Das Göttinger Heiserkeits-Diagramm - Entwicklung und Prüfung eines akustischen Verfahrens zur objektiven Stimmgütebeurteilung pathologischer Stimmen

Dissertation
zur Erlangung des Doktorgrades
der Mathematisch-Naturwissenschaftlichen Fakultäten
der Georg-August-Universität zu Göttingen

vorgelegt von
Dirk Michaelis
aus Braunschweig

Göttingen 1999

D 7
Referent: Prof. Dr. M. R. Schroeder
Korreferent: Prof. Dr. D. Ronneberger
Tag der mündlichen Prüfung: 27.1.2000

Was den Leser erwartet

Im ersten Teil dieser Arbeit wird nach einer Einführung in die qualitative Stimmanalyse das Heiserkeits-Diagramm vorgestellt. Das Heiserkeits-Diagramm ist eine grafische Darstellung der Stimmqualität in zwei Dimensionen. In der einen Richtung ist die Irregularität und in der anderen Richtung der Rauschanteil der Stimme aufgetragen. Besonderer Wert wurde darauf gelegt, dass sich jede gesunde und pathologische Stimme, auch solche mit schweren Stimmstörungen, in dem Diagramm darstellen lassen.

Die Messung des Rauschanteils beruht auf dem neuen akustischen Maß Glottal to Noise Excitation Ratio (GNE), dass in dieser Arbeit entwickelt wird. GNE zeigt gegenüber anderen Maßen, die den Rauschanteil messen, den großen Vorteil, dass er unabhängig gegenüber typischen Irregularitäten des Stimmsignals ist. Dies wird im Vergleich zu zwei Maßen aus der Literatur durch Messungen an synthetischen Signalen belegt.

Die Messung der Irregularität geschieht durch drei akustische Maße: Zwei statistische Maße zur Beschreibung der Periodenlängenschwankung (Jitter) und der Energieschwankung (Shimmer) sowie den mittleren Korrelationswert von je zwei aufeinanderfolgenden Perioden.

Die vier akustischen Maße des Heiserkeits-Diagramms wurden aus 22 Maßen selektiert. Dazu wurden die Korrelationen zwischen den Maßen gemessen, die Dimensionalität des Raumes der akustischen Maße bestimmt und mit einem informationstheoretischem Verfahren die geeignetste Viererkombination gefunden.

Am Ende des ersten Teiles wird der Einfluss des Vokaltraktes auf Jitter und Shimmer untersucht und das Verfahren zur Messung der Periodenlängen auf die Tauglichkeit für sehr unregelmäßige Stimmen getestet.

Dort wird gezeigt, dass der Vokaltrakt grundfrequenzabhängig Jitter und Shimmer wechselseitig ineinander umwandelt. Es wird eine Theorie für den durch Jitter induzierten Shimmer hergeleitet, die sehr gut mit den Messungen übereinstimmt.

In dem zweiten Teil der Arbeit geht es darum das Heiserkeits-Diagramm durch verschiedene Anwendungen zu testen. Hier wird zunächst gezeigt, dass die Vokale ein spezielles Muster im Heiserkeits-Diagramm bilden.

Daraufhin werden Patientengruppen mit gleicher Stimmpathologie analysiert. Es zeigt sich unter anderem, dass sich sechs Gruppen mit verschiedenen Phonationsmechanismen, darunter normale Stimmen und Flüsterstimmen, im Heiserkeits-Diagramm signifikant voneinander unterscheiden.

Es folgt ein Vergleich der akustische Maße mit perzeptiven Größen. Dabei stellt sich heraus, dass Jitter und Shimmer spezifisch für Rauhigkeit sind und GNE spezifisch für Behauchung.

Im darauffolgenden Teil wird die Frage untersucht, ob das bis dahin verwendete relativ umfangreiche Aufnahmeprotokoll verkürzt werden kann. Es wird gezeigt, dass die durchschnittliche Veränderung der Lage im Diagramm relativ gering ist, wenn man statt 28 nur drei Vokale verwendet. Andererseits zeigen Patienten während der Stimmtherapie Veränderungen, die in der gleichen Größenordnung liegen und sich als signifikant herausstellen, so dass hier das umfangreiche Protokoll gerechtfertigt erscheint.

Im letzten Teil wird auf den Katalog mit 48 Heiserkeitsdiagrammen im Anhang hingewiesen. Dort ist die Stimmgüteentwicklung einzelner Patienten zusammen mit den jeweiligen Tagesdiagnosen, basierend auf laryngoskopischen Beobachtungen, zusammengestellt. Dort sind z.B. erfolgreiche Therapieverläufe zu sehen. Nach Angaben der Ärzte, decken sich die klinischen Beobachtungen ausnahmslos mit den Informationen des Heiserkeits-Diagramms.

Inhaltsverzeichnis

Abkürzungen

ΔI_N	Normierte zusätzliche Information
A	(engl.) Asthenic, asthenische, geschwächte Stimme
AKF	Autokorrelationsfunktion
APF	Amplitude Perturbation Factor. Maß für die Unregelmäßigkeit der Amplitude. Wird aus einer Sequenz von Amplituden berechnet. Pro Periode eines Signals wird ein Amplitudenwert berücksichtigt. Der Amplitudenwert wird durch Bestimmung des (positiven oder negativen) Extremums festgelegt, oder durch die Differenz von Maximum und Minimum. Siehe Gleichung 3.12
APQ	Amplitude Perturbation Quotient. Siehe Gleichung 3.11. Siehe auch APF
B	Behauchung, Breathiness
CHNR	Cepstral Harmonic to Noise Ratio
CSL	Computer Speech Lab. Hard und Software zur Stimmaufnahme und Analyse mit einem PC. Hersteller: Kay Elemetrics
EPF	Energy Perturbation Factor. Siehe Gleichung 3.12. Wird aus einer Sequenz von Energie pro Periode berechnet. Die Energiewerte werden durch Summation aller Abtastwerte einer Periode berechnet. Siehe auch APF
EPQ	Energy Perturbation Quotient. Siehe Gleichung 3.11. Siehe auch EPF
G	(engl.) Grade. Ausmaß der Stimmstörung (perzeptiv)
GNE	Glottal to Noise Excitation Ratio. Maß für den Anteil glottaler Stimmanregung gegenüber Rauschanregung
H	Heiserkeit, Hoarseness
J2	Entspricht PPF wobei jedoch die Periodenlänge mit dem Waveform-Matching Verfahren berechnet wurden
J3, ..., J15	Entspricht PPQ sowie $K = 3, ..., 15$ in Formel 3.11 wobei die Periodenlänge mit dem Waveform-Matching Verfahren berechnet wurden
MDVP	Multi-Dimensional Voice Profile. Software zum CSL von Kay Elemetrics
MWC	Mean Waveform-Matching-Coefficient, Periodenkorrelation: mittlerer (Kurzzeit-) Korrelationswert aufeinanderfolgender Perioden
NNE	Normalized Noise Energy
PF	Perturbation Factor, siehe Gleichung 3.12
PPF	Pitch Perturbation Factor, siehe Gleichung 3.12
PPQ	Pitch Perturbation Quotient, siehe Gleichung 3.11; Periodenlängenbestimmung erfolgt bei MDVP durch Bestimmung des Maximums der Ableitung
PQ	Perturbation Quotient, siehe Gleichung 3.11

R	Rauhigkeit, Roughness
RMS	Root Mean Square (bei N Datenpunkten $x(n)$: RMS=$\sqrt{\frac{1}{N}\sum_{n=1}^{N} x(n)^2}$
S2	Shimmer Maß, entspricht EPF
S3, ..., S15	Shimmer Maß, entspricht EPQ und Formel 3.11 mit $K = 3, ..., 15$
S	(engl.) Strained, Anspannung bei der Phonation (perzeptiv)
SVD	Singular Value Decomposition, Methode zur Berechnung der Hauptrichtungen einer Verteilung sowie der Varianzen in den Hauptrichtungen; gleichzeitig werden die Koordinaten im Hauptachsensystem berechnet

Teil I.

Entwicklung des Heiserkeits-Diagramms

1. Stimmstörungen und Akustik

1.1. Einordnung der Stimmanalyse

Die Sprache ist eines der wichtigen Instrumente, mit denen Menschen miteinander kommunizieren. Neben den informativen Inhalten, die in einer bestimmten Sprache mit einer speziellen Grammatik kodiert sind, erhält der Zuhörer darüberhinaus weitere für den Sprecher charakteristische Mitteilungen. Diese Mitteilungen gehen bewusst oder unbewusst vom Sprecher aus und werden bewusst oder unbewusst vom Zuhörer wahrgenommen. Sie spiegeln eine Fülle von emotionalen Zuständen wider, wie Ruhe, Gelassenheit, Hektik, Gereiztheit, Sympathie, Antipathie usw., die bei einer inhaltlich gleichen, aber geschriebenen Äußerung nicht enthalten wären. Diese emotionalen Zustände werden dabei über die Sprechgeschwindigkeit, die Lautstärke, über die Sprechpausen oder Sprechunterbrechungen, wie z.B. Räuspern oder Schlucken, über die Höhe der Grundfrequenz, über die Deutlichkeit oder Undeutlichkeit der Artikulation und weitere Aspekte vermittelt. Darüberhinaus trägt die gesamte Körpersprache zur Kommunikation bei, wenn zwei Sprecher sich nicht nur hören, sondern auch sehen können.

Die uneingeschränkte Funktionstüchtigkeit des Sprechapparates ist notwendig, damit der Sprecher neben dem Inhalt die gesamte Bandbreite dieser Emotionen in angemessener Weise vermitteln kann. Zu diesem Sprachapparat ist fast der ganze Körper zu rechnen, wobei den Körperteilen verschiedene Aufgaben zugeordnet sind. Diese Aufgaben bei der Sprachproduktion sind: 1. Erzeugung eines Luftstroms aus den Lungen durch die Luftröhre mit Hilfe der Brust-, Rücken- und Bauchmuskulatur als Motor oder Energiequelle der Lauterzeugung 2. Zeitliche Modulation des Luftstromes zur Anregung akustischer Schwingungen an den Stimmlippen oder an Einengungen im sogenannten Stimmkanal als Quelle des Stimmsignals 3. Färbung dieses Stimmsignals in die charakteristischen Grundbausteine der Sprache, die Phoneme, und damit Kodierung der zu übermittelnden Sprache in ein akustisches Alphabet 4. Umsetzung des gewünschten sprachlichen Inhaltes in Sprache durch Steuerung bzw. Regelung der Atem- und Sprachmotorik bei gleichzeitiger akustischer Kontrolle der produzierten Laute durch das Gehirn, Nerven und Sensoren. Ist einer der benötigten Organkomplexe durch eine Verletzung oder Erkrankung beeinträchtigt, so geht die Möglichkeit zur Informationsübermittlung erst bei relativ schweren Krankheits- oder Verletzungsgraden verloren, wohingegen sich die Möglichkeit zur Kommunikation von emotionalen Zuständen schon bei leichteren Erkrankungs- und Verletzungsgraden stark einschränkt und den betroffenen Sprecher in seinen Eigenschaften zur feinabgestimmten Mitteilung stark beschneiden. Als Beispiel sei

hier daran erinnert, wie sehr eine zeitweise Heiserkeit die Möglichkeiten des sprachlichen Ausdrucks reduziert.

Von den genannten Organkomplexen stehen in dieser physikalischen Arbeit die Stimmlippen bzw. die Quellen der akustischen Schwingungen und ihre Eigenschaften bei krankheitsbedingten Einschränkungen im Vordergrund. Da die Stimmlippen sehr klein sind, wenige Zentimeter lang und nur einige Millimeter breit, können selbst gerade eben sichtbare organische Veränderungen wie Knötchen, Papillome, Ödeme, Tumore oder auch leichte Entzündungen zu einer signifikanten Einschränkung des Sprechvermögens führen. Eine solche Behinderung kann für Berufsgruppen wie Lehrer, Sänger, Telefonisten usw., bei denen es stark auf die Sprachfähigkeiten ankommt, Berufsunfähigkeit bedeuten. Deshalb ist es bei operativen Eingriffen an den Stimmlippen besonders wichtig, so weit wie möglich deren Struktur und Funktionsfähigkeit zu erhalten. Weiterhin muss besondere Sorgfalt auf die postoperative Stimmtherapie gelegt werden. Bei richtiger Therapie ist es möglich, dass sich aus operationsbedingten Stimmlippenresten wieder funktionsfähige Stimmanregungsorgane entwickeln. Wird die Therapie vernachlässigt oder in eine falsche Richtung geführt, so entwickelt der Patient unter Umständen kurzfristig Ersatzmechanismen zur Stimmproduktion, die ihm zwar relativ leicht fallen, mit denen er aber auf lange Sicht keine hohe Stimmqualität erreichen kann. Diese Ersatzmechanismen können z.B. eine antrainierte Flüsterstimme oder Ersatzschwingungen der über den Stimmlippen liegenden Taschenfalten sein. Ziel einer optimalen postoperativen Therapie muss es jedoch sein, im Rahmen der organischen Möglichkeiten des Patienten, mittel- oder langfristig wieder eine möglichst hohe Stimmqualität zu erreichen.

Ein wichtiger Teil der postoperativen Stimmtherapie ist deshalb die Beurteilung des Therapieverlaufs, um die Erfolgschancen und die weitere Vorgehensweise bei der Therapie einschätzen zu können. Für diese Beurteilung stehen zunächst die subjektiven Eindrücke des Patienten und des Therapeuten im Vordergrund, weiterhin können optische oder akustische Verfahren herangezogen werden. Wenn der Patient motiviert ist und von seinem Vorankommen überzeugt, bedarf es für die therapeutischen Belange des Patienten nur wenig Beurteilung durch Therapeuten oder anderer Verfahren. Wenn es jedoch um eine Langzeitbewertung, um den Vergleich verschiedener therapeutischer Methoden, um Studien mit großen Patientenzahlen oder mehreren behandelnden Therapeuten geht, so reicht die rein subjektive akustische Stimmbewertung nicht aus. Hier beginnt das Einsatzgebiet von anderen, zum Teil technischen Hilfsmitteln zur Stimmanalyse, die hier zunächst im Überblick vorgestellt werden sollen.

2. Physikalische Methoden zur Beurteilung der Stimmgüte

2.1. Physikalische Stimmanalyse

Unter dem Begriff physikalische Stimmanalyse werden alle Messmethoden zusammengefasst, die einen physikalischen Effekt oder eine physikalische Messgröße verwenden, um Aussagen über das Stimmschallsignal zu bekommen. Im folgenden werden drei Verfahren vorgestellt, die unterschiedliche physikalische Messgrößen benutzen.

2.2. Aerodynamische Verfahren

Aerodynamische Verfahren messen die Volumengeschwindigkeit der Luft beim Sprechen. Mit ihnen kann man sowohl den Gleich- als auch den Wechselanteil der Volumengeschwindigkeit bestimmen. Damit kann man unter anderem durch Integration der Volumengeschwindigkeit Aussagen über das Atemverhalten und die Effiziens der Stimmgebung bekommen.

Der Gleichanteil der Volumengeschwindigkeit ist z.B. auch für die Sprachsynthese mit physikalisch motivierten Modellen von Interesse und kann mit rein akustischen Methoden nicht bestimmt werden. Die Volumengeschwindigkeit stellt diesen fehlenden Parameter zur Sprachsynthese bereit.

Rothenberg [112] hat 1973 die Volumengeschwindigkeit mit einer speziell dafür konstruierten pneumatografischen Maske gemessen, um daraus durch inverse Filterung die Volumengeschwindigkeit an der Glottis zu rekonstruieren. Das Ergebnis dieser Rekonstruktion ist in Abb. 2.1 zu sehen.

Hierbei wurde die Volumengeschwindigkeit bei verschiedenen Grundfrequenzen rekonstruiert. In der Abbildung ist der Knick nach der abfallenden Flanke der Volumengeschwindigkeit gut zu erkennen. Er ist für die Anregung des höherfrequenten Anteils im Sprachsignal verantwortlich. Die Kenntnis dieses Zeitverlaufs der glottischen Volumengeschwindigkeit ist z.B. für die Konstruktion von Sprachsynthesemodellen wichtig.

Die Rothenberg-Maske wird in vielen Studien in Kombination mit inverser Filterung verwendet, um unter bestimmten Fragestellungen den glottalen Fluss zu approximieren [24, 25, 111, 125, 126, 135].

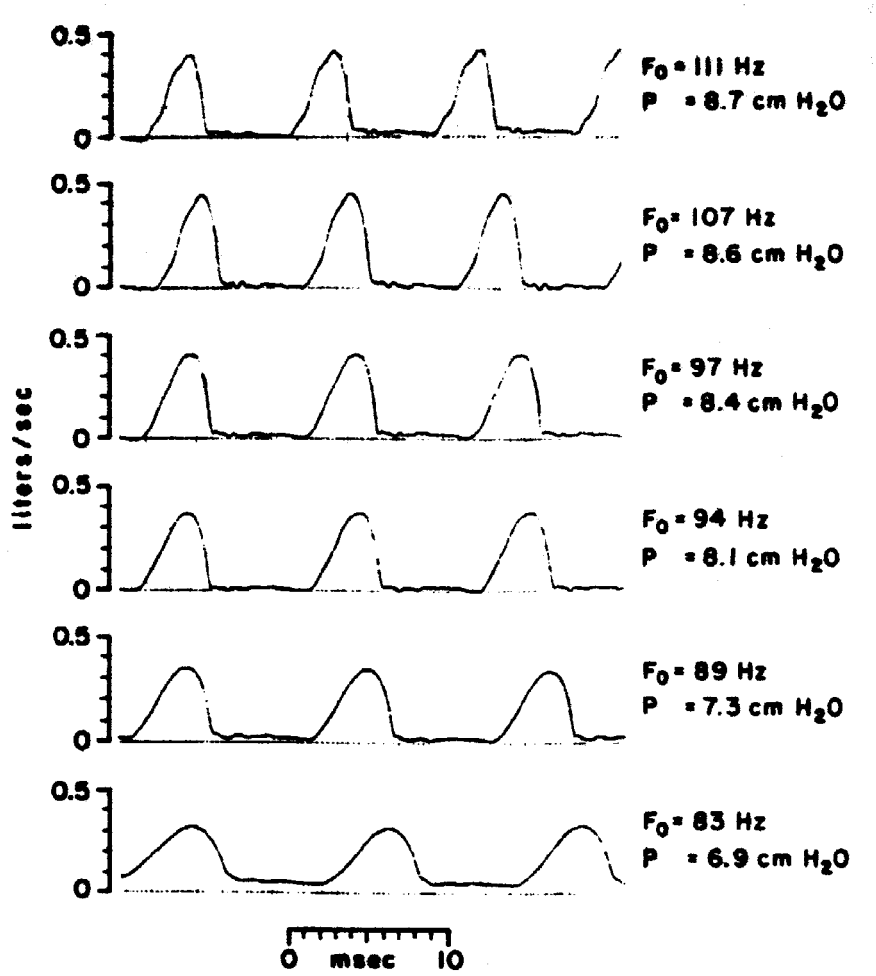


Abbildung 2.1.: Inversgefilterte glottale Volumengeschwindigkeit bei absteigender Grundfrequenz. Rothenberg 1973 [112].

2.3. Elektroglottographie

Eine Übersicht über die Elektroglottographie gibt Baken [10] in seiner Abhandlung. Bei der Elektroglottographie werden zwei Elektroden in der Höhe der Stimmbänder rechts und links an den Kehlkopf gebracht. Mit Hilfe der Elektroden wird bei einer Trägerfrequenz von 300 kHz bis einigen MHz ein Strom mit etwa 10mA durch den Kehlkopf geleitet. Die Änderung der Kontaktfläche zwischen den Stimmlippen bei der Phonation, die durch die Schwingungen hervorgerufen wird, führt zu einer Änderung des Leitwertes des Kehlkopfes um ca. 1 bis 2 Prozent durch Öffnen und Schließen der Stimmlippen. Dieser Leitwert wird gemessen und von der relativ zur Grundfrequenz langsamen Abdrift der Leitfähigkeit des Gewebes während der Phonation durch Hochpassfilterung bereinigt.

In der Abbildung 2.2 aus einem Artikel von Hess [38] ist in der Mitte ein Elektroglottogramm (EGG) gezeigt. Die obere Kurve zeigt zum Vergleich das akustische Schallsignal. Im unteren Bildteil ist das differenzierte Elektroglottogramm (DEGG) dargestellt. Der steile Anstieg des Elektroglottogramms, der sich im differenzierten Elektroglottogramm als Spitze deutlich zeigt, lässt sich zur Detektion der Grundperioden benutzen.

Abbildung 2.2.: (a) Akustisches Signal, (b) Elektroglottogramm und (c) differenziertes Elektroglottogramm eines männlichen Sprechers. Gezeigt ist der Übergang bei einem /ja/. Hess 1987 [38].

Man sieht in Abb. 2.2, dass diese Spitze im Elektroglottogramm schon bei der ersten Schwingung gut ausgeprägt ist, wohingegen das akustische Signal zeigt, dass sich nach dem /j/ das /a/ über mehrere Perioden einschwingt. Das Elektroglottogramm liefert also ein einfacheres Bild als das akustische Signal. Es wird in einigen neuen Arbeiten von Hess [38] und Schoentgen [118] zur genauen Bestimmung der einzelnen Grundperioden benutzt. Schoentgen zeigt, dass die Grundperioden, die er durch Spitzen des differenzierten Elektroglottogramms und durch Nulldurchgänge im akustischen Signal bestimmt, bis auf ca. 0,01 ms übereinstimmen. Das EGG wird in vielen Studien verwendet um z.B. Perturbationsmaße wie Jitter und Shimmer zu berechnen, aber auch um den prinzipiellen Zusammenhang des EGG-Signals mit den Schwingungsvorgängen der Stimmlippen

besser zu verstehen [12, 15, 34, 47, 99, 100, 113, 137, 139, 146].

In dieser Arbeit wird in Kapitel 8 das EGG eingesetzt, um den Einfluss des Vokaltraktes auf Jitter und Shimmer zu untersuchen.

2.4. Visuelle Stimmbewertung

Die sich langsam etablierende digitale Hochgeschwindigkeitsglottographie (HGG), gestattet es einige Fragen zur Stimmproduktion zu beantworten, über die bisher nur Vermutungen oder vage Messungen angestellt werden konnten. Hier wird, nach einer kurzen Einführung, ein Überblick über den Stand der Forschung gegeben, da die Ergebnisse auch für diese Arbeit interessant sind.

Bei der visuellen Bewertung der Stimmgebung wird die Funktionsfähigkeit von Stimmlippen und Kehlkopf mit Hilfe eines optischen Instrumentes,dem Laryngoskop, das einen Einblick in diese Halsregion ermöglicht, bewertet. Weiterhin können Ersatzphonationsmechanismen oder postoperative Heilungsprozesse beobachtet werden. Will man die Stimmlippen in Funktion beobachten, so muss man den Nachteil in Kauf nehmen, dass die Lautbildung durch die optischen Instrumente mehr oder weniger stark beeinflusst wird. Diese Einschränkung ist aber vertretbar, wenn hauptsächlich organische und nicht so sehr artikulatorische Vorgänge beobachtet werden sollen.

Eine wichtiges Mittel bei der optischen Stimmlippenuntersuchung ist die Stroboskopie [26]: Die Blitzfrequenz einer geeigneten Glühlampe wird z.B. mit einem Pedal ein wenig ober- oder unterhalb der Grundfrequenz der Stimmlippenschwingung eingeregelt, so dass jeweils nur kurze, wenig voneinander verschiedene Phasen von einzelnen Stimmlippenschwingungen sichtbar sind. Bei regelmäßiger Stimmlippenschwingung kann man so die Schwingung quasi in Zeitlupe beobachten. Der Vorteil dieser Methode ist, dass sie nur wenig technischen Aufwand erfordert, so dass die Apparatur relativ kostengünstig ist. Ein Nachteil ist, dass man bei unregelmäßigen Schwingungen, wie sie unter anderem bei den oben genannten Ersatzmechanismen auftreten, das zeitliche Schwingungsverhalten nicht gut untersuchen kann. Wird die Laryngoskopie mit einer Filmkamera kombiniert, so bestehen weitere Möglichkeiten zur Dokumentation von Stimmlippenschwingungen bei verschiedenen organischen Zuständen und zur detaillierteren Auswertung von Standbildern oder Zeitlupenaufnahmen.

Eine detaillierte optische Untersuchung der einzelnen Stimmlippenschwingungen ist nur mit Hochgeschwindigkeitsaufnahmen bei 2000 bis 9000 Bildern pro Sekunde möglich. Frühe Untersuchungen dieser Art beschreibt Lieberman [72]. Dort wurden die Aufnahmen der Hochgeschwindigkeitskamera Bild für Bild ausgewertet. Dabei wurden die Stimmlippenränder auf jedem einzelnen Bild genutzt, um die Öffnungsweite (im dort gezeigten Beispiel ca. 2,5mm bis 4,5mm) und die Öffnungsfläche (15 bis 25 Quadratmillimeter) im zeitlichen Verlauf darzustellen. Gleichzeitig wurde der Zeitverlauf des Luftdruckes an den Lippen aufgenommen. Das Ergebnis einer solchen Messung ist für fünf Perioden in Abb. 2.3 dargestellt.

Damit zeigte Lieberman, dass sich Unregelmäßigkeiten in der Glottisflächenzeitfunktion auf den Zeitverlauf des Luftdruckes übertragen, wenn keine plötzlichen Änderungen im Schwingungsmuster der Stimmlippen auftreten, und legte damit nahe, dass man unter solchen Umständen diese Unregelmäßigkeiten ebensogut im akustischen Schallsignal wie im optischen Glottisbild messen kann.

Baer [7] verglich 1983 vier Methoden zur Messung der Glottisschwingungen: Elektroglottographie, Photoglottographie (PGG), Akustik und Hochgeschwindigkeitsaufnah-

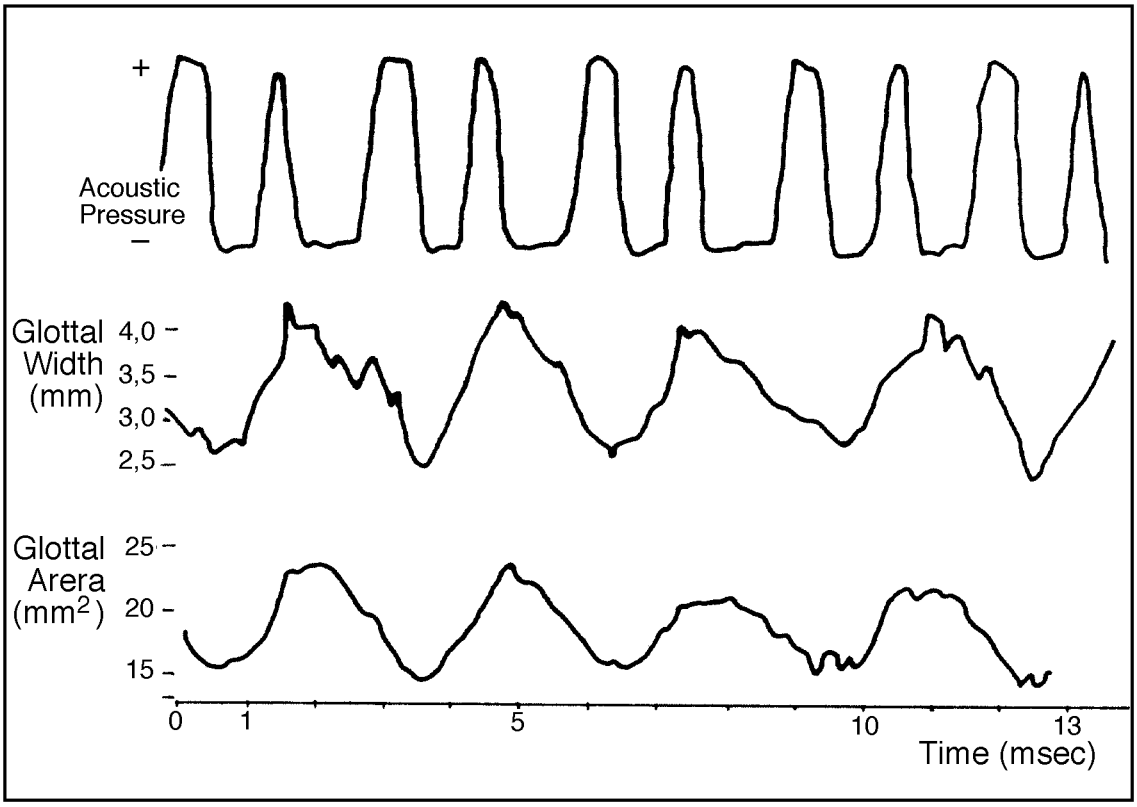


Abbildung 2.3.: Ergebnisse der Auswertung von akustischen Aufnahmen und gleichzeitigen optischen Hochgeschwindigkeitsaufnahmen (Lieberman 1963 [72]).

men. Bei der Photoglottographie wird ein lichtempfindlicher Sensor von außen am Kehlkopf unterhalb der Stimmlippen mit direktem Hautkontakt angebracht. Nach außen wird der Sensor lichtdicht abgeschirmt. In dieser Untersuchung wurde die Beleuchtung der Hochgeschwindigkeitskamera als Lichtquelle benutzt. Entsprechend der Öffnungsfläche der Glottis fällt somit mehr oder weniger Licht auf den Lichtsensor. Die Zeitfunktion der Öffnungsfläche kann auf diese Weise gemessen werden. Die Untersuchung wurde durchgeführt, um die Konsistenz der verschiedenen Messmethoden zu überprüfen.

Das Ergebnis einer solchen, von Baer durchgeführten Messung ist in Abbildung 2.4 dargestellt. Die PGG und Glottal Width Kurven stimmen gut überein. Bemerkenswert ist, dass der optische Verschluss erst kurz nach der Stelle auftritt, an der das EGG die stärkste Steigung aufweist. Dieser Zeitpunkt wird im Allgemeinen mit dem Anregungszeitpunkt der höheren Frequenzen im akustischen Signal gleichgesetzt. Die laufzeitkorrigierte Audioaufnahme bestätigt diese Annahme. Aus der Abbildung ist also ersichtlich, dass zu dem Zeitpunkt, an dem sich die Kontaktfläche am schnellsten erhöht, die Glottis noch nicht geschlossen ist.

Die Autoren folgern aus den Messungen, dass man mit PGG und EGG bei weit geringerem Aufwand einen Großteil der Information von Hochgeschwindigkeitsfilmen erhalten kann.

In einer neueren Arbeit von Kiritani et al. [56] wurde eine digitale Hochgeschwindig-

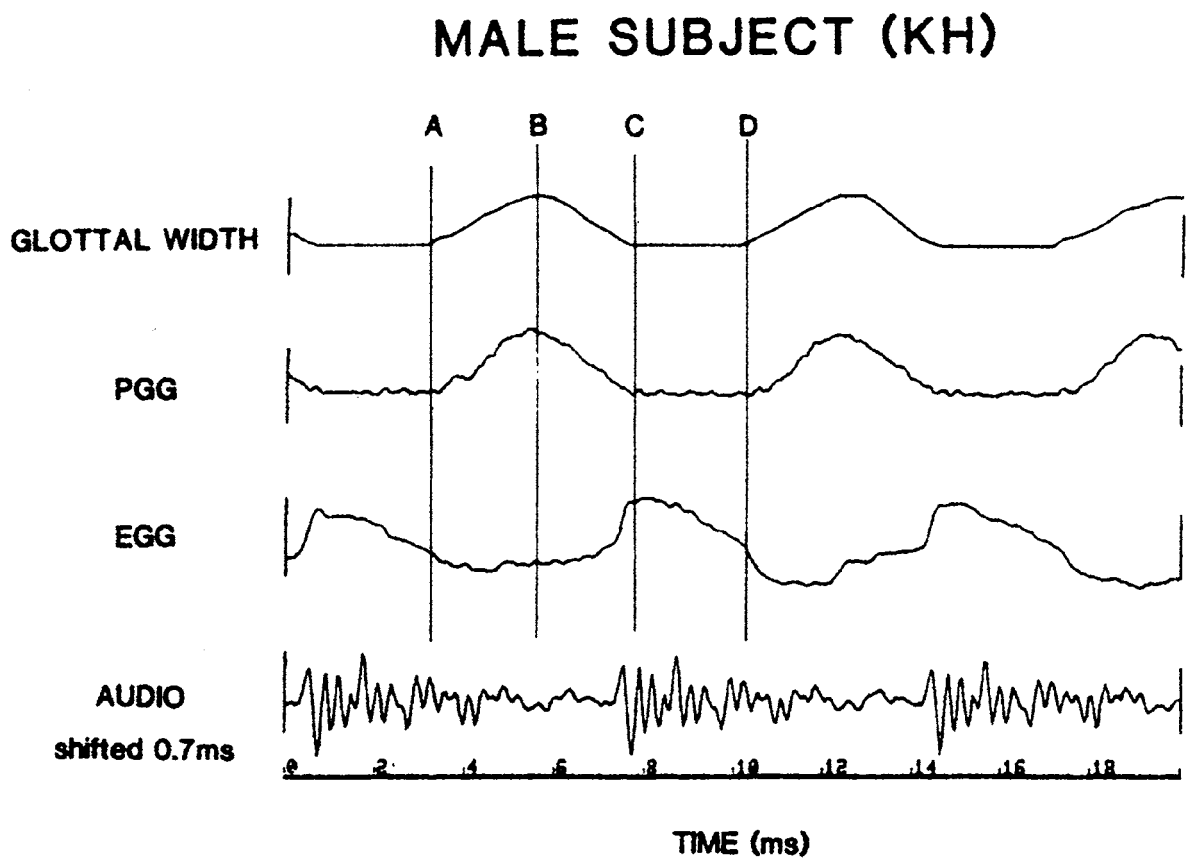


Abbildung 2.4.: Aufnahme einer normalen, männlichen Stimme. Glottal Width: Schwingungsweite, die aus Hochgeschwindigkeitsaufnahmen bestimmt wurde, PGG: Photoglottogramm, EGG: Elektroglottogramm, Audio: akustisches Signal. A: Beginn der Glottisöffnung, B: Maximale Öffnung, C: Verschlusszeitpunkt, D: erneuter Öffnungsbeginn (Baer 1983 [7])

keitskamera benutzt. Der Vorteil der Digitalisierung liegt in der einfacheren Automatisierbarkeit der bildweisen Bestimmung der Glottisöffnungsflächen. Der Nachteil ist die relativ niedrige Bildfrequenz von nur 2000 Hz. Hier wurde nicht nur die Öffnungsweite der Stimmlippen bestimmt, sondern die unilaterale Auslenkung der einzelnen Stimmlippen, so dass rechts-links- Asymmetrien analysiert werden können.

In Abb. 2.5 ist das akustische Signal und die Öffnungsfläche für eine normale Stimme abgebildet. In Abb. 2.6 ist das Verhalten einer Glottis bei einseitiger Stimmlippenlähmung zu erkennen. Dieses Beispiel macht einen Vorteil der Bildverarbeitung des HGG gegenüber anderen Verfahren (Messung des akustischen Signals oder des EGGs) deutlich, denn nur durch die Darstellung der Schwingung der linken und der rechten Stimmlippe kann man das akustische Signal (im Bild oben) richtig interpretieren.

Arndt und Schäfer führten 1994 [1] den Weiten-Längen-Quotient als Maß für die Amplitudengröße ein, um die Videoaufnahmen in einer Kenngröße zusammenzufassen. Es wurden dabei Einzelbilder aus stroboskopischen Aufnahmen mit maximaler Schwin-

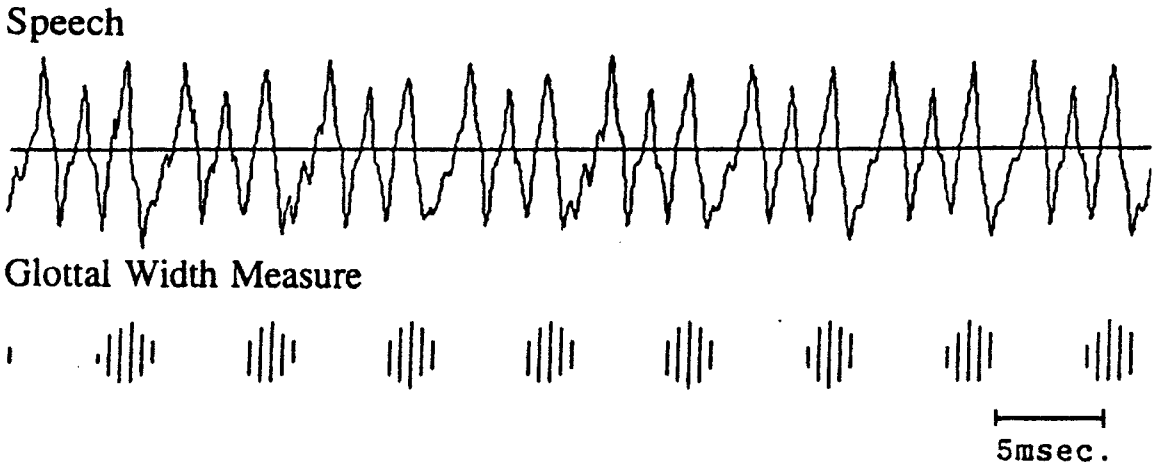


Abbildung 2.5.: Das obere und untere Ende der senkrechten Striche im unteren Bildteil markieren die Position der rechten und linken Stimmlippe in der zeitlichen Abfolge. Die Verschlusszeit und die symmetrische Schwingung ist zu erkennen. (Kiritani 1993 [56]).

gungsweite ausgedruckt, aus denen die Länge und die Schwingungsweite bestimmt wurden. Die Aufnahmen von 41 Männern und 41 Frauen wurden auf diese Art analysiert. Der Quotient liegt bei Männern (0,31) etwas über dem von Frauen (0,26). Die Messungen an drei Patienten mit einer hyperfunktionellen Dysphonie ergeben deutlich kleinere, die Messungen an zwei Patienten mit hypofunktioneller Dysphonie deutlich größere Quotienten als für die Normalstimmen.

Hertegard und Gauffin untersuchten 1995 [37] eine Methode zur Berechnung der Glottisöffnungsfläche durch Inversfilterung des Flusses. Der Fluss wurde dabei mit einer Maske gemessen. Die maximale und, bei inkompletten Schluss, die minimale Öffnungsfläche wurden aus dem inversgefilterten Fluss berechnet und an stroboskopischen Bildern gemessen. Die Kalibrierung der Längenmessung erfolgte mit CT Aufnahmen bei Phonationen mit entsprechenden Grundfrequenzen. Die Autoren fanden signifikante Korrelationen zwischen den gemessenen und den berechneten Glottisflächen. Die Auftragung der gemessenen gegenüber den berechneten Flächen zeigte jedoch, dass die Flächenwerte der verschiedenen Methoden jeweils sehr unterschiedlich waren. Als Ersatz der Messung der Flächen mit optischen Methoden kann man diese Methode deshalb nicht bezeichnen. Die statistische Analyse der gemessenen Werte bei normaler, gepresster und behauchter Phonation ergab einen signifikant höheren minimalen Fluss der behauchten Phonation gegenüber der normalen Phonation. Dadurch wird die Annahme bestätigt, dass behauchte Phonation mit einer erhöhten minimalen Fläche und damit mit einem inkompletten Verschluss einhergeht.

Sundberg verglich 1995 [136] den glottalen Fluss (Inversfilterung mit Flussmaske) mit der Schwingungsweite der Stimmlippen, die aus Hochgeschwindigkeitsaufnahmen berechnet wurde. Er untersuchte verschiedene Phonationsmoden, unter anderem normal, gepresst und behaucht. Sowohl im glottalen Fluss als auch in den Hochgeschwindigkeits-

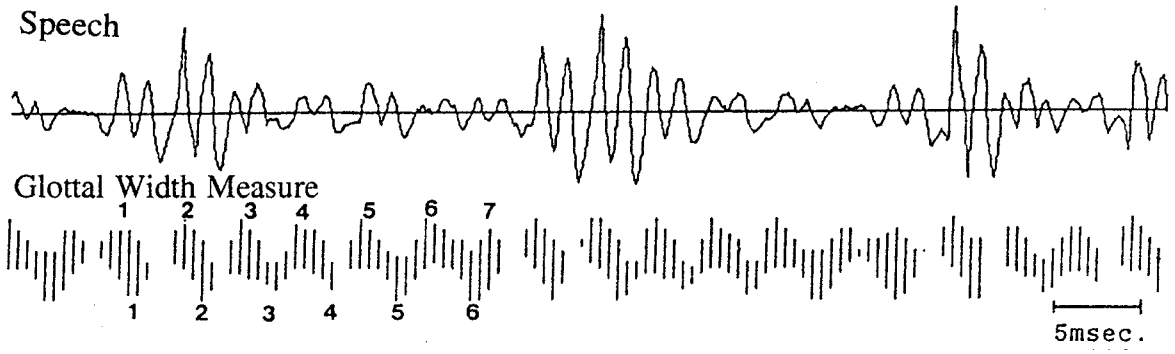


Abbildung 2.6.: Hier ist eine Stimmlippe gelähmt, die Stimmlippen schwingen nicht synchron: Nach sieben Perioden der oberen Stimmlippe hat das untere gerade sechs Schwingungen ausgeführt. Im akustischen Signal kann man dieses Schwingungsverhalten nicht erschließen (Kiritani 1993 [56]).

aufnahmen zeigte die behauchte Phonation einen inkompletten Glottisschluss.

Woo führte 1996 [154] quantitative Messungen an videostroboskopischen Aufnahmen der schwingenden Glottis durch. Dazu wurde jeweils in 20 aufeinander folgenden Bildern die Öffnungsfläche vermessen (durch manuelle Markierung eines Punktes innerhalb der Glottis und daraufolgender automatisierter Berechnung der Glottisränder mit der „Luminescence Shift“ Methode). 22 von 33 Frauen zeigten beim [e:] bei normaler Lautstärke keinen kompletten Glottisverschluss. Bei den Männern war dies bei 12 von 32 der Fall. Der Autor stellte weiterhin normatives Datenmaterial zur Charakterisierung der Flächen-Zeitfunktion zur Verfügung.

Eysholdt und andere setzten 1996 [22] eine digitale Hochgeschwindigkeitskamera ein, die bei einer Auflösung von 128 mal 16 Pixel 5590 Bilder pro Sekunde aufnehmen kann. Aufnahmen mit einer Länge bis zu einer Sekunde sind möglich. Die Autoren entwickelten eine Software, die automatisiert die Schwingungsweite der linken und der rechten Stimmlippe als Zeitfunktion aus dem Bild berechnet. Sie konnten so beispielhaft bei einer normalen Stimme den präphonatorischen Schluss zeigen. Bei einem Patienten mit einem Polypen erfolgte kein Glottisschluss und die Schwingungsamplitude der Stimmlippe mit dem Polypen betrug ca. nur ein Drittel der Amplitude der gesunden Stimmlippe. Bei einem dritten Patienten mit einem Kontaktgranulom war deutlich zu erkennen, dass bei beiden Stimmlippen schon vor dem Abschluss der Adduktion die Stimmlippenschwingung einsetzte.

Mergell und andere berechneten 1998 [77] den Verlauf der Schwingungsamplitude beim Einsatz der Schwingung für ein Zwei-Massen-Modell der Glottis. Dabei wurde bei dem dynamischen System die Hopf-Bifurkation untersucht, die durch Veränderungen von Systemparametern den Übergang von einer gedämpften Schwingung zu einer stationären Schwingung charakterisiert. Die so gewonnene Modellkurve kann durch die Messung von nur zwei Punkten der Einhüllenden der Stimmlippenschwingung an ein reales System angepasst werden. Aus dieser angepassten Modellkurve lassen sich Rückschlüsse auf die Parameter des realen Systems ziehen. Die Einhüllenden der Stimmlippenschwingung

wurden aus Hochgeschwindigkeitsaufnahmen der Glottis berechnet.

Kobler und andere verbesserten 1998 [59] das Verfahren von Hertegart und Gauffin [37]: 1) Sie modifizierten die Maske zur Flussmessung derart, dass das Endoskop luftdicht durch die Maske geführt werden konnte und verminderten so den Messfehler. 2) Veränderungen der Position des Endoskopes wurden automatisch verfolgt und die Videobilder entsprechend kalibriert. 3) Optische Verzerrungen wurden aus den Bildern herausgerechnet. 4) akustische Aufnahmen wurden verbessert, indem das Lichtleiterkabel verlängert wurde und die gesamte Untersuchung in einem schallgeschützten Raum durchgeführt wurde.

Sie konnten so die Übereinstimmung von optisch und aerodynamisch gemessenen Glottisöffnungsflächen wesentlich verbessern.

Die beschriebenen Untersuchungen zeigen das Forschungspotential auf, das hinter der (digitalen) Bildverarbeitung von Hochgeschwindigkeitsaufnahmen steckt. Bei der digitalen Bildverarbeitung von Hochgeschwindigkeitsaufnahmen werden jedoch große Finanz-, Rechen- und Speicherplatzkapazitäten benötigt, die den Rahmen einer phoniatrischen Klinik im Allgemeinen sprengen. In absehbarer Zeit wird sich dieses Verfahren aber etablieren, wenn die Leistungsfähigkeit von Computern und Zubehör in bisher gewohnter Weise anwachsen. Der Nachteil bei dieser Methode ist, dass es sich um eine semi-invasive Methode handelt, die keine ungestörte Artikulation zulässt.

Die beschriebenen Untersuchungen belegen zum Teil die plausiblen und dennoch vieldiskutierten Zusammenhänge von perzeptiver Behauchung, dem Glottisrestspalt und dem Gleichanteil des Volumenflusses an der Glottis. Dieser Zusammenhang wird im Rahmen der Korrelationen zwischen perzeptiver Behauchung und dem im Signal gemessenen Rauschanteil in Kapitel 14 noch von Bedeutung sein.

2.5. Akustische Stimmanalyse

Die akustische Stimmanalyse verarbeitet den Stimmschall, wie er von akustischen Wandlern, d.h. Mikrofonen, aufgenommen wird. Historisch stehen die analogen Methoden am Anfang der akustischen Stimmanalyse. Das bekannteste damalige Gerät zur Stimmanalyse besteht aus einer Vielzahl von analogen Bandpassfiltern, deren Mittenfrequenzen linear auf der Frequenzachse verteilt sind. Die Intensität des Eingangssignals in den entsprechenden Frequenzbändern wurde entweder durch verschieden starke Schwärzung auf Papier gebracht oder in Echtzeit auf eine nachleuchtende, sich drehende Bildtrommel geschrieben. Diese sogenannten Sonagramme waren lange Zeit ein Standardwerkzeug für die Stimmforscher. An ihnen kann man die Grundfrequenz und ihre Harmonischen, die Formanten sowie den Rauschanteil zwischen den Harmonischen ablesen und dadurch mit einiger Übung auf die Stimmqualität zurückschließen. Quantitative Untersuchungen des so gewonnenen Datenmaterials waren mit diesen Analysemethoden kaum möglich.

2.6. Akustische Stimmanalyse mit dem Computer

Eine neue Ära der Stimm- und Sprachforschung im Allgemeinen und der Stimmanalyse im Speziellen wurde durch die Entwicklung von leistungsfähigen Digitalrechnern und Analog-Digital-Wandlern eingeläutet. Zusammen mit der Möglichkeit, Stimmdaten in digitaler Form abzulegen, entwickelte sich sprunghaft der Zweig der diskreten Signalverarbeitung. Wichtige Stichworte zu diesem Forschungsgebiet sind: Diskrete Fourier-Transformation, Linear Predictive Coding (LPC) und digitale Filter. Die Methoden der akustischen Stimmanalyse werden seitdem so gut wie ausschließlich auf dem Computer realisiert.

3. Methoden der akustischen Stimmanalyse mit dem Computer

Es folgt ein kurzer Überblick über Arbeiten und Methoden der Sprachverarbeitung mit dem Computer, die für die Stimmanalyse relevant sind.

3.1. Bestimmung der Periodenlängen

Die Grundperiode und deren Kehrwert, die Grundfrequenz sind die Grundlage vieler abgeleiteter Größen zur Erfassung der Stimmqualität. Im Folgenden wird von Grundperioden gesprochen, wenn es um Methoden geht, die im Zeitbereich arbeiten, und von Grundfrequenz, wenn die Methoden im Frequenzbereich operieren. Es wurden und werden zahlreiche Arbeiten zu dem Thema Grundfrequenz und deren Bestimmung aus dem Zeitsignal veröffentlicht [2, 8, 9, 39, 48, 60, 90, 95, 97, 101, 121, 128, 142, 145]. Hier wird nur an gegebener Stelle auf Arbeiten bezug genommen, die mit in dieser Arbeit verwendeten Algorithmen in Verbindung stehen.

3.1.1. Definition von Periodizität

Die physiologische Definition von Grundfrequenz und Grundperiode geht direkt auf das Schwingungsverhalten der Stimmlippen zurück. Bei streng periodischen Vorgängen kann im Grunde jeder Zeitpunkt des sich periodisch wiederholenden Schwingungsmusters benutzt werden, um die Periodenlänge als die zeitliche Differenz des Wiedereintreten eines bestimmten Schwingungszustandes zu definieren.

Die Periodenlänge T ist so eigentlich nur für exakt periodische Vorgänge $s(t)$ definiert nämlich als

$$T = \mathrm{Min}\{T : (\forall t : -\infty < t < \infty) : s(t) = s(t+T); T > 0; \} \tag{3.1}$$

Gemessene Signale sind erstens endlich, so dass die Bedingung $-\infty < t < \infty$ auf ein endliches Intervall eingeschränkt werden muss. Zweitens wird aber gerade die exakte Gleichheit $s(t) = s(t+T)$ so gut wie nie erfüllt sein, so dass Gleichung 3.1 nicht direkt benutzt werden kann um die Periodenlänge zu bestimmen. Beispielsweise sind die tatsächlich auftretenden Stimmlippenschwingungen nicht streng periodisch. Die Periodenlänge schwankt von Schwingung zu Schwingung, da keiner der physikalischen Parameter, die das Schwingungsverhalten beeinflussen, konstant ist. Diese nichtkonstanten

Parameter sind etwa: Die Stimmlippenspannung, die Volumengeschwindigkeit der Luft beim Ausatmen und die Form des Mund- und Rachenraumes sowie die Lippenhaltung. Durch die Steuerung und die Schwankungen all dieser und weiterer Parameter entsteht gerade erst die Fülle der Ausdrucksmöglichkeiten in der Sprache.

Bei der Bestimmung der Periodenlängen gemessener, endlicher Signale können zwei Verfahren unterschieden werden: Einerseits Verfahren, die auf Zeitfenstern arbeiten, die mehrere (ca. drei bis zu 500 oder mehr) Perioden enthalten und die für jedes *Fenster* einen Wert der Periodenlänge liefern, andererseits Verfahren, die die Länge *jeder Periode* ermitteln. Die Grenze zwischen den Methoden ist jedoch nicht scharf. Die Methoden der ersten Klasse kann man zu solchen der zweiten Klasse machen, indem man keinen konstanten Fenstervorschub wählt, sondern jeweils den Wert der zuletzt gefundenen Periodenlänge.

3.1.2. Fensterweise Mittelung über mehrere Perioden

Autokorrelationsfunktion

Die Autokorrelationsfunktion wurde schon 1962 zur Grundperiodenbestimmung benutzt. Die theoretischen Grundlagen der Kurzzeitautokorrelation sind von Schroeder und Atal ebenfalls 1962 [122] untersucht worden.

Bei einem diskreten periodischen Signal $s(t); t = 1, ..., N$ der Periode T nimmt die diskrete Autokorrelationsfunktion

$$a(\tau) = \sum_{t=1}^{N-\tau} s(t)s(t+\tau) \tag{3.2}$$

nach einer Periode $\tau = T$ ein relatives Maximum an. Wenn von einem Signal bekannt ist, dass es annähernd periodisch ist, und wenn man den Bereich der Periodenlängen kennt, so braucht man nur das Maximum der diskreten Autokorrelationsfunktion in dem betreffenden Bereich zu suchen und hat damit ein Maß für die Periodenlänge in dem Signal.

Gehaltene Vokale von stimmgesunden Sprechern sind annähernd periodisch und die in Frage kommenden Periodenlängen sind empirisch bekannt: sie liegen bei der Sprechstimme von ca. 3ms (hohe Frauenstimme) bis ca. 14ms (tiefe Männerstimme). Der Nachteil dieser Methode ist jedoch, dass man auch bei kurzen Signalabschnitten keine Aussage über den exakten Anfangs- und Endpunkt einzelner Perioden in diesem Signalabschnitt erhält. Es wird nur eine lokal gemittelte Periodenlänge berechnet, die als Ausgangspunkt für detailliertere Methoden benutzt werden kann. Deshalb ist die Autokorrelationsmethode bei gehaltenen Vokalen eine gute Methode, um einen ersten Anhaltspunkt für den Wert der Grundperiode zu bekommen.

Der Suchbereich für die Periodenlänge muss an die Aufgabenstellung angepasst werden. Ist z.B. von dem Sprecher nur bekannt, dass es sich um einen Erwachsenen handelt (Mann oder Frau), der in Brusttonlage spricht, so sind Periodenlängen von 3,0ms bis 14ms zu erwarten. Eine Schwierigkeit ergibt sich nun daraus, dass der Suchbereich mehr als eine Oktave umfasst, denn wie viele andere Verfahren zur Grundperiodenbestimmung

ist auch die Autokorrelationsfunktion anfällig für Oktavfehler. Dies bedeutet, dass z.B. durch leichte Instationaritäten des Signals oder sogar durch spezielle Signaleigenschaften (Periodenverdopplung) die Autokorrelationsfunktion bei der halben oder bei der doppelten (wahrgenommenen) Grundfrequenz ein Maximum zeigt.

Kombination von inverser Filterung und Autokorrelation

Sondhi kombiniert die inverse Filterung (siehe unten) und die Autokorrelation [128]: Eine Kurzzeitautokorrelation des Fehlersignals führt zu sehr scharfen Spitzen bei der Grundperiode. Diese Methode wird auf überlappende kleine Zeitbereiche angewandt, um so eine lokale Grundperiode zu finden. Das Verfahren kombiniert aber leider auch die Nachteile der beiden Methoden.

Cepstrum

Das Cepstrum wurde von Noll und Schroeder [96–98] zur Grundperiodenbestimmung vorgeschlagen. Als Cepstrum $c(t)$ eines Signales $s(t)$ bezeichnet man die Fourierrücktransformierte $\mathcal{F}^{-1}$ des logarithmierten Leistungsspektrums des Signals:

$$c(t) = \mathcal{F}^{-1}\left\{\log\left|\mathcal{F}\left\{s(t)\right\}\right|^2\right\} \tag{3.3}$$

Die Grundfrequenz und die Harmonischen geben dem logarithmierten Leistungsspektrum eine periodische Gestalt, die sich in einem scharfen Maximum des Cepstrums bei der Grundfrequenz widerspiegelt. Der Zeitpunkt dieser Spitzen ist hier ein Maß für die Grundperiode. Die Methode ist mit der Autokorrelation eng verwandt und zeigt auch stellenweise Oktavfehler.

3.1.3. Bestimmung einzelner Periodenlängen

Zur Bestimmung der Periodenlänge von jedem einzelnen Schwingungszyklus muss man zunächst definieren, was unter der Periodenlänge bei realen, endlichen Signalen verstanden werden soll. Da man bei gemessenen Signalen nicht mit der exakten Wiederholung ganzer Zeitabschnitte rechnen kann, reduziert man die Definition der Periodenlänge oft auf das Wiedereintreten ausgezeichneter Schwingungszustände. Sei das Signal $s(t)$ nun endlich ($0 \leq t \leq t_{\max}$), so kann man diese ausgezeichneten Schwingungszustände t_a anordnen und nummerieren: $t_a[i] < t_a[i+1]; i = 1, ..., i_{\max}$. Die Periodenlänge des i-ten Schwingungszyklus $T[i]$ kann dann als

$$T[i] = t_a[i+1] - t_a[i]; i = 1, ..., i_{\max} - 1 \tag{3.4}$$

definiert werden. Damit haben wir die Schwierigkeit der Periodenlängenbestimmung auf die Definition und das Auffinden der ausgezeichneter Schwingungszustände t_a verlagert. Die Bestimmung der Periodenlänge anhand von ausgezeichneten Schwingungszuständen wird auch als ereignisbasierte (event based) Methode bezeichnet [141].

Peakpicking, Zerocrossing

Die einfachsten ausgezeichneten Schwingungszustände, die häufig zur Periodenlängenbestimmung benutzt werden, sind der Durchgang des Signals durch die Nulllinie (zerocrossing) und die Bestimmung des (positiven oder negativen) Maximums in einem Schwingungszyklus (peakpicking).

Je stärker das betrachtete Signal jedoch von einem periodischen Signal abweicht, umso größer werden die Schwierigkeiten, die man beim Auffinden der ausgezeichneten Schwingungszustände bekommt. Ein extremes Beispiel möge dies verdeutlichen: Bei selbstähnlichen Signalen (Fraktalen) [123] tritt jeder ausgezeichnete Signalzustand auf jeder Längenskala auf, so dass ohne zusätzliche Einschränkungen die Definition 3.4 nicht angewendet werden kann. Dass es sich hier nicht um ein rein akademisches Problem handelt wird schon daran deutlich, dass fraktale Methoden zur Kompression von Sprachdaten mit Erfolg angewendet werden [115]. Selbstähnlichkeit tritt also auch bei Sprachsignalen auf.

Die Kunst in der Anwendung von Definition 3.4 besteht deshalb darin, die möglichen Zeitpunkte der ausgezeichneten Ereignisse t_a von vornherein einzuschränken. Dazu gibt es prinzipiell zwei Möglichkeiten: Die eine besteht darin das, Signal so vorzubearbeiten (filtern), dass die gesuchten Schwingungszustände eindeutig werden. Die andere Möglichkeit ist, aufgrund von Vorwissen über das Signal den Bereich einzuschränken, in dem nach den Ereignissen gesucht wird. Häufig werden auch beide Möglichkeiten kombiniert angewandt.

Beispiele für die erste Möglichkeit sind: 1) Tiefpassfilterung mit Grenzfrequenzen knapp oberhalb der erwarteten Grundfrequenz. Die Filterung führt dazu, dass das Signal nur noch wenige, bestenfalls nur noch zwei Nulldurchgänge pro Periode besitzt, die sich durch die Richtung des Nulldurchgangs eindeutig identifizieren lassen. Probleme sind hierbei erstens, dass die Grenzfrequenz abhängig vom zu analysierenden Signalstück gewählt werden muss, und zweitens, dass durch die Filterung die so gefundenen Periodenlängen von denen des Originalsignals abweichen, da das Signal „verschmiert“ wird. 2) Festlegung eines Schwellenwertes beim Peakpicking. Hier muss die Schwelle z.B. in Abhängigkeit von der Kurzzeitenergie des analysierten Signalstücks laufend angepasst werden.

Ein Beispiel für die Einschränkung des Suchbereiches ist die Bestimmung der mittleren Periodenlängen T_M in einem mehrere Schwingungszyklen enthaltenden Segment mit einer der oben beschriebenen Methoden (Autokorrelation, Cepstrum oder andere). Ist die mittlere Periodenlänge bekannt, so kann man ausgehend von einem geeignet gewählten Anfangsereignis $t_a[1]$ den jeweiligen Suchbereich auf ein Gebiet der Breite $2aT_M$ einschränken:

$$t_a[i+1] \in \{t : t_a[i] + (1-a)T_M \leq t \leq t_a[i] + (1+a)T_M; 0 < a < 1\}. \tag{3.5}$$

Der freie Parameter a wird dem Signal entsprechend gewählt. Bei gehaltener Phonation (d.h. stimmhafte Phonation bei gleichbleibender Tonhöhe) ist oft die Wahl von $a = 0.49$ sinnvoll, da so der Bereich groß genug ist, damit die Periodenlänge auch bei

Schwankungen der Tonhöhe in dem Suchintervall bleibt, und klein genug, um Oktavfehler auszuschließen.

Ist die Tonhöhe nicht konstant, so ist es sinnvoll, bei festem a den Suchbereich von Periode zu Periode anzupassen. Dies geschieht, indem man für $t_a[i+1]$

$$T_M = t_a[i] - t_a[i-1] \tag{3.6}$$

setzt, also T_M entsprechend der zuletzt gefundenen Periodenlänge anpasst.

Inverse Filterung

Als wiederkehrender Schwingungszustand bietet sich besonders der Verschlusszeitpunkt der Stimmlippen an. Denn beim Schließen der Glottis entsteht der Knick in der Glottisöffnungsflächenfunktion und damit im Volumenstrom, der dazu führt, dass im Verschlussmoment der Vokaltrakt bei seinen Resonanzstellen (im Wesentlichen im Frequenzbereich von 0-5 kHz) zu akustischen Schwingungen angeregt wird. Ohne die Theorie der linearen Prädiktion (Anwendung der linearen Prädiktion zur Sprachkodierung: [4–6], Übersicht: [74]) hier aufzuführen, sei hier eine Konsequenz der Theorie erwähnt, nämlich, dass zum Zeitpunkt des Glottisverschlusses das Frequenzspektrum von einem linearen Modell M-ter Ordnung mit Koeffizienten a_i

$$s(t) + e(t) = \sum_{i=1}^{M} a_i s(t-i) \tag{3.7}$$

nur sehr schlecht vorausgesagt werden kann, so dass der Voraussagefehler $e(t)$ zur Bestimmung dieses Zeitpunktes herangezogen werden kann (siehe auch Strube 1974 [132]). Bei der inversen Filterung werden die Koeffizienten a_i dazu benutzt, um aus dem Signal das Fehlersignal $e(t)$ zu berechnen. Das Fehlersignal hat ein annähernd glattes Spektrum. Die Resonanzstellen des Vokaltraktes sind herausgefiltert. Im Fehlersignal treten die Verschlusszeitpunkte als deutliche Spitzen hervor, wie weiter hinten in Abb. 5.11 zu sehen ist.

Der Nachteil dieser Methode ist, dass die inverse Filterung nur dann zufriedenstellend arbeitet, wenn die Sprachsignale kein Rauschen im hochfrequenten Bereich beinhalten, in dem keine Stimminformation mehr enthalten ist. Das bedeutet, dass man sich bei der Abtastfrequenz auf ca. 10 kHz beschränken muss, denn das Anheben der Frequenzen größer als 5 kHz führt zu einem Fehlersignal, bei dem die Spitzen, die vom Glottisverschluss herrühren, nicht mehr zu erkennen sind. Da man also die Abtastfrequenz auf 10 kHz herabsetzt, ist die zeitliche Auflösung dieses Verfahrens vergleichsweise gering.

Oversampling

Durch Einfügen von Nullen zwischen den Abtastwerten und digitale Tiefpassfilterung steigert Hess in [38] die Abtastrate von 16 kHz auf 128 kHz. Er bestimmt die Periodenlängen aus den Abständen von zwei Punkten des Elektroglottogramms, an denen die Steigung maximal ist (Abb. 2.2). Er zeigt, dass die Abweichung der bei 128 kHz

gefundenen Perioden von den bei 16 kHz gefundenen normalverteilt, also nicht systematisch, ist und schließt deshalb auf die Berechtigung der Methode.

