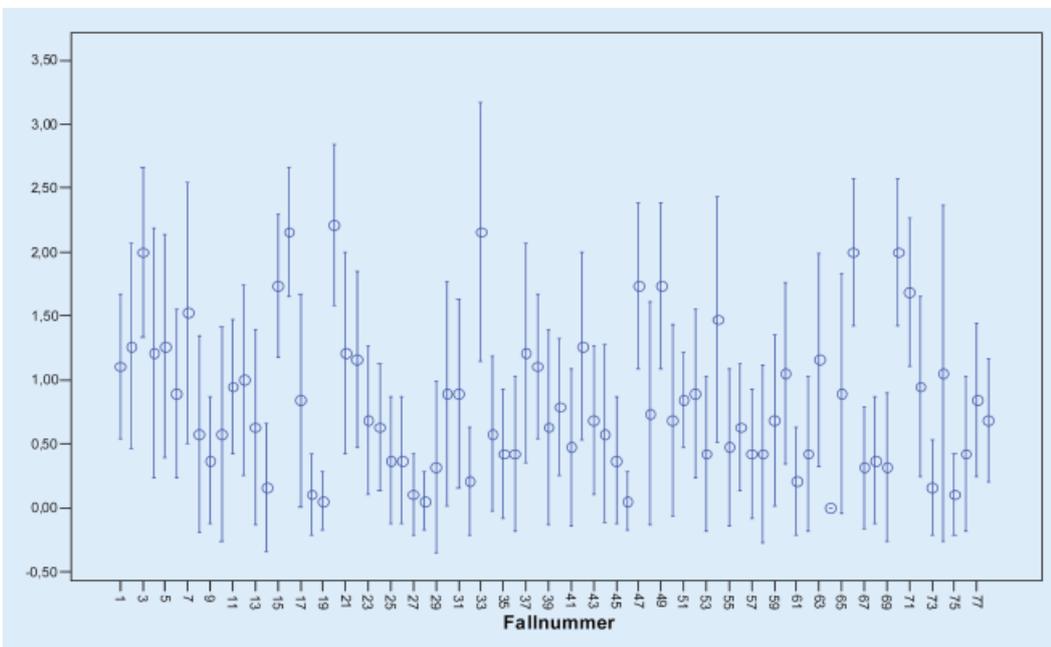


Abb. 2 ◀ Darstellung der Mittelwerte der Rauigkeitsbeurteilungen \pm Standardabweichung (durch die „-“ Abtragung ergeben sich auch die Werte kleiner 0). Je größer die Standardabweichung, desto weniger „einig“ waren sich die Beurteiler

Hier steht eine Anzeige.



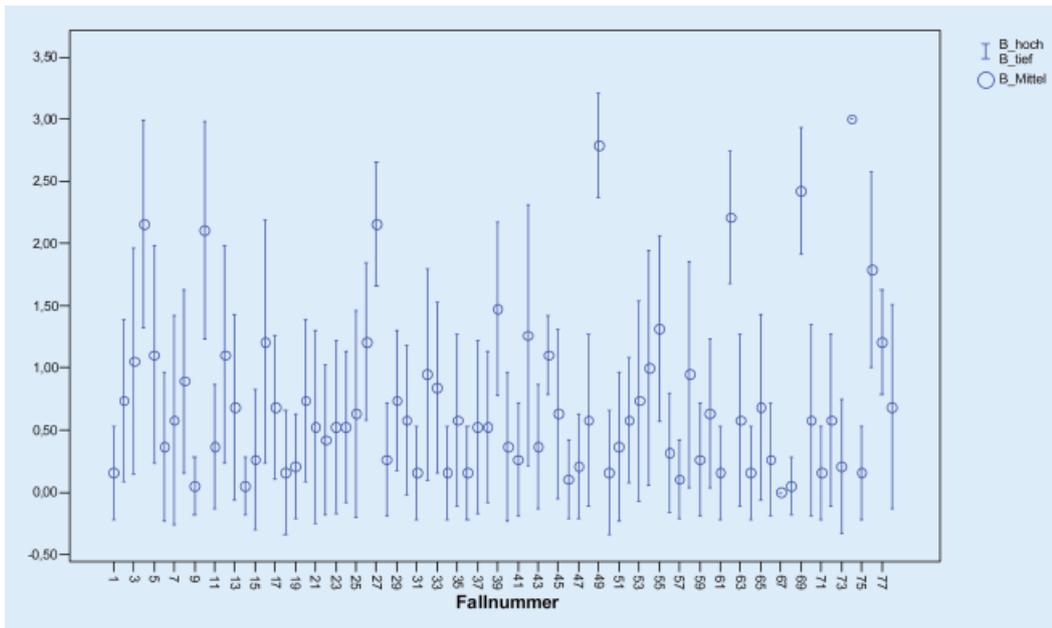


Abb. 3 ◀ Darstellung der Mittelwerte der Behauchtheitsurteilungen ± Standardabweichung

