

Aus der
Klinik und Poliklinik für
Hals-Nasen-Ohrenheilkunde,
Universitätsklinikum, Düsseldorf
(Direktor: Universitätsprofessor Dr. U. Ganzer)
und der
Klinik für
Hals-Nasen-Ohrenheilkunde
Kopf- und Halschirurgie, plastische Operationen
Phoniatrie und Pädaudiologie,
Lukas Krankenhaus, Neuss
(Chefarzt: Prof. Dr. med. Hans-Jürgen Schultz-Coulon)

**Zur Anwendbarkeit der Stimmanalysesoftware „Dr.
Speech®“ zur prae- und postoperativen Dokumentation der
Stimmqualität**

Dissertation

**Zur Erlangung des Grades eines Doktors der
Medizin**

Der Medizinischen Fakultät der Heinrich-Heine Universität

Düsseldorf

vorgelegt von

David Kasper

2003

Als Inauguraldissertation gedruckt mit Genehmigung der Medizinischen Fakultät der
Heinrich-Heine-Universität Düsseldorf

gez.: Univ.-Prof. Dr. med. Dr. phil. Alfons Labisch, M.A.

Dekan

Referent: Prof. Dr. Bier

Korreferent: Univ.-Prof. Dr. Steingrüber

Inhaltsverzeichnis

	Seite
1. Einleitung	5
1. 1. Die auditive Bewertung der Heiserkeit	6
1. 2. Die elektroakustische Analyse der Heiserkeit	9
1. 2. 1. Spektralanalytische Methoden	11
1. 2. 2. Periodizitätsanalyse	19
1. 2. 3. Kombinationsverfahren	21
1. 2. 4. Zur Stimmanalyse Software „Dr. Speech ®“	23
1. 3. Fragestellung	26
2. Material und Methode	27
2. 1. Technischer Aufbau	27
2. 2. Probanden und Durchführung der Untersuchung	27
2. 3. Auswahlverfahren und statistische Auswertung	28
3. Ergebnisse	31
3. 1. Ergebnisse der Einzelfall-Analyse	31
3. 1. 1. Fall1: beidseitige Parese des N. recurrens	31
3. 1. 2. Fall2: T1a Karzinom einer Stimmlippe	42
3. 2. Ergebnisse der statistischen Auswertung der Studie	51
3. 2. 1. Statistische Auswertung der Bewertungsergebnisse der subjektiven und objektiven Stimmklangbewertung im RBH-Schema	51
3. 2. 2. Statistischer Vergleich der subjektiven Behauchtheitsbewertung und der gemessenen Werte für die NNE	70
4. Diskussion	76
5. Anhang (Tabellen)	84
6. Literaturverzeichnis	90
Lebenslauf, Danksagung, Zusammenfassung	

Abkürzungen

APQ	Amplitude Pertubation Quotient (Shimmer)
Cohens Kappa	Konkordanzkoeffizient nach Cohen (im Zusammenhang auch κ)“
dB	Dezibel
EG	Gesamtenergie
EH	harmonische Energie
ER	Geräuschenergie
F(0)	Grundfrequenz
FFT	Fast Fourier Transformation
GNE	glottis to noise excitation
HNR	harmonic to noise ratio, H/N-ratio
Hz	Hertz
KHz	Kilohertz
LTAS	Long-time-average-spectra (Summationsspektrum)
MSSL	mittlere Sprechstimmlage
NNE	normalized noise energy
p	Irrtumswahrscheinlichkeit
postB	postoperative Behauchtheit
postH	postoperative Heiserkeit
postR	postoperative Rauhigkeit
PPQ	Pitch Pertubation Quotient (Jitter)
praeB	praeoperative Behauchtheit
praeH	praeoperative Heiserkeit
praeR	praeoperative Rauhigkeit
RBH-Schema	Schema zur Heiserkeitsbewertung (R auhigkeit, B ehauchtheit, H eiserkeit)
SD	Standardabweichung
Spearman's R	Korrelationskoeffizient nach Spearman (im Zusammenhang auch R_s)“

1. Einleitung

Die Stimme ist als persönlichstes Ausdrucksmittel des Menschen (Mathelitsch und Friedrich, 1995) durch die Sprache Träger von Information und durch sich selbst unmittelbarer Ausdruck von Gefühl und Stimmung.

In der klinischen Arbeit unterscheidet und beurteilt man verschiedene allgemeine Stimmleistungen. Neben der Beurteilung des Tonhaltevermögens, des Stimmeinsatzes, der mittleren Sprechstimmlage, des Tonhöhenumfangs und der dynamischen Steigerungsfähigkeit ist vor allem die Bewertung des Stimmklanges von Bedeutung, wenn es um Stimmfunktionsdiagnostik und Verlaufskontrolle nach therapeutischer Intervention geht. Meist führt der veränderte Stimmklang, wie z. B. eine bleibende Heiserkeit, und/oder eine Abnahme der stimmlichen Belastbarkeit den Patienten zum HNO-Arzt oder Phoniater. Den Erfolg einer Therapie misst auch der Patient an der Verbesserung des Klanges seiner Stimme und der zunehmenden stimmlichen Belastbarkeit, da beides seine individuellen Ausdrucksmöglichkeiten im Alltag unmittelbar beeinflusst. Deshalb ist im täglichen Umgang mit stimmgestörten Patienten die objektive, qualitative und quantitative Beurteilung des Stimmklanges und der Stimmleistung von großer Bedeutung. Diese Beurteilungsmethode sollte dabei vor allem drei Forderungen erfüllen:

1. zuverlässige Beurteilung des Stimmklanges,
2. für die tägliche Routine zumutbarer zeitlicher und materieller Aufwand,
3. krankensblattfähige Dokumentation der Ergebnisse.

Stimmklangveränderungen im Sinne der Heiserkeit sind ein Hauptsymptom von Stimmstörungen (Friedrich, 1996). Prinzipiell stehen zwei Methoden der Analyse und Bewertung des gestörten Stimmklanges zur Verfügung, einerseits die auditive und andererseits die elektroakustische.

1. 1. Die auditive Bewertung der Heiserkeit

Die Bewertung mit dem Ohr ist in starkem Maße von den Fähigkeiten des Untersuchers abhängig und kann oft fehlerhaft sein (Wendler, 1996), jedoch ist nach Friedrich das geschulte Ohr das sensibelste und letztlich einzig relevante Messinstrument "für die Heiserkeit (Friedrich, 1996).

Da es sehr schwer ist, die Stimme und ihre Klangphänomene nur mit sprachlichen Mitteln unverwechselbar und präzise zu definieren (Friedrich, 1996), begann man schon in den 50er Jahren Skalen zu erstellen, um die Stimmklangbeurteilung zu präzisieren (s.u.).

Bei allen perceptiven Verfahren steht die Abhängigkeit vom Untersucher, und die damit möglichen inter- und intraindividuellen Schwankungen der Beurteilung prinzipiell im Gegensatz zur Forderung der Objektivität. Angesichts großer Variabilität der subjektiven Ergebnisse, wünscht man sich objektive Daten mit geringen Schwankungsbreiten. Obwohl viele Arbeiten eine gute Reliabilität und Validität von perceptiver, d.h. auditiver Heiserkeitsbeurteilung bei erfahrenen Untersuchern (Dejonckere et al., 1993; Fex, 1992) aber auch bei Laien (Anders et al., 1988; Wendler et al., 1986) nachweisen konnten, sprechen andere Studienergebnisse und Übersichten gegen eine zuverlässige auditive Stimmbeurteilung (Kreimann et al., 1992 und 1993; Hammarberg, 1994). Hier spielt die Vorgehensweise bei der Beurteilung und das Beurteilungsschema selbst eine entscheidende Rolle (s.u.).

Der Zeitaufwand für die auditive Stimmklangbeurteilung ist sehr unterschiedlich in Abhängigkeit vom gewählten Beurteilungsschema. Beispielsweise ist die perceptuelle Bewertung von 28 verschiedenen stimmbeschreibenden Parametern "in einer 5-stufigen Skala, wie sie Hammarberg et al. 1980 vorschlägt, sicherlich sehr präzise und von hohem wissenschaftlichen Wert, jedoch in der täglichen Routine einer HNO-Klinik oder Praxis kaum praktikabel (Hammarberg et al., 1980). Die Verlaufsbeobachtung der stimmlichen Entwicklung während oder nach Therapie ist zwar genau, die Befunderhebung benötigt jedoch hohen Zeitaufwand. Als Ergebnis zahlreicher Arbeiten reduzierte Hammarberg die Anzahl der stimmbeschreibenden Parameter , "die routinemäßig bewertet werden sollten, von 28 auf 13 (*Stockholm voice evaluation approach*, Hammarberg et al., 1994). Außerdem, und dies wird auch von anderen Autoren herausgehoben (Dejonckere, 1998; Hirano, 1989), erfolgte die Definition der stimmbeschreibenden Parameter " in mehreren Konsensuskonferenzen (Consensus Model). Hammarberg konnte mit dieser Methode eine hohe inter- und intraindividuelle Reliabilität bei der auditiven Beurteilung erreichen.

Man sieht hieran das Dilemma jeder auditiven Bewertungsmethode. Je mehr Parameter des Stimmklangs routinemäßig beurteilt werden, desto genauer ist zwar die Aussage über die stimmliche Qualität, desto unpraktikabler wird jedoch die Methode in der täglichen Routine. Man wird sich daher auf das Notwendige beschränken müssen, indem man die Frage stellt, welche Stimmeigenschaften bzw. Stimmleistungen für die Alltagsroutine wirklich von Belang sind. Geht es beispielsweise bei mikrolaryngoskopischen Eingriffen um die prä- und postoperative Dokumentation der Stimmqualität, so reicht eine verlässliche und reproduzierbare Heiserkeitsbewertung aus.

Das RBH- Schema

Skalen zur auditiven Heiserkeitsbeurteilung wurden unter anderem von Nessel (1960), Isshiki (1969) und Habermann (1975) vorgestellt. Heutzutage hat sich im deutschsprachigen Raum vor allem das **RBH-Schema**, das von Wendler et al. 1986 vorgeschlagen wurde, als subjektive Beurteilungsmethode in der Praxis bewährt. Im RBH-Schema werden die drei Stimmqualitäten **Rauhigkeit**, **Behauchtheit** und **Heiserkeit** nach Schweregraden in einer 4-stufigen Skala (0 = nicht vorhanden, 1 = geringgradig, 2 = mittelgradig, 3 = hochgradig) klassifiziert. Heiserkeit (H) ist der Basisbegriff. Darunter sind alle Stimmklänge mit pathologischen Geräuschanteilen zu verstehen. Die heisere ‘Stimmqualität wird näher spezifiziert durch die Klassifizierung der Behauchtheit (B; Strömungsgeräusche durch wilde Luft und unerwünschte Turbulenzen infolge eines unvollständigen Glottisschlusses) und die Bewertung der Rauhigkeit (R), d. h. Rauschanteile durch Irregularitäten der Stimmlippenschwingungen (sog. *perturbation*):“ Da Heiserkeit als Basisbegriff verstanden wird, steht ihr Betrag mit dem der Rauhigkeit und Behauchtheit in Zusammenhang. Friedrich schreibt 1996 in seinem Artikel zur Qualitätssicherung in der Phoniatrie: Der Index von H, “ steht somit mit jenen von R, und B, in Beziehung... ,“und fordert folgerichtig, d.h. der Index H, muss zumindest so groß wie jener von R, oder B, sein ... (Friedrich, 1996).

Das RBH-Schema leitet sich aus japanischen Heiserkeitsskalen, die seit Ende der 60er Jahre, unter anderem von Isshiki und Takeuchi (1970) sowie von Hirano (1981) entwickelt wurden und allgemein als **GRBA-** und **GRBAS-Skalen** bezeichnet werden, ab. Hier werden, neben Grad der Heiserkeit (G), Rauhigkeit (R) und Behauchtheit (B), zusätzlich noch die Stärke bzw. Schwäche der Stimme (A=Asthenie) und in der GRBAS-Skala noch die Spannung (S) in einer vierstufigen Skala bewertet.

Obwohl sich das RBH-Schema im klinischen Alltag durchgesetzt hat, ist es auch mit Unsicherheiten behaftet. Die Tatsache, dass im RBH-Schema nur drei Standardqualitäten (Rauhigkeit, etc) vorgegeben sind, zwingt den Untersucher oft, andere akustische Klangeindrücke in diese 3 Qualitäten einzuordnen und engt ihn somit ein (Fex 1992). Zudem sind sie letztlich nicht klar zu definieren, so dass jeder Untersucher damit andere Klangphänomene verbindet (Hirano, 1988). Außerdem ist nicht auszuschließen, dass der Untersucher dazu neigt, den Stimmklang bezüglich der Heiserkeit posttherapeutisch besser bzw. geringgradiger einzustufen. Das könnte auch ein Hinweis darauf sein, dass der Untersucher nicht frei urteilt, sondern unbewusst durch Erwartungen (z. B. erwarteter Therapieerfolg) beeinflusst wird. Weiterhin sind die oft bei der auditiven RBH-Beurteilung verwendeten Zusätze (aphon, nasal, etc.) wenig objektiv und erschweren die Vergleichbarkeit (Ananthapadmanabha und Hirano, 1996).

Die Zuverlässigkeit des RBH-Schemas wird jedoch in zahlreichen Arbeiten bestätigt. So zeigte eine Arbeit von Rabinov et al (1995) eine deutlich höhere inter- und intraindividuelle Zuverlässigkeit von auditiven Rauheitsbeurteilungen gegenüber der Zuverlässigkeit von Jittermessungen verschiedener Stimmanalysesysteme. Auch Nawka (1994, 2000) erwähnt, dass mehrere Test-Retest-Studien für die Zuverlässigkeit des RBH-Systems sprechen, ebenso Anders (1985). Hammarberg erhielt ebenfalls sehr zuverlässige Ergebnisse der auditiven Beurteilung. Von allen Autoren wird betont, dass die Festlegung und Definition der Höreindrücke unerlässlich für die zuverlässige und reproduzierbare Beurteilung ist. Hammarberg hat durch ihr Consensus-Modell sehr zuverlässige Beurteilungen erzielt.

Der zeitliche Aufwand einer RBH-Klassifizierung ist wie oben erwähnt gering, da im klinischen Alltag als Prüfmaterial meist das anamnestische Gespräch verwendet wird. Über die Art des Stimmmaterials bei der auditiven Heiserkeitsbewertung besteht keine einheitliche Meinung, insbesondere, ob es Einfluss auf die Beurteilung hat (Bergbreiter, 1992; Fex, 1992; Friedrich 1996). Fex fordert 1992 fortlaufende Sprache und bestmögliche Aufnahmebedingungen zur perzeptiven Stimmbeurteilung. Hammarberg (1980) hält für die perzeptive Stimmanalyse gehaltene Vokale für zu kurz, da Klangphänomene in fortlaufender Sprache, wie z. B. Stimmeinsätze oder Stimmabbrüche bei der Analyse einzelner Vokale nicht gehört werden können. Sie nimmt als Grundlage für die perzeptive Beurteilung gelesene Standarttexte von 40 sec. Dauer. In der vorliegenden Arbeit wurde die Stimme anhand der spontanen Sprache und an gehaltenen Vokalen beurteilt. Dies hat den Vorteil, dass anhand der gehaltenen Vokale Klangphänomene wie Behauchung besser gehört oder noch einmal überprüft werden können. Zudem fühlt sich der Patient bei der Beurteilung von seiner

spontanen Sprache unbeobachtet und es entsteht keine Prüfungssituation,“ bei der emotionaler Stress die Stimme verändern könnte. Messungen von Schultz-Coulon (1979) zeigten zum Beispiel eine um ca. 50 Hz erhöhte MSSL beim Vorlesen gegenüber spontanem Sprechen.

Im klinischen Alltag hat sich das RBH-Schema bewährt, da es einerseits eine semiquantitative Heiserkeitsgraduierung vornimmt und eine qualitative Charakterisierung des Stimmklangs ermöglicht, und andererseits schnell durchzuführen ist. In der täglichen Arbeit mit Patienten geht es nicht darum, ein möglichst präzises Bild des Stimmklangs zu dokumentieren, sondern darum, Verbesserungen oder Verschlechterungen übersichtlich darstellen zu können. Unerlässlich ist jedoch die regelmäßige Übung und die Diskussion über verschiedene Höreindrücke und deren Definition. Bei der vorliegenden Arbeit wurde die RBH-Beurteilung durch den erfahrenen Untersucher als Goldstandard “für den Vergleich mit der computergestützten RBH-Beurteilung (s.u.) zugrundegelegt.

1. 2. Die elektroakustische Analyse der Heiserkeit

Durch die explosionsartige Entwicklung und Verbreitung der Computertechnik haben sich auch die objektiven und semiobjektiven Verfahren zur elektroakustischen Analyse des Stimmchalls so schnell vermehrt, dass es schwierig ist, hier einen geordneten Überblick zu behalten und vor allem die klinische Relevanz der Verfahren richtig zu beurteilen.

Auch die gestörte menschliche Stimme setzt sich, wie jeder Klang, aus einem Grundton und verschieden stark repräsentierten Obertönen, die ganzzahlige Vielfache des Grundtons sind, zusammen. Der Grundton (F_0) der menschlichen Stimme entspricht der subjektiv empfundenen Tonhöhe und seine Frequenz der Schwingungsfrequenz der Stimmlippen (Schultz-Coulon und Klingholz, 1988). Durch Modulation dieses Stimmlippensignals im veränderlichen Ansatzrohr (Übertragungsfunktion) können Obertonbereiche verschieden stark energetisch angehoben werden (Formanten). Im Spektrum stimmhafter Laute finden sich neben harmonischen Anteilen immer aber auch Geräuschanteile, die zumindest teilweise als akustisches Korrelat der Heiserkeit angesehen werden können (Klingholz, 1986). Heiserkeit entsteht nach Klingholz durch zwei unterschiedliche Störungsformen, die meist gemeinsam

vorliegen. Zum einen durch Schwingungsunregelmäßigkeiten der Glottis die zu einem aperiodischen Glottissignal führen und zum anderen durch Strömungsgeräusche überschüssiger Expirationsluft infolge Glottisinsuffizienz.

Grundsätzlich stehen zwei verschiedene Verfahren zur quantitativen Messung von Geräuschanteilen im harmonischen Stimmklangspektrum zur Verfügung: die *Spektralanalyse* und die *Periodizitätsanalyse* (s.u.). Durch diese Verfahren erhält man eine Fülle elektroakustischer Parameter, die unterschiedliche Informationen über die Stimme und ihren Klang vermitteln.

Diese Parameter eines Stimmsignals sind gut reproduzierbar und sehr präzise. Damit liegen Vorteile gegenüber einer perceptiven Heiserkeitsbeurteilung scheinbar auf der Hand, denn die auditive Heiserkeitsbewertung eines Untersuchers kann einigen Studien zufolge unzuverlässige, inter- und intraindividuelle abweichende Urteile zeigen (Kreimann, 1993; Hammarberg, 1994).

Bei der Interpretation vermeintlich objektiver elektroakustischer Parameter sollte aber nicht außer Acht gelassen werden, dass es eine hohe intraindividuelle Variabilität im Stimmklang eines Patienten gibt (Reker, 1991). Denn auch wenn die Aufzeichnungen der Stimmproben bei definierter Tonhöhe und Lautstärke erfolgen, können diese bezüglich Timbre, Tonhöhen- und Lautstärkenschwankungen, Einsatz- und Absatzverhalten erheblich variieren. Jitter (Frequenzschwankungen zwischen benachbarten Perioden, s.u.) und Shimmer (Amplitudenschwankungen zwischen benachbarten Perioden, s.u.) können z.B. von Faktoren wie persönlicher Stimmung und Dauer des Nachtschlafes (Bough, 1996), sowie von der Tageszeit und der dadurch anderen Stimmbelastung abhängig sein. Schon eine ausgelassene Party am Vorabend der Stimmbewertung mit Alkohol- und Nikotinkonsum (auch passiv), sowie übermäßigem Stimmgebrauch oder auch nur Nervosität vor und bei der Untersuchung beeinflusst die Ergebnisse am nächsten Tag. Ein erfahrener Untersucher ist jedoch, wenn auch nur begrenzt, fähig, solche Faktoren bei der auditiven Beurteilung zu berücksichtigen. Optimal vergleichbare Stimmproben erhält man nur, wenn man bestimmte Rahmenbedingungen (Tageszeit der Aufnahme, Aufnahmeort, 24h-Nikotin- und Alkoholkarenz, Haltung des Patienten, Tonhöhe, Lautstärke, etc.) annähernd gleich hält. Dieses ist unter Routinebedingungen schwierig bis unmöglich. Es darf also sowohl bei der elektroakustischen als auch bei der auditiven Heiserkeitsbewertung nie außer Acht gelassen werden, dass immer nur eine Momentaufnahme der Stimme vorliegt.

Weiterhin ist bei der Objektivität "computergestützter Stimmanalyse die technische Aufnahmenqualität von entscheidender Bedeutung (Klingholz, 1986). Auch hier spielen die

strikt einzuhaltenden Aufnahmebedingungen (Mikrophontyp, Aufnahmesoftware bzw. -geräte, Aufnahmeort, etc.) eine wichtige Rolle. Titze et al. (1993) stellt Abhängigkeiten der Periodizitätsanalyseergebnisse von Mikrophontyp und Abstand fest. Bei der auditiven Beurteilung der Stimme fallen diese Probleme weg. Die auditive Beurteilung anhand von Stimmaufnahmen, seien es gehaltene Vokale oder fließende Sprache, unterliegt allerdings den gleichen Gefahren. Hier spielt beispielsweise die Qualität der Wiedergabe-Lautsprecher bei der Bewertung der Heiserkeit (Titze 1987) eine Rolle.

1. 2. 1. Spektralanalytische Methoden

Mit spektralanalytischen Methoden wurde es als erstes möglich, den Stimmschall in seine Teiltöne bzw. Geräuschkomponenten zu zerlegen und ihre Energien zu messen. Damit konnten verbal nur schwer definierbare Höreindrücke pathologischer Stimmklangphänomene quantifiziert und visuell dargestellt werden. Dadurch wurden sie krankenblattfähig dokumentierbar und erleichtern zudem die stimmlichen Verlaufsbeobachtungen während logopädischer und nach chirurgischer Therapie.

1. Filteranalyse

Reiht man elektrische Schwingkreise (Filter) aneinander, deren Durchlassbereiche sich möglichst nahtlos über den gesamten zu untersuchenden Frequenzbereich erstrecken, fungiert diese Filterkette als Sieb für ein Schallsignal (Siebkette). Die ausgehenden Wechselspannungsamplituden der Filter sind proportional zu den vorhandenen Energien des Stimmschalls im zugehörigen Spektralbereich. So gelingt es ein visuelles Abbild des Stimmklangs zu erstellen. Verschiedenste Verfahren wie z. B. das Visible-Speech-Verfahren (Dudley et Gruenz, 1946) basieren auf der Filteranalyse. Auch für die Langzeitspektrographie war die Filteranalyse zunächst Grundlage, indem die Ausgangssignale der Filter zu einem Amplituden-Frequenz Diagramm summiert wurden (Sedláček, 1967; Frøkjær-Jensen u. Prytz, 1974).

Das Suchtonverfahren nach Grützmacher (1927) basiert ebenfalls auf filteranalytischen Methoden, hat aber den Vorteil, dass Zahl und Amplitude der tatsächlich vorhandenen Teiltöne genau bestimmt werden können. Die oben erwähnte Siebkette kann lediglich den

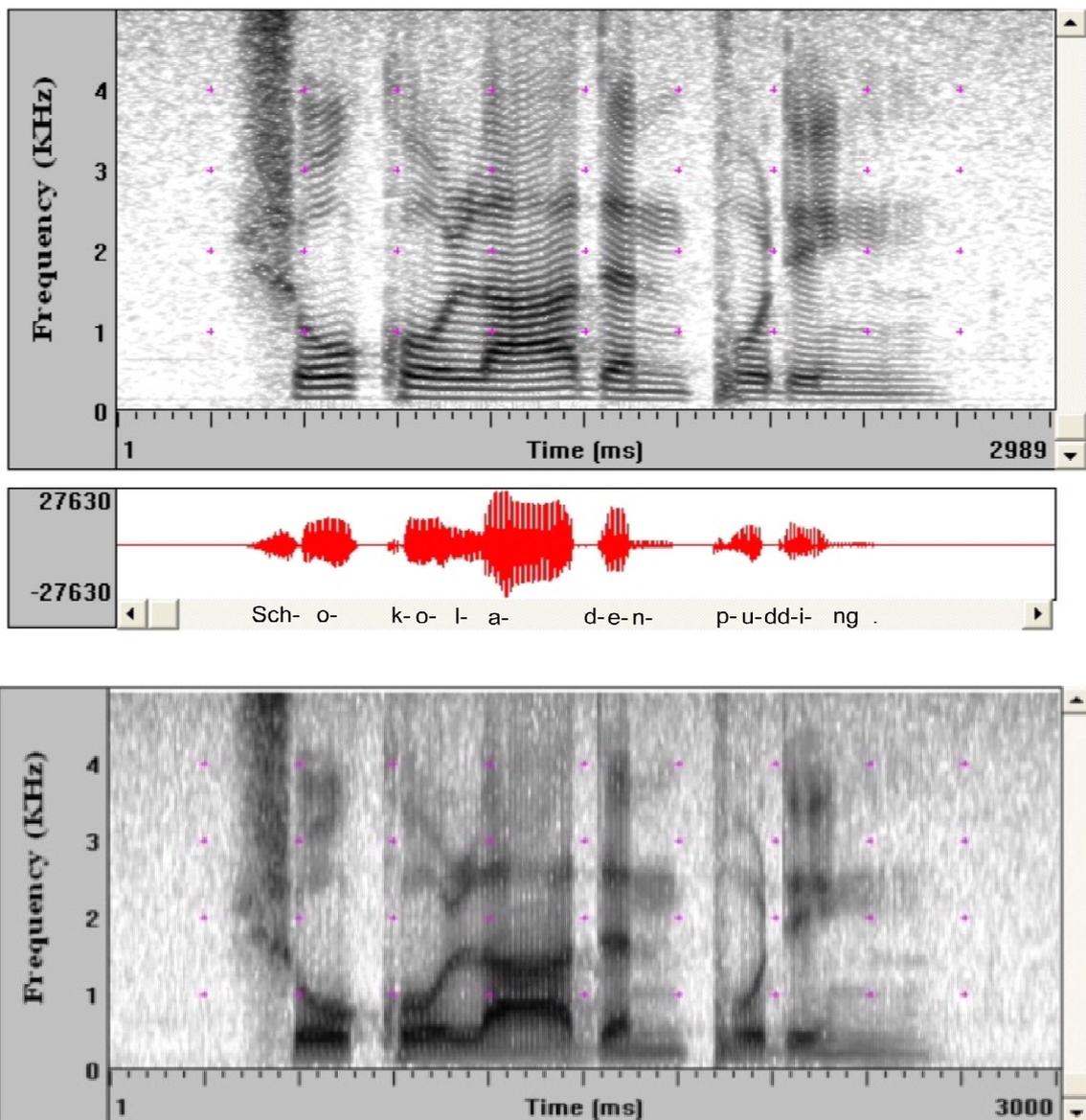
Gesamtenergiegehalt eines bestimmten Spektralbereiches bestimmen. Das Suchtonverfahren ist Grundlage der noch heute zur Stimmanalyse verwandten Sonographie.

Das zu analysierende Stimmsignal wird hierbei von einem Suchton mit langsam steigender Frequenz durchfahren. Trifft der Suchton auf einen Teilton des Stimmsignals so entstehen neben anderen Kombinationstönen auch bestimmte Differenztöne. Das in Wechselspannung verwandelte Signalgemisch (Suchton + Stimmsignal) durchläuft einen Filter, der seinen Durchlassbereich im Bereich des entsethenden Differenztones hat. Am Filterausgang erscheint nur dann eine Spannung, wenn zwischen gesuchtem Teilton und augenblicklicher Suchtonfrequenz ein Differenzton im Filterbereich entsteht. Die Höhe der ausgehenden Filterspannung ist der Teiltonamplitude proportional, die Frequenzlage des Teiltones ist durch den Filterbereich identifiziert. So liefert das Suchtonverfahren bei stimmhaften Lauten ein Linienspektrum, welches die tatsächlichen Teiltonenergien des Stimmsignals widerspiegelt.

Durch eine ausgeklügelte Registrieremethode von Koenig et al. (1946) fand das Suchtonverfahren in Form der Sonographie weite Verbreitung. Das auf einer rotierenden Magnetscheibe aufgezeichnete Stimmsignal (maximal 2,5 s Länge) wird mit einem konstanten Suchton pro Umdrehung der Scheibe durchsucht. „Bei jeder erneuten Umdrehung wird der Suchton um einen bestimmten Betrag erhöht. Mit jeder Suchtonerhöhung verschiebt sich auch der Stift um eine Strecke nach oben. Die Filterausgangsspannung heizt einen Metallstift, der ein Papier auf einer sich drehenden Trommel, je nach Spannungsgröße unterschiedlich schwärzt. Die dadurch entstehenden Sonogramme (spectrogram) geben Aufschluss über Verteilung und Energieverhältnisse der Formanten, sowie deren zeitliche Änderung während des Sprechens (Wirth, 1995).

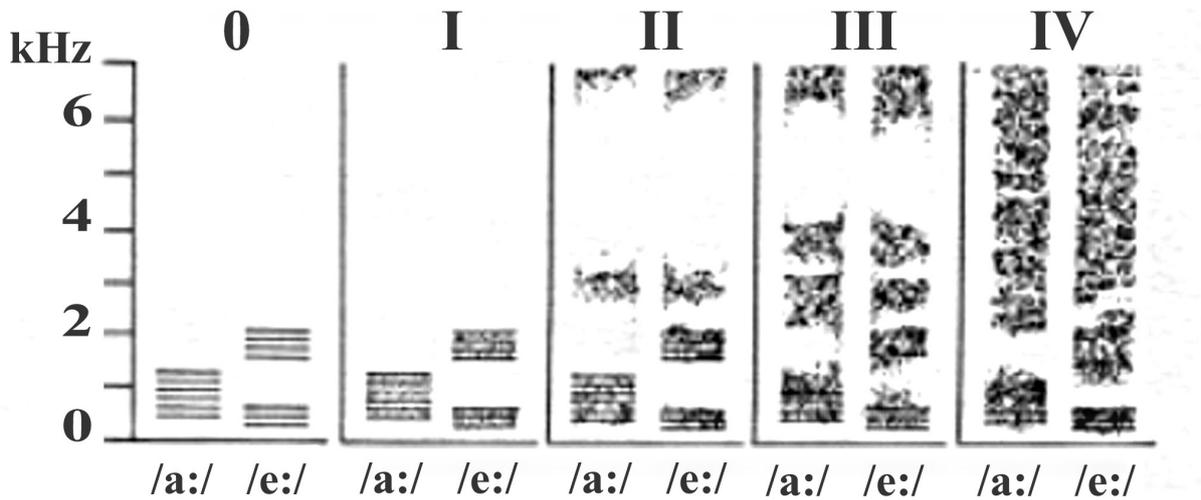
Durch die Digitalisierung des Verfahrens mithilfe der Fast-Fourier-Transformation gibt es heute keine maximale Länge des Aufnahmesignals mehr, sodass auch Analysen längerer Textabschnitte sonographisch möglich sind. Wählt man einen schmalbandigen Filter (Abb.1 oben), erreicht man aufgrund der besseren Frequenzauflösung eine schärfere Teiltondarstellung. Eine bessere zeitliche Auflösung wird mit breitbandigen Filtern erreicht (Abb.1 unten), jedoch leidet die genaue Teiltondarstellung wegen des umgekehrt-proportionalen Verhältnisses zwischen Bandbreite und Einschwingzeit eines Filters.

Abbildung 1: *Oben:* Schmalband-Sonographie mit Hilfe der Stimmanalysesoftware „Dr.Speech®“ (45 Hz Durchlassbreite) des Wortes „Schokoladenpudding“; *unten:* Breitband-Sonographie (300 Hz)



Die Sonographie eignet sich hervorragend zur semiobjektiven Heiserkeitsbeurteilung (siehe Kapitel 3.1.1.). Dies erkannte schon Yanagihara, der bereits 1967 seine klassischen 4 sonographischen Heiserkeitstypen entwickelte (Abb.2):

Abbildung 2: sonographische Heiserkeitstypen Typ I-IV nach Yanagihara (1967), ergänzt durch Typ 0 nach Silvermann und Zimmer (1978). aus: Die Diagnostik der gestörten Stimmfunktion, Schultz-Coulon, 1980



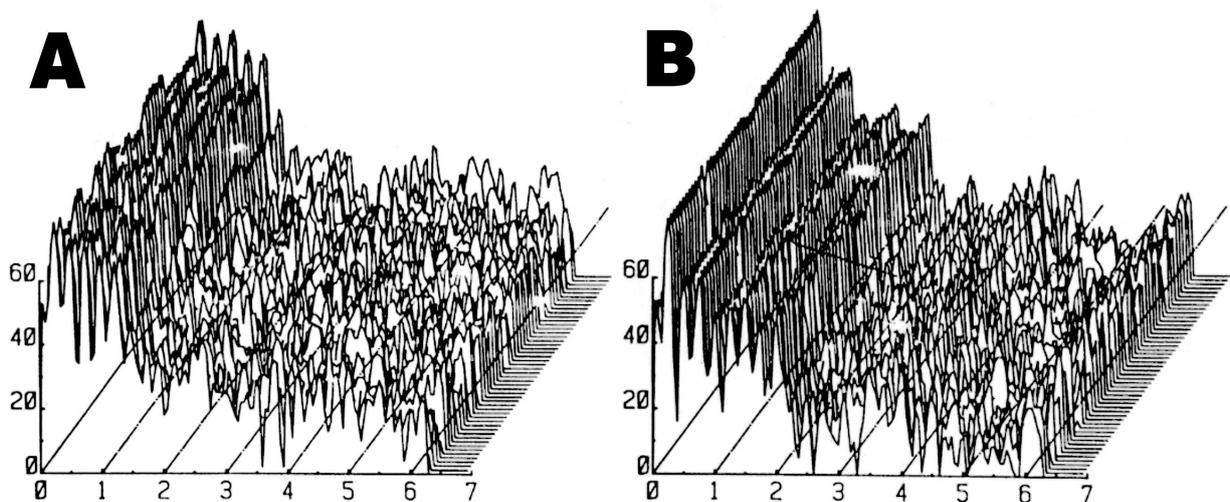
Ausgehend von der Beobachtung, dass ein heiserer Stimmklang durch zusätzliche Geräuschanteile im Stimmsignal bedingt ist (siehe auch Klingholz, 1978), entsprechen leicht überhauchte “Stimmen Typ I, bei dem sich nur schwache Geräuschanteile im Formantenbereich finden. Typ II zeichnet sich zusätzlich durch Geräuschbänder zwischen 3-5 kHz aus. Bei Typ III verbreitern sich die Geräuschbänder derart, dass sie die zweiten Formanten von [e:] und [i:] bereits komplett verwischen, bis schließlich in Typ IV auch die ersten Formanten ihre Periodik und Schärfe verlieren. Für klare, nicht heisere Stimmen ergänzten Silvermann et al. 1978 noch Typ 0.