Ähnlichkeitsmodell: Waveform-Matching

Neben der ereignisbasierten Periodenlängenbestimmung hat sich in letzten Jahren mehr und mehr ein sogenanntes integrales Verfahren durchgesetzt, das sogenannte Waveform-Matching [76, 89]. Diese Verfahren nutzt die gesammte Information von zwei aufeinanderfolgenden Schwingungszyklen, um die Periodenlänge zu berechnen.

Die Berechnung der Periodenlängen anhand des Waveform-Matching- Verfahren mit parabolischer Interpolation soll kurz skizziert werden: Ein zeitdiskretes Signal $s(t); t \in \mathbb{Z}$ werde untersucht. Der Startpunkt der Analyse liege bei $t = t_0$. Gesucht wird eine lokale Periodenlänge T (im Kontrast zu einer globaleren Periodenlänge, die über viele Perioden mittelt). Außerdem sei aufgrund plausibler Annahmen der Bereich der möglichen Periodenlängen auf $T_{\min} \leq T \leq T_{\max}$ eingeschränkt. Weiterhin seien

$$x(t_0, \tau, t) = s(t); t_0 \leq t < t_0 + \tau \tag{3.8}$$

und

$$y(t_0, \tau, t) = s(t); t_0 + \tau \leq t < t_0 + 2\tau \tag{3.9}$$

Signalabschnitte, die bei t_0 bzw. $t_0 + \tau$ beginnen und jeweils die Länge τ haben. Diese gleich langen Abschnitte x und y des Signales $s(t)$ seien als τ-dimensionale Vektoren aufgefasst. Dann wird die Periodenlänge T als

$$T = \underset{T_{\min} \leq \tau \leq T_{\max}}{\operatorname{argmax}} \left(\frac{xy}{|x||y|}\right) \tag{3.10}$$

definiert. Durch eine parabolische Interpolation des Maximums zur Bestimmung von T kann die Genauigkeit der Periodenlängen weit über die Abtastperiode hinaus gesteigert werden. Dies ist in Abbildung 3.1 daran zu erkennen, dass die so definierten Periodenlängen (mittlere Kurve) konsistent mit denen sind, die durch Peakpicking (obere Kurve), also durch einfache Maximalwertbestimmung und parabolischer Interpolation in jeder Periode des Zeitsignals, bestimmt wurden. Der Maximale Korrelationswert in Gleichung 3.10 gestattet neben der Messung der Periodenlänge auch eine Aussage über die Ähnlichkeit aufeinanderfolgender Perioden. Der maximale Korrelationswert umso kleiner, je unähnlicher die aufeinanderfolgenden Perioden sind. Dieser Wert wird hier als ein weiteres wichtiges Stimmgütemaß definiert und mit Periodenkorrelationwert oder Mean Waveform-Matching Coefficient (MWC) bezeichnet.

Der große Nachteil dieses Verfahrens liegt darin, dass die Lage von Beginn und Ende des Schwingungszyklus nicht stabil ist, wenn man es auf sehr viele Schwingungszyklen anwendet. Die Lage der Grenzen innerhalb der Schwingungszyklen (d.h. die Phasenlage der Grenzen) können dabei über den kompletten Zyklus variieren. Dies wurde besonders bei eigenen Versuchen mit Waveform-Matching am EGG festgestellt, die im Kapitel 8

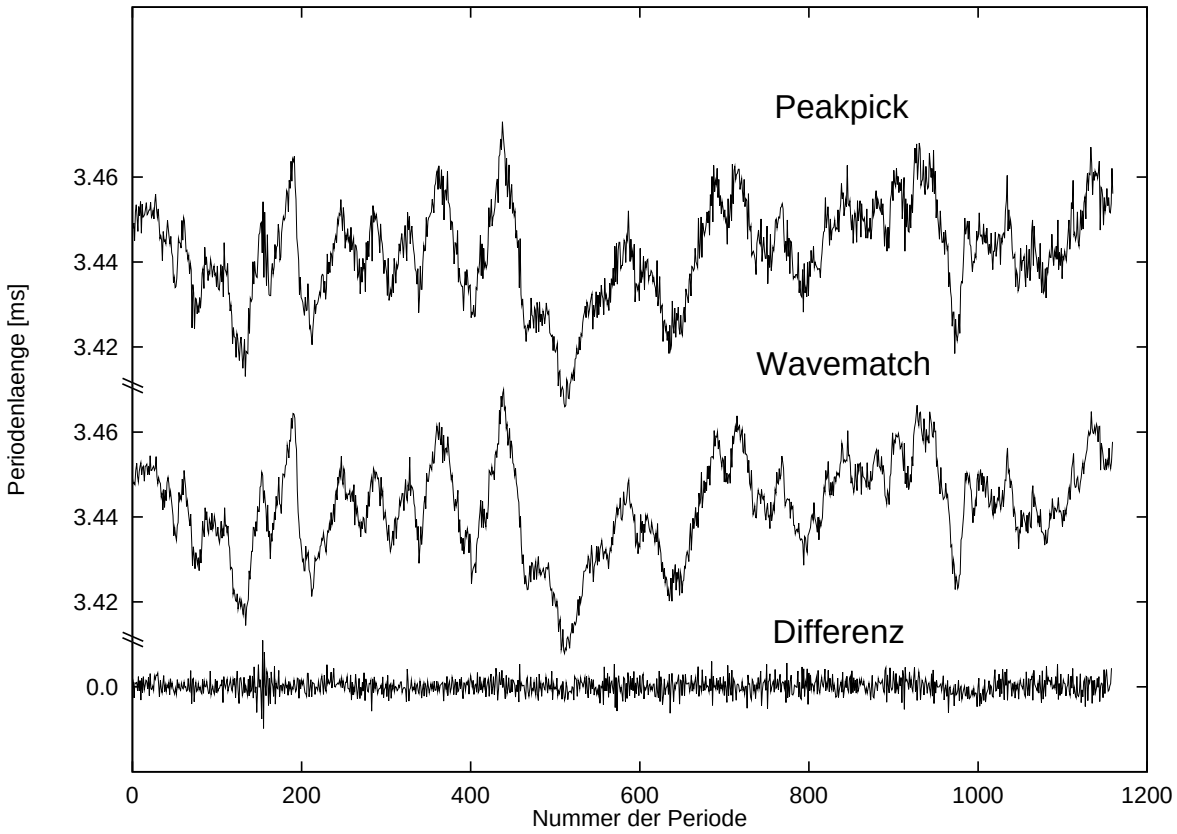


Abbildung 3.1.: Veranschaulichung der Genauigkeit und der Übereinstimmung des Waveform-Matching und des Peakpicking. Abtastperiode: 0,02ms.

beschrieben werden. Hier bietet sich (bei hinreichend gutmütigen Signalen) eine mehrstufige Methode an: 1) Berechnung der Periodengrenzen mit einer ereignisbasierten Methode. 2) Ausgehend von den Periodengrenzen Berechnung der Periodenlänge mit dem Waveform-Matching. Dabei kann Schritt 1) zum Beispiel relativ robust am differenzierten EGG durchgeführt werden, und Schritt 2) dann am Mikrofonsignal. Dabei muss man jedoch ggf. den Laufzeitunterschied zwischen EGG und Mikrofonsignal berücksichtigen. Dieses mehrstufige Verfahren wird im Kapitel 8 angewendet.

Mit den beschriebenen Methoden lassen sich jeweils für gehaltene Vokale Sequenzen von Periodenlängen ermitteln. Aus diesen Periodenlängensequenzen werden nun Größen abgeleitet, die über das statistische Verhalten der Periodenlängen Auskunft geben. Für die Stimmanalyse ist es von Interesse, die Zusammenhänge zwischen der Stimmqualität und diesen statistischen Größen zu finden.

3.2. Akustische Messgrößen zur Quantifizierung der Unregelmäßigkeit der Stimme

Bei den oszillierenden Stimmlippen handelt es sich um einen Schwingungsprozess, an dem lebendiges Gewebe beteiligt ist. Viele physiologische Parameter wie z.B. die Anspannung der Kehlkopfmuskulatur beeinflussen wichtige Kenngrößen (Amplitude, Periodenlänge) des Schwingungsvorgangs. Diese physiologischen Parameter sind im Allgemeinen aber zeitlich nicht konstant. Beispielsweise ergibt sich der Muskeltonus aus der Summe der Spannungen der momentan kontrahierten Muskelfasern. Die einzelnen Muskelfasern werden jedoch in einem stochastischen Prozess durch ihre zugehörigen Neurone stets so zur Kontraktion angeregt, dass sich ein gewünschter mittlerer Spannungszustand einstellt. Da es sich um eine Überlagerung vieler, aber eben nur endlich vieler, Einzelprozesse handelt stellt sich eine statistische Schwankung des Muskeltonus ein. Diese Schwankung überträgt sich auf die charakteristischen Kenngrößen wie Amplitude und Periodenlänge so dass sich zwei aufeinanderfolgende Schwingungszyklen weder in der Amplitude noch in der Periodenlänge exakt gleichen. Wenn diese Schwankungen fehlen, wie es bei manchen Sprachsynthesizern der Fall ist, so klingt die Sprache hart und unnatürlich. Andererseits treten bei normalen Stimmen nur relativ kleine Schwankungen auf. Die Güte einer Stimme hängt unter anderem mit dem Ausmaß der Perioden-, Amplituden- und Formschwankungen des Zeitsignals von Periode zu Periode zusammen. Um diese Schwankungen zu quantifizieren, werden Maßzahlen für den Grad der Schwankungen berechnet. Im Folgenden werden Verfahren zur Erfassung und Quantifizierung dieser Schwankungen beschrieben. Zur Beschreibung der Periodenlängenschwankungen hat sich der Terminus Jitter und zur Beschreibung der Amplitudenschwankung der Terminus Shimmer eingebürgert.

Folgende Arbeiten beschäftigen sich mit verschiedenen Aspekten der Messung von Jitter und Shimmer: [51, 71, 72, 104, 107, 116, 118–120, 138, 140, 144, 148, 150, 151].

3.2.1. Jitter und Shimmer

Es gibt keine Definition für den Begriff Jitter, die vorschreibt, wie dieser ermittelt wird. Ein erster Anhaltspunkt für eine Definition ist etwa: die Breite der Häufigkeitsverteilung der Differenzen von je zwei aufeinanderfolgenden Periodenlängen. Der Jitter hängt von der Art der Grundperiodenbestimmung ab, wie Titze gezeigt hat [141]. Weiterhin wird der Jitter entweder auf die Periodenlänge bezogen und dann meist in Prozent angegeben, oder als sog. absoluter Jitter berechnet [20].

Der Jitter wurde schon 1961 von Lieberman [71] in fließender Sprache untersucht. In einer weiteren Arbeit 1963 [72] untersucht er den Zusammenhang von Jitter und Stimmstörungen bei pathologischen Stimmen. Bemerkenswert ist die Methode, mit der Liebermann einige tausend Periodenlängen bestimmte: Oszillographenbilder des Stimmsignals wurden gefilmt und dann auf Mikrofilm gebracht. Unter einem Mikrofilmsichtgerät wurden dann mit einem Lineal die Periodenlängen einzeln von Amplitudenmaximum zu Amplitudenmaximum ausgemessen. Als Maß für die Stimmgüte benutzte

Liebermann die Zahl der Periodenlängenunterschiede, die größer als 0,5ms waren.

Als Shimmer werden die Schwankungen der Amplituden der einzelnen Grundperioden bezeichnet. Shimmer basiert deshalb ebenso wie Jitter auf dem Auffinden der einzelnen Grundperioden, in denen dann zum Beispiel jeweils die Energie berechnet oder das Maximum der Amplitude gesucht wird.

Da auch für diese Arbeit Methoden zur Periodenstatistik angewandt wurden, werden hier zunächst Arbeiten aus der Literatur, die sich mit Jitter und Shimmer beschäftigen, vorgestellt. Sie lassen sich in zwei Gruppen trennen: Die eine untersucht und entwickelt Methoden zur Bestimmung von Jitter und Shimmer, die andere wendet diese Methoden zur Stimmanalyse an. Neuere methodische Arbeiten: Schoentgen und de Guchteneere zur Bestimmung des Jitters aus dem akustischen Stimmsignal und dem Elektroglottogramm [118], Kroeger über den Einfluss der Vokaltrakt-Glottis-Kopplung auf Jitter und Shimmer [66], Titze und Winholtz über den Einfluss von Mikrofontyp und Mikrofonpositionierung [143, 150] sowie Titze über den Methodeneinfluss der Grundperiodenbestimmung [141].

Die letztgenannte Arbeit von Titze verwendet drei verschiedene Methoden, um die Grundperiode zu bestimmen: Bestimmung der Nulldurchgänge des tiefpassgefilterten Signales mit linearer Interpolation, Peakpicking des Periodenmaximums mit Interpolation durch eine Parabel, Waveform-Matching mit parabolischer Interpolation des Maximums (entspricht dem Ähnlichkeitsmodell von Medan et al. mit parabolischer Interpolation des Maximums des Skalarproduktes). Das Ergebnis der Arbeit ist, dass die Methoden zwar verschieden starke Werte für den Jitter liefern, das aber die Relationen zwischen dem Jitter verschiedener Stimmen bei den drei Methoden gleich bleiben. In der Praxis wählt man deshalb die robusteste Methode aus.

Arbeiten, die Jitter und Shimmer in klinischen Studien als Stimmgüteparameter einsetzen, sind: Kasuya et al. in [54] und [52], Banci et al. in [11], Laver et al. in [69], Peppard et al. in [102], Verstraete et al. in [148] und Plante et al. in [105]. Sowohl die Patientengruppen als auch die Zielsetzungen dieser Arbeiten sind unterschiedlich und sollen hier nicht näher besprochen werden.

3.2.2. Perturbationsmaße

Seit Liebermann 1961 das erste Maß für die Periodenschwankungen oder Periodenperturbationen eingeführt hat, sind unter verschiedenen Namen Perturbationsmaße veröffentlicht worden, die jeweils den Jitter der akustischen Signale messen sollen. Pinto und Titze haben 1990 in [104] eine Arbeit zur Vereinheitlichung von Perturbationsmaßen vorgestellt, in der die Perturbationsmaße aus der Literatur auf mathematische Begriffe zurückgeführt werden.

Im Allgemeinen wird mit den Perturbationsmaßen eine Abweichung der einzelnen Periodenlängen von einem lokalen Periodenlängenmittelwert gemessen und wiederum über diese lokale Abweichung der Perioden gemittelt. Diese Maße sind zum Beispiel bei

Kasuya et al. [51] aufgeführt. Der Perturbation Quotient (PQ) ist wie folgt definiert:

$$\mathrm{PQ} = \frac{1}{N-K} \sum_{n=\frac{K-1}{2}}^{N-\frac{K-1}{2}-1} \left| \frac{u(n) - \frac{1}{K}\sum_{k=-\frac{K-1}{2}}^{\frac{K-1}{2}} u(n+k)}{\frac{1}{K}\sum_{k=-\frac{K-1}{2}}^{\frac{K-1}{2}} u(n+k)} \right| \times 100\%. \tag{3.11}$$

Dabei ist K die Zahl der Perioden, über die gemittelt wird, wobei K ungerade ist, so dass es stets eine zentrale Periodenlänge gibt. N ist die Anzahl der Perioden, und $u(n)$ steht nicht nur für die Periodenlängen, sondern kann auch bei der Bestimmung des Shimmers die Amplitude oder die Energie der einzelnen Perioden bedeuten. Deshalb wird zwischen einem Pitch Perturbation Quotient (PPQ) und einem Amplitude (bzw. Energy) Perturbation Quotient APQ (EPQ) unterschieden.

Daneben wird der Perturbation Faktor (PF) (entsprechend PPF, APF und EPF) erwähnt:

$$\mathrm{PF} = \frac{1}{N-1} \sum_{n=1}^{N-1} \left| \frac{u(n) - u(n-1)}{u(n)} \right| \times 100\%. \tag{3.12}$$

Hier wird über lokale normierte Abweichungen von nur zwei Perioden gemittelt. Dieses Perturbationsmaß ist im Gegensatz zu PQ sensitiv für einen Anstieg oder Abfall der Grundfrequenz z.B. aufgrund der Satzmelodie. In dieser Arbeit werden Perturbationen mit PF und PQ, $K = \{3, 5, 7, 11, 15\}$ Perioden untersucht und folgende Abkürzungen verwendet:

Jittermaße

- Pitch Perturbation Factor, Gleichung 3.12, Abkürzung in dieser Arbeit: J2, Periodenlängenbestimmung mit dem Waveform-matching Verfahren und Interpolation des Maximums, Gleichung 3.10.
- Pitch Perturbation Quotient, Gleichung 3.11, Abkürzung J3 für $K = 3$, J5 für $K = 5$ usw., Periodenlängenbestimmung wie oben.

Shimmermaße

- Energy Perturbation Factor [3.12] (Abkürzung in dieser Arbeit: S2),
- Energy Perturbation Quotient [3.11] (Abkürzung S3 für $K = 3$, S5 für $K = 5$ usw).

3.2.3. Modelle des Jitters

1993 wurden von Kasuya et al. in [51] ein ARMA- Modell (autoregressive moving average) des Jitters und von Schoentgen und de Guchteneere in [116] ein autoregressives Modell (AR) benutzt, um Aussagen über die Statistik der Periodenlängensequenzen

zu erhalten. Kasuya et al. charakterisieren mit dem ARMA Modell das Spektrum der Periodensequenz und finden für gesunde und pathologische Sprecher verschiedene charakteristische Modellparameter.

Schoentgen und de Guchteneere wollen durch ihr Modell die systematischen Periodenlängenschwankungen eliminieren, um nur statistische Schwankungen zur Berechnung des Perturbationsmaßes heranzuziehen. Sie stellen fest, dass sowohl zwischen männlichen und weiblichen Sprechern als auch zwischen gesunden und kranken Stimmen die Ordnung des Modells verschieden hoch sein muß, um eine statistische Verteilung zu erreichen. Auch das Perturbationsmaß dieser statistischen Verteilung ist für die jeweiligen Gruppen unterschiedlich. Beide Arbeiten benutzen leider nur wenige Stimmproben um, ihre Methoden zu testen, so dass die Vorteile dieser Methoden bei der klinischen Anwendung der Stimmanalyse nur schwer einzuschätzen sind.

3.3. Maße für den turbulenten Rauschanteil

Neben der Ünregelmäßigkeit kann Rauschen als weitere Abweichung im Signal auftreten [28]. Im Folgenden werden Methoden besprochen, die diesen Rauschanteil messen.

Yanagihara

Die erste bekannte Arbeit zu diesem Thema von Yanagihara wurde 1967 [155] vorgestellt. Er beschreibt drei Faktoren, um den subjektiven optischen Eindruck von Sonagrammen zu klassifizieren: 1. Rauschkomponenten in den Hauptformanten der Vokale, 2. Hochfrequentes Rauschen über 3000 Hz und 3. der Abfall der harmonischen Komponenten zu höheren Frequenzen. Diese Art der Klassifizierung ist bis heute bei den Phoniatern gebräuchlich, aber nicht leicht durch Automatisierung zu objektivieren.

Harmonics-to-Noise Ratio (HNR)

Ein Ansatz, um die Heiserkeit mit einem Computer zu bestimmen, stammt von Yumoto [156]. Der Stimmparameter Harmonics-to-Noise Ratio (HNR) gibt die relative Stärke der harmonischen Signalenergie zur Energie des Rauschanteils an. Dazu werden $n = 50$ Perioden $f_i(\tau)$ der Periodendauern T_i gemittelt und zur mittleren Periode

$$f_A(\tau) = \sum_{i=1}^{n} \frac{f_i(\tau)}{n}; 0 <= \tau <= T \tag{3.13}$$

mit der Periodendauer $T = \max\{T_i; i = 1, ..., n\}$ zusammengefasst. Dabei wird $f_i(\tau) = 0$ gesetzt, wenn $T_i <= \tau <= T$ ist. Die Energie der mittleren Periode

$$H = n \int_0^T f_A^2(\tau) d\tau \tag{3.14}$$

steht für den harmonischen Signalanteil. Die Rauschenergie wird durch die Abweichung der einzelnen Perioden von der mittleren Periode definiert:

$$N = \sum_{i=1}^{n} \int_0^T [f_i(\tau) - f_A(\tau)]^2 \, d\tau \tag{3.15}$$

Der Heiserkeitsparameter HNR ist der Quotient H/N. Yumoto schreibt, dass sich der Jitter wegen der Annahme $f_i(\tau) = 0$ für $T_i <= \tau <= T$ auf dem Rauschwert niederschlägt, d.h. der Parameter HNR wird zu klein, wenn ein relativ starker Jitter vorliegt. Diese Methode wurde bei dem kommerziellen „Computer Speech Lab“, einem System zur Stimmanalyse, implementiert. Es stellte sich jedoch heraus, dass diese Methode bei manchen Stimmen HNR-Werte liefert, die im Widerspruch zur auditiven Einschätzung der Stimmqualität stehen.

Harmonische Intensität

Hiraoka [45] benutzt die auf die Gesamtintensität des Spektrums P bezogene Summe der Intensitäten der harmonischen Komponenten p_i ausschließlich der Grundfrequenz p_1

$$H_r = \left(\frac{\sum_{i \geq 2} p_i}{P} \right) 100(\%) \tag{3.16}$$

zur Analyse von normalen und heiseren Stimmen. Er findet einen kritischen Wert von 67,2%, unterhalb dessen sich nur noch heisere Stimmen finden. Wie beim HNR ist jedoch auch dieser Parameter vom Aufsuchen der Grundfrequenz und vom Jitter abhängig.

Spektrale Rekonstruktion

Klingholtz rekonstruiert in [58] den harmonischen Anteil des Spektrums aus gaußförmigen Komponenten, wobei die Bandbreite der Komponenten für jede Stimme angepasst wird. Die hierfür benötigte Grundfrequenz wird mit Hilfe des Produktspektrums [121] berechnet. Der Quotient aus der Energie des rekonstruierten harmonischen Anteils und dem restlichen Rauschanteil wird als Signal-to-Noise Ratio bezeichnet und dient als Heiserkeitsparameter. Die Grenzen der Einsetzbarkeit sind auch hier durch Stimmen gegeben, die keine harmonische Struktur im Spektrum zeigen.

Periodensynchrone und cepstrale Methoden

Gleiches gilt für die Methode von Muta und Baer [94]. Sie benutzen genau vier Perioden, um dann im Spektrum aus der Tiefe der Täler zwischen den Harmonischen auf den Rauschanteil zu schließen. Diese Methode basiert wiederum auf der Harmonizität der Stimmen und ist so in ihrem Einsatz begrenzt.

Normalized Noise Energy (NNE)

Um die Nachteile des HNR-Parameters auszugleichen, benutzt Kasuya 1986 [53] die Faltung mit einem adaptiven Kammfilter im Zeitbereich zur Mittelung der Perioden, wobei die Zinken des Kammfilters nicht äquidistant sind, sondern den Abstand der jeweiligen Periodenlänge haben. Außerdem werden der Start- und Endpunkt der Perioden in einem iterativen Verfahren so linear angepasst, dass die Perioden möglichst ähnlich werden. Der Energieunterschied von gefiltertem und ungefiltertem Signal bildet als Normalized-Noise-Energy (NNE) ein Maß für die Heiserkeit. Da die Periodengrenzen genau bestimmt werden müssen, ist die Anwendbarkeit dieser Methode auf Stimmen mit definierbarer Grundperiode beschränkt.

Der Stimmgütemesswert Normalized Noise Energy (NNE) wird 1986 ein weiteres mal definiert [55]. Hierbei wird der Quotient aus einem Schätzwert der Rauschenergie und der gesamten Signalenergie im Spektralbereich berechnet.

Unterteilt man das Spektrum eines (harmonischen) Signals in Spitzen und Täler, so ergibt sich der Schätzwert der Rauschenergie aus der Summe der spektralen Energie in den Tälern und dem geschätzten Rauschenergieanteil an den (harmonischen) Spitzen.

Der Rauschanteil an den Spitzen wird dabei einfach als Mittelwert der Energie der angrenzenden Täler angenommen.

Zur Analyse wird zunächst auf dem gesamten Vokal in 40ms–Fenstern bei 20ms Fenstervorschub jeweils das erste Nebenmaximum der Autokorrelationsfunktion bestimmt. Der Medianwert dieser Maxima legt die Periodenlänge fest. Nun werden Fenster analysiert, die genau 7 Perioden enthalten (M Abtastwerte). Diese Fenster werden mit einem Hamming–Fenster versehen und bei einer Abtastfrequenz von 10kHz auf 102,4ms (entsprechend N=1024 Abtastwerten) mit Nullen aufgefüllt. Nach einer diskreten Fourier–Transformation (DFT) wird nun im Spektrum die Gesamtenergie und die Rauschenergie berechnet. Die Breite der Spitzen wird dabei als Breite des Hamming-Fensters mit $2N/M$ angenommen.

Zur Beurteilung der Leistung des NNE wurden die NNE–Werte von 250 Stimmen, von denen 64 Normalstimmen und der Rest pathologische Stimmen waren, berechnet. Als Gütekriterium diente die Diskriminationsfähigkeit des NNE bei verschiedenen Frequenzbereichen: 0-1kHz, 0-3kHz, 0-5kHz, 1-3kHz, 1-5kHz. Die beste Diskriminationsgüte ergab sich für den Frerquenzbereich 1-5kHz (13 Fehler).

Cepstral Harmonic to Noise Ratio CHNR

Einen anderen Ansatz zur Bestimmung des Verhältnisses von harmonischer Energie zu der Rauschenergie wurde 1993 von de Krom [17] vorgestellt: Der Rauschanteil wird hier in zwei Stufen approximiert: Zuerst werden die so genannten Rahmonischen aus dem Cepstrum (hier als die Cosinustransformierte des logarithmierten Leistungsspektrums) entfernt. Dazu werden nach einer initialen Grundperiodenschätzung (wie beim NNE) die Spitzen im Cepstrum bei Vielfachen der Periodenlänge auf Null gesetzt, und zwar vom Spitzenwert ausgehend nach links und rechts, bis sich das Vorzeichen der Ableitung umkehrt. Nach Rücktransformation des gelifterten Cepstrums erhält man eine erste Schätzung des Rauschanteils. Auf der zweiten Stufe wird eine sogenannte Basislinienkorrektur durchgeführt, so dass bei keiner Frequenz das geschätzte Rauschen über dem Originalspektrum liegt. Aus der Differenz von diesem Rauschen und dem originalen logarithmierten Leistungsspektrum wird nun der Signal-Rauschabstand berechnet. CHNR wird wie auch NNE in verschiedenen Frequenzbereichen berechnet. In dieser Arbeit werden der im Spektrum berechnete NNE und CHNR zu Vergleichszwecken verwendet.

3.4. Computer Speech Lab (CSL), Multidimensional Voice Profile (MDVP)

In dieser Arbeit wurden unter anderem auch akustische Stimmgütemesswerte eines kommerziellen Systems zur Stimmanalyse herangezogen. Dieses System ist das Computer Speech Lab von Kay Elemetrics, ein System zur Aufnahme und Analyse von Stimmen an einem computergestützten Arbeitsplatz. Dieses System wird weltweit häufig eingesetzt. Die enthaltene Analysesoftware Multidimensional Voice Profile (MDVP) stellt beinahe einen de facto Standard zur Stimmanalyse dar. Mit MDVP können folgende Messgrößen für turbulentes Rauschen bei der Stimmanregung gemessen werden:

- Noise to Harmonic Ratio (NHR): Durchschnittliche Energie der unharmonischen Komponenten im Bereich von 1500-4500Hz bezogen auf die Energie der harmonischen Komponenten im Bereich von 70-4500Hz.
- Voice Turbulence Index (VTI): Durchschnittliche Energie der unharmonischen Komponenten im Bereich von 2800-5800Hz bezogen auf die Energie der harmonischen Komponenten im Bereich von 70-4500Hz.
- Soft Phonation Index (SPI): Durchschnittliche Energie der harmonischen Komponenten im Bereich von 70-1600Hz bezogen auf die Energie der harmonischen Komponenten im Bereich von 1600-4500Hz.

MDVP stellt folgende Jitter- und Shimmermaße, die auf ereignisbasierter Periodenlängenbestimmung und auf der Bestimmung der Maximalamplitude beruhen, zur Verfügung:

- Jitter (Jita, Jitter absulute): mittlere Differenz aufeinanderfolgender Periodenlängen. Geht auf Lieberman [71] zurück.
- Jitter Percent (Jitt): mittlere Differenz aufeinanderfolgender Periodenlängen bezogen auf die mittlere Periodenlänge.
- Relative Average Perturbation (RAP): entspricht PQ mit $K = 3$.
- Pitch Perturbation Quotient (PPQ): entspricht PQ mit $K = 5$.

Die Shimmermaße der MDVP Software sind:

- Shimmer in dB (ShdB): mittlere Differenz aufeinanderfolgender Amplituden.
- Shimmer Percent (Shim): mittlere Differenz aufeinanderfolgender Amplituden bezogen auf die mittlere Amplitude.
- Amplitude Perturbation Quotient (APQ): entspricht PQ mit $K = 11$.

4. Datenmaterial

In dieser Arbeit werden ausschließlich natürliche oder synthetische Vokale untersucht, da die hier angewandten akustischen Methoden zunächst nur auf gehaltene Phonation ausgelegt ist. Unter gehaltener Phonation soll hier die Phonation eines Vokales verstanden werden, bei dem sich die Grundfrequenz und die Lautstärke nicht oder nur langsam und wenig ändern.

4.1. Synthetische Signale

Da hier neue Methoden zur Stimmgütebeurteilung entwickelt und mit bisherigen Methoden verglichen werden sollen, ist es nötig Testdaten zu generieren, bei denen sich bestimmte Eigenschaften (Jitter, Shimmer, Signal-Rauschverhältnis, Anregungsfunktion, Spektrum) exakt vorgeben lassen.

4.1.1. Rosenberg-Glottispuls

Rosenberg stellte 1970 [110] ein einfaches Modell der Anregungsfunktion an der Glottis vor. Die Glottisfläche in Abhängigkeit von der Zeit $g_R(t)$ wird wie folgt in drei Zeitabschnitten modelliert:

$$g_R(t) = a \begin{cases} \frac{1}{2}(1 - \cos\frac{\pi t}{T_O}) & : \quad 0 \leq t < T_O \\ \cos\frac{\pi(t-T_O)}{2T_C} & : \quad T_O \leq t < T_O + T_C \\ 0 & : \quad T_O + T_C \leq t < T \end{cases} \tag{4.1}$$

Dabei beschreibt a die Amplitude, T_O die Öffnungszeit (die Zeit, die bis zur maximalen Öffnung vergeht) und entsprechend T_C die Verschlusszeit. Wichtig ist, dass die Ableitung an der Stelle $T_O + T_C$ einen Sprung besitzt. Dadurch wird der rapide Glottisschluss modelliert, der zur Anregung der höheren Harmonischen führt.

4.1.2. Resonanzfilter

Als erste Annäherung an das Formantspektrum von Vokalen dient im Folgenden ein Resonanzfilter zweiter Ordnung, das einen Formanten simuliert. Die Übertragungsfunktion in der z-Ebene lautet:

$$H(z) = \frac{r\sin(\gamma)z}{z^2 - 2r\cos(\gamma)z + r^2}\,. \tag{4.2}$$

Das Filter ist durch r und γ vollständig bestimmt. Die genaue Lage der Resonanzfrequenz sowie die Bandbreite des Filters wird im Anhang A beschrieben (diese entsprechen bei Resonanzfrequenzen, die klein gegen die Abtastfrequenz sind, *nicht exakt* dem Winkel γ und dem Abstand des Poles vom Einheitskreis $1 - r$).

4.1.3. Sprachsynthetisator „Speech Maker“

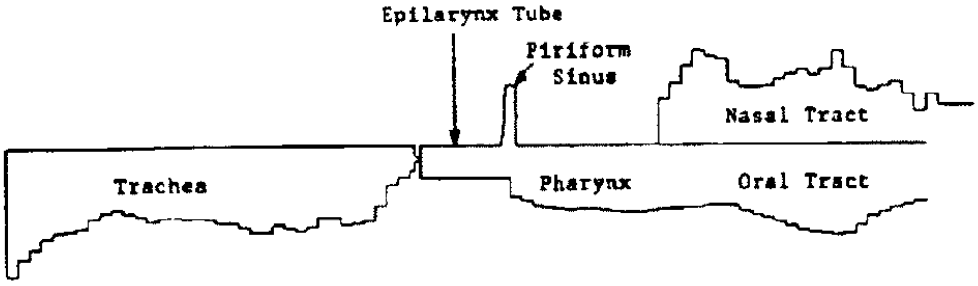


Abbildung 4.1.: Aufbau des „Speech Makers“

Der Sprachsynthetisator „Speech Maker“ wurde von Story [130] an der Universität Iowa entwickelt und freundlich für diese Arbeit zur Verfügung gestellt. Der Synthetisator basiert auf zeitlich veränderlichen Wellendigitalfiltern von Strube [134] (Eine gute Zusammenfassung findet in [124]). Das verwendete Röhrenmodell wurde unter anderem von Ishizaka und Flanagan [50], Sondhi und Schroeter [129] und Liljencrants [73] entwickelt. Der „Speech Maker“ zeichnet sich gegenüber seinen Vorgängern durch folgende Merkmale aus:

- Hohe Abtastfrequenz von 44100Hz
- „moderne“ Querschnittsfunktionen aus Kernspintomografie (siehe auch Story [131])
- Modellierung subglottaler Segmente (Trachea)
- Modellierung der piriformen Sinus
- Energieaufnahme und Abstrahlung über die „Haut“ (dadurch ist ein kompletter Lippenschluss möglich)

Weiterhin stellt der „Speech Maker“ verschiedene Glottismodelle zur Verfügung (Einmassenmodell, Zweimassenmodell, Parametrische Modelle). Diese wurden jedoch in dieser Arbeit nicht benutzt, die Anregungsfunktion wurde außerhalb des „Speech Maker“ generiert.

4.2. Stimmaufnahmen

Die Stimmaufnahmen für diese Arbeit wurden in der Abteilung Phoniatrie und Pädaudiologie der Universitätsklinik Göttingen aufgenommen. Dazu stand ein schallgedämmter Raum und ein reflexionsfreier Raum zur Verfügung. In der Tabelle 4.1 sind die Gruppen zusammengefasst, die im folgenden unter verschiedenen Fragestellungen analysiert werden. Der Eintrag „normal“ in der Spalte „Grundfrequenz“ steht für eine bequeme Tonhöhe (comfortable pitch). Die Gruppen eins bis sieben und zehn wurden mit dem Computer Speech Lab (CSL) und einem Standmikrofon bei ca. 30cm Mikrofonabstand aufgenommen. Die Gruppen acht und neun wurden mit einem Kopfmikrofon (bayerdynamics) bei einem Mund-Mikrofonabstand von ca. 10cm mit einem DAT-Recorder aufgenommen. Dabei wurde darauf geachtet, dass sich das Mikrofon seitlich und unterhalb des Mundes befand, da sonst Störgeräusche durch Pusten auftreten. Bei Gruppe neun wurde außerdem ein Elektroglottogramm aufgenommen. Die Daten wurden vom DAT–Band digital auf die Festplatten des Computers überspielt, so dass keine zusätzlichen Verluste durch AD-DA Wandlung auftraten.

Im folgenden wird auf die jeweiligen Gruppen nur noch mit der Nummer der Gruppe verwiesen, also z.B. Gruppe 2 oder Gruppe 8. Diese Gruppenangabe bezieht sich, wenn nicht anders angegeben, auf Tabelle 4.1.

Tabelle 4.1.: Übersicht Stimmaufnahmen. Die Aufnahmen mit 50kHz Abtastfrequenz (f_s) wurden mit dem CSL aufgenommen. Die Aufnahmen mit 48kHz Abtastfrequenz wurden mit einem DAT Recorder durchgeführt.

Nr.	Gruppe	f_s [kHz]	Anzahl	Vokal	Länge [s]	Grundfrequenz
1	Pathologische Stimmen (Mehrfachaufnahmen möglich), ein Segment pro Aufnahme, 1s Segmentlänge	50	459	[ɛ:]	>1	normal
2	Anästhesie-Patienten (Mehrfachaufnahmen möglich), ein Segment pro Aufnahme, 1s Segmentlänge	50	101	[ɛ:]	>1	normal
3	Normale Sprecher (Mehrfachaufnahmen möglich), ein Segment pro Aufnahme, 1s Segmentlänge	50	23	[ɛ:]	>1	normal
4	Wie Gruppe 1, jedoch jeder Sprecher nur einmal, ein Segment pro Aufnahme, 1s Segmentlänge	50	447	[ɛ:]	1	normal
5	Wie Gruppe 2 und 3, jedoch jeder Sprecher nur einmal, ein Segment pro Aufnahme, 1s Segmentlänge	50	88	[ɛ:]	1	normal
6	Gruppe 1,2,3; bis zu 4 nichtüberlappende Segmente pro Aufnahme, 1s Segmentlänge	50	1799	[ɛ:]	1	normal
7	5 normale Sprecher, 3 normale Versionen, 1 Mal geflüstert, ein Segment pro Aufnahme, 1s Segmentlänge	50	200	[ɛ:]	1	98-247Hz in 10Hz Schritten
8	3 normale Sprecher, Mikrofon und EGG, reflexionsfreier Raum	48	51	[ɛ:]	>3	Tonhöhenanstieg ca. 70-500Hz, mit Atemunterbrechungen
9	Normale und pathologische Sprecher, Segmente mit 0,5s Länge und 0,25 s Vorschub	48	293817	$[ɛ:]_1$, [a:], [e:], [i:], [o:], [u:], $[ɛ:]_2$	>0,5	jeweils normal 1, tief, hoch, normal 2
10	Gruppe 4 plus zusätzliche Aufnahmen der Sprecher, 0,5s Segmentlänge, 0,25s Segmentvorschub	50	13414	[ɛ:]	0,5	normal

5. Korrelation zwischen Hilberteinhüllenden in verschiedenen Frequenzbändern als Stimmgütemaß

In diesem Kapitel wird ein neues akustisches Maß zur Stimmgütebeschreibung vorgestellt, der Glottal to Noise Excitation Ratio (GNE) [84, 88]. Das Maß soll beschreiben, inwieweit die Stimme durch einen Verschluss der Stimmlippen oder durch turbulentes Rauschen an der Glottis angeregt wird.

5.1. Motivation eines neuen Maßes

Zur Motivation des neuen Maßes werden noch einmal die prinzipiell möglichen Abweichungen eines Signals von der exakten Periodizität zusammengefasst. Diese Abweichungen sind:

1. Jitter (Frequenzmodulationsrauschen): Abweichungen der Periodenlänge von Periode zu Periode.

2. Shimmer (Amplitudenmodulationsrauschen): Abweichungen der Amplitude von Periode zu Periode.

3. Veränderung der Signalform von Periode zu Periode, z.B. durch Veränderungen der Artikulation oder der Glottisöffnungsfunktion.

4. Additives Rauschen, das durch turbulente Strömung an einer Vokaltrakteinengung entsteht. (Bei Vokalen kann turbulentes Rauschen durch inkompletten Verschluss der Stimmlippen entstehen. Zusatzlich kann bei Vokalen Rauschen an den Lippen erzeugt werden. Dies ist häufig bei den Vokalen [o:] und [u:] der Fall, wenn bei der Artikulation die Lippen fast geschlossen werden.)

Im ersten Teil dieser Arbeit wird das Ziel verfolgt, Maße zu implementieren, die in der Lage sind, die einzelnen Arten der Abweichungen von der Periodizität *möglichst unabhängig* von den anderen Arten zu messen. Folgende Überlegungen sollen zeigen, dass dies schon aus methodischen Gründen bei vielen bekannten Maßen nicht der Fall ist.

Betrachten wir zunächst die Jittermessung. In vielen Studien werden zur Jittermessung peakpicking oder zerocrossing verwendet, um die Periodenlängen zu bestimmen. Addiert man zu einem Signal ohne Jitter (einem Sinus oder einem synthetischen Vokal) ein Rauschsignal (weißes oder rosa Rauschen), so wird die Lage des Maximums oder des Nulldurchgangs zufällig verschoben sein, und zwar umso stärker, je stärker der relative Rauschpegel ist. Sind bei dem Signal schon Periodenlängenschwankungen vorhanden, so wird der mit peakpicking und zerocrossing gemessene Jitter ebenfalls mit dem Pegel des Rauschens zunehmen. Die gleiche Rauschanfälligkeit gilt entsprechend für Shimmer, wenn dieser aus den minimalen bzw. maximalen Amplituden pro Periode berechnet wird oder aus deren Differenz.

Der mittlere Korrelationswert aufeinander folgender Perioden (Mean Waveform–Matching Coefficient MWC) kann verwendet werden, um Abweichungen der Signalform zu quantifizieren. Der Korrelationswert ist durch die Normierung nicht vom Shimmer abhängig. Enthält das Signal jedoch Jitter, so sinkt der Korrelationswert da die einzelnen Abtastwerte zeitlich nicht mehr exakt zusammenpassen.

Bisherige Maße zur Bestimmung des additiven Rauschens sind nicht nur von dem Pegel des Rauschens abhängig, sie sind auch sensitiv für Jitter. Dies wurde von zwei Autoren bei der Vorstellung neuer Maße gezeigt. In Abbildung 5.1 aus der Arbeit von de Krom [17] und in Abbildung 5.2 aus der Arbeit von Muta [94], ist zu erkennen, dass die Maße HNR und N/S Ratio nicht nur den Rauschanteil messen, sondern auch auf Jitter reagieren. Eine getrennte Messung von additivem Rauschen und Jitter ist mit HNR und N/S Ratio nicht möglich.

Beide Maße beziehen die Energie in den Tälern des Spektrums auf die Energie der harmonischen Spektralkomponenten. Durch Jitter steigt die Energie in den Tälern zwischen den Harmonischen, so dass der gemessene Signal–Rauschabstand sinkt. Außerdem muss man darauf achten, dass die Form der Harmonischen und die Tiefe der Täler zusätzlich von der Fensterfunktion abhängen.

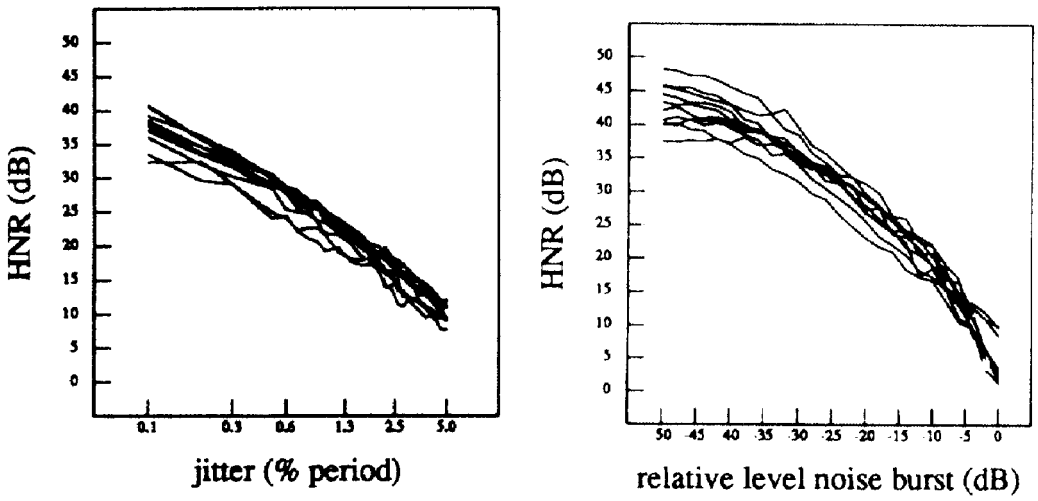


Abbildung 5.1.: Vergleich von Jitter- und Rauschabhängigkeit des Stimmgütemaßes von de Krom 1993 [17].

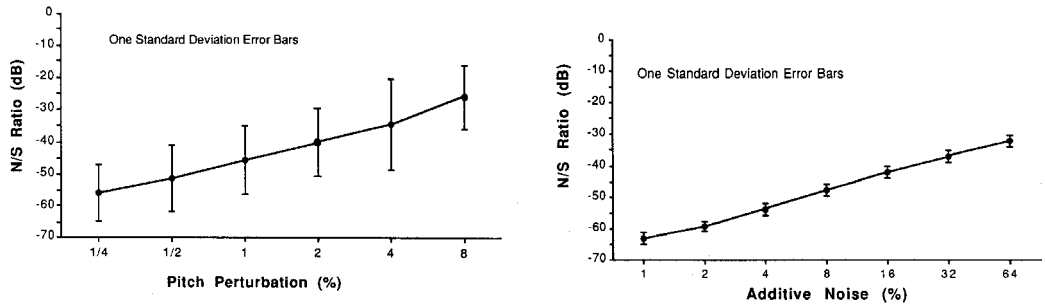


Abbildung 5.2.: Vergleich von Jitter- und Rauschabhängigkeit des Stimmgütemaßes von Muta 1988 [94].

Ein weiterer Nachteil bisheriger Rauschmaße ist, dass zu ihrer Berechnung die Periodenlänge in dem analysierten Signalstück bestimmt werden muss. Bei einem regelmäßigen Signal, wie in Abbildung 5.3, bereitet diese Aufgabe keine Schwierigkeit.

Ist das Signal jedoch sehr unregelmäßig, wie in Abbildung 5.4, so ist eine sinnvolle Bestimmung der Periodenlänge schwierig. Zu der Stimme in Abbildung 5.4 ist noch zu sagen, dass perzeptiv deutlich Pulse wahrnehmnbar sind, die auf unregelmäßig schwingendes Gewebe schließen lassen.

Die Beispielstimme in Abbildung 5.5 entspricht weitgehend reinem Rauschen. Die Stmme klingt aphon, sie enthält keinen stimmhaften Anteil. Bei solchen Stimmen ist keine Periodenlänge im ursprünglichen Sinn mehr vorhanden. Später wird auch für solche Signale eine Periodenlänge definiert und daraus Jitter und Shimmer berechnet, um nicht die Signale in analysierbare und nicht analysierbare Signale trennen zu müssen. Die praktische Anwendung muss zeigen ob ein solches Vorgehen gerechtfertigt ist.

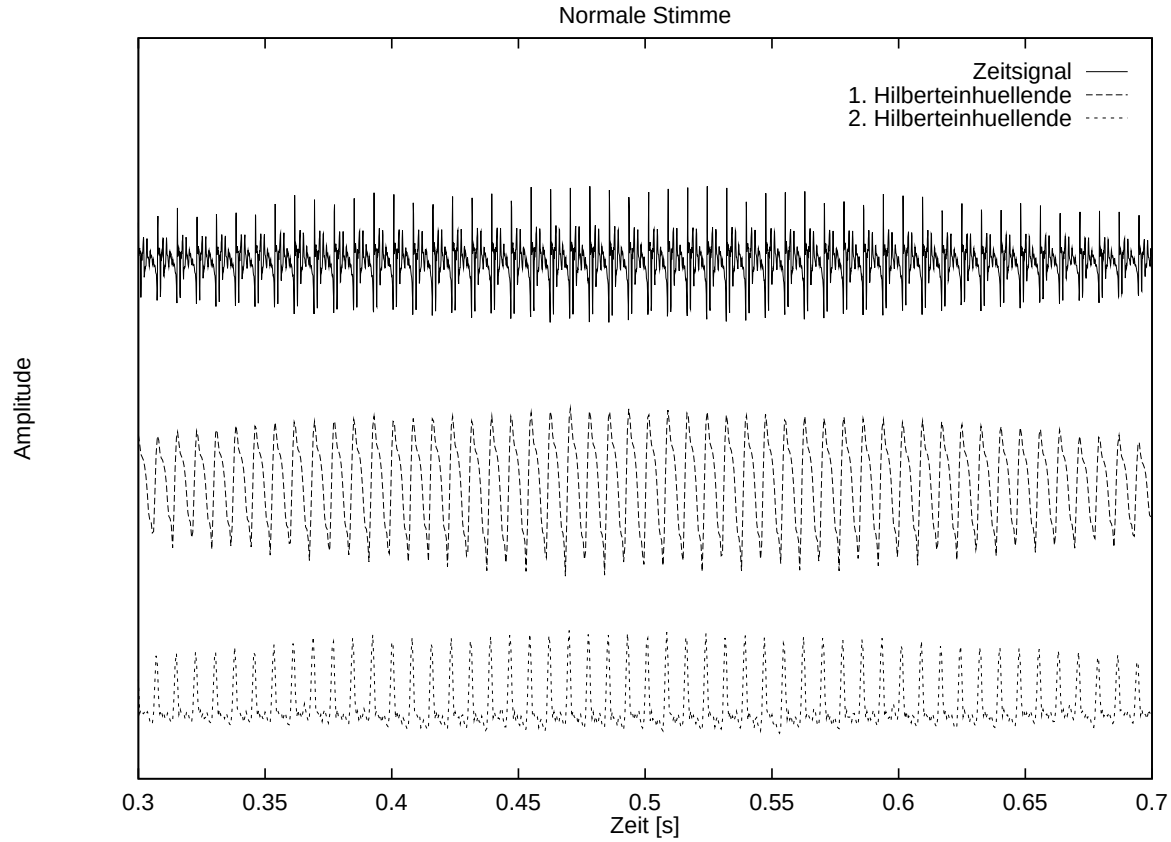


Abbildung 5.3.: 400 ms aus einem gehaltenen [ɛː] eines normalen Sprechers (oben) und seine kanalweisen Hilberteinhüllenden. Aufgrund der Gleichmäßigkeit der aufeinanderfolgenden Grundperioden ist die Bestimmung der Periodenlänge relativ einfach und zuverlässig möglich.

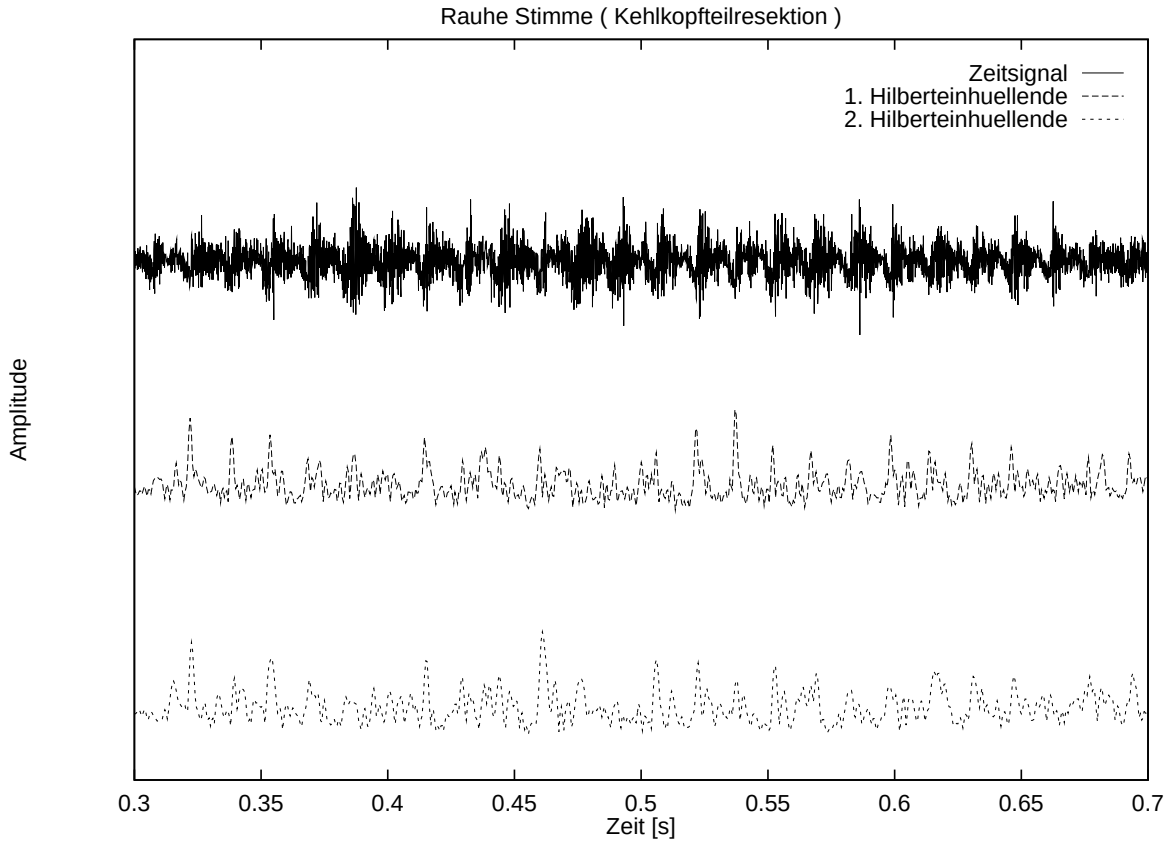


Abbildung 5.4.: 400 ms aus einem gehaltenen [ɛ:] eines Patienten mit einer Kehlkopfteilresektion. Eine bestimmte Periodenlänge ist nicht erkennbar. Die Anregung ist aber noch pulsartig.

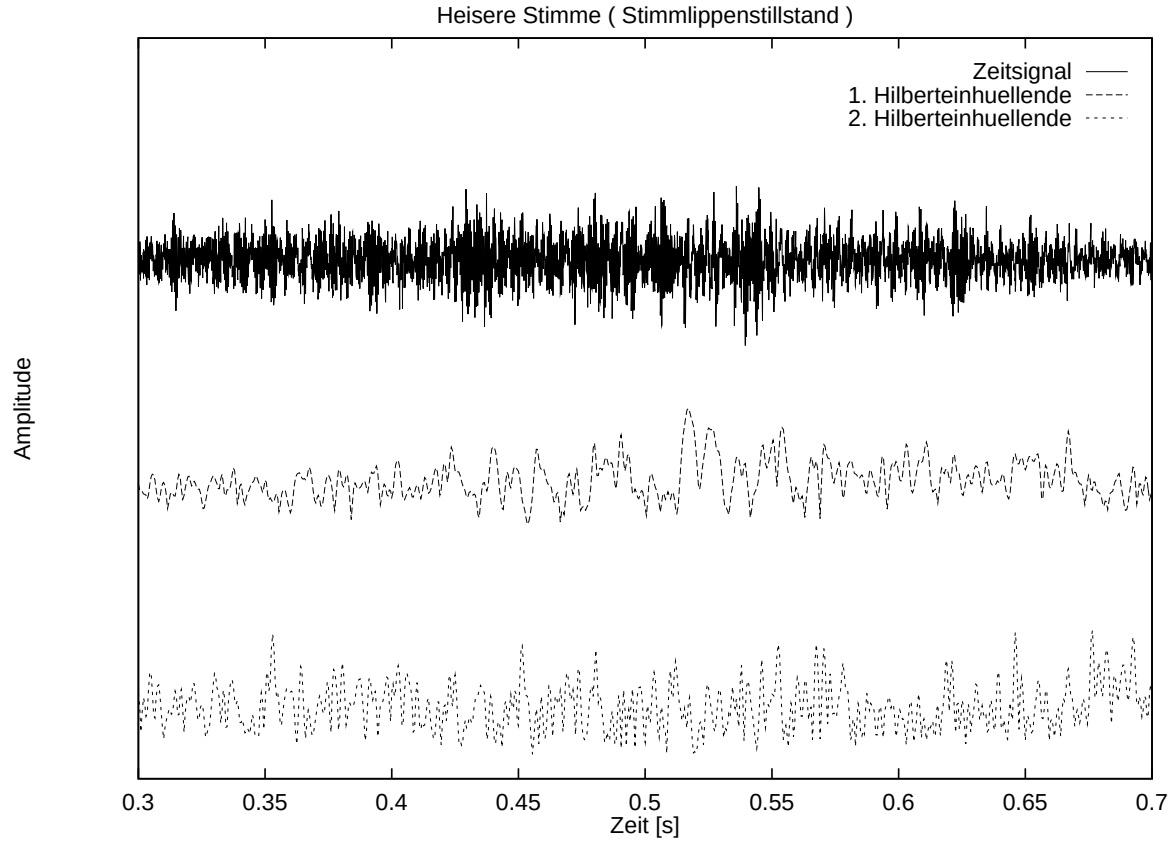


Abbildung 5.5.: 400 ms aus einem gehaltenen [ɛ:] eines Patienten mit einem Stimmlippenstillstand infolge einer Lähmung nach einer Kropfoperation. Das Zeitsignal sieht wie ein Rauschen aus, es sind keine Pulse, die einen Hinweis auf einen bestimmten Anregungszeitpunkt geben könnten, zu erkennen.

5.2. Hilberteinhüllende einer Pulsfolge und einer Rauschfolge

Um den neuen Ansatz für das Rauschmaß GNE zu erklären, ist in Abbildung 5.6 zunächst eine Pulsfolge statt eines Vokals dargestellt. Das Wesentliche an dem neuen Ansatz ist, dass das Signal in verschiedenen Frequenzbändern betrachtet wird. In Abbildung 5.6 sind unter dem Zeitsignal der Pulsfolge die Einhüllenden in den zwei Frequenzbereichen von 0-1000Hz und von 500-1500Hz abgebildet. Die Einhüllenden werden im Folgenden als Hilberteinhüllende bezeichnet, da sie aus dem Betrag des analytischen Signals berechnet werden, dessen Imaginärteil durch Hilberttransformation aus dem Signal gewonnen wird.

Bei der Pulsfolge sind die Hilberteinhüllenden der beiden Frequenzbereiche identisch.

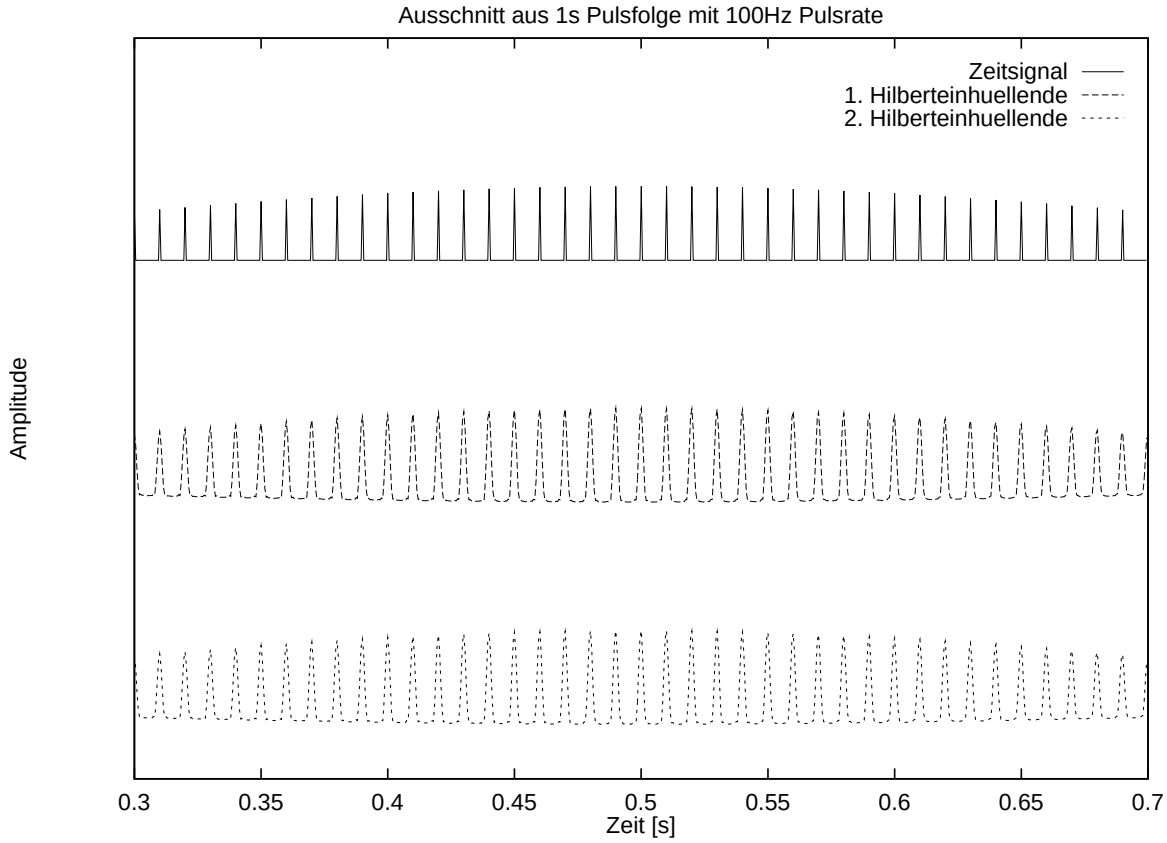


Abbildung 5.6.: Die Hilberteinhüllende des 1. Frequenzbereiches (0-1000 Hz) und des 2. (500-1500 Hz) sind bei einer gefensterten Pulsfolge (100 Hz Pulsfrequenz) identisch. Der Korrelationskoeffizient ist 1.

Eine Rauschfolge, wie in Abbildung 5.7, besitzt jedoch sehr unähnliche Hilberteinhüllende in den beiden Frequenzbereichen. Berechnet man den Korrelationswert zwischen den Einhüllenden als Maß der Ähnlichkeit, so ergibt sich für die Pulsfolge der Korrelationswert 1 und für diese Rauschfolge der Korrelationswert 0,05.