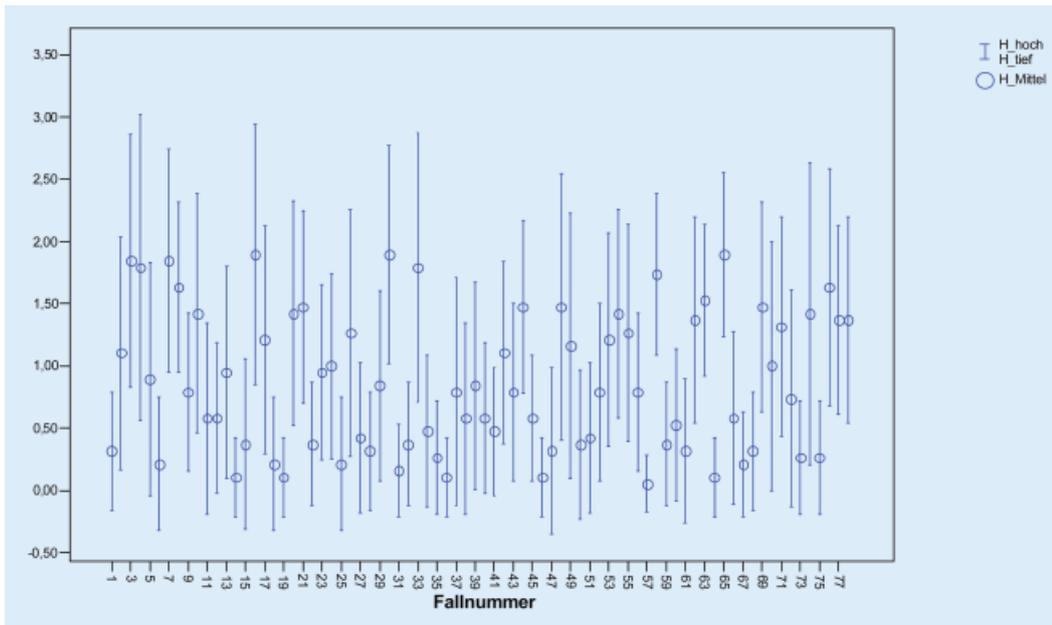


Abb. 4 ◀ Darstellung der Mittelwerte der Heiserkeitsurteilungen ± Standardabweichung

diagnostisches Verfahren seit Jahren international bewährt (z. B. [2, 5, 6, 29, 37, 39, 41]). Von verschiedenen Gruppen wurde die Zuverlässigkeit der Rauigkeits-, Behauchtheits- und der globalen Heiserkeitsbewertung betont (■ **Tab. 2, 3, 4**). Eine Differenzialdiagnose, d. h. ein Rückschluss auf zugrunde liegende Erkrankungen wie z. B. die Unterscheidung zwischen entzündlichen und neoplastischen Veränderungen, war nicht möglich.

Reliabilitätsanalysen zeigten eine hohe Übereinstimmung zwischen verschiedenen Bewertungen, die von professionellen, aber auch von nichtprofessionellen

Stimmeurteilern abgegeben wurden. Allerdings muss eine 4-stufige Bewertung vorgenommen werden [39]. Auch die Übereinstimmung zwischen den Urteilen von professionellen versus nichtprofessionellen Stimmeurteilern kann als zufriedenstellend bis sehr gut bezeichnet werden [3].