2. Fourier-Transformation

Mit Hilfe der Fourier-Transformation ist es rechnerisch möglich, jede periodische Schwingungsbewegung in eine Superposition aus einzelnen, verschieden großen harmonischen Sinusschwingungen zu zerlegen. Hierbei wird also die Teiltonstruktur eines Schallsignals durch eine rein rechnerische Spektralanalyse dargestellt. Die unter anderem von

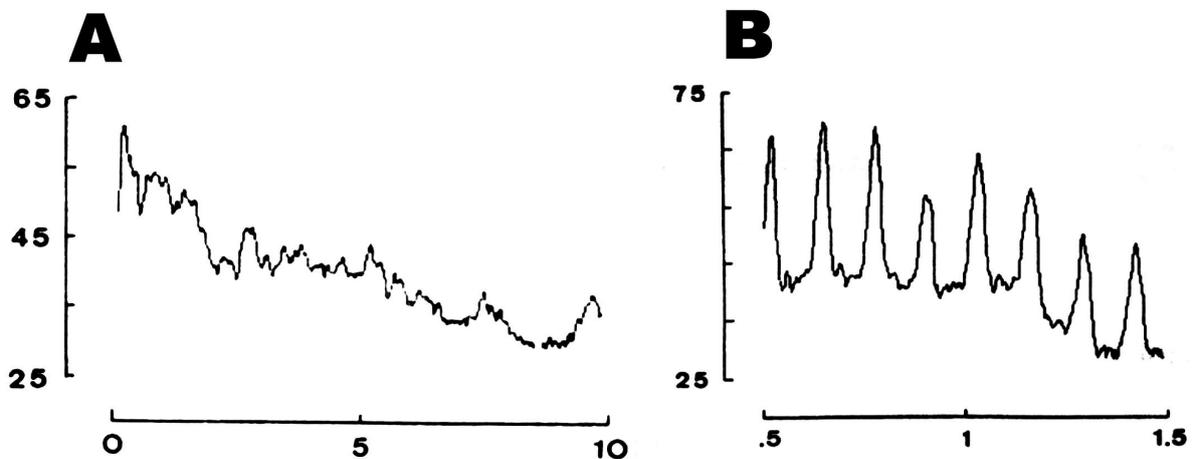
Cooley und Tukey (1965) entwickelten Computer-Algorithmen zur schnellen Fourier-Transformation (Fast-Fourier-Transformation, FFT) bilden die Grundlage der heute gebräuchlichen spektralanalytischen Verfahren. Da annähernd periodische Signale der Stimmanalyse mittels FFT zugrundezulegen sind, konnten zunächst nur sehr kurze Abschnitte des Stimmsignals mit FFT analysiert werden (Kurzeitspektren, Abb. 3).

Abbildung 3: mehrere Kurzeitspektren einer 29-jährigen Patientin mit hyperfunktioneller Dysphonie im dreidimensionalen Amplituden-Frequenz-Zeitdiagramm. A: praetherapeutisch, B: posttherapeutisch; Vokal /a/, ca. 80dB, 1s Signaldauer, Ordinate in dB, Abszisse in kHz. (aus: Objektive und semiobjektive Untersuchungsmethoden der Stimme, H-J Schultz-Coulon und F Klingholz, 1988)



Eine Analyse von längeren Stimmsignalen wurde durch die Summationsspektrographie (LTAS) möglich, die durch Summation von Kurzeitspektren gemittelte Langzeitspektren berechnet, ähnlich den filteranalytischen Langzeitspektren. LTAS verschiedener Autoren oder Analysesysteme sind nicht miteinander vergleichbar, da die Berechnung und Darstellung der LTAS stark von der Methode abhängig ist. Auch macht es große Unterschiede, ob nur stimmhafte Laute erfasst werden oder alle Laute eines Textes (Löfquist u. Manderson, 1987; Abb. 4)

Abbildung 4: A: LTA-Spektrum eines gelesenen Textes, 45s; B: LTA-Spektrum des gehaltenen Vokal /a/, 1s;
Ordinate in dB, Abszisse in kHz (aus: Harmonic emergence in formant zone of a sustained /a/ as a parameter for evaluating hoarseness, P H Dejonckere und J Lebacqz, 1987)



Durch die Analyse von gestörten und gesunden Stimmen mit Kurzzeitspektren und LTA-Spektren in zahlreichen Arbeiten verschiedener Autoren, konnten sich verschiedenen stimmlichen Eigenschaften und Klangphänomenen bestimmte akustische bzw. spektralanalytische Korrelate zuordnen lassen. Hierzu einige Beispiele:

- LTAS-Untersuchungen zeigten, dass sich behauchte Stimmen durch niedrige Amplituden im Bereich von 2-5 kHz im Verhältnis zur Amplitude (Energie) des ersten Formanten auszeichnen. (Hammarberg et al., 1980 ; 1986 ; 1994)
- Im Einklang mit der sonographischen Stimmanalyse zeigen behauchte Stimmen in Summationsspektrogrammen hohe Energien zwischen 5-8 kHz (noise-energy). (Klatt und Klatt, 1990; Hammarberg et al., 1994)
- Übereinstimmend mit den sonographischen Analysen Yanagiharas (1962), konnte Dejonckere (1987 und 1991) eine Korrelation von abnehmender harmonischer Energie im Bereich der ersten beiden Formanten (gehaltener Vokal /a/, LTAS) und zunehmend heiserem Stimmklang nachweisen.
- Die Tragfähigkeit der Stimme drückt sich in der relativen Erhöhung der Amplitude (Frequenzabhängigkeit!) der Teiltöne im 3-kHz-Bereich des Vokalspektrums aus (Sänger-Formant). (Winckel, 1971)
- Die geschulte Stimme unterscheidet sich von der Normalstimme durch eine reduzierte Obertonzahl. Legt man das diatonale System (d.h. Dur-Moll Tonsystem) zugrunde, nehmen die Dissonanzen innerhalb des Teiltonspektrums ab dem 7. Teilton stark zu. Besonders obertonreiche Stimmen klingen deshalb oft rau und scharf. Allzu

obertonarme Stimmen klingen aber matt und glanzlos, weshalb eine bestimmte Anzahl Obertöne (Frequenzabhängigkeit!) für einen sauberen Stimmklang als optimal gelten. (Schultz-Coulon und Wenn, 1986)

- Hypofunktionelle Stimmen können von hyperfunktionellen durch eine relativ hohe Amplitude des Grundtons gegenüber dem ersten Formanten unterschieden werden. Bei hyperfunktionellen Stimmen ist dies umgekehrt. Hier ist die Grundtonenergie deutlich niedriger gegenüber dem ersten Formanten. (Hammarberg et al., 1994)

Um Geräuschkomponenten in stimmhaften Lauten als Korrelat der Heiserkeit quantifizieren zu können, müssen die harmonische Energie (EH) und die Geräuschenergie (ER) aus der Gesamtenergie (EG) des Stimmschalls getrennt voneinander bestimmt werden. Dabei werden die jeweiligen Energien als Verhältnisse ausgedrückt, z. B. Gesamtenergie zu Geräuschenergie (Signal-Rausch-Verhältnis, SRV) oder Harmonischenenergie zu Geräuschenergie (Harmonischen-Rausch-Verhältnis, HRV).

Seit Ende der 60er Jahre sind verschiedene methodische Wege beschrrieben worden um die Geräuschenergie im Stimmsignal auszudrücken. Im Folgenden werden beispielhaft drei heute gebräuchliche Methoden (HNR, NNE und GNE) herausgegriffen.

HNR (harmonic to noise ratio oder H/N ratio)

Obwohl die Methode auch auf Prinzipien der Periodizitätsanalyse, die später erläutert wird, beruht, soll sie an dieser Stelle erläutert werden, da sie sich aus den obigen Überlegungen ergibt. Yumoto et al. entwickelten zu Beginn der 80er Jahre (1982, 1983) dieses Verfahren, um das Verhältnis von harmonischer (EH) und Rauschenergie (ER) auszudrücken. Ausgehend von der Annahme, dass die Harmonischen im Spektrum in gleichmäßigen Abständen folgen (Grundton = f_0 ; 1. Oberton = $2f_0$; usw.) und sich Geräusche im Spektrum als unregelmäßig auftretende Zwischenpeaks zeigen (siehe auch spektrale Rekonstruktion “ von Klingholz, 1987) übertrugen Yumoto et al. dies auf die unzerlegte “ Stimmschallschwingung und ermittelten näherungsweise die Harmonischenenergie (EH oder H), indem sie aus 50 Schwingungszyklen einen Durchschnittsschwingungszyklus errechneten (sog. averaging): “Die Rauschenergie (ER oder N) entsprach nun der Differenz aus einem tatsächlichen Schwingungszyklus und dem Durchschnittszyklus. Das Verhältnis EH/ER bzw. H/N wird in Dezibel ausgedrückt. Zahlreiche Arbeiten ergaben Korrelationen der HNR mit der wahrgenommenen Behauchung und/oder Heiserkeit (Yumoto et al., 1982, 1983, 1984; Zwirner, 1996; Bielałowicz, 1996). Nachteilig wird jedoch die Abhängigkeit von minimalen

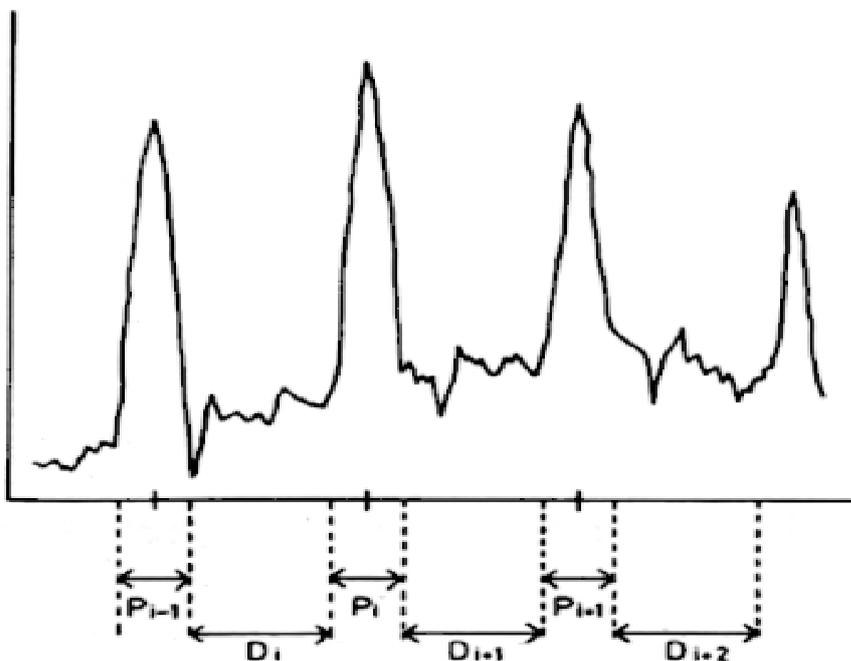
Frequenzschwankungen von Periode zu Periode (Jitter, siehe unten) und Amplitudenschwankungen (Shimmer) gesehen (Shoji et al. 1992; Kasuya, 1986, Fröhlich et al. 1998; Michaelis, 1998).

NNE (normalized-noise-energy)

Sie ist ein Parameter, der das Verhältnis von Geräuschenergie zur Gesamtenergie im Stimmsignal angibt (Signal-Rausch-Verhältnis) (Mitev et al. 2000). Die 1986 von Kasuya et al. (1986a+c) vorgestellte NNE wird durch ein kompliziertes, quasi rein mathematisches Filterverfahren errechnet und drückt das Verhältnis von Geräuschenergie zur Energie des Gesamtsignals üblicherweise im Frequenzbereich von 1-5 kHz aus. Basis der Berechnungen sind FFT-Spektren im Gegensatz zur HNR, der das ursprüngliche Stimmsignal als Grundlage dient.

Ein Zeitfenster von (meist 7) Schwingungsperioden wird mittels FFT spektral zerlegt. Anschließend erfolgt über zahlreiche Algorithmen die Bestimmung der Frequenzbreite und -Lage der harmonischen Maxima (harmonic peak region, P_i) und harmonischen Minima, die Geräuschäquivalenten entsprechen (harmonic dip region, D_i). Nach Filterung des Signals durch ein Filtersieb mit den Bandbreiten von D_i kann die Geräuschenergie berechnet werden und mit der Gesamtenergie ins Verhältnis gebracht werden. Die Einheit ist ebenfalls dB.

Abbildung 5: Identifikation von harmonischen Minima (D_i) und Maxima (P_i) im Stimmspektrum als Grundlage der NNE-Berechnung



Viele Arbeiten untersuchten die Anwendung und Aussagekraft der NNE. Es wurden deutliche Korrelationen mit der wahrgenommenen Behauchung oder Heiserkeit einer Stimme festgestellt (Huang, 1998; Kasuya 1986; Hirano, 1988). Kasuya selbst (1993) stellt fest, dass die NNE besser als die HNR mit der Wahrnehmung pathologischer Stimmen korreliert, und bemängelt die Abhängigkeit der HNR von Jitter und Shimmer. Nach ihm hat die NNE (1986) Vorteile gegenüber der HNR, da sie vom Jitter und Shimmer unabhängiger zu sein scheint. Jedoch zeigen Untersuchungen von Michaelis et al. (1997b) und Zwirner et al. (2000), dass auch die NNE nicht von Jitter und Shimmer unabhängig ist. Auch Mitev et al. (2000) errechneten deutliche Korrelation von Jitter und Shimmer mit der NNE. Da eine Änderung im Schwingungszyklus (Jitter- und Shimmerveränderung) auch immer das Ergebnis der Spektralanalyse beeinflusst, also die Basis der NNE Berechnung, ist eine vollkommene Unabhängigkeit der NNE von Jitter und Shimmer auch nicht zu erwarten.

GNE (Glottal to noise excitation ratio)

Michaelis und seine Mitarbeiter suchten nach einer Möglichkeit, die Rauschenergie der Stimme in einer Weise darzustellen, die unabhängig von Jitter und Shimmer, also von Daten aus der Periodizitätsanalyse sind. Sie stellten 1996 (1996, 1997b) den GNE Faktor vor, der widerspiegelt, inwieweit die Stimmgebung durch Glottisaktivität (glottal) oder durch turbulentes Rauschen (noise) im Stimmkanal angeregt wird, und der somit ein Maß für die Behauchtheit ist (Zwirner, Michaelis, Kruse, 1996). Der GNE-Faktor wird durch komplizierte Algorithmen berechnet und basiert auf Kreuzkorrelationen zwischen den Einhüllenden von verschiedenen Frequenzbändern des Stimmspektrums. Nach Michaelis et al. ist er unabhängig von Jitter oder Shimmer. Arbeiten von Zwirner et al. (1998) konnten eine gute Korrelation des GNE-Faktors mit der wahrgenommenen Behauchung nachweisen. Weiterhin wurde deutlich, dass mit dem GNE Faktor behauchte Stimmen wesentlich besser von rauhen Stimmen zu unterscheiden sind als mit NNE oder HNR.

1. 2. 2. Periodizitätsanalyse

Die *Periodizitätsanalyse*, die sich seit Liebermann Anfang der 60er Jahre kontinuierlich weiterentwickelt hat, hat heute einen festen Platz in der quantitativen Heiserkeitsanalyse. Auch wenn nach Baken (1987) unser stimmbildendes System keine perfekte Maschine ist, und kein Schwingungszyklus dem anderen gleicht, wissen wir jedoch, dass eine Zunahme von solchen Irregularitäten zwischen den einzelnen Schwingungsperioden als zunehmende

Stimmklangverschlechterung wahrgenommen wird (Laver, 1992). Generell geht man davon aus, dass periodische Stimmlippenschwingungen mit periodischen Stimmsignalen einhergehen und zunehmende Unregelmäßigkeiten in den Schwingungszyklen der Stimmlippen zu hörbaren Geräuschanteilen im Stimmschallsignal führen (Laver, 1992; Zwirner 1996). Zu unterscheiden sind vor allem Zeitdifferenzen (Syn: Periodenzeitdifferenzen, Jitter, pitch pertubation) und Amplitudenschwankungen benachbarter Schwingungsperioden (Syn: Shimmer, amplitude pertubation).

Interperiodische Zeitdifferenzen (Jitter)

Es stehen verschiedenste Konzepte zur Quantifizierung des Jitters zur Verfügung. Liebermann (1961) berechnete als erstes die absoluten interperiodischen Zeitdifferenzen und drückte sie in Millisekunden aus. Kittel (1977) drückte die absoluten Frequenzdifferenzen in Herz aus. Weiterhin besteht die Möglichkeit die relativen Zeit- oder Frequenzdifferenzen in Prozent darzustellen (Schultz-Coulon, 1979; Hammarberg, 1986). Die vorliegende Arbeit legt als Maß für den Jitter den pitch pertubation quotient (PPQ) zugrunde. Er drückt die relativen Frequenzschwankungen in Prozent aus und geht auf Arbeiten von Davis et al. (1976, 1978) zurück. Davis konnte zeigen, dass der PPQ die höchste Klassifikationspotenz zur Unterscheidung normaler von pathologischen Stimmen hat. Auch neuere Arbeiten verwenden den PPQ als zuverlässiges Jittermaß (Boyanov et al, 1997 und 1992; Hirano et al 1988; Bough, 1996; Martin et al, 1995). Allgemein gilt der Jitter bzw. PPQ als Maß für die Rauhigkeit.

Interperiodische Amplitudenschwankungen (Shimmer)

Der Shimmer spielt heute gegenüber dem Jitter eine untergeordnete Rolle, da sich der Shimmer in gleicher Richtung wie der Jitter ändert (Klingholz, 1983; Hirano, 1988; Zwirner, 1996), jedoch sein Beitrag am Störungsgrad des Stimmsignals gegenüber dem Jitter gering ist (Klingholz, 1983). Weiterhin ist die zuverlässige Messung des Shimmer von zahlreichen Variablen abhängig (Case, 1999). Klingholz (1986) zeigte zudem, dass der Shimmer pathologische von gesunden Stimmen schlechter trennen kann.

Die Quantifizierungsmethoden entsprechen denen des Jitters, dabei wird der Shimmer entweder in Prozent oder dB ausgedrückt. Die vorliegende Arbeit verwendet den amplitude pertubation quotient (APQ), der analog dem PPQ auf Arbeiten von Davis et al. zurückgeht. Er drückt die relativen Amplitudenschwankungen in Prozent aus.

Heute werden im allgemeinen Shimmer- und Jittermaß gemeinsam verwendet, da ihre Kombination die Trennschärfe in der Graduierung pathologischer Stimmklänge erhöht (Klingholz, 1986).

In Arbeiten Huangs (1998) scheint der Shimmer bzw. der APQ besonders mit der Wahrnehmung von Heiserkeit zu korrelieren. Jedoch weisen die meisten Arbeiten annähernd gleich gute Korrelationen des Shimmers mit der wahrgenommenen Heiserkeit oder Rauigkeit oder Behauchtheit auf (Zwirner 1996+1998; de Krom, 1994).

Bei all diesen erwähnten elektroakustischen Verfahren gilt die Objektivität der Erfassung bestimmter akustischer Parameter als hoch (Fröhlich et al. 1998), doch der zeitliche Aufwand variiert beträchtlich.

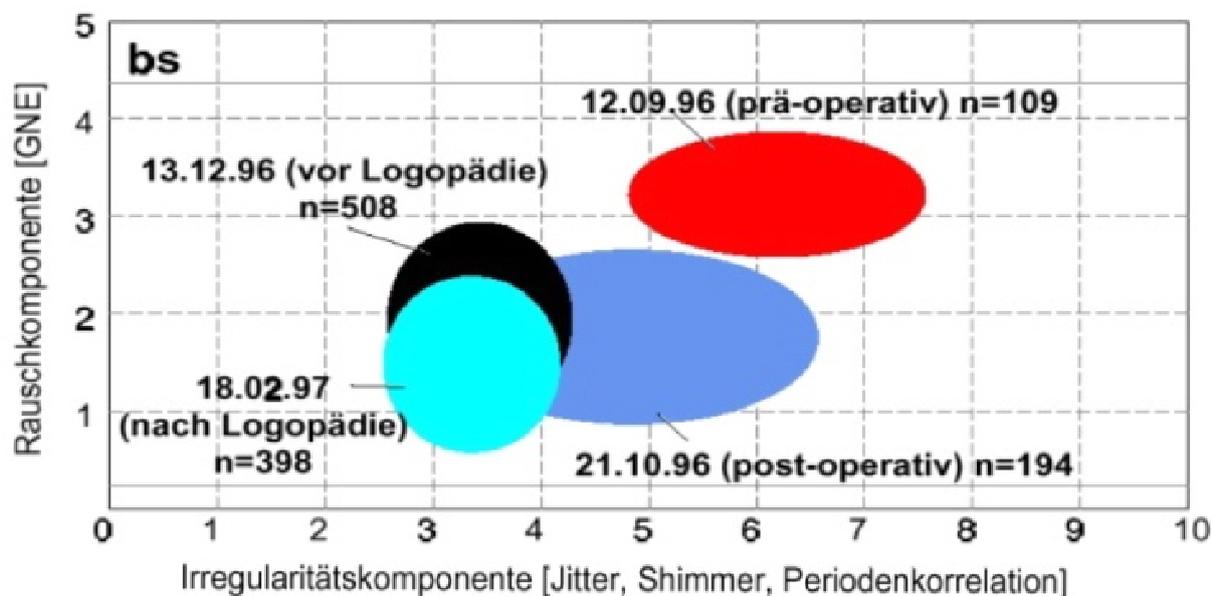
1. 2. 3. Kombinationsverfahren

Da Heiserkeit eine Mischung aus Strömungsgeräusch und Schwingungsunregelmäßigkeit ist, war es naheliegend, die als physikalisch-akustische Korrelate für die auditiven Geräuschphänomene ermittelten akustischen Messdaten zusammenzufassen, um ein möglichst vollständiges objektives Dokument des Stimmklangs zu erhalten. Exemplarisch seien hier zwei unterschiedliche Ansätze vorgestellt.

Der jüngste Vorschlag stammt von Wuyts et al. (2000): Sie schlagen einen Dysphonia-Severity-Index (DSI) vor. Er setzt sich aus höchstmöglicher Frequenz in Hz, leisestem Ton in dB, maximalem Tonhaltevermögen in s und Jitter in % zusammen und soll Werte zwischen +5 (normal) und -5 (maximale Heiserkeit) ergeben. Untersuchungen von Nawka (2000) ergaben, dass der DSI eine nützliche Maßzahl zur Beschreibung der Stimmqualität im klinischen Alltag ist. Da der DSI auch das Tonhaltevermögen und den Tonhöhenumfang der Stimme miteinbezieht, sollte er jedoch nicht in erster Linie als Maß des Stimmklangs, sondern eher als ein Index für die allgemeine stimmliche Leistungsfähigkeit aufgefasst werden.

Ein sehr vielversprechendes Kombinationsverfahren, welches ausschließlich den Stimmklang bewertet, scheint das Mitte der 90er Jahre von Michaelis und Fröhlich vorgestellte Heiserkeitsdiagramm zu sein.

Abbildung 6: Heiserkeitsdiagramm nach Michaelis und Fröhlich Es zeigt die Stimmqualität einer Patientin (bs) nach Operation eines Reinke-Ödems und anschließender Logopädie (n= Anzahl der Datenpunkte für die jeweilige Ellipse). Deutlich erkennbar ist die Verbesserung der Stimmqualität durch die Ellipsenverschiebung von rechts oben (schlechtere Qualität) nach links unten (bessere Qualität). (aus: Fröhlich M, Michaelis D, Strube H, Kruse E (2000) Acoustic voice analysis by means of the hoarseness diagram. *J. Speech Lang Hear Res* (43): 706-720)



Im Heiserkeitsdiagramm werden Daten der Periodizitätsanalyse (sog. Irregularitätskomponente, die Jitter, Shimmer und die Periodenkorrelation zusammenfasst) mit Daten auf Grundlage der Spektralanalyse (GNE/glottal-to-noise-excitation-ratio) graphisch vereint. In verschiedenen Arbeiten konnten sie die Validität der akustischen Stimmgütebeschreibung durch das Heiserkeitsdiagramm belegen (Fröhlich et al., 1997; 2000; Michaelis et al., 1997; 1998). Als Ausgangsmaterial für die jeweilige Gütebeschreibung müssen 28 Vokalaufnahmen in verschiedenen Tonlagen und Lautstärken der zu bewertenden Stimme erstellt werden. 1998 konnten Michaelis et al. jedoch zeigen das auch bei reduziertem Umfang der Stimmaufnahmen auf nur drei Vokalaufnahmen noch valide Aussagen gemacht werden können. Das Heiserkeitsdiagramm eignet sich sehr gut für die Verlaufsbeobachtung der Stimmklangveränderungen unter logopädischer Therapie oder nach Operationen am Kehlkopf.

Abschließend sei jedoch erwähnt, dass auch in der computerisierten HNO-Klinik die meisten elektroakustischen Verfahren nicht zur alltäglichen Routine gehören, da sie zum einen noch zuwenig standardisiert sind und häufig einen erhöhten Zeitaufwand mit sich bringen, und andererseits ihre klinische Relevanz teils noch unklar ist (Friedrich et al., 2000).

1. 2. 4. Zur Stimmklanganalyse Software „Dr. Speech ®“ 3.0

Wie erwähnt, ist die elektroakustische Analyse des Stimmklangs eines Patienten vor, während und nach Therapie aus verschiedenen Gründen häufig keine Routine, obwohl sich gerade solche objektiven Analysen zur Kontrolle der Stimmqualität nach operativen Eingriffen anbieten würden. Die Stimmanalysesoftware Dr. Speech ® der Firma Tiger Electronics Inc., U.S.A., im folgenden nur als Dr. Speech ® bezeichnet, bietet neben Unter-Programmen zur Stimmsynthese, zum Sprach- und Stimmtraining in Echtzeit und zur EGG-Beurteilung eine Reihe von elektroakustischen Analysestandards zur Heiserkeitsbewertung an.

Dr. Speech ® unterstützt zum einen die Berechnung der Tonhöhen- und Amplitudenperturbation (APQ, PPQ) und der Geräuschanteile (NNE), sowie die Darstellung von LTA-Spektren und Sonographien auf der Basis der Fast Fourier Transformation.

Zum anderen steht durch Dr. Speech ® ein weiteres Kombinationsverfahren zur Verfügung, welches eigentliches Thema der vorliegenden Arbeit ist. Es bietet die Möglichkeit einer computergestützten Stimmklangbeurteilung im RBH-Schema an.

Computergestützte RBH-Beurteilung

Dieses Kombinationsverfahren legt Parameter der elektroakustischen Analyse, insbesondere APQ, PPQ und NNE zugrunde, um den Stimmklang in einer 4-stufigen Skala in die Qualitäten Rauhigkeit, Behauchtheit und Heiserkeit einzustufen. Jede der 3 Qualitäten (VQ = voice quality) setzt sich nach Huang (2001), einem der Entwickler dieses Kombinationsverfahrens, aus mehreren Faktoren zusammen:

$$VQ = k_1x_1 + k_2x_2 + k_3x_3 \dots$$

Hierbei entspricht VQ einer der drei Qualitäten Rauhigkeit (harsh voice) oder Behauchtheit (breathy voice) oder Heiserkeit (hoarse voice).

k_1, k_2, k_3, \dots sind Gewichtungsfaktoren, wobei

$$k_1 + k_2 + k_3 + \dots = 1 \quad \text{ist.}$$

x_1, x_2, x_3, \dots sind verschiedene akustische Parameter:

- Für die Behauchung ist $x_1 = \text{NNE}$ und k_1 hat den höchsten Wert. Die computergestützte Beurteilung der Behauchtheit ist also hauptsächlich von der NNE abhängig, aber nicht ausschließlich. Die Parameter für x_2, x_3, \dots werden vom Hersteller nicht mitgeteilt.
- Für die Rauigkeit ist $x_1 = \text{PPQ}$ (Jitter) und k_1 hat den höchsten Wert. Die computergestützte Beurteilung der Rauigkeit ist also hauptsächlich von PPQ abhängig, aber nicht ausschließlich. Die Parameter für x_2, x_3, \dots werden vom Hersteller nicht mitgeteilt.
- Für die Heiserkeit ist $x_1 = \text{APQ}$ (Shimmer) und k_1 hat den höchsten Wert. Die computergestützte Beurteilung der Heiserkeit ist also hauptsächlich von APQ abhängig, aber nicht ausschließlich. Die Parameter für x_2, x_3, \dots werden vom Hersteller nicht mitgeteilt.

Zusammengefasst legt die computergestützte RBH-Beurteilung durch Dr. Speech® bei der Rauigkeitsbeurteilung vor allem den Jitter (PPQ), bei der Behauchtheitseinstufung vor allem die NNE, und bei der Bewertung der Heiserkeit in erster Linie den Shimmer (APQ) zugrunde. Diese Algorithmen entstanden auf der Grundlage der Untersuchungen Huangs. Dabei wurde der synthetisch erzeugte Vokal [ae] durch jeweils nur einen akustischen Parameter verändert und dann perzeptiv in der GRABS-Skala (s.o.) beurteilt und so die Korrelationen von akustischen zu perzeptiven Parametern festgestellt (Huang, 1995 und 1998).

Weder über die genauen Algorithmen noch über die Grenzwerte zur computergestützten RBH-Beurteilung gab die Herstellerfirma weitere Auskünfte.

Vor allem diese elektroakustische RBH-Beurteilung scheint zunächst aus verschiedenen Gründen besonders für den klinischen Alltag geeignet. Erstens könnte die gewählte Darstellung der computergestützten Stimmklangbeurteilung im RBH-Schema die Vergleichbarkeit mit der perzeptiven Beurteilung im oft verwendeten RBH-Schema erleichtern. Zweitens ist es durch diese Darstellungsform (Abb. 7) auch für den Laien möglich, sich eine ungefähre Vorstellung vom Stimmklang zu machen, auch ohne dass die Stimme vorher gehört wurde. Werden ausschließlich akustische Parameter bei der Ergebnisdarstellung verwendet, wie z.B. beim Heiserkeitsdiagramm, sind diese beiden Punkte nicht zutreffend. Drittens ist die computergestützte RBH-Beurteilung sehr schnell

durchführbar. Ob diese computergestützte RBH-Beurteilung auch brauchbare und zuverlässige Ergebnisse liefert, ist Gegenstand dieser Arbeit.

Abbildung 7: Beispielbild der computergestützten RBH-Beurteilung durch „Dr. Speech“[®]

Der Stimmklang wird computergestützt, übersichtlich im RBH-Schema (harsh voice, breathy voice und hoarse voice) auf einer 4-stufigen Skala (normal=0, slight=1, moderate=2, extreme=3) klassifiziert.

Hoarse Voice	Normal	Slight	Moderate	Extreme
Harsh Voice	Normal	Slight	Moderate	Extreme
Breathy Voice	Normal	Slight	Moderate	Extreme

1. 3. Fragestellung

Habermann wies 1975 darauf hin, dass die Anwendung der sich entwickelnden Methoden zur akustischen Stimmanalyse der Realität des praktizierenden Facharztes unerreichbar fern ist. „Zwar sind heutzutage Computer allgegenwärtig, wodurch die Bestimmung akustischer Parameter aus einem Stimmsignal deutlich erleichtert wurde, jedoch gibt es nach wie vor ein deutliches Missverhältnis zwischen aufwendigen diagnostischen Prozeduren in phoniatischen Institutionen auf der einen Seite, und dem meist völligen Verzicht auf Diagnostik und Dokumentation auf der anderen Seite“ (Friedrich 1996). Aus diesem Missverhältnis heraus entstand der Wunsch nach einem täglich einsetzbaren elektroakustischen Verfahren zur Bestimmung der Stimmqualität, insbesondere des Stimmklangs.

Die vorliegende Arbeit untersucht, ob die computergestützte Heiserkeitsbewertung im RBH-Schema durch Dr. Speech® ein solches Verfahren ist, d. h. ob es die Forderungen nach Zuverlässigkeit, geringem Zeitaufwand und guter Dokumentierbarkeit erfüllt.