Man kann sich leicht vorstellen, dass der Korrelationswert einer Pulsfolge kleiner wird, wenn ein bestimmter Rauschanteil zu der Pulsfolge addiert wird. Die Abhängig-

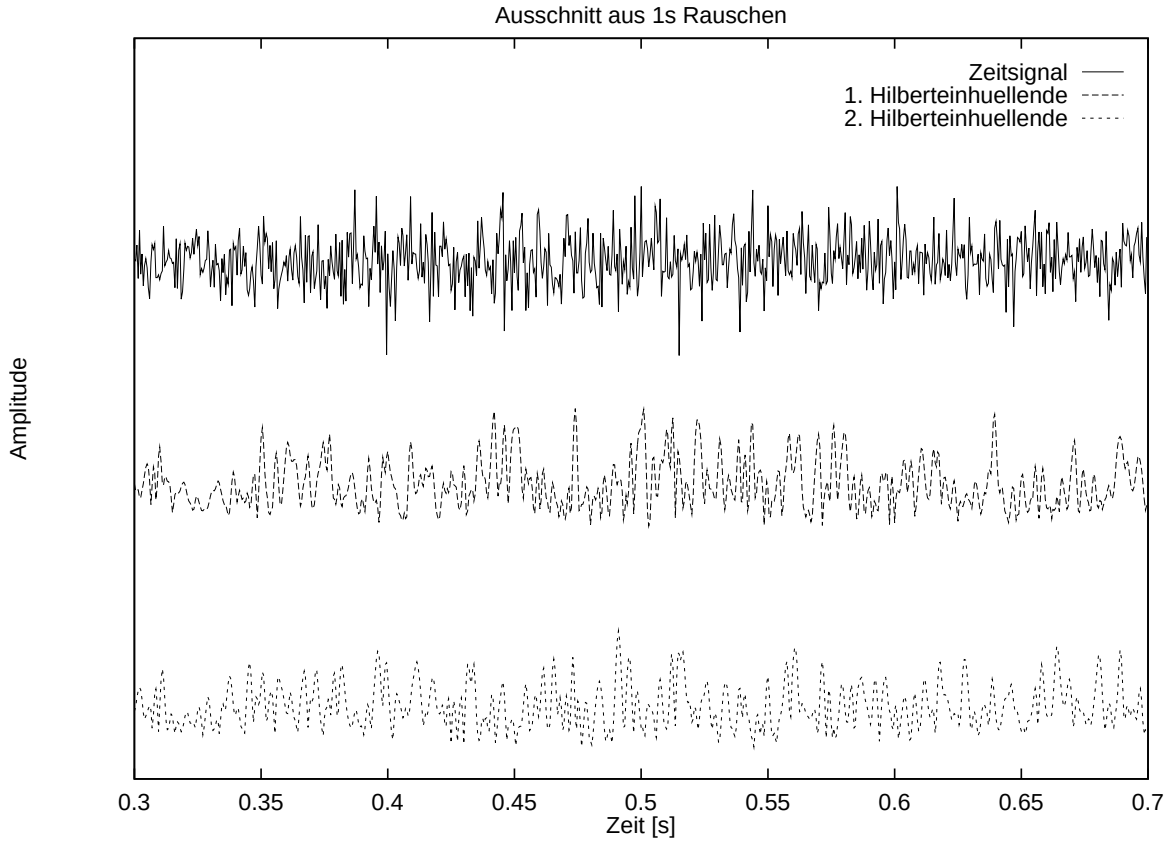


Abbildung 5.7.: Bei einer Rauschfolge sind die Hilberteinhüllenden unkorreliert. Der Korrelationskoeffizient beträgt 0,05

keit des Korrelationskoeffizienten vom RMS-Wert des Rauschens ist in Abbildung 5.8 dargestellt.

Die Ähnlichkeit der beiden Einhüllenden ist bei einer Pulsfolge auch dann noch gegeben, wenn der Pulsfolge 30% Jitter hinzugefügt wird (Abbildung 5.9).

Wird bei verschiedenen Grundfrequenzen der Pulsfolge der Jitter der Signale variiert, so ist bei 120Hz Grundfrequenz und 220Hz Grundfrequenz der Korrelationswert bei bis zu 30% Jitter nahe bei 1 (Abbildung 5.10). Bei höheren Grundfrequenzen (320Hz und 420Hz) nimmt der Korrelationswert mit zunehmendem Jitter ab. Der Grund dafür ist, dass bei höheren Grundfrequenzen und hohen Jitter Werten die Wahrscheinlichkeit wächst, dass zwei aufeinanderfolgende Pulse sehr dicht hintereinander liegen. Im Anhang B wird gezeigt, dass die Hilberteinhüllenden von zwei dicht beieinander liegenden Pulsen in verschiedenen Frequenzbereichen sehr unterschiedliche Form haben können. Diese Unterschiede führen zu der Verminderung des Korrelationskoeffizienten. Da es bei diesem Effekt jedoch nur auf das Verhältnis des Abstandes zweier Pulse und der Bandbreite der Frequenzbereiche ankommt, nimmt die Jitterabhängigkeit stark ab, wenn die Bandbreite der Frequenzbereiche von 1000Hz auf 2000Hz vergrößert wird (Abbildung 5.10 420Hz (2)). Der Abstand der Mittenfrequenzen wurde dabei von 500Hz auf 1000Hz erhöht, so dass die Einhüllenden in den Frequenzbereichen 0-2000Hz und 1000-3000Hz berechnet

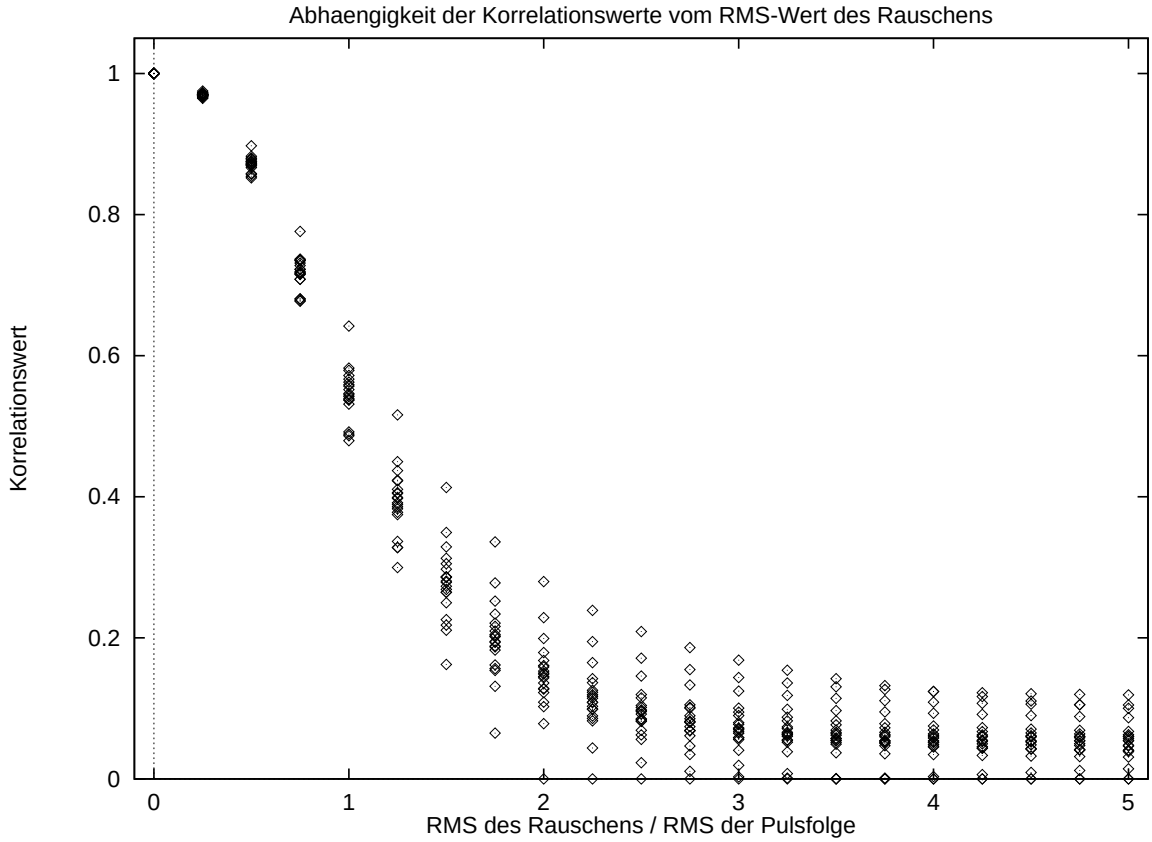


Abbildung 5.8.: Die Abhängigkeit des Korrelationskoeffizienten vom RMS-Wert des Rauschens bezogen auf den RMS-Wert der Pulsfolge ist im Mittel eine monoton fallende Funktion.

wurden.

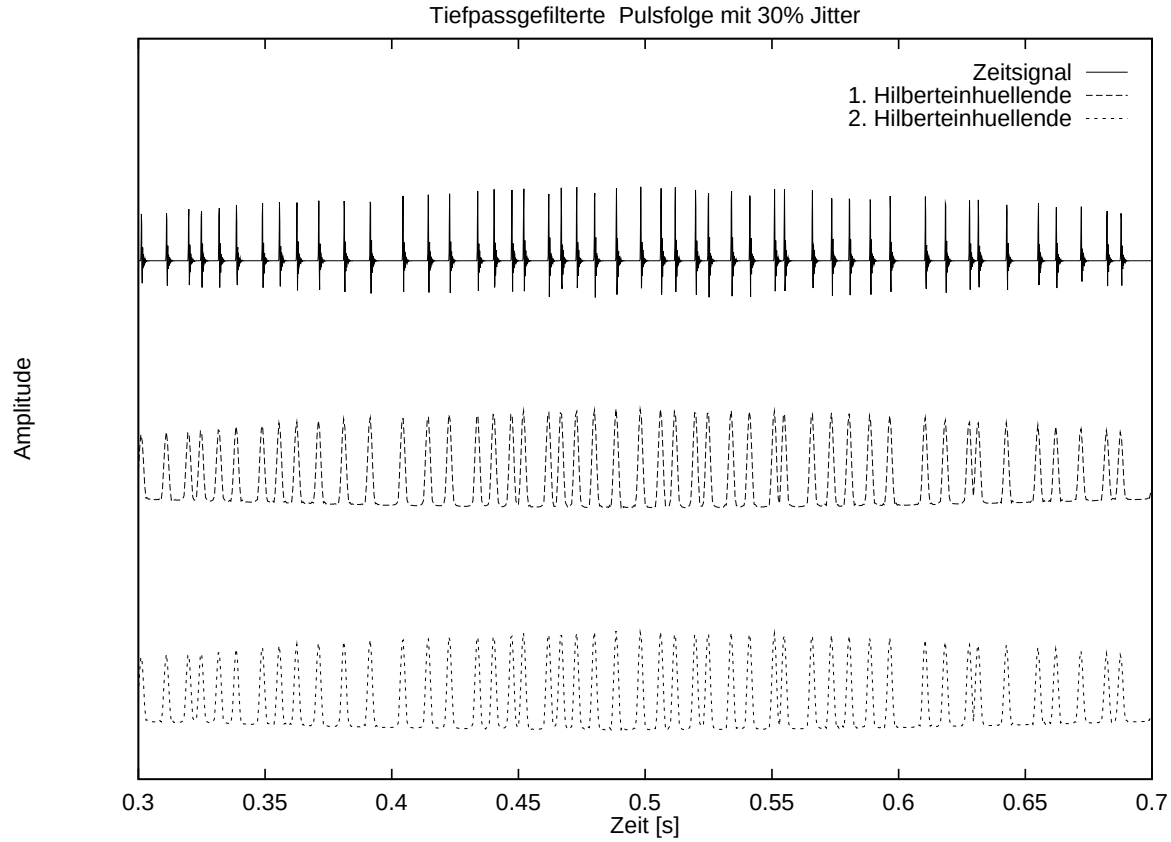


Abbildung 5.9.: Eine Pulsfolge mit 120 Hz und 30% Jitter. Man sieht, dass trotz der großen Unregelmäßigkeit der Periodenlängen die Hilberteinhüllenden gleich aussehen.

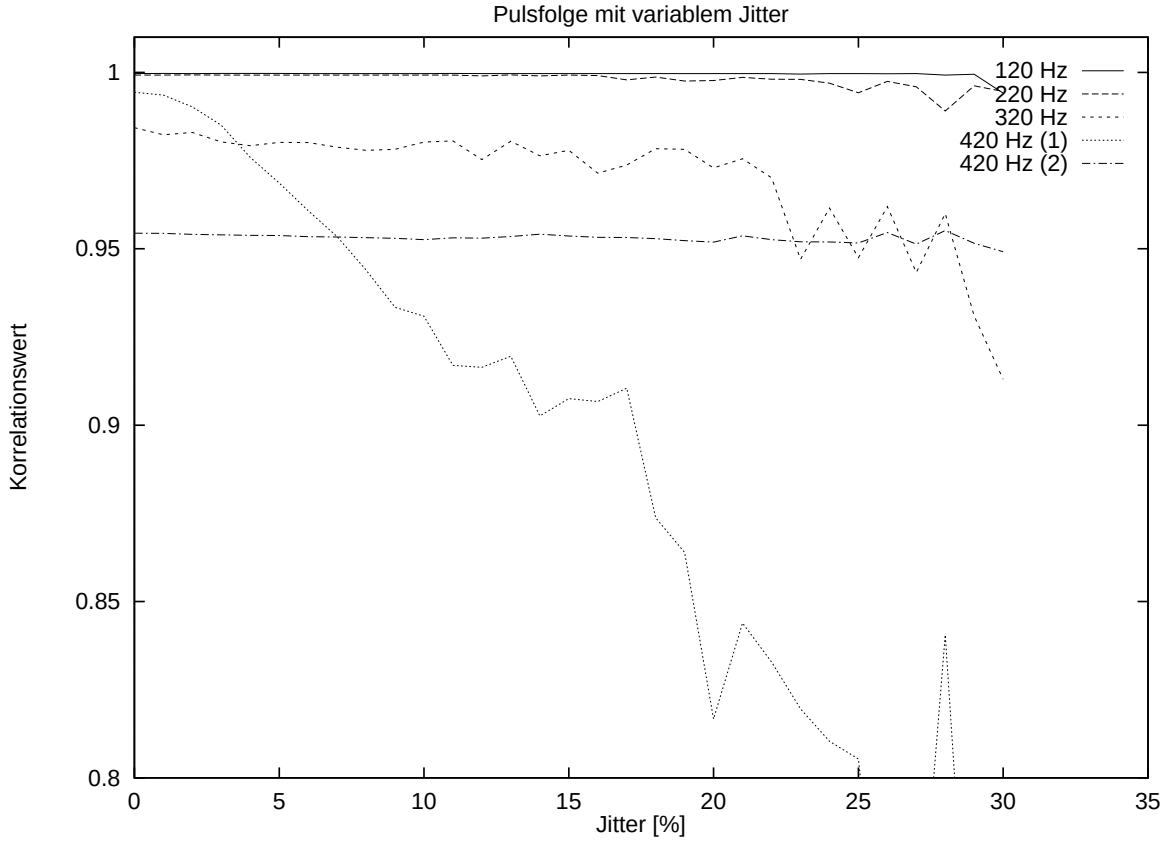


Abbildung 5.10.: Der Korrelationskoeffizient ist nahezu konstant, wenn die Pulsfolgen variablen Jitter enthalten. Bei hoher Grundfrequenz (420Hz) sinkt der Korrelationswert mit steigendem Jitter ab. Wird die Bandbreite der Frequenzbereiche von 1000Hz (420Hz (1)) auf 2000Hz heraufgesetzt, so verschwindet die Jitterabhängigkeit.

5.3. Inverse Filterung

Die Pulsfolge war eine Abstraktion vom Sprachsignal, die jetzt genauer diskutiert werden soll. Der Korrelationskoeffizient zwischen den beiden Hilberteinhüllenden des regelmäßigen Vokals [ɛ:] (Abbildung 5.3) beträgt 0,7 und ist damit deutlich kleiner als eins. Nun handelt es sich bei dem Signal ja auch nicht um eine Pulsfolge, sondern um eine Pulsfolge, die mit der Glottisfunktion und der Übertragungsfunktion des Vokaltraktes gefaltet ist. Durch inverse Filterung kann jedoch das Spektrum des Signals geglättet werden. Bei geschickter Wahl der Ordnung des linearen Filters und nach einer Unterabtastung der Signale auf 10kHz kann man aus dem Signal (Abbildung 5.3) eine Pulsfolge rekonstruieren (Abbildung 5.11).

Da die Anregung in verschiedenen Frequenzbereiche zu verschiedenen Zeiten erfolgen kann, wird nicht nur ein Korrelationswert zwischen den Einhüllenden berechnet, sondern die Korrelationsfunktion im Bereich von -0,3ms bis 0,3ms berechnet und deren Maximum bestimmt.

Der daraus resultierende Korrelationswert steigt von 0,7 auf 0,98 an und liegt damit sehr dicht an dem theoretischen Höchstwert von 1. Im Vergleich zu dieser gesunden Stimme ist der Korrelationswert der unregelmäßigen Stimme (Abbildung 5.4) nach inverser Filterung mit 0,42 deutlich geringer. Er liegt aber seinerseits noch deutlich über dem Korrelationwert der aphonen Stimme (Abbildung 5.5) der nur noch 0,15 beträgt.

Der Korrelationswert zwischen den Einhüllenden in verschiedenen Frequenzbereichen des inversgefilterten Signals lässt also eine deutliche Unterscheidung dieser drei Stimmarten zu und ist deshalb ein guter Kandidat für ein neues Stimmgütemaß.

Die einzelnen Schritte der Signalverarbeitung zur Berechnung des neuen Maßes seien noch einmal zusammengefasst:

1. Inverse Filterung der Signale (0,5s Fenster, 10kHz Abtastfrequenz, Autokorrelationsmethode, 13. Ordnung, 30ms Hanning–Fenster bei 10ms Fenstervorschub)
2. Fouriertransformation
3. Ausschneiden von zwei Hanning-Frequenzfenstern
4. Auffüllen der Fenster mit derselben Länge Nullen (negative Frequenzen auf Null)
5. Fourierrücktransformation dieser beiden Bänder
6. Betragsbildung
7. Mittelwertbefreiung
8. Berechnung der Kreuzkorrelationsfunktion (im Bereich von -0,3ms bis 0,3ms)
9. Bestimmung des Maximums der Kreuzkorrelationsfunktion

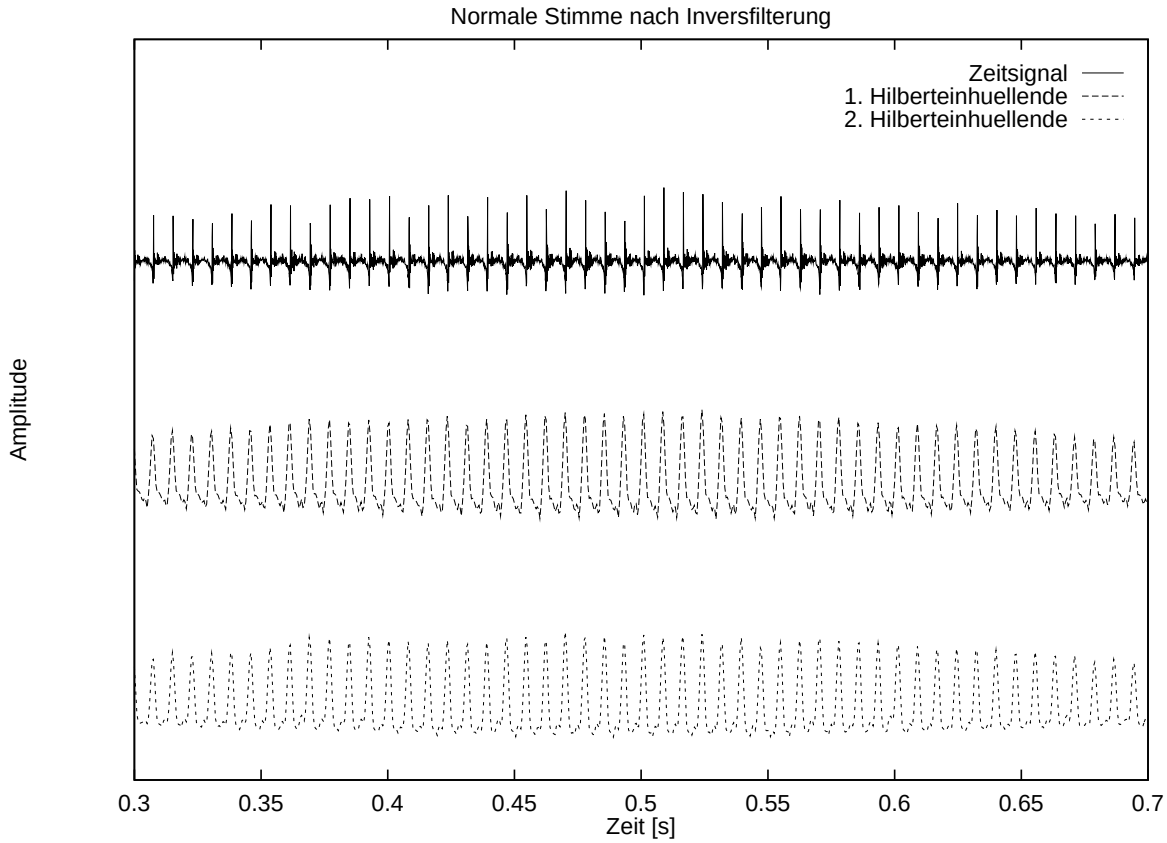


Abbildung 5.11.: 400 ms aus einem gehaltenen [ɛ:] eines normalen Sprechers (oben) und seine kanalweisen Hilberteinhüllenden nach inverser Filterung. Die Ähnlichkeit der Hilberteinhüllenden hat gegenüber der ungefilterten Stimme deutlich zugenommen.

5.4. Messungen bei männlichen, gesunden Sprechern

Um die Anwendbarkeit des Korrelationsparameters auf menschliche Stimmen eingehender zu prüfen, wurden Aufnahmen von 5 männlichen Sprechern bei Grundfrequenzen von 98 Hz bis 247 Hz gemacht und außerdem bei der für die jeweilige Grundfrequenz charakteristischen Haltung von Kehlkopf, Zunge und Lippen die Flüsterstimme aufgenommen.

Die Bandbreite und der Fensterabstand wurden so wie oben gewählt. Der Fensterabstand der beiden Frequenzfenster beträgt eine halbe Fensterbreite. Das Ergebnis der Messungen ist in Abb. 5.12 und Abb. 5.13 dargestellt. Die Erwartungen an den Korrelationskoeffizient werden im Wesentlichen bestätigt. In Abb. 5.12 zeigt sich genau das erwartete Bild: Die Korrelationskoeffizienten der jeweils drei normalen Äußerungen liegen für alle Frequenzen nahe bei 1, die der Flüsterstimme hingegen alle bei sehr kleinen Korrelationskoeffizienten.

In Abb. 5.13 sieht man jedoch einige Ausreißer bei den normalen Stimmen zu niedrigen Korrelationskoeffizienten, die im Folgenden diskutiert werden: Die Ausreißer im tiefen Frequenzbereich um 100Hz stammen von einem Sprecher, der sich bei diesen Frequenzen sichtbar unwohl fühlte und meinte, er sei nun mal keine Bassstimme. Die Ausreißer bei hohen Frequenzen hingegen sind ja, wie oben erwähnt, durch die Fensterbreite im Frequenzbereich bedingt. Bei den Ausreißern um 165 Hz hatte der Sprecher das Gefühl, dass ein Übergang im Schwingungsmodus der Stimmlippen derart stattfand, dass durch Überlagerung zweier Moden die beiden Frequenzbänder unterschiedlich angeregt wurden.

Trotz der stellenweise niedrigen Korrelationskoeffizienten bei einzelnen Frequenzen bietet der Überblick über alle Frequenzen bei jedem einzelnen Sprecher die Möglichkeit, ohne jeden Zweifel zwischen geflüsterter und normaler Stimme zu unterscheiden. Bei mehrfachen Aufnahmen eines Sprechers kann man sich am Maximum der erreichten Korrelation orientieren, denn eine normale Stimme kann zwar bei einzelnen Aufnahmen niedrige Korrelationen haben, umgekehrt gibt es bei heiseren und geflüsterten keine Ausreißer zu hohen Werten.

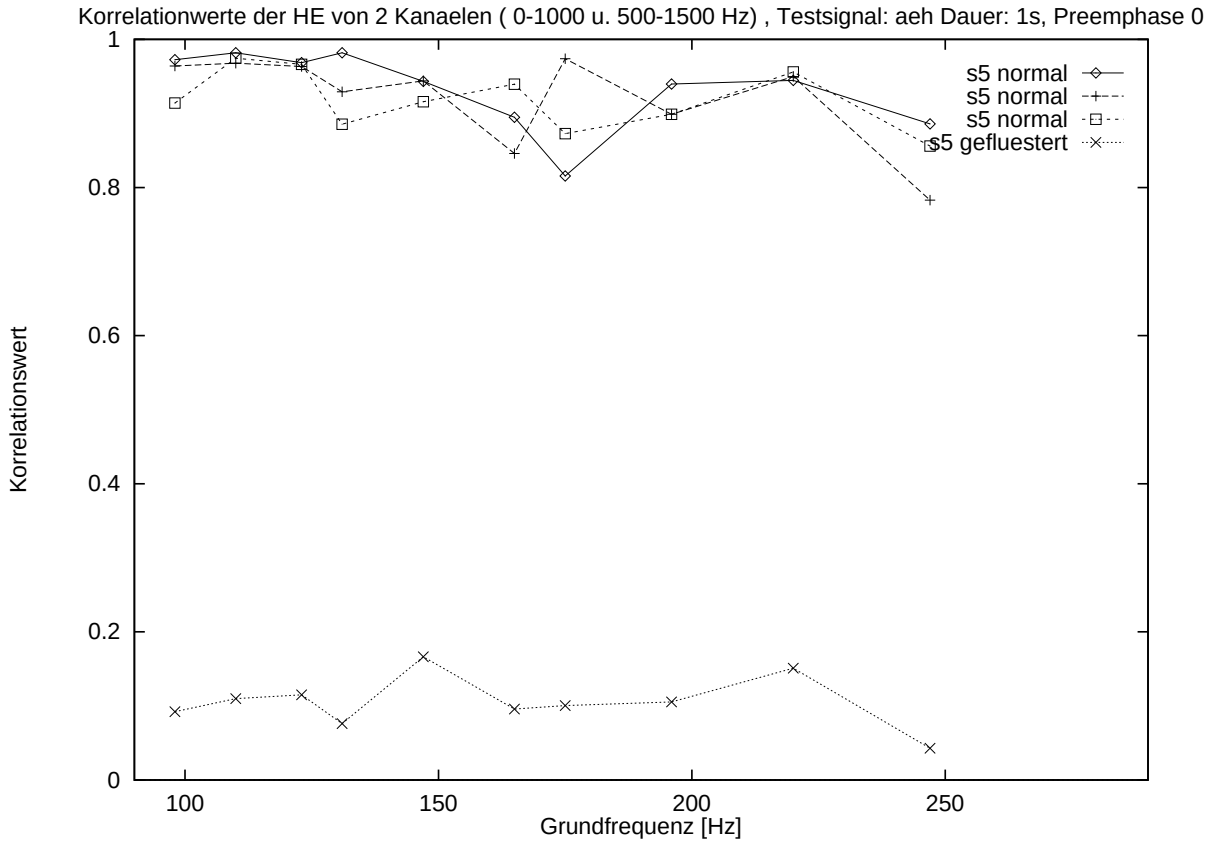


Abbildung 5.12.: Der Korrelationskoeffizient eines Sprechers bei normaler Äußerung liegt nahe bei 1 und entspricht damit einer Pulsanregung. Die Korrelation der geflüsterten Äußerungen liegt bei 0 und entspricht somit einem Rauschen.

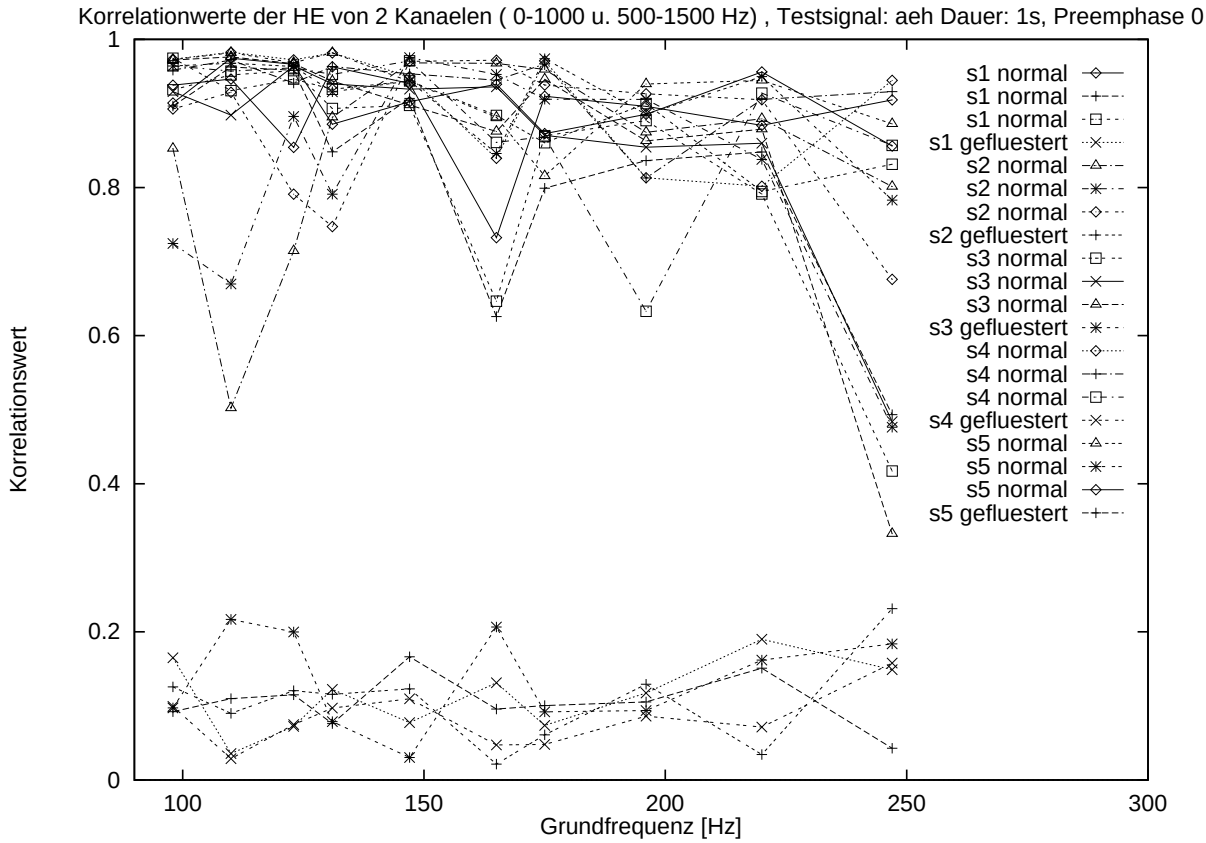


Abbildung 5.13.: Der Korrelationskoeffizient liegt bei allen Sprechern bei normaler Äußerung im Mittel über alle Grundfrequenzen bei 1. Die Korrelation der geflüsterten Äußerungen liegt bei allen Sprechern bei 0.

5.5. Optimierung des Parameters

Die Ausreißer der Korrelationskoeffizienten von normalen Stimmen zu niedrigen Werten in Abb. 5.13 waren der Anlass, den Einfluss der Mittenfrequenzen der Kanäle auf den Korrelationskoeffizienten genauer zu untersuchen. Dazu wurden nicht nur zwei, sondern eine größere Anzahl (30 bis 80) von kanalweisen Hilberteinhüllenden gebildet und der Korrelationskoeffizient zwischen jeweils allen Einhüllenden berechnet. Die Mittenfrequenzen lagen nun von 500 Hz bis 4500Hz linear verteilt. Für die Abbildungen 5.14, 5.15 und 5.16 wurden 41 Hilberteinhüllende berechnet, die Mittenfrequenz wurde in 100 Hz-Schritten variiert. In den Abbildungen sind die Korrelationskoeffizienten über den Mittenfrequenzen der zwei zu korrelierenden Kanäle dargestellt. Die Korrelationswerte auf der Diagonalen sind alle eins, da hier jeder Kanal mit sich selbst korreliert wird. Auf der einen Seite der Diagonalen sind alle Korrelationswerte auf Null gesetzt, da sonst alle Werte bis auf die in der Diagonalen doppelt auftreten würden. Je weiter ein Korrelationswert von der Diagonalen entfernt ist, desto weiter liegen die Mittenfrequenzen der beiden korrelierten Kanäle auseinander. In Abb. 5.14 sind die Korrelationskoeffizienten des gesunden Sprechers alle sehr dicht bei eins. Insbesondere auch bei Korrelationen zwischen Hilberteinhüllenden mit weit auseinanderliegenden Mittenfrequenzen. Bei diesem Sprecher ist die Wahl der Kanäle unkritisch, man erhält für jede Zweierkombination einen hohen Korrelationskoeffizienten, der die rauscharme, glottale Anregung der Stimme dokumentiert.

Die Korrelationskoeffizienten für ein geflüstertes /ä/ sind in Abb. 5.15 gezeigt. Die Korrelationswerte fallen zunächst von der Hauptdiagonalen her steil ab, bis sie bei einem gewissen Abstand von der Hauptdiagonalen mit einer geringen Streuung konstant unter einem relativ niedrigen Maximalwert bleiben. Die Korrelationswerte nahe der Hauptdiagonalen fallen mit dem Abstand ab, da die Kanäle durch den Überlapp der Frequenzfenster korreliert sind. Wenn man nun ein Maß für den Rauschanteil sucht, so muss man den Teil der Matrix, der aufgrund der Frequenzüberlappung korreliert, abschneiden. Dieser Bereich, der abgeschnitten werden muss, liegt mindestens bei der halben Breite der Frequenzfenster. Dann liefert der Maximalwert der verbleibenden Korrelationsmatrix ein gutes Maß für den Heiserkeitsgrad der Stimme.

Die Notwendigkeit, das Maximum der Korrelationskoeffizienten als Maß für den Rauschanteil zu benutzen, ersieht man aus Abb. 5.16. Hier liegt das Minimum aller Korrelationskoeffizienten im Bereich des Minimums der Flüsterstimme. Es treten relativ viele Kanalpaare auf, deren Korrelationkoeffizienten gering sind, deshalb liegt auch der Mittelwert deutlich unter eins.

Die beste Trennung der normalen Stimmen mit Glottisanregung von den Flüsterstimmen wird erreicht, wenn man das Maximum aller Korrelationskoeffizienten heraussucht, denn es gibt zwar niedrige Korrelationskoeffizienten bei normalen Stimmen, doch bei Flüsterstimmen tritt nie ein relativ hoher Korrelationskoeffizient auf.

In der Abb. 5.17 ist das Ergebnis der in dieser Weise optimierten Methode gezeigt. Man sieht, dass im direkten Vergleich zu Abb. 5.13 die Lücke zwischen Normalstimmen und Flüsterstimmen deutlich angewachsen ist, die Untescheidbarkeit zwischen Glottisanregung und Anregung durch turbulentes Rauschen hat zugenommen. Damit wird bei

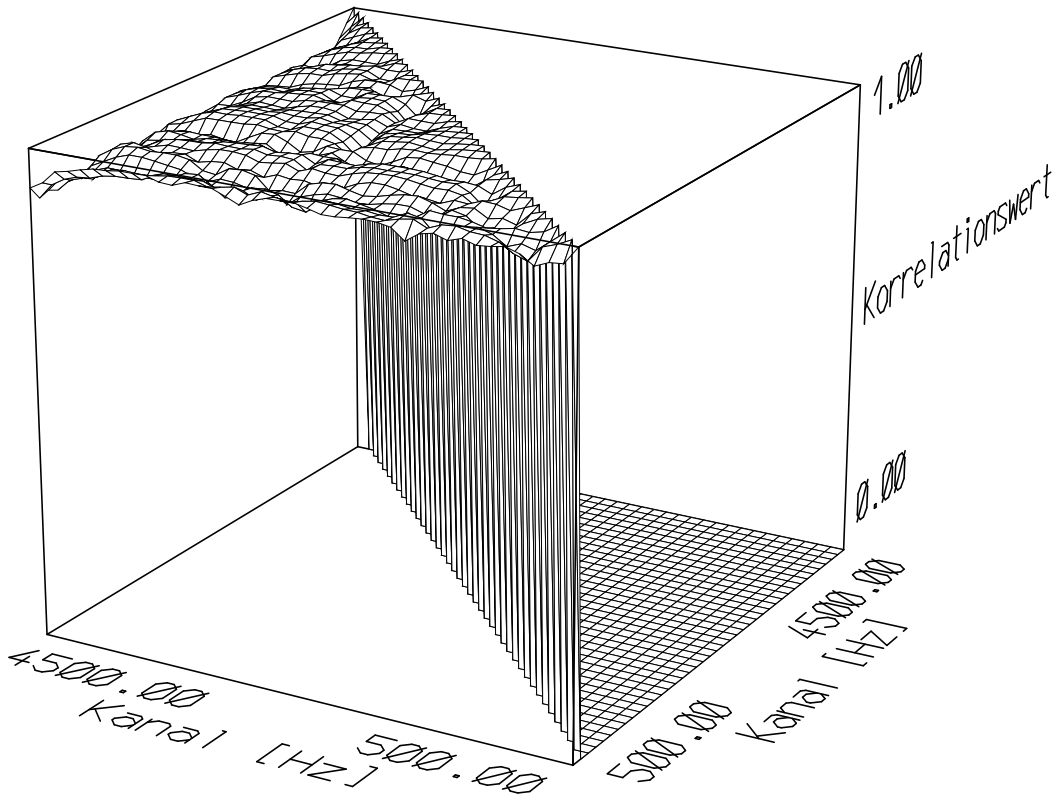


Abbildung 5.14.: Matrix der Korrelationskoeffizienten zwischen Hilberteinhüllenden mit variierenden Mittenfrequenzen von einem mustergültigen normalen Sprecher.

pathologischen Stimmen besser zwischen den Anregungsarten unterschieden und Mischformen der beiden Anregungsarten besser aufgelöst.

Da der Name: „Maximum der Matrix der Maxima der Kreuzkorrelationsfunktionen von den kanalweisen Hilberteinhüllenden des inversgefilterten akustischen Zeitsignales“ etwas unhandlich ist, wurde die Abkürzung GNE für das englische „Glottal to Noise Excitation“, also Glottale- zu Rauschanregung, eingeführt, um die Eigenschaften des neuen Parameters zusammenzufassen.

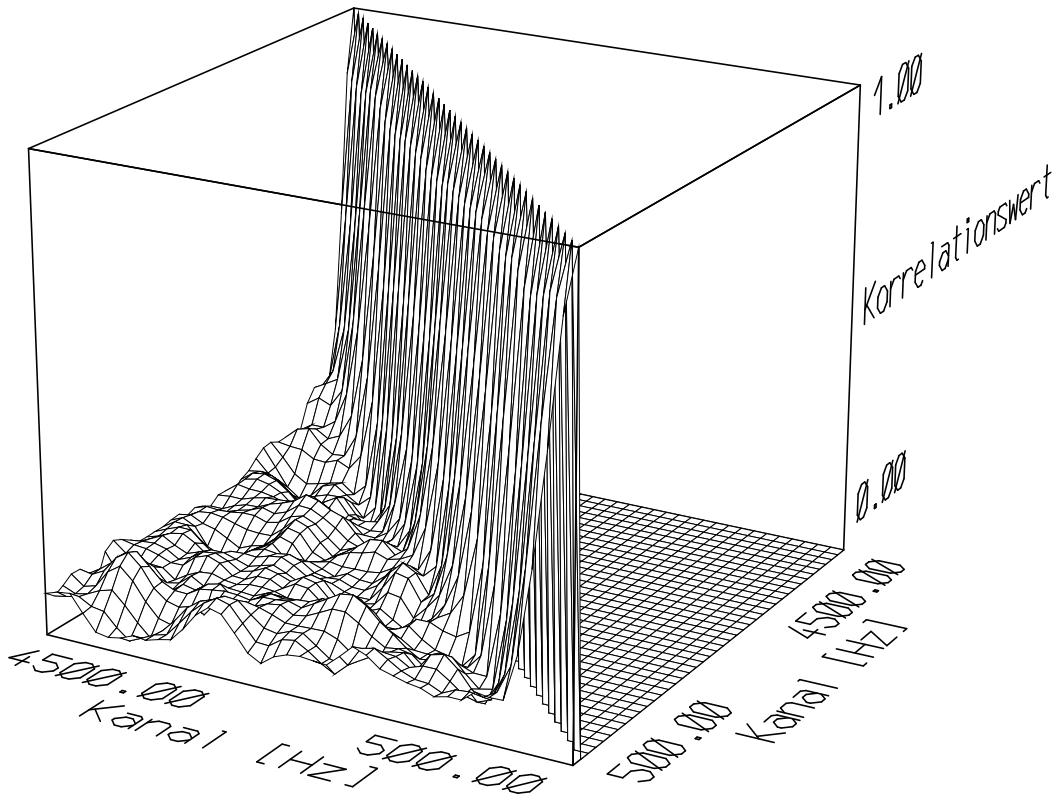


Abbildung 5.15.: Matrix, wie vorige Abbildung, von einem geflüsterten /ä/.

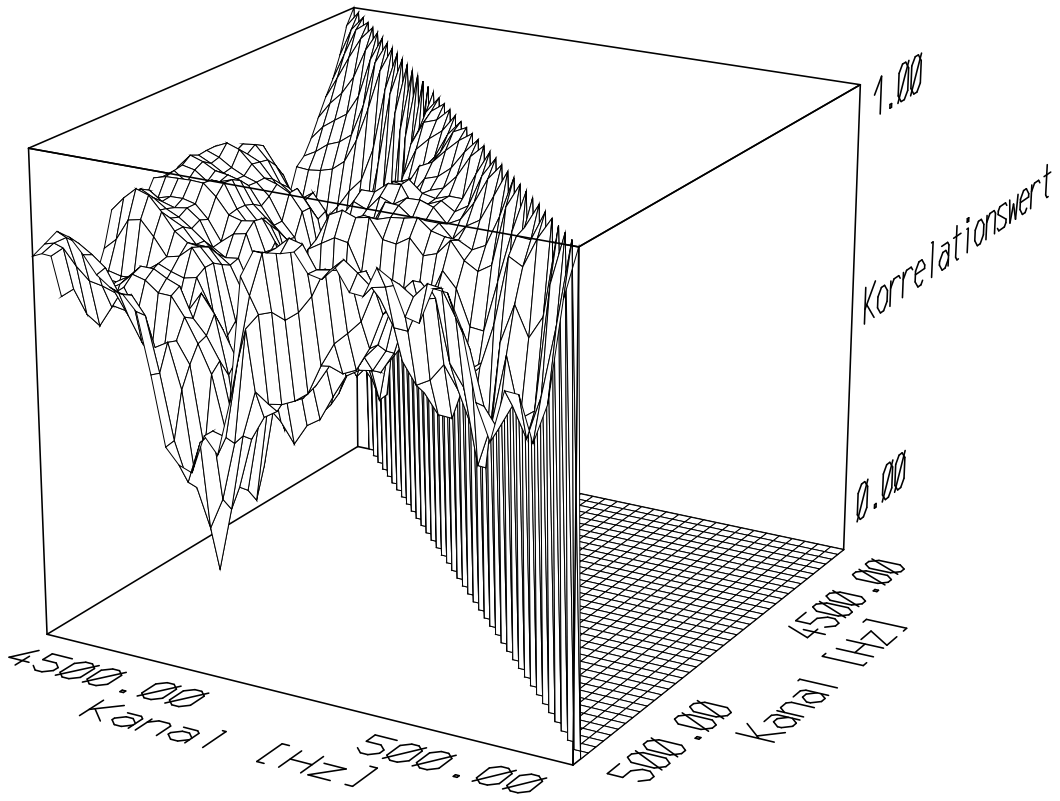


Abbildung 5.16.: Matrix, wie vorige Abbildungen, von einem normalen Sprecher. Es treten einige sehr niedrige Korrelationskoeffizienten auf.

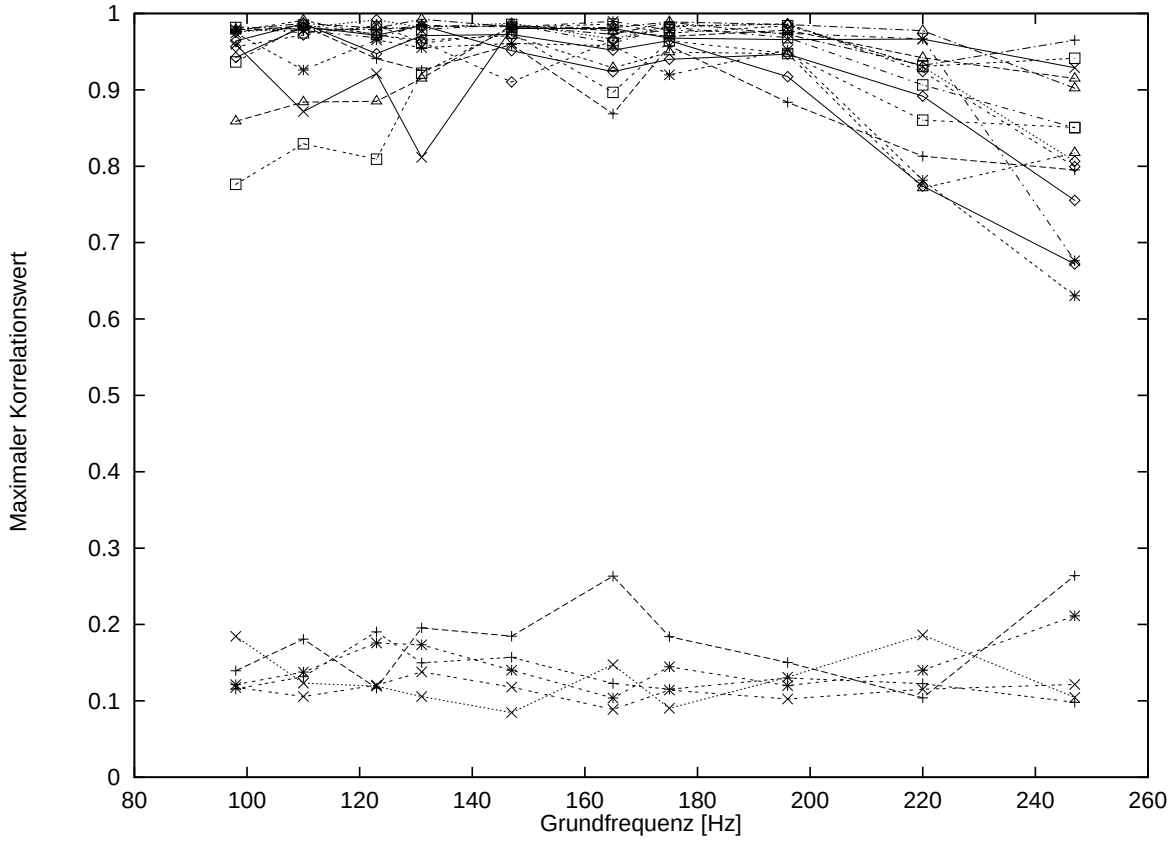


Abbildung 5.17.: Der maximale Korrelationskoeffizient erlaubt eine noch deutlichere Trennung von normalen und geflüsterten Stimmen.

5.6. Differenz der Mittenfrequenzen

Bisher wurden noch keine systematischen Untersuchungen über den notwendigen Abstand der Mittenfrequenzen von zwei bandpassgefilterten Hilberteinhüllenden durchgeführt. Ist der Abstand der Mittenfrequenzen klein, so ergeben sich Korrelationen zwischen den Einhüllenden aufgrund der Überlappung im Frequenzbereich, auch wenn es sich bei dem analysierten Signal um Rauschen handelt. Deshalb sollte der Abstand der Mittenfrequenzen möglichst groß sein.

Zur Berechnung des GNE werden aus dem inversgefilterten Signalen Hilberteinhüllende mit unterschiedlicher Mittenfrequenz bei gleicher Bandbreite (1000Hz, 2000Hz oder 3000Hz) berechnet. Bei 1000Hz Bandbreite werden 51 Einhüllende (Mittenfrequenzen von 500Hz bis 4500Hz in 80Hz Schritten), bei 2000Hz 31 Einhüllende (Mittenfrequenzen von 1000Hz bis 4000Hz in 100Hz Schritten) und bei 3000Hz 21 Einhüllende (Mittenfrequenzen von 1500Hz bis 3500Hz in 100Hz Schritten) berechnet.

Im folgenden wird der GNE mit Bandbreiten bis zu 3000Hz berechnet. Da die Signale zur Berechnung des GNE mit 10kHz abgetastet sind, stehen bei 3000Hz Bandbreite aber nur Mittenfrequenzen von 1500Hz bis 3500Hz zur Verfügung. Die Differenz der Mittenfrequenzen kann deshalb maximal 2000Hz betragen. Die minimale Differenz, für die Korrelationswerte zwischen den Einhüllenden sinnvoll sind, soll nun untersucht werden.

Dazu wurden jeweils 10 0,5s lange Rauschsignale erzeugt und wie die Vokale vorverarbeitet. Für die verschiedenen Bandbreiten 1000, 2000 und 3000Hz wurden nun die maximalen Korrelationswerte der Hilberteinhüllenden bestimmt, wobei der minimale Abstand der Mittenfrequenzen variiert wurde.

Die Ergebnisse (jeweils 1-GNE) sind in den Abbildungen 5.18 bis 5.20 dargestellt. Bei allen drei Bandbreiten ist die Korrelation aufgrund der Frequenzüberlappung bei einer Mittenfrequenzendifferenz von ca. einer halben Bandbreite hinreichend weit abgesunken, so dass das Kriterium gerechtfertigt ist, im Weiteren jeweils die Korrelationen von Hilberteinhüllenden zu berücksichtigen, deren Mittenfrequenzen sich um mindestens eine halbe Bandbreite unterscheiden.

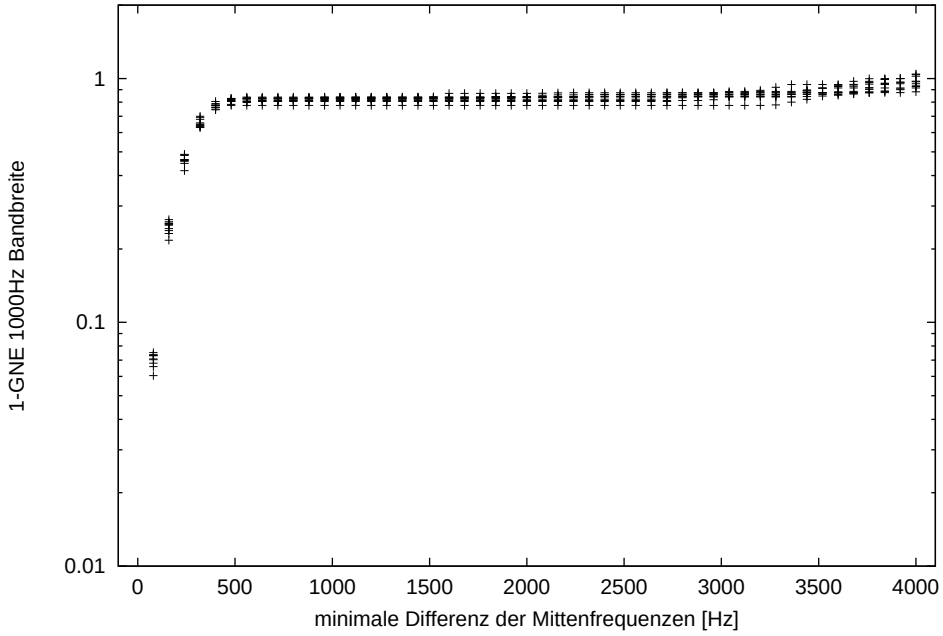


Abbildung 5.18.: Maximaler Korrelationswert der Hilberteinhüllenden bei 1000Hz Bandbreite in Abhängigkeit von der minimalen Differenz der Mittenfrequenzen

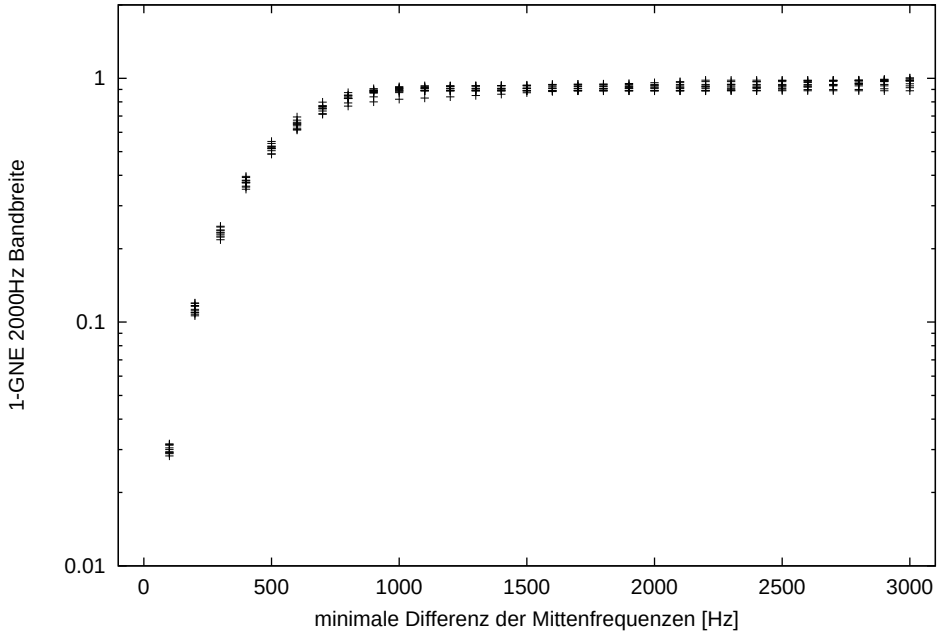


Abbildung 5.19.: Maximaler Korrelationswert der Hilberteinhüllenden bei 2000Hz Bandbreite in Abhängigkeit von der minimalen Differenz der Mittenfrequenzen

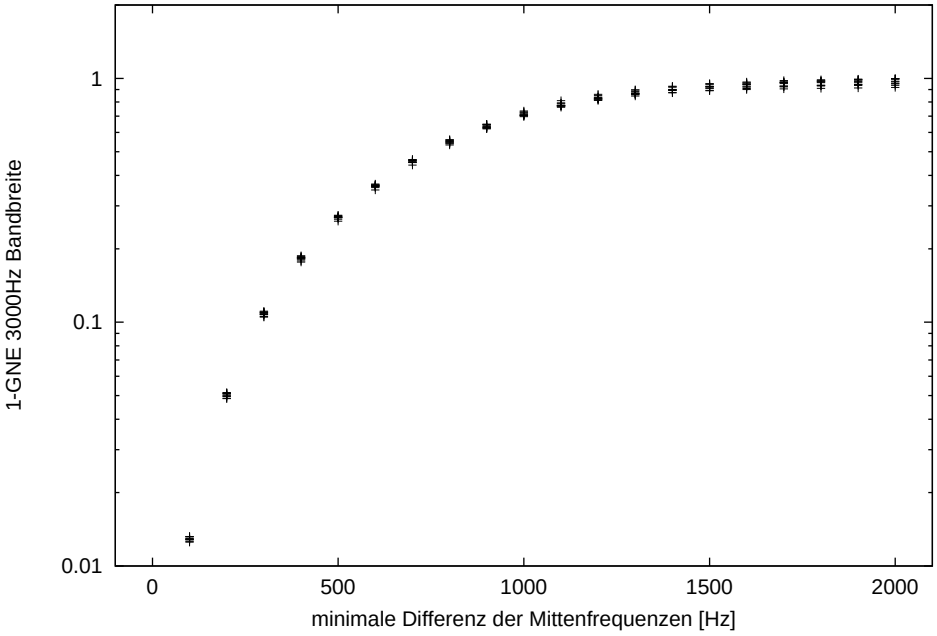


Abbildung 5.20.: Maximaler Korrelationswert der Hilberteinhüllenden bei 3000Hz Bandbreite in Abhängigkeit von der minimalen Differenz der Mittenfrequenzen

6. Vergleich des GNE mit anderen Rauschparametern

Im Folgenden wird genauer analysiert, wie GNE im Vergleich zu NNE und CHNR von dem Signal-Rauschverhältnis, Jitter und Shimmer abhängen. Dazu werden Signale mit vorgegebenem Jitter, Shimmer oder Rauschanteil erzeugt und daran GNE, NNE und CHNR gemessen.

GNE, CHNR und NNE werden jeweils in drei Varianten berechnet. Beim GNE wird die Bandbreite der Einhüllenden und beim NNE und CHNR der Frequenzbereich variiert. Die genauen Parameter sind in Tabelle 6.1 aufgelistet.

Tabelle 6.1.: Die neun Maße zur Messung des Rauschanteils

Messgröße (x)	Symbol	Beschreibung	Einheit	Anzahl Kanäle	Mittenfrequenzdifferenz [Hz]
GNE	GNE1	1000Hz Bandbreite		51	80
	GNE2	2000Hz Bandbreite		31	100
	GNE3	3000Hz Bandbreite		11	100
NNE	NNE1	60-5000Hz	dB		
	NNE2	60-2000Hz	dB		
	NNE3	1000-5000Hz	dB		
CHNR	CHNR1	60-5000Hz	dB		
	CHNR2	60-2000Hz	dB		
	CHNR3	1000-5000Hz	dB		

Im folgenden wurden Signale $s(t)$ generiert, die sich aus einem Rauschanteil $r(t)$ und einer Sequenz von Deltafunktionen $D(t)$ zusammensetzen:

$$s(t) = D(t) + r(t) \tag{6.1}$$

Außerdem wurden Periodenlängenschwankungen (Jitter J in Prozent) und Amplituden-

schwankungen (Shimmer S in Prozent) wie folgt eingeführt:

$$D(t) = \sum_i (1 + \frac{S}{100\%} z_{1i})\, \delta(t - t_i), \quad t_i - t_{i-1} = (1 + \frac{J}{100\%} z_{2i})\, T \tag{6.2}$$

mit der mittleren Periodendauer T, der Deltafunktion δ und den beiden normalverteilten Zufallszahlen z_{1i} und z_{2i} mit der Standardabweichung $\sigma = 1$. Der Wertebereich der Zufallszahlen wurde auf $(-3\sigma \leq z_{1i}, z_{2i} \leq 3\sigma)$ beschränkt, damit bei ca. 30% Jitter keine Veränderung der Pulsreihenfolge auftritt.

Das Signal $D(t)$ wurde mit einer Abtastfrequenz von 200kHz generiert, um eine ausreichende zeitliche Auflösung für kleine Jitter-Werte zu erreichen. Dann wurde $D(t)$ tiefpassgefiltert (30. Ordnung 4.5kHz Grenzfrequenz) und auf 10kHz unterabgetastet. Der Rauschanteil $r(t)$ ist eine Folge von gleichverteilten Zufallszahlen mit 10kHz Abtastfrequenz. Die im Folgenden analysierten Signale $s(t)$ sind jeweils 1s (10000 Abtastwerte) lang.

Die folgenden Experimente wurden jeweils mit verschiedenen mittleren Grundfrequenzen durchgeführt: 100Hz, 150Hz, 200Hz, 250Hz und 300Hz. Da numerische Analysen bisweilen zu *günstig* verlaufen, wenn die Periodenlänge ein ganzzahliges Vielfaches der Abtastfrequenz bzw. ein ganzzahliger Teiler der Signallänge ist, wurden nicht einfach die genannten Grundfrequenzen verwendet, sondern es wurde jeweils eine zufällige mittlere Grundfrequenz aus einem Bereich von ± 5% um die oben genannten Grundfrequenzen gewählt. Das bedeutet z.B., dass Grundfrequenzen, die in folgenden Abbildungen mit 100Hz bezeichnet sind Werte von 95Hz bis 105Hz annehmen können (entsprechend bei den anderen Grundfrequenzen).

6.1. Abhängigkeit vom Rauschpegel

Da die drei Maße GNE, NNE und CHNR entwickelt wurden, um die Pegelverhältnisse von Signal und Rauschanteil zu messen, wird im folgenden die Abhängigkeit des GNE, NNE und CHNR vom Rauschpegel untersucht. Dazu wurde der Pegel der Pulsfolge auf 0dB normiert. Dann wurde Rauschen mit einem Pegel, der zufällig aus dem Bereich von -50dB bis 20dB gewählt wurde, addiert. Jeweils 1000 Signale pro Messgröße und Grundfrequenz wurden generiert und vermessen. In den folgenden Abbildungen steht jeder Punkt für den Messwert eines 1s langen Signals. Jitter und Shimmer wurden für jedes Signal zufällig auf einen sehr kleinen Wert in dem Intervall 0,001% bis 0,01% (logarithmisch gleichverteilt) gesetzt.

Abbildung 6.1 zeigt NNE und CHNR, die jeweils in einem Frequenzbereich von 50Hz bis 5000Hz gemessen wurden. Die Messungen für die Frequenzbereiche 50Hz bis 2000Hz und 1000Hz bis 5000Hz sind nicht dargestellt, da die Ergebnisse fast identisch sind. In Abbildung 6.1 ist zu erkennen, dass sowohl NNE als auch CHNR ein monotones Maß für den Rauschpegel sind. Beim CHNR oben in der Abbildung nimmt die Streuung des Messwertes mit fallendem Rauschpegel zu. Beim NNE treten einige Ausreißer auf. Die Ursache hierfür konnte nicht geklärt werden. Sieht man von diesen beiden Punkten ab,

so können beide Maße als sinnvolle Größen zur Quantifizierung des Rauschanteils in einem Bereich von ca. -30dB bis 0dB angesehen werden.

In Abbildung 6.2 ist jeweils 1-GNE in logarithmischer Darstellung aufgetragen. GNE zeigt abhängig von der Bandbreite in Abbildung 6.2 unterschiedliche minimale Rauschpegel, für die der GNE noch sensitiv ist. Unterschreitet der Rauschpegel diesen minimalen Pegel, so erhöht sich der GNE nicht mehr. Dieser Pegel liegt bei ca. -50dB (1000Hz Bandbreite), bei ca. -35dB (2000Hz Bandbreite) und bei ca. -25dB (3000Hz Bandbreite). Bei allen drei GNE wird ab ca. 5-7dB Rauschpegel der maximale Korrelationswert nicht mehr größer (1-GNE nicht mehr kleiner). Beim GNE mit 1000Hz Bandbreite ist bei einem festen Rauschpegel eine deutliche Variation mit der Grundfrequenz zu erkennen. Diese Variation wird für 2000Hz und 3000Hz Bandbreite jeweils kleiner. Speziell bei 3000Hz Bandbreite verhält sich GNE monoton zum Rauschpegel im Bereich von -25dB bis 7dB und ist dabei weitgehend unabhängig von der Grundfrequenz und zeigt auch keine Ausreißer.