Visuelle Analogskalen

Eine stufenlose Unterteilung des Ausprägungsgrades wird auf visuellen Analogskalen (VAS) vorgenommen. Auch für diese Art der Stimmeurteilung wurden

sehr hohe Reliabilitätswerte gefunden. In einer Studie aus 3 Ländern (England, Dänemark, Deutschland) und 4 Zentren beurteilten 11 Personen den gleichen deutschen Text „Der Nordwind und die Sonne“ bei 156 Aufnahmen mit 142 verschiedenen Stimmen von deutschen Muttersprachlern. Cronbach- α lag bei 0,96 [7]. Interessant an dieser Arbeit ist, dass die dänischen und englischen Hörer die Heiserkeit in der für sie fremden deutschen Sprache genauso empfinden wie die deutschen. Trotz der hohen Übereinstimmung in dieser Untersuchung mit der VAS ist die einfachere Klassifikation in

4 Schweregrade vorzuziehen, weil sie besser mittelbar ist.

Sehr viel kritischer sehen Kreiman u. Garratt die Reliabilität [9, 10, 19, 21, 22, 23, 24, 32] und sehen durch eine schlechte Reliabilität sogar die Validität gefährdet [20].

Die hier vorgestellten Daten bestätigen die hohe Interrater-Reliabilität: sie lag sowohl für die Rauigkeits-, Behauchtheits- wie die gesamte Heiserkeitsbewertung sehr hoch, obwohl bei einigen Stimmproben deutlich unterschiedliche Beurteilungen abgegeben wurden.

Dieser scheinbare Widerspruch liegt darin begründet, dass die Interrater-Reliabilität als Schätzwert für den unbekannt wahren Wert der Merkmale R, B und H den Mittelwert von allen Urteilen verwendet. Weiterhin ist für Cronbach- α charakteristisch, dass es ein adjustiertes Reliabilitätsmaß ist. Damit werden die unterschiedlichen Mittelwertstendenzen der einzelnen Beurteiler eliminiert [38].

Vergleicht man die hier vorgestellten Ergebnisse mit den Ergebnissen von Koreman et al. [18], die die Interrater-Reliabilität bei gehaltenen Vokalen untersuchten, kann geschlossen werden, dass die Beurteilungen von fortlaufender Sprache stärker als die von gehaltenen Vokalen übereinstimmen. Das stützt das Vorgehen von Nawka et al. [26, 27].

Hierbei trug jede Beurteilerin zu der hohen Reliabilität bei, d. h. es konnten lediglich bei der Heiserkeitsbeurteilung 2 „Ausreißer“ identifiziert werden.

Bewertung der Heiserkeit

Die Analyse der Einzeldaten zeigt allerdings auch, dass durch das hier eingesetzte methodische Vorgehen „Heiserkeit“ als eine eigene und den beiden anderen Stimmqualitäten gleichwertige perzeptuelle Kategorie und keine übergeordnete oder „kognitiv errechnete“ Kategorie erachtet wurde.

Ähnlich hierzu meinen Pützer et al. [31], dass Stimmen durchaus den auditiven Eindruck des Behauchtheits erwecken können, ohne zwangsläufig als heiser charakterisiert werden zu müssen. Sie diskutieren, dass die Dimension Heiserkeit nicht notwendigerweise als ein den beiden anderen Dimensionen „übergeord-

Tab. 2 Gesamt-Item-Statistik für die Rauigkeitsbewertung

	Skalenmittelwert, wenn Item weggelassen	Skalenvarianz, wenn Item weggelassen	Korrigierte Item-Skala-Korrelation	Quadierte multiple Korrelation	Cronbach- α , wenn Item weggelassen
R1	14,38	99,149	0,719	0,679	0,933
R2	14,46	102,563	0,695	0,676	0,933
R3	14,82	104,409	0,698	0,646	0,933
R4	14,83	105,621	0,628	0,598	0,935
R5	14,63	106,678	0,506	0,459	0,937
R6	15,09	105,070	0,652	0,604	0,934
R7	14,42	103,728	0,638	0,594	0,934
R8	14,50	104,799	0,689	0,665	0,934
R9	14,63	102,938	0,695	0,690	0,933
R10	14,76	105,252	0,557	0,591	0,936
R11	14,79	103,074	0,700	0,613	0,933
R12	15,06	107,385	0,558	0,551	0,936
R13	15,06	105,152	0,681	0,680	0,934
R14	15,06	104,970	0,694	0,688	0,934
R15	14,79	102,659	0,756	0,663	0,932
R16	14,90	107,184	0,496	0,434	0,937
R17	14,69	102,242	0,661	0,741	0,934
R18	14,87	104,165	0,666	0,784	0,934
R19	15,08	106,747	0,570	0,622	0,936