Die Arbeit gliedert sich in zwei Teile:

Zunächst wird anhand von 2 Einzelfällen die Aussagekraft der computergestützten RBH-Beurteilung durch Dr. Speech® und einzelner akustischer Parameter überprüft.

Im zweiten Teil wird durch eine Studie die Zuverlässigkeit der Ergebnisse der computergestützten RBH-Beurteilung untersucht. Folgende Fragen werden beantwortet:

1. *Stimmt die computergestützte RBH-Beurteilung mittels „Dr. Speech®“ mit der auditiven Heiserkeitsbewertung überein, kann sie sie ggf. ersetzen oder sie zumindest in der täglichen Arbeit unterstützen?*
2. *Ist die computergestützte RBH-Beurteilung mittels „Dr. Speech®“ in der alltäglichen Routine zur Dokumentation von Stimmqualitätsveränderungen vor und nach operativer Therapie geeignet ?*

2. Material und Methode

2.1. Technischer Aufbau

Dr. Speech ® 3.0 wurde auf einem handelsüblichen Multimedia-PC unter Windows (ab Version 3/11) installiert. Für die Stimmaufnahmen wurde ein externes Mikrofon (Firma Sennheiser ®) mit nierenförmiger Frequenz-Charakteristik verwendet. Die Basis für die computergestützte RBH-Beurteilung durch Dr. Speech ® sind Stimmproben des Vokals [ae:] (wie z.B. in Ähre), von dem 3-5 Aufnahmen pro Patient erstellt wurden. Dabei betrug der Abstand des Mikrofons zum Patienten 10 cm, wie es auch der Hersteller der Stimmanalysesoftware empfiehlt. Der Prüfraum hatte Wohnzimmerakustik. Während der Aufnahme der Stimmproben (weiteres siehe unten) erfolgte gleichzeitig die Bestimmung der Lautstärke in dB(A) mit einem handelsüblichen Schalldruckpegelmessers (Fa. Brüel & Kjaer, Type 2226, Abstand zum Patienten 30 cm). Die Aufnahmen erfolgten am stehenden Patienten.

2.2. Probanden und Durchführung der Untersuchung

Insgesamt wurden die prae- und postoperative Stimmaufnahmen von 62 Patienten (n=62) verglichen und statistisch ausgewertet.

Ursprünglich wurden von 92 Patienten praeoperative Stimmproben aufgezeichnet, jedoch reduzierte sich die Anzahl aus verschiedenen Gründen auf die erwähnten 62 Patienten:

- 12 Patienten erschienen nicht zu den mehrfach anberaumten postoperativen Kontrollterminen.
- 3 Patienten wurden wenige Wochen nach der praeoperativen Stimmaufnahme laryngektomiert.

Somit würden 77 Patienten verbleiben, von denen jedoch aus technischen Gründen weitere 15 ausgeschlossen werden mussten. Diese Patienten hatten ausnahmslos sehr heisere bis aphone Stimmen, bei denen wegen zu hoher Geräuschbeimengungen und Periodizitätsschwankungen (Unmöglichkeit der F0-Bestimmung) keine Periozitätsanalyse und damit auch keine elektroakustische RBH-Klassifizierung möglich war. Dies wurde durch die Meldung: Warning! No perturbation measures can be reported from this vowel, because of the extremely pathological condition. Please record again. angezeigt. Hierauf wird in der Diskussion näher eingegangen.

Unter den 62 ausgewerteten Patienten (44 Männer und 18 Frauen im Alter von 23 –78 Jahren) befanden sich 10 Patienten mit einseitigem Stimmlippenkarzinom (T1a), 13 mit Leukoplakien auf den Stimmlippen, 7 mit Reinke-Ödem, 3 mit Stimmlippenzyste, 14 mit Stimmlippenpolyp, 2 mit beiderseitiger Stimmlippenparese, 7 mit Kontaktgranulom und 6 mit einer Kehlkopfpapillomatose. Die Tabelle 1 des Anhangs gibt einen Überblick über Diagnosen und Therapie und enthält die subjektive und objektive Heiserkeitsbewertung durch den Untersucher bzw. *Dr. Speech®* .“

Bei allen 62 Patienten wurde die Stimmqualität am Tag vor der operativen Behandlung von einem erfahrenen Untersucher auditiv nach dem RBH-Schema eingestuft. Anschließend wurden mehrere Stimmproben des gleichen Patienten aufgezeichnet.

Eine postoperative Kontrolluntersuchung fand nach 6 –17 Wochen (Mittelwert ~58 Tage) statt. Nach Larynxspiegelung und Feststellung einer abgeschlossenen Wundheilung erfolgte wiederum in gleicher Weise die subjektive und objektive Bewertung des Stimmklangs.

Bei der Aufzeichnung der Stimmproben wurde der Patient aufgefordert, den Laut [ae:] in entspannter, normaler Sprechstimmlage und Sprechlautstärke zu halten. Wie in den Untersuchungen Kasuyas (1986) wurde die Lautstärke bei jeder Aufnahme schrittweise leicht erhöht. Davon wurden vom Computer je 3 sec (ohne Stimmeinsatz) unter standardisierten Bedingungen (s.o.) mit einer Abtastfrequenz von 44,1 kHz aufgezeichnet.

Die auditive Beurteilung im RBH-Schema erfolgte anhand des Anamnesegesprächs und des Weiteren anhand von gehaltenen Vokalen, dabei wurde unter **Rauhigkeit** alle Geräuschbeimengungen mit aperiodischem Pulscharakter zusammengefasst. Auch Stimmabbrüche und Frequenzsprünge sind Rauigkeitsphänomene. Die **Behauchung** subsumiert kontinuierliches Begleitrauschen. Bei der Beurteilung der **Heiserkeit** stand der Gesamteindruck des Stimmklangs, mit allen pathologischen Stimmklangphänomenen im Vordergrund.

2. 3. Auswahlverfahren und statistische Auswertung

Für den abschließenden Vergleich der prae- und postoperativen Stimmaufnahmen musste aus den 3-5 prae- und den 3-5 postoperativen Stimmaufnahmen je eine ausgewählt werden. Um eine gute Vergleichbarkeit der Stimm-Samples zu gewährleisten, sollten definierte Parameter prae- und postoperativ annähernd gleich sein:

1. Die Grundfrequenz (F0) sollte möglichst übereinstimmen.
2. Die mit dem externen dB-Meter gemessene Lautstärke sollte möglichst gleich sein.
3. Um annähernd gleiche Vokale zu vergleichen sollte die Formantenlage der ersten beiden Formanten (ermittelt durch Spektralanalyse) möglichst identisch sein.

Die Kriterien wurden in der angegebenen Reihenfolge berücksichtigt.

Nach Abschluss der Datensammlung wurde nach dem oben beschriebenen Verfahren je eine prae- und postoperative Stimmufnahmen (3-sec-samples) ausgewählt und durch Dr. Speech ® die computergestützte RBH-Beurteilung durchgeführt.

Abschließend ergaben sich 62 prae- und 62 postoperative computergestützte RBH-Klassifizierungen und die gleiche Anzahl (2x62) an auditiven Beurteilungen, die statistisch gut zu vergleichen sind. Die Tabelle 1 des Anhangs zeigt die RBH-Bewertungen von Dr. Speech ® und dem Untersucher.

In Tabelle 2 des Anhangs wurde die prae- und postoperativ gemessene NNE der subjektiv eingeschätzten Behauchung gegenübergestellt.

Um zu untersuchen inwieweit das objektive ‘Urteil durch Dr. Speech ® ‘und die subjektive Bewertung durch den Untersucher miteinander korrelieren, wurden im folgenden neben der graphischen Darstellung der Auswertungsergebnisse der Korrelationskoeffizient nach Spearman und der Konkordanzkoeffizient Kappa, sowie der Wilcoxon-Test für Paardifferenzen verwendet. Das ROC-Verfahren nach Metz wurde verwendet, um die Korrelation des Behauchtheitsmaßes NNE mit der subjektiv eingeschätzten Behauchung darzustellen.

Der **Korrelationskoeffizient nach Spearman (Spearman's R)** testet, ob höhergradige Beurteilungen von Dr. Speech ® mit höhergradigen Beurteilungen des Untersuchers einhergehen und analog niedergradige Beurteilungen mit niedergradigen, und zwar unabhängig davon, ob beide Beurteilungen jeweils gleich sind. Wenn z. B. Dr. Speech ® nur die Beurteilungen 0 und 1 abgibt und der Untersucher nur 2 und 3, kann trotzdem ein hohes Spearman R berechnet werden, solange die Beurteilungen weitgehend parallel verlaufen. Spearman's R kann Werte zwischen +1 und -1 annehmen. Je größer R dem Betrag nach ist, desto stärker ist der Zusammenhang zwischen den untersuchten Größen. Für R-Werte bis +/-0.5 spricht man von einem schwachen Zusammenhang, bis 0.75 von einem deutlichen Zusammenhang. R-Werte bis 0.99 deuten auf einen straffen Zusammenhang und ein R von 1.0 zeigt einen gesetzmäßigen Zusammenhang zwischen den Variablen an. Ein positives R bedeutet, dass der Zusammenhang direkt proportional ist (wenn die eine Beurteilung größer wird,

wird auch die andere Beurteilung größer), negative R-Werte sind ein Zeichen für einen umgekehrt proportionalen Zusammenhang (wenn die eine Beurteilung größer wird, wird die andere kleiner).

Der **Konkordanzkoeffizient Kappa (Cohens KAPPA)** ist ein Maß für die Stärke der Übereinstimmung zwischen beiden Beobachtern. Er untersucht stärker als Spearman R, ob beide Beurteilungen wirklich gleich sind, was sich bei einer nur 4stufigen Skala, wie in unserem Fall, besser anbietet. Es werden deutlich niedrigere Werte berechnet, wenn die Beurteilungen von Dr. Speech® und dem Untersucher voneinander abweichen, auch dann wenn z. B. die Tendenz besteht, dass hohe Dr. Speech® Beurteilungen mit hohen Untersucher Beurteilungen einhergeht. KAPPA kann Werte zwischen 0 (keine Übereinstimmung) und 1 (vollständige Übereinstimmung) annehmen.

Deutliche Unterschiede zwischen Spearmans R und Cohens KAPPA für die Untersuchung des selben Zusammenhanges sind auf dieses Phänomen zurückzuführen.

Der **Wilcoxon-Test für Paardifferenzen** gibt Auskunft darüber, ob sich zwei Variablen in ihrer zentralen Tendenz statistisch signifikant unterscheiden. Z.B. untersucht er, ob sich die Beurteilung der Heiserkeit von Dr. Speech® und dem Untersucher in ihrer zentralen Tendenz unterscheiden. Dieser Test weist eine ähnlich hohe Effizienz auf wie der ebenfalls anwendbare t-Test, hat diesem jedoch voraus, dass keine Anforderungen hinsichtlich der Verteilungsfunktion der beiden Variablen gestellt werden.

Das **ROC-Verfahren (Receiver-Operating-Characteristics) von Metz** (Metz, 1978) ist geeignet die Fähigkeit eines Parameters (hier: NNE) bezüglich seines Differenzierungsvermögens zu beurteilen. In der vorliegenden Arbeit wurde mit dem ROC-Verfahren der Wert der NNE (Trennungspunkt) berechnet, der zwischen behauchten und nicht behauchten Stimmen unterscheidet. Diese Methode stützt sich auf Berechnungen der Sensitivität (Wahrscheinlichkeit, dass eine behauchte Stimme als solches erkannt wird) und der Spezifität (Wahrscheinlichkeit, dass eine nicht behauchte Stimme als solche erkannt wird) für jeden Punkt der Beurteilungsskala (hier: jeder gemessene NNE-Wert), sowie den Anteil fehlklassifizierter Fälle. Der (NNE-) Wert, an dem die Summe aus Sensitivität und Spezifität am höchsten ist und die Fehlklassifikationen am niedrigsten, ist der Trennungspunkt.

3. Ergebnisse

3. 1. Einzelfall-Analysen

3. 1. 1. Fall 1: Laterofixation wegen beidseitiger N. recurrens-Parese

Als erster Fall sei hier ein 45jähriger Lagerarbeiter demonstriert, bei dem wegen einer posttraumatischen beidseitigen Stimmlippenparese (Zustand nach Thoraxquetschung mit nachfolgender Langzeitintubation) eine mikrolaryngochirurgische Glottiserweiterung erforderlich wurde. Der Patient war präoperativ mit dem Klang seiner Stimme zufrieden, war jedoch körperlich kaum belastbar wegen eines erheblichen Belastungsstridors. Die Laryngoskopie zeigte einen beidseitigen Stimmlippenstillstand mit einem dorsalen Glottisrestspalt von etwa 2 mm.

Die Glottiserweiterung erfolgte im Sinne einer modifizierten Thornell'schen Operation, das heißt CO² Laser-Arytänoidektomie mit Laterofixation durch Naht.

Acht Wochen nach dem Eingriff war der Patient körperlich deutlich belastbarer; sein Stridor war verschwunden. Dafür aber hatte sich seine Stimmfunktion deutlich verschlechtert: Der Stimmeinsatz war jetzt stark verhaucht, das Tonhaltevermögen auf 3 Sekunden verkürzt. Der erhebliche Luftverbrauch bei der Stimmgebung schränkte darüber hinaus die Steigerungsfähigkeit der Stimme (Stimmdynamik) stark ein.

3. 1. 1. 1. *Auditive prae- und postoperative Heiserkeitsbewertung im RBH-Schema (Fall1)*

Vom Untersucher wurde der Stimmklang praeoperativ als geringgradig heiser (praeoperative Heiserkeit = 1) und nicht behaucht (praeoperative Behauchtheit = 0) bewertet. Diese leichte Heiserkeit entstand dem auditiven Eindruck nach offenbar hauptsächlich durch Aperiodizitäten und nicht durch wilde Luft- bzw. Strömungsgeräusche, weshalb die Rauigkeit mit 1 und die Behauchtheit mit 0 bewertet wurde. Praeoperativ war also der Heiserkeitsgrad gering als Zeichen eines noch ausreichenden Glottisschluss. Postoperativ verschlechterte sich infolge der Glottiserweiterung der auditive Eindruck des Stimmklangs

erwartungsgemäß deutlich. Der Stimmklang war maximal behaucht (B=3) und heiser (H=3) als Zeichen eines jetzt insuffizienten Glottisschlusses. Der Rauigkeitseindruck hatte sich nicht verändert (prae- und postoperative Rauigkeit = 1).

Tabelle 3: auditive RBH-Beurteilung des Stimmklangs des Patienten von Fall 1 (Patient 51 aus Tabelle 1)

	praeoperativ	postoperativ
Rauhigkeit	1	1
Behauchtheit	0	3
Heiserkeit	1	3

3. 1. 1. 2. Vergleich der prae- und postoperativen Oszillogramme (Fall 1)

Vergleicht man einen Ausschnitt von sieben Perioden aus den prae- und postoperativen Oszillogrammen (Abbildung 8a und b), fallen im postoperativen Oszillogramm die stärkeren Unregelmäßigkeiten der aufeinanderfolgenden Schwingungsperioden auf. Einerseits wird es durch eine wesentlich stärkere Oberwelligkeit gestört, andererseits sind die Zeitdifferenzen aufeinanderfolgender Perioden wesentlich größer. Dies drückt sich in einer postoperativen Zunahme von Jitter (PPQ) und Shimmer (APQ) aus (s.u.). Weiterhin lässt sich bei der postoperativ aufgezeichneten Stimmprobe eine stark verkürzte Tonhaldedauer (durch den höheren Luftverbrauch) erkennen (Abbildung 8b, untenstehend). Das postoperative Gesamtsignal zeigt im Gegensatz zum praeoperativen eine kontinuierliche Amplituden-Verkleinerung, was vor allem bei der Bewertung des Shimmers berücksichtigt werden muss (s.u.).

Abbildung 8a: Oszillogramm der praeoperativen Stimmaufnahme von Fall 1 (vor Laterofixation), unten: Gesamtsignal (3 sec); oben: Ausschnitt mit sieben Perioden

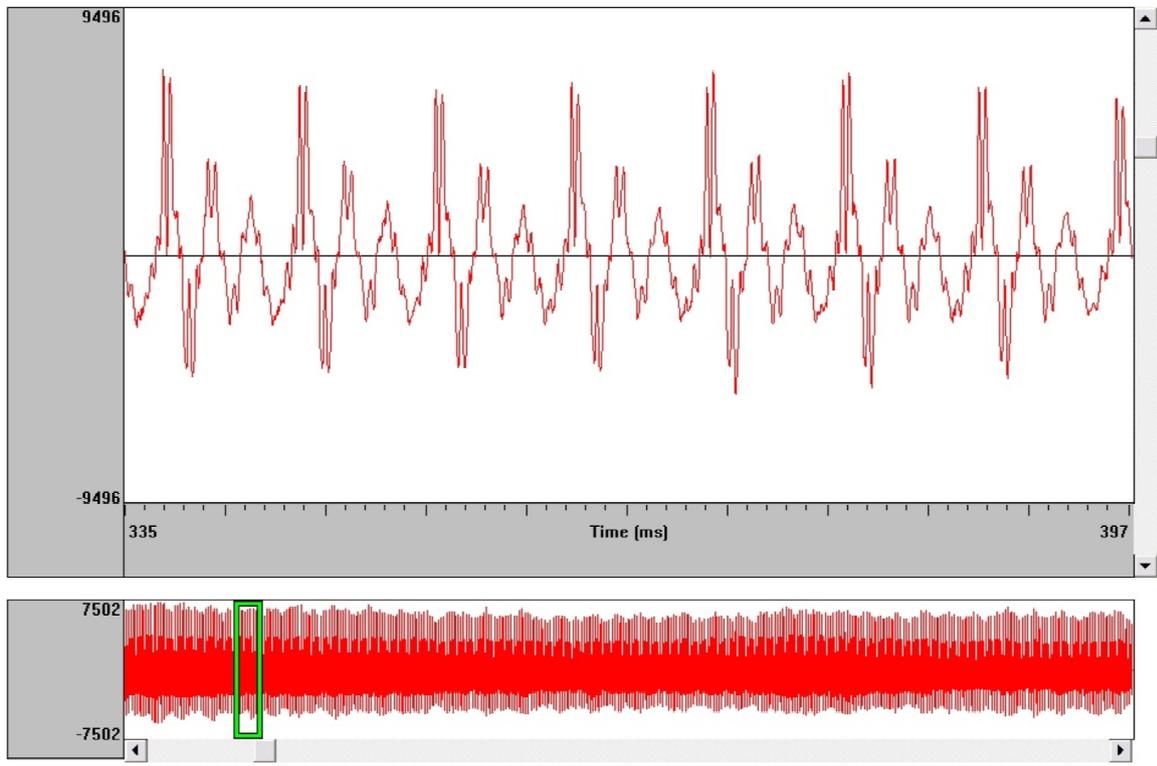
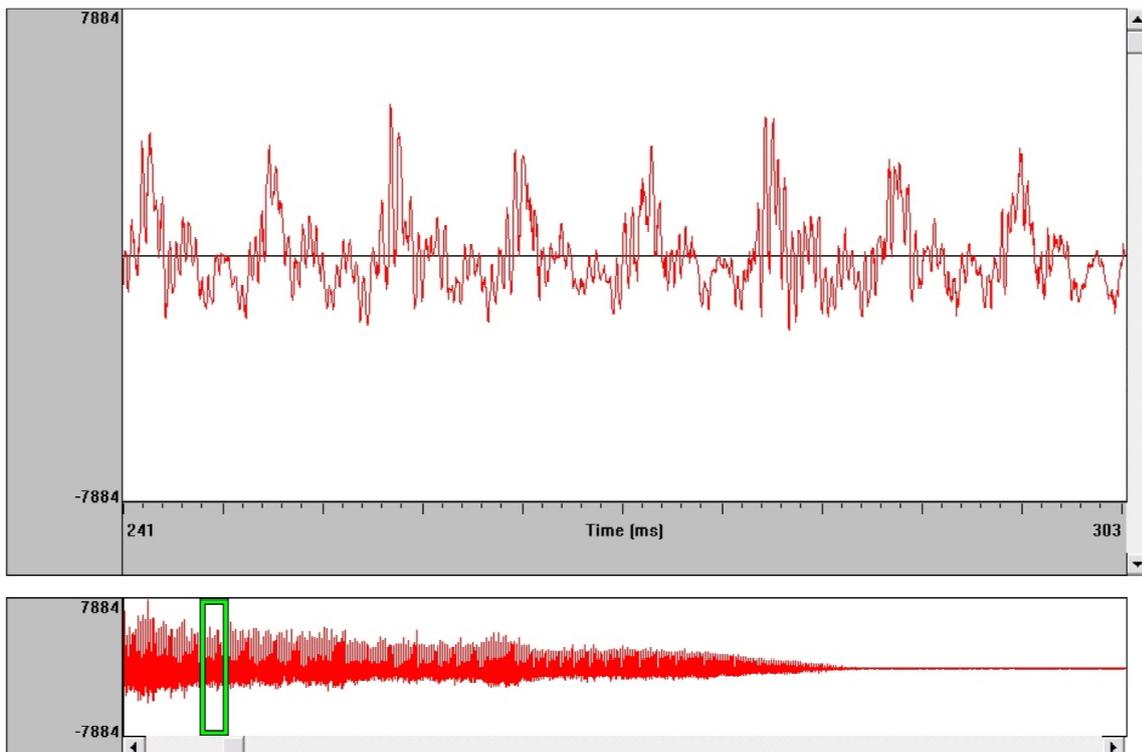


Abbildung 8b: Oszillogramm der postoperativen Stimmaufnahme von Fall 1 (nach Laterofixation), unten: Gesamtsignal (<3 sec, deutlich verkürzte Phonationszeit); oben: Ausschnitt mit sieben Perioden



3. 1. 1. 3. Vergleich der prae- und postoperativen Spektralanalysen (Fall 1)

Wie in Kapitel 1.2.1. erwähnt, gibt es zur Interpretation von Spektrogrammen zahlreiche Ansätze:

- 1) Im praeoperativen LTA-Spektrum (Abbildung 9a) sind die Harmonischen bis zu einem Frequenzbereich von etwa 3,5 KHz deutlich ausgeprägt. Postoperativ (Abbildung 9b) liegt eine Verarmung an Obertönen vor (bis ca. 1,2 KHz).
- 2) Die Formantregionen sind praeoperativ klarer gezeichnet als postoperativ.
- 3) Nach Hammarberg (1994) und Fant (1993) haben behauchte Stimmen gegenüber normalen Stimmen einen geringeren maximalen Energiegehalt (entspricht der maximalen Amplitude) im Bereich zwischen 2-5 kHz, im Verhältnis zur Energie der Region des ersten Formanten. Vergleicht man die vorliegenden LTA-Spektren erkennt man deutlich, dass die Energie zwischen 2-5 kHz postoperativ geringer ist gegenüber dem praeoperativen Spektrum. Dies ist sowohl absolut gesehen der Fall, als auch im Verhältnis zur Energie des ersten Formanten. Zur Verdeutlichung sind die Abstände der Energiemaxima in grün angedeutet. A_{prae} ist deutlich geringer als A_{post} .“
Außerdem steigt postoperativ der Energiegehalt des Frequenzbereiches zwischen 5-8 kHz im Verhältnis zur Energie zwischen 2-5 kHz an ($B_{prae} > B_{post}$):“ Auch absolut gesehen ist postoperativ eine Energiezunahme oberhalb 5 kHz zu verzeichnen, d.h. der Anteil an akustischer Energie durch Unharmonische (Geräuschanteile) ist im postoperativen Spektrum wesentlich größer.

Diese drei Beobachtungen sprechen für eine postoperativ verschlechterte Stimmqualität, im Sinne einer postoperativen Abnahme an harmonischen Klanganteilen und einer Zunahme von Geräuschanteilen, die als zunehmende Heiserkeit bzw. Behauchung wahrgenommen wird.

Abbildung 9a: Spektralanalyse der praeoperativen Stimmaufnahme von Fall 1 (vor Laterofixation), rote Kurve: LTAS-Spektrum; die schwarzen Linien zeigen die Energiemaxima in den Bereichen 0-2 kHz, 2-5 kHz und 5-8 kHz; die grünen Linien deuten die Abstände zwischen den Energiemaxima an (A prae und B prae)

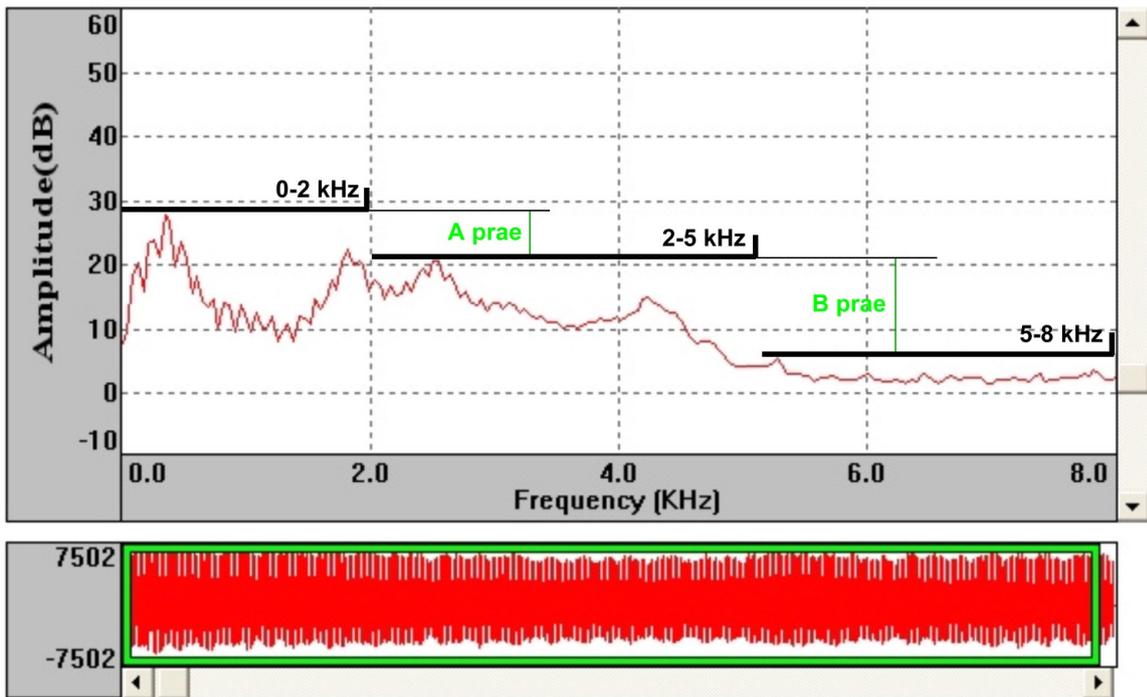
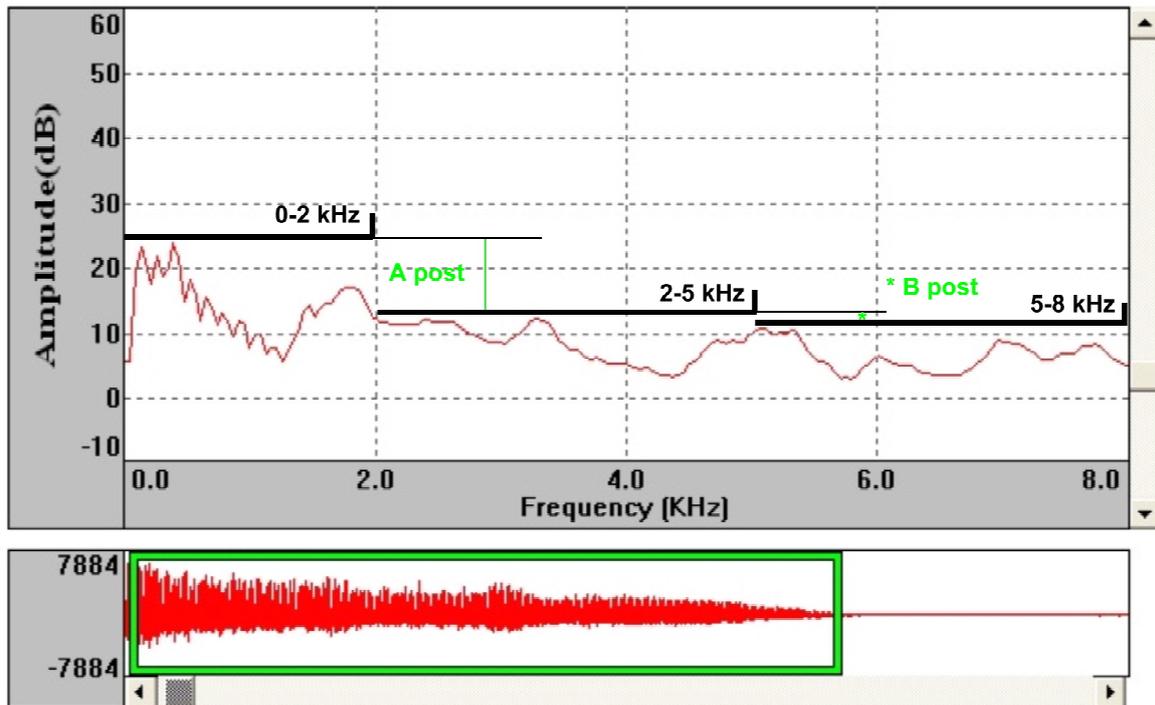


Abbildung 9b: Spektralanalyse der postoperativen Stimmaufnahme von Fall 1 (nach Laterofixation), rote Kurve: LTAS-Spektrum; die schwarzen Linien zeigen die Energiemaxima in den Bereichen 0-2 kHz, 2-5 kHz und 5-8 kHz; die grünen Linien deuten die Abstände zwischen den Energiemaxima an (A post und B post)

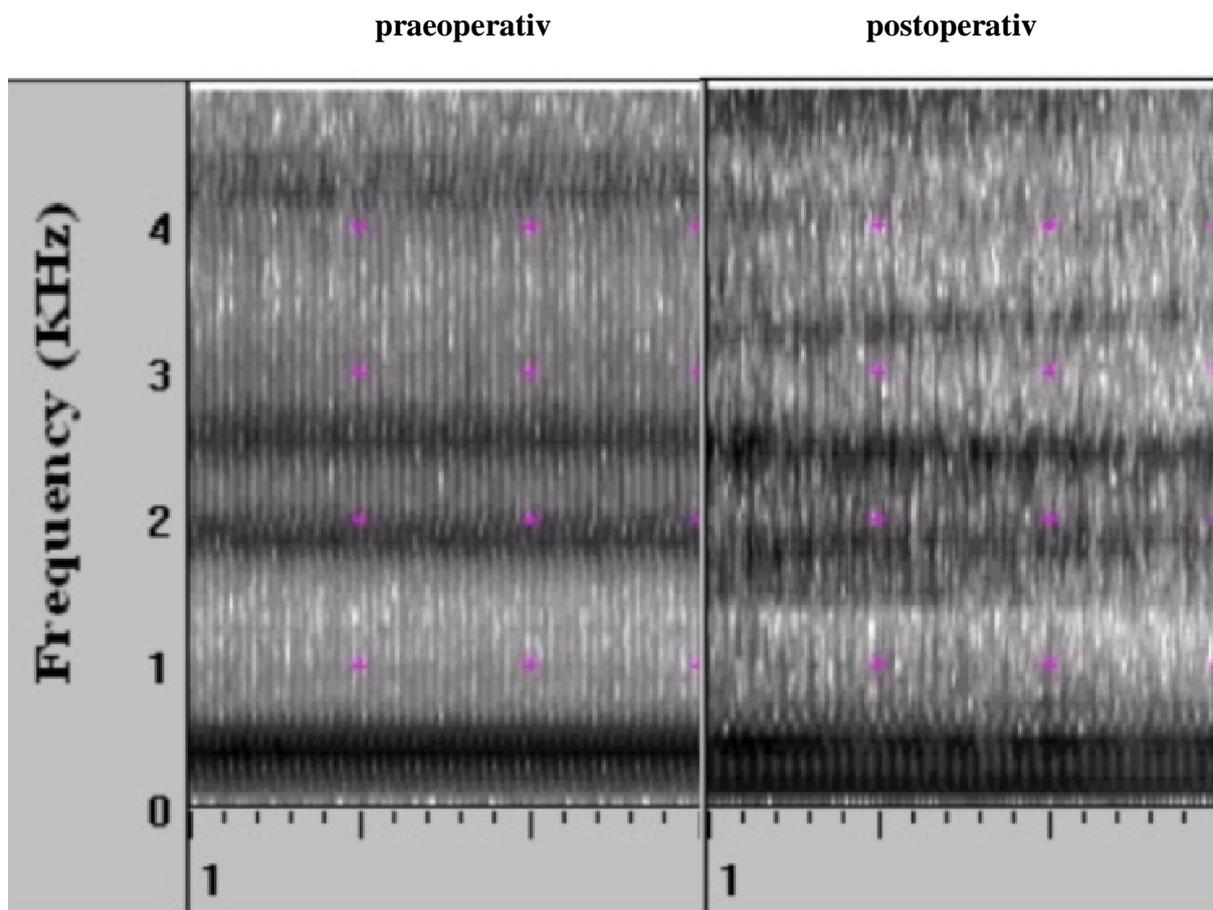


3. 1. 1. 4. Vergleich der prae- und postoperativen Sonogramme von Fall 1

Ähnliches wie in den LTA-Spektren sieht man im Sonogramm:

Praeoperativ lassen sich die Formantbänder im Bereich der ersten beiden und auch des dritten Formanten von [ae] gut erkennen. Geräuschkomponenten im Formantenbereich finden sich kaum (Typ 0 nach Silvermann und Zimmer (1978)). Postoperativ haben die Geräuschkomponenten im Bereich des 2. und 3. Formanten deutlich zugenommen, es treten Geräuschbänder oberhalb von 3 kHz hinzu (Typ 2-3 nach Yangiahara). Die Sonographie eignet sich hier hervorragend zur visuellen Dokumentation der auditiv konstatierten Stimmverschlechterung.