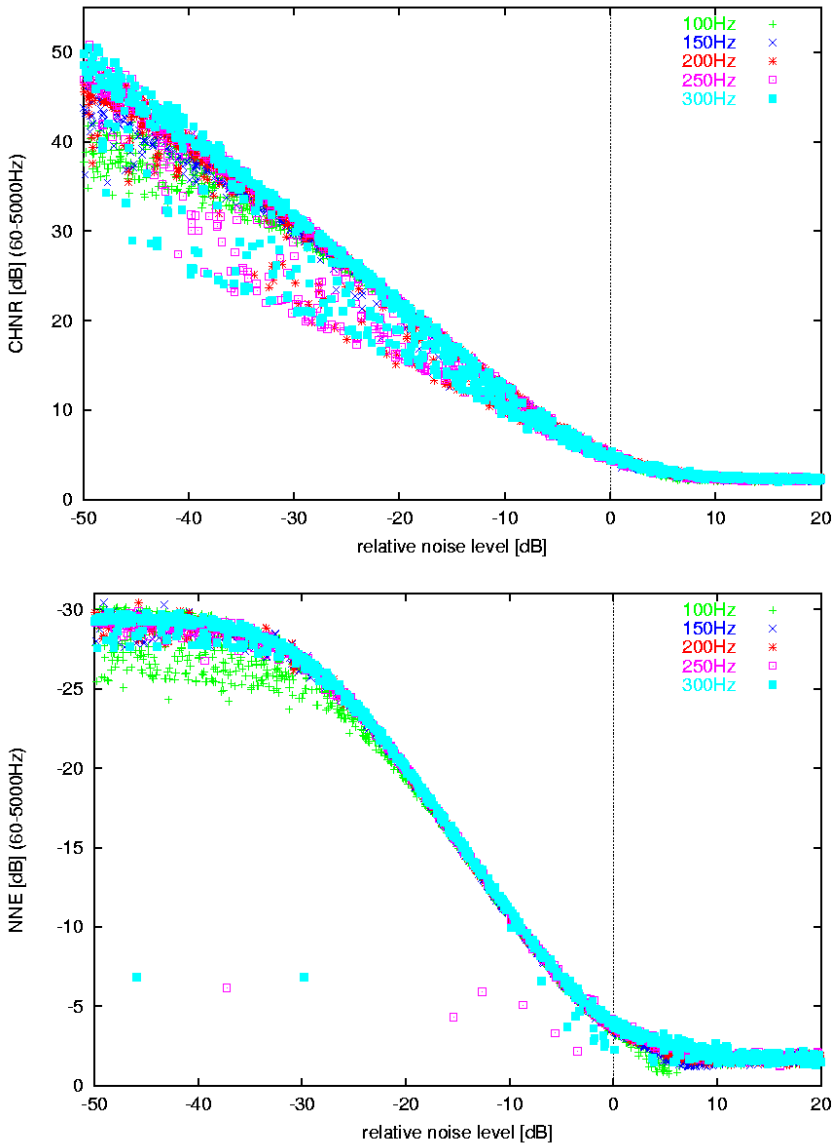


Abbildung 6.1.: Abhängigkeit des CHNR und des NNE vom Rauschanteil

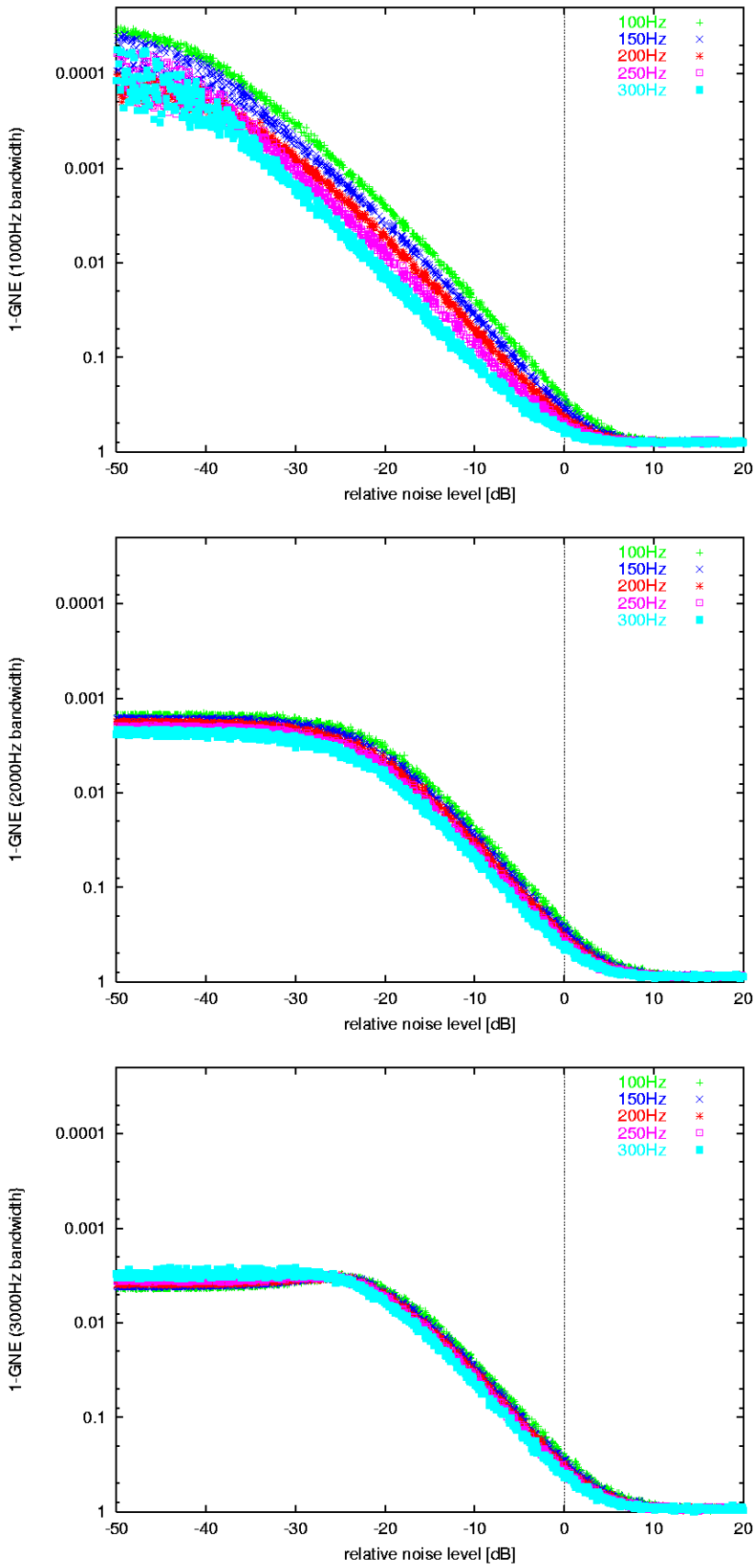


Abbildung 6.2.: Abhängigkeit des GNE vom Rauschanteil

6.2. Abhängigkeit vom Jitter

Abbildung 6.3 zeigt den gemessenen Zusammenhang von NNE bzw. CHNR und Jitter. Der Rauschpegel der Testsignale wurde zufällig auf -50dB bis -49dB gesetzt. Shimmer wurde zufällig auf 0,0001% bis 0,001% gesetzt.

NNE und CHNR zeigen eine sehr starke Jitterabhängigkeit: bereits bei 1% Jitter (1% Jitter wird auch bei Normalstimmen gemessen) sind die Maße so stark abgefallen wie bei den vorigen Test bei einem Rauschpegel von ca. -10dB bis -5dB. Die Jitterabhängigkeit nimmt mit kleineren Grundfrequenzen deutlich zu. Sie ist bei 100Hz und ca. 0,5% Jitter ca. 10dB stärker als bei 300Hz Grundfrequenz.

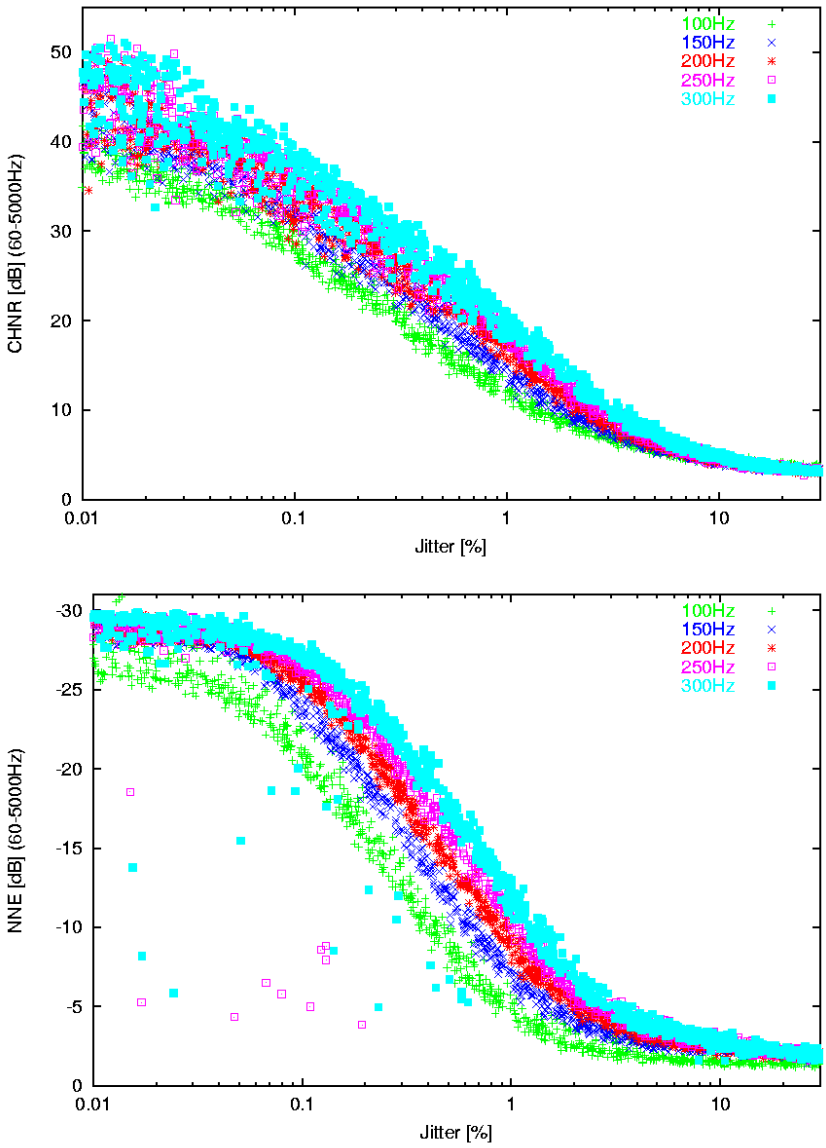


Abbildung 6.3.: Abhängigkeit des CHNR und des NNE vom Jitter

In Abbildung 6.4 nimmt GNE bei 1000Hz Bandbreite oberhalb von 1% Jitter deutlich, mit der Grundfrequenz stärker werdend, ab. Diese Abhängigkeit ist bei 2000Hz Bandbreite und erst recht bei 3000Hz Bandbreite kaum noch erkennbar. Der Grund für diese Abhängigkeit liegt darin, dass schnell aufeinander folgende Pulse bei zu kleiner Bandbreite zu sehr verschiedenen Einhüllenden in verschiedenen Kanälen führen (siehe auch Anhang B). GNE ist also bei den betrachteten Grundfrequenzen bei 2000Hz und 3000Hz Bandbreite weitgehend robust gegenüber Jitter.

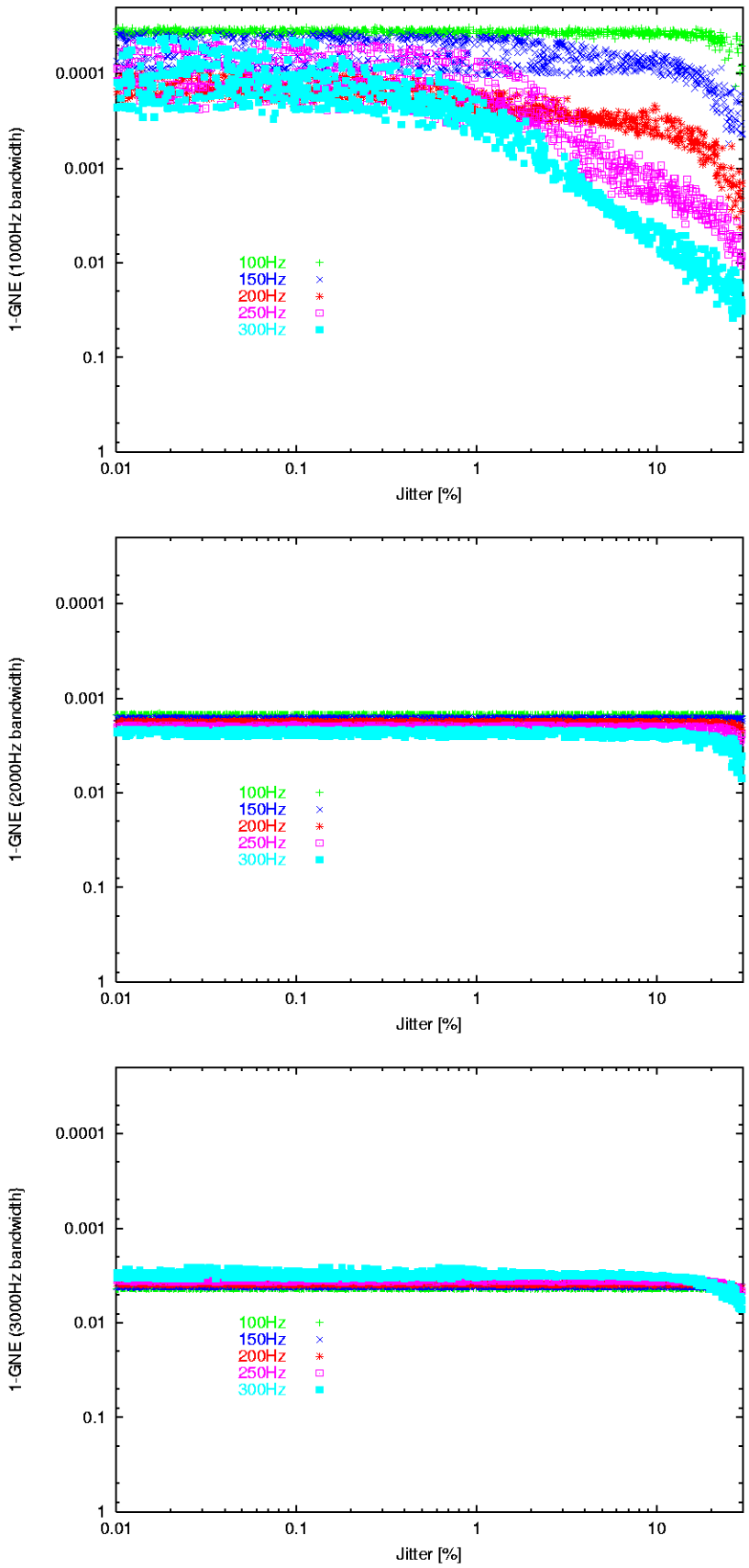


Abbildung 6.4.: Abhängigkeit des GNE vom Jitter

6.3. Abhängigkeit vom Shimmer

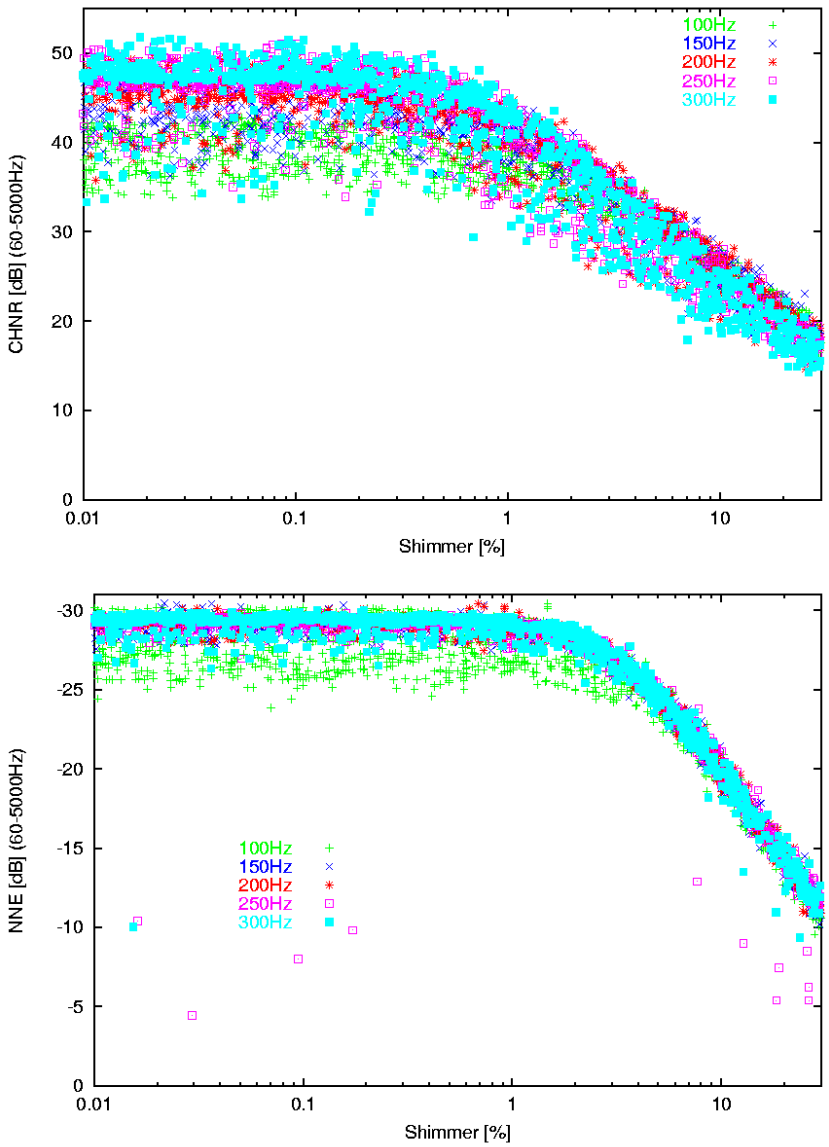


Abbildung 6.5.: Abhängigkeit des CHNR (oben) und des NNE (unten) vom Shimmer

Die Shimmer-Abhängigkeit von NNE und CHNR in Abbildung 6.5 ist deutlich geringer als beim Jitter, aber dennoch gut sichtbar. In Anhang C werden die spektralen Konsequenzen des Shimmers hergeleitet, aus denen sich die Shimmerabhängigkeit von NNE und CHNR erklären lässt.

In Abbildung 6.6 zeigt GNE bei allen Bandbreiten keine Abhängigkeit vom Shimmer.

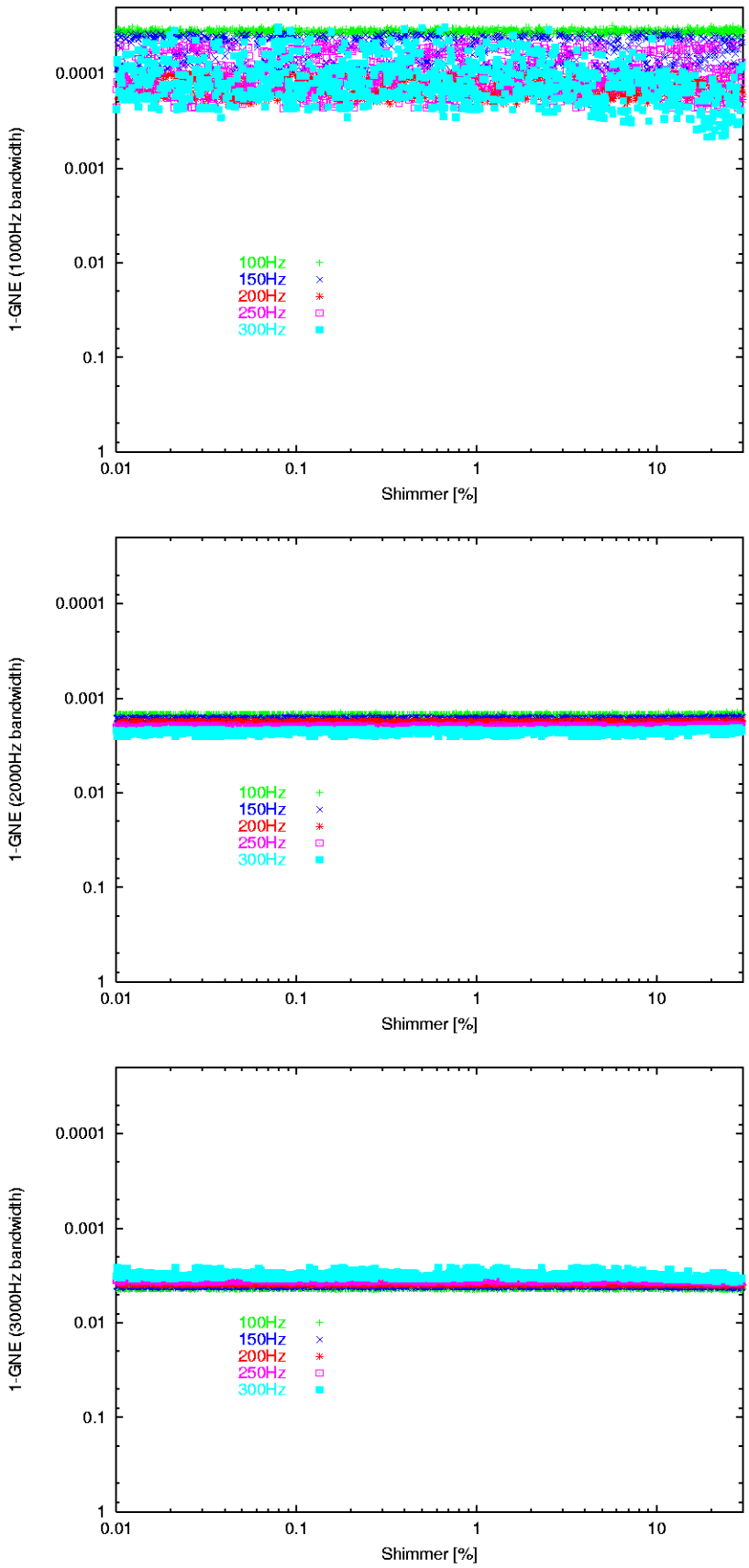


Abbildung 6.6.: Abhängigkeit des GNE vom Shimmer

7. Analyse des Datenraumes der akustischen Stimmgütemaße

7.1. Statistische Methoden

Nachdem bisher der GNE vorgestellt und mit NNE und CHNR verglichen wurde, stellt sich die Frage, wie dieses neue Maß am besten mit anderen akustischen Stimmgütemaßen kombiniert werden soll [80]. Um alle Aspekte der Stimmgüte zu erfassen, reicht ein Maß bestimmt nicht aus. Der GNE soll hauptsächlich für den im Signal vorhandenen Rauschanteil sensitiv sein. Es gibt andere Maße, die genau das auch sollen, wie z.B. NNE und CHNR, die jeweils noch in verschiedenen Frequenzbereichen gemessen werden können. Darüber hinaus existieren viele weitere Maße zur Quantifizierung des Signal-Rauschabstandes (z.B. die MDVP Maße NHR, VTI, SPI).

Die Maße die die Unregelmäßigkeit des Signals in Bezug auf Amplitude, Periodenlänge und Signalform beschreiben, bilden eine weitere große Gruppe, die Perturbationsmaße. Auch hier sind sehr zahlreiche Definitionen und Berechnungsverfahren in der Literatur beschrieben.

Die Maße, die statistische Aussagen über den Verlauf der Periodenlänge machen, wie Mittelwert, Minimum, Maximum und Standardabweichung, können in einer dritten Gruppe akustischer Maße zusammengefasst werden.

Im folgenden werden die wechselseitigen Beziehungen zwischen Maßen aus diesen drei Gruppen untersucht. Dazu werden die Maße jeweils für eine Gruppe von Patienten oder Normalstimmen berechnet. Die Werte der Maße für jeden Probanden (bzw. sogar für jedes analysierte Segment, je nach Fragestellung) bilden den Datenraum, der dann genauer analysiert werden soll.

Da die Anzahl der veröffentlichten Stimmgütemaße inzwischen sehr groß geworden ist, musste eine sinnvolle Vorauswahl getroffen werden. Aus der Gruppe der Maße zur Bestimmung des Signal-Rauschverhältnisses werden der hier neu vorgestellte GNE sowie NNE und CHNR (in verschiedenen Frequenzbereichen) verglichen. Die letzten beiden werden in der neueren Literatur oft verwendet und stellen jeweils Weiterentwicklungen und Verbesserungen vorangegangener Maße dar. Deshalb brauchen deren Vorgänger nicht noch einmal betrachtet zu werden.

In der Gruppe der Perturbationsmaße (Jitter und Shimmer) existieren ebenfalls zahlreiche verschiedene Methoden zur Quantifizierung. Die Unterschiede in den Definitionen sind oft gering und ein Versuch der Vereinheitlichung ist in [104] unternommen wor-

den. Die Anzahl der Quantifizierungsmethoden muss noch mit der Zahl der Methoden zur Bestimmung der einzelnen Periodenlängen multipliziert werden, um alle möglichen Perturbationsmaße zu erhalten. Einen weiteren Faktor zwei in der Anzahl bekommt man, wenn man diese Methoden jeweils am Mikrofonsignal und am EGG anwendet. Ein kompletter Vergleich ist deshalb ein aussichtsloses Unterfangen, eine Vorauswahl muss getroffen werden.

Aus der Gruppe der Perturbationsmaße werden deshalb hier nur Jitter und Shimmer Maße verglichen, die aus den Gleichungen 3.12 und 3.11 berechnet werden. Dabei findet bei der Datenraumanalyse nur die Waveform-Matching-Methode zur Periodenlängenberechnung Anwendung, da sich diese Methode schon an anderen Stellen als robust und zuverlässig herausgestellt hat [101, 147].

Im Rest dieses Abschnitts werden kurz die statistischen Methoden zusammengefasst, die zur Datenraumanalyse verwendet werden.

7.2. Korrelationen

7.2.1. Pearson's r

Einfache (lineare) Korrelation zwischen zwei Datensätzen (Skalarprodukt der Mittelwertbefreiten und normierten Datensätze). Die Implementierung und die Berechnung des Signifikanzniveaus erfolgte mit Hilfe der Numerical Recipes [106].

7.2.2. Spearmans Rangkorrelationen

Hier werden die Daten durch ihre Ränge ersetzt und dafür Pearson's r berechnet. Die Rangkorrelationen werden verwendet, falls nicht von einer Normal- oder Gleichverteilung ausgegangen werden kann. Insbesondere ist die Rangkorrelation robust gegen Ausreißer. Bei großer Anzahl von Datenpunkten sind Rangkorrelation und lineare Korrelation nahezu gleichwertig.

7.2.3. Korrektur nach Bonferoni und Holm

Führt man N Tests mit einer Irrtumswahrscheinlichkeit p durch, so ist die gesamte Irrtumswahrscheinlichkeit Np. Führt man also z.B. 20 Test mit $p \leq 0,05$ durch, so ist fast mit Sicherheit ein Irrtum vorhanden. Dieses Problem wurde zuerst von Bonferoni korrigiert, indem er ein neues $p_{\text{neu}} = p/N$ für alle Tests brerechnete und später mit einer weniger konservativen, aber dennoch exakten Methode (Methode der sequentiellen Verwerfung [46]) von Holm verfeinert.

Diese Korrektur findet in dieser Arbeit immer dann Anwendung, wenn mehrere Tests auf demselben Datenmaterial berechnet werden. Die Anzahl der Tests N wird jeweils angegeben. Das multiple Signifikanzniveau wird stets auf $p \leq 0,05$ gesetzt.

7.3. Singulärwertzerlegung SVD

Jede Matrix A mit m Zeilen und n Spalten ($m \geq n$, Rang von $A = n$) lässt sich eindeutig in das Produkt dreier Matrizen U (m Zeilen und n Spalten), D (Diagonalmatrix mit n Zeilen und Spalten) und V (mit n Zeilen und Spalten) zerlegen

$$A = UDV. \tag{7.1}$$

Dabei sind die Spalten von U und die Zeilen von V orthonormal. Gleichung 7.1 kann als Koordinatentransformation in ein Hauptachsensystem gelesen werden, wobei die neuen Koordinaten in U stehen, die Basisvektoren in V und die Varianz in den Hauptachsen proportional zu den Quadraten der entsprechenden Diagonalelement von D ist.

Gleichung 7.1 ist aber auch zur Datenanalyse nützlich: Stehen in den Zeilen von A die n normalisierten Datensätze mit je m Daten pro Datensatz, so kann man an den Werten der Diagonalelemente von D die *tatsächliche* Dimension des Datenraumes ablesen, indem man nur Hauptrichtungen bis zu einem gewissen kleiner Schwellenwert der Varianz berücksichtigt. Außerdem kann man an den Basisvektoren die Bedeutung einzelner Daten für eine bestimmte Hauptrichtung ablesen. In dieser Arbeit wird die SVD auf Datensätze von mehreren akustischen Stimmgütemesswerten pro Stimmsegment (bzw. pro Patient) angewandt.

7.4. Relativer Informationszuwachs

Der Nachteil der SVD-Datenraumanalyse ist, dass es sich un eine lineare Methode handelt, die im strengen Sinne nur für normal- oder gleichverteilte Daten gültig ist. Ist die Verteilung der Daten unbekannt oder weist die Verteilung viele Ausreißer auf, so werden die Ergebnisse der SVD unzuverlässig. Es würde z.B. die SVD eines zweidimensionalen Raums, bei dem die Datenpunkte auf einer Kreislinie gleichverteilt sind, einen zweidimensionalen Datenraum anzeigen, obwohl die Daten nur auf einem eindimensionalen Unterraum (einer Kreislinie) verteilt sind. Hier bietet die Informationstheorie eine Alternative, die auf der Transinformation bzw. auf deren n-dimensionalen Verallgemeinerung beruht [78].

Verallgemeinerung der Transinformation

Wir gehen zunächst von zwei (akustischen Stimmgüte-)Messgrößen x_1 und x_2 aus. Diese Größen seien an einer großen Zahl N von Probanden bzw. Stimmsegmenten gemessen worden, so dass genügend Daten vorliegen, um eine Wahrscheinlichkeitsverteilung zu berechnen. Dazu wird der Wertebereich von x_1 und x_2 in M gleichgroße Intervalle aufgeteilt. Sei $n_i(x_1)$ die Anzahl der Messwerte von x_1 im Intervall i und entsprechend $n_j(x_2)$. Sei weiterhin $n_{ij}(x_1, x_2)$ die Anzahl von Datensätze deren x_1 Messwert im Intervall i und deren x_2 Messwert im Intervall j liegt, so können wir nun die empirischen,

diskreten Häufigkeitsverteilungen $p_i(x_1)$, $p_j(x_2)$ und $p_{ij}(x_1, x_2)$ berechnen:

$$\begin{aligned} p_i(x_1) &= \frac{n_i(x_1)}{N} \\ p_j(x_2) &= \frac{n_i(x_2)}{N} \qquad\qquad i,j = 1, \cdots, M \\ p_{ij}(x_1, x_2) &= \frac{n_{ij}(x_1, x_2)}{N^2} \end{aligned} \tag{7.2}$$

Daraus lassen sich sofort die empirischen Entropien $H(x_1)$, $H(x_2)$ und $H(x_1, x_2)$ berechnen:

$$H(x_1) = -\sum_{i=1}^{M} p_i(x_1)\mathrm{ld}(p_i(x_1)) \tag{7.3}$$

$$H(x_2) = -\sum_{j=1}^{M} p_j(x_2)\mathrm{ld}(p_j(x_2)) \tag{7.4}$$

$$H(x_1, x_2) = -\sum_{i,j=1}^{M} p_{ij}(x_1, x_2)\mathrm{ld}(p_{ij}(x_1, x_2)) \tag{7.5}$$

wobei ld für den Logarithmus zur Basis Zwei steht.

Mit Hilfe dieser Gleichungen 7.3, 7.4 und 7.5 kann man die Transinformation $I_2 = I(x_1, x_2)$ folgendermaßen schreiben (siehe z.B. [23]):

$$I_2 = I(x_1, x_2) = H(x_1) + H(x_2) - H(x_1, x_2) \qquad [\mathrm{Bit}]. \tag{7.6}$$

I_2 kann man anschaulich als die Anzahl der Bits interpretieren, die man (abhängig von M) über x_2 vorhersagen kann, wenn man x_1 kennt. Der Wertebereich von I_2 reicht von 0 Bit (kein Bit kann vorausgesagt werden, die Daten x_1 und x_2 sind unabhängig) bis $\mathrm{ld}(M)$ Bit (alle Bits können vorausgesagt werden, die Daten x_1 und x_2 sind informationstheoretisch identisch).

Die Transinformation kann leicht auf μ-dimensionale Daten verallgemeinert werden:

$$I_\mu = I(x_1, .., x_\mu) = \sum_{i=1}^{\mu} H(x_i) - H(x_1, ..., x_\mu)\,. \tag{7.7}$$

Bei vorgegebener Zahl der Datensätze N und Dimension der Daten μ, muss man bei der Wahl der Anzahl der Intervalle M einen Kompromiss zwischen hoher Auflösung der Häufigkeitsverteilungen, entsprechend großen Werten von M, und der Anzahl der zur Verfügung stehenden Datensätze N finden. Denn die gesamte Anzahl der μ-dimensionalen Intervalle M^μ sollte höchstens so groß sein wie die Anzahl der Datensätze, um durchschnittlich mindesten einen Datenpunkt pro μ-dimensionalem Intervall zu garantieren.

Normalisierter Informationszuwachs ΔI_N

Wird zu z.B. drei Messgrößen x_1, x_2, x_3 eine vierte Messgröße hinzugenommen, so werden sich i.A. nicht alle Bits der neuen Messgröße aus den vorherigen vorhersagen lassen. Je weniger Bit sich vorhersagen lassen, umso mehr zusätzliche Information steckt in der vierten Messgröße.

$I_4 - I_3$ misst nun die zusätzliche Transinformation durch den vierten Messwert. Bildet man die Differenz von $B = \mathrm{ld}(M)$ und $I_4 - I_3$, und bezieht diese auf B, so erhält man eine Zahl zwischen 0 und 1, die den Informationsgewinn durch den zusätzlichen Messwert anzeigt.

$$\Delta I_N = \frac{B - (I_4 - I_3)}{B}\,. \tag{7.8}$$

Dabei bedeutet $\Delta I_N = 0$ keine zusätzliche Information (z.B. wenn man probeweise eine der drei Messgrößen nocheinmal hinzufügt) und $\Delta I_N = 1$ den größtmöglichen Informationszuwachs.

7.5. SVD mit 20 Messgrößen

Nach der Übersicht über die statistischen Methoden wird nun in diesem Abschnitt der akustische Datenraum analysiert. Dieser wird zuerst einmal aus den folgenden fünf Klassen verschiedener Maße gebildet: Grundperiode, mittlere Periodenkorrelation, Jitter, Shimmer, GNE. Von den drei Maßen Grundperiode, mittlere Periodenkorrelation und GNE werden neben dem Mittelwert weitere Kenngrößen - nämlich Minimum, Maximum und Standardabweichung- berechnet. Diese erste Analyse soll zeigen, welche dieser Kenngrößen zur Beschreibung der Pathologie realer Stimmen am wichtigsten sind.

In Tabelle 7.1 sind die Maße aufgelistet, die von 1799 Aufnahmen des Vokals [ɛ:] von pathologischen und normalen Stimmen berechnet wurden (Gruppe 6 aus Tabelle 4.1). Die Daten wurden wie in der Tabelle angegeben transformiert, mittelwertbefreit und durch die Standardabweichung dividiert. Die SVD ergab bei einem Schwellenwert von 5% einen vierdimensionalen Datenraum mit den Varianzen 60%, 16%, 9% und 6%.

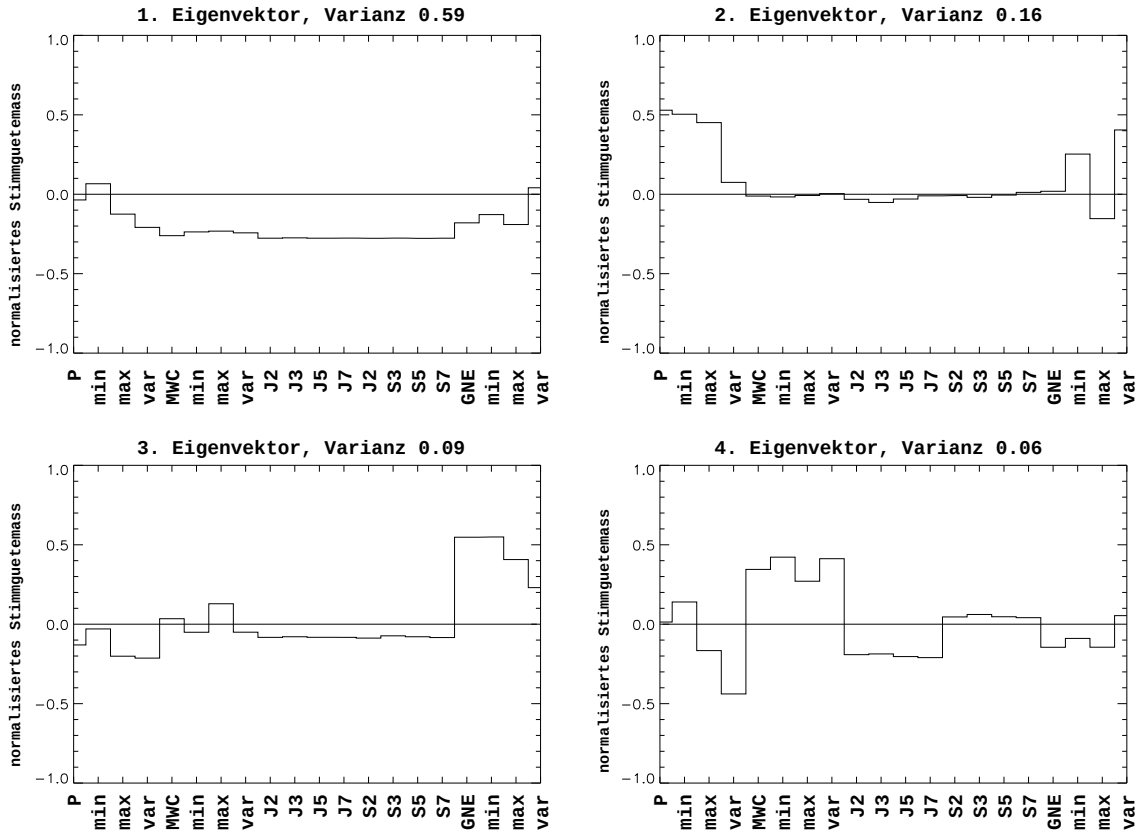


Abbildung 7.1.: Die vier Hauptrichtungen der 20-dimensionalen SVD

In Abbildung 7.1 sind die Komponenten der ersten vier Hauptrichtungen dargestellt. In der ersten Hauptrichtung mit 60% Varianz sind die Jitter-, Shimmer- und Korrelationskomponenten etwa gleichstark. Etwas geringer sind die Beträge der Varianz der Grundperiode und der GNE-Werte (Mittelwert, Minimum, Maximum). Die Varianz des

GNE, die mittlere Grundperiode, deren Minimum und Maximum haben nur kleine Beträge in der ersten Hauptrichtung. In der zweiten Hauptrichtung sind die Grundperiodenlänge sowie deren Minimum und Maximum dominant. Ebenso haben die Komponenten des minimalen GNE sowie dessen Varianz große Beträge. Die dritte Hauptrichtung mit 9% Varianz wird vom GNE, dessen Minimum, Maximum und Varianz bestimmt. In der vierten Hauptrichtung dominieren die Varianz der Grundfrequenz, die Korrelationsmaße sowie die Jitter-Maße.

An einer Darstellung der Daten im Raum der ersten beiden Hauptrichtungen in Abbildung 7.2 ist zu erkennen, dass sich in der zweiten Hauptrichtung bei den Normalstimmen zwei Gruppen bilden. Diese Gruppen stellen sich als eine Männer- und eine Frauengruppe heraus. Aufgrund der unterschiedlichen Grundperioden werden in der zweiten Hauptrichtung Männer von Frauen getrennt. Weiterhin zeigt sich, dass die auf der linken Seite liegenden pathologischen Stimmen gerade diejenigen mit den schwersten Stimmstörungen sind (Aphonie, Lähmungen, schwere Kehlkopfoperationen). Auf der rechten Seite liegen Normalstimmen und Stimmen mit leichten pathologischen Veränderungen.

Tabelle 7.1.: Die 20 Messgrößen zur SVD-Analyse

Nr.	Name	Transformation	Kürzel
1	Grundperiode	-	P
2	Minimum	-	min
3	Maximum	-	max
4	Varianz	-	var
5	mittlerer Kurzzeitkreuzkorrelationswert (mean waveform-matching coefficient)	log 1 - ( )	MWC
6	Minimum	log 1 - ( )	min
7	Maximum	log 1 - ( )	max
8	Varianz	log( )	var
9	Jitter im PF Maß	log( )	J2
10	Jitter im PQ3 Maß	log( )	J3
11	Jitter im PQ5 Maß	log( )	J5
12	Jitter im PQ7 Maß	log( )	J7
13	Shimmer im PF Maß	log( )	S2
14	Shimmer im PQ3 Maß	log( )	S3
15	Shimmer im PQ5 Maß	log( )	S5
16	Shimmer im PQ7 Maß	log( )	S7
17	GNE: Mittlerer Korrelationswert der kanalweisen Hilberteinhüllenden. Bis 200 Hz Grundfrequenz: n (Zahl der Einhüllenden)=51 shift (Abstände der Mittenfrequenzen)=80Hz skip (minimaler Mittenfrequenzabstand für Korrelation)=560Hz; ab 200 Hz: n=31 shift=100Hz skip=1000Hz	1 - ( )	GNE
18	Minimum	1 - ( )	min
19	Maximum	1 - ( )	max
20	Varianz	-	var

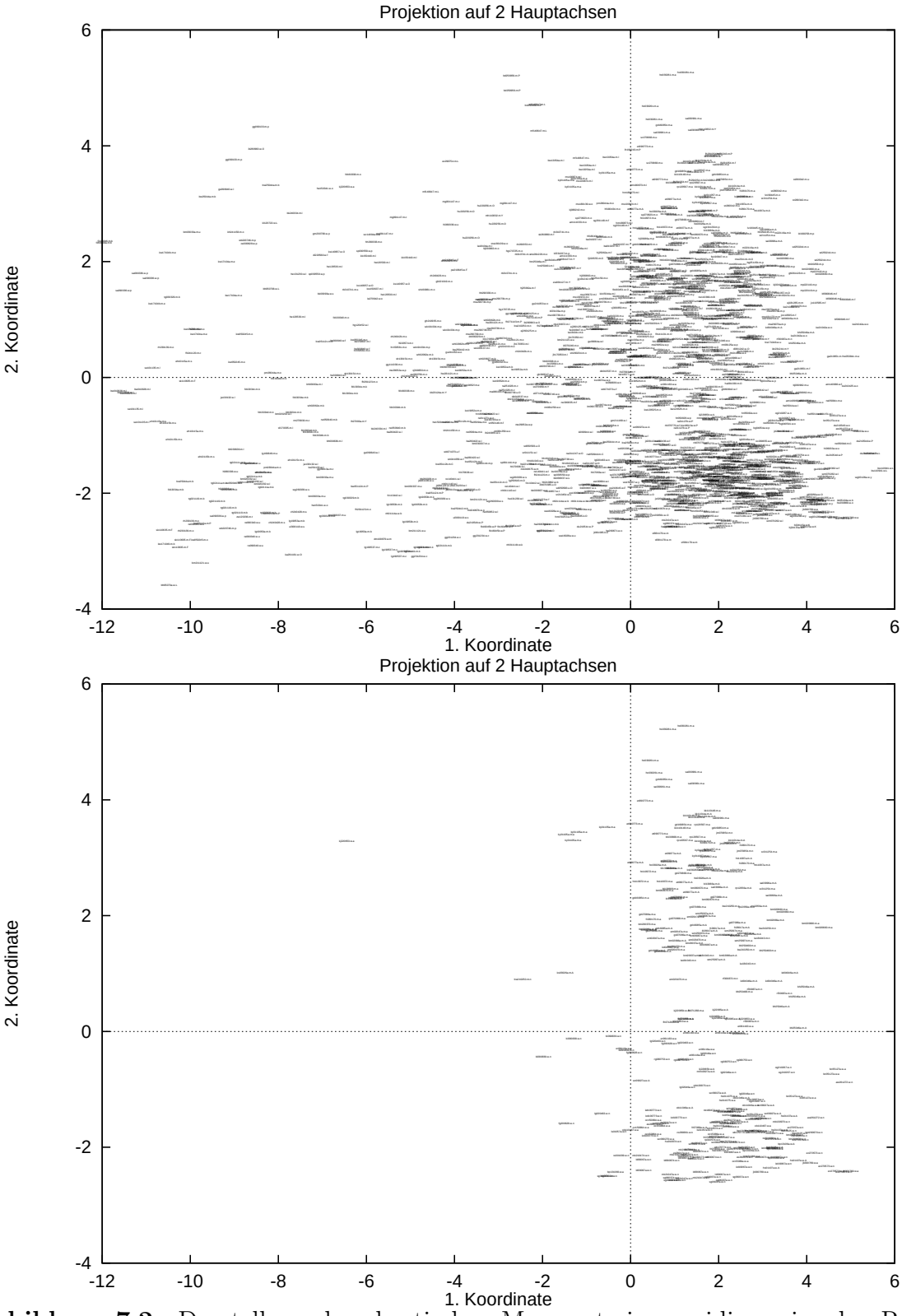


Abbildung 7.2.: Darstellung der akustischen Messwerte im zweidimensionalen Raum der ersten beiden Hauptrichtungen. Oben: pathologische Stimmen, unten Normalstimmen.

7.6. SVD mit vier Messgrößen

Da es hier um die Entwicklung neuer, aussagekräftiger Darstellungen der Stimmgüte gehen soll [79, 85, 86], ist der hohe Anteil der Grundperiodenlänge an der zweitwichtigsten Hauptachse nicht wünschenswert, denn die Verschiedenheit der Grundperioden bei Männern und Frauen gestattet keine Aussage über den Grad der Stimmstörung. Deshalb werden die Grundperiodenmaße im Folgenden weggelassen. Da die Jitter und Shimmermaße jeweils mit fast gleichen Beiträgen der Komponenten zu den Hauptachsen beitragen, reicht es aus, nur ein Jitter- und nur ein Shimmermaß zu betrachten. Gleiches gilt für die Korrelationsmaße: Hier wird im Folgenden nur noch der Mittelwert verwendet, Minimum, Maximum und Varianz werden ausgelassen. Weiterhin wird wegen der oben besprochenen Eigenschaften nur noch das Maximum des GNE betrachtet.

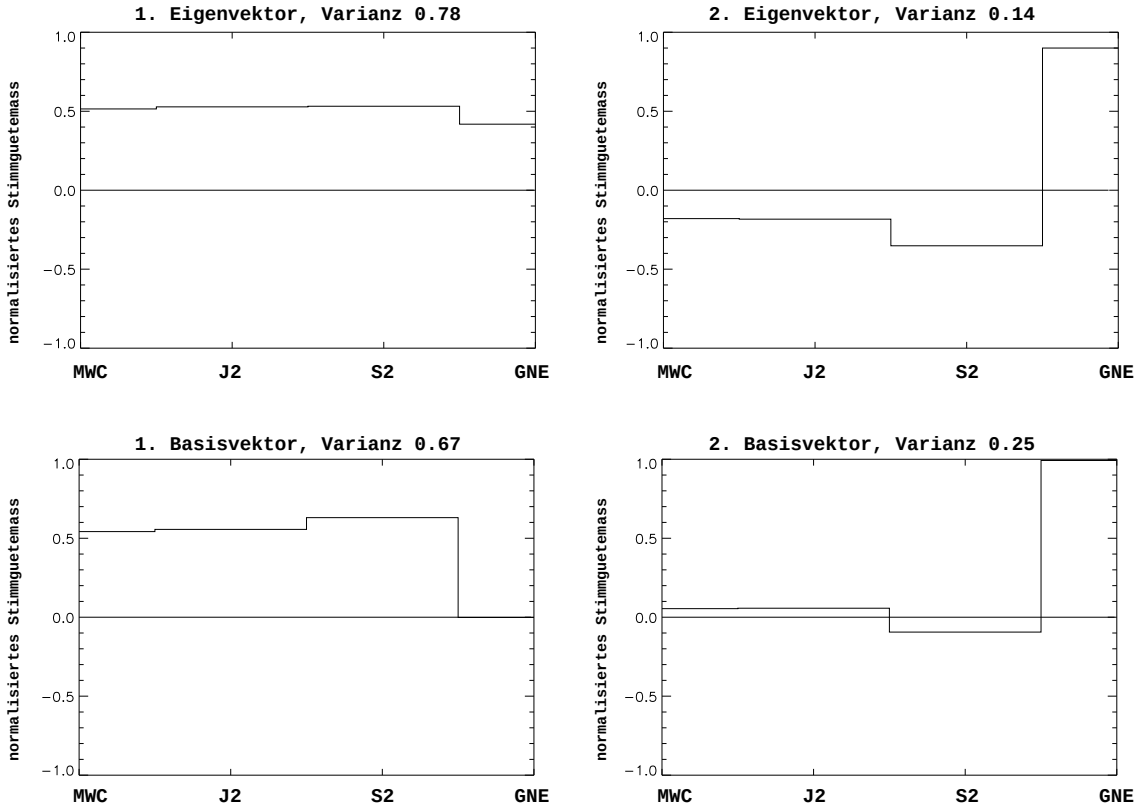


Abbildung 7.3.: Oben: Die ersten beiden Hauptrichtungen; unten: Nach Rotation des Koordinatensystems.

Es bleiben also vier Maße, auf die erneut die SVD angewandt wird. Die Wahl von J2 (entspricht PPF) und S2 (entspricht EPF) als Repräsentanten der Jitter- und Shimmermaße ist zunächst noch etwas willkürlich und wird später genauer untersucht werden.

Die Hauptachsentransformation ergibt, dass der Datenraum im Wesentlichen zweidimensional ist: In der ersten Hauptrichtung liegen 78% und in der zweiten liegen 14% der Varianz. In Abbildung 7.3 sind oben die ersten beiden Hauptvektoren abgebildet.

In Abbildung 7.4 sind die Datenpunkte aller pathologischen Sprecher im zweidimensionalen Unterraum dargestellt. Jeder Datenpunkt wird durch einen (bei dieser Vergrößerung nicht lesbaren) Bezeichner markiert. Auffällig ist die leicht abfallende Gerade, die die Datenpunkte nach unten begrenzt. Die Datenverteilung sieht wie ein leicht gedrehtes Rechteck aus. Man zeigt leicht, dass sich durch eine Rotation des zweidimensionalen Unterraums um den Winkel

$$\varphi = \arctan \frac{v_{2i}}{v_{1i}} \tag{7.9}$$

wobei v_{1i} die i-te Komponente des ersten Hauptvektors und v_{2i} die i-te Komponente des zweiten Hauptvektors ist, stets erreichen lässt, dass die i-te Komponente in der rotierten ersten Hauptrichtung verschwindet. Durch eine solche Rotation (25 Grad) entstehen die Basisvektoren in Abbildung 7.3 unten, bei denen die GNE-Komponente im ersten Basisvektor verschwindet. Die Basisvektoren zeigen jetzt nicht mehr in die Hauptrichtungen, sind jedoch immer noch orthogonal. Da man bei der Multiplikation von drei Matrizen beliebig klammern darf, ist es gleichwertig, ob erst die Daten in den zweidimensionalen Unterraum projiziert werden und dann die Rotation erfolgt, oder ob man gleich die Hauptrichtungsvektoren mit der Rotationsmatrix multipliziert.

Durch diese Rotation wird eine Trennung zwischen den Maßen erreicht: In der ersten Basisrichtung sind die Maße MWC, J2, S2 etwa gleich stark vertreten, in der zweiten Hauptrichtung bleibt nur der GNE übrig. Der zweidimensionale Unterraum der ersten beiden Hauptrichtungen wird deshalb von zwei orthogonalen Maßen aufgespannt: einer Linearkombination von MWC, J2, S2 und einer Lineartransformation des GNE. Diese zwei Maße sind die Grundlage des Heiserkeits-Diagramms. Bei diesem wird auch die erste Achse (die Irregularitätskomponente) aus dem Korrelationswert MWC, einem Jittermaß und einem Shimmermaß berechnet. Die zweite Achse ist auch eine Lineartransformation des GNE.

In Abbildung 7.5 sind in der oberen Abbildung die Messdaten von Normalstimmen (links unten) und Flüsterstimmen (rechts oben) zu sehen. Die beiden Gruppen sind sowohl in der horizontalen als auch in der vertikalen Richtung deutlich getrennt. Eine eindeutige Klassifikation ist möglich.

In Abbildung 7.4 unten fällt auf, dass die Datenpunkte in der vertikalen Richtung nicht normalverteilt sind, sie häufen sich im unteren Teil. Da diese Achse im Wesentlichen durch den GNE bestimmt wird, bedeutet das, dass der GNE nicht normalverteilt ist. Die Grundlage der SVD sind aber normal- oder gleichverteilte Daten. Deshalb wurde der GNE so transformiert, dass die Daten annähernd gleichverteilt sind ($\log(1 - \mathrm{GNE})$).

Das Ergebnis für Normal- und Flüsterstimmen nach SVD (mit den Daten der pathologischen Gruppe) und Rotation ist in Abbildung 7.5 unten zu sehen: Die Trennung in horizontaler Richtung hat sich nicht verändert. Der Abstand zwischen den Gruppen in der Vertikalen hat sich jedoch stark verkleinert. Da die zu entwickelnde Stimmgütedarstellung eine möglichst klare Trennung zwischen normalen und pathologischen Stimmen ermöglichen soll, wird deshalb auf die Transformation des GNE verzichtet, obwohl die Werte dann nicht gleichverteilt sind.

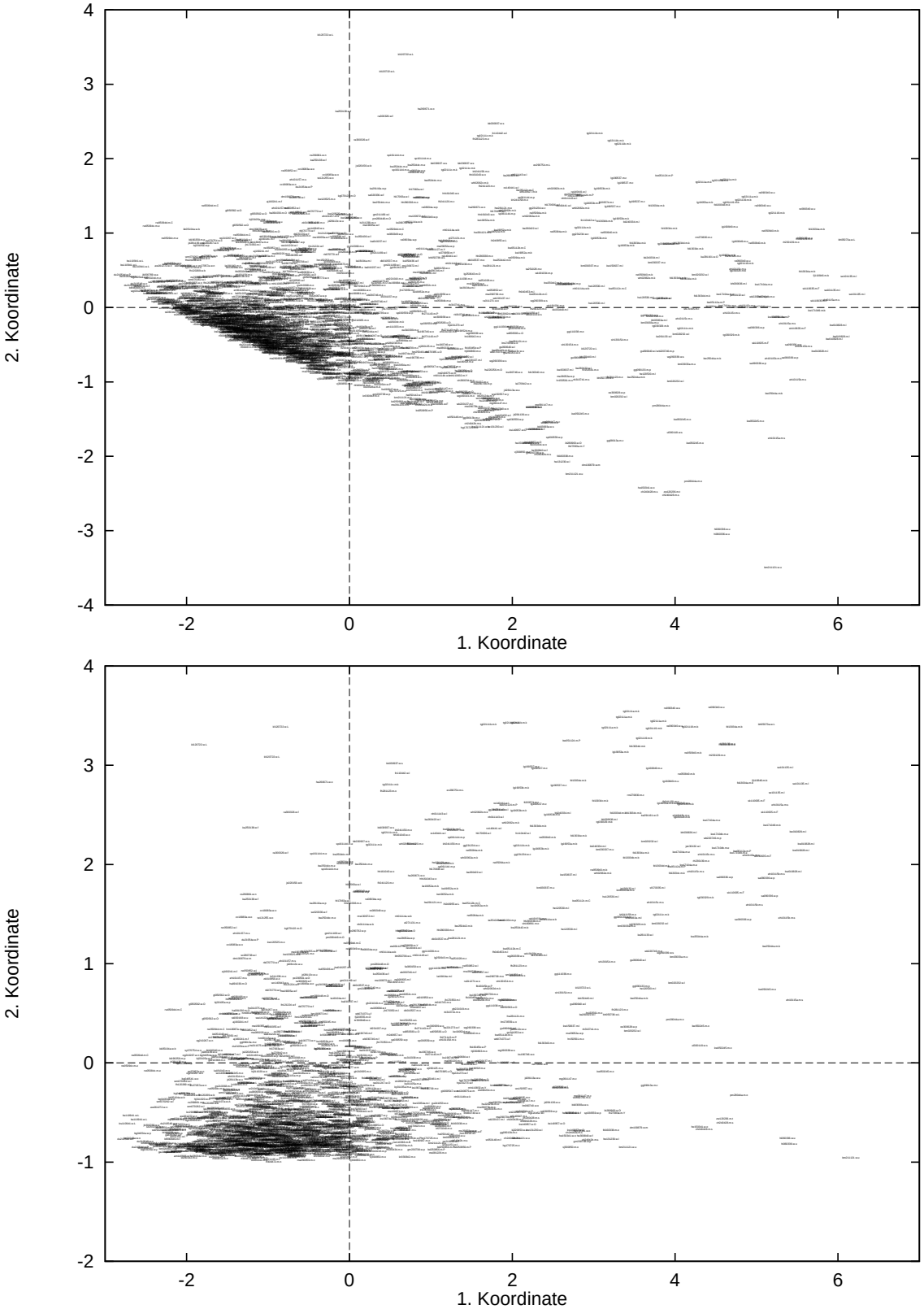


Abbildung 7.4.: Oben: Datenpunkte im Koordinatensystem der ersten beiden Hauptachsen; unten: nach Rotation des Koordinatensystems.

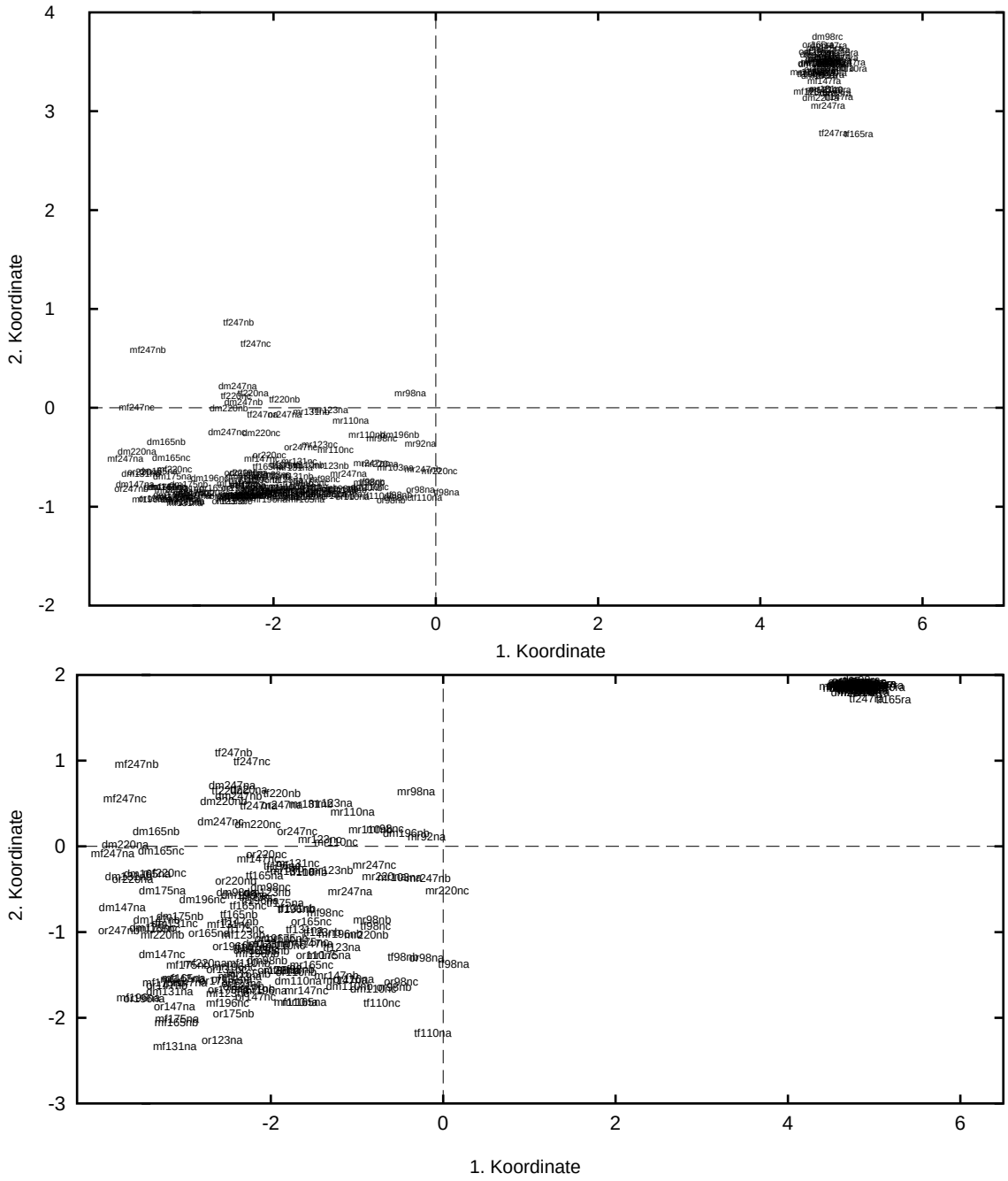


Abbildung 7.5.: Oben: Datenpunkte im Koordinatensystem der rotierten ersten beiden Hauptachsen. Normalstimmen liegen links unten,Flüsterstimmen rechts oben; unten: Darstellung nach SVD mit log(1-GNE). Varianzen: 1. Koordinate 67%, 2. Koordinate 25%.

7.7. Korrelationen zwischen akustischen Stimmgütemessgrößen

Neben den methodenimmanenten Abhängigkeiten zwischen verschiedenen Stimmgütemaßen (z.B. die oben gezeigte Jitter- und Shimmer-Abhängigkeit von NNE und CHNR) ist in der Praxis die empirische Abhängigkeit von Maßen wichtig. Für ein umfassendes Bild der Stimmgüte mit möglichst wenig Redundanz werden Maße benötigt, die einzeln aussagekräftig sind, aber untereinander so wenig Redundanz wie möglich zeigen. Im folgenden werden Korrelationswerte zwischen verschiedenen Maßen an einem repräsentativen Patientenkollektiv bestimmt, um die empirischen Abhängigkeiten zwischen den verschiedenen Maßen zu quantifizieren.

Insbesondere werden auch die Korrelationen von GNE, NNE und CHNR mit Jitter- und Shimmer-Maßen betrachtet, um festzustellen, ob sich die Ergebnisse der Messungen mit synthetischen Signalen auf reale Messungen übertragen lassen. Denn anders als bei den synthetischen Signalen müssen Jitter und Shimmer hier selbst erst gemessen werden und sind deshalb potentiell fehlerbehaftet. Die Frage wird sein, ob sich auch bei den realen Messungen bestätigt, dass der GNE im Gegensatz zum NNE und CHNR nicht von Jitter und Shimmer abhängig ist.

7.7.1. Datenmaterial und Diagnosen

Das Datenmaterial für diese Untersuchung wurde noch einmal nach strengeren Kriterien als bisher ausgewählt: es wurde von jedem Patienten (bzw. von jeder Normalstimme) genau eine Sekunde aus dem stabilen mittleren Teil des gehaltenen Vokals [ɛ:] analysiert. Von jeder Person wurde nur eine Aufnahme zugelassen. Die so gewählten Gruppen umfassten 88 Normalstimmen im Alter von 18 bis 90 Jahren (Mittelwert 47 Jahre) und 447 pathologische Stimmen. Das Patientenkollektiv bestand aus 447 Patienten mit unterschiedlichen Stimmstörungen im Alter von 10 bis 80 Jahren (Mittelwert 48 Jahre).