In der ersten Spalte sind die verschiedenen Beurteiler benannt (z. B. bedeutet R1: Rauigkeitsbewertung der ersten Beurteilerin); die Berechnungen zeigen, wie sich die Gesamtwerte ändern würden, wenn man die Bewertung der jeweiligen Beurteilerin nicht berücksichtigen würde (z. B. in der ersten Zeile, wie sich die Gesamtwerte ändern würden, wenn man die R-Bewertungen der ersten Beurteilerin nicht berücksichtigen würde). Cronbach- α von allen 19 Bewertungen lag bei 0,938. Jede Bewertung trägt zu einer Erhöhung des Gesamt-Cronbach- α bei, selbst wenn sie eine niedrige Korrelation zum Gesamtwert hat. Ließe man z. B. die Beurteilerin R16 weg, weil sie eine Korrelation zu den anderen unter 0,5 hat, würde Cronbach- α von 0,938 auf 0,937 sinken.

netes Urteil“ und die beiden anderen Dimensionen als „untergeordnete Faktoren“ bei der Bewertung des Stimmklangs allgemein betrachtet werden müssen.

Die geringe Korrelation zwischen R und B weist darauf hin, dass es sich um unabhängig wahrnehmbare Komponenten der Heiserkeit handelt. Zu ähnlichen Ergebnissen kamen Koreman et al. [18]. Die hohe Korrelation der H-Werte mit den B-Werten deutet an, dass sich die Beurteiler bei der Bewertung des Gesamtheiserkeitsgrades am Behauchtheitsgrad orientiert haben. Für H wurde auch in dieser Studie kein Argument für eine unabhängige gleichrangige Bewertung erbracht. Die übergeordnete Kategorie Gesamteindruck der Heiserkeit H behält ihre Berechtigung.

Wie erwähnt, stehen die guten Reliabilitätswerte im Gegensatz zu den Ergebnissen von Kreiman u. Garratt (s. oben). Als Erklärung hierfür bieten sich einer-

seits an, dass Kreiman u. Garratt überwiegend die Phonation gehaltener Vokale untersuchten. Außerdem sahen auch wir das von dieser Arbeitsgruppe angesprochene Phänomen der offensichtlich sehr schwierig zu kategorisierenden Stimmen. Auch die Verwendung der statistischen Methode war unterschiedlich. Kreiman u. Garratt bestimmten den Intraklassenkorrelationskoeffizient für mehrere Beurteiler über eine zweifaktorielle Varianzanalyse [35]. Dieser unjustierte Wert registriert intraindividuell verschobene Maßstäbe (ein Beurteiler urteilt regelmäßig strenger als der andere) als Zeichen für fehlende Übereinstimmung.

Bei Betrachtung der Werte fällt auf, dass deutlich die Hälfte der Mittelwerte unter dem Wert 1 und nur sehr wenige Mittelwerte über dem Wert 2 liegen. Dies ist deshalb überraschend, weil es sich um Stimmproben von Patienten handelte, bei denen auch stärkere Heiserkeits-

Tab. 3 Gesamt-Item-Statistik für die Behauchtheitsbewertung. Das Gesamt-Cronbach- α lag bei 0,961

	Skalenmittelwert, wenn Item weggelassen	Skalenvarianz, wenn Item weggelassen	Korrigierte Item-Skala-Korrelation	Quadrierte multiple Korrelation	Cronbach- α , wenn Item weggelassen
B1	12,67	142,017	0,693	0,616	0,960
B2	12,74	144,765	0,593	0,546	0,961
B3	13,35	140,047	0,847	0,847	0,958
B4	13,37	140,237	0,839	0,846	0,958
B5	13,50	145,032	0,658	0,613	0,960
B6	13,41	141,336	0,820	0,783	0,958
B7	12,73	140,771	0,679	0,613	0,960
B8	12,91	138,213	0,782	0,851	0,958
B9	12,54	145,550	0,585	0,465	0,961
B10	13,18	138,227	0,805	0,880	0,958
B11	13,27	139,238	0,836	0,802	0,958
B12	13,14	140,512	0,723	0,653	0,959
B13	13,15	136,521	0,825	0,834	0,958
B14	13,19	140,287	0,770	0,821	0,959
B15	13,12	142,259	0,708	0,715	0,959
B16	12,99	141,337	0,623	0,572	0,961
B17	12,99	137,727	0,757	0,720	0,959
B18	13,12	140,129	0,776	0,833	0,958
B19	12,95	138,127	0,738	0,803	0,959

Erläuterungen analog zu Tab. 2.