Abbildung 10: prae- und postoperative Breitband-Sonogramme (Bandbreite: 300 Hz, Ausschnitt) von Fall 1.



3. 1. 1. 5. Vergleich der prae- und postoperativen Werte für APQ, PPQ und NNE von Fall 1

Die optisch erfasste postoperative Zunahme der Schwingungsunregelmäßigkeiten in den oben besprochenen Abbildungen 8 a und b, lässt sich durch ein postoperatives Ansteigen der Periodizitätsmaße PPQ und APQ quantifizieren. Auch die NNE nimmt nach Laterofixation deutlich pathologischere Werte an (Tabelle 4).

Tabelle 4: prae- zu posttherapeutische Veränderung von 3 akustischen Parametern in Fall 1.

Parameter	prae	post	normal (männlich)	Pathologisch (männlich)
PPQ(Jitter) [%]	0,32	0,86	0,631 (mean) ¹	1,264 (mean) ¹
APQ(Shimmer) [%]	3,87	9,91	3,027 (mean) ¹	4,811 (mean) ¹
NNE[dB]	-15,5	-2,47	< -11 ²	> -11 ²

¹ nach Boyanov et al. 1997 ² nach Kasuya et al. 1986

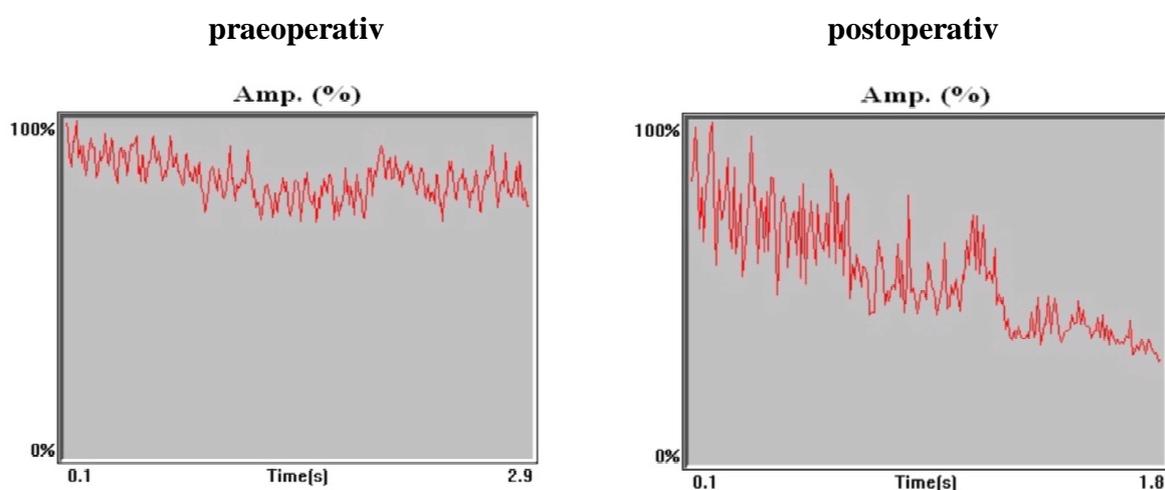
Vor allem die Veränderung der NNE korreliert deutlich mit der subjektiv starken Zunahme der Behauchung. Praeoperativ klang die Stimme nur schwach überhaucht. Postoperativ ist sie stark heiser bzw. verhaucht. Das spiegelt sich in einer normwertigen NNE vor der Operation und in der postoperativ deutlichen Verschiebung der NNE zu pathologischen Werten wieder.

Untersuchungen von Hirano et al. (1988) und Zwirner et al. (1996) ergaben, dass die NNE, aber auch APQ und PPQ valide Parameter zur Dokumentation von postoperativer Stimmverbesserung (Verschlechterung) sind. In Übereinstimmung dazu lässt sich aus Tabelle 4 entnehmen, dass postoperativ diese drei akustischen Parameter bei auditiv festgestellter Stimmverschlechterung zu pathologischeren Werten verschoben sind. Besonders die Zunahme der Behauchtheit drückt sich hier deutlich in der schlechteren NNE aus.

Allerdings sollten diese elektroakustischen Parameter nicht allzu kritiklos hingenommen werden. Vor allem zur Bewertung des Jitters (PPQ) und des Shimmers (APQ) sollte immer der zeitliche Amplituden- und Frequenzverlauf des gesamten Stimmsignals mitberücksichtigt werden. Aus der Tabelle 4 entnehmen wir eine Shimmererhöhung (APQ über 5 Perioden) von 3,87% praeoperativ auf 9,91% (APQ) postoperativ. Die Shimmererhöhung könnte aber

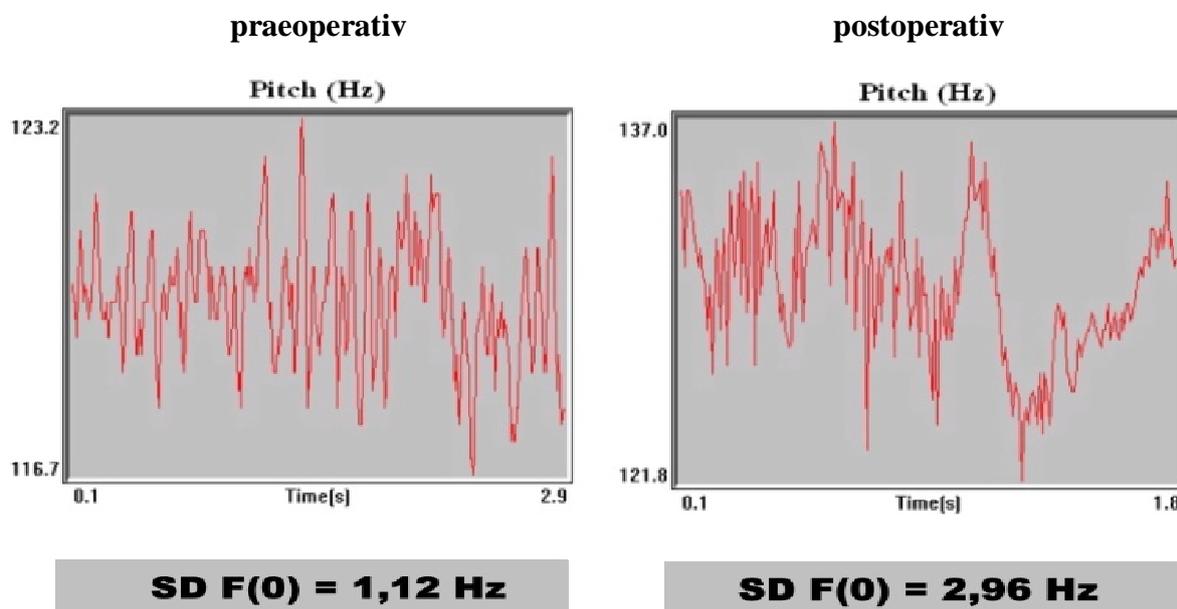
gerade in diesem Fall durch die in der postoperativen Aufnahme vorliegende kontinuierliche Abnahme der Intensität des Stimmsignals (Abnahme der Lautstärke, respektive der Amplitude) durch die deutlich verkürzte Tonhaldedauer nach Laterofixation bedingt sein (siehe hierzu Oszillogramm der Stimmaufnahme in Abb. 8b unten; und den zeitlichen Verlauf der Amplitude in Abb. 11a). Es ist erkennbar, wie schwierig die Trennung von Amplitudenschwankungen benachbarter Perioden (Shimmer) von längerfristigen Lautstärkeschwankungen des Stimmsignals ist.

Abbildung 11a: zeitlicher Amplitudenverlauf über das gesamte Stimmsignal, vor und nach Laterofixation (Fall 1)



Ähnlich verhält es sich in diesem Fall mit dem Jitter (PPQ). Nach Laterofixation nimmt der Jitter von praeoperativ 0,32% (PPQ über 5 Perioden) auf postoperativ 0,86% (PPQ) (Tabelle 4) zu. Betrachtet man jetzt den Frequenzverlauf des Gesamtsignals und auch die Standardabweichung von $F(0)$ (Abbildung 11b) postoperativ, fällt zum einen eine stärkere Unregelmäßigkeit des Frequenzverlaufes und eine höhere Schwankungsbreite der Grundfrequenz postoperativ auf, sowie zum anderen eine dadurch bedingte höhere Standardabweichung von der Grundfrequenz (SD von $F(0)$). Dies und auch die zunehmende NNE bzw. Behauchung könnte zur (falschen) „Erhöhung des Jitters beitragen und somit auch die elektroakustische Einstufung der Rauigkeit verfälschen. Schon Untersuchungen von Schultz-Coulon 1979 und anderen zeigten, dass die Messung des Jitters auch von physiologischen und krankhaften Tonhöhenschwankungen und auch von überlagerndem Rauschen beeinflusst wird und nicht nur von kurzzeitigen Periodenunregelmäßigkeiten beim Vergleich von Periode zu Periode.

Abbildung 11b: zeitlicher Frequenzverlauf über das gesamte Stimmsignal, vor und nach Laterofixation (Fall1), darunter stehen jeweils die prae- und postoperativen Standardabweichungen von der Grundfrequenz (SD F(0)). Beachte auch, dass der gewählte Frequenzbereich (Ordinate) in der postoperativen Darstellung deutlich größer ist. (praeoperativ=6,5 Hz; postoperativ=15,2 Hz)



3. 1. 1. 6. **Vergleich von auditiver und computergestützter Heiserkeits-Bewertung im RBH-Schema (Fall1)**

Die elektroakustische RBH-Beurteilung durch Dr. Speech ® "stimmt bezüglich der Bewertung der Behauchtheit mit der auditiven überein (siehe Tabelle 5 und Abb. 12a und b). Beide, Dr. Speech ® und der Untersucher, stellen eine postoperative Zunahme der Behauchtheit in identischem Ausmaß fest. Da die Bestimmung des Behauchtheitsgrades durch Dr. Speech ® hauptsächlich auf den Werten für die NNE basiert (s.o.), ist die vom Computer angezeigte starke Zunahme der Behauchtheit auf die postoperative NNE-Erhöhung zurückzuführen.

Ebenfalls konstatieren sowohl die auditive als auch die computergestützte Analyse eine postoperative Zunahme der Heiserkeit. Durch den Untersucher wird die Zunahme jedoch stärker bewertet. Auditiv beruht die Zunahme der Heiserkeit vor allem auf der stärkeren Behauchtheit durch insuffizienten Glottisschluss nach Laterofixation. Diese Zunahme der Behauchtheit stellt auch Dr. Speech ® "(anhand der NNE Erhöhung) fest. Für die Bestimmung des Heiserkeitsgrades legen die internen Algorithmen der Software Dr. Speech

® über hauptsächlich den Shimmer (APQ) zugrunde (siehe Einleitung). Die postoperative Shimmererhöhung könnte aber wie oben beschrieben durch die kontinuierliche Abnahme der Intensität des Stimmsignals bedingt sein (s.o.). So ist die übereinstimmend festgestellte Zunahme der Heiserkeit einerseits auf die auditiv wahrgenommene Erhöhung der Behauchung gegründet und andererseits vor allem auf die gemessene Shimmererhöhung.

Tabelle 5: Gegenüberstellung der auditiven und computergestützten RBH-Beurteilung des Stimmklangs des Patienten von Fall 1 (Originalabbildung der computergestützten RBH-Beurteilung durch „Dr.Speech ®“ in Abbildung 12a und b)

	praeoperativ (auditiv)	praeoperativ (computer)	postoperativ (auditiv)	postoperativ (computer)
Rauhigkeit	1	0	1	2
Behauchtheit	0	0	3	3
Heiserkeit	1	1	3	2

Unterschiedlich ist die Beurteilung der Rauhigkeit. Während für den Untersucher die Stimme prae- und postoperativ bezüglich der Rauhigkeit gleich bleibt, wird sie durch die elektroakustische RBH-Beurteilung von Dr. Speech ® ‘postoperativ wesentlich rauher bewertet (praeoperative Rauhigkeit = 0 ? postoperative Rauhigkeit = 2). Die hier fälschlicherweise ‘erhöhten Jitterwerte (siehe voriges Kapitel) könnten erklären, warum auditive und elektroakustische Einstufung der Rauhigkeit in Fall 1 voneinander abweichen. Zusammenfassend lässt sich feststellen, dass bezüglich Fall 1 in erster Linie die Behauchtheitsbewertung von Untersucher und Dr. Speech ® übereinstimmen. Auch bei der Beurteilung der Heiserkeit kommen beide zu ähnlichen Ergebnissen. Die Rauhigkeit wird jedoch von beiden deutlich unterschiedlich eingestuft.

Abbildung 12a: praeoperative (vor Laterofixation) computergestützte Heiserkeitsbeurteilung im RBH-Schema (Fall1) durch „Dr. Speech“[®] “(normal=0, slight=1, moderate=2, extreme=3) Die Beurteilung ist auch in Tabelle 5 aufgeführt.

Hoarse Voice	Normal	Slight	Moderate	Extreme
Harsh Voice	Normal	Slight	Moderate	Extreme
Breathy Voice	Normal	Slight	Moderate	Extreme

Abbildung 12b: postoperative (nach Laterofixation) computergestützte Heiserkeitsbeurteilung im RBH-Schema (Fall1) durch „Dr. Speech“[®] “(normal=0, slight=1, moderate=2, extreme=3) Die Beurteilung ist auch in Tabelle 5 aufgeführt.

Hoarse Voice	Normal	Slight	Moderate	Extreme
Harsh Voice	Normal	Slight	Moderate	Extreme
Breathy Voice	Normal	Slight	Moderate	Extreme

3. 1. 2. Fall 2: Laserchordektomie bei Stimmlippenkarzinom (T1a)

Als zweites sei ein 69 jähriger Raucher vorgestellt, bei dem wegen eines Stimmlippenkarzinoms eine Laserchordektomie links durchgeführt wurde. Den Patient führte eine seit einem Jahr zunehmende Heiserkeit zum Arzt. Spiegelbefundlich erkannte man breite leukoplakische Auflagerungen der linken Stimmlippe.

3. 1. 2. 1. *Auditive prae- und postoperative Heiserkeitsbewertung im RBH-Schema (Fall2)*

Vor allem die maximale Rauigkeit (R=3) und eine dezente Behauchtheit (B=1) führten zu einer sehr heiseren praeoperativen Stimme (H=3) (siehe Tabelle 6). Diesmal war der Stimmklang nicht durch ein Rauschen von unmodulierter Luft bestimmt sondern es kam offenbar durch die beginnende Karzinominfiltration zu Unregelmäßigkeiten und Asynchronizitäten in den Stimmlippenschwingungen, die als Rauigkeit imponierten. Zudem klang die Stimme sehr gepresst, im Sinne einer zusätzlichen hyperfunktionellen Komponente. 9 Wochen nach dem Eingriff war der Patient deutlich zufriedener mit seinem Stimmklang, und auch auditiv ließ sich vom Untersucher eine Besserung feststellen (R,B,H=1). Die Stimme war jedoch noch gepresst und zu hoch.

Tabelle 6: auditive RBH-Beurteilung des Stimmklangs des Patienten von Fall 2 (Patient 10 aus Tabelle 1)

	praeoperativ	postoperativ
Rauhigkeit	3	1
Behauchtheit	1	1
Heiserkeit	3	1

3. 1. 2. 2. Vergleich der prae- und postoperativen Oszillogramme (Fall 2)

Die postoperative Verbesserung der Stimme ist in Fall 2 ebenfalls bereits aus den Oszillogrammen (Abb. 13a und b) ablesbar, wenn auch weniger deutlich als in Fall 1. Die Schwingungsunregelmäßigkeiten der Stimmlippen stellen sich durch Unterschiede der einzelnen Perioden im praeoperativen Oszillogramm dar. Jede Periode unterscheidet sich deutlich in ihrer maximalen Amplitudengröße, Periodenlänge und Wellenform von den sie umgebenden Perioden. Diese Schwingungsunregelmäßigkeiten der Perioden, die akustisch als Heiserkeit auffallen, sind postoperativ geringer ausgeprägt. Die einzelnen Perioden im postoperativen Oszillogramm (Abb. 13b) sind sich untereinander wesentlich ähnlicher, sowohl in ihrer Maximalamplitude als auch in Periodenlänge und Wellenform. Die schwarzen Balken (A und B) in den Oszillogrammen deuten die Differenz der größten und der kleinsten Periodenamplitude über diese 7 Perioden an. A, (praeoperativ) ist deutlich größer als B, “ (postoperativ).

Abbildung 13a: *Oszillogramm der praeoperativen Stimmaufnahme von Fall 2 (vor Laserchordektomie), unten: Gesamtsignal (3 sec); oben: Ausschnitt mit sieben Perioden*

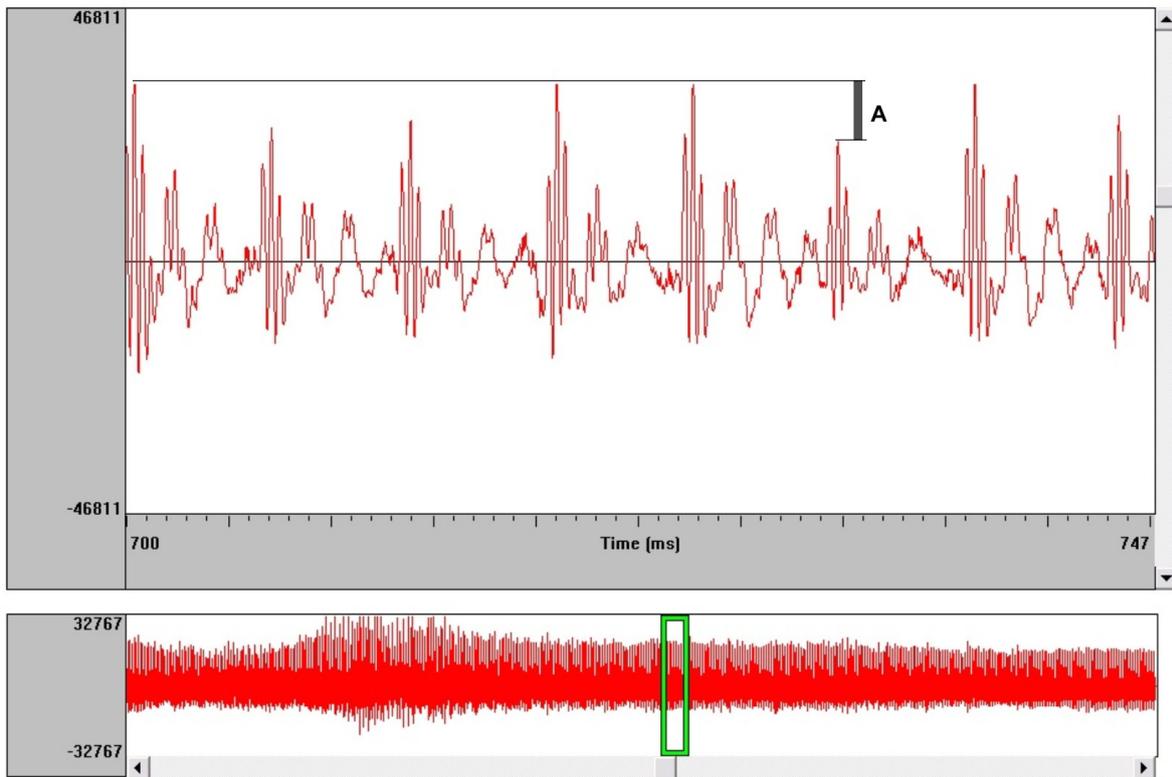
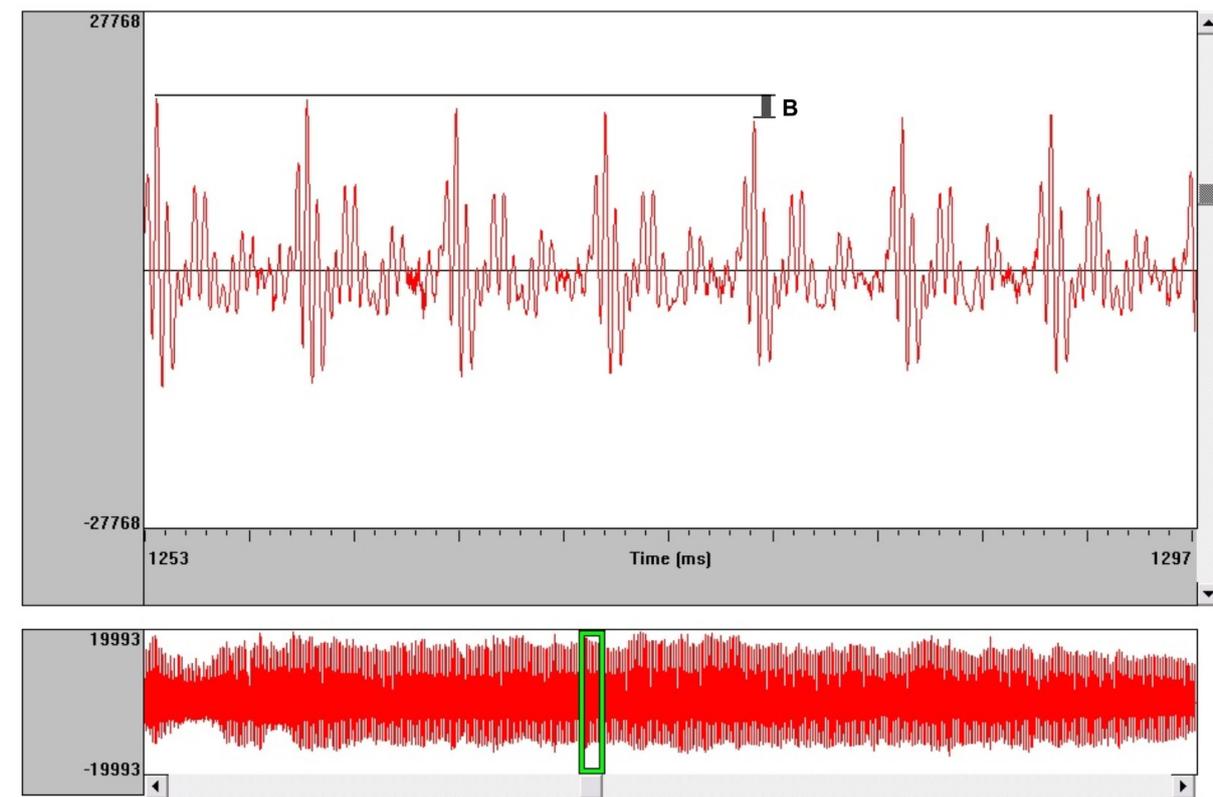


Abbildung 13b: *Oszillogramm der postoperativen Stimmaufnahme von Fall 2 (nach Laserchordektomie), unten: Gesamtsignal (3 sec); oben: Ausschnitt mit sieben Perioden*



3. 1. 2. 3. Vergleich der prae- und postoperativen Spektralanalysen (Fall 2)

Die postoperative Verbesserung des Stimmklangs lässt sich auch anhand der LTA-Spektren nachweisen.

- 1) Postoperativ lassen sich Harmonische bis zu einem Frequenzbereich von ca. 2,5 KHz abgrenzen. Im Gegensatz zum praeoperativen Spektrum, in dem die Harmonischen nur schwach ausgeprägt sind (schwarze Punkte in den Abbildungen 14a und b). Die Struktur der Teiltöne ist hier fast völlig zerstört.
- 2) Auch die Formantregionen sind postoperativ besser abzugrenzen.
- 3) Die postoperative Abnahme der absoluten Energie oberhalb von 5 kHz spricht ebenfalls für eine Verringerung der Rauschanteile im postoperativen Stimmklang

Abbildung 14a: Spektralanalyse der praeoperativen Stimmaufnahme von Fall 2 (vor Chordektomie), rote Kurve: LTAS-Spektrum; schwarze Punkte: Harmonische; schwarze Linie: Energiemaxima von 5-8 kHz

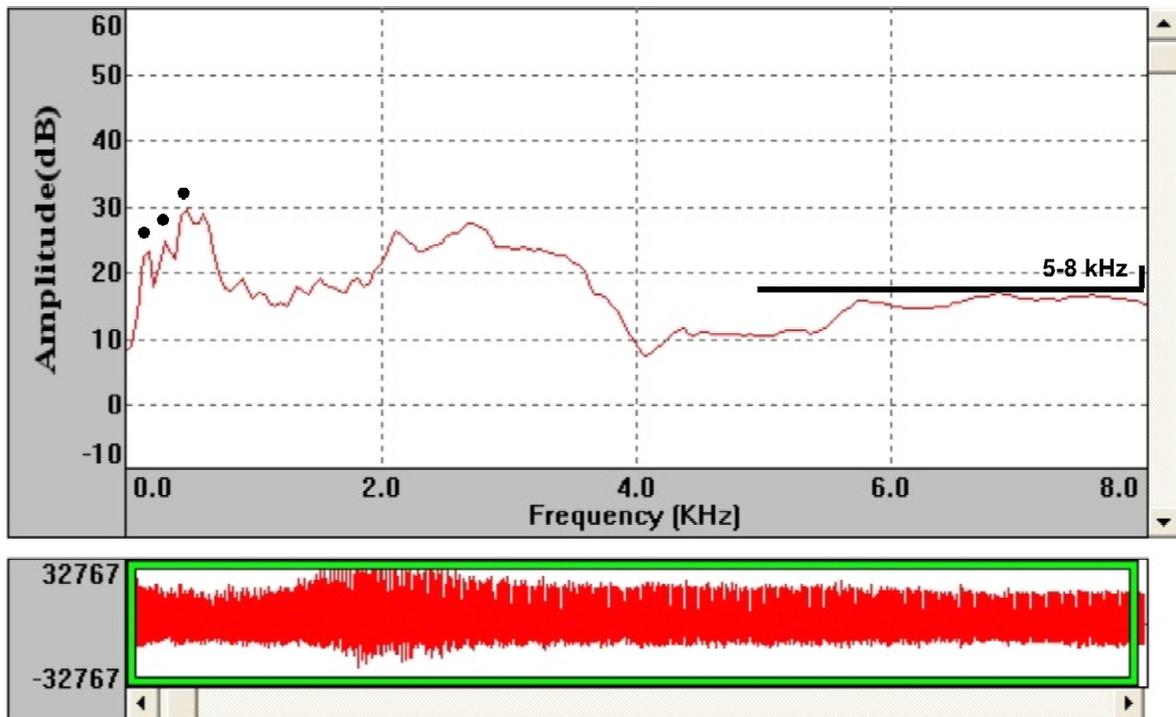
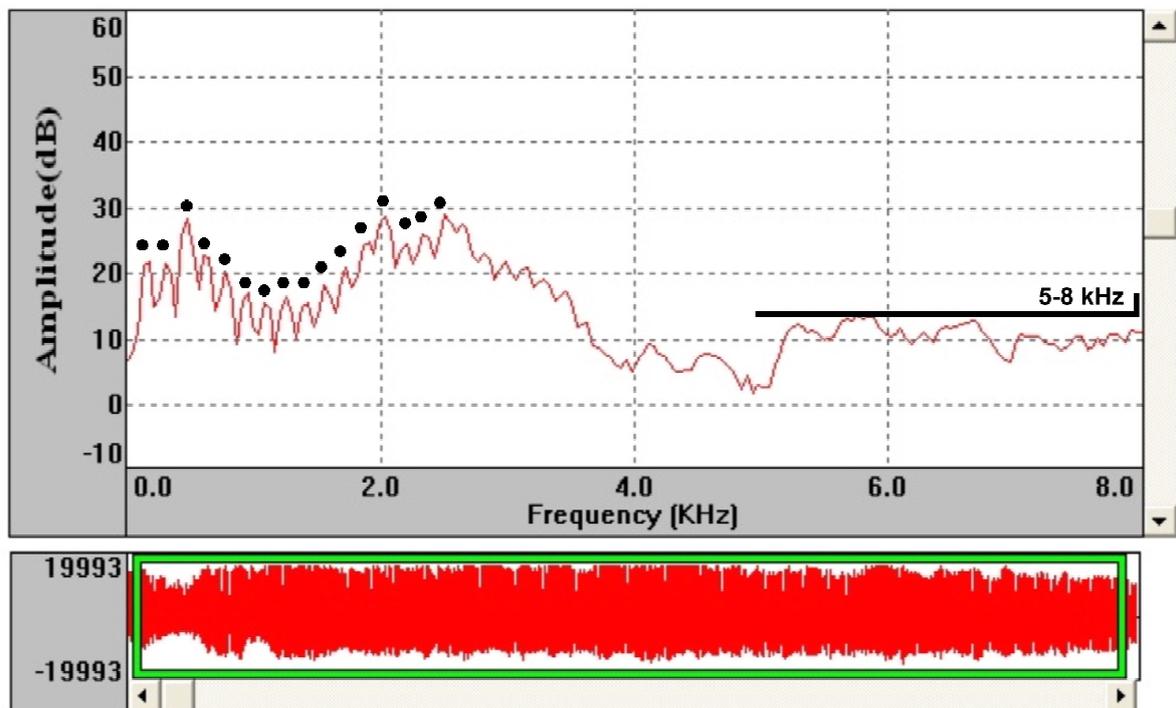


Abbildung 14b: Spektralanalyse der postoperativen Stimmaufnahme von Fall 2 (nach Chordektomie), rote Kurve: LTAS-Spektrum; schwarze Punkte: Harmonische; schwarze Linie: Energiemaxima von 5-8 kHz



3. 1. 2. 5. Vergleich der prae- und postoperativen Werte für APQ, PPQ und NNE von Fall 2

Aus Tabelle 7 ist zu entnehmen, dass sich APQ und PPQ postoperativ zu besseren, d.h. kleineren Werten verschieben. Dies bestätigt auch die postoperativ wahrgenommene Heiserkeits- und Rauigkeitsabnahme. Hirano zeigte (1988) anhand von 10 Patienten mit T1a Stimmlippenkarzinomen, dass APQ und PPQ (sowie NNE) nur bei den Patienten postoperativ erniedrigt waren, bei denen sich auch die Stimmqualität verbessert hatte. Dies bestätigt dieser Fall.

Analog zur gleichbleibenden auditiven Behauchtheitsbewertung verändert sich die NNE, die allgemein als Behauchtheitsmaß gilt, kaum. Ihr Wert verschiebt sich postoperativ geringfügig, jedoch tendenziell zu normaleren Werten. Weiterhin erwähnenswert ist die erhöhte Lage der Grundfrequenz (F(0)) prae- und postoperativ (Tabelle 7). Auditiv wurde die Stimme als zu hoch und gepresst im Sinne einer Hyperfunktion wahrgenommen. Arbeiten von Hammarberg et al. (1980) zeigen eine Erhöhung der MSSL (mittlere Sprechstimmlage) durch erhöhten Muskeltonus der Stimmlippen wie bei der hyperfunktionellen Stimme. Die erhöhte F0 kann hier ein Hinweis auf die Hyperfunktion geben, jedoch kann man die MSSL und die anhand eines gehaltenen Vokals ermittelte F0 nicht gleichsetzen, da die MSSL korrekterweise durch verschiedene Verfahren aus fortlaufender Sprache (gelesen oder spontan) ermittelt wird. Außerdem hängt die MSSL von zahlreichen Faktoren ab (z. B. Stimmung des Patienten und Art des Stimmsignals), und neuere Untersuchungen von Baken (1996) ergaben keinen signifikanten Zusammenhang bestimmter Stimmstörungen und der Höhe der MSSL. Desweiteren unterliegt die MSSL hohen inter- und intraindividuellen Schwankungen (Atkinson 1976).

Tabelle 7: prae- zu posttherapeutische Veränderung von 4 akustischen Parametern in Fall 2.

Parameter	prae	post	normal (männlich)	pathologisch (männlich)
PPQ(Jitter) [%]	1,58	0,39	0,631 (mean) ¹	1,264 (mean) ¹
APQ(Shimmer) [%]	7,81	5,38	3,027 (mean) ¹	4,811 (mean) ¹
F (0) [Hz]	152	159	98-131 (MSSL) ³	außerhalb Norm ³
NNE[dB]	-10,59	-11,61	< -11 ²	> -11 ²

¹ nach Boyanov et al. 1997 ² nach Kasuya et al. 1986 ³ nach Schultz-Coulon 1980

3. 1. 2. 6. Vergleich von auditiver und computergestützter Heiserkeits-Bewertung im RBH-Schema (Fall2)

Die computergestützte Beurteilung (siehe Tabelle 8 und Abb. 16a und b) zeigt genaue Übereinstimmung mit den Ergebnissen des Untersuchers bei der Bewertung der prae- und postoperativen Rauigkeit und Behauchtheit. Die praeoperative Heiserkeit bewerten der Untersucher und Dr. Speech[®] jedoch sehr unterschiedlich. Während der Untersucher die Heiserkeit praeoperativ als maximal (3) einstuft, bewertet sie Dr. Speech[®] „als geringgradig. Die postoperative Heiserkeit wird von beiden wieder übereinstimmend als geringgradig bewertet.