Die Diagnosen der pathologischen Sprecher sind in Tabelle 7.2 zusammengefasst.

Tabelle 7.2.: Liste der Stimmstörungen

Anzahl	Diagnose
56	Stimmlippenlähmung
22	Fixation der Stimmlippen
34	Zustand nach partieller Kehlkopfentfernung bei Krebs
21	Tumor vor der Operation
39	Polypen
16	Knötchen
19	Granulome
28	Zysten
12	Dysphonie bei Stimmbruch
11	Papillomae
24	Reinke-Ödeme
47	hypofunktionelle Dysphonie
45	Zustand nach Entfernung eines gutartigen Tumors
11	Laryngitis
62	andere (weniger als 5 Patienten pro Diagnose)

7.7.2. Akustische Maße und Transformationen

Tabelle 7.3.: Abkürzungen

Messgröße (x)	Symbol	Beschreibung	Einheit	Transformation y=f(x)
MWC	MWC			log (1-x)
Jitter	J2	PPF	%	log x
	J3	PPQ K=3	%	log x
	J5	PPQ K=5	%	log x
	J7	PPQ K=7	%	log x
	J11	PPQ K=11	%	log x
	J15	PPQ K=15	%	log x
Shimmer	S2	EPF	%	log x
	S3	EPQ K=3	%	log x
	S5	EPQ K=5	%	log x
	S7	EPQ K=7	%	log x
	S11	EPQ K=11	%	log x
	S15	EPQ K=15	%	log x
GNE	GNE1	1000Hz		log (1-x)
		Bandbreite		x
	GNE2	2000Hz		log (1-x)
		Bandbreite		x
	GNE3	3000Hz		log (1-x)
		Bandbreite		x
NNE	NNE1	60-5000Hz	dB	x
	NNE2	60-2000Hz	dB	x
	NNE3	1000-5000Hz	dB	x
CHNR	CHNR1	60-5000Hz	dB	x
	CHNR2	60-2000Hz	dB	x
	CHNR3	1000-5000Hz	dB	x

Da Mittelwerte und Standardabweichungen bei Normalverteilungen am Aussagekräftigsten sind, wurden die gemessenen Werte wie in Tabelle 7.3 angegeben transformiert, um bei dem jeweiligen Maß bezüglich der Gruppe mit pathologischen Sprechern annähernd eine Normalverteilung zu erreichen. Beim GNE wird durch die Transformation y=log(1-x) die Normalverteilung angenähert. In den folgenden Tabellen ist jedoch auch der nichttransformierte GNE angegeben, da sich im praktischen Einsatz gezeigt hat, dass der transformierte GNE (optisch) keine so gute Unterscheidung zwischen pathologischen und normalen Stimmen erlaubt wie der untransformierte Wert.

7.7.3. Mittelwerte und Standardabweichungen

In Tabelle 7.4 sind die Mittelwerte und Standardabweichungen der akustischen Maße von der pathologischen und der normalen Sprechergruppe zusammengefasst. Die Mittelwerte der Jittermessungen bei Normalstimmen (ca. $10^{-0,6}\% = 0,25\%$) und der Shimmermessungen (ca. $10^{0,4}\% = 2,5\%$) entsprechen den in der Literatur angegebenen Werten. In der letzten Spalte ist unter dem Namen „relative Differenz“ der Differenzbetrag der Mittelwerte von normaler und pathologischer Gruppe, geteilt durch die Standardabweichung der normalen Gruppe eingetragen. Bei allen Maßen sind die Mittelwerte von normaler und pathologischer Gruppe ca. eine Standardabweichung (der normalen Gruppe) voneinander entfernt. Das bedeutet, dass alle Maße eine Tendenz zur Unterscheidung von normalen und pathologischen Stimmen zeigen. Die Unterscheidbarkeit von Gruppen wird an anderer Stelle noch genauer untersucht werden.

Tabelle 7.4.: Mittelwerte und Standardabweichungen für normale und pathologische Sprecher. Die relative Differenz in der letzten Spalte ist der Betrag der Differenz der Mittelwerte der normalen und der pathologischen Gruppe, geteilt durch die Standardabweichung der normalen Gruppe.

		normal $n = 88$		pathologisch $n = 447$		
Symbol	Beschreibung	Mittelw.	Standard-abw.	Mittelw.	Standard-abw.	relative Differenz
MWC		−2,021	0,335	−1,614	0,574	1,21
J2	PPF	−0,492	0,234	−0,089	0,611	1,72
J3	PPQ K=3	−0,792	0,246	−0,374	0,645	1,70
J5	PPQ K=5	−0,734	0,203	−0,323	0,626	2,02
J7	PPQ K=7	−0,673	0,211	−0,276	0,610	1,88
J11	PPQ K=11	−0,588	0,206	−0,210	0,584	1,83
J15	PPQ K=15	−0,522	0,199	−0,161	0,563	1,81
S2	EPF	0,572	0,212	0,848	0,421	1,30
S3	EPQ K=3	0,268	0,224	0,550	0,424	1,26
S5	EPQ K=5	0,347	0,199	0,617	0,407	1,36
S7	EPQ K=7	0,403	0,204	0,662	0,398	1,27
S11	EPQ K=11	0,476	0,203	0,717	0,384	1,19
S15	EPQ K=15	0,531	0,204	0,757	0,368	1,11
GNE1	1000Hz	−1,612	0,291	−1,062	0,515	1,89
	Bandbreite	0,969	0,022	0,834	0,189	6,14
GNE2	2000Hz	−1,360	0,331	−0,870	0,485	1,48
	Bandbreite	0,940	0,056	0,768	0,222	3,07
GNE3	3000Hz	−1,120	0,345	−0,690	0,428	1,25
	Bandbreite	0,892	0,106	0,695	0,242	1,86
NNE1	60-5000Hz	−19,425	3,634	−16,025	5,853	0,94
NNE2	60-2000Hz	−22,831	3,606	−19,492	6,830	0,93
NNE3	1000-5000Hz	−11,715	3,734	−7,441	4,448	1,14
CHNR1	60-5000Hz	25,169	3,649	20,088	7,001	1,39
CHNR2	60-2000Hz	29,157	3,833	23,877	8,261	1,38
CHNR3	1000-5000Hz	17,345	4,123	11,609	5,397	1,39

7.7.4. Rangkorrelationen zwischen den Irregularitätsmaßen Jitter und Shimmer

Tabelle 7.5.: Rangkorrelationen zwischen Maßen zur Bestimmung der Irregularität, pathologische Gruppe

	J3	J5	J7	J11	J15	S2	S3	S5	S7	S11	S15
J2	0,98	0,99	0,99	0,95	0,92	0,85	0,85	0,84	0,84	0,82	0,80
J3		0,98	0,96	0,89	0,84	0,82	0,84	0,81	0,80	0,76	0,74
J5			0,99	0,94	0,90	0,85	0,85	0,84	0,84	0,81	0,79
J7				0,98	0,94	0,86	0,85	0,85	0,86	0,84	0,82
J11					0,99	0,86	0,84	0,85	0,87	0,87	0,86
J15						0,86	0,82	0,84	0,86	0,87	0,87
S2							0,99	0,99	0,99	0,97	0,95
S3								0,98	0,97	0,93	0,90
S5									0,99	0,97	0,94
S7										0,99	0,97
S11											0,99

Die Rangkorrelationen in den Tabellen 7.5 und 7.6 zwischen allen Jitter- und Shimmermaßen sind signifikant (multiples Signifikanzniveau $p \leq 0,05$). Bei der pathologischen Gruppe sind die Korrelationen zwischen den Jittermaßen sehr hoch (0,84 bis 0,99). Sie werden von den Korrelationen zwischen den Shimmermaßen noch übertroffen (0,90 bis 0,99). Die Korrelationen zwischen Jitter- und Shimmermaßen sind immer noch relativ hoch, aber niedriger als die vorigen (0,74 bis 0,87). Die niedrigste Korrelation besteht zwischen J3 uns S15 (0,74).

Bei der Normalgruppe sind die Korrelationen insgesamt niedriger als bei der pathologischen Gruppe, aber auch signifikant. Sie reichen von 0,66 bis 0,98 bei Jitter, 0,73 bis 0,98 bei Shimmer und 0,46 bis 0,76 zwischen Jitter und Shimmer. Wieder ist der niedrigste Korrelationswert der zwischen J3 und S15 (0,46).

Die hohen Korrelationen innerhalb der Jittermaße und Shimmermaße zeigen, dass die genaue Wahl für die Anzahl lokal zu mittelnder Perioden (K in Gleichung 3.11) von zweitrangiger Bedeutung für die Quantifizierung von Jitter und Shimmer sind.

Zudem belegen die relativ hohen Korrelationen zwischen Jitter und Shimmer, dass beide Maße ähnlichen Stimmeigenschaften entsprechen. Wie gezeigt wird, stehen Jitter und Shimmer durch die Eigenschaften des Vokaltrakts direkt in Beziehung.

Andererseits zeigen die Korrelationen von z.B. J3 und S15 (0,46 bzw. 0,74), dass Jitter und Shimmer nicht notwendig identisch sind. Keines der beiden Maße kann gegenüber dem anderen vorgezogen werden.

Tabelle 7.6.: Rangkorrelationen zwischen Maßen zur Bestimmung der Irregularität, Normalgruppe

	J3	J5	J7	J11	J15	S2	S3	S5	S7	S11	S15
J2	0,96	0,97	0,95	0,89	0,82	0,64	0,62	0,60	0,61	0,59	0,60
J3		0,93	0,87	0,75	0,66	0,56	0,59	0,53	0,51	0,46	0,46
J5			0,98	0,90	0,82	0,65	0,63	0,61	0,63	0,62	0,62
J7				0,96	0,90	0,69	0,64	0,65	0,68	0,69	0,69
J11					0,98	0,68	0,59	0,63	0,69	0,73	0,75
J15						0,65	0,54	0,59	0,65	0,72	0,76
S2							0,95	0,98	0,97	0,92	0,88
S3								0,94	0,90	0,79	0,73
S5									0,98	0,90	0,85
S7										0,96	0,92
S11											0,98

7.7.5. Rangkorrelationen zwischen Maßen zur Bestimmung des Rauschanteils

Tabelle 7.7.: Rangkorrelationen zwischen Maßen zur Bestimmung des Rauschanteils, pathologische Gruppe

	GNE2	GNE3	NNE1	NNE2	NNE3	CHNR1	CHNR2	CHNR3
GNE1	0,95	0,89	0,53	0,53	0,79	−0,68	−0,71	−0,86
GNE2		0,96	0,49	0,50	0,78	−0,64	−0,67	−0,85
GNE3			0,45	0,45	0,75	−0,60	−0,62	−0,82
NNE1				0,93	0,76	−0,91	−0,86	−0,69
NNE2					0,72	−0,83	−0,88	−0,64
NNE3						−0,83	−0,83	−0,96
CHNR1							0,94	0,83
CHNR2								0,83

Für die pathologische Gruppe sind alle Korrelationen zwischen Maßen zur Bestimmung des Rauschanteils in Tabelle 7.7 signifikant. Die Korrelationswerte zwischen den GNEs mit verschiedenen Mittenfrequenzen sind bei der pathologischen Gruppe sehr hoch (0,89 bis 0,96). Der kleinste Korrelationswert entspricht dem größten Unterschied der Bandbreiten. Die NNE- und CHNR Maße zeigen ebenfalls sehr hohe Korrelationen untereinander. Dabei fällt auf, dass NNE und CHNR bei gleichen Frequenzbereichen höher miteinander korreliert sind als z.B. NNE1 und NNE3. Das bedeutet, dass sich NNE und CHNR sehr ähnlich verhalten, wenn sie im gleichen Frequenzbereich gemessen werden. Die Korrelationen der GNEs sind jeweils am höchsten zu NNE3 und CHNR3. Da NNE3 und CHNR3 den höherfrequenten Rauschanteil messen, muss auch der GNE sensitiv für diese Frequenzregion sein. Der Grund dafür ist, dass der GNE auch dann einen niedrigen Wert hat, wenn die Einhüllende im tieffrequenten Bereich pulsartig, in höheren Frequenzbereichen jedoch rauschartig ist.

Die Rangfolge der Korrelationswerte für die Normalgruppe entspricht insgesamt der pathologischen Gruppe. Auffällig ist jedoch, dass nur GNE1 und GNE2 mit CHNR3 signifikant korreliert sind. Die anderen Korrelationen zwischen den GNEs und NNE bzw. CHNR sind nicht signifikant. GNE misst also etwas anderes als NNE und CHNR.

Tabelle 7.8.: Rangkorrelationen zwischen Maßen zur Bestimmung des Rauschanteils, Normalgruppe. *:Korrelation nicht signifikant.

	GNE2	GNE3	NNE1	NNE2	NNE3	CHNR1	CHNR2	CHNR3
GNE1	0,81	0,67	−0,09*	−0,12*	0,33*	−0,09*	−0,08*	−0,48
GNE2		0,89	−0,24*	−0,21*	0,27*	0,06*	0,02*	−0,44
GNE3			−0,27*	−0,22*	0,14*	0,14*	0,10*	−0,31*
NNE1				0,86	0,60	−0,91	−0,79	−0,44
NNE2					0,46	−0,73	−0,86	−0,28*
NNE3						−0,71	−0,65	−0,94
CHNR1							0,85	0,66
CHNR2								0,58

7.7.6. Rangkorrelationen zwischen Maßen zur Bestimmung des Rauschanteils und den Irregularitätsmaßen Periodenkorrelation, Jitter und Shimmer

Tabelle 7.9.: Rangkorrelationen zwischen akustischen Maßen für den Rauschanteil und Jittermaßen, pathologische Gruppe

	J2	J3	J5	J7	J11	J15
GNE1	0,66	0,65	0,66	0,66	0,65	0,64
GNE2	0,62	0,62	0,63	0,63	0,61	0,59
GNE3	0,58	0,58	0,59	0,59	0,57	0,55
NNE1	0,82	0,77	0,81	0,83	0,83	0,82
NNE2	0,79	0,74	0,78	0,81	0,83	0,83
NNE3	0,81	0,78	0,81	0,82	0,80	0,77
CHNR1	−0,84	−0,80	−0,84	−0,84	−0,84	−0,82
CHNR2	−0,81	−0,78	−0,82	−0,83	−0,84	−0,83
CHNR3	−0,79	−0,77	−0,79	−0,79	−0,77	−0,75

Tabelle 7.10.: Rangkorrelationen zwischen akustischen Maßen für den Rauschanteil und Shimmermaßen, pathologische Gruppe.

	S2	S3	S5	S7	S11	S15
GNE1	0,60	0,60	0,60	0,60	0,58	0,57
GNE2	0,55	0,56	0,56	0,56	0,54	0,53
GNE3	0,52	0,53	0,53	0,52	0,50	0,49
NNE1	0,85	0,84	0,85	0,86	0,85	0,83
NNE2	0,84	0,81	0,83	0,84	0,85	0,84
NNE3	0,72	0,71	0,72	0,72	0,71	0,70
CHNR1	−0,87	−0,86	−0,88	−0,87	−0,85	−0,83
CHNR2	−0,86	−0,85	−0,86	−0,86	−0,85	−0,83
CHNR3	−0,70	−0,70	−0,71	−0,70	−0,69	−0,67

Die Korrelationen der GNEs mit den Jittermaßen in Tabelle 7.9 sind alle niedriger (0,55 bis 0,66) als von den NNEs und CHNRs (0,74 bis 0,84). Alle Korrelationen sind signifikant.

Das gleiche gilt für die Korrelationen zu den Shimmermaßen in Tabelle 7.10: Hier liegen die Korrelationen zu den GNEs im Bereich von 0,49 bis 0,60. NNE und CHNR zeigen jedoch Korrelationen im Bereich von 0,70 bis 0,86.

Tabelle 7.11.: Rangkorrelationen zwischen akustischen Maßen für den Rauschanteil und Jittermaßen, Normalgruppe. *:Korrelation nicht signifikant.

	J2	J3	J5	J7	J11	J15
GNE1	0,11*	0,14*	0,10*	0,07*	0,00*	−0,02*
GNE2	0,00*	0,05*	0,02*	−0,03*	−0,11*	−0,14*
GNE3	−0,09*	−0,05*	−0,08*	−0,13*	−0,20*	−0,22*
NNE1	0,59	0,48	0,60	0,66	0,68	0,64
NNE2	0,57	0,44	0,58	0,64	0,68	0,66
NNE3	0,70	0,65	0,72	0,72	0,67	0,61
CHNR1	−0,62	−0,56	−0,64	−0,68	−0,67	−0,62
CHNR2	−0,62	−0,53	−0,66	−0,70	−0,69	−0,65
CHNR3	−0,60	−0,58	−0,62	−0,61	−0,53	−0,47

Noch deutlicher wird der Unterschied von GNE gegenüber NNE und CHNR an den Korrelationen zu Jitter und Shimmer der Normalstimmen in Tabelle 7.11 und 7.12: Alle Korrelationen zu den GNEs sind nicht signifikant, alle Korrelationen zu NNE und CHNR sind signifikant.

Die berechneten Korrelationen entsprechen den Ergebnissen der Messungen mit synthetischen Signalen: GNE ist weniger (bzw. nicht signifikant) mit Jitter und Shimmer korreliert als NNE und CHNR. Im Gegensatz zu den Messungen an synthetischen Signalen sind die GNEs jedoch bei pathologischen Stimmen mit Jitter und Shimmer korreliert. Der Grund dafür ist, dass offensichtlich Jitter und Shimmer bei pathologischen Stimmen nicht getrennt von einem erhöhten Rauschanteil auftreten. Dies ist jedoch bei den normalen Stimmen nicht der Fall.

Neben der Periodenlänge und der Amplitude (bzw. Energie) kann auch die Wellenform einer Signalschwingung von Periode zu Periode schwanken. Diese dritte Art der Abweichung von der strengen Periodizität wird von dem Korrelationswert aufeinanderfolgender Perioden quantifiziert. Sind die Perioden identisch, ist dieser Wert 1, sind sie spiegelbildlich, ist der Wert -1 (negative Werte treten praktisch nie auf, da ja das Maximum dieses Wertes gerade zur Definition der Periodenlänge dient). Sind sie sehr unähnlich (z.B. im Falle von Rauschen), so ist der Wert nahe bei null. Werden diese Korrelationen über alle Perioden gemittelt, so erhält man den Mean Waveform Matching Coefficient (MWC). Die Korrelationen von Jitter- und Shimmermaßen, GNE, NNE und CHNR mit dem MWC sind in Tabelle 7.13 zusammengefasst. Auch hier sind die Korrelationen des GNE bei der Normalgruppe im Gegensatz zu NNE und CHNR nicht signifikant. Es sei darauf hingewiesen, dass der kleinste Korrelationswert zwischen MWC und Jitter bei J3 und zwischen MWC und Shimmer bei S3 und S15 auftritt.

Tabelle 7.12.: Rangkorrelationen zwischen akustischen Maßen für den Rauschanteil und Shimmermaßen, Normalgruppe. *:Korrelation nicht signifikant.

	S2	S3	S5	S7	S11	S15
GNE1	0,15*	0,21*	0,17*	0,11*	0,05*	0,03*
GNE2	0,00*	0,07*	0,02*	−0,03*	−0,09*	−0,12*
GNE3	−0,05*	−0,00*	−0,04*	−0,08*	−0,13*	−0,17*
NNE1	0,71	0,63	0,69	0,73	0,73	0,71
NNE2	0,72	0,62	0,70	0,74	0,76	0,74
NNE3	0,45	0,43	0,43	0,44	0,45	0,46
CHNR1	−0,73	−0,70	−0,73	−0,74	−0,71	−0,69
CHNR2	−0,74	−0,70	−0,74	−0,76	−0,74	−0,72
CHNR3	−0,39	−0,41	−0,39	−0,38	−0,35	−0,36

Tabelle 7.13.: Rangkorrelationen zwischen mittlerem Periodenkorrelationswert (MWC) und den übrigen Maßen. Nicht signifikante Werte sind mit *gekennzeichnet.

	MWC			MWC	
	pathol.	normal		pathol.	normal
J2	0,81	0,41	S2	0,88	0,65
J3	0,78	0,34*	S3	0,87	0,59
J5	0,81	0,42	S5	0,89	0,65
J7	0,82	0,46	S7	0,88	0,67
J11	0,82	0,46	S11	0,86	0,63
J15	0,81	0,43	S15	0,84	0,59
GNE1	0,65	0,06*	CHNR1	−0,95	−0,80
GNE2	0,61	−0,07*	CHNR2	−0,87	−0,66
GNE3	0,58	−0,12*	CHNR3	−0,77	−0,43
NNE1	0,91	0,78			
NNE2	0,82	0,61			
NNE3	0,77	0,48			

7.8. Optimale Kombination von Stimmgütemessgrößen mit einem informationstheoretischen Optimierungskriterium

Aus den vorherigen Untersuchungen bietet sich zur weiteren Betrachtung ein vierdimensionaler Stimmgüteraum an, da auch die Abweichung von der Periodizität bei gehaltenen Vokalen im Wesentlichen auf vier Arten erfolgen kann. Diesen vier Arten der Abweichung entsprechen jeweils bestimmte akustische Maße.

Tabelle 7.14.: Arten der Periodizitätsabweichung und die entsprechenden akustischen Maße

Periodizitätsabweichung	Akustischen Maße
Periodenlänge	Jitter (J2,J3, ..., RAP, Jitt, Jitta, PPQ)
Amplitude / Energie	Shimmer (S2, S3, ..., ShdB, Shim, APQ)
Wellenform	Periodenkorrelation (MWC)
additives Rauschen	GNE, NNE, CHNR

Mit Korrelationen lassen sich nur paarweise Abhängigkeiten bestimmen. Mit der in 7.8 eingeführten normalisierten zusätzlichen Information ΔI_N wird untersucht, welche Kombination der Maße Jitter, Shimmer, MWC die günstigste ist und welcher der Rauschmaße GNE, NNE oder CHNR zusätzlich die meiste Information liefert.

7.8.1. Die beste Kombination von {Periodenkorrelation, Jitter, Shimmer}

Von den Gruppen 4, 9 und 10 (Tab.4.1)wurde für die dreidimensionalen Kombinationen Periodenkorrelation (MWC), Jitter, Shimmer die auf den Bereich 0 bis 1 normierte dreidimensionale Entropie berechnet. Dabei bedeutet ein kleiner Wert wenig Unordnung, d.h. die Maße haben eine wechselseitige Ordnung und sind deshalb nicht sehr informationsreich. Ein hoher Wert der normierten Entropie hingegen zeigt „unordentliche“ Maße mit wenig Redundanz.

Die drei Gruppen wurden ausgewählt, da sie jeweils unterschiedliche Vor- und Nachteile zeigen. Gruppe 4 ist von der Zusammensetzung die „sauberste“ Gruppe. Von jedem Patienten wurde genau ein mittleres Segment analysiert. Jeder Patient kommt nur einmal in der Gruppe vor, aber mit 447 Patienten, ist die Gruppe relativ klein (im Sinne der Informationstheorie, nicht im Vergleich zu anderen Studien). Gruppe 9 ist eine ausgeweitete Version von Gruppe 4 (siehe Tabelle 4.1, die aufgrund ihrer höheren Anzahl von Datenpunkten besser für diese Analyse geeignet ist. Gruppe 10 eignet sich schließlich am besten, da sie sehr groß ist. Sie enthält die akustischen Maße verschiedener Vokale, wohingegen die anderen beiden Gruppen nur den Vokal [ɛː] enthalten. Die Definition des

Heiserkeits-Diagramms wird später nur auf der Grundlage der Vokale [ɛ:] erfolgen. Hier zeigt sich, ob die Wahl der Maße auch für alle Vokale gerechtfertigt ist.

Die Entropiewerte sind in Tabelle 7.15 für jede Gruppe nach absteigenden Entropiewerten sortiert, so dass die informationsreichste Kombination oben steht. Je größer die Gruppe ist, desto größer kann auch die Anzahl der Intervalle gewählt werden, so dass immer noch genügend Datenpunkte pro dreidimensionalem Intervall zur Verfügung stehen.

Die Unterschiede der Entropiewerte von der informationsreichsten bis zur informationsärmsten Kombination sind bei allen Gruppen relativ klein. Die genaue Kombination ist also nicht kritisch. Wenn man jedoch die Wahl hat, so kann man sich für die beste Kombination, die bei allen Gruppen {MWC, J3, S15} ist, entscheiden. Insbesondere ergibt sich die gleiche beste Kombination für die Vokale [ɛ:] und für die Gruppe mit allen Vokalen.

Tabelle 7.15.: Normierte Entropie bei allen Kombinationen aus 1. MWC, 2. J2-J15, 3. S2-S15

	n=447 4 Intervalle		n=13414 8 Intervalle		n=293817 50 Intervalle	
Rang	j, s	Normierte Entropie	j, s	Normierte Entropie	j, s	Normierte Entropie
1	J3 S15	0.771	J3 S15	0.815	J3 S15	0.869
2	J5 S15	0.767	J2 S15	0.809	J2 S15	0.864
3	J2 S15	0.762	J5 S15	0.806	J15 S15	0.864
4	J7 S15	0.761	J3 S11	0.805	J15 S3	0.864
5	J3 S11	0.757	J15 S15	0.802	J3 S11	0.864
6	J5 S11	0.752	J7 S15	0.802	J5 S15	0.863
7	J2 S11	0.749	J11 S15	0.800	J11 S15	0.862
8	J7 S11	0.746	J15 S3	0.799	J7 S15	0.862
9	J3 S7	0.744	J2 S11	0.798	J15 S11	0.861
10	J11 S15	0.744	J5 S11	0.796	J15 S5	0.861
11	J15 S15	0.739	J15 S11	0.795	J15 S2	0.860
12	J2 S7	0.739	J3 S7	0.793	J15 S7	0.860
13	J5 S7	0.739	J11 S3	0.793	J2 S11	0.859
14	J3 S5	0.737	J15 S2	0.792	J11 S3	0.859
15	J15 S3	0.736	J15 S5	0.792	J11 S11	0.857
16	J7 S7	0.736	J7 S11	0.792	J5 S11	0.857
17	J2 S5	0.734	J15 S7	0.791	J3 S7	0.856
18	J7 S3	0.733	J11 S11	0.791	J7 S11	0.856
19	J11 S11	0.733	J3 S5	0.789	J11 S2	0.855
20	J5 S3	0.733	J3 S2	0.789	J11 S5	0.855
21	J5 S5	0.733	J3 S3	0.789	J11 S7	0.855
22	J11 S3	0.733	J2 S7	0.788	J3 S2	0.853
23	J3 S2	0.733	J2 S3	0.787	J2 S7	0.853
24	J3 S3	0.733	J7 S3	0.786	J3 S5	0.852
25	J2 S3	0.731	J11 S5	0.786	J7 S3	0.852
26	J7 S5	0.731	J11 S2	0.786	J2 S3	0.851
27	J15 S11	0.730	J11 S7	0.786	J2 S5	0.851
28	J5 S2	0.730	J5 S7	0.785	J7 S2	0.850
29	J2 S2	0.729	J5 S3	0.785	J7 S7	0.850
30	J7 S2	0.727	J2 S5	0.785	J2 S2	0.850
31	J15 S5	0.726	J2 S2	0.784	J5 S7	0.850
32	J15 S7	0.724	J7 S7	0.783	J3 S3	0.850
33	J15 S2	0.723	J5 S2	0.782	J7 S5	0.849
34	J11 S5	0.722	J5 S5	0.781	J5 S2	0.849
35	J11 S2	0.722	J7 S2	0.781	J5 S3	0.848
36	J11 S7	0.722	J7 S5	0.781	J5 S5	0.846

7.8.2. Zusätzliche Information durch Rauschmaße

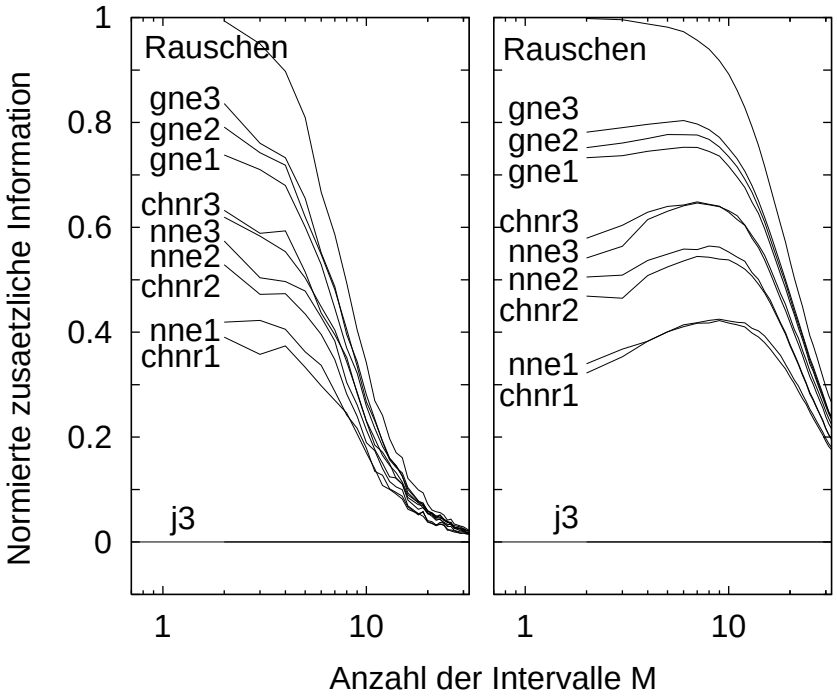


Abbildung 7.6.: Informationstheoretischer Ansatz

Die normierte zusätzliche Information ΔI_N wurde nun für die drei Gruppen jeweils für eine Kombination von {MWC, J3, S15} und eines der Maße zur Messung des Rauschanteils berechnet, um zu messen, welches der Rauschmaße der optimale Partner für die drei anderen Maße ist.

Die berechnete zusätzliche Information ΔI_N hängt stark von der durchschnittlichen Anzahl der Datenpunkte pro vierdimensionalem Intervall ab. Je größer die Anzahl der Intervalle ist, desto kleiner wird die normierte zusätzliche Information. Um diesen systematischen Einfluss abzuschätzen, wurde neben den Rauschmaßen eine Rauschfolge als zufällige Messwerte hinzugenommen. Außerdem wurde überprüft, welchen minimalen Wert die zusätzliche Information annimmt, wenn ein redundanter Messwert hinzugenommen wird, indem der Wert J3 als viertes Maß gewählt wurde. Kein Rauschmaß sollte mehr zusätzliche Information bieten, als das Rauschen, und keines weniger als J3.

Die gemessenen normierten zusätzlichen Informationen der Gruppen 4 und 10 sind in der Abbildung 7.6 und die der Gruppe 9 in Abbildung 7.7 in Abhängigkeit von der Anzahl der Intervalle pro Achse dargestellt. Wie erwartet liegen alle Werte der Rauschmaße zwischen denen für J3 und für die Rauschfolge. Die Reihenfolge der Werte für die verschiedenen Rauschmaße ist aber weder von der Gruppe noch von der Anzahl der Intervalle abhängig (mit der Ausnahme, dass bei der Gruppe 9 NNE3 und CHNR3 die Plätze gegenüber Gruppe 4 und 10 tauschen). Die Aussage ist bei allen Gruppen die gleiche: Die GNE-Maße insbesondere GNE3 bieten die höchste zusätzliche Information

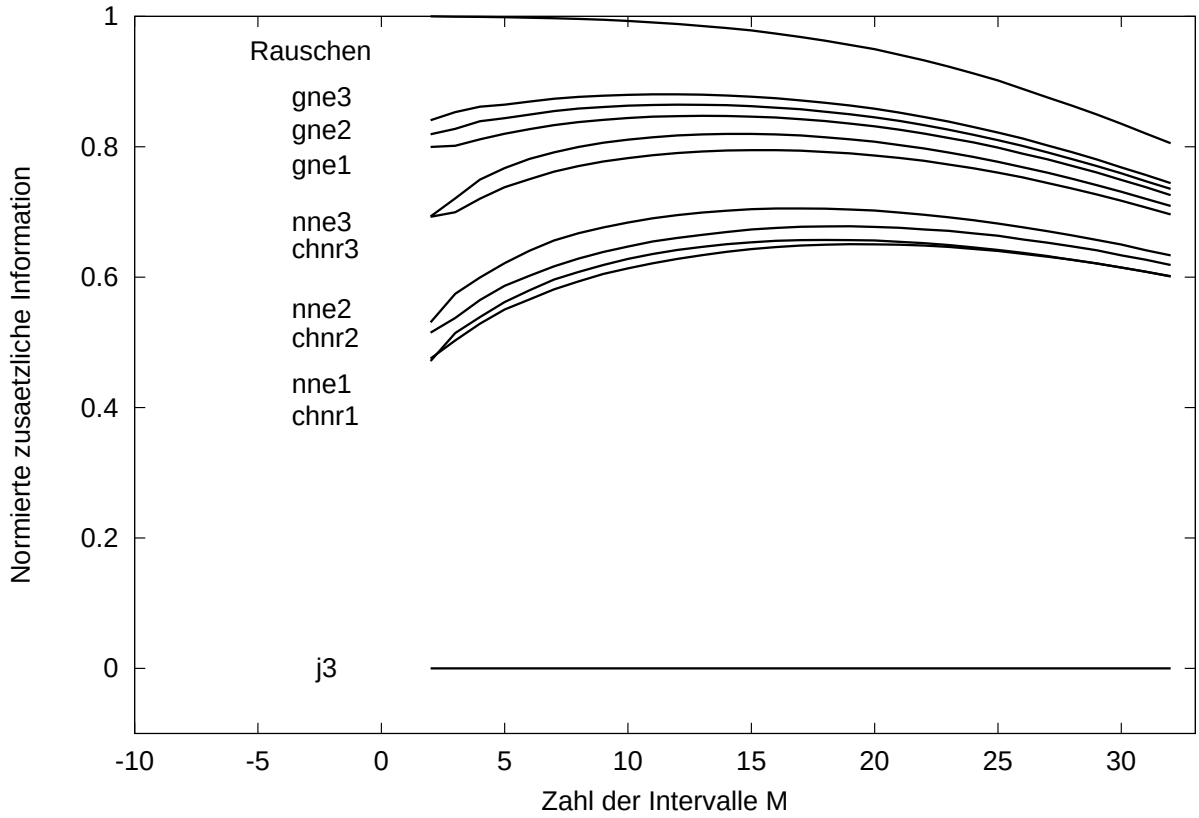


Abbildung 7.7.: Informationstheoretischer Ansatz

zu der Kombination {MWC, J3, S15}.

Bemerkenswert ist, dass die Rauschmaße, die relativ wenig Zusatzinformation bieten, in Abbildung 7.7 ein Maximum bei einer höheren Anzahl der Intervalle zeigen, als die Maße, die relativ viel Zusatzinformation bieten.

Für einen quantitativen Vergleich der normierten zusätzlichen Information wurde deshalb in Tabelle 7.16 eine Intervallanzahl von 16 gewählt. Hier hat das „Rauschen“ noch einen Wert relativ nahe bei 1 und im Zweifelsfall werden die Rauschmaße mit niedriger Zusatzinformation bevorteiligt. Hier zeigt sich, dass GNE3 gegenüber CHNR1 ca. 30% mehr Zusatzinformation liefert. Anders als bei der Kombination von Jitter und Shimmer hat die Wahl des Rauschmaßes einen großen Einfluss auf den Informationsgehalt der akustischen Analyse. Hier schneiden die GNE-Maße am besten ab, gefolgt von den NNE- und CHNR-Maßen, die den höherfrequenten Rauschanteil messen.

Tabelle 7.16.: Normierte zusätzliche Information: n=293817, 16 Intervalle pro Achse (ca. 5 Datenpunkte pro Intervall)

	ΔI_N
GNE3	0,87
GNE2	0,86
GNE1	0,84
NNE3	0,82
CHNR3	0,79
NNE2	0,71
CHNR2	0,68
NNE1	0,66
CHNR1	0,65

7.9. Zweidimensionale Projektion des Raumes der Stimmgütemessgrößen

Obwohl nun schon klar ist, dass der GNE die beste Ergänzung zu den drei Perturbationsmaßen ist, soll noch einmal die (lineare) Struktur des Datenraumes betrachtet werden. Denn zur späteren Darstellung der akustischen Stimmgüte wird ein möglichst niedrigdimensionaler Raum angestrebt, der jedoch alle wichtigen Informationen enthalten soll. Oben wurde schon gezeigt, dass der vierdimensionale Datenraum im Wesentlichen durch zwei Richtungen aufgespannt wird: eine GNE-Richtung und eine Jitter-, Shimmer- und MWC- Richtung. Hier wird nun verglichen, welchen Anteil die zweite Hauptrichtung erhält, wenn GNE durch NNE und CHNR ersetzt wird. Denn je höher die Varianz in der zweiten Hauptrichtung ist (solange die Varianz in der dritten und vierten Hauptrichtung vernachlässigbar ist), umso mehr Information steckt in den ersten beiden Dimensionen. Ist die Varianz in der zweiten Hauptrichtung relativ klein, so sind alle Maße relativ stark voneinander abhängig und könnten auf ein einziges Maß reduziert werden.

Tabelle 7.17.: Varianzen in den vier Hauptrichtungen der akustischen Datenräume, die aus den Werten {MWC, J3, S15} und einem vierten Maß zusammengesetzt sind. Pathologische Gruppe 4 ($n = 447$) und erweiterte pathologische Gruppe 10 ($n = 13414$). Varianzen $\geq 10\%$ in der zweiten Hauptrichtung sind fett gedruckt.

	($n = 447$)				($n = 13414$)			
Maß	1	2	3	4	1	2	3	4
GNE1	0,80	**0,12**	0,05	0,03	0,79	**0,12**	0,05	0,03
GNE2	0,78	**0,14**	0,05	0,03	0,77	**0,15**	0,05	0,03
GNE3	0,76	**0,16**	0,05	0,03	0,76	**0,16**	0,05	0,03
NNE1	0,88	0,06	0,03	0,02	0,89	0,06	0,04	0,01
NNE2	0,87	0,06	0,04	0,03	0,88	0,05	0,04	0,03
NNE3	0,82	**0,10**	0,05	0,03	0,82	**0,10**	0,05	0,03
CHNR1	0,89	0,06	0,04	0,01	0,89	0,06	0,04	0,01
CHNR2	0,88	0,06	0,04	0,03	0,88	0,06	0,04	0,02
CHNR3	0,83	0,09	0,05	0,03	0,83	0,09	0,05	0,03

Die Varianzen in den vier Hauptrichtungen sind für die Gruppen 4 und 10 in Tabelle 7.17 und für jeden einzelnen Vokal aus Gruppe 9 in Tabelle 7.18 gezeigt. Die Varianzen der großen Gruppe 9 (Tabelle 7.18) für die Vokale $[\varepsilon:]_1$ und $[\varepsilon:]_2$ stimmen in allen Hauptrichtungen mit einer Toleranz von einem Prozent mit denen der kleineren Gruppe 4 (n=447) überein (Ausnahmen sind lediglich GNE2 und GNE3, die sich bei den Werten der ersten Hauptrichtung um zwei Prozent unterscheiden: 0,80 statt 0,78 und 0,78 statt 0,76). Statistisch sind somit die Ergebnisse beider Gruppen nahezu identisch, obwohl die Gruppen disjunkt sind, die Aufnahmeapparaturen verschieden waren (Standmikrophon

versus Kopfmikrofon und direkte AD-Wandlung am PC versus DAT-Aufnahmen) und in der großen Gruppe auch Aufnahmen mit verschiedenen Grundfrequenzen von einem Sprecher enthalten sind. In allen Gruppen und bei allen Vokalen zeigen die Kombinationen, die eines der GNE-Maße enthalten, die höchste Varianz in der zweiten Hauptrichtung (stets in der Reihenfolge GNE3, GNE2, GNE1). In den Tabellen sind die Varianzen in der zweiten Hauptrichtung hervorgehoben, die größer oder gleich 10% sind. Die Ergebnisse der Hauptkomponentenanalyse stimmen mit den Ergebnissen der normierten zusätzlichen Information weitgehend überein.

Tabelle 7.18.: Varianzen der vier Hauptrichtungen von Gruppe 9: Für die Vokale $[\varepsilon:]_1$ (n=43337), [a:] (n=42459), [e:] (n=42621), [i:] (n=42297), [o:] (n=41019), [u:] (n=41379), $[\varepsilon:]_2$ (n=40705) und alle Vokale (28, n=293817). Varianzen $\geq$ 10% in der zweiten Hauptrichtung sind fett gedruckt.

Maß	Vokal	1	2	3	4	Vokal(e)	1	2	3	4
GNE1	$[\varepsilon:]_1$	0,81	**0,12**	0,04	0,03	[o:]	0,82	**0,12**	0,03	0,03
GNE2	$[\varepsilon:]_1$	0,80	**0,14**	0,04	0,03	[o:]	0,80	**0,15**	0,03	0,03
GNE3	$[\varepsilon:]_1$	0,78	**0,15**	0,04	0,03	[o:]	0,78	**0,16**	0,03	0,03
NNE1	$[\varepsilon:]_1$	0,88	0,05	0,04	0,02	[o:]	0,89	0,06	0,03	0,02
NNE2	$[\varepsilon:]_1$	0,88	0,05	0,05	0,03	[o:]	0,89	0,06	0,03	0,02
NNE3	$[\varepsilon:]_1$	0,83	**0,10**	0,04	0,03	[o:]	0,82	**0,12**	0,03	0,02
CHNR1	$[\varepsilon:]_1$	0,89	0,06	0,03	0,02	[o:]	0,91	0,05	0,03	0,02
CHNR2	$[\varepsilon:]_1$	0,89	0,06	0,03	0,02	[o:]	0,91	0,05	0,03	0,02
CHNR3	$[\varepsilon:]_1$	0,84	0,09	0,04	0,03	[o:]	0,84	**0,10**	0,03	0,02
GNE1	[a:]	0,82	**0,12**	0,04	0,02	[u:]	0,81	**0,14**	0,03	0,03
GNE2	[a:]	0,81	**0,13**	0,03	0,03	[u:]	0,78	**0,17**	0,03	0,03
GNE3	[a:]	0,80	**0,14**	0,04	0,03	[u:]	0,75	**0,19**	0,03	0,03
NNE1	[a:]	0,90	0,05	0,03	0,01	[u:]	0,89	0,06	0,03	0,02
NNE2	[a:]	0,90	0,05	0,03	0,02	[u:]	0,89	0,06	0,03	0,02
NNE3	[a:]	0,86	0,08	0,03	0,03	[u:]	0,80	**0,14**	0,03	0,03
CHNR1	[a:]	0,90	0,05	0,03	0,01	[u:]	0,89	0,06	0,03	0,02
CHNR2	[a:]	0,90	0,05	0,03	0,01	[u:]	0,89	0,06	0,03	0,02
CHNR3	[a:]	0,87	0,07	0,03	0,03	[u:]	0,82	**0,12**	0,03	0,03
GNE1	[e:]	0,82	**0,11**	0,04	0,03	$[\varepsilon:]_2$	0,81	**0,12**	0,04	0,03
GNE2	[e:]	0,81	**0,13**	0,04	0,03	$[\varepsilon:]_2$	0,80	**0,14**	0,04	0,03
GNE3	[e:]	0,80	**0,14**	0,04	0,03	$[\varepsilon:]_2$	0,78	**0,15**	0,04	0,03
NNE1	[e:]	0,88	0,06	0,04	0,02	$[\varepsilon:]_2$	0,89	0,06	0,04	0,02
NNE2	[e:]	0,87	0,06	0,04	0,03	$[\varepsilon:]_2$	0,88	0,05	0,04	0,03
NNE3	[e:]	0,82	**0,11**	0,04	0,03	$[\varepsilon:]_2$	0,83	**0,10**	0,04	0,03
CHNR1	[e:]	0,90	0,05	0,03	0,02	$[\varepsilon:]_2$	0,90	0,06	0,03	0,02
CHNR2	[e:]	0,89	0,05	0,03	0,03	$[\varepsilon:]_2$	0,89	0,05	0,03	0,02
CHNR3	[e:]	0,84	0,09	0,04	0,03	$[\varepsilon:]_2$	0,84	0,09	0,04	0,03
GNE1	[i:]	0,82	**0,11**	0,04	0,03	alle	0,79	**0,13**	0,05	0,03
GNE2	[i:]	0,80	**0,13**	0,04	0,03	alle	0,77	**0,15**	0,05	0,03
GNE3	[i:]	0,79	**0,14**	0,04	0,03	alle	0,75	**0,17**	0,05	0,03
NNE1	[i:]	0,87	0,07	0,04	0,03	alle	0,88	0,06	0,04	0,03
NNE2	[i:]	0,86	0,06	0,04	0,03	alle	0,87	0,06	0,05	0,03
NNE3	[i:]	0,81	**0,12**	0,04	0,03	alle	0,80	**0,13**	0,05	0,03
CHNR1	[i:]	0,89	0,05	0,03	0,03	alle	0,89	0,05	0,03	0,03
CHNR2	[i:]	0,88	0,05	0,04	0,03	alle	0,88	0,05	0,04	0,03
CHNR3	[i:]	0,83	**0,10**	0,04	0,03	alle	0,81	**0,12**	0,05	0,03

7.10. Definition des Heiserkeits-Diagramms

Die Varianzen der Gruppe 4 in der dritten und vierten Hauptrichtung sind bei allen Kombinationen kleiner oder gleich 5% und können somit ohne zu großen Informationsverlust vernachlässigt werden. Damit erhalten wir die Möglichkeit, den vierdimensionalen Raum in eine weit anwenderfreundlichere, zweidimensionale Projektion zu überführen.

Tabelle 7.19.: Komponenten der Basisvektoren 1 und 2 nach Rotation sowie die idealisierten Faktoren für das Heiserkeits-Diagramm.

Basisrichtung	MWC	J3	S15	GNE3
1	0,499	0,606	0,619	0,000
2	−0,172	0,078	0,063	0,980
Faktoren des „Heiserkeits-Diagrams“:				
1	0,577	0,577	0,577	0,000
2	0	0	0	1

Da der GNE mit 3000Hz Bandbreite sowohl die niedrigsten Korrelationen zu den Perturbationsmaßen zeigt als auch die meiste zusätzliche Information bietet und zu der höchsten Varianz in der zweiten Hauptrichtung führt, wird die Kombination der Maße Jitter (PPQ $K = 3$), Shimmer (PPQ $K = 15$), mittlere Periodenkorrelation (MWC) und GNE3 zu der Definition des Heiserkeits-Diagramms herangezogen.

In Tabelle 7.19 sind die Komponenten der ersten beiden Basisrichtungen, die durch Rotation aus den ersten beiden Hauptrichtungen hervorgehen, gezeigt. In der ersten Basisrichtung sind MWC, J3 und S15 etwa gleich stark vertreten. In der zweiten Basisrichtung dominiert GNE. Zur Definition des Heiserkeits-Diagramms werden nicht die exakten Komponenten verwendet, sondern eine idealisierte, ausbalancierte Kombination, die im unteren Teil der Tabelle zu sehen ist.

Der Nullpunkt dieses zweidimensionalen Unterraums wird so verschoben, dass in der Praxis keine negativen Werte vorkommen, da negative Stimmgütewerte zu Irritationen bei der Interpretation führen. Die x-Achse und y-Achse des Heiserkeits-Diagramms werden also wie folgt definiert:

$$\text{IK} = 5 + \frac{1}{\sqrt{3}}\left(\frac{\log J3 + 0,374}{0,645} + \frac{\log S15 - 0,757}{0,368} + \frac{\log(1 - MWC) + 1,614}{0,574}\right) \quad (7.10)$$

$$\text{RK} = 1,5 + \frac{GNE3 - 0,695}{2,42}. \quad (7.11)$$

Dabei sind: lg der Zehnerlogarithmus, J3, S15, MWC, GNE3 die untransformierten Werte, IK die Irregularitätskomponente, RK die Rauschkomponente. Die Zahlenwerte zur Mittelwertbefreiung und Normierung auf Standardabweichung 1 stammen aus Tabelle 7.4.

7.11. Datenraum bei Normalstimmen

Das Heiserkeits-Diagramm stellt die Gruppe der pathologischen Stimmen in einer zweidimensionalen Projektion von vier akustischen Maßen ohne großen Informationsverlust dar. In Tabelle 7.20 sind die Ergebnisse der Hauptkomponentenanalyse der Normalgruppe dargestellt. Im Gegensatz zu den pathologischen Gruppe ist hier die Varianz in der dritten und vierten Hauptrichtung nicht vernachlässigbar. Die Varianzen betragen selbst in der vierten Hauptrichtung noch bis zu 9%. Auch bei den Normalstimmen sind bei Kombinationen mit dem GNE stets die Varianzen der höheren Hauptachsen am größten. Das heißt, dass auch bei den Normalstimmen der GNE den Datenraum am besten ausweitet.

Tabelle 7.20.: SVD bei Normalstimmen ($n = 88$)

Maß	1	2	3	4
GNE1	0,51	0,25	0,16	0,09
GNE2	0,50	0,26	0,15	0,08
GNE3	0,51	0,25	0,16	0,08
NNE1	0,68	0,18	0,09	0,05
NNE2	0,67	0,18	0,10	0,06
NNE3	0,63	0,18	0,12	0,08
CHNR1	0,69	0,17	0,09	0,04
CHNR2	0,68	0,17	0,09	0,06
CHNR3	0,60	0,19	0,13	0,07

8. Vokaltrakteinfluss auf Jitter und Shimmer

Nach der Definition des Heiserkeits-Diagramms soll einerseits die Methode zur Periodenlängenbestimmung eingehender getestet werden und außerden der Vokaltrakteinfluss auf Jitter und Shimmer untersucht werden. Wir sind bisher davon ausgegangen, dass die im Mikrofonsignal gemessenen akustischen Maße den Werten am Ort der Stimmlippen entsprechen. Ob dies der Fall, ist soll nun untersucht werden.

In der Literatur wird häufig darauf hingewiesen, dass der Vokaltrakt einen Einfluss auf Jitter und Shimmer habe: Jitter und Shimmer an der Glottis können durch Interaktion mit dem Vokaltrakt im abgestrahlten Signal andere Werte annehmen. Bisher ist nur eine Arbeit von Kroeger [66] bekannt, in der dieses Phänomen mit Hilfe eines Sprachsynthetisators untersucht wurde.

8.1. Messungen bei realen Stimmen

8.1.1. Vorüberlegungen zur Messung

Um den Einfluss des Vokaltrakts auf Jitter und Shimmer an realen Stimmen zu messen, muss man sich zunächst einmal überlegen, wie und aus welchen Signalen man Jitter und Shimmer jeweils vor und hinter dem Vokaltrakt messen kann. Dabei ist man durch die Wahl praktikabler Lösungen sehr beschränkt. Ideal wäre es, das akustische Signal am Ort der Glottis zu messen. Diese Art der Messung ist jedoch nicht zumutbar und mit vertretbarem Aufwand zu leisten. Jitter und Shimmer an der Glottis könnten auch mit Hochgeschwindigkeitskameras gemessen werden. Diese Methode wird gegenwärtig jedoch relativ selten angewandt, da Hochgeschwindigkeitskameras noch nicht als Standardausrüstung vorhanden sind. Außerdem liefern die derzeitigen Systeme zeitlich und räumlich zu geringe Auflösungen, um Jitter und Shimmer zu messen. Ein weiterer Nachteil der Hochgeschwindigkeitskameras ist, dass sie die Phonation behindern und selbst relativ laut sind, also Störgeräusche verursachen. Prinzipiell wären Jitter und Shimmer auch mit der Photoglottographie (PGG) messbar. Hier ist zumindest die zeitliche Auflösung nicht begrenzt und die räumliche Auflösung relativ hoch. Leider stand für diese Arbeit kein PGG zur Verfügung. Deshalb wird hier auf die vierte Möglichkeit ausgewichen: Messung von Jitter und Shimmer mit dem Elektroglottogramm (EGG). Das EGG muss wegen des relativ geringen Signal-Rauschabstandes tiefpassgefiltert werden.

Die Ableitung des EGG (DEGG) gibt Auskunft über die Geschwindigkeit der Änderung der Stimmlippenkontaktfläche. Diese Änderung ist im Moment des Glottisschlusses am größten, zu dem Moment also, an dem die höheren Frequenzen angeregt werden. Es ist deshalb sinnvoll, eine Periodenlänge von einem Minimum des DEGG bis zum nächsten zu messen. Schoentgen rät bei der Differenzierung und Tiefpassfilterung des EGG's ein Filter mit linearer Phase (also keine relativen Phasenverschiebungen bei verschiedenen Frequenzen) zu verwenden. Deshalb wurde hier ein Filter mit reeller Übertragungsfunktion, einer Gaußfunktion, im Spektralbereich benutzt. Diese wurde differenziert, so dass die Übertragungsfunktion lautete:

$$\begin{aligned} Re(H(f)) &= 0 \\ Im(H(f)) &= 2\pi f e^{\ln(2)\frac{f^2}{f_G^2}} \end{aligned} \tag{8.1}$$

mit der im Folgenden als Grenzfrequenz bezeichneten Frequenz f_G.

Das Ergebnis von Differenzierung und Tiefpassfilterung eines EGG ist beispielhaft in Abbildung 8.1 gezeigt. Die differenzierten und tiefpassgefilterte EGG zeigen ein sehr deutliches Minimum. Oben im Zeitsignal ist zu erkennen, dass der Zeitpunkt der Minima mit dem Anregen der Formantschwingungen zusammenfällt. Neben den Minima in den DEGG ist auch ein schwächeres Maximum zu erkennen, dass bei der Öffnung der Stimmlippen auftritt. Aus dem Abstand dieser Extrema kann die Verschlusszeit und damit auch die Öffnungszeit gemessen werden.

Der Rauschanteil in den DEGG nimmt mit kleiner werdender Grenzfrequenz des Filters ab. Während das Signal bei 5000Hz Grenzfrequenz noch relativ stark verrauscht ist, erscheint das Signal bei 2000Hz Grenzfrequenz vergleichsweise glatt.

Die Spektren in Abbildung 8.2 zeigen, dass das EGG Störfrequenzen um 8kHz und 16kHz enthält, die bei einer Grenzfrequenz von 5000Hz noch deutlich sichtbar sind, die jedoch bei 2000Hz Grenzfrequenz weitgehend unterdrückt werden. Im folgenden wird deshalb stets mit einer Grenzfrequenz von 2000Hz gearbeitet.

Trotz Tiefpassfilterung zeigen die Minima des DEGG zuweilen Doppelspitzen wie in Abbildung 8.3. Da diese Doppelpeaks über mehrere Perioden konsistent auftreten, kann es sich hierbei nicht um zufälliges Rauschen handeln. Bei der Frequenz, die dem reziproken Zeitabstand der Doppelpeaks entspricht (ca. 3000Hz), ist im Zeitsignal ein Formant zu erkennen. Es ist außerdem bekannt, dass bei der Glottisschwingung eine Wanderwelle auf der Stimmlippenschleimhaut von unten nach oben läuft. Durch zwei Wellenberge dieser „mucosal wave“ könnte es beim Verschluss zu dem beobachteten Doppelpeak kommen. In der Literatur konnten bisher keine Angaben zu einem solchen Doppelpeak gefunden werden.

Die Doppelpeaks sind nicht nur theoretisch interessant, sondern sie erschweren auch die Bestimmung der Periodenlängen. Ohne solche Doppelpeaks wäre Peakpicking ein gutes Verfahren zur Bestimmung der Periodenlänge, es erfasst genau den Zeitraum von einer Anregung des akustischen Signals bis zur nächsten. Die Doppelpeaks führen jedoch zu falschen Periodenlängen (Abbildung 8.4) oben. Mit einem iterativen Ansatz kann man jedoch trotzdem noch Nutzen aus den Maxima des DEGG ziehen: Zuerst

wird durch Peakpicking grob die Lage des Minimums bestimmt. Dann wird mit einem dem Waveform Matching sehr ähnlichen Verfahren die genaue Periodenlänge bestimmt: Es wird die Zeitdifferenz gesucht, bei der der Korrelationswert zwischen einem Zeitsegment mit 1ms Länge (zentriert um den aktuell gefundenen Peak) und einem zweiten 1ms langem Segment, das im Bereich der zu erwartenden Grundperiode vom aktuellen Peak entfernt liegt, maximal ist. Die Segmentlänge von 1ms hat sich als günstig herausgestellt, da sie den zentralen Teil der Minima überdeckt. Die Breite der Peaks ist auch bei stark variierender Grundperiode relativ konstant, sie entspricht in etwa der reziproken Frequenzbandbreite des Filters. Mit dieser Methode wird eine sehr gute Übereinstimmung der Periodenlänge erreicht (Abbildung 8.4 Mitte und unten).

In der Abbildung 8.5 sind Energie und Periodelängen jeweils im EGG und vom Mikrofonsignal in einem Bereich alternierender Periodenlängen eines Sprechers dargestellt. Hier sieht man, dass die schnellen Wechsel von kurzer und langer Periodenlänge sowohl im EGG als auch im Mikrofonsignal gemessen werden können. Das Gleiche gilt für die Energie. Während jedoch der zeitliche Verlauf beim EGG relativ glatt ist, zeigen sich im Mikrofonsignal deutliche Schwankungen der mittleren Werte und der Differenzen der aufeinanderfolgenden Werte.

Um den iterativen Ansatz der Periodenlängenberechnung im EGG zu testen, wurden Jitterwerte einmal aus einer Periodenlängensequenz berechnet, die durch Peakpicking bestimmt wurden, und ein zweites Mal mit der oben angegebenen Methode (ähnlich dem Waveform Matching). Im oberen Teil der Abbildung 8.6 ist der Quotient der Jitterwerte die analysierten Segmente dargestellt. Es wurden jeweils 100 Periodenlängen zur Berechnung eines Jitterwertes herangezogen. Es ist zu erkennen, dass der Quotient im Großen und Ganzen in der Nähe von eins liegt. Die Methoden liefern also vergleichbaren Jitter. Bei manchen Segmenten steigt jedoch der Quotient auf große Werte an. Das bedeutet hohe Jitterwerte bei der Peakpicking-Methode. Diese hohen Werte stammen aus der fehlerhaften Periodenlängenbestimmung mit dem Peakpicking bei Doppelpeaks.