Tab. 4 Gesamt-Item-Statistik für die Heiserkeitsbewertung. Das Gesamt-Cronbach- α lag bei 0,924

	Skalenmittelwert, wenn Item weggelassen	Skalenvarianz, wenn Item weggelassen	Korrigierte Item-Skala-Korrelation	Quadrierte multiple Korrelation	Cronbach- α , wenn Item weggelassen
H1	14,96	104,271	0,563	0,613	0,921
H2	15,12	106,129	0,497	0,405	0,923
H3	15,79	104,503	0,635	0,650	0,920
H4	15,73	103,290	0,714	0,688	0,918
H5	15,95	107,686	0,481	0,442	0,923
H6	15,97	109,194	0,337	0,462	0,926
H7	15,41	99,518	0,712	0,751	0,918
H8	15,37	104,652	0,627	0,633	0,920
H9	14,96	104,453	0,574	0,599	0,921
H10	15,91	110,576	0,266	0,533	0,927
H11	15,73	106,407	0,552	0,606	0,922
H12	15,55	102,510	0,632	0,552	0,920
H13	15,50	99,552	0,810	0,840	0,915
H14	15,60	99,515	0,726	0,801	0,917
H15	15,65	101,788	0,712	0,660	0,918
H16	15,77	104,959	0,516	0,464	0,922
H17	15,85	103,924	0,587	0,633	0,921
H18	15,58	101,780	0,688	0,779	0,918
H19	15,67	100,433	0,725	0,803	0,917

Erläuterungen analog zu Tab. 2. Die Bewertung durch H6 und H10 „verschlechterte“ das Gesamt-Cronbach- α .

grade vorliegen können. Eine Erklärung ist, dass „naive“ Beurteiler eher zu „konservativen“ Bewertungen mit Mittentendenzen neigen. Unabhängig von der absoluten Einstufung in die Grade 0, 1, 2 oder 3, zeigt jedoch die hohe Interrater-Reliabilität, dass die Schweregrade der Heiserkeit und ihrer beiden bewerteten Komponenten differenziert werden konnten.

Stichprobenauswahl

Zu den hier vorgelegten Daten kann kritisch angemerkt werden, dass es sich nicht um Stimmproben handelte, die über alle „Schweregrade“ von R, B und H gleichmäßig verteilt waren, also für jede Kategorie von R, B und H je 20 Stimmproben. Eine solche Stichprobe auszuwählen, ist allerdings unrealistisch, weil es kein Außenkriterium gibt – und somit auch keine Verifizierung möglich ist, dass „richtig“ ausgewählt wurde. Alternativ hätte man eine Stichprobe nehmen können, bei der als „Außenkriterium“ für die „Richtigkeit“ der Einstufung eine allgemein akzeptierte subjektive Beurteilung vorliegt. Dies hätte allerdings durch die Selektion an sich kaum den klinischen Alltag widerspiegelt.

Weiterhin könnte kritisch angemerkt werden, dass die Stimmproben nicht verwürfelt dargeboten wurden und somit ein Sukzessivkontrast nicht ausgeschlossen werden kann. Hierzu kann allerdings entgegnet werden, dass die Stimmproben chronologisch nach Besuchstermin in der Klinik „geordnet“ waren, also keinesfalls nach Stimmqualitätskriterien. Der zweifellos vorhandene Sukzessivkontrast führt zu einem systematischen Fehler, der allerdings die Interrater-Reliabilität nicht beeinflusst.

Die „anchor stimuli“, wie von Gerratt u. Kreiman beschrieben, können die Streuung der Urteile im Bereich mittlerer Störungsgrade vermindern. In unserer Studie wurde aber die Reliabilität der Gruppenurteile, also der Mittelwerte aus allen Bewertungen, untersucht und nicht eine absolute (wahre, unbekannt) Klassifizierung postuliert. Das erklärt die gefundenen hohen Werte.

In der hier vorgelegten Studie wurde nur die Interrater-Reliabilität überprüft. Für weitere Studien ist es wünschenswert,

auch andere Maße der Reliabilität wie die Intrarater-Reliabilität oder die Test-Retest-Reliabilität zu überprüfen.