Tabelle 8: Gegenüberstellung der auditiven und computergestützten RBH-Beurteilung des Stimmklangs des Patienten von Fall 2 (Originalabbildung der computergestützten RBH-Beurteilung durch „Dr.Speech[®]“ in Abbildung 16a und b)

	praeoperativ (auditiv)	praeoperativ (computer)	postoperativ (auditiv)	postoperativ (computer)
Rauhigkeit	3	3	1	1
Behauchtheit	1	1	1	1
Heiserkeit	3	1	1	1

Abbildung 16a: praeoperative computergestützte Heiserkeitsbeurteilung im RBH-Schema (Fall 2)

Hoarse Voice	Normal	Slight	Moderate	Extreme
Harsh Voice	Normal	Slight	Moderate	Extreme
Breathy Voice	Normal	Slight	Moderate	Extreme

Abbildung 16b: postoperative computergestützte Heiserkeitsbeurteilung im RBH-Schema (Fall 2)

Hoarse Voice	Normal	Slight	Moderate	Extreme
Harsh Voice	Normal	Slight	Moderate	Extreme
Breathy Voice	Normal	Slight	Moderate	Extreme

3. 2. Ergebnisse der statistischen Auswertung der Studie

3. 2. 1. Statistische Auswertung der Bewertungsergebnisse der subjektiven und objektiven (computergestützten) Stimmklangbewertung im RBH-Schema

3. 2. 1. 1. Übereinstimmungsquoten zwischen subjektiver und computergestützter RBH-Beurteilung

Den folgenden Berechnungen, Tabellen und Abbildungen dient die Tabelle 1 des Anhangs als Datengrundlage. Dort sind alle Bewertungsergebnisse des Untersuchers und der Computeranalyse im Einzelnen aufgeführt.

1) Rauigkeit

Die Rauigkeit (Abbildung 17) wird in 32 von insgesamt 124 Bewertungen (post- und praeoperativ gemeinsam) identisch von Dr. Speech® und dem Untersucher beurteilt“ (25,8% der Fälle). Praeoperativ sind 14 von 62 Beurteilungen identisch (22,6%) und postoperativ 18 von 62 (29%).

Für die Rauigkeitsbeurteilungen des Untersuchers und die Werte der computergestützten Rauigkeitsbeurteilung besteht ein statistisch signifikanter Zusammenhang bezüglich des Korrelationsquotienten R_s (nach Spearman), da die Irrtumswahrscheinlichkeit p ,“ deutlich $< 0,05$ ist (Tabelle 9). Für R_s ergibt sich ein Wert von 0,6 ($n=124$; prae- und postoperative Werte gemeinsam). Der Zusammenhang ist also direkt, d.h. geringgradige Beurteilungen des Untersuchers gehen tendenziell auch mit geringgradigen Werten der computergestützten Rauigkeitsbewertung einher.

Berechnet man aber den Konkordanzkoeffizienten Kappa (nach Cohen), der mehr die exakten Übereinstimmungen berücksichtigt, ergibt sich kein statistischer Zusammenhang (Irrtumswahrscheinlichkeit $p=0,221639$, Tabelle 10).

Zusammenfassend wird die Rauigkeit subjektiv und computergestützt tendenziell ähnlich beurteilt, seltener (statistisch nicht signifikant) werden die Stimmklänge identisch bewertet.

Tabelle 9: Korrelationsquotient Spearman R für die Bewertungsdaten des Merkmals „Rauhigkeit“ von „Dr. Speech ®“ und Untersucher; „Speech R“ entspricht hier den prae- und postoperativen Beurteilungen der Rauhigkeit von „Dr. Speech ®“; „Unters R“ entspricht hier den prae- und postoperativen Beurteilungen der Rauhigkeit des Untersuchers

Spearman's Rangkorrelation				
Variablenpaar	Gültige N	Spearman R	t(N-2)	p-Niveau
Speech R & Unters R	124	,600308	8,29067	,000000

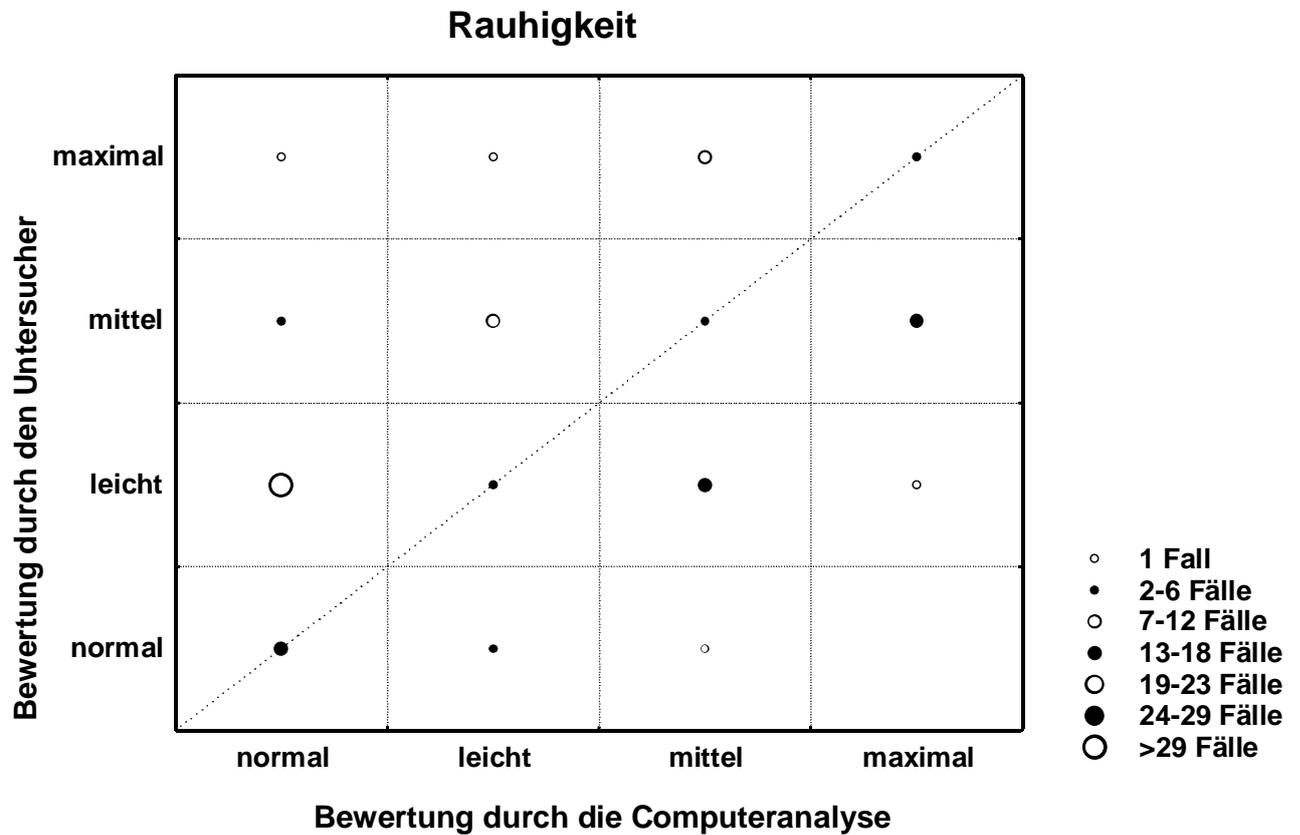
(Für die prae- und postoperativ getrennte Berechnung von Spearman's R ergaben sich ähnliche Ergebnisse, die hier nicht aufgeführt werden)

Tabelle 10: Cohens KAPPA für die Bewertungsdaten des Merkmals „Rauhigkeit“ von „Dr. Speech ®“ und Untersucher; n=124 prae- und postoperative Beurteilungen; „Speech R“ = Rauhigkeitsbeurteilungen durch „Dr. Speech ®“; „Unters R“ = Rauhigkeitsbeurteilungen durch den Untersucher

Häufigkeitstabelle Speech R x Unters R						Cohen's Kappa = 0.053514
	Unters R normal	Unters R leicht	Unters R mittel	Unters R maximal	Zeile Gesamt	Prüfgröße Z = 1.22218
Speech R normal	16	35	5	1	57	mit p = 0.221639
Speech R leicht	3	5	7	1	16	
Speech R mittel	1	15	5	9	30	
Speech R maximal	0	1	14	6	21	
Alle	20	56	31	17	124	

(Für die prae- und postoperativ getrennte Berechnung von Cohens Kappa ergaben sich ähnliche Ergebnisse, die hier nicht aufgeführt werden)

Abbildung 17: Darstellung der subjektiv und computergestützt ermittelten Bewertungsdaten für die Stimmqualität „Rauhigkeit“; unabhängig davon, ob die Befunde prae- oder postoperativ erhoben wurden (n=124). Bei identischer Bewertung liegen die Punkte auf der gestrichelten Diagonale



2) Behauchtheit

Die Behauchtheit (Abbildung 18) wird in wesentlich mehr Fällen als die Rauigkeit identisch von Untersucher und Dr. Speech[®] bewertet. Von 124 Bewertungen (Prae- und postoperativ gemeinsam) sind 105 übereinstimmend (84,7% aller Fälle). Prae- und postoperativ unterscheiden sich die Anteile der identischen Beurteilungen kaum (praeoperativ: 52 von 62 Beurteilungen identisch (83,9%); postoperativ: 53 von 62 (85,5%)).

Für die Behauchtheitsbeurteilung des Untersuchers und die Werte der computergestützten Behauchtheitseinstufung existiert ein straffer und direkter statistischer Zusammenhang für Spearmans R ($R=0,87$; $p<0,0001$; $n=124$; Tabelle 11). Auch die Konkordanz, d.h. die tatsächlich identischen Bewertungen ist sehr hoch (Cohens Kappa= $0,78$; $p<0,0001$; $n=124$; Tabelle 12).

Abbildung 18 stellt die hohe Zahl an exakt übereinstimmenden Bewertungen grafisch dar. Die meisten Punkte liegen auf der diagonalen Linie, die die genaue Übereinstimmung markiert.

Zusammenfassend wird die Behauchtheit eines Stimmklangs computergestützt und subjektiv sehr häufig identisch beurteilt .“

Tabelle 11: Korrelationsquotient Spearman R für die Bewertungsdaten des Merkmals „Behauchtheit“ von „Dr. Speech[®]“ und Untersucher; „Speech B“ entspricht hier den prae- und postoperativen Beurteilungen der Behauchtheit von „Dr. Speech[®]“; „Unters B“ entspricht hier den prae- und postoperativen Beurteilungen der Behauchtheit des Untersuchers

Spearman's Rangkorrelation				
Variablenpaar	Gültige N	Spearman R	t(N-2)	p-Niveau
Speech B & Unters B	124	,872860	19,75748	0,000000

(Für die prae- und postoperativ getrennte Berechnung von Spearmans R ergaben sich nahezu identische Ergebnisse, die hier nicht aufgeführt werden)

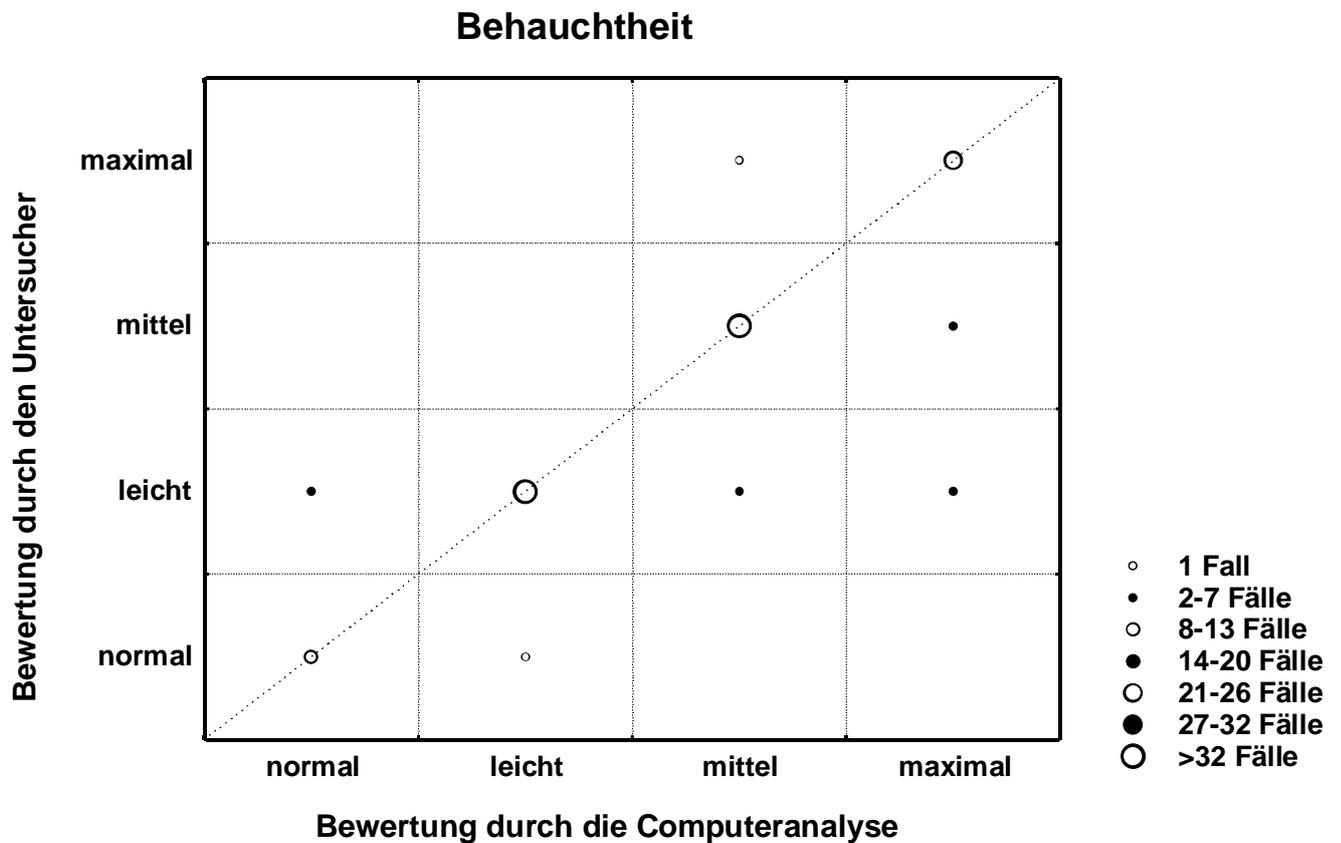
Tabelle 12: Cohens KAPPA für die Bewertungsdaten des Merkmals „Behauchtheit“ von „Dr. Speech®“ und Untersucher; n=124 prae- und postoperative Beurteilungen; „Speech B“ = Behauchtheitsbeurteilungen durch „Dr. Speech®“; „Unters B“ = Behauchtheitsbeurteilungen durch den Untersucher

Häufigkeitstabelle Speech B x Unters B					
	Unters B normal	Unters B leicht	Unters B mittel	Unters B maximal	Zeile Gesamt
Speech R normal	8	2	0	0	10
Speech R leicht	1	39	0	0	40
Speech R mittel	0	5	35	1	41
Speech R maximal	0	3	7	23	33
Alle	9	49	42	24	124

Cohen's Kappa = 0.782094
 Prüfgröße Z = 13.90720
 mit p = 0.000000

(Für die prae- und postoperativ getrennte Berechnung von Cohens Kappa ergaben sich nahezu identische Ergebnisse, die hier nicht aufgeführt werden)

Abbildung 18: Darstellung der subjektiv und computergestützt ermittelten Bewertungsdaten für die Stimmqualität „Behauchtheit“; unabhängig davon, ob die Befunde prae- oder postoperativ erhoben wurden (n=124). Bei identischer Bewertung liegen die Punkte auf der gestrichelten Diagonale



3) Heiserkeit

Eine übereinstimmende Bewertung der Heiserkeit (Abbildung 19) liegt in 44 von 124 Fällen (prae- und postoperativ gemeinsam) vor (35,5% der Fälle). Postoperativ werden mehr Fälle übereinstimmend vom Untersucher und Dr. Speech ® beurteilt als praeoperativ (praeoperativ: Übereinstimmung in 30,6% der Fälle ; postoperativ: 40,3%). Im Vergleich zu den Beurteilungen der Rauhigkeit und der Behauchtheit sind die Übereinstimmungen in den Bewertungsstufen zwischen Untersucher und Dr. Speech ® bezüglich des Merkmals Heiserkeit am geringsten ausgeprägt. Der Zusammenhang ist zwar statistisch signifikant ($p < 0,0001$, $n=124$), aber mit einem Korrelationsquotienten von $R = 0,56$ am geringsten ausgeprägt (gegenüber Rauhigkeit und Behauchtheit). Eine statistisch signifikante Konkordanz läßt sich gar nicht nachweisen ($p=0,46$). Die geringe Zahl der übereinstimmenden Bewertungen stellt sich auch in Abbildung 19 dar. Hier fällt außerdem auf, dass Dr. Speech ® die Stimmklänge tendenziell als geringergradig heiser einstuft (siehe 3.2.1.2.).

Tabelle 13: Korrelationsquotient Spearman R für die Bewertungsdaten des Merkmals „Heiserkeit“ von „Dr. Speech ®“ und Untersucher; „Speech H“ entspricht hier den prae- und postoperativen Beurteilungen der Heiserkeit von „Dr. Speech ®“; „Unters H“ entspricht hier den prae- und postoperativen Beurteilungen der Heiserkeit des Untersuchers

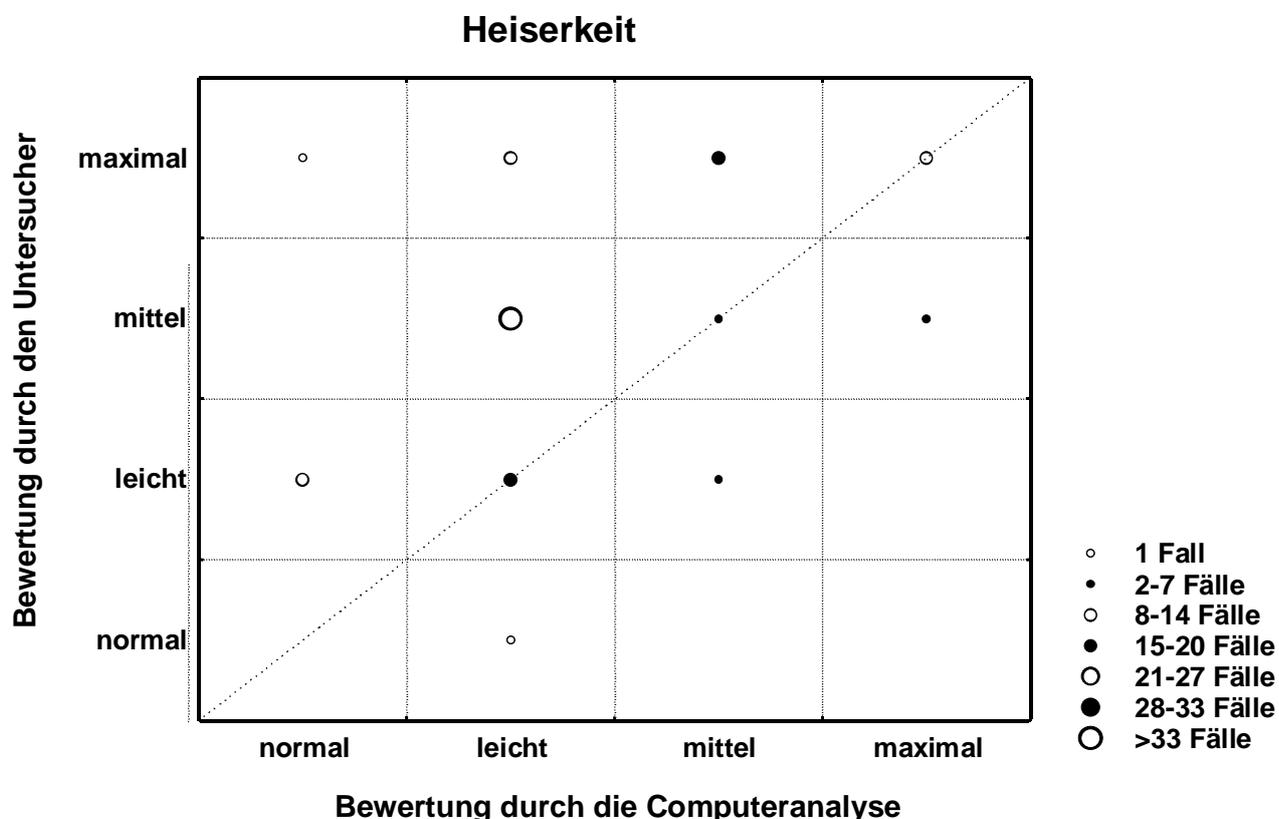
Spearman's Rangkorrelation				
Variablenpaar	Gültige N	Spearman R	t(N-2)	p-Niveau
Speech H & Unters H	124	,562146	7,50765	,000000

Tabelle 14: Cohens KAPPA für die Bewertungsdaten des Merkmals „Heiserkeit“ von „Dr. Speech ®“ und Untersucher; $n=124$ prae- und postoperative Beurteilungen; „Speech H“ = Heiserkeitsbeurteilungen durch „Dr. Speech ®“; „Unters H“ = Heiserkeitsbeurteilungen durch den Untersucher

Häufigkeitstabelle Speech H x Unters H						Cohen's Kappa = 0.036356
	Unters H normal	Unters H leicht	Unters H mittel	Unters H maximal	Zeile Gesamt	Prüfgröße Z = 0.73923
Speech H normal	0	9	0	1	10	mit p = 0.459768
Speech H leicht	1	17	40	12	70	
Speech H mittel	0	4	6	17	27	
Speech H maximal	0	0	3	14	17	
Alle	1	30	49	44	124	

(Für die prae- und postoperativ getrennte Berechnung von Cohens Kappa und Spearman's R ergaben sich geringfügig unterschiedliche Ergebnisse, die hier nicht aufgeführt werden)

Abbildung 19: Darstellung der subjektiv und computergestützt ermittelten Bewertungsdaten für die Stimmqualität „Heiserkeit“; unabhängig davon, ob die Befunde prae- oder postoperativ erhoben wurden (n=124). Bei identischer Bewertung liegen die Punkte auf der gestrichelten Diagonale



Zusammenfassend bestehen für alle drei Heiserkeitsparameter tendenzielle Übereinstimmungen (statistisch signifikant) zwischen den Beurteilungen des Untersuchers und von Dr. Speech®. „Die Korrelationskoeffizienten sind alle größer als 0, die Zusammenhänge sind also direkt. Geringgradige Beurteilungen des Untersuchers gehen tendenziell auch mit geringgradigen Beurteilungen des Computerprogramms einher und umgekehrt. Dabei ist der Zusammenhang für die Beurteilung der Behauchtheit mit Spearmans $R=0,87$ (straffer Zusammenhang) am höchsten. Sehr häufig kommt es dabei zu identischen subjektiven und computergestützten Bewertungen (Konkordanzkoeffizient Kappa =0,78 (Behauchtheit)). Die statistischen Zusammenhänge der Beurteilungen durch den Untersucher und Dr. Speech® bezüglich der Rauigkeit (Spearmans $R=0,60$) und der für die Heiserkeit (Spearmans $R=0,56$) sind geringer ausgeprägt. Das bedeutet, dass die identischen Beurteilungen hier seltener sind (keine statistische Signifikanz), die Stimmklänge jedoch tendenziell ähnlich bewertet werden.

3. 2. 1. 2. *Abweichungsquoten zwischen subjektiver und computergestützter RBH-Beurteilung*

Nicht nur die Übereinstimmungen zwischen den subjektiven und computergestützten Bewertungen sind von Interesse, sondern auch, in welchen Fällen und bei welchen Stimmqualitäten die Computeranalyse abweichend urteilt. Hier ist der Wilcoxon-Test für Paardifferenzen hilfreich. Er gibt Auskunft darüber, inwieweit sich die Beurteilungen in ihrer zentralen Tendenz signifikant unterscheiden. Die Unterschiede der zentralen Tendenz erkennt man durch Vergleich der Mittelwerte aller Beurteilungen, die Signifikanz an der Irrtumswahrscheinlichkeit p . Im Folgenden wurde bei den Berechnungen die Gesamtmenge der Beurteilungen, also prae- und postoperative RBH-Beurteilungen gemeinsam, zugrundegelegt ($n=124$), da sich für die getrennte prae- und postoperative Auswertung des Wilcoxon Tests ähnliche Ergebnisse ergaben.

1) **Rauhigkeit**

Höhergradig raue Stimmklänge (Rauhigkeit = 2 oder 3) werden subjektiv und computergestützt nahezu identisch beurteilt. Bei nur leichter Rauhigkeit (oder keine) treten jedoch deutliche Unterschiede auf (Tabelle 15a und b). Vom Untersucher werden nur 20 Stimmproben ($n=124$) als nicht rau (normal) eingestuft, dafür aber 56 als leicht rau ($R=1$). Die computergestützte Bewertung ist quasi genau umgekehrt. Hier werden 57 Stimmproben ohne Rauhigkeit bewertet und 16 sind leicht rau ($R=1$). Offenbar werden nur schwach ausgeprägte Störungen im Stimmklang (bezüglich der Rauhigkeit) von der computergestützten Analyse geringgradiger bewertet als vom Untersucher. Auch der Wilcoxon-Test für Paardifferenzen zeigt, dass es statistisch signifikante Unterschiede in der zentralen Tendenz der subjektiven und computergestützten Beurteilungen gibt (Irrtumswahrscheinlichkeit $p=0,016$, $n=124$). Erkennbar ist die geringgradigere Bewertung durch Dr. Speech® auch durch den niedrigeren Mittelwert aus allen Bewertungen gegenüber dem Mittelwert aus den Bewertungen des Untersuchers (Tabelle 15c: Mittelwert \pm Standardabweichung: Rauhigkeit Untersucher= $1,36 \pm 0,91$; Rauhigkeit Speech= $1,12 \pm 1,17$) Dies stellt sich auch in Abbildung 20 deutlich dar.

Tabelle 15a, 15b und 15c: Häufigkeitsverteilung der Bewertungsdaten des Merkmals „Rauhigkeit“ vom Untersucher (15a) und von „Dr. Speech®“ (15b); Tabelle 15c: Deskriptive Statistik: Mittelwert, Median, Standardabweichung (n=124, prae- und postoperativ gemeinsam)

Tabelle 15a

Untersucher Rauhigkeit (prae und post, n=124):				
Wert	Häufigk.	Summen-Häufigk	Prozent	Kumul. Prozent
normal	20	20	16,12903	16,1290
leicht	56	76	45,16129	61,2903
mittel	31	107	25,00000	86,2903
maximal	17	124	13,70968	100,0000

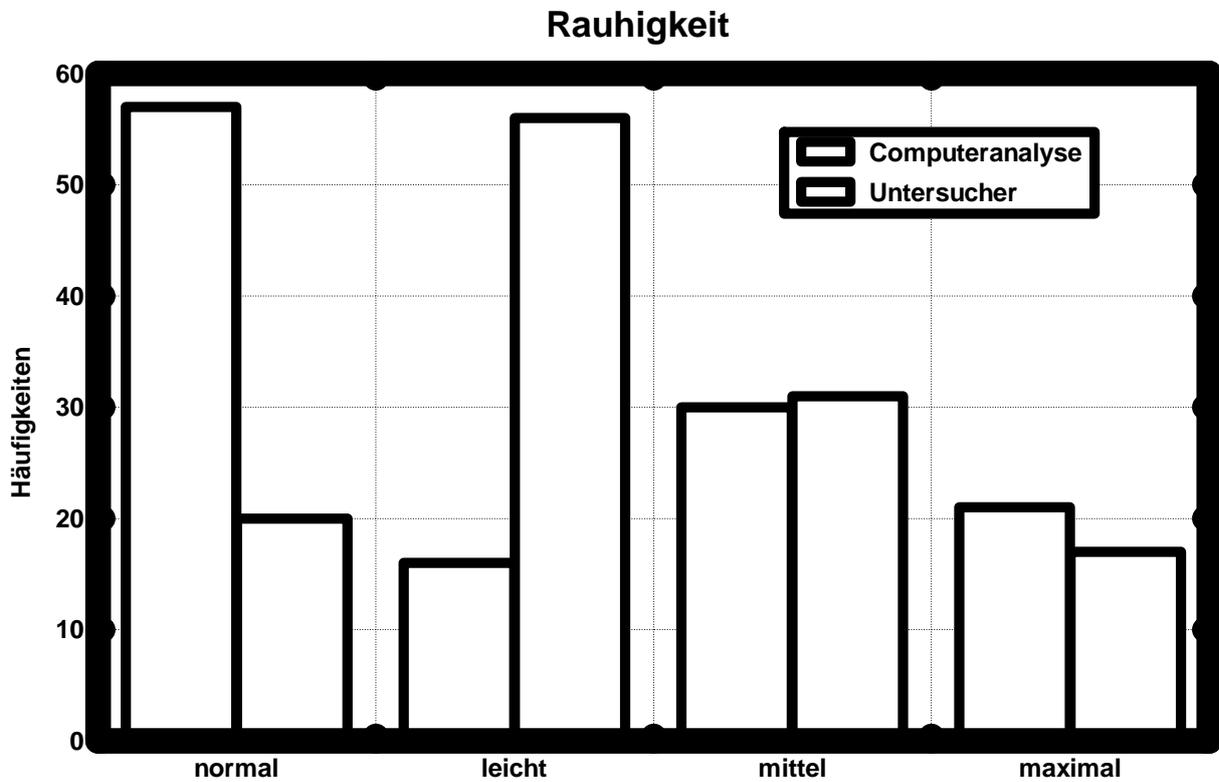
Tabelle 15b

Speech Rauhigkeit (prae und post, n=124)				
Wert	Häufigk.	Summen-Häufigk	Prozent	Kumul. Prozent
normal	57	57	45,96774	45,9677
leicht	16	73	12,90323	58,8710
mittel	30	103	24,19355	83,0645
maximal	21	124	16,93548	100,0000

Tabelle 15c

Deskriptive Statistik									
Variable	Gült. N	Mittelw.	Median	Minimum	Maximum	Unteres Quartil	Oberes Quartil	Quartil Abstand	Stdabw.
Speech_R	124	1,120968	1,000000	0,00	3,000000	0,000000	2,000000	2,000000	1,172814
Unters_R	124	1,362903	1,000000	0,00	3,000000	1,000000	2,000000	1,000000	,913625

Abbildung 20: Häufigkeitsverteilung der Bewertungsdaten der „Rauhigkeit“ von „Dr. Speech®“ und dem Untersucher (n=124).



2) Behauchtheit

Auch hier zeigt der Wilcoxon-Test für Paardifferenzen signifikante Unterschiede in der zentralen Tendenz der subjektiven und computergestützten Beurteilungen an (Irrtumswahrscheinlichkeit $p=0,0052$, $n=124$). Im Mittel ist hier, im Gegensatz zur Rauigkeit (s.o.) und auch zur Heiserkeit (s.u.), die computergestützte Bewertung der Behauchtheit höhergradiger gegenüber den Beurteilungen des Untersuchers (Tabelle 16c: Mittelwert \pm Standardabweichung: Behauchtheit Untersucher= $1,65 \pm 0,87$; Behauchtheit Speech= $1,78 \pm 0,93$). Der Mittelwert aller Beurteilungen des Untersuchers und aller Bewertungen von Dr. Speech® liegen aber deutlich näher beieinander (gegenüber Rauigkeit und Heiserkeit): „Die Unterschiede sind zwar signifikant, jedoch kommt es in nur 15,3% der Fälle überhaupt zu unterschiedlichen Bewertungen (84,7% identische Beurteilungen). Abbildung 21 zeigt, in Übereinstimmung zu Abbildung 18, dass die Unterschiede in den Beurteilungen der Behauchtheit zwischen dem Untersucher und Dr. Speech® wesentlich geringer ausgeprägt sind als bei der Beurteilung der Rauigkeit (Abbildung 20, s.o.) und der Heiserkeit (Abbildung 22, s.u.).