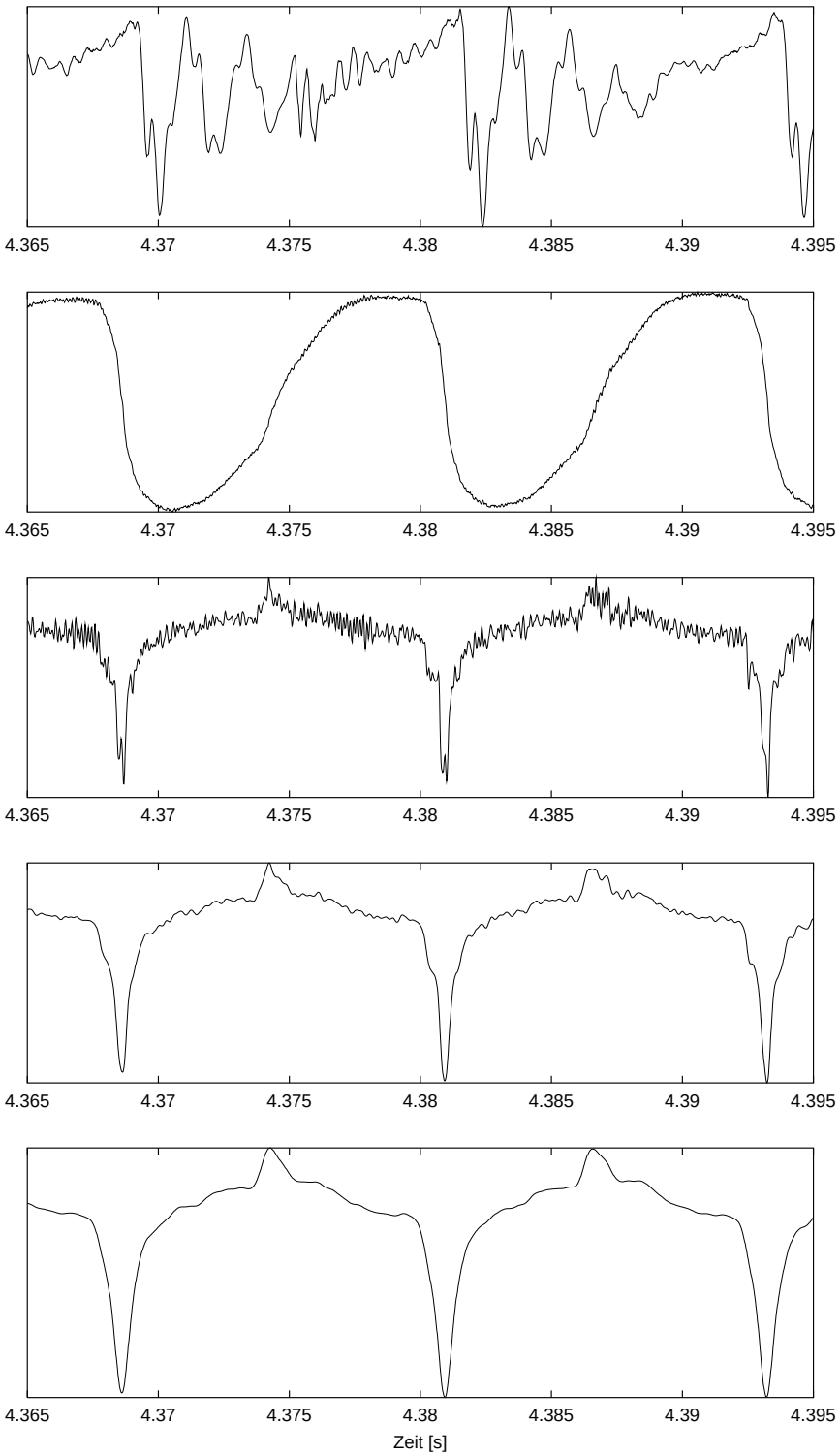


Abbildung 8.1.: Auswirkung von Differenziation und Tiefpassfilterung auf das EGG. Von oben nach unten: Zeitsignal ([ɛ:]), EGG, differenzierte und tiefpassgefilterte EGGs: 5000Hz, 2000Hz und 1000Hz Grenzfrequenz, Gaußfilter. Der Abstand zwischen den Doppelminima bei bei 5000Hz Grenzfrequenz beträgt ca. 8 Sample entsprechend 6000Hz

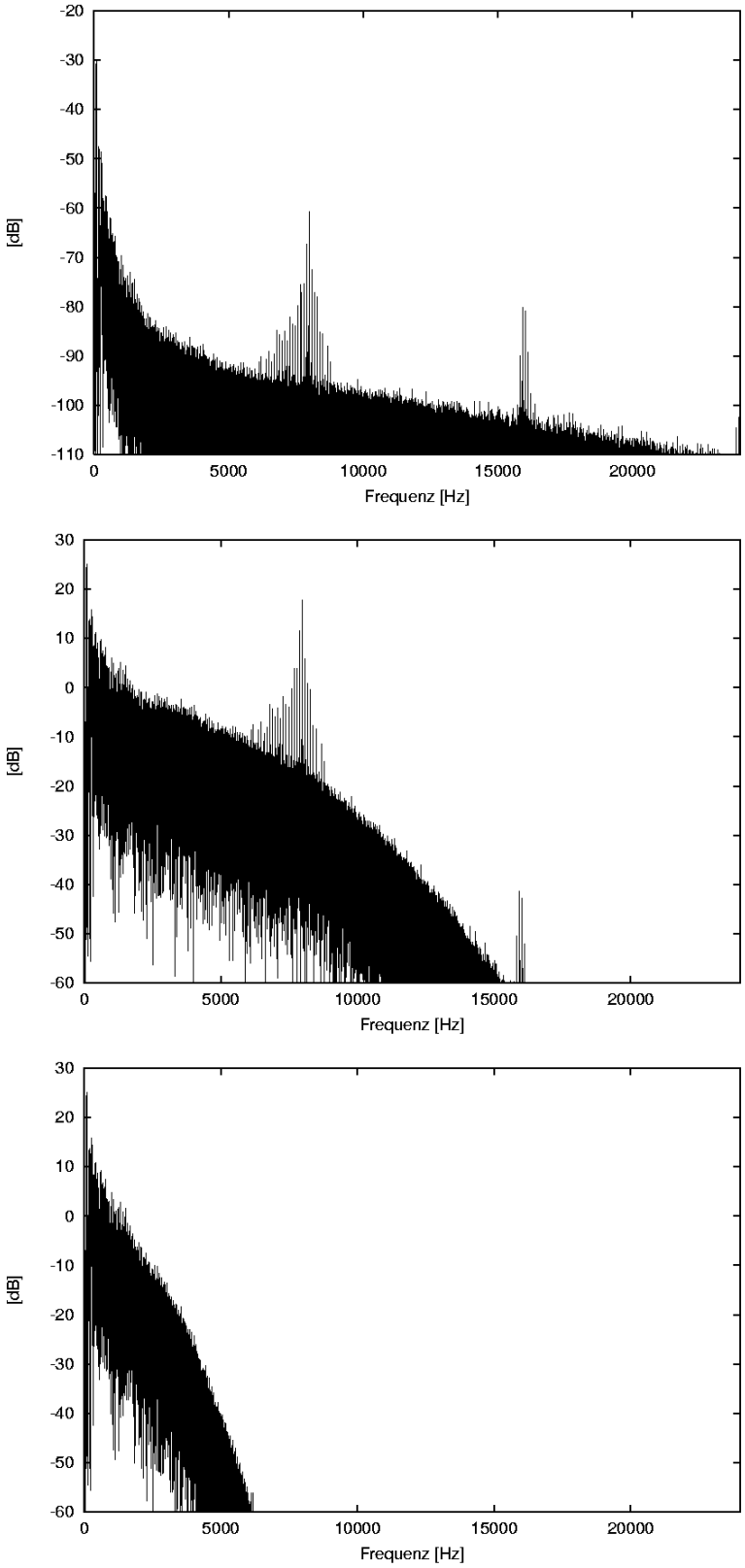


Abbildung 8.2.: Spektren des EGG. Von oben nach unten: EGG, differenzierte und tiefpassgefilterte EGG's: 5000Hz und 2000Hz Grenzfrequenz, Gaußfilter

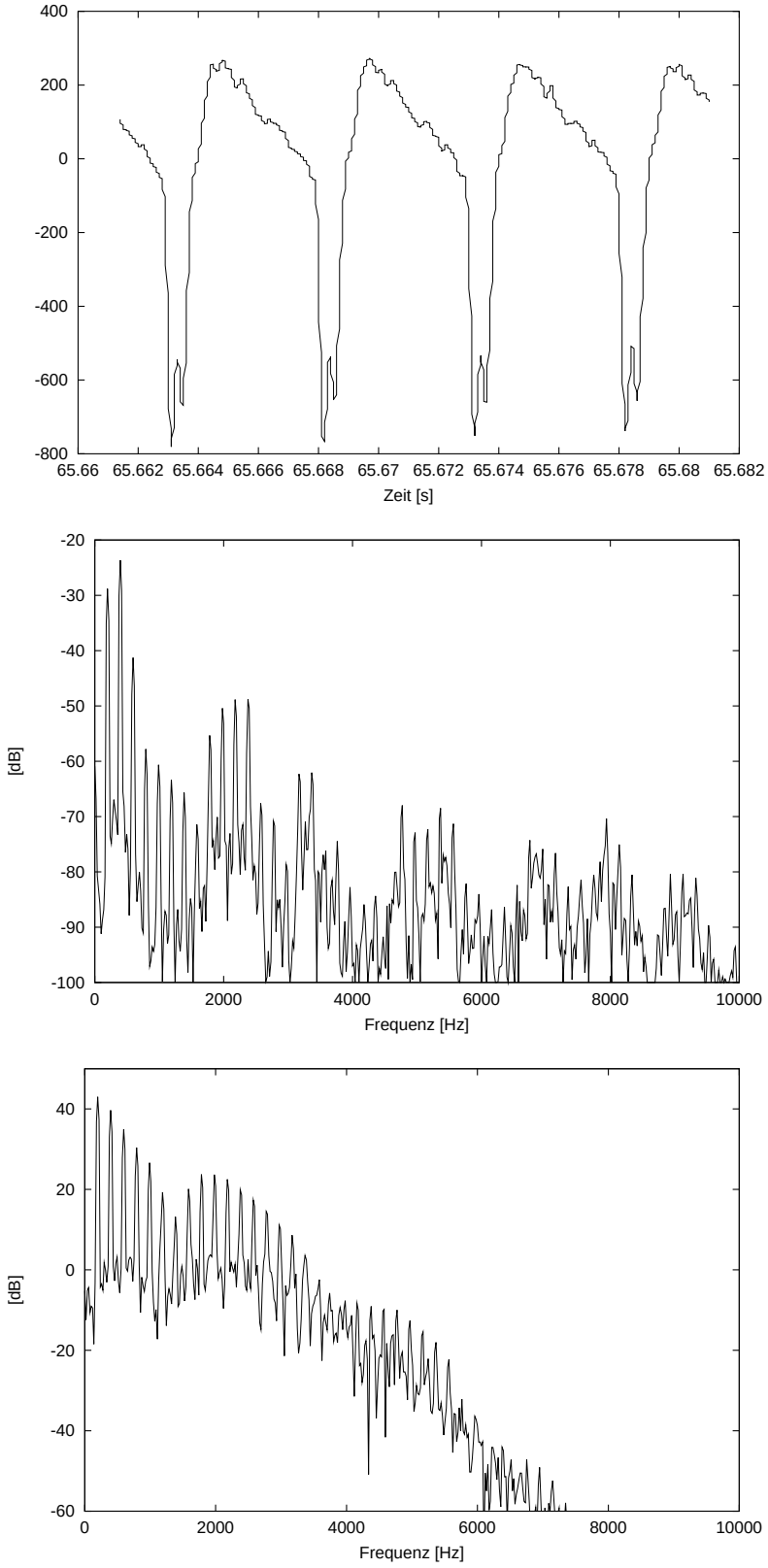


Abbildung 8.3.: Von oben nach unten: Zeitsignal, Spektrum des Zeitsignals, Spektrum des differenzierten und tiefpassgefilterten EGGs bei 2000Hz Grenzfrequenz, Gaußfilter. Der Abstand zwischen den Doppelminima in der oberen Kurve beträgt ca. 16 Sample entsprechend 3000Hz

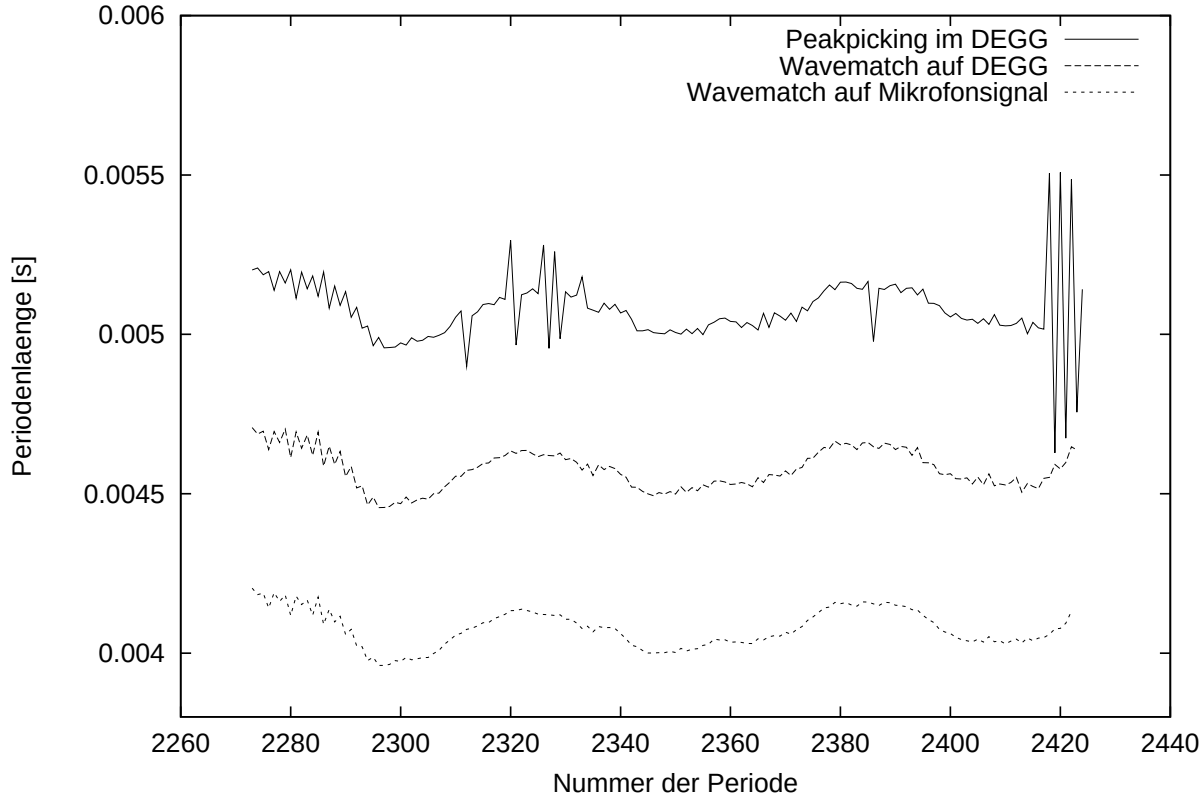


Abbildung 8.4.: Oben: Periodenlängenberechnung durch Peakpicking auf dem differenzierten EGG (bei 2000Hz Grenzfrequenz, Gaußfilter). Mitte: Wavematching auf dem vorigen Signal mit fester Korrelationsfensterlänge von 1ms, Startpunkte zur Korrelationsberechnung sind die Peaks des DEGG. Referenzfenster ist um den Peak zentriert. Unten: Wavematching auf dem Mikrofonsignal. Startpunkte zur Korrelationsberechnung sind die Peaks des DEGG. Von dem mittleren Verlauf wurden 0,5ms und von dem unteren Verlauf 1ms zur Darstellung subtrahiert.

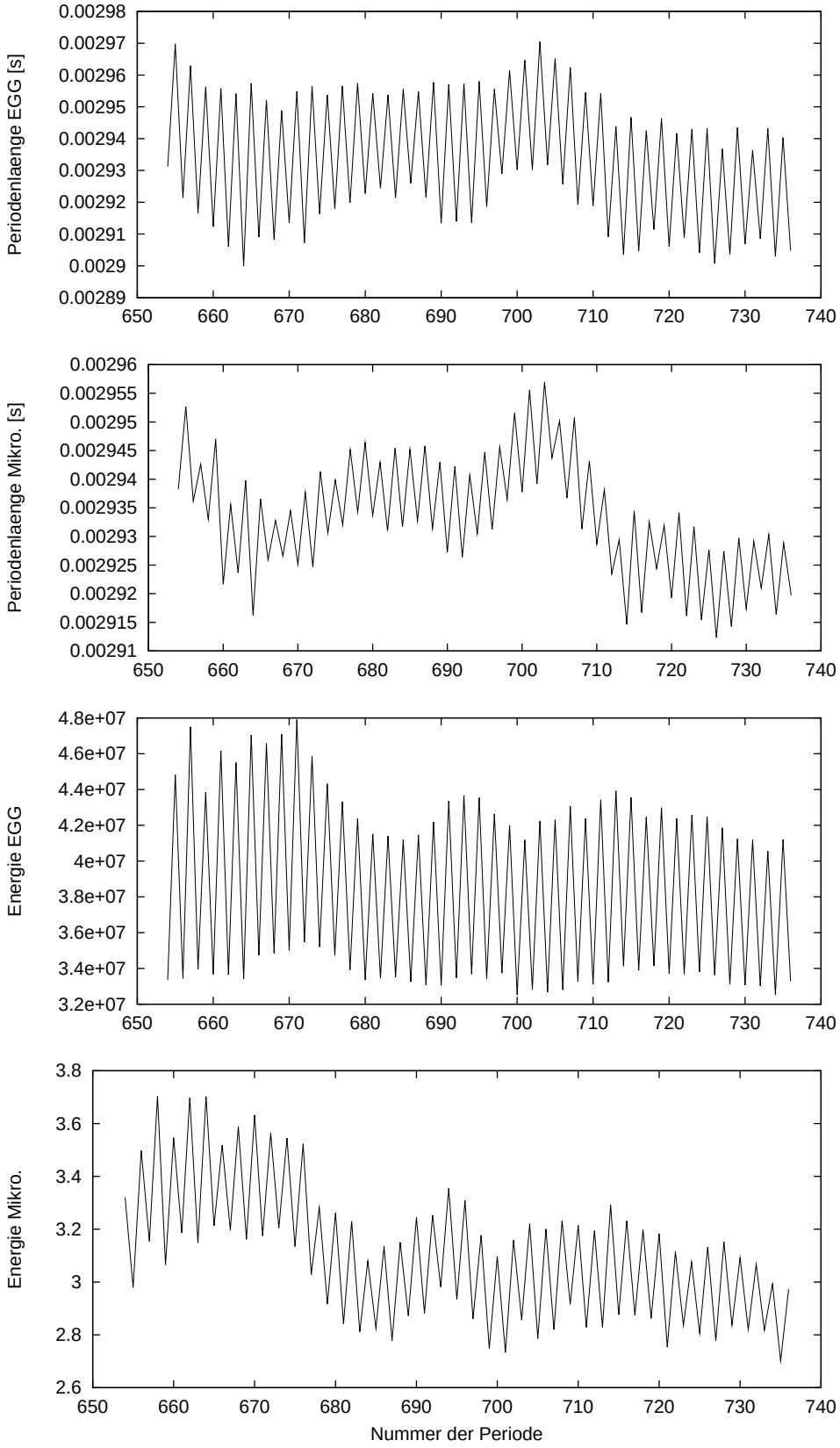


Abbildung 8.5.: Alternierende Periodenlängen und Periodenenergien treten im Signal und im EGG auf.

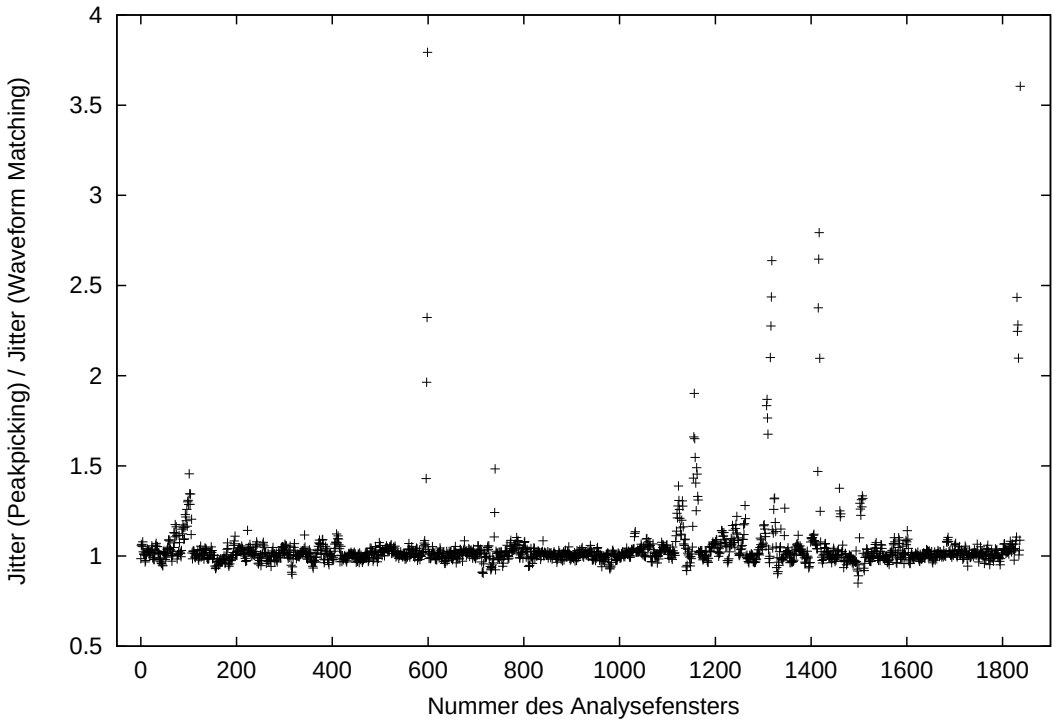


Abbildung 8.6.: Quotient der Jitterwerte mit der Peakpicking Methode und mit der Waveform-Matching-Methode (beide am DEGG).

8.1.2. Messungen

Mit der beschriebenen Methode zur Periodenlängenbestimmung im EGG wurden die Jitter- und Shimmerwerte im Mikrofonsignal und im EGG von drei Sprechern berechnet. Um den Einfluss verschiedener Grundfrequenzen einschätzen zu können, wurden von den Sprechern jeweils fünf Folgen des Vokals [ɛ:] aufgenommen, bei denen die Grundfrequenz möglichst kontinuierlich von dem kleinstmöglichen Wert auf den größten Wert innerhalb des Brustregisters erhöht wurde. Dabei sollte darauf geachtet werden, die Artikulationsstellung nicht zu verändern. Während der Grundfrequenzerhöhung wurde zum Teil eine Atempause eingelegt und neu angesetzt, so dass von jedem Sprecher etwa 15 Aufnahmen mit einer mittleren Länge von je 13 Sekunden vorhanden waren. Von diesen Aufnahmen wurden Jitter und Shimmer jeweils im EGG und Mikrofonsignal berechnet. Dabei wurde der Shimmer im EGG aus der Energie im DEGG berechnet, und zwar in dem Bereich um den Peak des DEGG bis zu den umgebenden Nulldurchgängen. Jitter und Shimmer wurden jeweils für 100 aufeinander folgende Periodenlängen (PPQ, EPQ mit $K = 3$) mit 25 Perioden Fenstervorschub bestimmt.

Der gemessene Grundfrequenzverlauf der Perturbationswerte ist in den Abbildungen 8.7 bis 8.9 dargestellt. Auffällig ist bei allen Sprechern ein Abfall der Perturbationsmaße im unteren, etwa 50Hz breiten Grundfrequenzbereich. In diesem Bereich ist auch die Variation bei allen Sprechen am geringsten. Dieser Bereich liegt in der normalen Sprechtonlage und ist vielleicht deshalb am stabilsten, da sich die Stimme am häufigsten in diesem Bereich bewegt und der Sprecher deshalb viel Übung darin besitzt, die Stimmbandschwingungen in diesem Bereich ohne große Variationen zu modulieren. Im Bereich höherer Grundfrequenzen nimmt die Variabilität besonders bei den Sprechern 2 und 3 deutlich zu. Die Stimmqualität wird instabiler. Perzeptiv wurden diese Instabilitäten als „Schnarren“ wahrgenommen.

Ein besonders wichtiger Effekt ist bei Sprecher 3 bei ca. 160Hz Grundfrequenz zu erkennen: Der „Haken“, der nach oben deutlich aus den gemessenen Jitterwerten des EGG und des Mikrofons herausragt, ist im Shimmer des EGG kaum sichtbar. Im Shimmer des Mikrofonsignals ist er jedoch deutlich zu erkennen. Später wird noch klar, dass sich hier der glottale Jitter in Shimmer umgewandelt hat.

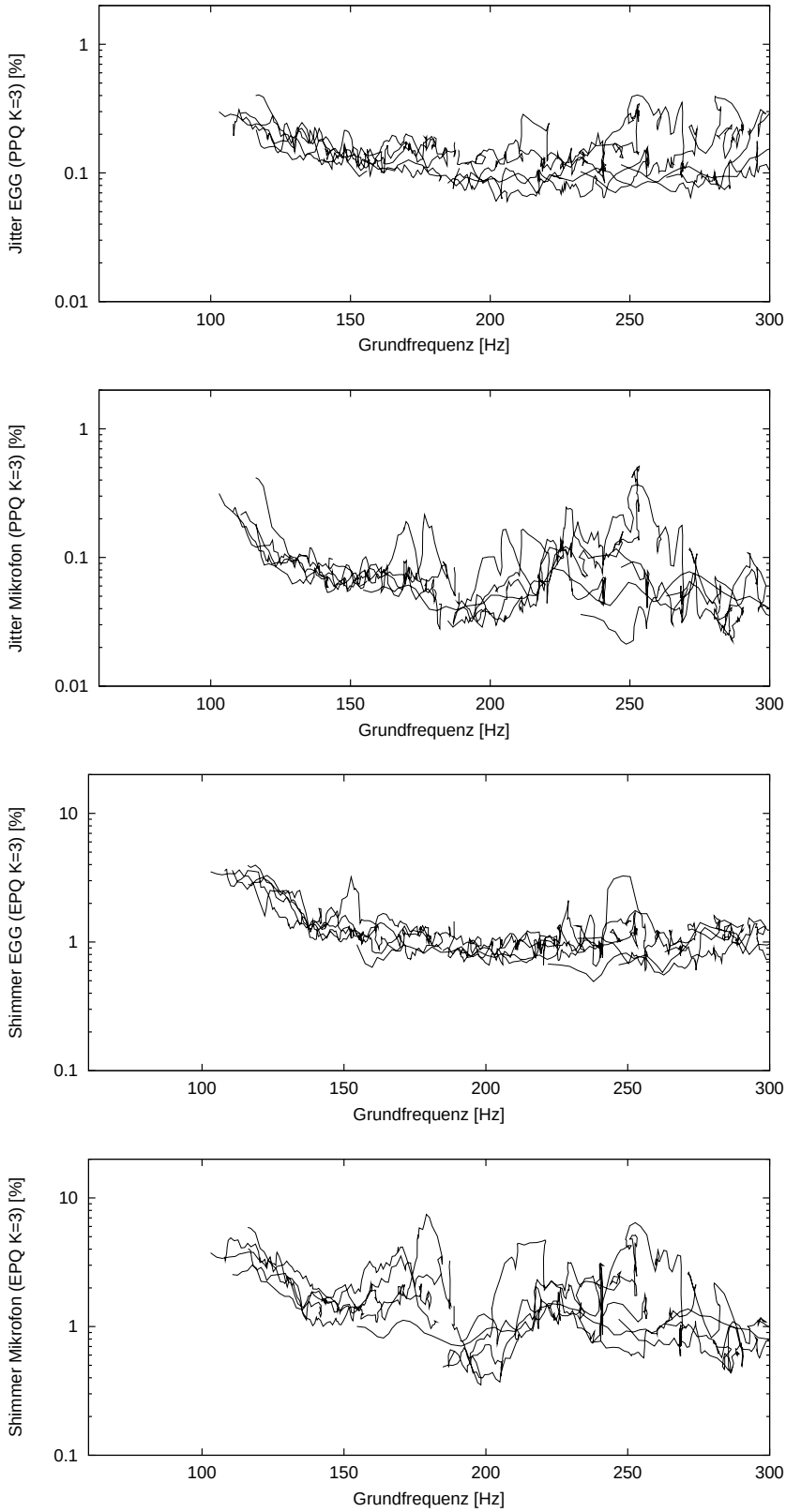


Abbildung 8.7.: Jitter und Shimmer im EGG und im Mikrofonsignal gemessen, Sprecher 1

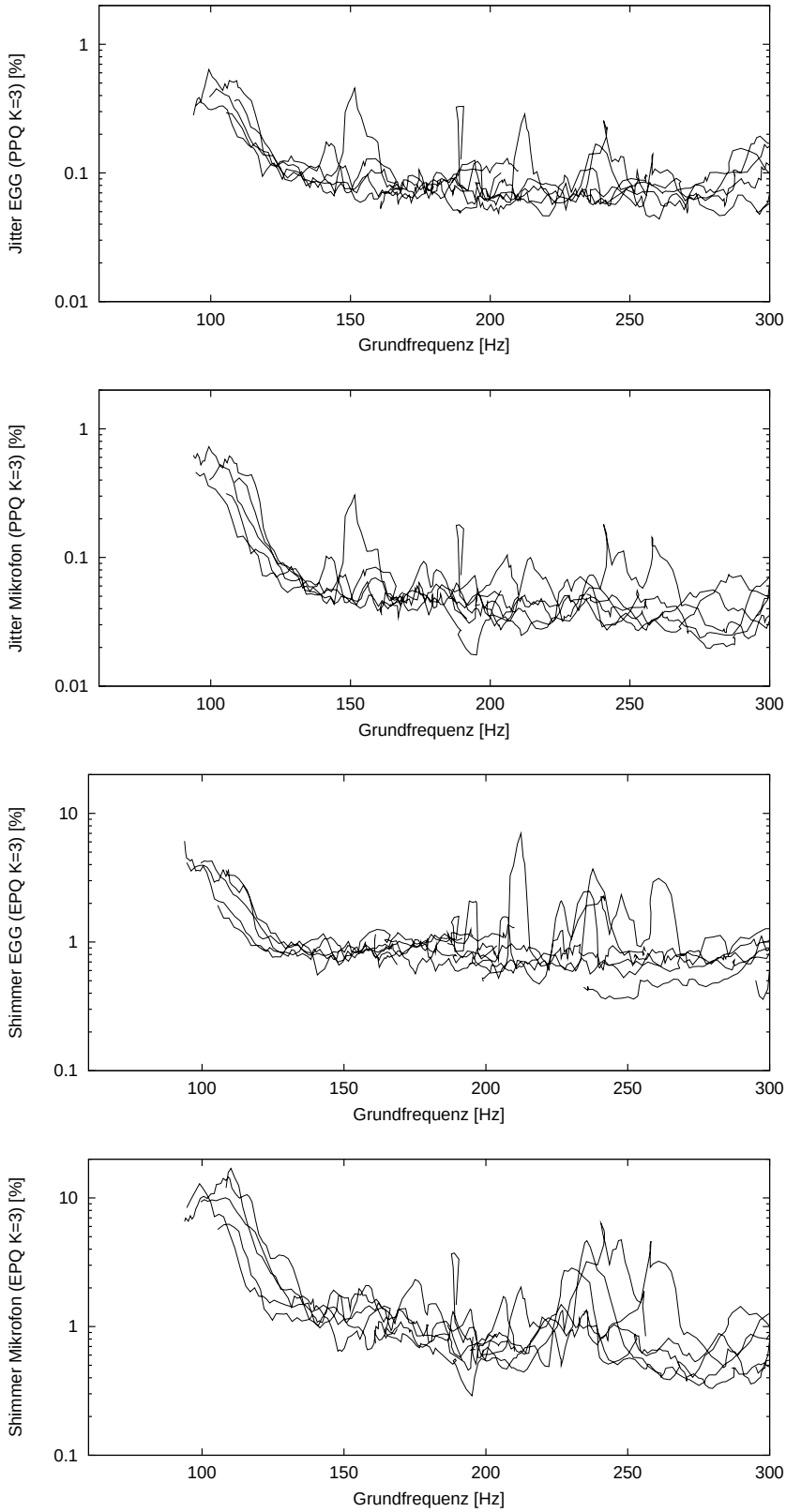


Abbildung 8.8.: Jitter und Shimmer im EGG und im Mikrofonsignal gemessen, Sprecher 2

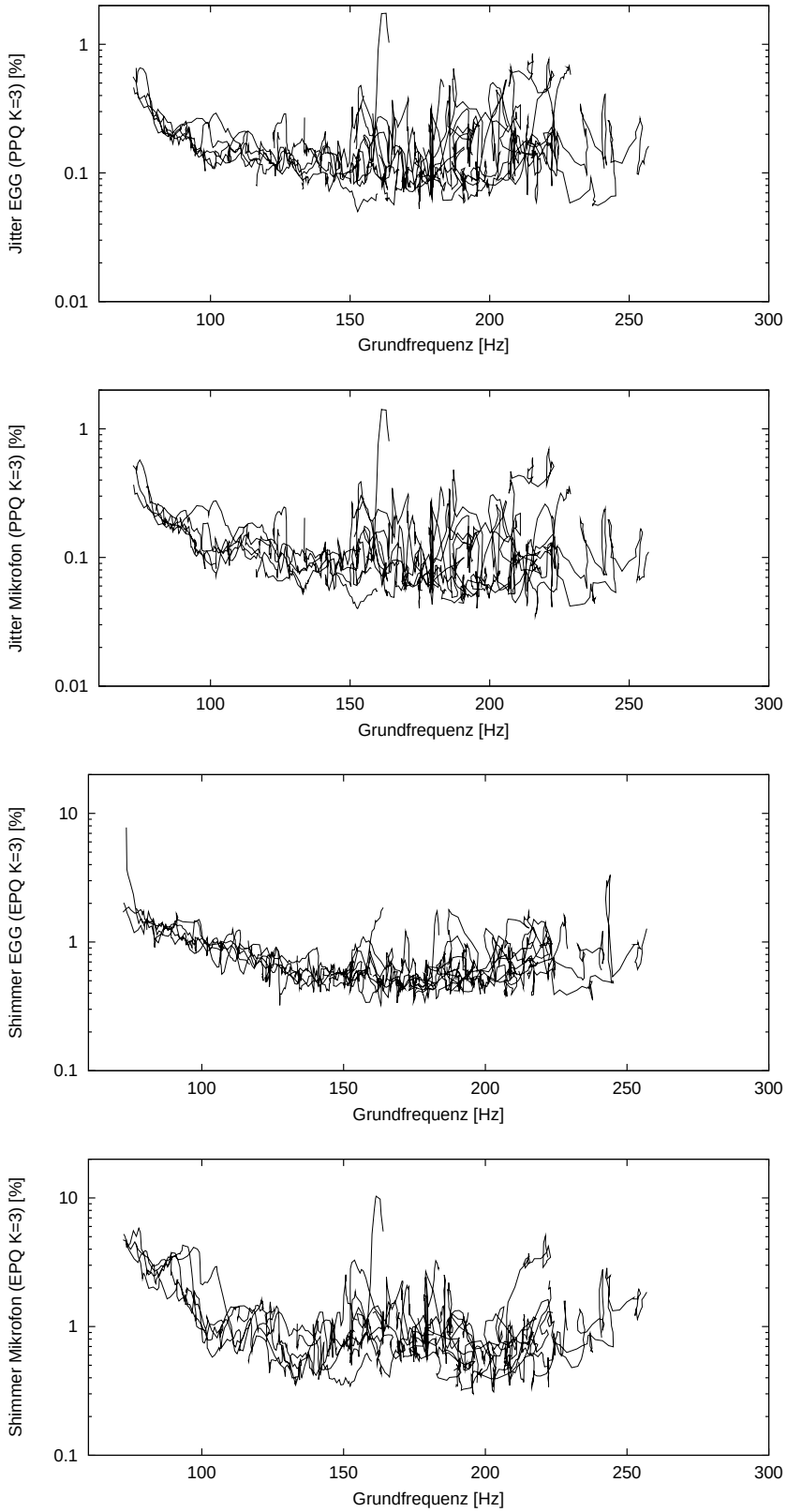


Abbildung 8.9.: Jitter und Shimmer im EGG und im Mikrofonsignal gemessen, Sprecher 3

8.1.3. Korrelationen zwischen Perturbationen im EGG und im Mikrofonsignal

Als Maß für die Ähnlichkeit der Jitter- und Shimmerwerte im EGG und Mikrofonsignal werden die Korrelationen zwischen diesen Werten für die drei Sprecher berechnet. Für die Shimmerberechnung des Mikrofonsignals wurde jeweils die Energie pro Periode berücksichtigt. Da jedoch nicht von vornherein klar ist, welches Energiemaß im EGG am ehesten der Energie pro Periode im Mikrofonsignal entspricht, wurden hier vier verschiedene Energiebereich verglichen: EGG 1: Energieshimmer: Energie von dem Nulldurchgang vor dem Peak des DEGG bis zu dem Nulldurchgang hinter dem Peak des DEGG; EGG 2: Amplitudenshimmer aus Amplitude des DEGG-Peaks; EGG 3: Energieshimmer: Energie aus den 20 Samples mit der größten Energie im Bereich des Peaks; EGG 4: Energieshimmer: Energie aus den 40 Samples mit der größten Energie im Bereich des Peaks.

Die größte Überraschung bei den Korrelationswerten in Tabelle 8.1 ist der Korrelationswert zwischen EGG- und Mikrofonjitter bei Sprecher 1: Während die anderen beiden Sprecher hier hohe Korrelationen zeigen (0,84 und 0,93) liegen die Korrelationen bei Sprecher 1 deutlich niedriger bei 0,44. Da bei den drei Sprechern der Wertebereich der gemessenen Jitterwerte vergleichbar ist, muss daraus geschlossen werden, dass Jitter im EGG und Jitter im Mikrofonsignal unter Umständen sehr unterschiedlich sein können. Weiterhin fällt auf, dass Jitter und Shimmer im Mikrofonsignal deutlich höher miteinander Korreliert sind als im EGG. Der Vokaltrakt verändert also Jitter und Shimmer derart, dass die Korrelation zwischen diesen Größen im Mikrofonsignal ansteigt. Die folgenden Experimente mit synthetischen Signalen werden zum Teil erklären, warum dies so ist.

Die Shimmermaße im EGG: EGG 1, EGG 3, EGG 4 zeigen bei allen Sprechern sehr hohe Korrelationen (über 0,94), bis auf eine Ausnahme. Die Korrelation von EGG 1 und EGG 3 bei Sprecher 1 beträgt nur 0,85. Das bedeutet, dass diese drei Shimmermaße im Wesentlichen austauschbar sind. Der genaue Bereich des DEGG zur Energieberechnung spielt also keine Rolle. Die Korrelationen des EGG Amplitudenshimmers (EGG 2) mit EGG 1 und EGG 4 sind relativ niedrig (0,76 bis 0,9), ein Anzeichen dafür, dass Amplitudenwert und Kurzzeitenergie nicht immer gleichgesetzt werden können.

Die Korrelationen zwischen Shimmer im EGG und Shimmer im Mikrofonsignal sind für EGG 4 am höchsten und für den Amplitudenshimmer EGG 2 am geringsten. Die Ursache könnte zum Teil darin liegen, dass im Mikrofonsignal auch der Energieshimmer und nicht der Amplitudenshimmer berechnet wird.

Tabelle 8.1.: Korrelationswerte zwischen Jitter und Shimmer. Im Mikrofonsignal und im EGG gemessen. Shimmer EGG 1: Energieshimmer: Energie von Nulldurchgang vor dem Peak bis Nulldurchgang hinter dem Peak des DEGG; EGG 2: Amplitudenshimmer aus Amplitude des DEGG-Peaks; EGG 3: Energieshimmer: Energie aus den 20 Samples mit der größten Energie im Bereich des Peaks; EGG 4: Energieshimmer: Energie aus den 40 Samples mit der größten Energie im Bereich des Peaks.

	Sprecher 1, 1381 Segmente					
	Jitter Mikrofon	Shimmer EGG 1	Shimmer EGG 2	Shimmer EGG 3	Shimmer EGG 4	Shimmer Mikrofon
Jitter EGG	0,44	0,68	0,69	0,73	0,69	0,55
Jitter Mikro.		0,41	0,30	0,34	0,38	0,78
Shimmer EGG 1			0,76	0,85	0,96	0,47
Shimmer EGG 2				0,95	0,86	0,42
Shimmer EGG 3					0,94	0,46
Shimmer EGG 4						0,48
	Sprecher 2, 1446 Segmente					
	Jitter Mikrofon	Shimmer EGG 1	Shimmer EGG 2	Shimmer EGG 3	Shimmer EGG 4	Shimmer Mikrofon
Jitter EGG	0,84	0,76	0,76	0,79	0,77	0,66
Jitter Mikro.		0,69	0,56	0,66	0,69	0,79
Shimmer EGG 1			0,84	0,96	1,00	0,69
Shimmer EGG 2				0,94	0,85	0,54
Shimmer EGG 3					0,97	0,64
Shimmer EGG 4						0,70
	Sprecher 3, 1838 Segmente					
	Jitter Mikrofon	Shimmer EGG 1	Shimmer EGG 2	Shimmer EGG 3	Shimmer EGG 4	Shimmer Mikrofon
Jitter EGG	0,93	0,67	0,54	0,60	0,66	0,66
Jitter Mikro.		0,66	0,58	0,62	0,67	0,75
Shimmer EGG 1			0,88	0,96	0,99	0,58
Shimmer EGG 2				0,95	0,90	0,55
Shimmer EGG 3					0,98	0,56
Shimmer EGG 4						0,59

8.1.4. Phasenabhängigkeit von Jitter- und Shimmer-Messungen

Die Waveform-Matching-Methode zur Periodenlängenbestimmung, wie sie im Heiserkeits-Diagramm implementiert ist, beginnt mit der Suche nach der Periodenlänge jeweils an einem Punkt, der gegenüber dem vorigen Startpunkt um eine Periodenlänge verschoben ist. Diese Verschiebung basiert also auf dem Verfahren selbst. Dadurch ist leider nicht garantiert, dass die Periodengrenzen eine feste Phasenbeziehung zum Signal haben. So wurde z.B. bei Signalen (aufgenommenen Vokalen) mit konstanter Grundfrequenz festgestellt, dass die Periodengrenzen von Maxima herunterwandern und dann im Bereich eines Nulldurchganges stabil bleiben. Bei den Aufnahmen der drei Sprecher mit ansteigender Grundfrequenz zeigte sich, das die Periodengrenzen beim Waveform Matching im EGG und im Mikrofonsignal nach (zeitlich) vorne wandern und dabei bei sehr langen Signale über einige Perioden hinweglaufen können.

Diese Wanderung der Periodengrenzen kann wie weiter oben beschrieben vermieden werden, wenn man die jeweiligen Startpunkte für das Waveform Matching vorgibt. Dies ist bei relativ guten Stimmen mit dem Peak des DEGG gut möglich. Leider ist das EGG bei stark behauchten oder sehr stark gestörten Stimmen so unregelmäßig, dass hier keine eindeutigen Peaks pro Periode gefunden werden können. Da die Methode der Periodenlängenbestimmung jedoch für alle Stimmen und vollautomatisch passieren soll, wurde auf die Vorgabe der Startpunkte beim Waveform Matching verzichtet und ein eventuelles Wandern der Phase in Kauf genommen.

Dennoch soll hier einmal überprüft werden, welche Auswirkung das Wandern der Phase auf die Berechnung von Jitter hat. Dazu eignen sich die Aufnahmen mit steigender Grundfrequenz besonders, sie stellen für das Waveform Matching sozusagen den „worst case“ dar, denn beim Waveform Matching wird ja vorausgesetzt, dass sich im Prinzip die aufeinander folgenden Perioden gleichen. Durch den Grundfrequenzanstieg wird jedoch diese Annahme systematisch verletzt, so dass es zu dem beschriebenen Wandern der Periodengrenzen über mehrere Perioden kommt, die bei gehaltener Tonhöhe nicht gefunden wurden. Durch diese Wanderung der Grenzen durch die Perioden ist sichergestellt, dass fast jeder Phasenwinkel zwischen Periodengrenzen und einem gegenüber der Schwingung charakteristischen Zeitpunkt (z.B. der Punkt im Mikrofonsignal, der dem Verschlusszeitpunkt der Glottis entspricht) einmal auftritt.

Jitter und Shimmer wurden im EGG und im Zeitsignal gemessen, und zwar jeweils mit fest vorgegebenen Startpunkten für das Waveform Matching (den Peaks des DEGG) und mit freien Grenzen, d.h. die Suche nach der nächsten Periodenlänge wurde immer gerade um die gefundene vorige Periodenlänge verschoben. Aus allen Periodenlängensequenzen wurde jeweils Jitter für je 100 Periodenlängen berechnet bei 25 Perioden Fenstervorschub. Da Jitter logarithmisch normalverteilt ist, wurde nicht die mittlere Differenz bestimmt, sondern der mittlere Quotient der Jitterwerte, die einmal mit freien und einmal mit vorgegebenen Startpunkten berechnet wurden. Der Quotient wurde immer so gebildet, dass der größere der beiden Messwerte im Zähler stand, sonst wäre bei der Mittelung ein Quotient nahe bei 1 herausgekommen. Die so berechneten mittleren Quotienten zeigt folgende Tabelle:

Sprecher	N	EGG	Mikrofon
1	1381	1,31	1,48
2	1446	1,38	1,27
3	1838	1,29	1,26

Der Quotient der Jittermesswerte und damit der zu erwartende maximale Fehlerfaktor liegt bei ca. 1,3. Da dies der „worst case“ ist und Jitter wie gesagt logarithmisch normalverteilt ist, d.h. es kommt in erster Linie auf die Größenordnung an, ist dieser Fehlerfaktor für die Berechnung beim Heiserkeits-Diagramm tragbar. Der Fehler bei der Berechnung des Heiserkeitsdiagramms dürfte im Allgemeinen noch geringer sein, denn dort werden ja nur gehaltene Vokale mit möglichst konstanter Grundfrequenz untersucht. Außerdem werden zur letztendlichen Charakterisierung der Stimmgüte mit dem Heiserkeits-Diagramm die Ergebnisse aller Vokale gemittelt, so dass sich etwaige Fehler dadurch nicht so schwer auswirken.

8.2. Messung der Perturbationsmaße im synthetischen Glottissignal

Der Vokaltrakteinfluss und das Waveform-Matching-Verfahren zur Bestimmung der Periodenlängen sollen hier genauer untersucht werden [82]. Zur Messung von Jitter und Shimmer werden die Periodenlängen jeder einzelnen Periode gebraucht. Die Frage ist, bis zu welchen minimalen Jitter- und Shimmerwerten und bis zu welchen maximalen Jitter- und Shimmerwerten das Waveform-Matching-Verfahren geeignet ist, um die benötigten Periodenlängen zu berechnen [83].

Titze [142] untersuchte die Genauigkeit des Waveform Matching zur hochpräzisen Berechnung von Perturbationen. Dabei wurden drei verschiedene Verfahren zur Periodenlängenbestimmung verglichen: Nulldurchgangsbestimmung (Zero-Crossing, ZC), Maximalwertbestimmung (Peak Picking, PP) und Waveform Matching (WM). Als Testsignale dienten Sinustöne und synthetische Vokale, die mit einem parametrischen Glottismodell und einem Vokaltrakt aus Resonanzfiltern mit 5 Formanten synthetisiert wurden.

Bei diskreten Grundfrequenzen von 50Hz bis 500Hz in 50Hz Schritten wurde Jitter (entsprechend PF in Gleichung 3.11) mit den drei Verfahren in Signalen ohne Jitter gemessen. Beim Signaltyp Vokal lag das WM mit einem maximalen Messwert von 0,001% Jitter deutlich besser als ZC und PP (beide ca. 0,09%). Wurde Rauschen zu dem Signal gegeben, so erhöhte sich bei allen Verfahren zunehmend der gemessene Jitter (ohne dass Jitter eingespeist wurde). Auch hier lagen die gemessenen Jitterwerte beim WM deutlich unter denen von ZC und PP (ca. um einen Faktor 10 besser).

Die drei Verfahren zeigten bei verschiedenen Abtastfrequenzen (20 bis 50kHz) nur minimale Veränderungen des gemessenen Jitters (wieder ohne dass Jitter eingespeist wurde). Hier ist jedoch kritisch anzumerken, dass nur die Grundfrequenz 150Hz gemessen wurde. Persönliche Kommunikation mit Titze bestätigte, dass die gemessenen minimalen Jitterwerte stets besonders gering sind, wenn die Anzahl der Abtastwerte in einer Periode (oder kleine Vielfache davon) gerade eine ganze Zahl ist.

Wurde dem Signal Shimmer hinzugefügt, so zeigte sich ab ca. 6% ein deutlicher Anstieg der minimalen gemessenen Jitterwerte von ca. 0,0001% bei 1% Shimmer bis auf etwa 0,01% Jitter bei 10% Shimmer.

Das Verhalten von eingeführtem Jitter und gemessenem Jitter wurde für Jitterwerte von 0,05% bis ca. 8% in 14 diskreten Schritten bei 150Hz Grundfrequenz untersucht. In diesem Bereich zeigte sich nur beim WM ein linearer Zusammenhang zwischen eingeführtem und gemessenem Jitter.

Parsa und Jamieson untersuchten 1999 [101] 7 Verfahren zur Messung von Jitter, darunter auch das WM. Sie zeigten, dass WM bei Grundfrequenzen von 100 bis 400Hz in synthetischen Vokalen Jitter kleiner als ca. 0,002% messen konnte. Auch hier stieg der gemessene Jitter bei der Zugabe von Rauschen an, und zwar von ca. 0,01% bei 40dB Signal-Rausch-Verhältnis bis auf 1% bei 0dB. Bei der Zugabe von Shimmer stieg der mit WM gemessene Jitter von 0,002% auf ca. 0,02% bei 5dB (ca. 1.8%) Shimmer an.

Der Zusammenhang von gemessenem und zugegebenem Jitter wurde von ca. 0,2% bis 8% Jitter gemessen. Bei diesen Messungen und bei Messungen an pathologischen

Stimmen war WM stets eines der besten Verfahren.

Bei beiden Untersuchungen wurden die Grenzen für sehr kleine und für sehr große Jitterwerte nicht ausgemessen. Weiterhin wurde die Auswirkung der gemessenen Jitterwerte auf den Shimmer nicht bestimmt. Shimmer muss ja in realen Situationen auf Grund der gefundenen Periodenlängen berechnet werden. Beim Energy Perturbation Quotient werden alle Abtastwerte einer Periode quadriert und aufsummiert. Deshalb hängt der Wert der Energie in jeder Periode von den Grenzen der einzelnen Perioden und damit von den Periodenlängen selbst ab. Treten hier Ungenauigkeiten auf, so werden diese auf die Shimmermessung übertragen.

Da bei der unüberwachten Anwendung des Waveform Matching auf pathologische Stimmen auch Signale auftreten, die kaum noch eine periodische Struktur zeigen, ist es notwendig zu wissen, wie sich WM verhält, wenn sehr große Unregelmäßigkeiten im Signal auftreten, die über die in den vorherigen Studien untersuchten 8% Jitter hinausgehen. Würde sich z.B. zeigen, dass WM bei sehr irregulären Signalen plötzlich wieder kleine Jitterwerte misst, so wäre das Verfahren nicht zur unüberwachten Anwendung auf pathologische Stimmen geeignet.

Bei der unüberwachten Berechnung von Jitter und Shimmer mit diesem Verfahren ist es nicht unbedingt nötig, dass bei stark unregelmäßigen Signalen die Periodengrenzen an „sinnvollen“ Punkten liegen, also etwa dort, wo man sie mit viel Mühe von Hand hinsetzen würde. Wichtig ist nur, dass mit den Periodenlängen, die gefunden werden Jitter und Shimmer konsistent berechenbar sind.

Ausgehend von den beschriebenen Arbeiten wird hier nun das genauere Verhalten des WM bei sehr kleinen und sehr großen Jitter- und Shimmerwerten untersucht. Dabei wird auch jeweils der Shimmer variiert und gemessen.

Bei ihren Messungen mit synthetischen Vokalen gingen die vorigen Studien davon aus, dass der Jitter, der in das Glottissignal eingeprägt wurde, dem Jitter im abgestrahlten Signal des Synthetisierers entspricht, denn die Messungen des Jitters bezogen sich stets auf diesen Jitterwert. Kroeger hat bereits 1991 [66] darauf hingewiesen, das sich der gemessene Jitter und Shimmer durch die Vokaltraktfilterung grundfrequenzabhängig verändern kann (siehe Abbildung 8.10).

Deshalb werden hier zuerst Messungen im Glottissignal selbst durchgeführt, um die Genauigkeit des WM zu testen. In der Praxis ist das entsprechende akustische Signal nur mit sehr großen Schwierigkeiten invasiv zu messen. Qualitativ ist jedoch das Elektroglottogramm dem Glottissignal sehr ähnlich: Es ist noch weitgehend frei von den Oszillationen der Formanten und zeigt in einer Periode nur zwei Nulldurchgänge. Die folgenden Aussagen können deshalb auf das (undifferenzierte) elektroglottographische Signal übertragen werden.

Dazu werden jeweils Folgen von Rosenbergpulsen mit $T_O = 0,4$ und $T_C = 0,16$ und 1s Dauer erzeugt. Da die analytische Definition des Rosenbergpulses das Abtasttheorem verletzt (die Funktion besitzt einen Knick und damit beliebig hohe Frequenzen. Dies macht sich als Aliasing bemerkbar, wenn die Periodenlänge kein ganzzahliges Vielfaches der Abtastzeit ist), wurde das Signal bei einer gegebenen Abtastfrequenz mit einem zehnfach höheren Abtastwert generiert und dann korrekt (d.h. inklusive Tiefpassfilter) unterabgetastet. Die Spektren zeigten dann kein Aliasing mehr.

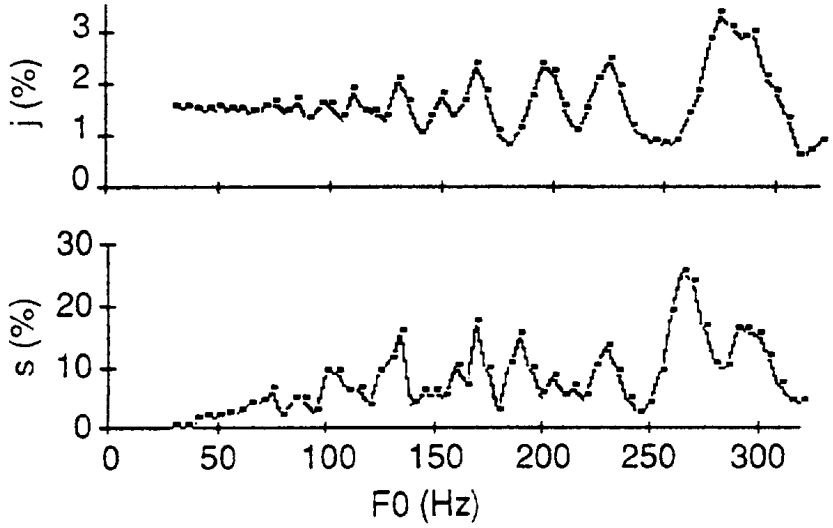


Abbildung 8.10.: Jitter und Shimmermessungen bei fest vorgegebenem glottalen Jitter eines synthetischen Vokals

Jitter und Shimmer (mit p Prozent) werden modelliert, indem die Periodenlänge bzw. die Energie pro Periode zufällig durch eine gleichverteilte Zufallszahl bestimmt wird. Dazu wird die mittlere Periodenlänge (bzw. Energie pro Periode) mit einer Zufallszahl aus dem Bereich $[1-p, 1+p]$ multipliziert. Die Werte, die im Folgenden als vorgegebener Jitter bzw. vorgegebener Shimmer bezeichnet werden, werden aus den tatsächlich realisierten Periodenlängen und Energien nach der Gleichung 3.11 berechnet ($K = 3$ beim Jitter, $K = 15$ beim Shimmer), so dass ein direkter Vergleich mit den gemessenen Werten möglich ist.

Die Grundfrequenzen, die im Folgenden untersucht werden (100Hz und 300Hz), stehen nur als Richtwert. Damit auch Periodenlängen auftreten, die kein ganzzahliges Vielfaches der Abtastperiode sind, wurde die mittlere Grundfrequenz um $\pm 5\%$ zufällig um die Werte 100Hz und 300Hz variiert.

Jitter und Shimmer wurden von $1 \times 10^{-5}\%$ bis 99.7% variiert und dann mit dem Waveform-Matching-Verfahren und dem PPQ ($K = 3$ beim Jitter) und dem EPQ ($K = 15$ beim Shimmer) gemessen. Dazu wurde hier und in den folgenden Abschnitten 0,5s aus dem mittleren Signalteil analysiert. Dabei wurde als Parameter die Abtastfrequenz auf folgende Werte gesetzt: 11025Hz, 22050Hz, 44100Hz, 88200Hz, 176400Hz. Für jede Grundfrequenz und für jede Abtastfrequenz wurden jeweils 3500 Signale erzeugt und analysiert. In den folgenden Abbildungen, in denen jeder Punkt den Messwert eines Signals darstellt, sind wegen der Übersichtlichkeit nur jeweils 1000 Punkte eingetragen.

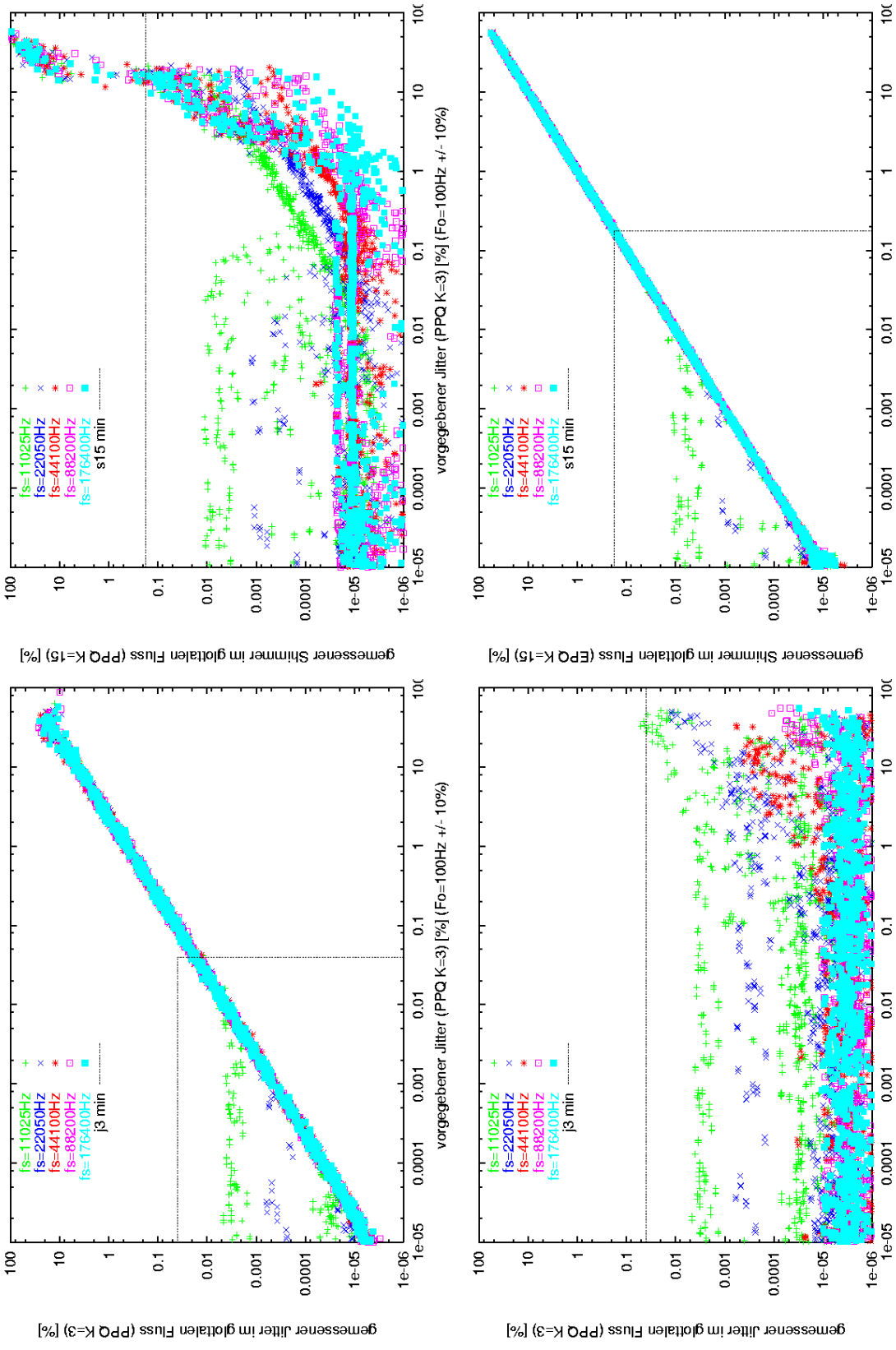


Abbildung 8.11.: Messung der Perturbationsmaße im glottischen Signal bei 100Hz Grundfrequenz

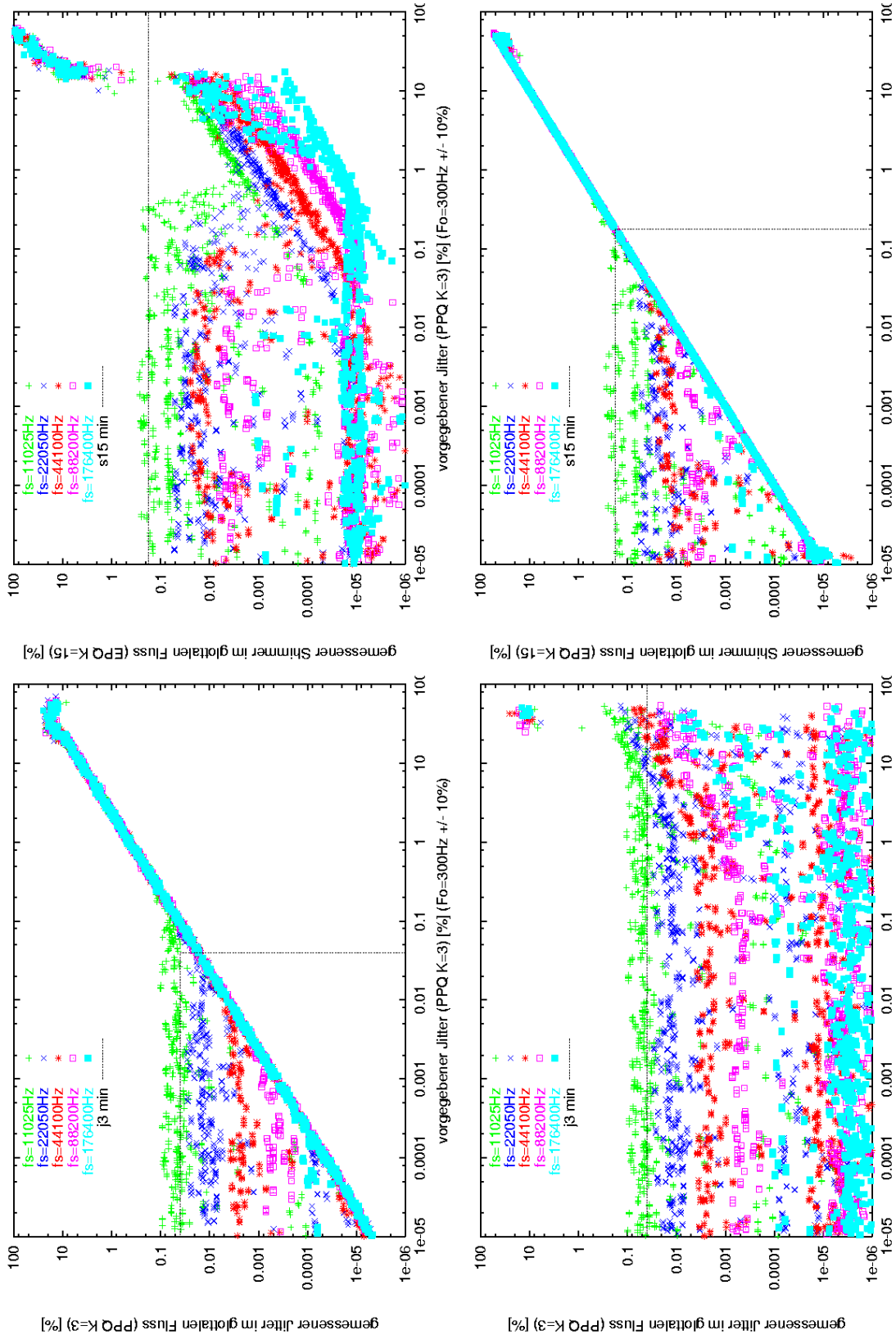


Abbildung 8.12.: Messung der Perturbationsmaße im glottischen Signal bei 300Hz Grundfrequenz