Fazit für die Praxis

Die hier vorgelegten Daten belegen, dass sich die RBH-Skala zur auditiven Bewertung der Heiserkeit mit ihren Teilaspekten Rauigkeit und Behauchtheit prinzipiell gut eignet. Allerdings besteht das Problem der „schwierigen Stimmen“, d. h. einzelner Stimmen, bei denen verschiedene Beurteiler zu unterschiedlichen Ergebnissen kommen und bei denen somit die Interrater-Reliabilität gering ist. Für die Praxis ist es wichtig zu bedenken, dass die aktuell zu beurteilende Stimme tatsächlich eine solche „schwierige Stimme“ sein könnte. Leider gibt es noch kein Kriterium, dies im Voraus abzuschätzen. Außerdem werden Aspekte der Stimmqualität wie Durchdringungsfähigkeit, Brillanz, Schönheit etc. mit dieser Skala nicht erfasst.

Korrespondierender Autor

Prof. Dr. Dr. h.c. M. Ptok

Klinik und Poliklinik für Phoniatrie und Pädaudiologie, Medizinische Hochschule Hannover
Carl-Neuberg-Straße 1, 30625 Hannover
ptok.martin@mh-hannover.de

Danksagung. Den beiden anonymen Gutachern sowie Herrn M. Jessen sei für konstruktive Kritik und Diskussionsbereitschaft nachdrücklich gedankt.

Literatur

- Anders LC, Hollien H, Hurme P, Sonninen A, Wendler J (1988) Perception of hoarseness by several classes of listeners. *Folia Phoniatri Logopaed* 40: 91–100
- Carding P, Carlson E, Epstein R, Mathieson L, Shewell C (2000) Formal perceptual evaluation of voice quality in the United Kingdom. *Logoped Phoniatri Vocol* 25: 133–138
- De-Bodt MS, Wuyts FL, Van-de-Heyning PH, Croux C (1997) Test-retest study of the GRBAS scale: influence of experience and professional background on perceptual rating of voice quality. *J Voice* 11: 74–80
- DeJonckere PH, Crevier-Buchman L, Marie JP, Moerman M, Remacle M, Woisard V (2003) Implementation of the European Laryngological Society (ELS) basic protocol for assessing voice treatment effect. *Rev Laryngol Otol Rhinol (Bord)* 124: 279–283
- Dejonckere PH, Obbens C, de-Moor GM, Wieneke GH (1993) Perceptual evaluation of dysphonia: reliability and relevance. *Folia Phoniatri Logopaed* 45: 76–83
- Dejonckere PH, Remacle M, Fresnel-Elbaz E, Woisard V, Crevier-Buchman L, Millet B (1996) Differentiated perceptual evaluation of pathological voice quality: reliability and correlations with acoustic measurements. *Rev Laryngol Otol Rhinol (Bord)* 117: 219–224
- Evans R, Nawka T, Gong Y, Gluud C (2004) Auditive Stimmbewertung nach dem CAPE-V-Protokoll in einer multizentrischen Studie. Jahrestagung der DGPP, <http://www.egms.de/en/meetings/dgpp2004/04dgpp75.shtml>
- Fujiu M, Hibi SR, Hirano M (1988) An improved technique for measurement of the relative noise level using a sound spectrograph. *Folia Phoniatri Logopaed* 40: 53–57
- Gerratt BR, Kreiman J (2001) Measuring vocal quality with speech synthesis. *J Acoust Soc Am* 110 (5 Pt 1): 2560–2566
- Gerratt BR, Kreiman J, Antonanzas-Barroso N, Berke GS (1993) Comparing internal and external standards in voice quality judgments. *J Speech Hear Res* 36: 14–20
- Halberstam B (2004) Acoustic and perceptual parameters relating to connected speech are more reliable measures of hoarseness than parameters relating to sustained vowels. *ORL J Otorhinolaryngol Relat Spec* 66: 70–73
- Hiki S, Imaizumi S, Hirano M, Matsushita H, Kakita Y (1976) Acoustical analysis for voice disorders. *Proc IEEE Icassp: Rome*, 613–616
- Hiki S, Kakita Y, Hirano M, Matsushita H (1976) Acoustic analysis for voice disorders. A basic conception for the use of acoustic measurements for the diagnosis in voice disorders. *Prakt Otol (Kyoto)* 69: 267–271
- Hiki S, Kakita Y, Hirano M, Matsushita H, Imaizumi S (1977) Correlation between parameters related to vocal cord vibration and acoustical parameters in voice disorders. *Prakt Otol (Kyoto)* 70: 393–403
- Hirano M (1989) Objective evaluation of the human voice: clinical aspects. *Folia Phoniatri Logopaed* 41: 89–144
- Imaizumi S, Hiki S, Hirano M, Matsushita H (1980) Analysis of pathological voices with a sound spectrograph. *J Acoust Soc Jap* 36: 9–16
- Isshiki N, Takeuchi Y (1970) Factor analysis of hoarseness. *Studia Phonologica* 5: 37–44
- Koreman J, Pützer M, Just M (2004) Correlates of varying vocal fold adduction deficiencies in perception and production: methodological and practical considerations. *Folia Phoniatri Logopaed* 56: 305–320
- Kreiman J, Gerratt BR (1996) The perceptual structure of pathologic voice quality. *J Acoust Soc Am* 100: 1787–1795
- Kreiman J, Gerratt BR (1998) Validity of rating scale measures of voice quality. *J Acoust Soc Am* 104 (3 Pt 1): 1598–1608
- Kreiman J, Gerratt BR (2000) Sources of listener disagreement in voice quality assessment. *J Acoust Soc Am* 108: 1867–1876