Tabelle 16a, 16b und 16c: Häufigkeitsverteilung der Bewertungsdaten des Merkmals „Behauchtheit“ vom Untersucher (16a) und von „Dr. Speech®“ (16b); Tabelle 16c: Deskriptive Statistik: Mittelwert, Median, Standardabweichung ($n=124$, prae- und postoperativ gemeinsam)

Tabelle 16a

Untersucher Behauchtheit (prae und post, n=124)				
Wert	Häufigk.	Summen-Häufigk	Prozent	Kumul. Prozent
normal	9	9	7,25806	7,2581
leicht	49	58	39,51613	46,7742
mittel	42	100	33,87097	80,6452
maximal	24	124	19,35484	100,0000

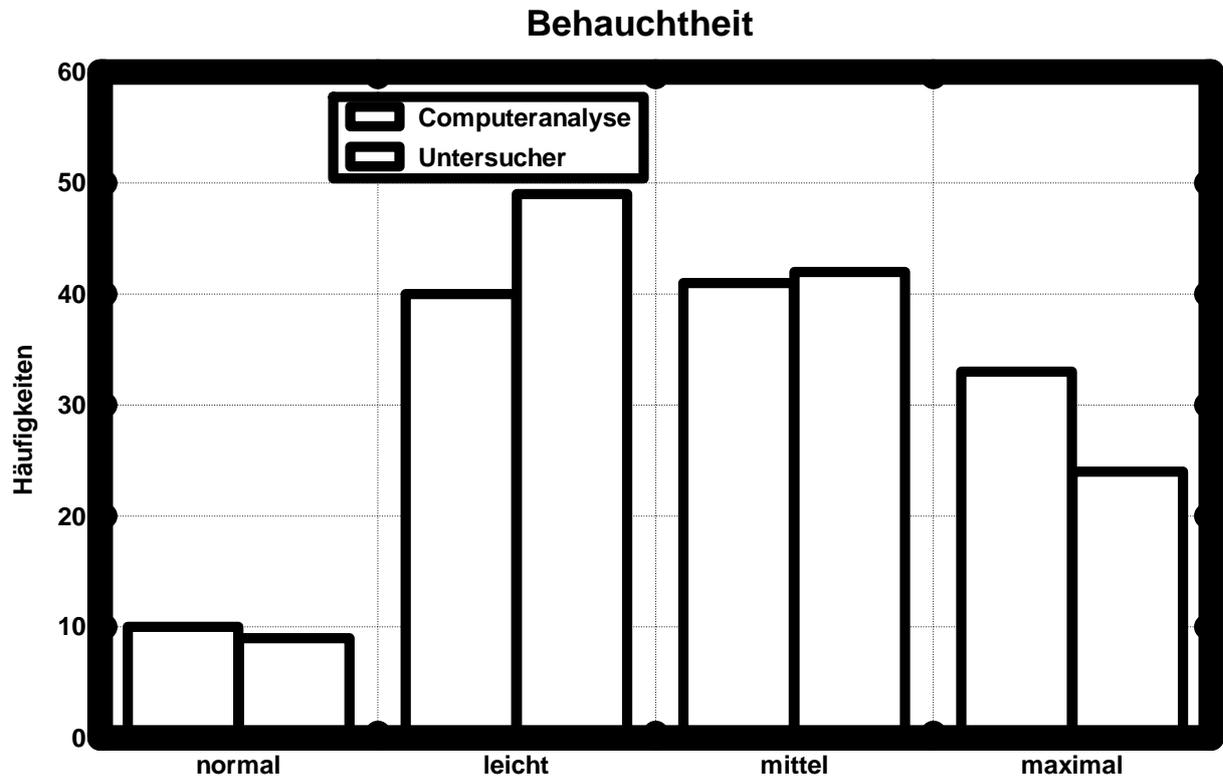
Tabelle 16b

Speech Behauchtheit (prae und post, n=124)				
Wert	Häufigk.	Summen-Häufigk	Prozent	Kumul. Prozent
normal	10	10	8,06452	8,0645
leicht	40	50	32,25806	40,3226
mittel	41	91	33,06452	73,3871
maximal	33	124	26,61290	100,0000

Tabelle 16c

Deskriptive Statistik									
Variable	Gült. N	Mittelw.	Median	Minimum	Maximum	Unteres Quartil	Oberes Quartil	Quartil Abstand	Stdabw.
Speech_B	124	1,782258	2,000000	0,00	3,000000	1,000000	3,000000	2,000000	,933221
Unters_B	124	1,653226	2,000000	0,00	3,000000	1,000000	2,000000	1,000000	,874614

Abbildung 21: Häufigkeitsverteilung der Bewertungsdaten der ‚Behauchtheit‘ von ‚Dr. Speech ®‘ und dem Untersucher (n=124).



3) Heiserkeit

Auch in Abbildung 22 sind, ähnlich wie in Abbildung 20 bei der Rauigkeit, „die Beurteilungen von Dr. Speech[®] bezüglich der Heiserkeit tendenziell geringgradiger als die des Untersuchers. Das heißt, Dr. Speech[®] „bewertet die Stimmproben durchschnittlich als weniger heiser als der Untersucher. Auch hier gibt der Wilcoxon-Test signifikante Unterschiede in der zentralen Tendenz der Beurteilungen an ($p < 0,0001$, $n = 124$). Dies drückt sich in den jeweils sehr unterschiedlichen Mittelwerten aller Beurteilungen des Untersuchers und aller Bewertungen von Dr. Speech[®] aus (Tabelle 17c: Mittelwert \pm Standardabweichung: Heiserkeit Untersucher: $2,10 \pm 0,79$; Heiserkeit Speech: $1,41 \pm 0,83$).

Tabelle 17a, 17b und 17c: Häufigkeitsverteilung der Bewertungsdaten des Merkmals „Heiserkeit“ vom Untersucher (17a) und von „Dr. Speech[®]“ (17b); Tabelle 17c: Deskriptive Statistik: Mittelwert, Median, Standardabweichung ($n = 124$, prae- und postoperativ gemeinsam)

Tabelle 17a

Untersucher Heiserkeit (prae und post, n=124)				
Wert	Häufigk.	Summen-Häufigk	Prozent	Kumul. Prozent
normal	1	1	,80645	,8065
leicht	30	31	24,19355	25,0000
mittel	49	80	39,51613	64,5161
maximal	44	124	35,48387	100,0000

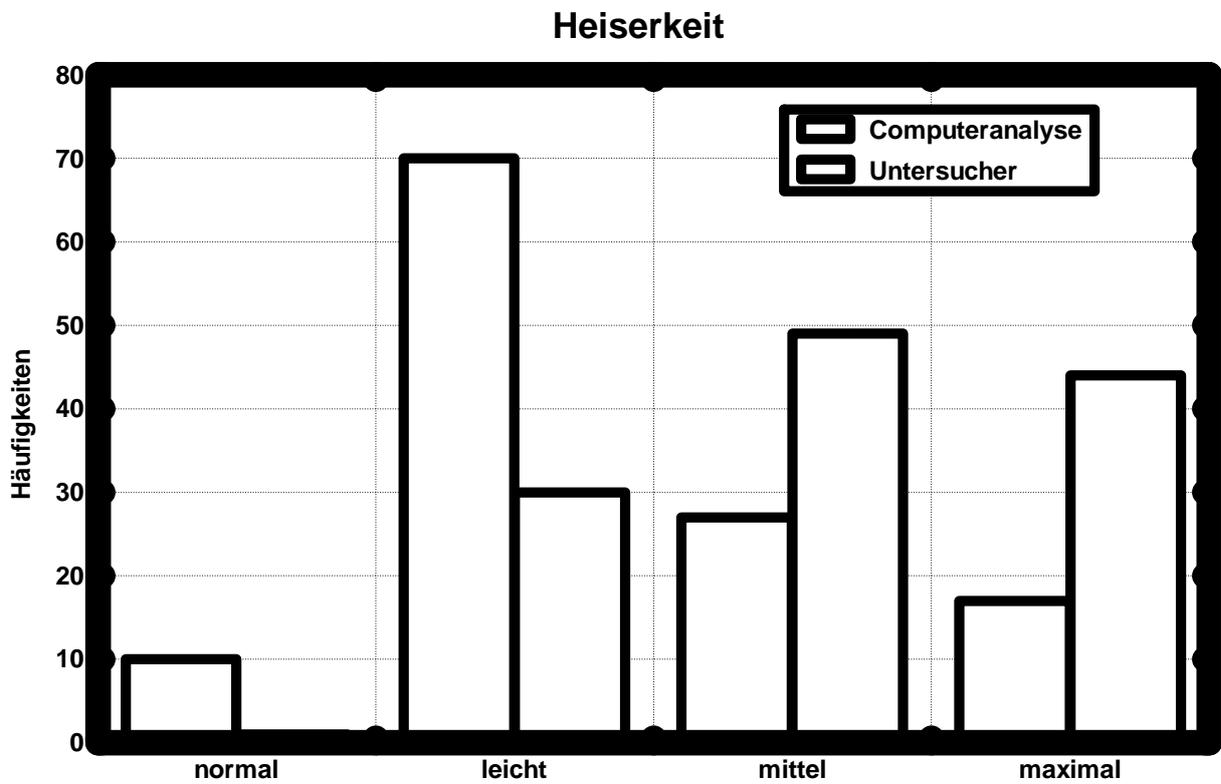
Tabelle 17b

Speech Heiserkeit (prae und post, n=124)				
Wert	Häufigk.	Summen-Häufigk	Prozent	Kumul. Prozent
normal	10	10	8,06452	8,0645
leicht	70	80	56,45161	64,5161
mittel	27	107	21,77419	86,2903
maximal	17	124	13,70968	100,0000

Tabelle 17c

Deskriptive Statistik									
Variable	Gült. N	Mittelw.	Median	Minimum	Maximum	Unteres Quartil	Oberes Quartil	Quartil Abstand	Stdabw.
Speech_H	124	1,411290	1,000000	0,00	3,000000	1,000000	2,000000	1,000000	,826513
Unters_H	124	2,096774	2,000000	0,00	3,000000	1,500000	3,000000	1,500000	,790383

Abbildung 22: Häufigkeitsverteilung der Bewertungsdaten der „Heiserkeit“ von „Dr. Speech®“ und dem Untersucher (n=124).



Zusammenfassend geht aus den Berechnungen und der Interpretation der Abbildungen hervor, dass Dr. Speech® vor allem hinsichtlich der Heiserkeitsbewertungen, aber auch bei der Rauigkeitsbewertung bessere (d.h. niedergradigere) Beurteilungen abgibt als der Untersucher. Für den Parameter Behauchtheit sind die Beurteilungen von Dr. Speech® im Mittel etwas schlechter (d.h. höhergradiger) als die des Untersuchers.

3. 2. 1. 3. Korrelationen in der subjektiven und apparativen Bewertung der postoperativen Veränderung des Stimmklangs

Wie bereits in der Einleitung erwähnt, ist es im klinischen Alltag vor allem wichtig Veränderungen des Stimmklangs nach einer Operation oder logopädischer Therapie feststellen und dokumentieren zu können. Tabelle 1 (Anhang) sollte abschließend auf Unterschiede und Übereinstimmungen zwischen dem Untersucher und Dr. Speech® hinsichtlich der postoperativen Veränderung der Stimmqualität untersucht werden.

Die Abbildungen 23 - 25 stellt die vom Untersucher und Dr. Speech® postoperativ festgestellte Veränderung des Stimmklangs in der jeweiligen Stimmqualität einander gegenüber (n=62). Besser heißt, dass die Stimme postoperativ in der jeweiligen Qualität besser, also niedergradiger, wurde. Wenn z.B. die Rauigkeit praeoperativ mit 3 bewertet wurde und postoperativ mit 2 oder 1 oder 0, dann ist die Stimme besser. 'Es ist aus den Grafiken keine Aussage über die Quantität der Besserung ablesbar. Eine analoge Bedeutung hat schlechter und gleich.'

1) Rauigkeit

Aus Tabelle 18 ist zu entnehmen, dass von den 62 bewerteten postoperativen Stimmklangveränderungen 32 (51,6% der Fälle) als gleichsinnig verändert von Dr. Speech® und dem Untersucher eingestuft werden. D. h. in gut der Hälfte der Fälle herrscht Übereinstimmung, ob sich die Rauigkeit nach der Operation verschlechtert oder verbessert hat oder ob sie annähernd gleich geblieben ist. In Abbildung 23 sind die Säulen der übereinstimmenden Bewertungen mit der grünen Raute (◆) gekennzeichnet. Die mit der roten Raute (◆) markierte Säule, deutet an, dass der Untersucher die Stimme postoperativ häufiger als verbessert wahrnimmt und bewertet, als Dr. Speech®.

Aus Tabelle 18 sind weiterhin die komplett gegenteiligen Beurteilung vom Untersucher und Dr. Speech® interessant. Das sind prae- und postoperative Bewertungs-paare bei denen der Untersucher die Stimme postoperativ als verbessert bewertet, Dr. Speech® aber als verschlechtert, oder umgekehrt. Insgesamt kommt es in 7 von 62 Fällen (11,3%) zu solch gegenteiligen Bewertungen.

Die Rangkorrelation (Spearman's R) ist mit $R=0,24$ direkt, aber nur schwach ausgeprägt, und die Konkordanz ist mit einem Cohens Kappa=0,2 ebenfalls nur schwach. Beide Ergebnisse sind auch nicht statistisch signifikant (beide $p>0,05$); d.h. insgesamt sind die

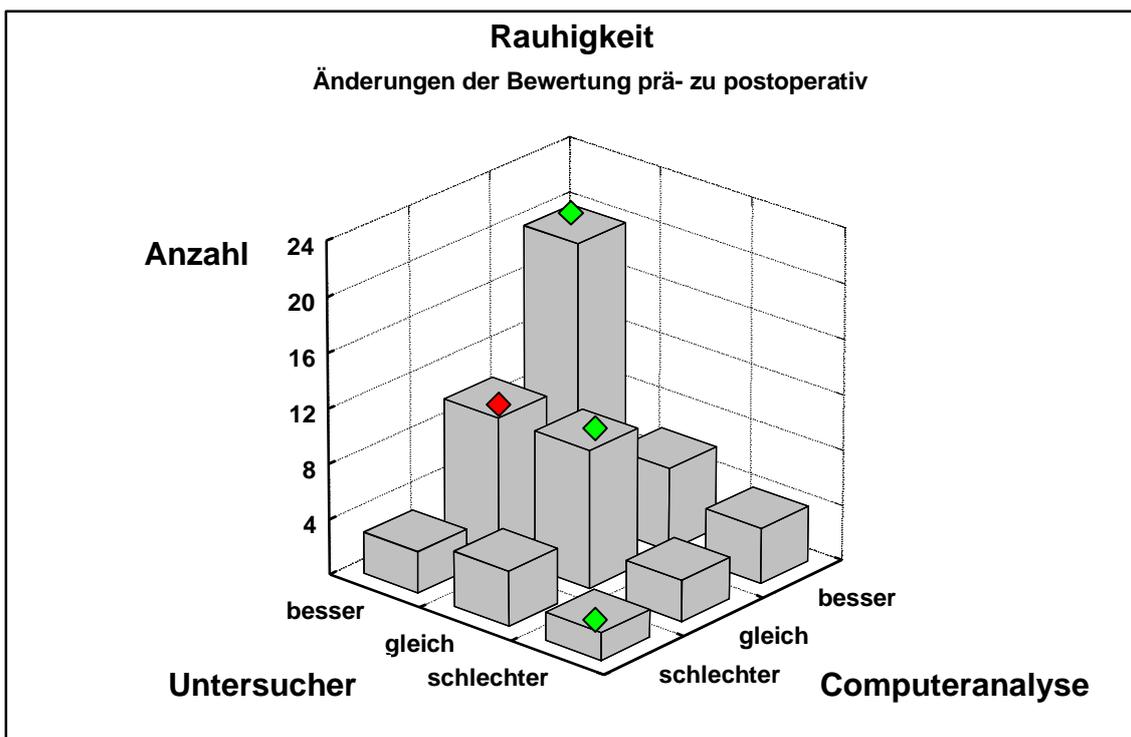
Übereinstimmungen zwischen dem Untersucher und Dr. Speech ® “bezüglich der Beurteilung der postoperativen Rauigkeitsveränderungen nur schwach ausgeprägt.

Tabelle 18: Vollständige Häufigkeitstabelle der postoperativen Änderung der Bewertungsdaten des Merkmals „Rauigkeit“ (V R S entspricht hier der postoperativen Veränderung der Rauigkeitsbewertung durch „Dr. Speech ®“; V R U entspricht analog der postoperativen Veränderung der Rauigkeitsbewertung durch den Untersucher.)

Vollständige Häufigkeitstabelle bezüglich der postop. Veränderung der Rauigkeit				
V_R_S	V_R_U schlechter	V_R_U gleich	V_R_U besser	Zeile Gesamt
schlechter	2	4	3	9
gleich	3	10	10	23
besser	4	6	20	30
Alle	9	20	33	62

Cohen's Kappa = 0,195850
 Prüfgröße Z = 2,04395
 mit p = 0,059959
 Spearmans R = 0,238001
 mit p = 0,062497

Abbildung 23: Gegenüberstellung der Änderungen der Bewertungsdaten des Merkmals „Rauigkeit“ prä- zu postoperativ von „Dr. Speech ®“ und Untersucher (n=62). Tabelle 18 gibt die vollständige Häufigkeitsverteilung der postoperativen Veränderung der Rauigkeit an, aus der sich die Abbildung herleitet. Die mit der grünen Raute (◆) gekennzeichneten Säulen liegen auf der Linie der exakten Übereinstimmung. Zur roten Raute (◆) siehe Text, oben.



2) Behauchtheit

In 49 der 62 prä- und postoperativen Stimmprobenpaaren stellen der Untersucher und Dr. Speech ® eine gleichsinnige postoperative Veränderung der Behauchtheit fest (79% der Fälle). Auch Spearmans R und Cohens Kappa zeigen einen statistisch signifikanten und deutlichen Zusammenhang zwischen den Bewertungen des Untersuchers und von Dr. Speech ® bezüglich der postoperativen Behauchtheitsveränderung (Spearmans R=0,85; Cohens Kappa=0,68 wobei beide p=0,00). Die Übereinstimmungen bei der Bewertung der postoperativen Veränderung der Behauchtheit sind also wesentlich deutlicher ausgeprägt als bei der Rauigkeit. Dies zeigt auch der Vergleich von Abbildung 23 und 24. Komplett gegensinnige Beurteilungen finden sich nicht.

Tabelle 19: Vollständige Häufigkeitstabelle der postoperativen Änderung der Bewertungsdaten des Merkmals „Behauchtheit“ (V B S entspricht hier der postoperativen Veränderung der Behauchtheitsbewertung durch „Dr. Speech ®“; V B U entspricht analog der postoperativen Veränderung der Behauchtheitsbewertung durch den Untersucher.)

Vollständige Häufigkeitstabelle bezüglich der postop. Veränderung der Behauchtheit				
V_B_S	V_B_U schlechter	V_B_U gleich	V_B_U besser	Zeile Gesamt
schlechter	14	2	0	16
gleich	4	11	3	18
besser	0	4	24	28
Alle	18	17	27	62

Cohen's Kappa = 0.676824
 Prüfgröße Z = 7.47362
 mit p = 0.000000
 Spearmans R = 0.847698
 mit p = 0.000000

3) Heiserkeit

Laut Tabelle 20 liegen 33 Fälle vor (n=62; 53,2% der Fälle) in denen der Untersucher und Dr. Speech ® die Veränderung der Heiserkeit gleich bewerten; d. h. in gut der Hälfte der Fälle haben beide das gleiche Ergebnis in der Bewertung, ob die Stimme nach der Operation heiserer oder weniger heiser geworden ist, oder ob sie gleich geblieben ist. In Spearmans R (0,45) und Cohens Kappa (0,27) drückt sich dieser eher schwache Zusammenhang aus (beide aber statistisch signifikant, denn beide $p < 0,05$). Komplette gegenteilige Bewertungen gibt es in nur 3 von 62 Fällen (4,8%).

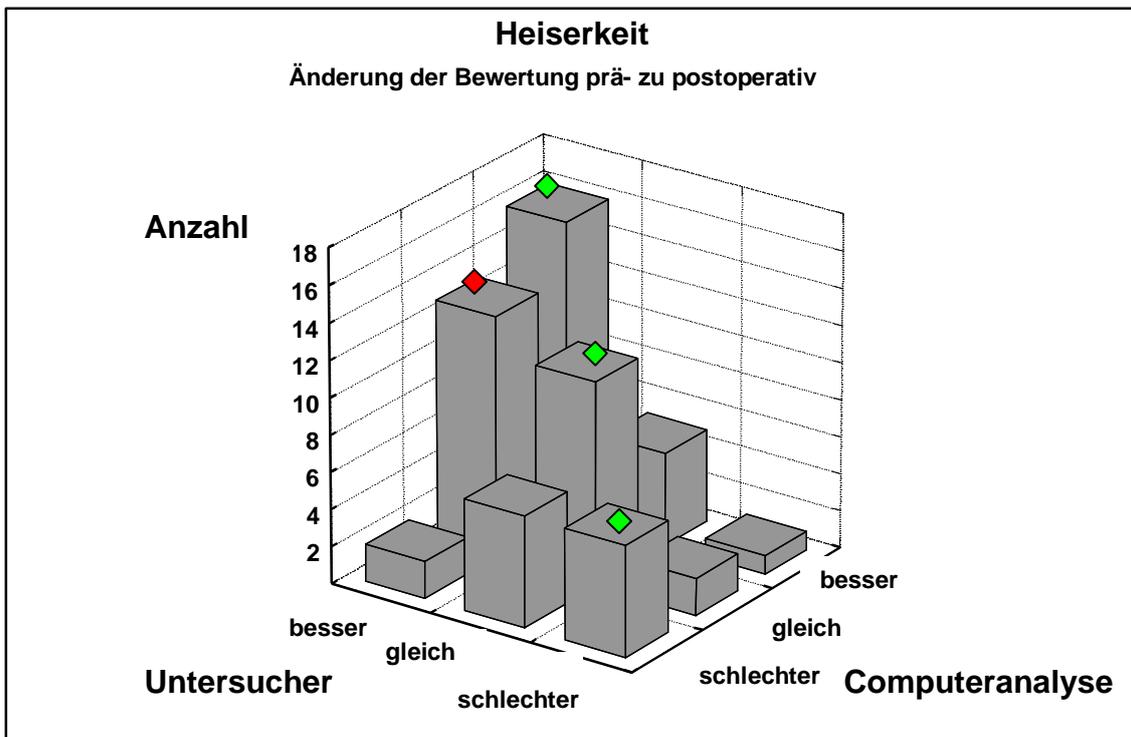
Auffällig ist die mit der roten Raute (♦) markierte Säule in Abbildung 25. In 13 Fällen (21%) die Dr. Speech ® postoperativ als unverändert (gleich) bewertet, stellt der Untersucher eine Abnahme der Heiserkeit, bzw. eine Verbesserung des Stimmklangs fest. Der Wilcoxon Test, der Unterschiede in der zentralen Tendenz feststellen kann, zeigt einzig bei der Heiserkeit, dass es Unterschiede gibt. Statistisch signifikant gibt der Untersucher postoperativ häufiger Besserungen an als Dr. Speech ® ($p = 0,039964$). Bei der Rauigkeit und der Behauchtheit kommt es nicht zu signifikanten Unterschieden.

Tabelle 20: Vollständige Häufigkeitstabelle der postoperativen Änderung der Bewertungsdaten des Merkmals „Heiserkeit“ (V H S entspricht hier der postoperativen Veränderung der Heiserkeitsbewertung durch „Dr. Speech ®“; V H U entspricht analog der postoperativen Veränderung der Heiserkeitsbewertung durch den Untersucher.)

Vollständige Häufigkeitstabelle bezüglich der postop. Veränderung der Heiserkeit				
V_H_S	V_H_U schlechter	V_H_U gleich	V_H_U besser	Zeile Gesamt
schlechter	6	6	2	14
gleich	2	11	13	26
besser	1	5	16	22
Alle	9	22	31	62

Cohen's Kappa = 0.270292
 Prüfgröße Z = 2.98513
 mit p = 0.002835
 Spearmans R = 0.467428
 mit p = 0.000128

Abbildung 25: Gegenüberstellung der Änderungen der Bewertungsdaten des Merkmals „Heiserkeit“ prä- zu postoperativ von „Dr. Speech®“ und Untersucher (n=62). Tabelle 20 gibt die vollständige Häufigkeitsverteilung der postoperativen Veränderung der Heiserkeit an, aus der sich die Abbildung herleitet. Die mit der grünen Raute (◆) gekennzeichneten Säulen liegen auf der Linie der höchsten Übereinstimmung. Zur roten Raute (◆) siehe Text, oben.



Zusammenfassend besteht für die postoperative Änderung der Behauchtheit wieder die höchste Übereinstimmung ($R=0,85$; $Kappa=0,68$, beide $p<0,05$) zwischen dem Untersucher und Dr. Speech®. Für die Heiserkeit sind die Übereinstimmungen zwar statistisch signifikant, jedoch mit $R=0,47$ und $Kappa=0,27$ deutlich geringer ausgeprägt. Bei der Rauigkeit kommt es nicht zu statistisch signifikanten Übereinstimmungen der subjektiven und computergestützten Bewertungen der postoperativen Rauigkeitsveränderungen. Auffällig ist weiterhin die Tendenz des Untersuchers postoperativ niedergradige, also verbesserte Bewertungen bezüglich der Veränderungen der Heiserkeit (statistisch signifikant) und der Rauigkeit (nicht statistisch signifikant) abzugeben, im Vergleich zu den Bewertungen von Dr. Speech®.

3. 2. 2. Statistischer Vergleich der subjektiven Behauchtheits-Bewertung und der gemessenen Werte für die NNE

Den folgenden Berechnungen, Tabellen und Abbildungen dient die Tabelle 2 des Anhangs als Datengrundlage. Dort sind die Ergebnisse der Behauchtheitsbewertung des Untersuchers (aus Tabelle 1, n=124) den, durch die Analysesoftware Dr. Speech ® gemessenen Werten der normalized noise energy (NNE) der gleichen Stimmproben gegenübergestellt.

3. 2. 2. 1. Zusammenhang zwischen subjektiver Behauchtheitsbeurteilung und objektiven NNE-Werten

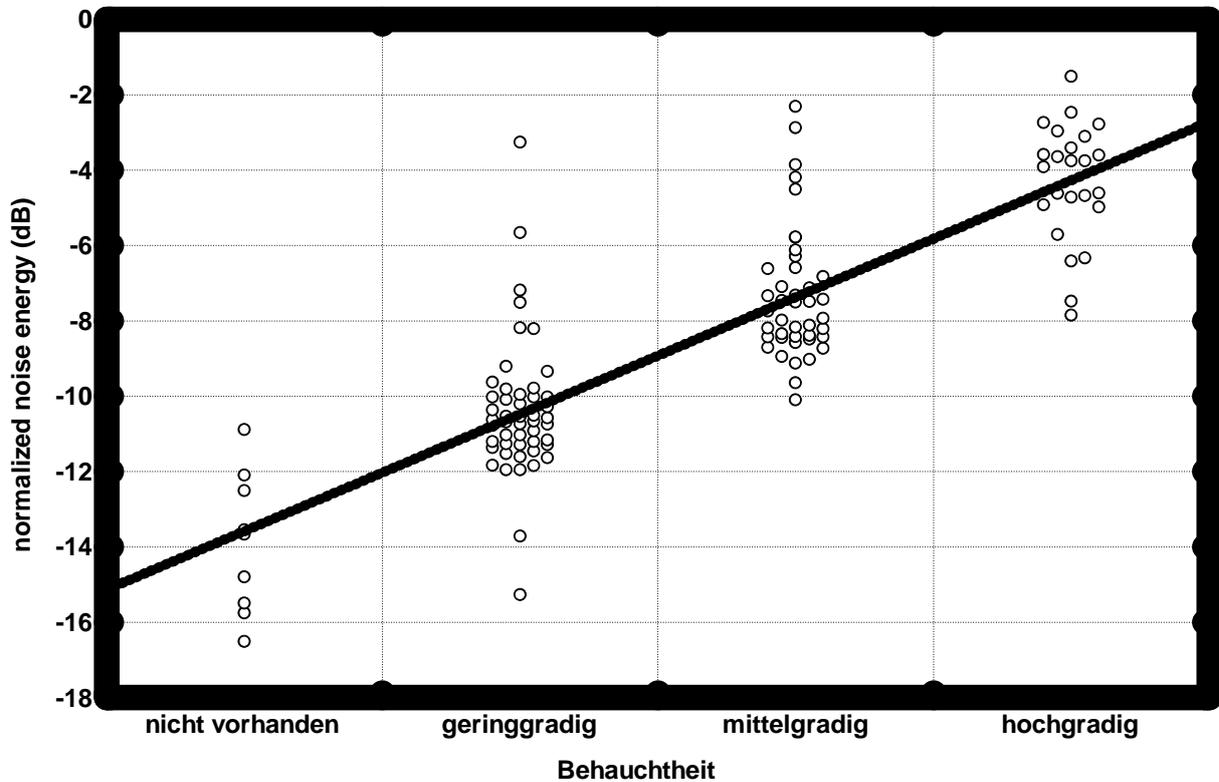
In den obigen Ergebnissen (siehe 3.2.1.) zeigten sich deutliche Übereinstimmungen der Beurteilungen der Behauchtheit durch den Untersucher und den Bewertungsmaßen der computergestützten Behauchtheits-Einstufung. Da die internen Algorithmen der Analysesoftware Dr. Speech ® bei der Behauchtheitseinstufung hauptsächlich die NNE zugrundelegen (siehe 1.2.4.), schien die statistische Untersuchung auf einen Zusammenhang zwischen subjektiv bewerteter Behauchtheit und objektiv gemessenen NNE-Werten interessant.

Die Errechnung des Spearmanschen Rangkorrelationskoeffizienten zeigt einen deutlichen und statistisch signifikanten Zusammenhang ($R=0,85$; $p<0,001$; Tabelle 21). Das bedeutet, dass Stimmproben, die als stark behaucht vom Untersucher eingestuft werden mit hohen NNE-Werten einhergehen und gering oder gar nicht behauchte Stimmen mit niedrigen NNE-Werten (Abbildung 26a).

Tabelle 21: Korrelationsquotient Spearman R für die Beurteilung des Merkmals „Behauchtheit“ durch den Untersucher und die durch „Dr. Speech ®“ gemessene NNE;

Variablenpaar	Spearman's Rangkorrelation gemessene NNE und subj. Behauchtheit			
	Gültige N	Spearman R	t(N-2)	p-Niveau
NNE & Behauchtheit	124	,847806	17,65830	0,0000

Abbildung 26a: Zusammenhang zwischen wahrgenommener Behauchtheit des Untersuchers und der durch „Dr. Speech ®“ errechneten normalized noise energy. Gepunktete Linie: approximierte „ideale“ Korrelationslinie.



3. 2. 2. 2. NNE-Trennwerte der 4 subjektiven Behauchtheitskategorien

Will man die NNE-Werte berechnen, die zwischen den 4 subjektiven Behauchtheitskategorien liegen (Trennungspunkt oder Trennwert), so ist hier das ROC-Verfahren (Receiver-Operating-Characteristics) von Metz (siehe 2.3.) anzuwenden. Jeder gemessene NNE-Wert wird als potenzieller Trennungspunkt zwischen zwei Behauchtheitskategorien (z. B. nicht vorhanden “vs. geringgradig vorhanden”) angesehen, und für jeden Wert wird die Sensitivität (Wahrscheinlichkeit, dass eine behauchte Stimme als solches erkannt wird) und die Spezifität (Wahrscheinlichkeit, dass eine nicht behauchte Stimme als solche erkannt wird) errechnet. Dann wird der Anteil der mit diesem Trennungspunkt fehlklassifizierten Fälle bestimmt. Der NNE-Wert bei dem die Spezifität und die Sensitivität am höchsten ist (Summe aus beiden) und der Anteil fehlklassifizierter Fälle am niedrigsten, ist der Trennungspunkt zwischen den beiden Kategorien.

- **NNE-Trennwert zwischen „nicht behaucht“ und „geringgradig behaucht,,**

Der NNE-Trennwert zwischen (subjektiv) nicht behauchten und (subjektiv) geringgradig behauchten Stimmen liegt bei $NNE = -12,03$ dB. Dies lässt sich aus Tabelle 22 deutlich ablesen (fett unterlegt). Das bedeutet, dass Werte oberhalb von $-12,03$ (z.B. -10 oder -5) geringgradig behaucht sind, und alle Werte unterhalb $-12,03$ (z.B. -14 oder -16) nicht behaucht sind.

Tabelle 22: Sensitivität, Spezifität zur Diskriminierung der Kategorien „nicht behaucht“ und „geringgradig behaucht“ und Anteil fehlklassifizierter Fälle; Trennwert der Kategorien bei $NNE = -12,03$ dB (fett unterlegt)

Trennpunkt NNE (dB)	Sensitivität	Spezifität	Anteil fehlklassifizierter Fälle	Sensitivität + Spezifität
-15,630	1,000	0,222	0,389	1,222
-15,385	1,000	0,333	0,333	1,333
-15,035	0,980	0,333	0,344	1,313
-14,260	0,980	0,444	0,288	1,424
-13,695	0,959	0,444	0,298	1,404
-13,610	0,959	0,556	0,243	1,515
-13,030	0,959	0,667	0,187	1,626
-12,305	0,959	0,778	0,132	1,737
-12,030	0,959	0,889	0,076	1,848
-11,905	0,918	0,889	0,096	1,807
-11,845	0,898	0,889	0,107	1,787
-11,740	0,878	0,889	0,117	1,766
-11,625	0,857	0,889	0,127	1,746
-11,570	0,837	0,889	0,137	1,726
-11,495	0,816	0,889	0,147	1,705
-11,415	0,796	0,889	0,158	1,685
-11,330	0,776	0,889	0,168	1,664
-11,280	0,735	0,889	0,188	1,624
-11,240	0,714	0,889	0,198	1,603
-11,190	0,673	0,889	0,219	1,562
-11,100	0,653	0,889	0,229	1,542
-10,975	0,612	0,889	0,249	1,501
-10,905	0,592	0,889	0,260	1,481
-10,835	0,592	1,000	0,204	1,592
-10,765	0,571	1,000	0,214	1,571
-10,745	0,551	1,000	0,224	1,551
-10,715	0,531	1,000	0,235	1,531
-10,675	0,510	1,000	0,245	1,510
-10,645	0,490	1,000	0,255	1,490
-10,605	0,469	1,000	0,265	1,469
-10,565	0,449	1,000	0,276	1,449
-10,545	0,429	1,000	0,286	1,429
-10,525	0,408	1,000	0,296	1,408

- **NNE-Trennwert zwischen „geringgradig behaucht“ und „mittelgradig behaucht,,**

Der NNE-Trennwert zwischen (subjektiv) geringgradig behauchten und (subjektiv) mittelgradig behauchten Stimmen liegt bei $NNE = -9,17$ dB. Dies lässt sich aus Tabelle 23 deutlich ablesen (fett unterlegt). Das bedeutet, dass Werte oberhalb von $-9,17$ (z.B. -8 - oder -5) mittelgradig behaucht sind, und Werte unterhalb $-9,17$ (z.B. -10 oder -11) geringgradig behaucht sind.