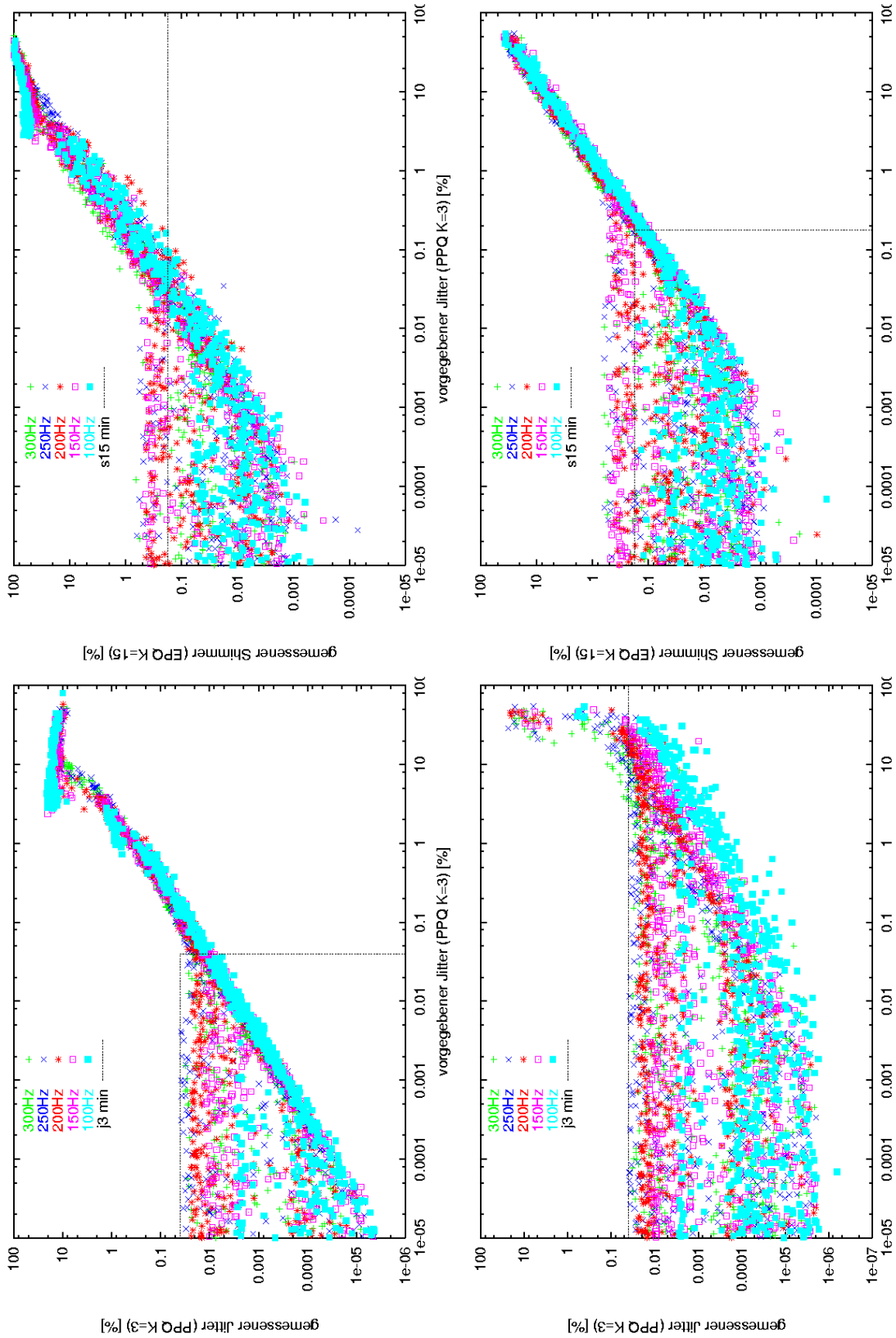


Abbildung 8.14.: Messung der Perturbationsmaße im abgestrahlten Signal

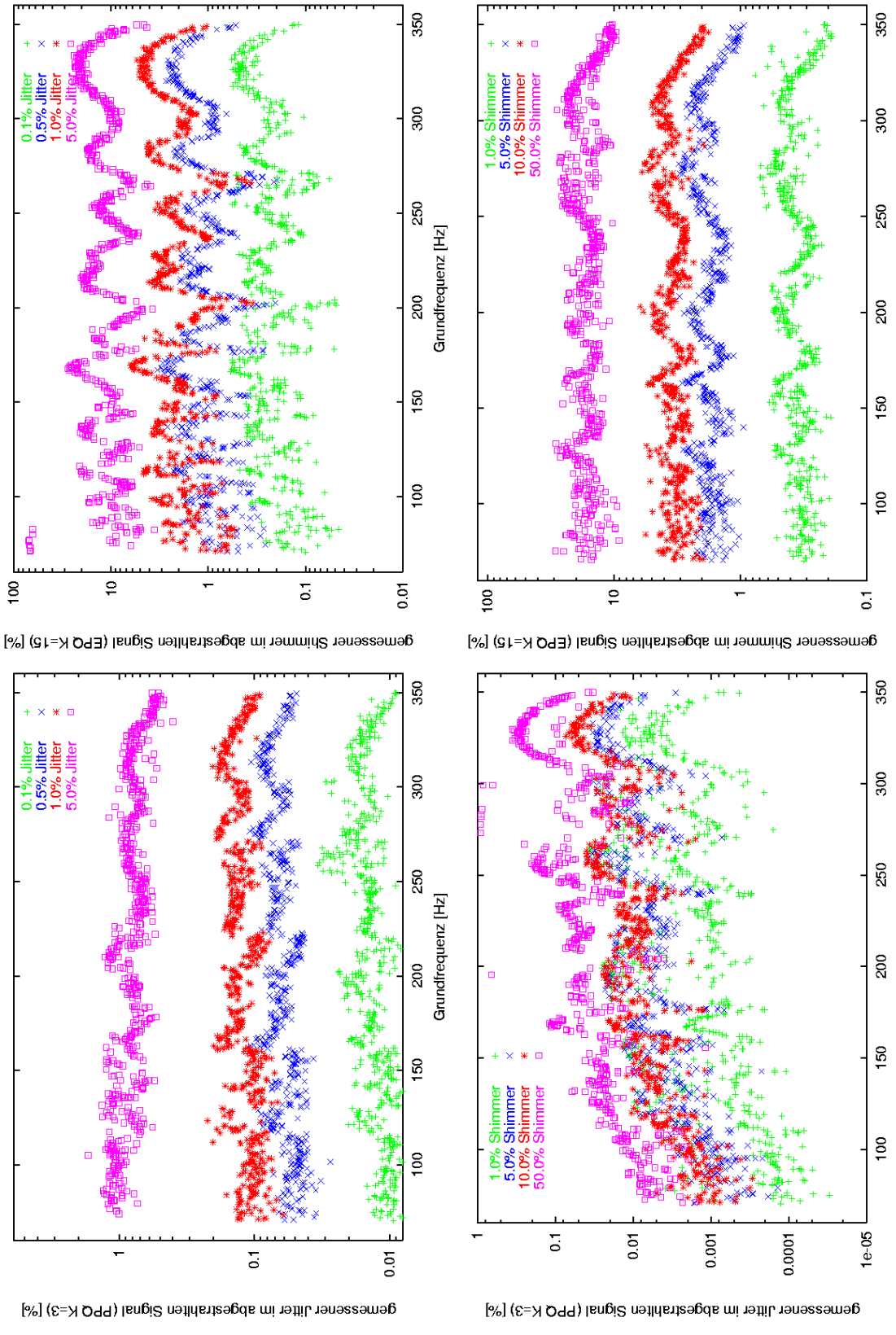


Abbildung 8.15.: Messung der Perturbationsmaße im abgestrahlten Signal in Abhängigkeit von der Grundfrequenz; /ɛ:/

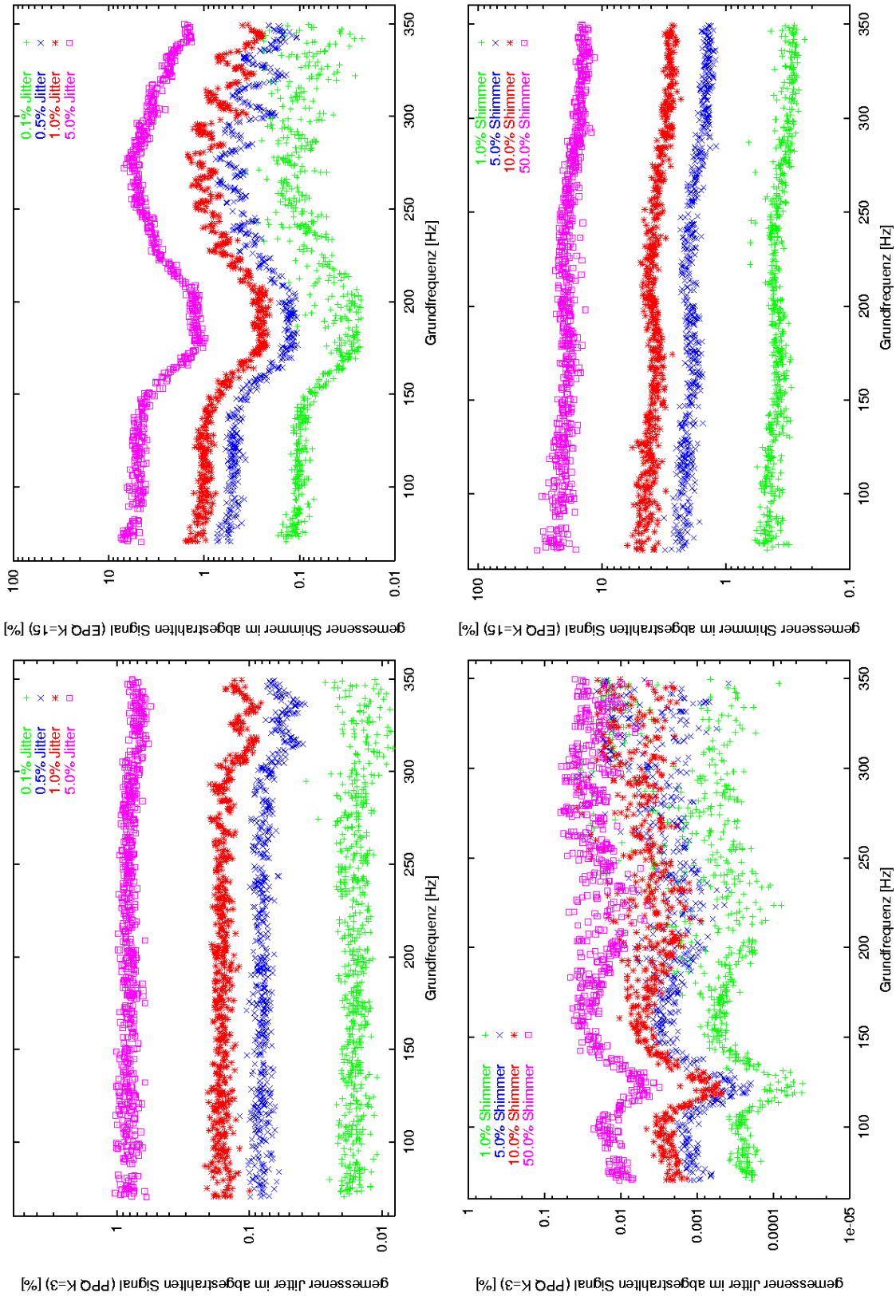


Abbildung 8.16.: Messung der Perturbationsmaße im abgestrahlten Signal in Abhängigkeit von der Grundfrequenz; /u:/

Jitter und Shimmer können durch den Vokaltraktfilter jeweils auf zwei Arten verändert werden: Der Wert des Jitters kann je nach Grundfrequenz verstärkt oder abgeschwächt werden und Jitter kann Shimmer induzieren. Shimmer kann kann je nach Grundfrequenz verstärkt oder abgeschwächt werden und Shimmer kann Jitter induzieren. Diese vier Möglichkeiten wurden bei jeweils fest vorgegebenem Jitter (0,1%, 0,5%, 1,0% und 5,0%, Abbildungen 8.15 für den Vokal [ɛ:], Abbildung 8.16 für den Vokal [u:], jeweils oben) und vorgegebenem Shimmer (1%, 5%, 10%, 50%, unten in den Abbildungen) bei Variation der Grundfrequenz von 60Hz bis 360Hz vermessen.

Die Grundfrequenzabhängigkeit der Perturbationsmaße hängt auf den ersten Blick nicht von der Stärke des vorgegebenen Jitters bzw. Shimmers ab. Als Ausnahme ist in Abbildung 8.16 das [u:] zu nennen. Hier ist die Feinstruktur, die bis 1% Jitter oberhalb von ca. 220Hz Grundfrequenz zu erkennen ist, bei 5% Jitter verschwunden.

Die Verläufe der gemessenen Jitter- und Shimmerwerte sind für einen Vokal bei den vier Messungen sehr verschieden und sehr komplex. In Abbildung 8.15 links unten und rechts oben kann man jedoch deutlich eine Struktur erkennen, die an die Übertragungsfunktion eines Nur-Nullstellenfilters erinnert: Es sind scharfe Spitzen als Minima zu erkennen und abgerundete Maxima.

Um den Einfluss der Grundfrequenz auf die gemessenen Jitter- und Shimmerwerte zu verdeutlichen, wurden bei 1% vorgegebenem Jitter und 5% vorgegebenem Shimmer jeweils der Quotient aus dem minimalen und maximalen gemessenem Jitter und Shimmerwert gebildet. Es zeigte sich, dass die Variation von Jitter bei vorgegebenem Jitter (Quotienten [ɛ:]: 3,6; [u:]: 2,6) und von Shimmer bei vorgegebenem Shimmer (Quotienten [ɛ:]: 3,2; [u:]: 2,6) noch verhältnismäßig klein sind. Die Variationen des gemessenen Jitters bei vorgegebenem Shimmer (Quotienten [ɛ:]: 18; [u:]: 7,8) und erst recht des gemessenen Shimmers bei vorgegebenem Jitter (Quotienten [ɛ:]: 290; [u:]: 190) sind dramatisch. Durch diese „Querterme“ kann man praktisch im abgestrahlten Signal bei einer beliebigen Mischung von Jitter und Shimmer im Glottissignal nicht mehr unterscheiden, ob der gemessene Jitter aus glottalem Jitter oder aus glottalem Shimmer herrührt, entsprechend für den gemessenen Shimmer. Es bleibt jedoch der Trost, dass die jeweils gemessenen Werte bei fester Grundfrequenz monoton mit den glottalen Werten zusammenhängen.

8.4. Messung der Perturbationsmaße nach Filterung mit einem Resonanzfilter

Zu einem Ansatz zum Verständnis der Beeinflussung von Jitter und Shimmer durch den Vokaltrakt gelangt man, wenn man das Problem gedanklich auf einen einfacheren Fall reduziert. Wenn man nur von der stärksten Harmonischen ausgeht, die sich meist in der Nähe des ersten Formanten befindet, so kann man drei Fälle der relativen Lage von Harmonischer und Formant unterscheiden: Die Harmonische liegt vor dem Maximum unterhalb des Formants, sie liegt genau im Maximum oder sie liegt hinter dem Maximum. Diese relative Lage bestimmt auch die Energieänderung, die eintritt, wenn sich z.B. die Frequenz der Harmonischen leicht erhöht. Im ersten Fall steigt sie an, im

zweiten Fall ändert sie sich in erster Ordnung nicht, im dritten Fall sinkt die Energie. Die Steigung der Übertragungsfunktion im Bereich des Formanten bestimmt dabei den Betrag der Energieänderung bei einer gewissen Frequenzänderung. Jitter entspricht nun einer zufälligen, kleinen, positiven oder negativen Frequenzänderung. Die daraus resultierende Energieschwankung und damit der Shimmer hängt deshalb ebenfalls von der Lage der Harmonischen ab: Ist die Steigung groß (hier ist es gleichwertig, ob die Übertragungsfunktion steigt oder fällt, da Shimmer nur die mittlere Energieschwankung misst), so wird durch den Jitter relativ starker Shimmer erzeugt. Ist die Steigung Null (im Maximum), so wird in erster Näherung kein Shimmer durch Jitter erzeugt.

Um diese Annahme zu prüfen, wurden die gleichen Experimente wie für die synthetischen Vokale [ɛ:] und [u:] mit zwei Resonanzfiltern anstelle des Synthetisators wiederholt. Die Polfrequenzen der Filter wurden entsprechend dem ersten Formanten des [ɛ:] (ca. 750Hz) und des [u:] (320Hz) gewählt. Im Anhang A wird dargestellt, warum die Resonanzfrequenz nicht exakt der Polfrequenz entspricht (siehe Gleichung A.2). Die exakten Resonanzfrequenzen liegen deshalb bei 312Hz und 746Hz.

Qualitativ zeigen die Resonanzfilter einen ähnlichen Einfluss auf Jitter und Shimmer wie die synthetischen Vokale. Die Variation des gemessenen Jitters bei vorgegebenem, festem Jitter mit der Grundfrequenz und die Variation des gemessenen Shimmers bei vorgegebenem, festem Shimmer mit der Grundfrequenz zeigen in den Abbildungen 8.17 und 8.18 einen relativ flachen Verlauf. Die gemessenen Jitter- bzw. Shimmerwerte hängen nicht sehr stark von der Grundfrequenz ab. Der Grundfrequenzverlauf der gemessenen Shimmerwerte bei festem, vorgegebenem Jitter ist bei dem 320Hz Resonanzfilter dem Verlauf bei dem Vokal [u:] sehr ähnlich. Der Verlauf des [u:] zeigt jedoch bei Grundfrequenzen über 200Hz zusätzliche Rippel (Maxima und Minima mit einem Abstand von ca. 10Hz). Die Annäherung durch einen Formanten ist beim [u:] gut, da die höheren Formanten bei diesem Vokal eine vergleichsweise kleine Amplitude besitzen.

Oben wurde erläutert, dass in erster Näherung kein Shimmer durch Jitter induziert wird, wenn eine Harmonische in den Formanten fällt. Dies wäre bei dem 320Hz-Resonanzfilter bei 312Hz (entspricht der exakten Resonanzfrequenz) und bei 156Hz (2. Harmonische) der Fall. In Abbildung 8.18 ist zu erkennen, dass in der Nähe dieser Frequenzen Minima liegen, also Stellen, wo sehr wenig Shimmer durch Jitter induziert wird. Die Annahme scheint also in grober Näherung richtig. Bei genauerem Hinsehen ist jedoch zu erkennen, dass die Minima bei etwas zu hohen Frequenzen liegen. Bei dem 750Hz-Resonanzfilter ist in Abbildung 8.17 rechts, oben ein Minimum bei ca. 250Hz zu sehen, das der Situation entspricht, in der die 3. Harmonische an der Resonanzstelle liegt (exakt wäre 249Hz). Bei beiden Resonanzfiltern treten jedoch zusätzliche Minima auf.

Weitere Minima können genau dann entstehen, wenn, vereinfacht gesagt, eine Harmonische links und eine rechts von der Resonanzstelle liegt. Und zwar genau so, dass in erster Ordnung z.B. eine kleine Frequenzerhöhung gerade die eine Harmonische so weit anhebt, wie die andere abgesenkt wird. Dazu müssen die Steigungen an den Stellen der Harmonischen entgegengesetzt gleich sein. Da man es mit mehr als zwei Harmonischen zu tun hat, muss gerade die Summe der Anhebungen auf der einen Seite der Summe der Absenkungen auf der anderen Seite entsprechen. Eine Approximation der gemessenen Shimmerwerte mit diesem Ansatz für ein Filter mit mehreren Resonanzen wurde in [82]

beschrieben. Sie führte nicht zu so guten Ergebnissen wie der weiter unten beschriebene Ansatz.

Für den quantitativen Verlauf der Grundfrequenzabhängigkeit des gemessenen Jitters bei gegebenem Shimmer konnte bislang keine befriedigende Theorie gefunden werden.

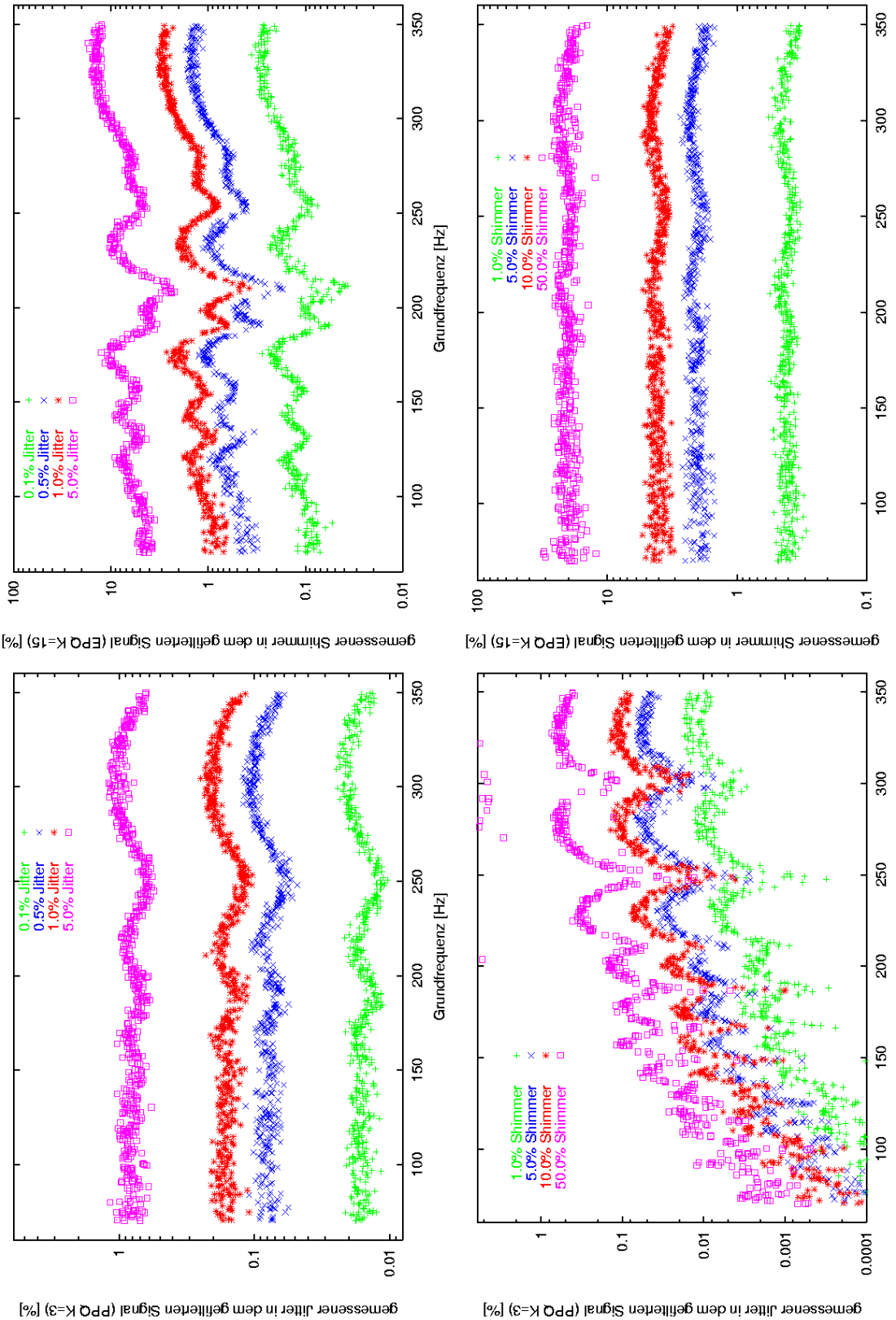


Abbildung 8.17.: Messung der Perturbationsmaße in dem mit einem 750Hz Resonanzfilter gefilterten Signal

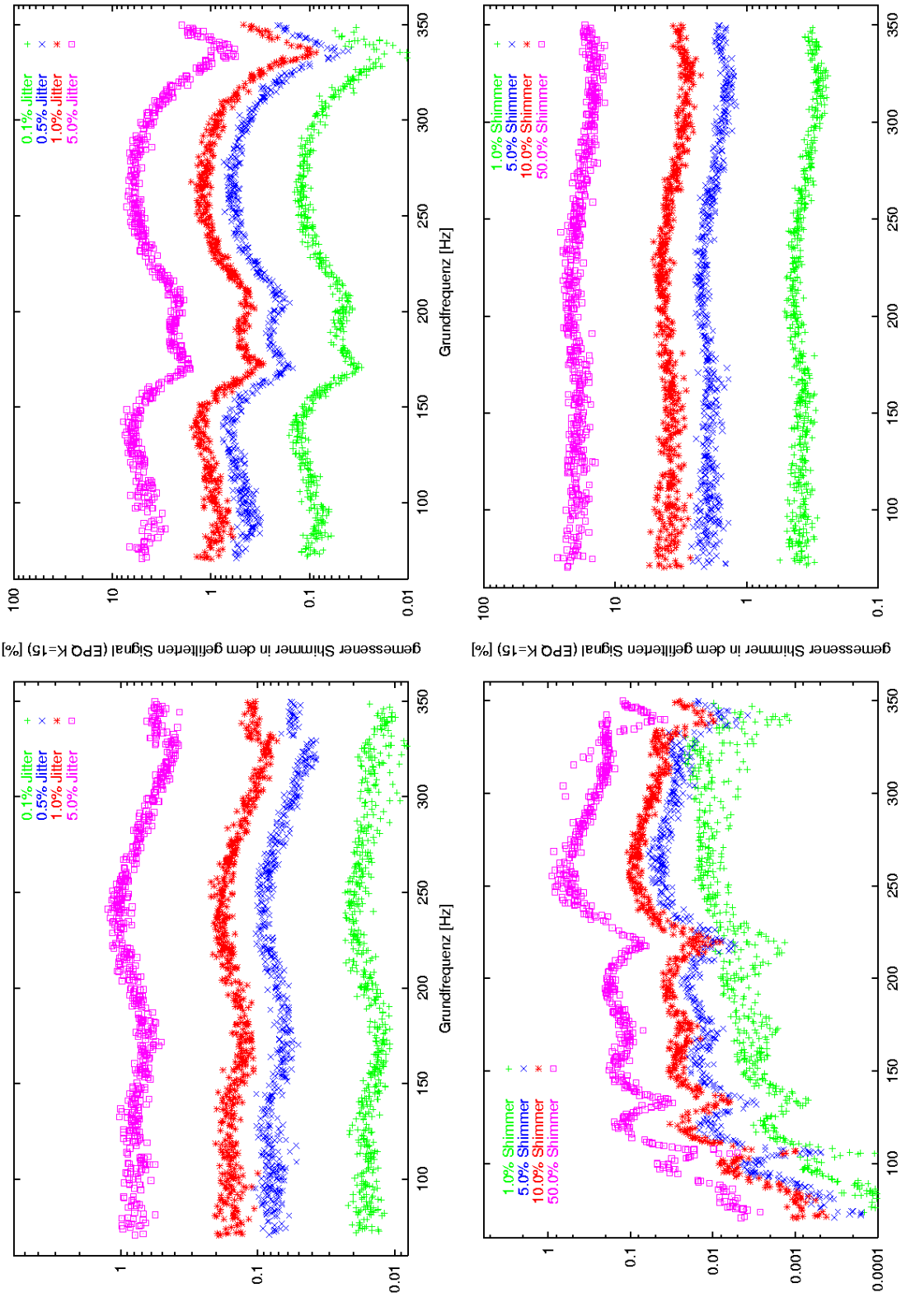


Abbildung 8.18.: Messung der Perturbationsmaße in dem mit einem 320Hz Resonanzfilter gefilterten Signal

8.5. Theoretische Beschreibung von Jitter-induziertem Shimmer

Im folgenden soll versucht werden, eine Erklärung für einen Teil der komplexen Kurven im vorherigen Abschnitt zu finden, in denen die wechselseitige Abhängigkeit von Jitter und Shimmer mit der Periodenlänge (bzw. Grundfrequenz) als Parameter deutlich wurde. Dieser Erklärungsversuch befasst sich dabei mit dem Shimmer, der im abgestrahlten (bzw. gefilterten) Signal entsteht, wenn die Anregungsfunktion shimmerfrei ist, dafür aber einen festen Jitterwert enthält.

Gesucht wird eine Gleichung, die den Wert des Energie-Perturbations-Quotienten (EPQ) in Abhängigkeit von der Grundperiode T darstellt. Da der Shimmer die mittleren relative Energieänderungen beschreibt, wenn sich die Periodenlänge von Periode zu Periode ändert, ist ein nahe liegender Ansatz:

$$\mathrm{EPQ} \propto \Delta T \frac{\frac{dE(T)}{dT}}{E(T)}. \tag{8.2}$$

Dabei beschreibt ΔT die mittlere Stärke der zufälligen Periodenlängenänderung, $E(T)$ die Energie pro Periode des Ausgangssignals in Abhängigkeit von der Grundperiode bei gleich bleibender Energie des Eingangssignals und $\frac{dE(T)}{dT}$ die Energieänderung bei Periodenlängenänderung. Bei den gemessenen Kurven wurde ΔT jeweils durch das Produkt aus einem vorgegebenem Jitter J mit der Grundperiode T und einer Zufallszahl realisiert. Deshalb kann man für 8.2 auch schreiben:

$$\mathrm{EPQ} \propto JT \frac{\frac{dE(T)}{dT}}{E(T)}. \tag{8.3}$$

Die Energie pro Periode in Abhängigkeit von der Periodenlänge $E(T)$ lässt sich berechnen, indem man den Eingangsimpuls $x(t)$ mit der (i.A. unendlichen) Impulsantwort des Systems (Resonanzfilter, Sprachsynthetisator) faltet: $y(t) = x(t) * h(t)$ und dann in folgender Weise aufsummiert:

$$E(T) = \sum_{t=0}^{T-1} [\sum_{i=-\infty}^{\infty} y(t + i\,T)]^2 \tag{8.4}$$

Da speziell für den Sprachsynthetisator die Impulsantwort schwer zu bestimmen ist, wurde für die folgenden Abbildungen $E(T)$ auf folgende Weise bestimmt: Es wurden jeweils 30 Perioden des Anregungssignals bei einer bestimmten Periodenlänge T generiert und das Ausgangssignal $y(t)$ berechnet. Nun wurden 10 eingeschwungene Perioden (von der zehnten bis zur zwanzigsten) benutzt, um daraus $E(T)$ durch Summation der quadrierten Abtastwerte zu berechnen. Die Periodenlänge T wurde zunächst für die Resonanzfilter mit 320Hz und 750Hz Resonanzfrequenz bei einer Abtastfrequenz von 44100Hz von 122 Abtastwerten bis auf 755 Abtastwerte um jeweils einen Abtastwert erhöht und so $E(T)$ berechnet. Es wurden hier nur ganzzahlige Periodenlängen genommen, um Rundungsfehler bei der Energieberechnung zu vermeiden. Um bei den synthetisierten Vokalen [ɛ:] und [u:] eine noch bessere Auflösung von $E(T)$ zu erhalten, wurde

hier die Periodenlänge bei gleicher Abtastfrequenz wie vorher von 122 Abtastwerten bis auf 755 Abtastwerten in Schritten von 0,1 Abtastwert erhöht, wobei sich bei der Energieberechnung über 10 Perioden gerade wieder keine Rundungsprobleme ergeben.

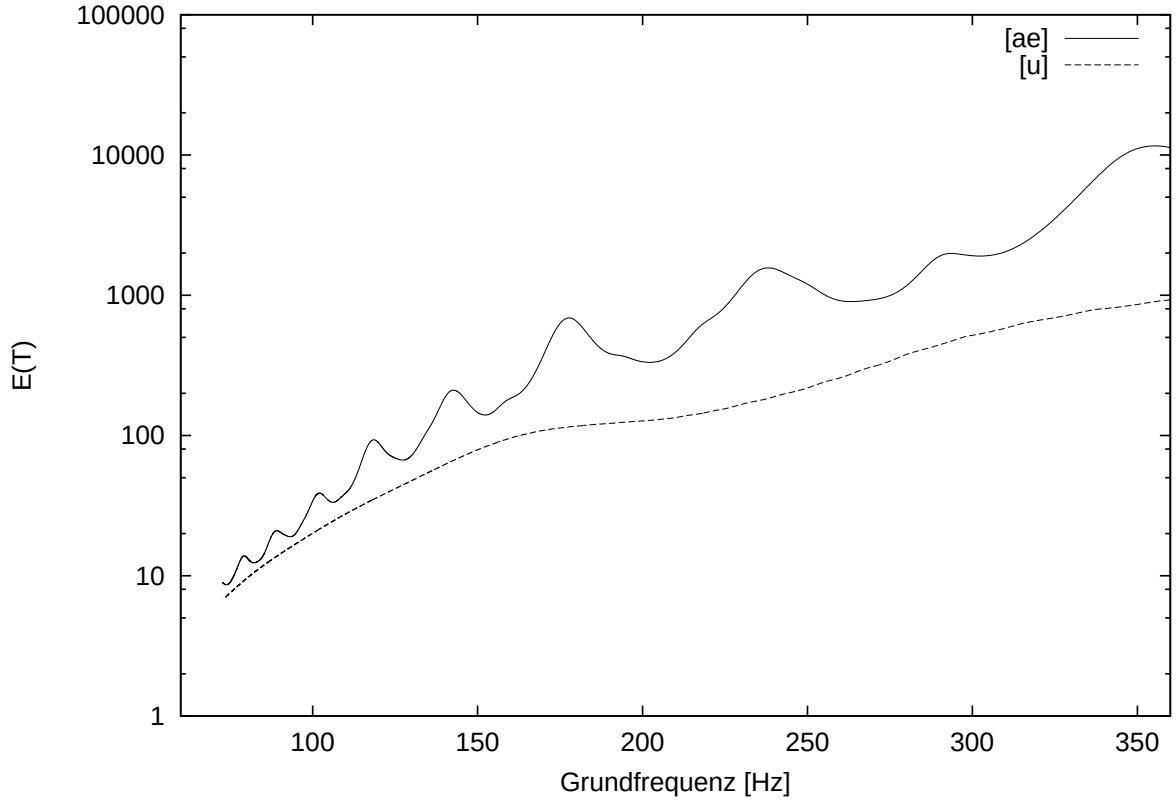


Abbildung 8.19.: Energie des Ausgangssignals pro Periode in Abhängigkeit von der Grundfrequenz $E(T)$ am Sprachsynthetisator für die Vokale [a:] (oben) und [u:] (unten) gemessen

Die Kurven $E(T)$ sind für die Vokale [ε:] und [u:] in Abbildung 8.19 dargestellt. Man beachte, dass sich die gemessene Energie pro Periode im dargestellten Grundfrequenzintervall über mehrere Zehnerpotenzen erstreckt. $\frac{dE(T)}{dT}$ wurde daraus durch numerische Differentiation berechnet. Der Proportionalitätsfaktor wurde (unabhängig von Jitter) heuristisch für die Abbildungen 8.20 bis 8.22 angepasst.

In Abbildung 8.20 sind noch einmal die Shimmermessungen bei Resonanzfilterung mit einer Resonanzfrequenz $f_r = 750$Hz bei 1% Jitter und $f_r = 320$Hz und 0,1% Jitter gezeigt. Zusätzlich sind die theoretischen Kurven (nach Gleichung 8.3 und den gemessenen $E(T)$) Kurven eingetragen.

Es ist zu erkennen, dass die gemessenen Kurven durch die hergeleitete Gleichung 8.3 sehr zufriedenstellend beschrieben werden. Speziell die Lage der Minima stimmt zwischen den berechneten und den gemessenen Kurven gut überein. Die gemessenen Kurven reichen jedoch nicht ganz in die Minima der berechneten Kurven hinab. Die Ursache

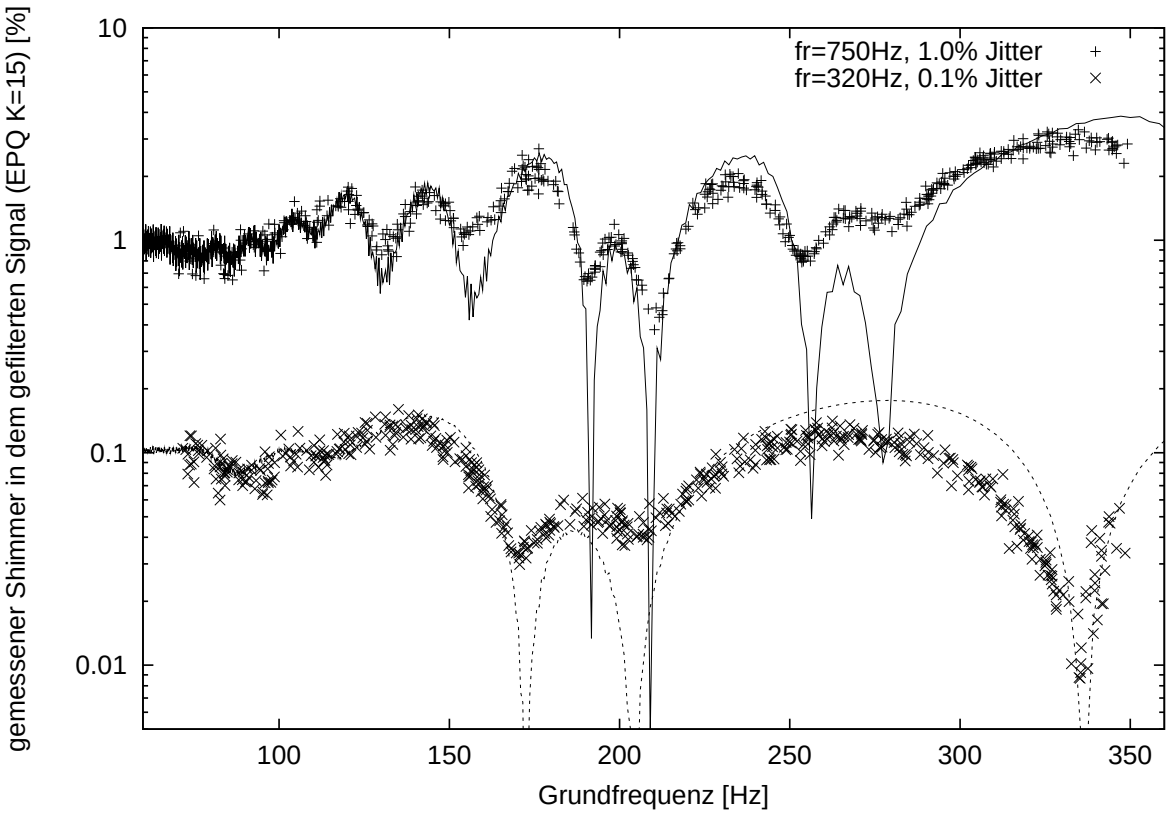


Abbildung 8.20.: Vergleich von Theorie und Messung

hierfür ist, dass die Periodenlänge T zwar bei der theoretischen Kurve exakt vorgegeben werden kann, während sie jedoch bei den gemessenen, durch den Zufallsprozess (Jitter) bedingten Kurven lokal schwankt und somit die theoretische Kurve *verschmiert* wird. Diese Verschmierung wird in den Minima am deutlichsten sichtbar. Die Stärke dieser Verschmierung hängt deshalb auch von der Stärke des Jitters ab. Dieser Effekt ist in Abbildung 8.16 bei der Shimmermessung mit 5% vorgegebenem Jitter (rechts oben) deutlich zu erkennen: Die lokalen Minima, die bei 1% Jitter noch deutlich zu erkennen sind, sind bei 5% Jitter verschwunden.

Abbildung 8.21 zeigt den gemessenen Shimmer des synthetisierten Vokals [ɛ:] bei fest vorgegebenem Jitter von 0,5% und die hierfür nach Gleichung 8.3 berechnete theoretische Kurve. Bis auf die Tatsache, dass die gemessene Kurve aus dem oben beschriebenem Grunde nicht ganz in die Minima hinabreicht (z.B. bei F_0=300Hz), kann man hier schon fast von einer erstaunlichen Übereinstimmung von Theorie und Messung sprechen. Die gerade bei niedrigen Grundfrequenzen zahlreichen Maxima und Minima werden sowohl nach Lage als auch nach ihrem Betrag sehr gut durch Gleichung 8.21 beschrieben.

In Abbildung 8.22 sind die Messkurve vom Shimmer bei festem Jitter von 1% und die theoretische Kurve für den Vokal [u:] dargestellt. Bei gleichem Proportionalitätsfaktor wie in Abbildung 8.21 ist auch hier eine gute Übereinstimmung der Lage von Minima

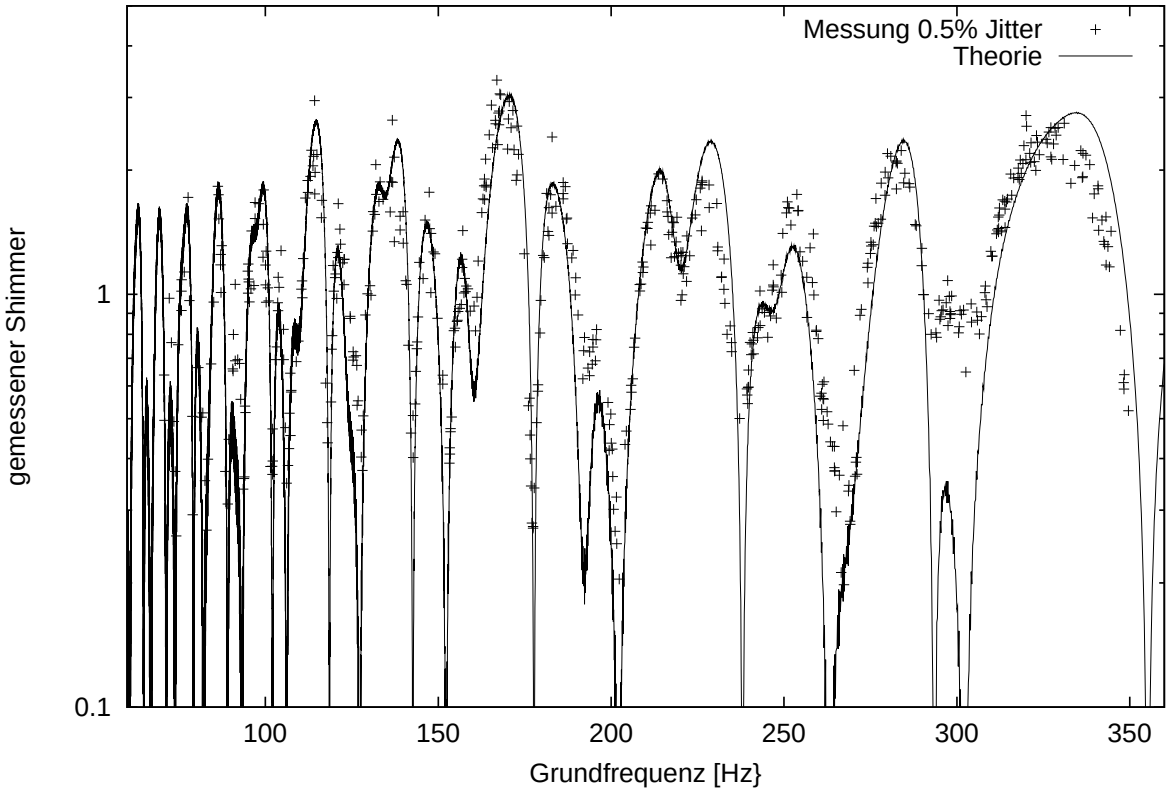


Abbildung 8.21.: Vergleich von Theorie und Messung: [ε:]

und Maxima zu erkennen. Die Anpassung des Betrages ist hier jedoch nicht ganz so gut wie bei dem [ε:] in Abbildung 8.21. Der Grund für diese etwas schlechtere Übereinstimmung könnte in dem relativ flachen Verlauf von $E(T)$ (Abbildung 8.19) liegen, der zu numerischen Schwierigkeiten bei der diskreten Differenzierung führen kann.

Neben numerischen Problemen ist aber auch prinzipiell ein Abweichen der Messungen von der Theorie bei hohen Grundfrequenzen zu erwarten. Denn der Ansatz, dass die mittlere relative Energieschwankung (gemessen durch EPQ) im Wesentlichen von $dE(T)/dT$ abhängt, geht implizit davon aus, dass die Energie in einer Periode nur durch den Ausschwingvorgang der vorherigen Periode beeinflusst wird. Bei sehr kurzen Grundperioden muss man jedoch davon ausgehen, dass nicht nur der Ausschwingvorgang der direkt vorhergehenden Periode in die momentane Periode hineinreicht, sondern auch noch der Ausschwingvorgang der vorvorherigen Periode. Hier hängt dann die Energie der momentanen Periode von den genauen Periodenlängenverhältnissen der beiden vorigen Perioden ab. Es gibt dann zahlreichere Möglichkeiten (z.B. kürzere-längere, kürzere-kürzere, längere-kürzere usw.), als wenn nur die vorhergehende Periode betrachtet werden muss. Dieses komplizierte Zusammenspiel der nun drei oder mehr beteiligten Periodenlängen wird durch Gleichung 8.22 nicht mehr korrekt erfasst. So kann vielleicht die Abweichung in Abbildungen 8.20, 8.21 und 8.22 bei hohen Grundfrequenzen gedeutet werden.

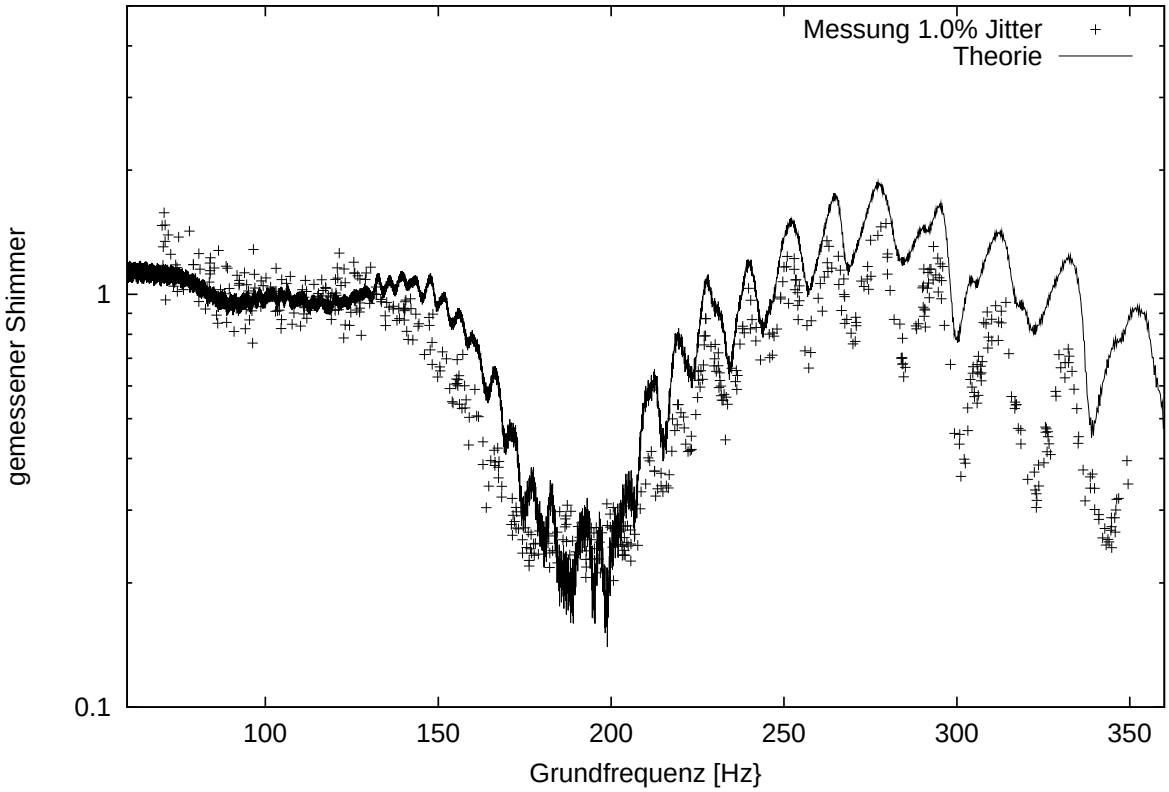


Abbildung 8.22.: Vergleich von Theorie und Messung [u:]

8.6. Folgerung

Insgesamt kann man aus der sehr guten Übereinstimmung von theoretischer Vorhersage des zu erwartenden Shimmers bei festem Jitter und bei Kenntnis der Energiekurve des Systems $E(T)$ sagen, dass mit Gleichung 8.3 erstmalig eine zufrieden stellende quantitative Aussage über den Einfluss des supraglottalen Systems auf Perturbationen im glottalen System gemacht werden konnte.

Durch die Untersuchung des Waveform Matching an synthetischen Signalen konnte sichergestellt werden, dass die Auflösung zur Berechnung sehr kleiner Jitter- und Shimmerwerte bei 44100Hz Abtastfrequenz ausreichend ist. Außerdem zeigte sich, dass bei sehr hohen Jitter- und Shimmerwerten keine genaue Differenzierung mehr zwischen Jitter und Shimmer möglich ist. Dies spielt jedoch bei der Berechnung des Heiserkeits-Diagramms nur eine untergeordnete Rolle, denn bei der Entwicklung des Heiserkeits-Diagramms hatte sich ja gerade gezeigt, dass Jitter und Shimmer bei pathologischen Stimmen sehr hoch korreliert sind, und zwar auch bei Jitter- und Shimmerwerten, die noch eindeutig mit dem Waveform Matching berechnet werden können.

Damit ist die Untersuchung der rein akustischen Methoden, die die Grundlage des Heiserkeits-Diagramms bilden, abgeschlossen. Im zweiten Teil folgen nun Anwendun-

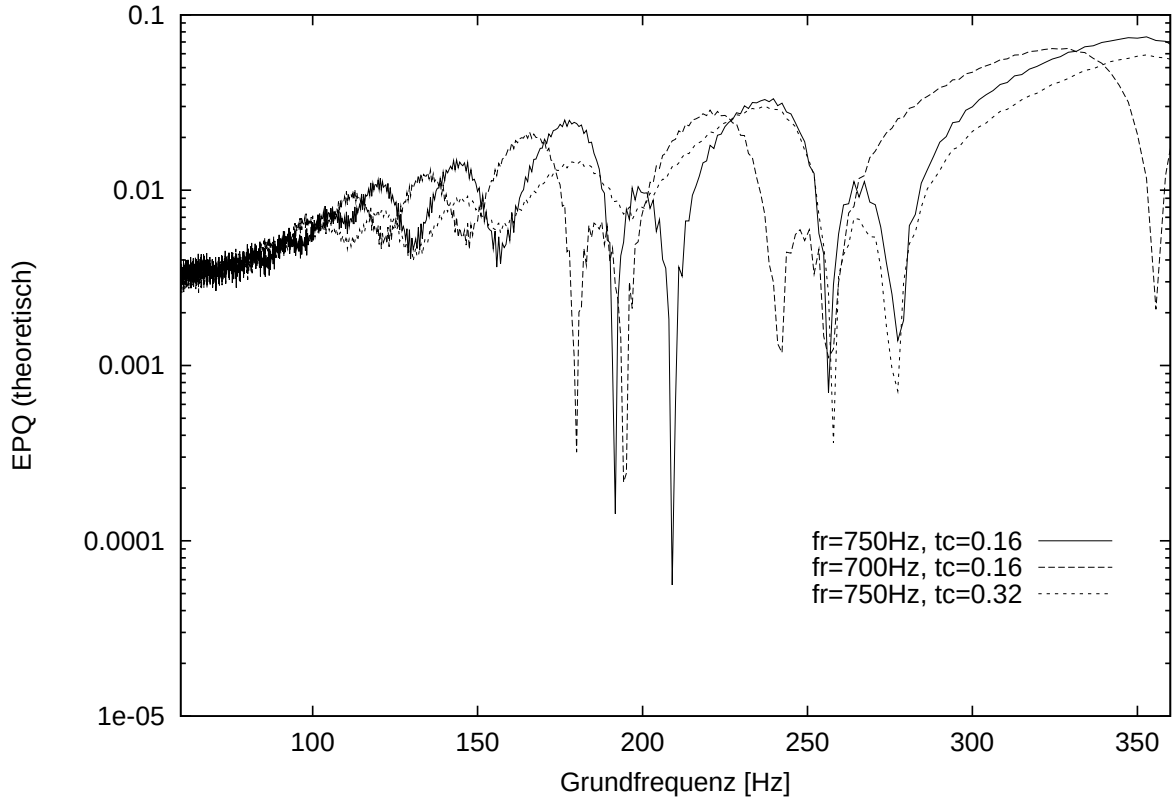


Abbildung 8.23.: Verschiedene Resonanzfrequenzen bzw. Verschlusszeiten (Tc) beim Resonanzfilter

gen des Heiserkeits-Diagramms, in denen sich zeigen soll, ob sich die Darstellung der Stimmgüte in den zwei Dimensionen „Irregularitätskomponente“ und „Rauschkomponente“ in der klinischen Anwendung und Forschung bewährt.

Teil II.

Anwendungen des Heiserkeits-Diagramms

9. Von der Signalverarbeitung zur interdisziplinären Forschung

Aus der Sicht des Signalverarbeiters könnte die Arbeit an dieser Stelle schließen. Für den Phoniater oder Sprachforscher beginnt jedoch jetzt der spannendere Teil der Arbeit: Der Einsatz der entwickelten Methoden in der Praxis [29, 31, 68, 133]. Das getrennte Nacheinander von Entwicklung und Anwendung des Heiserkeits-Diagramms, wie es in dieser Arbeit beschrieben wird täuscht darüber hinweg, dass die Methodik in enger Wechselwirkung mit dem klinischen Umfeld der Phoniatrie und Pädaudiologie in Göttingen entstanden ist. Die Entwicklung vollzog sich schrittweise von der ersten Implementation und Anwendung einiger Stimmgüteparameter bis zur Definition des Heiserkeits-Diagramms mit Hilfe der Datenraumanalyse der umfangreichen Stimmaufnahmen von zahlreichen Patienten. Auch liefert erst das klinische Umfeld die Rückmeldung, ob die Zwischenergebnisse in die richtige Richtung führten.

Nicht zuletzt wurde diese Arbeit wesentlich durch ein interdisziplinäres Projekt der Deutschen Forschungsgemeinschaft finanziell unterstützt, bei dem der Schwerpunkt auf der klinischen Anwendung liegt.

10. Statistische Methoden und mehrdimensinale Abbildungsverfahren

Zur Auswertung der Daten des Heiserkeits-Diagramms werden im Folgenden einige zusätzliche statistische Verfahren verwendet, die hier kurz erläutert werden. Darüber hinaus werden zwei mehrdimensinale Abbildungsverfahren vorgestellt, lineare Regression und Abbildung durch ein Backpropagation-Netzwerk, die in Kapitel 15 verwendet werden.

10.1. Zweidimensionaler Kolmogorov-Smirnov-Test

Dieser Test basiert auf Unterschieden der Verteilungsfunktionen und gibt Auskunft darüber, ob sich die Verteilungen von zwei zweidimensionalen Datensätzen signifikant unterscheiden. Dabei können sich die Datensätze jedoch in ihrer Varianz oder ihrem Mittelwert (logisches oder, d.h. oder auch in beidem) unterscheiden. Aus den Werten für Mittelwert und Varianz kann man jedoch beurteilen, worauf der Unterschied der Verteilungen basiert.

10.2. Lineare Regression und Abbildung durch ein „Backpropagation- Netzwerk“

In Kapitel 15 wird untersucht, inwieweit sich das umfangreiche Aufnahmeprotokoll, das dem Heiserkeits-Diagramm zugrunde liegt (viermal sieben gehaltene Vokale) reduzieren lässt. Dazu wird der Mittelwert der Irregularitätskomponente und der Rauschkomponente von nur einem, zwei oder drei Vokalen mit dem Gesamtmittelwert der 28 Vokale verglichen. Wie weiter unten noch gezeigt werden wird, liegen die Vokale jedoch im Mittel an ganz speziellen Positionen bezogen auf den Gesamtmittelwert, so dass vor dem Vergleich von einzelnen Vokalen mit dem Gesamtmittelwert erst ein systematischer Trend herausgerechnet werden soll. Dieses Herausrechnen erfolgt auf zwei Arten: Erstens wird eine multidimensionale lineare Regression durchgeführt, um den (nun linear angenommenen) systematischen Trend zu berücksichtigen. Weiterhin wird eine neuronales Netz zur Vorhersage des Gesamtmittelwertes aus nur einem, zwei oder drei Vokalen

benutzt, um auch eventuelle nichtlineare Zusammenhänge von Heiserkeits-Diagramm-Koordinaten der Einzelvokale und dem Gesamtmittelwert zu berücksichtigen.

10.2.1. Beschreibung des Lernalgorithmus des neuronalen Netzes

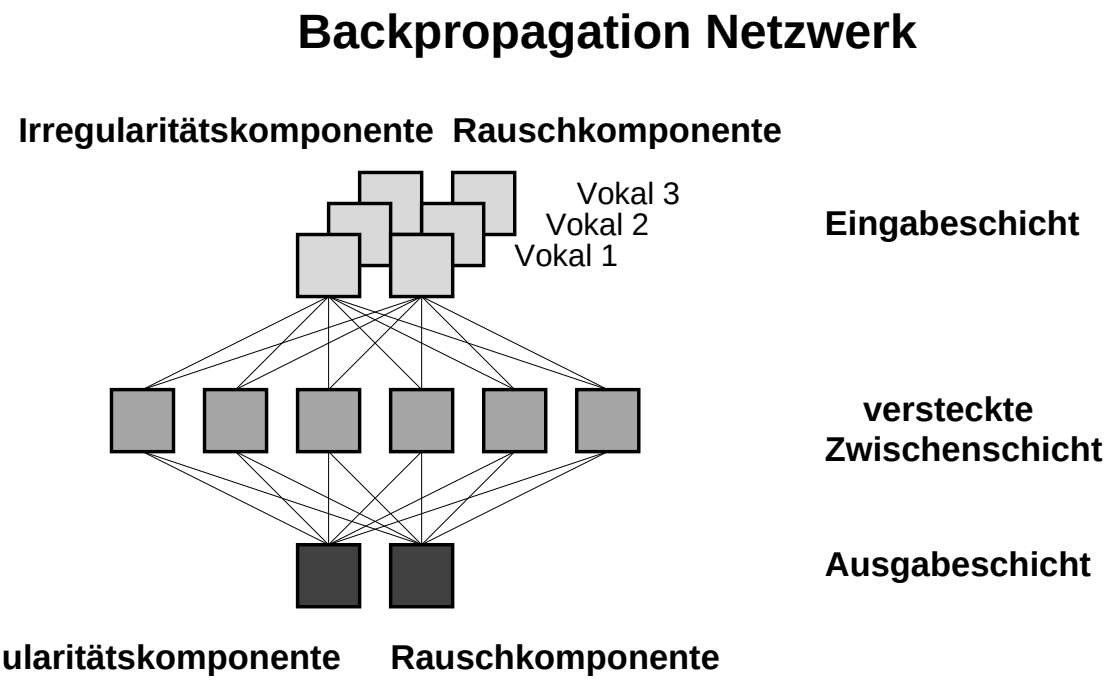


Abbildung 10.1.: Aufbau des verwendeten neuronalen Netzes: 2,4, oder 6 Eingabezellen, 6 versteckte Zellen und zwei Ausgabezellen.

Der Lernalgorithmus des Backpropagation-Netzwerks ist eine leicht modifizierte Version des von Müller in [91] vorgestellten Algorithmus.

Das Netz bestehe aus N_e Eingabeneuronen σ_k, N_v versteckten Neuronen s_j und N_a Ausgabeneuronen S_i. Die Trainingsmenge umfasse N_m Trainingsmuster für die Eingabe σ^μ mit jeweils N_k Komponenten σ_k^μ und M Trainingsmuster für die Ausgabe ζ^μ mit jeweils N_i Komponenten ζ_i^μ.

Die Aktivität der Ausgabeneuronen S_i ist dann:

$$S_i = f(h_i) \qquad h_i = \sum_{j=1}^{N_v} \omega_{ij} s_j - \vartheta_i, \tag{10.1}$$

wobei f die Aktivierungsfunktion und ϑ_i der Schwellenwert des i-ten Ausgabeneurons und ω_{ij} die Gewichtsmatrix ist. Entsprechend ist die Aktivität der versteckten Zwischenschicht

$$\bar{s}_j = f(\bar{h}_j) \qquad \bar{h}_j = \sum_{k=1}^{N_e} \bar{\omega}_{jk} \sigma_k - \bar{\vartheta}_j. \tag{10.2}$$

Die Gewichtsmatrizen und die Schwellenwerte sollen nun so gewählt werden, dass die Fehlerfunktion

$$D[\omega_{ij}, \vartheta_i, \bar{\omega}_{jk}, \bar{\vartheta}_j] = \frac{1}{2} \sum_{\mu}^{N_m} \sum_{i}^{N_a} [\zeta_i^\mu - f(h_i^\mu)]^2 \tag{10.3}$$

minimal wird. Zur Annäherung an dieses Minimum dient ein iteratives Gradientenverfahren, bei dem die Gewichte und die Schwellen gemäß

$$\begin{aligned} \omega'_{ij} &= \omega_{ij} + \delta\omega_{ij} \\ \vartheta'_i &= \vartheta_i + \delta\vartheta_i \\ \bar{\omega}'_{jk} &= \bar{\omega}_{jk} + \delta\bar{\omega}_{jk} \\ \bar{\vartheta}'_j &= \bar{\vartheta}_j + \delta\bar{\vartheta}_j \end{aligned} \tag{10.4}$$

angepasst werden, wobei

$$\begin{aligned} \delta\omega_{ij} &= -\epsilon \frac{\partial D}{\partial \omega_{ij}} = \epsilon \sum_{\mu}^{N_m} [\zeta_i^\mu - f(h_i^\mu)] f'(h_i^\mu) \frac{\partial h_i^\mu}{\partial \omega_{ij}} \\ &= \epsilon \sum_{\mu}^{N_m} \Delta_i^\mu s_j^\mu \\ \delta\vartheta_i &= -\epsilon \frac{\partial D}{\partial \vartheta_i} = \epsilon \sum_{\mu}^{N_m} [\zeta_i^\mu - f(h_i^\mu)] f'(h_i^\mu) \frac{\partial h_i^\mu}{\partial \vartheta_i} \\ &= -\epsilon \sum_{\mu}^{N_m} \Delta_i^\mu \\ \delta\bar{\omega}_{jk} &= -\epsilon \frac{\partial D}{\partial \bar{\omega}_{jk}} = \epsilon \sum_{i}^{N_a} \sum_{\mu}^{N_m} [\zeta_i^\mu - f(h_i^\mu)] f'(h_i^\mu) \frac{\partial h_i^\mu}{\partial s_j} \frac{\partial s_j}{\partial \bar{\omega}_{jk}} \\ &= \epsilon \sum_{i}^{N_a} \sum_{\mu}^{N_m} \Delta_i^\mu \omega_{ij} f'(\bar{h}_j^\mu) \frac{\partial \bar{h}_j^\mu}{\partial \bar{\omega}_{jk}} = \epsilon \sum_{\mu}^{N_m} \bar{\Delta}_j^\mu \sigma_k^\mu \\ \delta\bar{\vartheta}_j &= -\epsilon \frac{\partial D}{\partial \bar{\vartheta}_j} = \epsilon \sum_{i}^{N_a} \sum_{\mu}^{N_m} [\zeta_i^\mu - f(h_i^\mu)] f'(h_i^\mu) \frac{\partial h_i^\mu}{\partial s_j} \frac{\partial s_j}{\partial \bar{\vartheta}_j} \\ &= \epsilon \sum_{i}^{N_a} \sum_{\mu}^{N_m} \Delta_i^\mu \omega_{ij} f'(\bar{h}_j^\mu) \frac{\partial \bar{h}_j^\mu}{\partial \bar{\vartheta}_j} = -\epsilon \sum_{\mu}^{N_m} \bar{\Delta}_j^\mu \end{aligned} \tag{10.5}$$

sind mit

$$\begin{aligned} \Delta_i^\mu &= [\zeta_i^\mu - f(h_i^\mu)] f'(h_i^\mu) \\ \bar{\Delta}_j^\mu &= (\sum_{i}^{N_a} \Delta_i^\mu \omega_{ij}) f'(\bar{h}_j^\mu). \end{aligned} \tag{10.6}$$

Dabei ist ϵ der Lernparameter, der angibt, wie stark die Gewichte in einem Lernschritt angepasst werden.

Der Lernalgorithmus wurde dahingehend verändert, dass eine gewisse Trägheit in die Änderung der Gewichte implementiert wurde: Seien $\delta\omega'_{ij}$, $\delta\vartheta'_i$, $\delta\bar{\omega}'_{jk}$ und $\delta\bar{\vartheta}'$ die Gewichtsänderungen aus dem vorhergehenden Iterationsschritt, so lauten die modifizierten Gewichtsänderungen:

$$\begin{aligned} \delta\omega''_{ij} &= \eta\delta\omega'_{ij} + \delta\omega_{ij} \\ \delta\vartheta''_i &= \eta\vartheta'_i + \delta\vartheta_i \\ \delta\bar{\omega}''_{jk} &= \eta\delta\bar{\omega}'_{jk} + \delta\bar{\omega}_{jk} \\ \delta\bar{\vartheta}''_j &= \eta\bar{\vartheta}'_j + \delta\bar{\vartheta}_j \end{aligned} \tag{10.7}$$

mit dem Trägheitsparameter η, $\quad 0 \leq \eta \leq 1$.