Hier steht eine Anzeige.



22. Kreiman J, Gerratt BR, Berke GS (1994) The multi-dimensional nature of pathologic vocal quality. *J Acoust Soc Am* 96: 1291–1302
23. Kreiman J, Gerratt BR, Precoda K (1990) Listener experience and perception of voice quality. *J Speech Hear Res* 33: 103–115
24. Kreiman J, Gerratt BR, Precoda K, Berke GS (1992) Individual differences in voice quality perception. *J Speech Hear Res* 35: 512–520
25. Müller R (1995) Heiserkeit. *Ther Umsch* 52: 759–762
26. Nawka T, Anders LC (1996) Die auditive Bewertung heiserer Stimmen nach dem RBH-System – Doppel-Audio CD mit Stimmbeispielen. Thieme, Stuttgart
27. Nawka T, Anders LC, Wendler J (1994) Die auditive Beurteilung heiserer Stimmen nach dem RBH-System. *Sprache Stimme Gehör* 18: 130–133
28. Nawka T, Franz T, Voigt M, Cebulla M (1996) Beurteilung von Ergebnissen nach indirekter phono-chirurgischer Mikrolaryngoskopie. In: Gross M (Hrsg) Aktuelle phoniatisch-pädaudiologische Aspekte. Renate Gross, Berlin, S 82–84
29. Nunez-Batalla F, Corte-Santos P, Sequeiros-Santiago G, Senaris-Gonzalez B, Suarez-Nieto C (2004) Evaluacion perceptual de la disfonia: correlacion con los parametros acusticos y fiabilidad. *Acta Otorrinolaringol Esp* 55: 282–287
30. Pützer M, Barry WJ (2004) Methodische Aspekte der auditiven Beurteilung von Stimmqualität. *Sprache Stimm Gehör* 28: 188–197
31. Pützer M, Masarek K (2000) Differenzierung gesunder Stimmqualitäten und Stimmqualitäten bei Rekurrensparese mit Hilfe elektrolottographischer Messungen und RBH-System. *Sprache Stimme Gehör* 24: 154–163
32. Rabinov CR, Kreiman J, Gerratt BR, Bielamowicz S (1995) Comparing reliability of perceptual ratings of roughness and acoustic measure of jitter. *J Speech Hear Res* 38: 26–32
33. Revis J, Giovanni A, Triglia JM (2002) Influence of voice onset on the perceptual analysis of dysphonia. *Folia Phoniater Logopaed* 54: 19–25
34. Revis J, Giovanni A, Wuyts F, Triglia J (1999) Comparison of different voice samples for perceptual analysis. *Folia Phoniater Logopaed* 51: 108–116
35. Shrout PE, Fleiss JL (1979) Intraclass correlations: uses in assessing rater reliability. *Psych Bull* 86 (2): 420–428
36. Takahashi H, Koike Y (1976) Some perceptual dimensions and acoustical correlates of pathologic voices. *Acta Otolaryngol Suppl* 338: 1–24
37. Webb AL, Carding PN, Deary IJ, MacKenzie K, Steen N, Wilson JA (2004) The reliability of three perceptual evaluation scales for dysphonia. *Eur Arch Otorhinolaryngol* 261: 429–434
38. Wirtz M, Caspar F (2002) Beurteilerübereinstimmung und Beurteilerreliabilität. Hogrefe Verlag für Psychologie, Göttingen
39. Wuyts FL, De-Bodt MS, Van-de-Heyning PH (1999) Is the reliability of a visual analog scale higher than an ordinal scale? An experiment with the GRBAS scale for the perceptual evaluation of dysphonia. *J Voice* 13: 508–517
40. Yamaguchi H, Shrivastav R, Andrews ML, Niimi S (2003) A Comparison of Voice Quality Ratings Made by Japanese and American Listeners Using the GRBAS Scale. *Folia Phoniater Logopaed* 55: 147–157
41. Yu P, Revis J, Wuyts FL, Zanaret M, Giovanni A (2002) Correlation of instrumental voice evaluation with perceptual voice analysis using a modified visual analog scale. *Folia Phoniater Logopaed* 54: 271–281