Tabelle 23: Sensitivität, Spezifität zur Diskriminierung der Kategorien „geringgradig behaucht“ und „mittelgradig behaucht“ und Anteil fehlklassifizierter Fälle; Trennwert der Kategorien bei $NNE = -9,17$ dB (fett unterlegt)

Trennpunkt NNE (dB)	Sensitivität	Spezifität	Anteil fehlklassifizierter Fälle	Sensitivität + Spezifität
-10,675	1,000	0,490	0,255	1,490
-10,645	1,000	0,510	0,245	1,510
-10,605	1,000	0,531	0,235	1,531
-10,565	1,000	0,551	0,224	1,551
-10,545	1,000	0,571	0,214	1,571
-10,525	1,000	0,592	0,204	1,592
-10,440	1,000	0,612	0,194	1,612
-10,330	1,000	0,633	0,184	1,633
-10,250	1,000	0,653	0,173	1,653
-10,155	1,000	0,673	0,163	1,673
-10,095	0,976	0,673	0,175	1,650
-10,060	0,976	0,694	0,165	1,670
-9,995	0,976	0,755	0,134	1,731
-9,890	0,976	0,776	0,124	1,752
-9,805	0,976	0,796	0,114	1,772
-9,720	0,976	0,816	0,104	1,793
-9,640	0,952	0,816	0,116	1,769
-9,490	0,952	0,837	0,105	1,789
-9,280	0,952	0,857	0,095	1,810
-9,170	0,952	0,878	0,085	1,830
-9,080	0,929	0,878	0,097	1,806
-8,990	0,905	0,878	0,109	1,782
-8,840	0,881	0,878	0,121	1,759
-8,720	0,857	0,878	0,133	1,735
-8,645	0,833	0,878	0,145	1,711
-8,535	0,810	0,878	0,156	1,687
-8,470	0,786	0,878	0,168	1,663
-8,440	0,762	0,878	0,180	1,639
-8,425	0,714	0,878	0,204	1,592
-8,405	0,690	0,878	0,216	1,568
-8,370	0,667	0,878	0,228	1,544
-8,280	0,643	0,878	0,240	1,520
-8,205	0,619	0,898	0,241	1,517

- **NNE-Trennwert zwischen „mittelgradig behaucht“ und „hochgradig behaucht,,**

Der NNE-Trennwert zwischen (subjektiv) mittelgradig behauchten und (subjektiv) hochgradig behauchten Stimmen liegt bei $NNE = -5,75$ dB. Dies lässt sich aus Tabelle 24 deutlich ablesen (fett unterlegt). Das bedeutet, dass Werte oberhalb von $-5,75$ (z.B. -3 oder -1) hochgradig behaucht sind, und Werte unterhalb $-5,75$ (z.B. -7 oder -8) mittelgradig behaucht sind.

Tabelle 24: Sensitivität, Spezifität zur Diskriminierung der Kategorien „mittelgradig behaucht“ und „hochgradig behaucht“ und Anteil fehlklassifizierter Fälle; Trennwert der Kategorien bei $NNE = -5,75$ dB (fett unterlegt)

Trennpunkt NNE (dB)	Sensitivität	Spezifität	Anteil fehlklassifizierter Fälle	Sensitivität + Spezifität
-8,370	1,000	0,333	0,333	1,333
-8,280	1,000	0,357	0,321	1,357
-8,205	1,000	0,381	0,310	1,381
-8,185	1,000	0,405	0,298	1,405
-8,145	1,000	0,429	0,286	1,429
-8,055	1,000	0,452	0,274	1,452
-7,965	1,000	0,476	0,262	1,476
-7,895	1,000	0,500	0,250	1,500
-7,795	0,958	0,500	0,271	1,458
-7,625	0,958	0,524	0,259	1,482
-7,500	0,958	0,548	0,247	1,506
-7,485	0,958	0,571	0,235	1,530
-7,475	0,917	0,571	0,256	1,488
-7,450	0,917	0,595	0,244	1,512
-7,385	0,917	0,619	0,232	1,536
-7,330	0,917	0,643	0,220	1,560
-7,225	0,917	0,667	0,208	1,583
-7,115	0,917	0,690	0,196	1,607
-6,965	0,917	0,714	0,185	1,631
-6,725	0,917	0,738	0,173	1,655
-6,605	0,917	0,762	0,161	1,679
-6,505	0,917	0,786	0,149	1,702
-6,375	0,875	0,786	0,170	1,661
-6,310	0,833	0,786	0,190	1,619
-6,205	0,833	0,810	0,179	1,643
-5,955	0,833	0,833	0,167	1,667
-5,785	0,833	0,857	0,155	1,690
-5,745	0,833	0,881	0,143	1,714
-5,345	0,792	0,881	0,164	1,673
-4,950	0,750	0,881	0,185	1,631
-4,820	0,708	0,881	0,205	1,589
-4,700	0,667	0,881	0,226	1,548
-4,650	0,625	0,881	0,247	1,506
-4,615	0,583	0,881	0,268	1,464
-4,560	0,542	0,881	0,289	1,423
-4,350	0,542	0,905	0,277	1,446
-4,055	0,542	0,929	0,265	1,470
-3,890	0,500	0,929	0,286	1,429

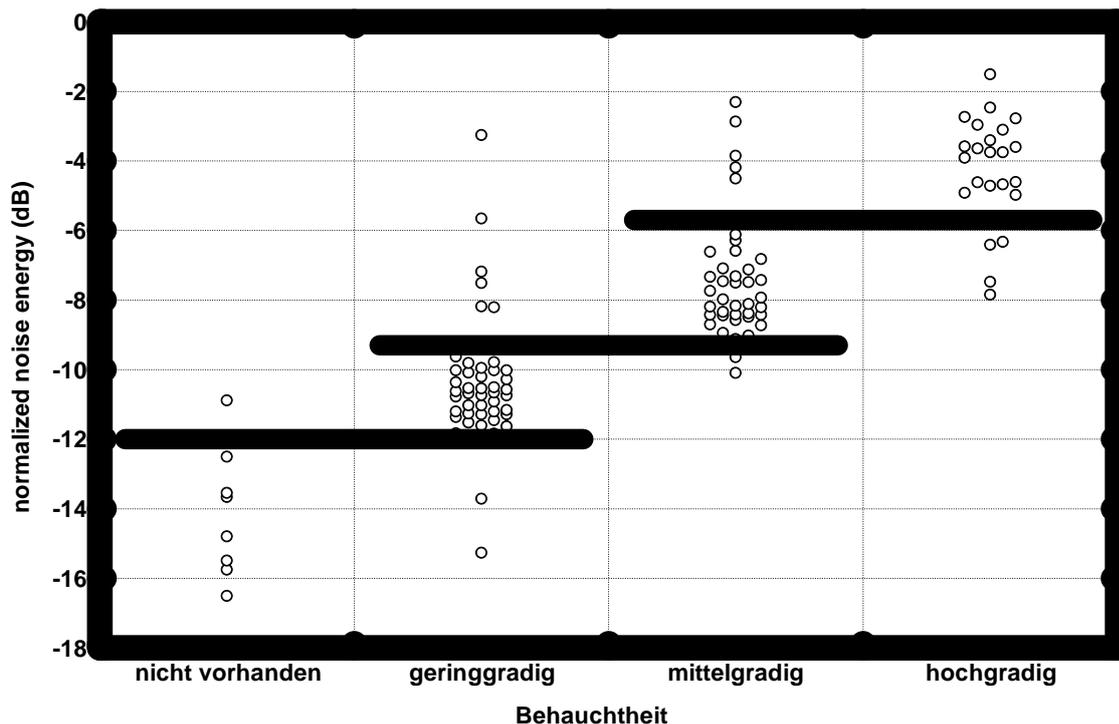
Trennpunkt NNE (dB)	Sensitivität	Spezifität	Anteil fehlassifizierter Fälle	Sensitivität + Spezifität
-3,810	0,500	0,952	0,274	1,452
-3,705	0,417	0,952	0,315	1,369
-3,630	0,375	0,952	0,336	1,327
-3,600	0,333	0,952	0,357	1,286
-3,500	0,292	0,952	0,378	1,244

Zusammenfassend ist festzustellen, dass die NNE in hohem Maße mit der subjektiven Behauchtheitswahrnehmung korreliert. In Tabelle 25 sind die NNE-Trennwerte zwischen den einzelnen Behauchtheitskategorien gemeinsam aufgeführt, und Abbildung 26b stellt die Kategorien graphisch dar. Zur Sicherung der gefundenen Trennpunkte müssten die Kategorien mit einem neuen Patientenkollektiv überprüft werden. Eine Überprüfung der neu gefundenen Kategorien am gleichen Patientenkollektiv verbietet sich aus statistischen Gründen.

Tabelle 24: Trennungspunkte der subjektiven Behauchtheitskategorien

Grenze zwischen...	Trennungspunkt bei NNE (dB)
keiner und geringer Behauchtheit	-12,0
geringer und mittlerer Behauchtheit	-9,2
mittlerer und hoher Behauchtheit	-5,8

Abbildung 26b: Die Trennungspunkte sind als schwarze Linien in Abbildung 26a eingezeichnet



4. Diskussion

In der Einleitung dieser Arbeit wurden 3 Voraussetzungen formuliert, die ein routinemäßig zur Stimmklangbeurteilung eingesetztes Verfahren erfüllen sollte:

- hohe Zuverlässigkeit der Stimmklangbeurteilung und vor allem der postoperativen Veränderung
- geringer Zeitaufwand
- gute, krankenblattfähige Dokumentierbarkeit und Vergleichbarkeit der Ergebnisse untereinander

Hieran muss sich das in dieser Arbeit verwendete Kombinationsverfahren der computergestützten RBH-Beurteilung (Dr. Speech ®) messen lassen. Wie anfangs erwähnt, hat sich das RBH-Schema zur auditiven Stimmklangbeurteilung bewährt, da es einen guten Kompromiss zwischen hoher Zuverlässigkeit und Reproduzierbarkeit einerseits und geringem zeitlichen Aufwand andererseits bildet. Zusätzlich ist das Ergebnis der Beurteilung leicht dokumentierbar. Deshalb wurde in dieser Arbeit die auditive RBH-Beurteilung als „Goldstandard“ für den Vergleich mit der computergestützten RBH-Beurteilung zugrundegelegt. In der beschriebenen Studie wird die aus gehaltenen Vokalen und spontaner Sprache gewonnene auditive RBH-Beurteilung mit der aus einem einzigen gehaltenen Vokal errechneten computergestützten RBH-Beurteilung verglichen.

Bei der elektroakustischen Stimmanalyse spielt das unter gleichbleibenden Aufnahmebedingungen zu erstellende stimmliche Prüfmaterial eine entscheidende Rolle (Titze et al., 1993), und zwar nicht nur für das Ergebnis der Analyse, sondern auch für den zeitlichen Aufwand der Methode. Prinzipiell stehen 2 verschiedene Stimmmaterialien zur Verfügung: gehaltene Vokale oder fortlaufende Sprache, wobei hier noch zwischen gelesener und spontaner Sprache unterschieden werden muss. Für die Untersuchung fortlaufender Sprache sind eher spektralanalytische oder sonographische Methoden geeignet. Vom zeitlichen Aspekt ist es jedoch einleuchtend, dass die Aufzeichnung von Standardtexten und -Sätzen oder spontaner Sprache ungleich aufwendiger ist, als die Aufnahme von gehaltenen Vokalen. Die hier verwendete computergestützte RBH-Beurteilung aus einem einzigen gehaltenen Vokal ist die zeitlich günstigste Methode. Für die in der vorliegenden Arbeit jeweils erstellten 3-5 Vokalaufnahmen von [ae] wurde eine Untersuchungszeit zwischen 3-5

min benötigt. Dies erscheint im Klinikalltag durchaus praktikabel, zumal man unter Alltagsbedingungen parallel zur Aufnahmeerstellung auch die auditive Bewertung der Spontansprache durchführen könnte.

Die übersichtliche grafische Darstellung der computergestützten RBH-Beurteilung von Dr. Speech® und die Möglichkeit, Befunde direkt auszudrucken, sorgen für eine leichte und auch krankenblattfähige oder digitale Dokumentation.

Somit erfüllt die computergestützte RBH-Beurteilung durch Dr. Speech® die Forderungen nach geringem zeitlichen Aufwand und guter Dokumentierbarkeit.

Die Übereinstimmungen in den Ergebnissen der computergestützten und der auditiven RBH-Beurteilung und somit die Zuverlässigkeit schwankt in Abhängigkeit vom bewerteten Merkmal aber beträchtlich.

Die elektroakustische Beurteilung der **Behauchung** durch Dr. Speech® stimmt statistisch signifikant mit der auditiven Beurteilung im Sinne eines straffen Zusammenhangs überein. Deshalb sind auch Unterschiede in den Behauchtheitsbewertungen von Untersucher und Dr. Speech® selten (Kapitel 3.2.1.1. und 3.2.1.2.). Dr. Speech® und der Untersucher stimmen also in hohem Maße exakt überein (84,7% identische Beurteilungen). Auch in den Einzelfällen 1 und 2 liegen nur exakte Übereinstimmungen bezüglich der prae- und postoperativen Behauchtheitsbeurteilungen des Untersuchers und von Dr. Speech® vor.

Wie in der Einleitung erwähnt, legen die programminternen Algorithmen der hier untersuchten Stimmanalyse Software bei der Bewertung der Behauchtheit vornehmlich die NNE zugrunde. Die Statistik in Kapitel 3.2.2. zeigt einen deutlichen, statistisch signifikanten und direkten Zusammenhang zwischen den NNE-Werten und der wahrgenommenen Behauchtheit. Die Korrelation von Behauchung und NNE wurde durch Huang 1998 bestätigt. Kasuya et al. ermitteln 1986 als Schwellenwert zur Trennung gesunder von kranken Stimmen eine NNE von 14 dB (Richtwert). D. h. Werte unter -11 (-12, -20, etc.) sind eher bei stimmgesunden und Werte darüber (-10, -2, etc.) eher bei stimmgestörten Patienten zu finden. Dieser Wert wird durch unsere Untersuchungen annähernd bestätigt (Trennwert: -12 dB; siehe Kapitel 3.2.2.). Der Unterschied zu Kasuyas Trennwert könnte durch die zu geringe Anzahl an nicht behauchten Stimmen in unserem Patientengut bedingt sein. Mit beiden Trennungspunkten (Kasuya: -11; vorliegende Arbeit: -12) übereinstimmend, hat in Fall 1 die praeoperativ (auditiv) nicht behauchte Stimme eine NNE von 15,53 dB und die postoperativ

(auditiv) maximal behauchte Stimme ein NNE von 2,47. In Fall 2 wird die Behauchtheit auditiv und computergestützt prae- und postoperativ als geringgradig (B=1) bewertet. Auditiv und elektroakustisch wurde im RBH-Schema prae- zu postoperativ also keine Veränderung der Behauchtheit konstatiert. Die NNE ist postoperativ aber geringfügig verkleinert (NNE (praeop):-10,59; NNE(postop): -11,61). Dies könnte ein Hinweis auf eine leichte postoperative Stimmklangverbesserung sein die sich im RBH-Schema aufgrund der 4-stufigen Skala nicht ausdrücken lässt. Wie auch schon vorher (siehe Einzelfallbesprechungen) erwähnt, kann ein Blick auf akustische R_{ph}-Parameter wie NNE oft zusätzliche Information geben und Tendenzen aufzeigen. Hier zeigt sich der Vorteil, bei der Heiserkeitsbewertung akustische Parameter mitzuberrücksichtigen, da geringfügige Änderungen, die auditiv nicht wahrgenommen, oder nicht präzise dokumentiert werden können, sichtbar werden und Tendenzen zu Besserung oder Verschlechterung noch dokumentiert werden. Dies könnte vor allem bei der Verlaufskontrolle bei konservativen (logopädischen) Therapien interessant sein, da hier die Stimmverbesserung nur langsam erfolgt und geringfügige Veränderungen akustischer Parameter dann den Therapieerfolg (oder Misserfolg) anzeigen.

Der in dieser Arbeit ermittelte Zusammenhang von NNE und wahrgenommener Behauchung sollte aber insofern sehr kritisch betrachtet werden, da er keinesfalls bedeutet, dass die NNE ein guter Parameter zur Diskrimination von Behauchtheit gegenüber Heiserkeit und Rauigkeit ist. Studien von Zwirner und Michaelis (1997) und auch von Hirano et al. (1988) zeigten, dass die NNE signifikant mit der perceptiven Rauigkeit (Zwirner) und zusätzlich auch mit der Wahrnehmung der Heiserkeit (Hirano) korreliert. Auch Kasuya (1986 und 1993) fand Zusammenhänge zwischen wahrgenommener Heiserkeit und der NNE.

Insgesamt ist festzustellen, dass es in ausreichend hohem Maße zu übereinstimmenden Bewertungen der Behauchtheit zwischen Dr. Speech® und dem Untersucher kommt. Eine zusätzliche Betrachtung der NNE kann aber hilfreich sein.

Die elektroakustische Beurteilung der **Rauigkeit** zeigt mehr tendenzielle als identische Übereinstimmungen mit der auditiven Bewertung (Kapitel 3.2.1.1. und 3.2.1.2.). Nur in 25,8% aller Fälle sind die Beurteilungen übereinstimmend zwischen Dr. Speech® und dem Untersucher. Dies zeigt sich auch in den Einzelfällen. In Fall 1 liegen deutliche Unterschiede in den Bewertungen (Erklärung siehe dort) vor und in Fall 2 exakte Übereinstimmungen.

Dr.Speech® beurteilt im Mittel die gleichen Stimmen prae- und postoperativ eher als geringgradiger rau, als es auditiv erscheint. Dies fällt vor allem bei niedergradig rauhen Stimmen auf (Abb. 20). Geringgradige Störungen der Rauigkeit werden von Dr. Speech® “

also nicht als solche erkannt und oft als nicht rauh (normal) eingestuft. Dr. Speech[®] legt bei der Rauigkeitseinstufung hauptsächlich den Jitter zugrunde. Hier mag die Erklärung der tendenziell besseren (niedergradigen) Rauigkeitsbeurteilungen durch Dr. Speech[®] im der Analyse zugrundeliegenden Stimmmaterial (Vokal [ae]) liegen. Nach Hammarberg (1994) haben die traditionellen, an gehaltenen Vokalen durchgeführten Jitter- und Shimmernessungen weniger Erfolg bei der Rauigkeitsbewertung gebracht als erwartet. Wie bereits in der Einleitung erwähnt, können bei der Analyse von gehaltenen Vokalen wichtige Klangphänomene verloren gehen (Hammarberg, 1980). Arbeiten von de Krom (1994) lassen den Schluss zu, dass sich gerade die Rauigkeit besser in Phänomenen der fortlaufenden Sprache, wie z. B. bei vokalen Stimmeinsätzen offenbart als in einem Vokalausschnitt. Dies könnte mit ein Grund dafür sein, dass bei der computergestützten RBH-Beurteilung vor allem leichte Rauigkeiten durch Dr. Speech[®] "tendenziell zu geringgradig bzw. als normal eingestuft werden.

Insgesamt reichen die Übereinstimmungen zwischen der auditiven und computergestützten Rauigkeitsbewertung nicht aus, um die computergestützte Bewertung der Rauigkeit zur routinemäßigen Dokumentation einzusetzen. Sinnvoller ist die Dokumentation und Interpretation der Periodizitätsmaße (Jitter und Shimmer) im Einzelfall.

Die elektroakustische Bewertung der **Heiserkeit** weist ebenfalls mehr tendenzielle Übereinstimmungen auf. Für tatsächlich übereinstimmende Beurteilungen ergab sich keine statistische Signifikanz (identische Bewertungen in 35,5% der Fälle). Die Heiserkeit wird von Dr. Speech[®] "tendenziell geringgradiger (d.h. niedrigere Werte) bewertet. Oder anders ausgedrückt: Dr. Speech[®] beurteilt die gleiche Stimme tendenziell weniger heiser als der Untersucher.

Die Wahrnehmung von Heiserkeit ist als eine Kombination von rauher und behauchter Stimmqualität zu betrachten (Huang, 1994; Klingholz, 1986). Insofern ist die Forderung von Friedrich (1996) sinnvoll, dass im RBH-Schema der Index von H, mindestens so groß sein sollte wie jener von R, oder B, "Somit steht H, in Beziehung zu R, und B, "Zahlreiche Befunde der computergestützten Heiserkeitsbeurteilung durch Dr. Speech[®] sind danach schwer nachzuvollziehen (z. B. Patient 20 R=0; B=3; H=1 oder Fall 1 und 2 präoperativ). Allein diese, dem auditiven RBH-Schema auferlegte Forderung erklärt das statistische Ergebnis, dass Dr. Speech[®] die Heiserkeit tendenziell geringgradiger einstuft als der

Untersucher, da für diesen der Begriff Heiserkeit als Oberbegriff aufzufassen ist und mit der Rauigkeit und der Behauchtheit in Zusammenhang steht. Vielleicht liegt die Ursache der geringgradigeren Heiserkeitseinstufung durch Dr. Speech ® über auch in einer hohen Gewichtung des Shimmers, der sich ähnlich wie der Jitter verhält (siehe Rauigkeit), in den softwareinternen Algorithmen und der zu geringen Berücksichtigung von Behauchtheitsmaßen (wie NNE) in den Algorithmen.

Wie bereits erwähnt, ist es für die tägliche Routine oft entscheidender, die Änderungen des Stimmklangs nach Therapie schnell zu erfassen, als ein exaktes Abbild des Stimmklangs zu erhalten. Für Arzt und Patient ist es vorrangig interessant, ob sich eine Stimme postoperativ verbessert oder verschlechtert hat, oder gleichgeblieben ist. Wichtiger als die exakten Übereinstimmungen zwischen Dr. Speech ® und dem Untersucher ist im alltäglichen Umgang also die Frage, ob beide Beurteilungssysteme (auditiv vs. computergestützt) die postoperative Veränderung des Merkmals gleichsinnig beurteilen (besser, schlechter, gleich). Die exakten (auditiven und computergestützten) Übereinstimmungen in den Beurteilungen mit dem jeweils verwendeten RBH-Schema ist von untergeordneter Bedeutung. Die Tendenz der Änderung ist klinisch entscheidend. Deshalb wurden die Ergebnisse der Bewertung der postoperativen Änderung eines Stimmerkmals untersucht und miteinander verglichen (Kapitel 3.2.1.3.).

Die festgestellten postoperativen Veränderungen durch Dr. Speech ® stimmen bezüglich der Beurteilung der Behauchtheit mit dem Untersucher am besten überein. Offensichtlich korreliert die aus spontaner Sprache und gehaltenen Vokalen wahrgenommene Behauchung gut mit der elektroakustischen Behauchtheitsbeurteilung, basierend auf dem Vokal [ae]. In knapp 80% der beurteilten Fälle sind sich der Untersucher und Dr. Speech ® einig, ob sich eine Stimme bezüglich der Behauchtheit verschlechtert oder verbessert hat, oder ob sie gleichgeblieben ist (Korrelation nach Spearman $R=0,85$). In keinem einzigen Fall kommt es zu komplett gegenteiligen Beurteilungen bezüglich stimmlicher Verbesserung oder Verschlechterung (bezogen auf die Behauchtheit). Die hohe Zahl der Übereinstimmungen und ein Fehlen von gegenteiligen Beurteilungen bezüglich der postoperativen Änderung der Behauchtheit würde für einen routinemäßigen Einsatz der computergestützten RBH-Beurteilung sprechen. Jedoch sind die übereinstimmenden Ergebnisse bei der Beurteilung der postoperativen Veränderungen bezüglich Rauigkeit und Heiserkeit bedeutend weniger. Die computergestützte Rauigkeitsbeurteilung durch Dr. Speech ® hat keine statistisch

signifikanten Übereinstimmungen mit dem Untersucher bezüglich der postoperativen Veränderung der Rauigkeit (52% gleichsinnige Veränderungen zwischen Untersucher und Dr. Speech ®):“ Die Frage ob eine Stimme postoperativ weniger rau, rauher oder gleichgeblieben ist, kann die computergestützte Rauigkeitsbewertung durch Dr. Speech ® nicht zuverlässig beantworten. Bei der Beurteilung der postoperativen Rauigkeitsveränderung kommt es sogar in 11% der Fälle zu entgegengesetzten Beurteilungen, das sind Fälle in denen z.B. die computergestützte Rauigkeitsbewertung eine postoperative Besserung einer Stimme angibt, während sich die Stimme auditiv bezüglich der Rauigkeit verschlechtert hat oder umgekehrt.

Bei der Beurteilung der postoperativen Veränderung der Heiserkeit sind zwar signifikante Übereinstimmungen zwischen Dr. Speech ® und dem Untersucher festzustellen (53% gleichsinnige Veränderungen), jedoch sind die Zusammenhänge schwach ausgeprägt (Spearman's R: 0,45; Cohens Kappa: 0,27).

Diese Ergebnisse machen deutlich, dass man bei routinemäßigem Einsatz der computergestützten RBH-Beurteilung durch Dr. Speech ® in Klinik oder Praxis nicht auf zusätzliche auditive Beurteilung zur Absicherung und Ergänzung der Computerbefunde verzichten sollte. Die computergestützte RBH-Beurteilung kann die auditive Beurteilung nicht ersetzen, da sich vor allem bei der Beurteilung der postoperativen Stimmveränderung, mit Ausnahme der Behauchtheitsbewertungen, zu wenig Übereinstimmungen ergeben. Zwar liefert die computergestützte Behauchtheitsbewertung zuverlässige Ergebnisse, jedoch steht auch mit der NNE ein wertvoller akustischer Parameter zur Verfügung, der ebenfalls sehr deutlich mit der wahrgenommenen Behauchtheit korreliert. Die computergestützte Behauchtheitsbewertung liefert nicht mehr Information, als die NNE.

Weiterhin sollte auch nicht außer Acht gelassen werden, dass in über 10% der Aufnahmen (siehe Kapitel 2.2.) eine computergestützte RBH-Beurteilung aufgrund von zu hohen Periodizitätsschwankungen gar nicht möglich war. Es handelte sich hier ausnahmslos um hochgradig heisere bis aphone Stimmen. Die Unmöglichkeit der Bewertung solcher Stimmen durch die computergestützte RBH-Beurteilung von Dr. Speech ®, ist jedoch kein für sie spezifisches Problem, sondern ein generelles und grundlegendes Problem bei elektroakustischen Analysen. Für die Berechnung der meisten akustischen Parameter (hier: APQ, PPQ und NNE) ist nämlich eine gewisse Periodizität Voraussetzung. Meist scheitern die Rechensysteme bei sehr heiseren oder aphonischen Stimmklängen an der Unmöglichkeit eine

Periodenlänge zu bestimmen, die sie mit den benachbarten Perioden ins Verhältnis setzen können um weitere Berechnungen durchzuführen (Boyanov et al. 1997; Hedelin et al 1990).

Die Analyse der Einzelfälle zeigt, dass die postoperative Veränderung der Werte verschiedener akustischer Parameter zur Untermauerung und Dokumentation von postoperativen Stimmveränderungen sehr aussagekräftig ist.

Die NNE bewährte sich in Übereinstimmung mit Ergebnissen von Hirano (1988), Kasuya (1986, 1993) und Huang (1998) auch in unserer Arbeit als hervorragenden Parameter um postoperative Stimmveränderungen (allgemeine Stimmqualität, insbesondere Behauchtheit) festzustellen und zu dokumentieren. Vor allem lassen sich auch geringfügige therapeutische Stimmveränderungen, zumindest im Einzelfall (Fall 2), durch die NNE dokumentieren. Für den klinischen Alltag scheint dieser Parameter besonders geeignet, da die Auswertung eines Vokals zuverlässige Ergebnisse liefert und die Erstellung zeitlich sehr schnell ist. Arbeiten von Michaelis et al. (1997) gaben ebenfalls Hinweise, dass zur zuverlässigen Berechnung von Rauschkomponenten (hier allerdings GNE, Glottal to noise excitation) reduzierte Stimmaufnahmen ausreichen. Er konnte mit der Aufnahme von 3 Vokalen ähnlich zuverlässige Ergebnisse erzielen wie vorher aus 18 Vokalen.

Auch APQ und PPQ erwiesen sich in unseren Einzelfallanalysen als gute Parameter, stimmliche Therapieeffekte sichtbar zu machen, wenn sie kritisch hinterfragt werden. Damit werden die Ergebnisse Hiranos (1988) und Zwirners (1996), die APQ und PPQ als valide Parameter zur Stimmklangdokumentation ermittelten, bestätigt.

Zusammenfassend lässt sich feststellen, dass die Auswertung von APQ, PPQ und NNE besser geeignet ist, den Stimmklang und vor allem seine postoperative Veränderung zu dokumentieren als die elektroakustische RBH-Einstufung durch Dr. Speech®. Jedoch zeigt die Analyse der Einzelfälle, dass auch akustische Parameter immer kritisch betrachtet werden müssen. Auch diese Parameter lassen sich mit dem gleichen geringen Zeitaufwand wie die computergestützte RBH-Beurteilung durch Dr. Speech® erstellen und sind ebenso gut dokumentierbar. Eine elektroakustische RBH-Beurteilung, wie sie die Software Dr. Speech® ermöglicht, ist deshalb nicht sinnvoll und - mit Ausnahme der Behauchtheitsbewertung - auch unzuverlässig. In der täglichen Routine sollte sich der Blick auf die oben erwähnten Parameter richten und nicht auf computerassistierte Einschätzungen von psychoakustischen Stimmqualitäten wie Rauigkeit, Behauchtheit und Heiserkeit. Einen guten Weg beschreitet hier letztlich das Heiserkeitsdiagramm von Michaelis und Zwirner (siehe Einleitung). Es werden lediglich die Intensitäten eines Behauchtheitsmaßes (GNE) und einer

Irregularitätskomponente (aus Jitter und Shimmer) angezeigt. Einstufungen in verschiedene Stimmqualitäten werden vermieden. Trotz der zahlreichen akustischen Untersuchungen am Stimmklang ist eine genaue und zuverlässige Zuordnung akustischer Parameter zu einzelnen Stimmklangqualitäten schlussendlich problematisch, da die Begrifflichkeiten schwierig zu definieren sind.

Akustische Parameter wie APQ, PPQ und NNE, die aus gleichbleibend guten Stimmaufnahmen ermittelt werden müssen, sollten zusätzlich zur auditiven RBH-Beurteilung interpretiert und mit ihr verglichen werden. Dann können sie die auditive Beurteilung der Stimmqualität in ihren einzelnen psychoakustischen Merkmalen, die in der tägliche Routine unverzichtbar bleibt, unterstützen und erhärten. Die vom Zeitaufwand her günstige Ermittlung der Parameter aus einem gehaltenen Vokal erleichtert die routinemäßige Einsetzbarkeit und liefert trotzdem zuverlässige Ergebnisse. Jedoch warnen selbst Baken et al. 1996 vor einem übermäßigen und kritiklosen Gebrauch von computerassistierter Stimmanalyse. Außerdem sollte man immer vor Augen haben, dass zum Zeitpunkt der Untersuchung, sei es bei der Aufzeichnung der Stimme oder bei der auditiven Beurteilung, immer nur eine Momentaufnahme der Patientenstimme zugänglich wird. Deshalb sollte die Zufriedenheit des Patienten mit seiner Stimme immer eruiert werden und als ebenso entscheidend für die Bewertung des Therapieerfolgs angesehen werden, wie die gemessene oder auditiv erfasste Veränderung des Stimmklangs.

Neben der Berücksichtigung elektroakustischer Befunde, steht immer noch das Ohr als wichtigstes und sensibelstes Untersuchungsorgan im Vordergrund.

5. Anhang

Tabelle 1: auditive und computergestützte RBH-Beurteilung des Stimmklangs von 62 Patienten, prae- und postoperativ.