J. Hirschberg, M. Gross
Velopharyngeale Insuffizienz mit und ohne Gaumenspalte

Diagnostik und Theorie der Hypernasalität

Heidelberg: Median-Verlag 2006, 239 S., (ISBN 978-922766-82), 75.00 EUR

Wer sich bisher mit Ätiologie, Diagnostik und Therapie des offenen Naselns infolge einer velopharyngealen Insuffizienz (VPI) etwas eingehender beschäftigen wollte, musste sich den gegenwärtigen Wissensstand mühsam aus vielen Büchern und Einzelpublikationen zahlreicher Fachgebiete wie Hals-Nasen-Ohren-Heilkunde, Kiefer-, Mund- und Gesichtschirurgie, Phoniatrie und Pädaudiologie, Neurologie, Pädiatrie, Logopädie etc. zusammensuchen. Jetzt haben zwei sehr renommierte Autoren – Jenő Hirschberg aus Budapest und Manfred Gross aus Berlin – eine umfassende Synopsis der klinischen und theoretischen Aspekte der velopharyngealen Insuffizienz zusammengestellt. Von der ersten Zeile an merkt man dem Buch an, dass es sich auf eine große klinische Erfahrung stützt, insbesondere auf die 43jährige Erfahrung des ungarischen Autors, die er an über 5.000 Patienten sammeln konnte. Nach kurzem historischem Rückblick schildern die Autoren Anatomie und Physiologie des velopharyngealen Sphinkters und definieren die VPI als dessen Unfähigkeit, den optimalen Verschluss zwischen Oro- und Nasopharynx zustande zu bringen. Es folgen epidemiologische Daten, die Klassifikation der VPI und eine ausführliche Darstellung der vielfältigen ätiologischen Möglichkeiten. Besonders hervorzuheben ist dabei die ausführliche und klare Darstellung neurogener Bewegungsstörungen des Gaumensegels, wie auch ein detailliertes Eingehen auf die velopharyngealen Funktionsstörungen im Rahmen von Fehlbildungssyndromen und ihre genetischen Aspekte. Hör-, Stimm- und Sprachstörungen im Rahmen der VPI werden ausführlich besprochen mit besonderer Berücksichtigung der gegenwärtigen Funktionstheorie der Tuba auditiva. Weiterhin erhält der Leser einen interessanten und umfassenden Überblick über die diagnostischen Möglichkeiten einschließlich der elektroakustischen, elektro-mechanischen und elektromyographischen Untersuchungsverfahren, deren Einsatzmög-

lichkeit kritisch bewertet werden. Im letzten, therapeutischen Teil der Buches werden die chirurgischen Möglichkeiten zur Beseitigung der VPI zusammengestellt und die konservativen Behandlungsnotwendigkeiten einschließlich der Behandlung der Schwerhörigkeit und der prothetischen Maßnahmen zur Sprachverbesserung behandelt.

Das Buch ist in einem sehr flüssigen und angenehm zu lesenden Stil geschrieben und mit zahlreichen instruktiven Abbildungen illustriert. Übersichtliche Tabellen unterstützen den Text, und die Lehrsätze am Ende vieler Kapitel, die als „Fazit für die Praxis“ wichtige Aussagen zusammenfassen, sind didaktisch hervorragend. Das sehr umfangreiche Literaturverzeichnis wird eine große Hilfe vor allem für den wissenschaftlich tätigen Leser sein. Insgesamt wird das Buch dem Anspruch, ein interdisziplinärer Leitfaden für die klinische und wissenschaftliche Beschäftigung mit der VPI sein zu wollen, voll gerecht und ist daher sehr zu empfehlen.

H.-J. Schultz-Coulon (Neuss)