Die Patienten 1-62 haben jeweils zwei Zeilen. Die Zeile „Dr. speech®“ zeigt unter „praeR“, „praeB“ und „praeH“ die praeoperative elektroakustische Beurteilung der Stimmqualitäten Rauhigkeit, Behauchtheit und Heiserkeit durch die Stimmanalyse Software „Dr. Speech®“. Unter „postR“, „postB“ und „postH“ finden sich die postoperativen Beurteilungen. Analog finden sich in der Zeile „Untersucher“ die auditiven RBH-Beurteilungen des Untersuchers. (0=normal; 1=geringgradig oder slight; 2=mittelgradig oder moderate; 3=hochgradig oder extreme)

Patient	Bewerter	Diagnose/Therapie	prae R	prae B	prae H	post R	post B	post H
1	Dr.Speech	CA/Chordektomie	2	1	2	3	3	3
1	Untersucher	CA/Chordektomie	3	1	3	3	3	3
2	Dr.Speech	CA/Chordektomie	3	3	3	1	2	2
2	Untersucher	CA/Chordektomie	2	3	3	3	2	3
3	Dr.Speech	CA/Chordektomie	2	2	1	2	3	2
3	Untersucher	CA/Chordektomie	3	2	3	3	3	3
4	Dr.Speech	CA/Chordektomie	0	2	1	1	3	2
4	Untersucher	CA/Chordektomie	1	2	3	2	3	3
5	Dr.Speech	CA/Chordektomie	0	3	1	2	3	2
5	Untersucher	CA/Chordektomie	2	2	2	2	3	3
6	Dr.Speech	CA/Chordektomie	3	2	3	3	3	3
6	Untersucher	CA/Chordektomie	2	2	3	3	3	3
7	Dr.Speech	CA/Chordektomie	2	1	1	3	2	3
7	Untersucher	CA/Chordektomie	3	1	3	2	2	3
8	Dr.Speech	LP/Chordektomie	2	2	2	0	3	1
8	Untersucher	LP/Chordektomie	1	2	2	1	3	3
9	Dr.Speech	LP/Chordektomie	2	1	2	2	2	2
9	Untersucher	LP/Chordektomie	3	1	3	3	2	3
10	Dr.Speech	CA/Chordektomie	3	1	1	1	1	1
10	Untersucher	CA/Chordektomie	3	1	3	1	1	1
11	Dr.Speech	Reinke bds/MLS	3	3	3	1	1	2
11	Untersucher	Reinke bds/MLS	2	3	3	0	1	1
12	Dr.Speech	Zyste/MLS	0	2	1	0	1	1
12	Untersucher	Zyste/MLS	1	2	2	1	1	1

Patient	Bewerter	Diagnose/Therapie	prae R	prae B	prae H	post R	post B	post H
13	Dr.Speech	Polyp/MLS	0	2	1	0	1	1
13	Untersucher	Polyp/MLS	2	2	2	1	1	1
14	Dr.Speech	Polyp/MLS	0	1	1	2	0	1
14	Untersucher	Polyp/MLS	2	1	2	1	0	1
15	Dr.Speech	Polyp/MLS	1	1	1	0	1	1
15	Untersucher	Polyp/MLS	2	1	2	1	1	1
16	Dr.Speech	SL-Parese/Laterofixation	0	1	1	2	3	2
16	Untersucher	SL-Parese/Laterofixation	1	1	2	1	3	3
17	Dr.Speech	LP/MLS	0	0	1	2	3	2
17	Untersucher	LP/MLS	1	0	1	1	3	3
18	Dr.Speech	Reinke bds./MLS	3	2	3	1	2	3
18	Untersucher	Reinke bds./MLS	2	3	3	1	2	2
19	Dr.Speech	Kontaktgranulom/MLS	3	3	3	0	2	1
19	Untersucher	Kontaktgranulom/MLS	2	3	3	1	2	2
20	Dr.Speech	LP/MLS	0	3	1	0	2	1
20	Untersucher	LP/MLS	0	3	3	0	3	3
21	Dr.Speech	Polyp/MLS	1	3	1	0	3	1
21	Untersucher	Polyp/MLS	1	1	2	1	2	2
22	Dr.Speech	Papillom/MLS	3	3	3	3	2	3
22	Untersucher	Papillom/MLS	2	2	3	1	2	3
23	Dr.Speech	Polyp/MLS	1	1	1	0	1	1
23	Untersucher	Polyp/MLS	1	1	2	1	0	1
24	Dr.Speech	Polyp/MLS	1	2	1	0	2	1
24	Untersucher	Polyp/MLS	2	2	3	1	2	2
25	Dr.Speech	Kontaktgranulom/MLS	0	1	0	0	3	1
25	Untersucher	Kontaktgranulom/MLS	1	1	1	0	1	1
26	Dr.Speech	LP/MLS	0	0	1	1	2	1
26	Untersucher	LP/MLS	1	0	1	0	2	2
27	Dr.Speech	LP/MLS	2	1	1	0	1	1
27	Untersucher	LP/MLS	1	1	2	3	1	3
28	Dr.Speech	Reinke bds/MLS	3	2	2	0	2	1
28	Untersucher	Reinke bds/MLS	2	2	2	1	2	2
29	Dr.Speech	SLZyste/MLS	0	3	1	0	2	1
29	Untersucher	SLZyste/MLS	1	3	2	0	2	2

Patient	Bewerter	Diagnose/Therapie	prae R	prae B	prae H	post R	post B	post H
30	Dr.Speech	LP bds/MLS	2	0	1	1	2	3
30	Untersucher	LP bds/MLS	1	0	1	2	2	2
31	Dr.Speech	Kontaktgranulom/MLS	0	3	1	0	2	1
31	Untersucher	Kontaktgranulom/MLS	0	3	2	0	2	2
32	Dr.Speech	Reinke bds/MLS	2	2	1	0	1	1
32	Untersucher	Reinke bds/MLS	1	2	2	1	1	1
33	Dr.Speech	Kontaktgranulom/MLS	0	3	1	0	2	1
33	Untersucher	Kontaktgranulom/MLS	1	3	3	0	1	1
34	Dr.Speech	Papillom/MLS	2	1	2	0	0	1
34	Untersucher	Papillom/MLS	3	1	3	1	0	1
35	Dr.Speech	Kontaktgranulom/MLS	0	2	1	0	0	2
35	Untersucher	Kontaktgranulom/MLS	0	2	2	1	0	1
36	Dr.Speech	Polyp/MLS	1	2	2	0	1	1
36	Untersucher	Polyp/MLS	2	2	3	0	1	1
37	Dr.Speech	Polyp/MLS	2	1	1	2	1	1
37	Untersucher	Polyp/MLS	2	1	2	0	1	1
38	Dr.Speech	LP/MLS	3	2	2	0	1	0
38	Untersucher	LP/MLS	2	2	3	1	1	1
39	Dr.Speech	Papillom/MLS	3	2	3	3	2	3
39	Untersucher	Papillom/MLS	2	2	3	2	2	3
40	Dr.Speech	Papillom/MLS	3	1	1	1	1	1
40	Untersucher	Papillom/MLS	2	1	2	1	1	2
41	Dr.Speech	LP/MLS	2	3	2	0	1	1
41	Untersucher	LP/MLS	2	3	3	1	1	2
42	Dr.Speech	Reinke bds/MLS	3	2	2	0	1	1
42	Untersucher	Reinke bds/MLS	2	1	2	1	1	1
43	Dr.Speech	Polyp/MLS	2	1	1	0	1	1
43	Untersucher	Polyp/MLS	1	1	2	0	1	1
44	Dr.Speech	Kontaktgranulom/MLS	0	3	1	0	2	2
44	Untersucher	Kontaktgranulom/MLS	1	3	2	1	2	2
45	Dr.Speech	LP/MLS	3	2	2	0	0	0
45	Untersucher	LP/MLS	3	2	3	0	1	1
46	Dr.Speech	Reinke/MLS	1	2	1	0	1	0
46	Untersucher	Reinke/MLS	2	1	2	0	1	1

Patient	Bewerter	Diagnose/Therapie	prae R	prae B	prae H	post R	post B	post H
47	Dr.Speech	Polyp/MLS	2	2	3	1	1	0
47	Untersucher	Polyp/MLS	2	2	2	0	1	1
48	Dr.Speech	CA/Chordektomie	2	2	2	2	3	1
48	Untersucher	CA/Chordektomie	2	2	3	1	2	3
49	Dr.Speech	LP/MLS	2	1	1	1	2	1
49	Untersucher	LP/MLS	1	1	2	2	2	2
50	Dr.Speech	Polyp/MLS	0	2	1	0	1	0
50	Untersucher	Polyp/MLS	1	2	2	1	1	1
51	Dr.Speech	SL-Parese/Laterofixation	0	0	1	2	3	2
51	Untersucher	SL-Parese/Laterofixation	1	0	1	1	3	3
52	Dr.Speech	Papillom/MLS	0	3	1	0	3	0
52	Untersucher	Papillom/MLS	1	3	3	0	1	1
53	Dr.Speech	Papillom/MLS	2	1	2	0	0	0
53	Untersucher	Papillom/MLS	1	1	2	0	1	1
54	Dr.Speech	LP/MLS	3	3	1	2	3	0
54	Untersucher	LP/MLS	3	2	3	3	3	3
55	Dr.Speech	CA/Chordektomie	2	3	1	2	3	2
55	Untersucher	CA/Chordektomie	1	2	2	3	3	3
56	Dr.Speech	LP/MLS	3	3	3	0	2	1
56	Untersucher	LP/MLS	3	2	3	2	1	2
57	Dr.Speech	Polyp/MLS	0	1	1	0	1	1
57	Untersucher	Polyp/MLS	2	1	2	1	1	2
58	Dr.Speech	Reinke bds./MLS	3	2	3	2	1	2
58	Untersucher	Reinke bds./MLS	2	2	3	1	1	1
59	Dr.Speech	Polyp/MLS	2	1	2	0	2	1
59	Untersucher	Polyp/MLS	1	1	2	1	2	2
60	Dr.Speech	SLZyste/MLS	0	3	1	0	1	0
60	Untersucher	SLZyste/MLS	1	3	3	0	1	1
61	Dr.Speech	Kontaktgranulom/MLS	0	2	1	0	0	2
61	Untersucher	Kontaktgranulom/MLS	1	1	2	1	0	1
62	Dr.Speech	Polyp/MLS	0	2	1	0	1	1
62	Untersucher	Polyp/MLS	1	2	2	1	1	1

Tabelle 2: Beurteilung der Behauchtheit durch den Untersucher und der von „Dr. Speech®“ ermittelten NNE in dB.

	praeoperativ		postoperativ	
	<i>Dr. Speech®</i> (NNE in dB)	<i>Untersucher</i> (wahrgenommene Behauchtheit)	<i>Dr. Speech®</i> (NNE in dB)	<i>Untersucher</i> (wahrgenommene Behauchtheit)
Patient 1	-11,27	1	-7,48	3
Patient 2	-6,42	3	-8,58	2
Patient 3	-8,43	2	-3,76	3
Patient 4	-8,95	2	-6,33	3
Patient 5	-7,51	2	-3,11	3
Patient 6	-8,43	2	-4,98	3
Patient 7	-10,03	1	-7,74	2
Patient 8	-6,59	2	-3,41	3
Patient 9	-11,29	1	-8,71	2
Patient 10	-10,54	1	-11,61	1
Patient 11	-4,68	3	-10,78	1
Patient 12	-9,65	2	-11,85	1
Patient 13	-7,32	2	-10,74	1
Patient 14	-11,21	1	-14,8	0
Patient 15	-10,75	1	-11,64	1
Patient 16	-9,82	1	-4,62	3
Patient 17	-13,55	0	-4,92	3
Patient 18	-7,85	3	-7,43	2
Patient 19	-3,76	3	-4,41	2
Patient 20	-2,97	3	-2,74	3
Patient 21	-3,26	1	-2,88	2
Patient 22	-2,31	2	-5,78	2
Patient 23	-10,09	1	-10,89	0
Patient 24	-7,99	2	-8,35	2
Patient 25	-9,79	1	-5,66	1
Patient 26	-15,76	0	-9,03	2
Patient 27	-11,21	1	-10,69	1
Patient 28	-10,1	2	-7,34	2
Patient 29	-3,59	3	-8,45	2
Patient 30	-13,67	0	-8,12	2
Patient 31	-2,78	3	-6,83	2
Patient 32	-7,47	2	-11,53	1
Patient 33	-4,72	3	-7,19	1
Patient 34	-11,46	1	-16,51	0
Patient 35	-8,39	2	-12,51	0
Patient 36	-8,42	2	-10,21	1
Patient 37	-9,35	1	-11,17	1

	praeoperativ		postoperativ	
	<i>Dr. Speech</i> ® (NNE in dB)	<i>Untersucher</i> (wahrgenommene Behauchtheit)	<i>Dr. Speech</i> ® (NNE in dB)	<i>Untersucher</i> (wahrgenommene Behauchtheit)
Patient 39	-7,1	2	-6,29	2
Patient 40	-10,51	1	-11,03	1
Patient 41	-5,71	3	-10,29	1
Patient 42	-11,96	1	-10,66	1
Patient 43	-11,03	1	-9,21	1
Patient 44	-4,61	3	-7,94	2
Patient 45	-9,13	2	-15,27	1
Patient 46	-8,19	1	-10,55	1
Patient 47	-8,37	2	-10,03	1
Patient 48	-8,2	2	-4,19	2
Patient 49	-11,84	1	-7,13	2
Patient 50	-8,49	2	-9,96	1
Patient 51	-15,5	0	-2,47	3
Patient 52	-3,65	3	-11,37	1
Patient 53	-10,92	1	-13,72	1
Patient 54	-6,62	2	-1,52	3
Patient 55	-5,79	2	-3,92	3
Patient 56	-3,86	2	-7,52	1
Patient 57	-10,58	1	-11,29	1
Patient 58	-8,21	2	-10,03	1
Patient 59	-9,63	1	-6,12	2
Patient 60	-3,61	3	-10,37	1
Patient 61	-8,21	1	-12,1	0
Patient 62	-8,17	2	-10,63	1

6. Literaturverzeichnis

1. Ananthapadmanabha T. V (1995) Acoustic factors determining perceived voice quality. (Discussion with Hirano M) In: Vocal fold physiology: voice quality control: proceedings of the VIIIth Vocal fold physiology conference, Kurume, 1994, Fujimura O, Hirano M. (Ed.) S. 113-126
2. Abberton E, Fourcin E (1984) Electrolaryngography. In: Code C, Ball M (ed) Experimental clinical phonetics. Croom Helm, London
3. Anders L C, Hollien H, Hurme P, Sonninen A, Wendler J (1988) Perception of hoarseness by several classes of listeners. *Folia Phoniatr* (40): 91-100
4. Anders, L C (1985) Auditive Beurteilung von Stimmen unterschiedlicher Heiserkeitsgrade. Diss., Halle (Saale)
5. Atkinson J (1976) Inter- and intraspeaker variability in fundamental voice frequency. *J Acoust Soc Am* 60 (2): 440-445
6. Baken R J (1996) Clinical measurements of speech and voice. Singular publishing group, San Diego, London
7. Bergbreiter R, Ptok M, Sesterhenn G, Arold R (1992) Untersuchungen zur Wertigkeit von Aufsprechmaterial für die Sonographie in der phoniatischen Praxis. *Eur Arch ORL, Suppl* 178-179
8. Bielamowicz S, Kreimann J, Gerratt B R, Dauer M S, Berke G S (1996) Comparison of voice analysis systems for perturbation measurement. *J Speech Hear Res* (39): 126-134
9. Bless D M, Baken R J (1992) Assessment of Voice. *J of Voice* 6: 95-97
10. Bough I D, Heurer R J, Sataloff R T, Hills J R, Cater J R (1996) Intrasubject variability of objective voice measures. *J of Voice* 10 (2): 166-174

11. Boyanov B, Hadjitodorov S (1997) Acoustic analysis of pathological voices. *IEEE Engineering in medicine and biology S*.74-82
12. Boyanov B, Doskov D, Ivanov T, Hadjitodorov S (1992) PC based system for analysis of the voice of patients with laryngeal disease. *Ear Nose and Throat J* (70): 767-772
13. Case J L (1999) Technologie in the assesment of voice disorder. *Seminars in speech and language* (20): 169-184
14. Cooley J W, Tukey J W (1965) An algorithm for the machine calculation of complex Fourier series. *Math of Comput* (19): 297-301
15. Dejonckere P H, Lebacq J (1987) Harmonic emergence in formant zone of sustained /a/ as a parameter for evaluating hoarseness. *Acta oto-rhino-laryngol belg* (41) 6:988-996
16. Dejonckere P H (1991) Spectral characteristics of dysphonia: relation with pitch perturbations, perceptual, aerodynamic and anatomical aspects. In: *Die Krankheit der Stimme die Stimme der Krankheit*. Gundermann H (Hrsg.) Gustav Fischer Verlag, Stuttgart, Jena, New York (1991) S 56-61
17. Dejonckere P H, Obbens C, de Moor G M, Wieneke G H (1993) Perceptual evaluation of dysphonia: Reliability and relevance. *Folia Phoniatr* (45):76-83
18. Dejonckere P H, Lebacq J (1996) Acoustic, perceptual aerodynamic and anatomical correlations in voice pathology. *ORL* (58) 326-332
19. De Krom G (1994) Acoustic correlates of breathiness and roughness: experiments on voice quality. Published doctoral dissertation. OTS, Research Institute for Language and Speech, Utrecht University, Netherlands
20. Dudley H, Gruenz O.O (1946) Visible speech translators with external phospor. *J Acoust Soc Am* 18: 62-73

21. Fant G (1993) Some problems in voice source analysis. *Speech communication* 13 : 7-22
22. Fex S (1992) Perceptual evaluation. *J of Voice* 6: 155-158
23. Friedrich G (1996) Qualitätssicherung in der Phoniatrie. *HNO* 44: 401-416
24. Friedrich G, Bigenzahn W, Zorowka P (2000) *Phoniatrie und Pädaudiologie*. Verlag Huber, Göttingen
25. Fröhlich M, Michaelis D, Strube H, Kruse E (1997) Acoustic voice quality description: case studies for different regions of the hoarseness diagram. In: Wittenberg T, Mergell P, Tigges M, Eysholdt U (eds) *Advances in quantitative laryngoscopy, 2nd round table*, Erlangen, S. 143-150
26. Fröhlich M, Michaelis D, Strube H, Kruse E (1998) Voice quality assessment by means of the hoarseness diagram. *J Speech Lang Hearing Res* (64):112-120
27. Fröhlich M, Michaelis D, Strube H, Kruse E (2000) Acoustic voice analysis by means of the hoarseness diagram. *J. Speech Lang Hear Res* (43): 706-720
28. Frøkjær-Jensen B, Prytz S (1974) Evaluation of speech disorders by means of long-time-average-spectra. *ARIPUC* 8: 227-237
29. Grützmacher N (1927) Eine neue Methode der Klanganalyse. *Elektr Nachr Techn* (4): 533-545
30. Habermann G (1976) Zur Bewertung krankhafter Stimmklänge mit dem Gehör. *Laryng Rhinol* (55):245-249
31. Hammarberg B (2000) Voice research and clinical needs. *Folia phoniatr Logop* (52): 93-102

32. Hammarberg B, Gauffin J, Sundberg J, Wedin L (1980) Perceptual and acoustic correlates of abnormal voice qualities. *Acta Otolaryngol* (90):441-451
33. Hammarberg B, Fritzell B, Gauffin J, Sundberg J (1986) Acoustic and perceptual analysis of vocal dysfunction. *J Phonetics* (14): 533-548
34. Hammarberg B, Gauffin J (1994) Perceptual and acoustic characteristics of quality differences in pathological voices as related to physiological aspects. In: *Vocal fold physiology: voice quality control: proceedings of the VIIIth Vocal fold physiology conference, Kurume, 1994*, Fujimura O, Hirano M. (Ed.) S. 113-12
35. Hedelin P, Huber D (1990) Pitch period determination of aperiodic signals. In: *Proceedings of ICASSP 1990, Vol. 1*, S. 361-364
36. Hirano M (1981) *Clinical examination of voice*. Springer, New York
37. Hirano M, Hibi S, Yoshida T, Hirade Y, Kasuya H, Kikuchi Y (1988) Acoustic analysis of pathological voice. *Acta Otolaryngol (Stockh)* 105: 432-438
38. Hirano M (1989) Objective evaluation of the human voice: Clinical aspects. *Folia Phoniatr* (41): 89-144
39. Horii Y (1982) Jitter and shimmer differences among sustained vowel phonations. *J Speech and Hearing Res* (25): 12-14
40. Huang, Z., Minifie, F., Kasuya, H., & Lin, X. (1995). Measures of vocal function during change in vocal effort level. *Journal of Voice*, 9(4): 429-438
41. Huang D (2001) persönliche e-mail
42. Huang D (1998) Relationship between acoustic measures of voice and judgments of voice quality. Artikel unter www.drspeech.com

43. Isshiki N, Okamura H, Tanabe M, Morimoto M (1969) Differential diagnosis of hoarseness. *Folia Phoniatr* (21): 9-19
44. Isshiki N, Takeuchi Y (1970) Factor analysis of hoarseness. *Studia phonologica* (5): 37-44
45. Kasuya H, Ogawa S, Mashima K, Ebihara S (1986a) Normalised noise energy as an acoustic measure to evaluate pathologic voice. *J Acoust Soc Am* 80 (5): 1329-1334
46. Kasuya H, Ogawa S, Kikuchi Y, Ebihara S (1986b) An acoustic analysis of pathological voice and its application to the evaluation of laryngeal pathology. *Speech Communication* 5 (2): 171-81
47. Kasuya H, Ogawa S, Kikuchi Y (1986c) An adaptive comb filtering method as applied to acoustic analysis of pathological voice. *Proc international conference on acoustics, speech, and signal proceedings* S. 669-672
48. Kasuya H, Zue W, Endo Y (1993) Measurements of laryngeal turbulence noise in pathological voice. *Annual ASHA Convention*
49. Klingholz F, Martin F (1983) Speech wave aperiodicities at sustained phonation in functional dysphonia. *Folia phoniatri* (35): 312-327
50. Klingholz F (1986) *Die Akustik der gestörten Stimme*. Thieme, Stuttgart, New York
51. Klingholz F (1987) The measurement of the signal-to-noise ratio (SNR) in continuous speech. *Speech Commun* (6): 15-26
52. Klingholz F (1978) Die Bewertung der Heiserkeit. *Folia phoniatri* (30): 257-265
53. Koenig W, Dunn H K, Lacey L Y (1946) The sound spectrograph. *J Acoust Soc Am* (28): 19-49

54. Kreimann J, Gerratt B R, Kempster G, Erman A, Berke G (1993) Perceptual evaluation of voice quality: Review, tutorial and a framework for future research. *J Speech and Hearing Res*: (36) 21-40
55. Kreimann J, Gerratt B R, Precoda K, Berke G S (1992) Individual differences in voice quality perception. *J Speech and Hearing Res* 35: 512-520
56. Laver J, Hiller S, Beck J M (1992) Acoustic waveform perturbations and voice disorders *J of Voice* (6): 115-126
57. Liebermann P (1961) Perturbation in vocal pitch. *J Acoust Soc Am* (33): 597-603
58. Löfquist A, Manderson B (1987) Long-time average spectrum of speech and voice analysis. *Folia phoniat.* 39: 221-229
59. Martin D, Wolfe V (1995) Pathological voice type and the acoustic prediction of severity. *J of Speech Hear Res* (38): 765-771
60. Mathelitsch L, Friedrich G (1995) *Die Stimme*. Springer, Berlin, Heidelberg, New York
61. Michaelis D, Fröhlich M, Strube H, Kruse E (1998) Reliabilität akustischer Stimmgütebeschreibungen bei reduziertem Umfang der Stimmaufnahmen. In: Gross M, Eysholdt U (Hrsg) *Aktuelle phoniatisch-pädaudiologische Aspekte 1997/98 Band 5*, Median, Heidelberg S. 48-53
62. Michaelis D, Fröhlich M, Strube H W (1998) Selection and combination of acoustic features for description of pathologic voices. *J Acoust Soc Am* 103 (3) 1628-1639
63. Michaelis D, Strube H, Kruse E (1997a) Reliabilität und Validität des Heiserkeits-Diagramms. In: Gross M, Eysholdt U (Hrsg) *Aktuelle phoniatisch-pädaudiologische Aspekte 1996 Band 4*, Median, Heidelberg S. 25-28

64. Michaelis D, Garamss T, Strube H W (1997b) Glottal to noise excitation ratio- A new measure for describing pathological voices. *Acta acust* (83): 700-706
65. Mitev P, Hadjitodorov S (2000) A method for turbulent noise estimation in voiced signals. *Med Biol Engin Comput* (38) 625-631
66. Moore, G P (1976) Observation on laryngeal disease, laryngeal behaviour and voice. *Ann Otol Rhinol Laryngol* (85): 553-564
67. Nawka T, Anders L C, Wendler J (1994) Die auditive Beurteilung heiserer Stimmen nach dem RBH-System. *Sprache Stimme Gehör* 18: 130-138
68. Nawka T (2000) Gibt es eine objektive Einschätzung der Stimmqualität. In: Gross E, Kruse E (Hrsg.) Aktuelle phoniatisch-pädaudiologische Aspekte 2000/2001 Band 8, Median, Heidelberg S. 57-61
69. Nessel E (1960) Über das Tonfrequenzspektrum der pathologische veränderten Stimme. *Acta Otolaryngol. Suppl.* 157
70. Ptok M (1998) Objektive Beschreibung der Stimmgüte unter Verwendung des Heiserkeits-Diagramms. *HNO* (46): 684-689
71. Rabinov C R, Kreiman J, Gerratt B R, Bielamowicz S (1995) Comparing Reliability of perceptual ratings of roughness and acoustic measures of jitter. *J Speech Hear Res* (38): 26-32
72. Reker U, Wesselmann U (1991) Probleme der objektiven Stimmbeurteilung. In: Gundermann H (Hrsg) Die Krankheit der Stimme die Stimme der Krankheit. Gustav Fischer Verlag, Stuttgart, Jena, New York
73. Schultz-Coulon H-J (1975) Bestimmung und Beurteilung der individuellen mittleren Sprechstimmlage. Experimentelle Studie. *Folia Phoniatic* (27): 375-386

74. Schultz-Coulon H-J, Battmer R D, Fedders B (1979) Zur quantitativen Bewertung der Tonhöenschwankungen im Rahmen der Stimmfunktionsprüfung. *Folia Phoniatr* (31): 56-59
75. Schultz-Coulon H-J, Wenn P (1986) Echtzeitspektrographische Untersuchungen an Sängern. *Sprache-Stimme-Gehör* 10: 9-14
76. Schultz-Coulon H-J, Klingholz F (1988) Objektive und semiobjektive Untersuchungsmethoden der Stimme. In: Kittel G, Schürenberg B (Hrsg.) *Objektive und semiobjektive Untersuchungsmöglichkeiten von Stimme, Sprache und Gehör*. Deutscher Ärzte Verlag, Köln
77. Schultz-Coulon H-J (1980) Die Diagnostik der gestörten Stimmfunktion. *Arch Otorhinolaryngol* 227: 1-169 (Kongressbericht)
78. Sedláček K (1967) Die Beurteilung des individuellen Spektrums der Stimme. *HNO* 15: 190
79. Shoji K, Regenbogen E, Yu J D, Blaugrund S M (1992) High-frequency power ratio of breathy voice. *Laryngoscope* (102):267-271
80. Silvermann E M, Zimmer C H (1978) Effect of the menstrual cycle on voice quality. *Arch Otolaryngol* (104) 7-10
81. Titze I, Horii Y, Scherer R (1987) Some technical consideration in voice perturbation measurements. *J Speech and Hearing Res* 30: 252-260
82. Titze I, Winholtz W S (1993) Effect of microphone type and placement on voice perturbation measurements. *J Speech and Hearing Res* 36: 1177-1190 (28): 26-32
83. Trendelenburg W, Franz E (1935) Untersuchungen an schnellveränderlichen Schallvorgängen. *Z Techn Phys* 16: 513-516
84. Wendahl R W (1966) Some parameters of auditory roughness. *Folia Phoniatr*

85. Wendler J, Anders L C (1986) Hoarse Voices: On the reliability of acoustic and auditory classifications. In: Proceedings XXth congress IALP Tokio, 1986
86. Wendler J, Rauhut A, Krüger H (1986) Classification of voice qualities. *J Phonetics* (14): 483-488
87. Wendler J (Hrsg.) Seidner S, Kittel G, Eyshold U (1996) Lehrbuch der Phoniatrie und Pädaudiologie. Thieme, Stuttgart, New York
88. Winkel F (1971) How to measure the effectiveness of stage singers voice. *Folia phoniatri* (23): 228-233
89. Wirth G (1995) Stimmstörungen. Deutscher Ärzte-Verlag Köln
90. Wuyts F L, de Bodt M S, Molengerghs G, Emacle M, Heylen L, Millet B, van Lierde K, Raes J, van de Heyning P H (2000) The Dysphonia Severity Index: An objective measure of vocal quality based on a multiparameter approach. *J Speech Lang Hearing Res*: (43)796-809
91. Yanagihara N (1962) Acoustic studies on hoarseness. *Oto-rhino-laryng. Clinic Kyoto* (55): 357-398
92. Yanagihara N (1967) Hoarseness: Investigation of the physiological mechanisms. *Ann Otol* (76): 472-488
93. Yumoto E, Gould W J (1982) Harmonic-to-noise ratio as an index of the degree of hoarseness. *J Acoust Soc Am* (71): 1544-1550
94. Yumoto E (1983) The quantitative evaluation of hoarseness. *Arch Otolaryngol* (109): 48-52
95. Yumoto E, Sasaki Y, Okamura H (1984) Harmonic-to-noise ratio and psychophysical measurement of the degree of hoarseness. *J speech hear res* (27): 2-6

96. Zwirner P, Michaelis D, Fröhlich M, Kruse E (1998) Korrelationen zwischen perzeptueller Beurteilung von Stimmen nach dem RBH-System und akustischen Parametern. In: Gross E, Kruse E (Hrsg.) Aktuelle phoniatriisch-pädaudiologische Aspekte 1997 Band 8, S. 63-67
97. Zwirner P, Michaelis D, Kruse E (1996) Akustische Stimmanalysen zur Dokumentation der Stimmrehabilitation nach laserchirurgischer Larynxkarzinomresektion. HNO (44):514-520

Lebenslauf

Name: David Christian Kasper

Geburtsdatum: 17.04.1969

Geburtsort: Orsoy, jetzt Rheinberg, Deutschland

Schulbildung: 7/75 - 6/79 Grundschule in Moers
7/79 - 6/88 Gymnasium in Moers
6/88 Abitur

Studium: 4/90 Immatrikulation an der Heinrich-Heine-
Universität in Düsseldorf, Beginn des Studiums
der Humanmedizin
11/97 3. Staatsexamen, Gesamtnote: gut

Tätigkeit: 4/98 AiP in der Klinik für Hals-Nasen-Ohren
Heilkunde, Kopf- und Halschirurgie, plastische
Operationen, Phoniatrie und Pädaudiologie,
Städtische Kliniken Neuss
Chefarzt: Prof. Dr. med. H.-J. Schultz-Coulon
10/99-11/03 seitdem Assistenzarzt ebendort

Danksagung

Besonders möchte ich mich bei Christine, meinen Eltern, vielen Freunden, und Kollegen für die ausdauernde Unterstützung und die immer wieder neue Motivationsförderung bedanken; bei Herrn Prof. Schultz-Coulon für die Überlassung des Themas, seine fachliche Kompetenz und Geduld; bei Herrn Dr. Hennig für das gelegentliche Erstellen von Stimmaufnahmen; bei Fr. Dr. Weigt für ihre Ungeduld; bei Herrn Prof. Mau für statistische Erklärungen und bei Marco Carvascero für die Einführung in die Akustik; und natürlich bei den vielen oft sehr interessierten Patienten, die mir durch ihre Nachuntersuchungen diese Arbeit ermöglichten.

Zusammenfassung

Zur Anwendbarkeit der Stimmanalysesoftware „Dr. Speech®“ zur prae- und postoperativen Dokumentation der Stimmqualität

Hintergrund und Fragestellung: Heiserkeit ist ein Leitsymptom von Kehlkopferkrankungen. Zur Therapiekontrolle ist es wünschenswert, bei der täglichen klinischen Arbeit ein Instrument zur Verfügung zu haben, welches die Qualität und Quantität der Heiserkeit zuverlässig und mit kurzem Zeitaufwand misst und bewertet. Neben den auditiven Bewertungsmethoden (Heiserkeits-Skalen, RBH-Schema), die teils hohen inter- und intraindividuellen Schwankungen unterliegen, existiert auch eine große Anzahl verschiedener elektroakustischer Methoden der Heiserkeitsbewertung. Die Firma Tiger electronics Inc. bietet mit der Stimmanalysesoftware Dr. Speech ® ‘heben der Berechnung etablierter akustischer Heiserkeitsparameter (Jitter, Shimmer, NNE) die Möglichkeit einer computerassistierten Bewertung des Stimmklangs bezüglich des Grades der Heiserkeit, der Rauigkeit und der Behauchtheit. Ob diese computergestützte RBH-Beurteilung ‘‘ zuverlässig und sinnvoll ist, und ob sie die auditive Bewertung im RBH-Schema ersetzen kann, ist Gegenstand dieser Arbeit.

Patienten und Methoden: Die Stimmproben von 62 Patienten mit verschiedenen benignen und malignen Kehlkopferkrankungen wurden vor und einige Wochen nach einem mikrolaryngoskopischen Eingriff an den Stimmlippen durch einen erfahrenen Untersucher und durch Dr. Speech ® bezüglich des Grades der Heiserkeit, der Rauigkeit und der Behauchtheit in einer 4stufigen Skala bewertet. Die auditive RBH-Beurteilung wurde mit der computergestützten RBH-Bewertung statistisch verglichen. Die Analyse zweier Einzelfälle ergänzt den statistischen Teil. Dabei werden auch andere akustische Parameter (Jitter (PPQ), Shimmer (APQ), Normalized-Noise-Energy (NNE)), Spektralanalysen und Sonagramme hinzugezogen.

Ergebnisse: Die Übereinstimmungen zwischen auditiver und computergestützter RBH-Bewertung sind für die drei bewerteten Qualitäten (Rauigkeit, Heiserkeit, Behauchtheit) sehr unterschiedlich. Lediglich bei der Bewertung der Behauchtheit ‘liegen ausreichende (statistisch signifikante) Übereinstimmungen vor. Knapp 85% aller bewerteten Stimmproben (n=124) werden vom Untersucher und durch den Computer identisch bezüglich des Grades der Behauchtheit bewertet. Die Bewertungen der Heiserkeit ‘‘(35,5% identische Bewertungen) und der Rauigkeit (25,8% identische Bewertungen) zeigten nur tendenzielle Übereinstimmungen, teils ohne statistische Signifikanz. Daneben wurden hohe statistisch signifikante Zusammenhänge des akustischen Parameters NNE mit der wahrgenommenen Behauchtheit festgestellt (Spearman's R = 0,85, p < 0,001) Die Einzelfallbesprechungen verdeutlichen, dass die Interpretation von APQ, PPQ und NNE sowie der Spektralanalysen und Sonagramme eine nützliche, sinnvolle und objektive Ergänzung zur auditiven Heiserkeitsbewertung darstellt.

Schlussfolgerung: APQ, PPQ und NNE sind wertvolle Parameter zur Dokumentation des Heiserkeitsgrades. Sie spiegeln die Ergebnisse der auditiven Bewertung wieder, erhärten und unterstützen diese. Die direkte Bewertung von Rauigkeit, Behauchtheit und Heiserkeit durch den Computer (Dr. Speech ®)‘‘ ist jedoch mit Ausnahme der computergestützten Behauchtheitsgraduierung nicht zuverlässig.