11. Datenmaterial

11.1. Stimmaufnahmen pathologischer und normaler Sprecher

Seit 1995 werden am Universitätsklinikum Göttingen in einem schallisolierten Raum Stimmen der Patienten mit einem DAT-Recorder aufgenommen. Bei dem Mikrofon handelt es sich um ein ein bayerdynamics Kopfmikrofon mit Kugelcharakteristik. Das Mikrofon sitzt am Ende eines Schwanenhalses und wird seitlich unterhalb des Mundes plaziert (ca. 10cm vom Mund entfernt) außerhalb des Luftstromes, der bei Plosiven oder z.B. beim [u:] erzeugt wird. Die Abtastfrequenz des DAT-Recorders beträgt 48kHz. Neben dem akustischen Signal wird auch ein EGG auf dem zweiten Kanal des DAT Recorders gespeichert. Das Aufnahmeprotokoll ist relativ umfangreich, um viele Eigenschaften der Stimme zu erfassen. Es umfasst:

- Die gehaltenen Vokale $[\varepsilon:]_1$, [a:], [e:], [i:], [o:], [u:], $[\varepsilon:]_2$, bei bequemer Grundfrequenz und Lautstärke
- Die gleiche Vokalreihe bei etwas tieferer Grundfrequenz
- Die Vokalreihe bei erhöhter Grundfrequenz
- Lesen des phonetisch ausgeglichenen Standardtextes „Nordwind und Sonne“
- Wiederholung der Vokalreihe bei bequemer Grundfrequenz und Lautstärke

Die Patienten werden angehalten, die Vokale möglichst lange, mindestens aber 3s zu halten. Bei manchen Stimmstörungen ist dies jedoch nicht möglich. In diesen Fällen phoniert der Patient, solange er es aufgrund der Stimmstörung vermag. Evtl. fallen dabei auch einige Vokalreihen weg. Die überwiegende Mehrzahl der Patienten ist jedoch in der Lage, das gesamte Aufnahmeprotokoll zu leisten.

Nach diesem Protokoll wurden seit 1995 etwa 1500 Stimmen aufgenommen, wobei von einigen Patienten mehrere Aufnahmen vorliegen, die es gestatten, den zeitlichen Verlauf der Stimmgüte zu verfolgen.

In den Abbildungen 11.1 bis 11.4 sind die Heiserkeits-Diagramme von zwei Normalsprechern dargestellt. Das erste Blatt zeigt jeweils das Ergebnis der Stimmanalyse für einen bestimmten Aufnahmetag, das zweite zeigt einen Verlauf der Stimmgüte über

einen Zeitraum von mehreren Jahren. Im ersten Datenblatt werden die Analyseergebnisse der einzelnen Vokale durch ein offenes Symbol, einem Linienzug und ein geschlossenes Symbol dargestellt. Das offene Symbol, jeder Eckpunkt des Linienzuges und das geschlossene Symbol stehen jeweils für ein analysiertes Signalsegment mit 0,5s Länge. Der Segmentvorschub beträgt 0,25s. In der Liste neben den Symbolen sind die Mittelwerte der akustischen Maße für je einen Vokal dargestellt. Diese Maße sind: Periodenkorrelation (MWC), Jitter (PPQ K=3, J3), Shimmer (EPQ K=15, S15), GNE, Grundfrequenz F_0, Sound Pressure Level (SPL), die Irregularitätskomponente (siehe 7.10) mit Standardabweichung und die Rauschkomponente (siehe 7.11) mit Standardabweichung. Am Ende der Liste sind die Mittelwerte über alle Vokale angegeben. Im Diagramm ist zusätzlich eine Ellipse eingezeichnet. Das Zentrum der Ellipse entspricht den Mittelwerten der Irregularitäts- und Rauschkomponente. Die Länge der Halbachsen der Ellipse entspricht einer Standardabweichung, so dass die gesamte Höhe bzw. Breite der Ellipse zwei Standardabweichungen in der jeweiligen Komponente entsprechen. Die hellgrau unterlegte Fläche in der linken unteren Ecke markiert einen Bereich, in dem 95% der Datenpunkte einer Stichprobe von 32 Normalsprechern und -sprecherinnen lagen.

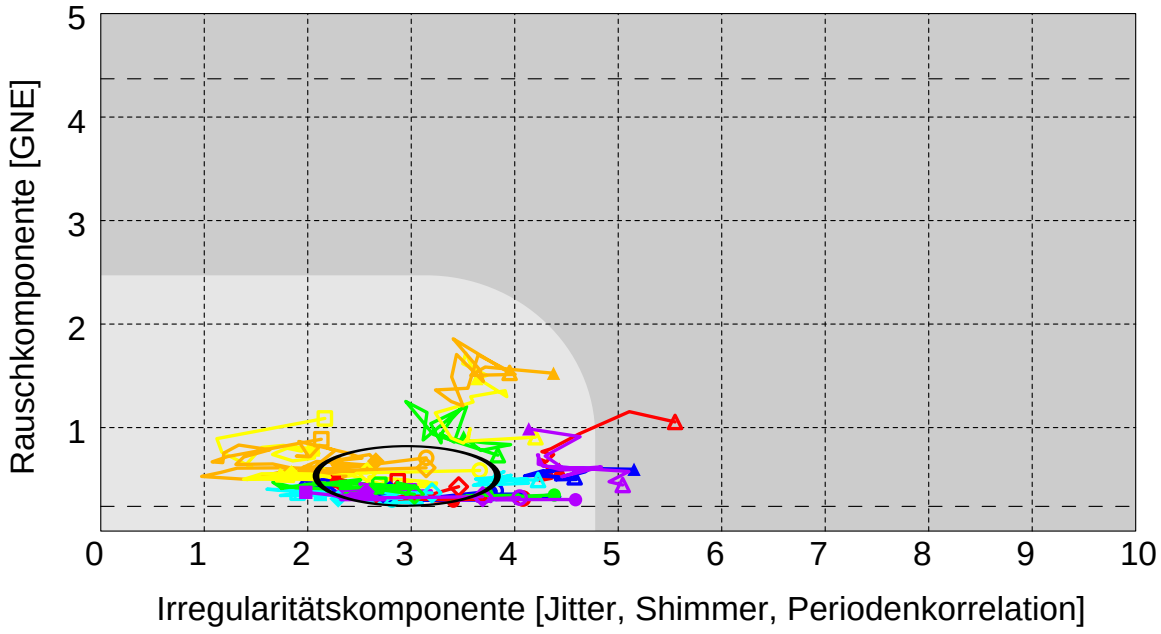
Diese beiden Datenblätter werden für jeden Patienten berechnet und in die Patientenakte aufgenommen.

Heiserkeits-Diagramm

Name: fm280867

Datum: fr.28.06.96
normal '0'

Modus	Perioden Korrelat.	Jitter [%]	Shimmer [%]	GNE	F0 [Hz]	SPL [dB]	Irreg.komp.	Rauschkomp.
normal, ä 1	0.990	0.10	2.5	0.98	124.4	68.3	3.4 (0.24)	0.3 (0.01)
normal, a	0.990	0.17	3.0	0.98	122.7	68.6	3.7 (0.35)	0.3 (0.03)
normal, e	0.994	0.06	1.9	0.99	121.5	66.9	2.8 (0.16)	0.3 (0.02)
normal, i	0.980	0.13	2.3	0.98	121.5	63.8	3.8 (0.25)	0.3 (0.01)
normal, o	0.998	0.21	2.0	0.94	121.8	67.1	2.7 (0.43)	0.5 (0.05)
normal, u	0.999	0.08	1.9	0.92	120.0	67.1	2.3 (0.28)	0.6 (0.07)
normal, ä 2	0.991	0.21	3.2	0.99	120.2	67.7	3.8 (0.30)	0.3 (0.01)
tief, ä 1	0.984	0.22	5.3	0.89	102.0	63.5	4.4 (0.35)	0.7 (0.19)
tief, a	0.971	0.14	4.8	0.92	100.8	63.6	4.5 (0.26)	0.6 (0.04)
tief, e	0.988	0.15	3.0	0.94	102.0	62.7	3.8 (0.19)	0.5 (0.03)
tief, i	0.995	0.14	3.0	0.83	102.4	62.3	3.4 (0.26)	1.0 (0.15)
tief, o	0.996	0.23	3.9	0.76	102.5	63.6	3.6 (0.22)	1.2 (0.28)
tief, u	0.997	0.25	4.6	0.69	102.5	62.2	3.6 (0.28)	1.5 (0.16)
tief, ä 2	0.985	0.23	6.0	0.90	101.7	63.2	4.5 (0.29)	0.6 (0.13)
hoch, ä 1	0.997	0.05	1.6	0.95	170.8	71.3	2.4 (0.27)	0.4 (0.04)
hoch, a	0.998	0.04	1.5	0.96	171.2	72.7	2.1 (0.11)	0.4 (0.03)
hoch, e	0.996	0.03	1.3	0.97	172.9	68.0	2.2 (0.29)	0.4 (0.03)
hoch, i	0.996	0.04	1.3	0.95	175.2	66.1	2.2 (0.33)	0.4 (0.03)
hoch, o	0.999	0.06	1.4	0.87	173.4	70.3	1.8 (0.31)	0.8 (0.10)
hoch, u	0.999	0.06	1.2	0.90	174.0	70.2	1.5 (0.37)	0.7 (0.10)
hoch, ä 2	0.996	0.10	1.5	0.97	172.3	72.0	2.7 (0.33)	0.4 (0.03)
belastet, ä 1	0.995	0.13	2.2	0.98	130.5	69.7	3.2 (0.28)	0.3 (0.03)
belastet, a	0.997	0.08	1.8	0.98	131.3	71.0	2.7 (0.15)	0.3 (0.02)
belastet, e	0.993	0.04	1.6	0.98	131.0	67.5	2.7 (0.22)	0.3 (0.03)
belastet, i	0.992	0.05	1.6	0.96	133.2	65.4	2.8 (0.20)	0.4 (0.04)
belastet, o	0.999	0.09	1.3	0.93	133.5	69.1	1.9 (0.27)	0.5 (0.04)
belastet, u	0.998	0.10	1.4	0.89	134.0	66.3	2.3 (0.31)	0.7 (0.09)
belastet, ä 2	0.997	0.09	1.8	0.98	132.5	70.1	2.6 (0.32)	0.3 (0.02)
	0.993	**0.12**	**2.4**	**0.93**	**132.9**	**67.2**	**3.0 (0.88)**	**0.5 (0.29)**



Drittes Physikalisches Institut und Abt. Phoniatrie und Pädaudiologie Universität Göttingen

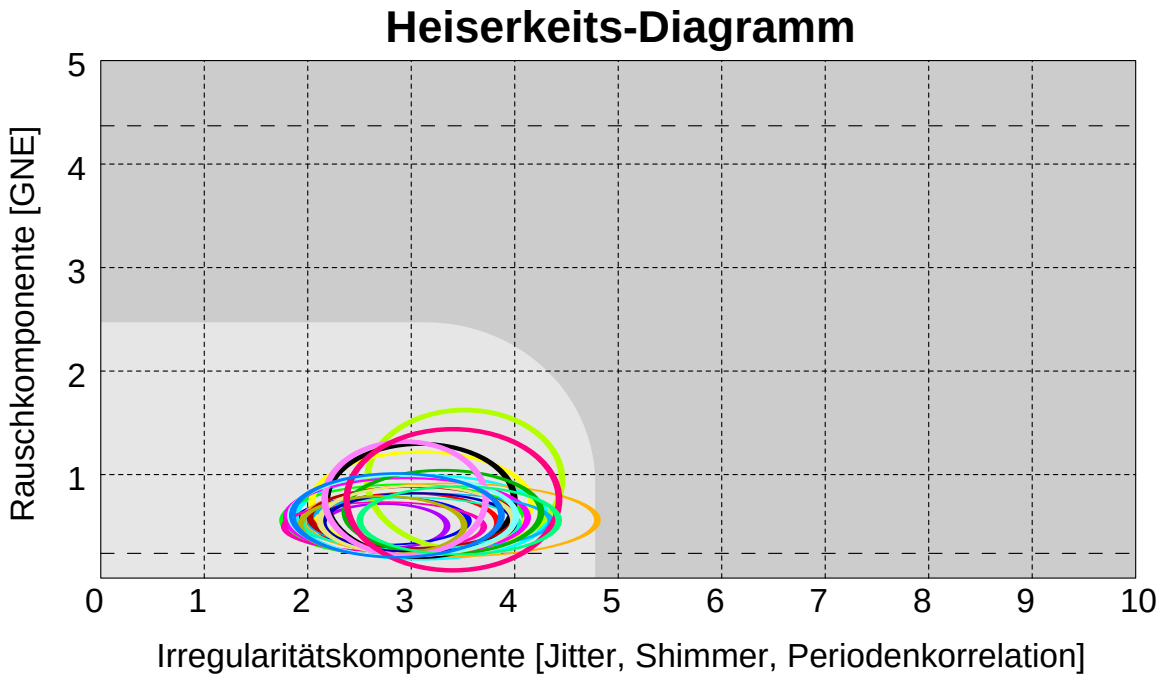
31. Mai 1999

Abbildung 11.1.: Normalstimme 1, Einzelaufnahme

Heiserkeits-Diagramm | Normalstimme

Datum: 08.11.99 Zeit: 20:29
normal

	Irreg.komp.	Rauschkomp.
28.06.96 n=491	3.0 (0.88)	0.5 (0.29)
18.10.96 n=412	2.7 (0.61)	0.5 (0.21)
08.11.96 n=544	2.7 (0.83)	0.6 (0.23)
28.11.96 n=490	3.1 (1.21)	0.6 (0.41)
18.12.96 n=375	2.9 (1.17)	0.6 (0.35)
20.01.97 n=320	3.5 (0.94)	0.9 (0.67)
24.01.97 n=343	3.1 (1.06)	0.7 (0.49)
05.03.97 n=394	3.4 (1.42)	0.6 (0.35)
11.04.97 n=350	3.1 (0.91)	0.8 (0.55)
04.07.97 n=356	3.0 (1.17)	0.6 (0.38)
25.07.97 n=378	2.7 (0.96)	0.5 (0.24)
15.08.97 n=305	3.0 (1.04)	0.5 (0.25)
17.10.97 n=324	3.2 (1.16)	0.5 (0.27)
07.11.97 n=356	3.0 (0.99)	0.6 (0.30)
28.11.97 n=303	3.3 (0.95)	0.6 (0.40)
18.12.97 n=341	3.1 (0.92)	0.5 (0.28)
12.02.98 n=307	2.7 (0.79)	0.5 (0.28)
23.06.98 n=312	3.1 (0.94)	0.6 (0.34)
16.07.98 n=248	2.9 (0.78)	0.8 (0.55)
25.09.98 n=321	2.9 (1.09)	0.6 (0.39)
03.12.98 n=333	3.4 (1.03)	0.8 (0.68)
11.01.99 n=320	3.5 (0.96)	0.6 (0.33)
03.03.99 n=350	2.9 (1.01)	0.6 (0.40)



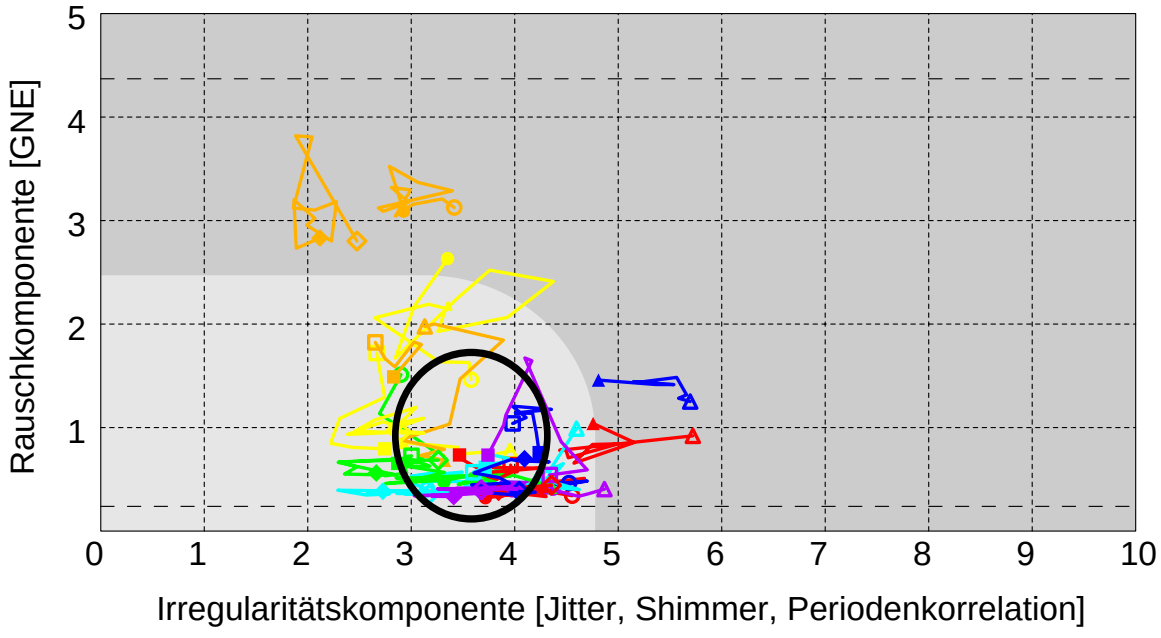
Drittes Physikalisches Institut und Abt. Phoniatrie und Pädaudiologie Universität Göttingen 8. Nov. 1999

Abbildung 11.2.: Normalstimme 1, Verlauf

Heiserkeits-Diagramm | Name: md060966

Datum: do.21.03.96
normale Stimme '0'

Modus	Perioden Korrelat.	Jitter [%]	Shimmer [%]	GNE	F0 [Hz]	SPL [dB]	Irreg.komp.	Rauschkomp.
normal, ä 1	0.991	0.28	3.8	0.97	88.0	134.0	4.1 (0.30)	0.4 (0.05)
normal, a	0.988	0.15	4.5	0.95	89.3	133.0	4.1 (0.27)	0.4 (0.03)
normal, e	0.994	0.12	2.8	0.94	90.9	131.6	3.3 (0.24)	0.5 (0.05)
normal, i	0.995	0.09	2.6	0.90	91.3	129.2	3.1 (0.25)	0.7 (0.29)
normal, o	0.998	0.18	4.4	0.56	90.6	134.9	3.4 (0.41)	2.1 (0.33)
normal, u	0.998	0.15	3.1	0.28	91.8	132.7	3.0 (0.23)	3.2 (0.13)
normal, ä 2	0.993	0.23	3.1	0.96	90.5	133.6	3.7 (0.33)	0.4 (0.04)
tief, ä 1	0.975	0.20	7.7	0.86	68.5	130.6	4.8 (0.39)	0.8 (0.11)
tief, a	0.963	0.33	10.6	0.72	67.7	132.4	5.4 (0.28)	1.4 (0.08)
tief, e	0.991	0.32	5.5	0.92	69.1	131.0	4.4 (0.16)	0.6 (0.17)
tief, i	0.992	0.21	3.1	0.93	70.3	129.6	3.8 (0.23)	0.5 (0.05)
tief, o	0.996	0.23	4.1	0.87	70.7	135.5	3.6 (0.31)	0.8 (0.05)
tief, u	0.998	0.23	3.5	0.75	71.9	133.4	3.3 (0.26)	1.3 (0.52)
tief, ä 2	0.980	0.19	5.1	0.96	72.2	135.4	4.4 (0.26)	0.4 (0.03)
hoch, ä 1	0.993	0.14	4.3	0.91	130.2	136.6	3.8 (0.24)	0.6 (0.07)
hoch, a	0.986	0.17	3.9	0.80	129.9	137.6	4.1 (0.13)	1.1 (0.14)
hoch, e	0.993	0.18	3.5	0.90	133.4	135.7	3.8 (0.10)	0.7 (0.05)
hoch, i	0.996	0.10	2.0	0.90	137.1	133.7	3.0 (0.11)	0.7 (0.04)
hoch, o	0.999	0.10	2.4	0.79	132.2	142.9	2.5 (0.21)	1.1 (0.33)
hoch, u	0.998	0.13	2.6	0.65	135.3	137.9	2.9 (0.15)	1.7 (0.13)
hoch, ä 2	0.990	0.26	4.2	0.81	132.0	139.2	4.2 (0.30)	1.0 (0.41)
belastet, ä 1	0.990	0.30	3.6	0.97	101.4	134.6	4.1 (0.22)	0.4 (0.03)
belastet, a	0.995	0.29	4.2	0.93	103.4	135.4	3.9 (0.21)	0.5 (0.13)
belastet, e	0.996	0.13	1.9	0.97	106.2	134.3	2.9 (0.33)	0.4 (0.02)
belastet, i	0.995	0.07	1.7	0.91	107.2	131.0	2.7 (0.27)	0.6 (0.06)
belastet, o	0.998	0.16	2.1	0.81	105.7	137.3	2.8 (0.22)	1.0 (0.07)
belastet, u	0.999	0.06	1.8	0.30	107.8	135.5	2.1 (0.18)	3.1 (0.35)
belastet, ä 2	0.995	0.16	3.7	0.97	108.2	136.5	3.5 (0.30)	0.4 (0.03)
	0.992	**0.18**	**3.7**	**0.83**	**97.6**	**134.1**	**3.6 (0.73)**	**0.9 (0.80)**



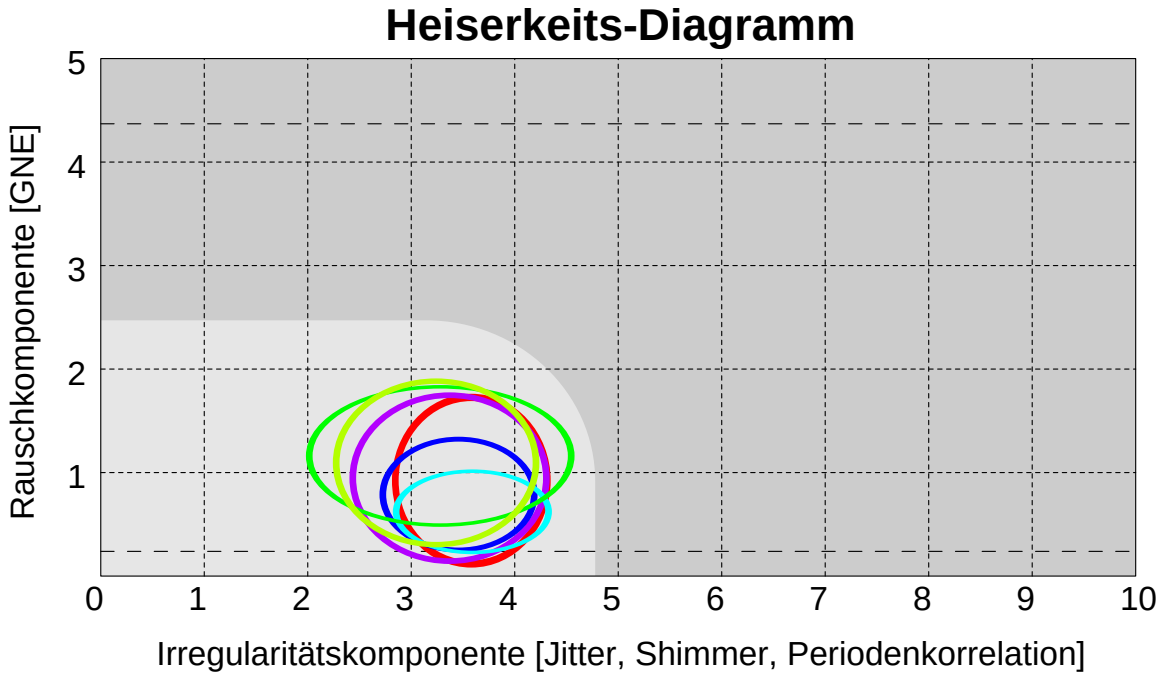
Drittes Physikalisches Institut und Abt. Phoniatrie und Pädaudiologie Universität Göttingen 10. Okt. 1997

Abbildung 11.3.: Normalstimme 2, Einzelaufnahme

Heiserkeits-Diagramm | Normalstimme

Datum: 08.11.99 Zeit: 20:31
normal

	Irreg.komp.	Rauschkomp.
21.03.96 n=315	3.6 (0.73)	0.9 (0.80)
17.04.97 n=341	3.4 (0.94)	0.9 (0.80)
17.10.97 n=438	3.5 (0.73)	0.8 (0.54)
10.06.98 n=244	3.6 (0.74)	0.6 (0.39)
03.03.99 n=548	3.3 (1.27)	1.2 (0.67)
22.10.99 n=641	3.2 (0.97)	1.1 (0.79)



Drittes Physikalisches Institut und Abt. Phoniatrie und Pädaudiologie Universität Göttingen 8. Nov. 1999

Abbildung 11.4.: Normalstimme 2, Verlauf

12. Charakteristisches Muster der Vokale

In den Abbildungen 11.1 und 11.3 des vorigen Kapitels kann man erkennen, dass bei den beiden Normalstimmen die Vokale [u:] und [o:] links oben vom Mittelwert liegen und die Vokale [ɛ:] und [a:] eher rechts vom Mittelwert. Hier soll nun an einer Gruppe von 383 pathologischen und 88 normalen Sprechern die Lage der einzelnen Vokale im Heiserkeits-Diagramm genauer untersucht werden. Aus dem gesamten Stimmkorpus wurden nur solche Aufnahmen zugelassen, die vollständig waren (alle 28 Vokale vorhanden) und bei denen alle Vokale mindesten 0,75s lang waren. Damit standen pro Stimme und Vokal mindestens zwei Messpunkte im Heiserkeits-Diagramm zu Verfügung. Die Werte Irregularität und Rauschkomponente für jeden einzelnen Vokal wurden gemittelt. Diese Mittelwerte dienten als Grundlage zur weiteren Auswertung.

In der Tabelle 12.1 sind die Mittelwerte über alle pathologischen und normalen Sprecher sowie die Standardabweichung dieser Gruppen aufgelistet. Außer den Werten für die einzelnen Vokale sind auch die Gesamtmittelwerte für die verschiedenen Phonationsmodi (normal, hoch, tief, belastet nach Lesen des Textes) eingetragen. Die Tabellendaten dienen als Referenzwerte. Die Ergebnisse sind noch einmal übersichtlicher in den Abbildungen 12.1 für die Normalsprecher und 12.2 für die pathologische Gruppe dargestellt. Hier ist deutlich zu sehen, dass bei allen Moden (normal, hoch, tief, belastet) und bei beiden Gruppen die Vokale auf einer Viertelellipse von links oben nach rechts unten liegen, und zwar in der Reihenfolge [u:], [u:], [i:], [e:], [ɛ:] und [a:].

Der zweidimensionale Kolmogorov-Smirnov-Test belegt, dass alle Unterschiede zwischen den Vokalen bei den Normalstimmen und den pathologischen Stimmen signifikant sind. Eine Ausnahme bildet das Vokalpaar $[\varepsilon:]_1$ und $[\varepsilon:]_2$, die sich nicht signifikant unterscheiden. Die Signifikanzniveaus wurden für alle Tests in diesem Kapitel nach Bonferoni und Holm korrigiert.

Das Muster der Vokale beruht also nicht auf bloßem Zufall sondern ist eine dem Heiserkeits-Diagramm und den Vokalen zugrunde liegende Eigenschaft.

Für die Rauschkomponente kommt der höhere Rauschanteil der Vokale [o:] und [u:] durch zwei Effekte: Die höheren Formanten dieser Vokale sind schwach, so dass die Harmonischen im oberen Frequenzbereich schon so weit abgeschwächt sind, dass in diesen Bereichen das Rauschen dominiert. Deshalb sind die Einhüllenden im tiefen und höheren Frequenzbereich unterschiedlich und der GNE ist entsprechend niedriger. Außerdem verschließen manche Probanden bei diesen Vokalen die Lippen so stark, dass hier tur-

bulentes Rauschen entsteht.

Für die unterschiedlichen Irregularitätskomponenten wurde bisher keine griffige Erklärung gefunden, die mehrfachem Nachfragen standhielte.

Die Mittelwerte der verschiedenen Phonationsmoden (normal, tief, hoch) unterscheiden sich in der Irregularitätskomponente bei der normalen und der pathologischen Gruppe: Der Modus „tief“ zeigt die stärkste, der Modus „hoch“ zeigt die geringste Irregularität, die der Modi „normal“ und „belastet“ liegen dazwischen. Das bedeutet, dass die Irregularität mit zunehmender Grundfrequenz abnimmt. Der t-Test für ungleiche Varianzen der Irregularitätskomponenten zeigt, dass die Unterschiede zwischen Moden mit verschiedenen Tonhöhen fast alle signifikant sind (bis auf den Unterschied hoch belastet bei der pathologischen Gruppe). Für die Rauschkomponente finden sich kaum signifikante Differenzen.

Die Ursache für dieses Verhalten ist nicht leicht zu ergründen. Eine Hypothese beruht auf der Feststellung, dass mit zunehmender Grundfrequenz die Schwingungsweite abnimmt. Genauer gesagt, verändert sich die Schwingung von einer gekoppelten Schwingung, an der der Vocalis-Muskel und die darüber liegende Schleimhautschicht beteiligt ist, hin zu einer reinen Schleimhautschwingung. Die Hypothese ist, dass das gekoppelte System eher zu Irregularitäten neigt als das einfachere System. Außerdem werden die Stimmlippen mit zunehmender Grundfrequenz immer stärker gespannt.

Tabelle 12.1.: Mittelwerte (M) und Standardabweichungen (S) der Irregularitätskomponente und der Rauschkomponente des Heiserkeits-Diagramms einzelner Vokale (Normalgruppe n=88, pathologische Gruppe n=383

	Normalgruppe				pathologische Gruppe			
	Irregk.		Rauschk.		Irregk.		Rauschk.	
	M	S	M	S	M	S	M	S
normal								
[ɛ:]$_1$	3,59	0,56	0,84	0,51	4,65	1,34	1,65	0,95
[a:]	3,77	0,68	0,86	0,56	4,92	1,39	1,77	0,99
[e:]	3,12	0,55	0,77	0,53	4,17	1,38	1,60	0,95
[i:]	2,95	0,61	0,88	0,56	3,88	1,53	1,76	0,98
[o:]	2,61	0,82	1,35	0,80	3,70	1,52	2,11	0,98
[u:]	2,72	1,17	2,19	0,90	3,47	1,55	2,74	0,78
[ɛ:]$_2$	3,58	0,73	0,83	0,58	4,62	1,37	1,66	0,96
hoch								
[ɛ:]$_1$	3,20	0,70	0,97	0,59	4,30	1,54	1,69	0,95
[a:]	3,33	0,72	0,95	0,59	4,56	1,58	1,78	0,99
[e:]	2,97	0,77	0,84	0,50	4,02	1,64	1,65	0,97
[i:]	2,49	0,85	0,95	0,53	3,60	1,67	1,76	0,99
[o:]	2,31	0,76	1,24	0,64	3,54	1,66	2,05	0,97
[u:]	2,28	0,96	1,87	0,89	3,30	1,72	2,57	0,85
[ɛ:]$_2$	3,06	0,74	0,93	0,61	4,37	1,65	1,70	0,98
tief								
[ɛ:]$_1$	3,86	0,82	0,84	0,51	4,93	1,46	1,77	0,99
[a:]	4,00	0,77	0,90	0,54	5,16	1,49	1,83	0,98
[e:]	3,36	0,78	0,84	0,50	4,53	1,48	1,68	0,98
[i:]	3,24	0,73	0,95	0,56	4,22	1,51	1,79	0,99
[o:]	3,09	0,97	1,43	0,74	4,16	1,63	2,20	0,98
[u:]	3,04	1,05	2,14	0,91	3,80	1,54	2,73	0,87
[ɛ:]$_2$	3,85	0,84	0,90	0,58	4,94	1,49	1,77	1,00
belastet								
[ɛ:]$_1$	3,43	0,58	0,78	0,48	4,46	1,39	1,56	1,00
[a:]	3,49	0,64	0,76	0,45	4,66	1,46	1,63	1,01
[e:]	3,05	0,63	0,69	0,37	4,05	1,48	1,53	0,98
[i:]	2,88	0,68	0,86	0,54	3,74	1,46	1,61	0,99
[o:]	2,50	0,75	1,09	0,57	3,59	1,64	2,01	1,00
[u:]	2,52	0,87	1,89	0,84	3,38	1,58	2,61	0,84
[ɛ:]$_2$	3,41	0,71	0,76	0,46	4,56	1,51	1,58	1,01
alle Modi								
[ɛ:]$_1$	3,59	0,56	0,84	0,51	4,65	1,34	1,65	0,95
[a:]	3,77	0,68	0,86	0,56	4,92	1,39	1,77	0,99
[e:]	3,12	0,55	0,77	0,53	4,17	1,38	1,60	0,95
[i:]	2,95	0,61	0,88	0,56	3,88	1,53	1,76	0,98
[o:]	2,61	0,82	1,35	0,80	3,70	1,52	2,11	0,98
[u:]	2,72	1,17	2,19	0,90	3,47	1,55	2,74	0,78
[ɛ:]$_2$	3,58	0,73	0,83	0,58	4,62	1,37	1,66	0,96
normal	3,19	0,87	1,11	0,81	4,20	1,53	1,90	1,02
hoch	2,80	0,89	1,11	0,71	3,96	1,70	1,88	1,01
tief	3,49	0,94	1,14	0,78	4,53	1,58	1,97	1,03
belastet	3,04	0,80	0,98	0,67	4,06	1,58	1,79	1,04

Tabelle 12.2.: Verschiedene Vokale normal. *: nicht signifikant.

Zweidimensionaler Kolmogorov-Smirnov-Test						
	[a:]	[e:]	[i:]	[o:]	[u:]	$[\varepsilon:]_2$
$[\varepsilon:]_1$	0,18	0,48	0,64	1,11	1,61	0,02*
[a:]		0,66	0,82	1,27	1,70	0,20
[e:]			0,20	0,77	1,47	0,46
[i:]				0,58	1,33	0,63
[o:]					0,85	1,10
[u:]						1,61

t-Test für ungleiche Varianzen der Irregularitätskomponenten						
	[a:]	[e:]	[i:]	[o:]	[u:]	$[\varepsilon:]_2$
$[\varepsilon:]_1$	−0,18	0,47	0,64	0,98	0,87	0,01*
[a:]		0,66	0,82	1,17	1,06	0,20*
[e:]			0,16*	0,51	0,40	−0,46
[i:]				0,34	0,24*	−0,62
[o:]					−0,11*	−0,97
[u:]						−0,86

t-Test für ungleiche Varianzen der Rauschkomponenten						
	[a:]	[e:]	[i:]	[o:]	[u:]	$[\varepsilon:]_2$
$[\varepsilon:]_1$	−0,01*	0,07*	−0,04*	−0,51	−1,35	0,02*
[a:]		0,09*	−0,02*	−0,49	−1,33	0,03*
[e:]			−0,11*	−0,58	−1,42	−0,05*
[i:]				−0,47	−1,31	0,06*
[o:]					−0,84	0,52
[u:]						1,36

Tabelle 12.3.: Verschiedene Vokale pathologisch. *: nicht signifikant.

Zweidimensionaler Kolmogorov-Smirnov-Test						
	[a:]	[e:]	[i:]	[o:]	[u:]	$[\varepsilon:]_2$
$[\varepsilon:]_1$	0,29	0,49	0,79	1,06	1,61	0,04*
[a:]		0,77	1,04	1,26	1,74	0,32
[e:]			0,33	0,69	1,33	0,45
[i:]				0,40	1,07	0,75
[o:]					0,67	1,02
[u:]						1,58
t-Test für ungleiche Varianzen der Irregularitätskomponenten						
	[a:]	[e:]	[i:]	[o:]	[u:]	$[\varepsilon:]_2$
$[\varepsilon:]_1$	−0,26	0,48	0,78	0,95	1,18	0,04*
[a:]		0,75	1,04	1,22	1,45	0,30
[e:]			0,29	0,47	0,70	−0,45
[i:]				0,17*	0,40	−0,74
[o:]					0,23	−0,92
[u:]						−1,15
t-Test für ungleiche Varianzen der Rauschkomponenten						
	[a:]	[e:]	[i:]	[o:]	[u:]	$[\varepsilon:]_2$
$[\varepsilon:]_1$	−0,12*	0,04*	−0,11*	−0,47	−1,10	−0,01*
[a:]		0,16	0,01*	−0,35	−0,97	0,11*
[e:]			−0,15	−0,51	−1,14	−0,05*
[i:]				−0,36	−0,99	0,10*
[o:]					−0,63	0,46
[u:]						1,08

Tabelle 12.4.: Verschiedene Modi. *: nicht signifikant.

Zweidimensionaler Kolmogorov-Smirnov-Test	Normalgruppe			pathologische Gruppe		
	hoch	tief	belastet	hoch	tief	belastet
normal	0,39	0,30*	0,20*	0,25	0,34	0,18
hoch		0,69	0,27		0,58	0,14
tief			0,48			0,50

t-Test für ungleiche Varianzen der Irregularitätskomponenten	Normalgruppe			pathologische Gruppe		
	hoch	tief	belastet	hoch	tief	belastet
normal	0,39	−0,30	0,15*	0,25	−0,33	0,14*
hoch		−0,69	−0,24		−0,58	−0,11*
tief			0,45			0,47

t-Test für ungleiche Varianzen der Rauschkomponenten	Normalgruppe			pathologische Gruppe		
	hoch	tief	belastet	hoch	tief	belastet
normal	−0,00*	−0,04*	0,13*	0,01*	−0,07*	0,11
hoch		−0,03*	0,13*		−0,08*	0,10*
tief			0,17			0,18

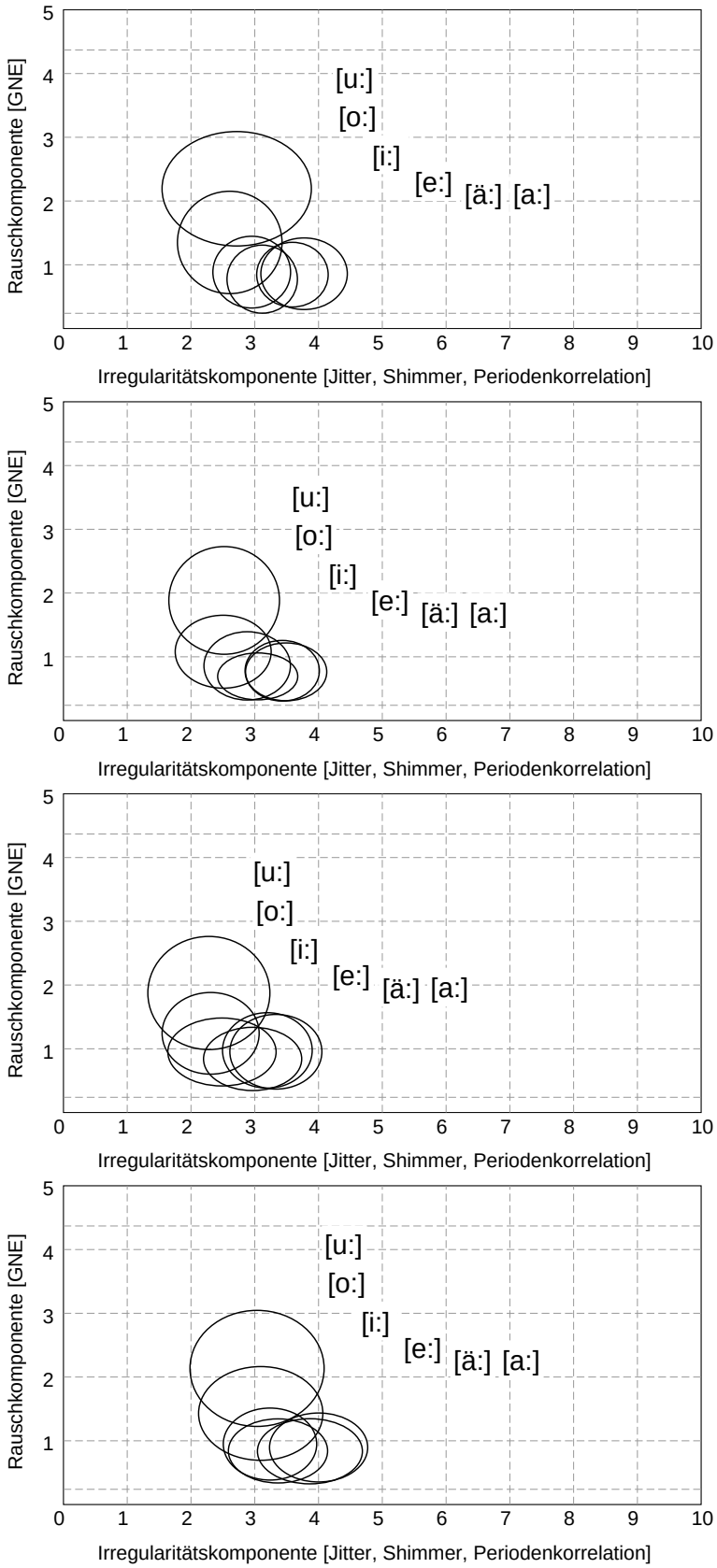


Abbildung 12.1.: Die Lage der verschiedenen Vokale der pathologischen Gruppe im Heiserkeits-Diagramm. Von oben nach unten: normal 1, normal 2, hoch, tief.

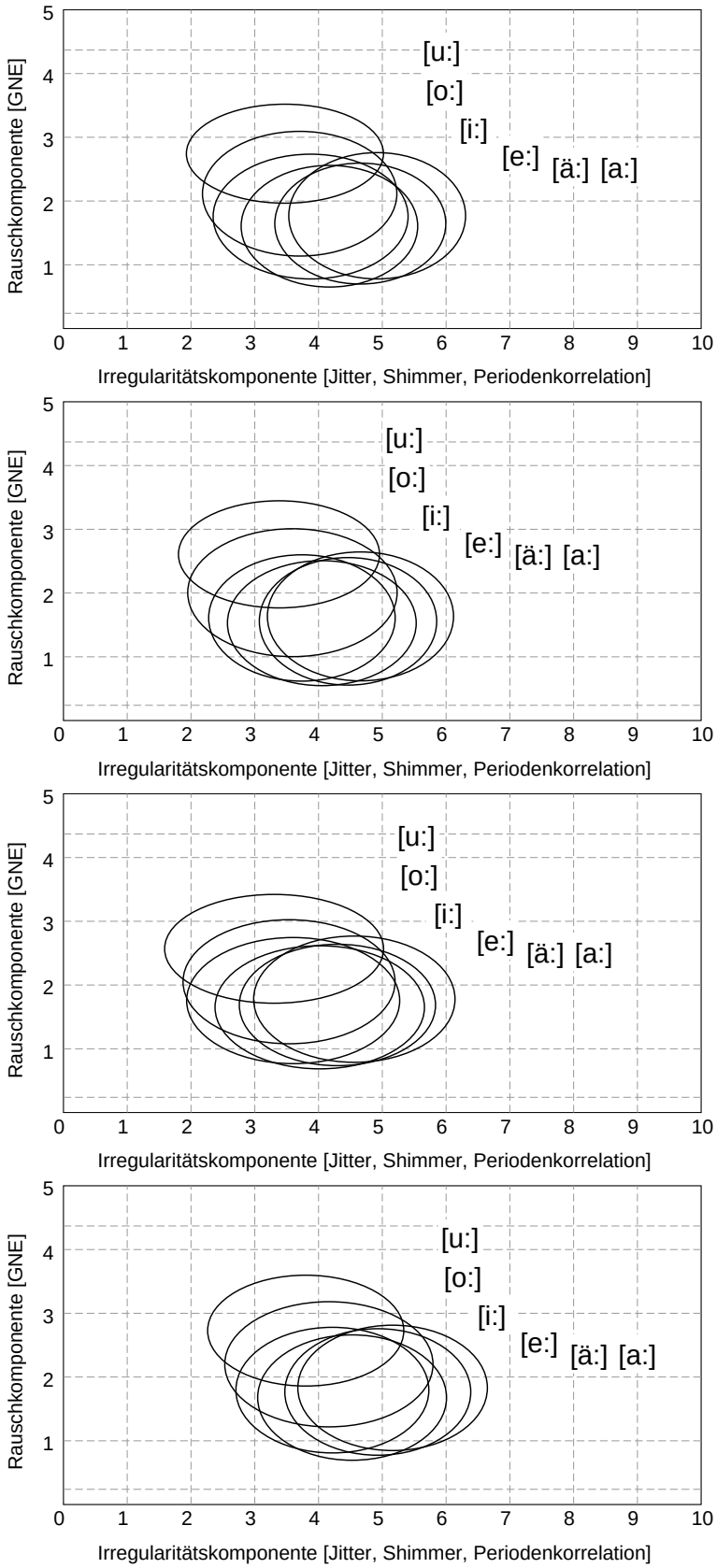


Abbildung 12.2.: Die Lage der verschiedenen Vokale der Normalstimmen im Heiserkeits-Diagramm. Von oben nach unten: normal 1, normal 2, hoch, tief.

12.1. Klassifikation der Stimmstörungen

Zur Validierung der Ergebnisse aus den akustischen Messungen wird zusätzliches Datenmaterial benötigt. Die wertvollste Hilfe, um perzeptiv wahrgenommene Stimmstörungen oder Missempfindungen des Patenten genauer zu klassifizieren, ist die Laryngoskopie und die Laryngostroboskopie des Kehlkopfes und der (schwingenden) Stimmlippen. Die Laryngoskopie bzw. Laryngostroboskopie erfolgt mit einer starren Optik durch den Mundraum oder mit einer flexiblen Optik (Endoskopie) durch den Nasenraum. Durch die Laryngostroboskopie kann bei hinreichend stationärer Schwingung das Schwingungsverhalten der Stimmbänder ohne aufwendige Hochgeschwindigkeitskameras in Echtzeit beobachtet werden. Mit Hilfe der Laryngostroboskopie können die Stimmstörungen in Klassen zusammengefasst werden. Die Stimmstörungen werden in die folgenden sechs Klassen eingeteilt:

12.1.1. Bösartige Tumore

Die schwerste und potentiell folgenreichste Störung des Stimmapparates ist die Bildung eines bösartigen Tumors an den Stimmlippen oder den umgebenden Strukturen. Die Ursache für bösartige Stimmlippentumore ist fast ausnahmslos das Rauchen. Je nach Größe und Eindringtiefe des Tumors in die Stimmlippe(n) werden verschiedene Tumortypen unterschieden. Bei den großen und tief reichenden Tumoren besteht durchaus Lebensgefahr, da sich der Tumor durch Metastasenbildung über den Blutkreislauf auf den gesamten Körper ausdehnen kann. Der Verlust der Stimme ist dann die zweitschlimmste Konsequenz eines bösartigen Stimmlippentumors. Bis vor einigen Jahren (und an manchen Kliniken auch bis heute) wurde wegen der vitalen Gefahr eines bösartigen Tumors dieser weiträumig entfernt, um erstens die Gefahr zu minimieren, dass der Tumor angeschnitten wird, und um zweitens sicherzustellen, dass alle Tumorzellen entfernt werden. Dabei wurde mindestens gleich die komplette befallene Stimmlippe entfernt, wenn nicht sogar der komplette Kehlkopf. Dies führte häufig zum Verlust der Arbeitstelle und zu einer starken Einschränkung der sprachlichen Kommunikation. Es gibt zwar Hilfsmittel zur externen Erzeugung eines Stimmanregungssignals, doch diese Stimme hat viel von ihrer „menschlichen“ Form eingebüßt.

Wegen dieser weitreichenden Konsequenzen wird seit einigen Jahren an der Göttinger Universitätsklinik eine erfolgversprechende Alternative zu diesen Totalresektionen angewandt und erforscht: Die sogenannte minimalinvasive Laserchirurgie. Hierbei werden zunächst das Ausmaß der Resektion und vor allem auch die zu erwartende postoperative Struktur durch Phoniater und Chirurg geplant, um schon bei der Operation das spätere Schwingungsverhalten des Restgewebes möglichst günstig zu gestalten. Außerdem wird nur soviel Gewebe entfernt, wie nötig ist, um den Tumor unter histologischer Sicherung komplett zu entfernen. Demgegenüber wird gesundes Restgewebe weitgehend erhalten (Vermeidung eines overtreatment). Dadurch kann oft die stimmgebende Funktion des Kehlkopfes erhalten werden. Die Gruppe dieser Patienten wird im Folgenden noch eine besondere Rolle spielen, wenn es um die praktische Anwendung der akustischen Methoden geht.

12.1.2. Lähmungen

Die zweite Gruppe der Stimmstörungen umfasst alle Beeinflussungen der Beweglichkeit einer oder beider Stimmlippen. Dazu gehören zunächst alle Arten der Lähmungen, d.h. Funktionsstörungen der laryngealen Muskulatur durch Beeinträchtigung oder sogar Ausfall der Funktion der entsprechenden Nerven. Ein Ausfall der Nervenfunktion kann dabei von einem Unfall oder einer Durchtrennung des Nerven bei der operativen Entfernung eines Kropfes (Strumektomie) herrühren. Auch bei intakten Nerven kann die Mobilität der Stimmlippen gestört sein. Hier ist häufig eine dauerhafte mechanische Belastung der Stimmbänder z.B. durch den Druck eines Intubationstubus oder eine kurzzeitige hohe Belastung (Traumatisierung) durch Unfälle die Ursache.

12.1.3. Gutartige Neubildungen

In der dritten Gruppe sollen alle Veränderungen der Stimmlippen durch gutartige Neubildungen zusammengefasst werden. Hierunter fallen Polypen, Zysten, Reinke-Ödeme, Papillome, Knötchen und Kontaktgranulome. Allen Neubildungen ist gemeinsam, dass sie nicht lebensgefährlich sind und dass der Patient gute Chancen hat, je nach Größe der Neubildung entweder durch eine logopädische Therapie oder durch eine Operation langfristig von seinen Leiden geheilt zu werden.

12.1.4. Funktionelle Stimmstörungen

Eine vielumstrittene, aber volkswirtschaftlich sehr relevante Art der Stimmstörungen bildet die vierte Gruppe: die funktionellen Stimmstörungen. Hier ist erst einmal keine eindeutige primärorganische Ursache der Störung feststellbar. Die Symptome der Störung reichen von Missempfinden im Kehlkopfbereich bis zu extremer Heiserkeit und Aphonie (die Stimmlippen schwingen oder schließen nicht mehr). Häufig sind hier Menschen aus Sprechberufen mit zusätzlichen Stressfaktoren betroffen (Lehrerinnen, Kindergärtner). Diese Stimmstörung kann die Arbeitsfähigkeit in den genannten und ähnlichen Berufen stark einschränken oder sogar zur Arbeitsunfähigkeit führen. In der Literatur werden am häufigsten die hyperfunktionelle Dysphonie und die hypofunktionelle Dysphonie erwähnt. Die eindeutige Abgrenzung dieser Störungsbilder ist schwierig und uneinheitlich.

12.1.5. Zentrale Stimmstörungen

Die fünfte Gruppe wird von den zentralen Stimmstörungen gebildet, also Störungen, die auf einer Fehlfunktion des zentralen Nervensystems beruhen. Hier ist vor allem die spasmodische Dysphonie zu nennen. Hierbei wird die Lautbildung durch spastische Kontraktion der Kehlkopfmuskulatur regelrecht abgewürgt. Das Hauptmerkmal dieser abgewürgten Stimmlaute sind Stimmunterbrechungen. Da die folgenden Algorithmen zur Stimmgütebeschreibung auf gehaltener Phonation beruhen, sind sie ausdrücklich nicht

geeignet um die spezifische Störung bei der spasmodischen Dysphonie richtig zu beschreiben.

12.1.6. Verschiedene

In der sechsten Gruppe sollen alle anderen Störungen zusammengefasst werden, wie Probleme beim Stimmbruch, Stimmstörungen bei Syndromen (Krankheiten, die eine Vielzahl von Störungen und/oder Missbildungen umfassen), Kehlkopfentzündungen usw.

13. Pathologische Gruppen im Heiserkeits-Diagramm

In diesem Kapitel werden Gruppen von Patienten untersucht, die jeweils pathologische Veränderungen an den Stimmlippen zeigen. Bei allen Patienten wurde eine laryngoskopische und wenn möglich auch eine videostroboskopische Untersuchung im Klinikum Göttingen durchgeführt. Darüber hinaus wurden die Stimmen aller Patienten nach dem oben beschriebenen Protokoll aufgenommen und analysiert.

Obwohl bereits 1988 und 1995 Scherer und andere [114] zeigten, dass für eine zuverlässige Analyse der Stimmgüte 7-15 Stimmproben pro Patient und Aufnahmedatum analysiert werden sollten, ist nach meinem Wissen noch keine Arbeit erschienen, in der diese Forderung erfüllt worden wäre.

Zahlreiche Arbeiten zum Thema der Unterscheidbarkeit von Normalstimmen und pathologischen Stimmen mit akustischen Maßen sind bereits erschienen, unter anderem: [11, 14, 43, 44, 52, 54, 69, 70, 75, 102, 103, 105, 117, 152, 153]. In diesen Arbeiten wurde im Unterschied zu folgenden Untersuchungen nicht versucht, mehrere Gruppen voneinander zu trennen, sondern jeweils nur die Trennung zwischen normalen und pathologischen Stimmen oder die Unterscheidung vor und nach Behandlung der Patienten mit einer bestimmten Methode. In den genannten Arbeiten wurde außerdem stets Jitter mit ereignisbasierten Methoden bestimmt und Shimmer als Amplitudenshimmer quantifiziert. Damit wurde die bereits beschriebene Rauschanfälligkeit dieser Methoden und deren geringere Messgenauigkeit im Vergleich zum Waveform-Matching-Verfahren in Kauf genommen.

Durch das umfangreiche Aufnahmeprotokoll kann hier davon ausgegangen werden, dass sich die charakteristischen Eigenschaften der Stimme des Patienten tatsächlich in den Aufnahmen wiederfinden. Außerdem gewinnt die Analyse des einzelnen Patienten durch die Mittelung über viele analysierte Segmente an statistischer Sicherheit.

Insofern sind die folgenden Analyseergebnisse neuartig, auch wenn schon in anderen Studien pathologische Gruppen verglichen wurden.

Neuartig ist auch, dass keine Stimmen von der Analyse ausgeschlossen werden mussten, so dass im Folgenden erstmalig auch ein quantitativer Vergleich gegenüber aphonen Stimmen möglich ist. Dies ist ein wichtiger Aspekt für den Phoniater, denn auch wenn eine Stimme sehr behaucht und unregelmäßig ist bzw. eine starke Rausch- und Irregularitätskomponente im Heiserkeits-Diagramm zeigt, so ist die Kommunikationstauglichkeit doch höher als bei der aphonen Stimme. Dass eine akustische Differenzierung zwischen

sehr starken Stimmstörungen und Aphonie möglich ist, wird unter anderem im Folgenden gezeigt.

13.1. Normalstimmen, Aphonie und gutartige Neubildungen

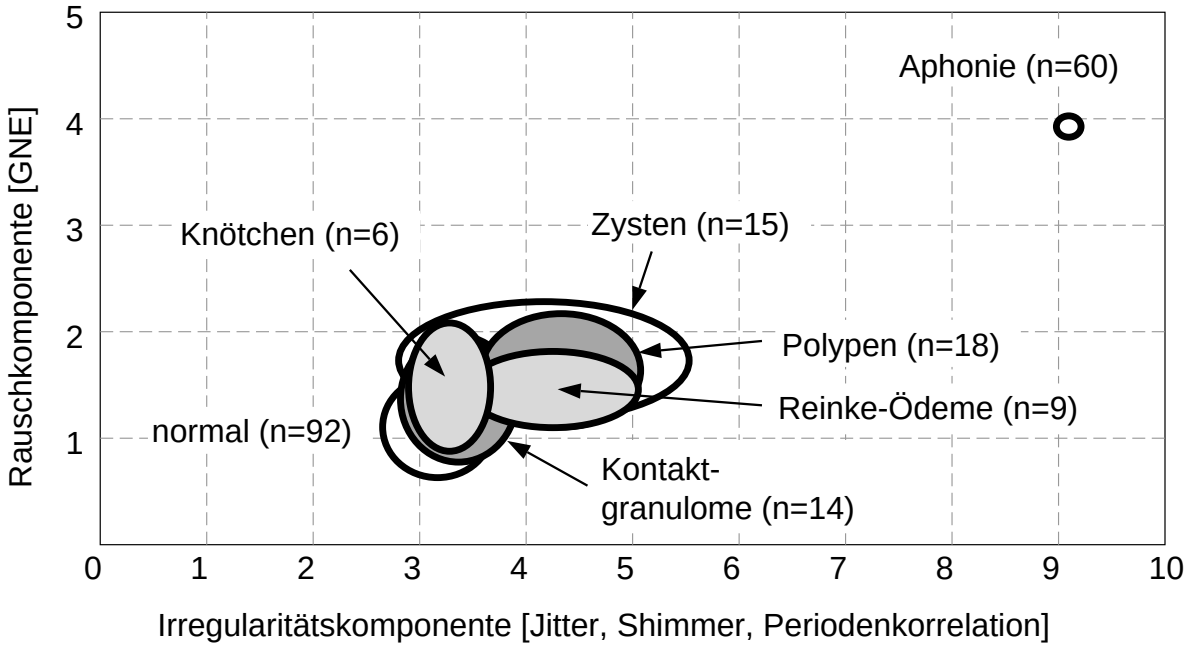


Abbildung 13.1.: Gruppen mit verschiedenen gutartigen Neubildungen auf den Stimmlippen im Heiserkeits-Diagramm: Sprecher mit normaler Stimme, mit Kontaktgranulomen, Knötchen, Polypen, Zysten, Reinke-Ödemen und aphone Sprecher (simulierte Aphonie durch Flüstern)

Zuerst wird die Lage der Stimmaufnahmen von 92 Normalstimmen und 60 aphonen Stimmen (durch Flüstern simulierte Aphonie) im Heiserkeits-Diagramm betrachtet: Die Mittelwerte (Tabelle 13.1) der Normalgruppe sind 3,2 für die Irregularitätskomponente und 1,1 für die Rauschkomponente. Die aphone Gruppe hat eine mittlere Irregularitätskomponente von 9,1 und eine mittlere Rauschkomponente von 3,9. Die beiden Gruppen sind in Abbildung 13.1 als Ellipsen dargestellt. Das Zentrum der Ellipse entspricht den Mittelwerten der Irregularitäts- und Rauschkomponente der Gruppen. Die Länge der Halbachsen der Ellipsen entspricht einer Standardabweichung der jeweiligen Gruppe. Die Normalstimmen liegen links unten, die aphonen Stimmen rechts oben im Diagramm. Die Streuung der aphonen Gruppe ist relativ gering. Diese beiden Gruppen dienen im Folgenden jeweils als Referenzgruppen zum Vergleich mit Gruppen spezieller Dysphonien.

Tabelle 13.1.: Normalstimmen, aphone Stimmen und gutartige Neubildungen auf den Stimmlippen als Gruppen im Heiserkeits-Diagramm

		Irregularitätskomponente		Rauschkomponente	
	n	Mittelwert	Standabw.	Mittelwert	Standabw.
normal	92	3,1678	0,5157	1,1044	0,4748
Kontaktgranulom	14	3,3648	0,5449	1,3679	0,5924
Knötchen	6	3,2808	0,3854	1,4798	0,6024
Polyp	18	4,3293	0,7460	1,6371	0,5326
Zyste	15	4,1664	1,3643	1,7293	0,5531
Reinke-Ödeme	9	4,2515	0,8025	1,4571	0,3576
Aphonie	60	9,0958	0,1158	3,9266	0,1003

Der erste Vergleich erfolgt mit Gruppen von Patienten mit gutartigen Neubildungen [16]. In Tabelle 13.1 und der Abbildung 13.1 sind die Mittelwerte und Standardabweichungen der Gruppen Knötchen, Kontaktgranulome, Reinke-Ödeme, Polypen und Zysten dargestellt. Die gutartigen Neubildungen behindern die Stimmgebung nicht sehr stark solange die Neubildung einen relativ kleinen Umfang besitzt. In wenigen Fällen werden aber auch große Neubildungen beobachtet, die die Stimme entsprechend stark beeinflussen. Die geringfügigste Neubildung sind Knötchen (nur bei Frauen) und Kontaktgranulome (nur bei Männern). Diese beiden Arten der Neubildungen umfassen nur das oberste Epithel der Stimmlippen.

Die Neubildungen Polyp, Zyste und Reinke-Ödeme reichen tiefer in die Stimmlippe hinein oder erfassen die gesamte Stimmlippe. Deshalb ist für diese Neubildungen eine größere Beeinflussung der Stimmqualität zu erwarten.

Die Lage der Gruppen im Heiserkeits-Diagramm entspricht der Schwere der Veränderung der Stimmlippen: Die Knötchen und die Kontaktgranulome sind der Normalgruppe dicht benachbart. Die Gruppen Zysten, Polypen und Reinke-Ödeme liegen weiter rechts bzw. weiter oben. Alle gutartigen Neubildungen liegen relativ weit von der aphonen Gruppe entfernt. Diese Anordnung der Gruppen ist ein Hinweis darauf, dass die Topologie des Stimmgüteraumes, wie er in dem Heiserkeits-Diagramm dargestellt wird, eine Entsprechung zu dem Grad der Stimmveränderung hat: Je schwerer die Veränderung, desto weiter oben und links liegen die Gruppen.

Tabelle 13.2.: Statistische Tests für gutartige Neubildungen. †: Unterschied nicht signifikant.

Zweidimensionaler Kolmolgorov-Smirnov Test

	Kontaktg.	Knötchen	Polyp	Zyste	Reinke-Ö.	Aphonie
normal	$-0{,}26^{\dagger}$	$-0{,}38^{\dagger}$	$-0{,}53$	$-0{,}62$	$-0{,}35^{\dagger}$	$-2{,}82$
Kontaktgranulom		$-0{,}11^{\dagger}$	$-0{,}27^{\dagger}$	$-0{,}36^{\dagger}$	$-0{,}09^{\dagger}$	$-2{,}56$
Knötchen			$-0{,}16^{\dagger}$	$-0{,}25^{\dagger}$	$0{,}02^{\dagger}$	$-2{,}45$
Polyp				$-0{,}09^{\dagger}$	$0{,}18^{\dagger}$	$-2{,}29$
Zyste					$0{,}27^{\dagger}$	$-2{,}20$
Reinke-Ödeme						$-2{,}47$

t-Test für ungleiche Varianzen der Rauschkomponente

	Kontaktg.	Knötchen	Polyp	Zyste	Reinke-Ö.	Aphonie
normal	$0{,}33^{\dagger}$	$0{,}39^{\dagger}$	$1{,}28$	$1{,}18^{\dagger}$	$1{,}14^{\dagger}$	$6{,}57$
Kontaktgranulom		$0{,}14^{\dagger}$	$1{,}00^{\dagger}$	$0{,}88^{\dagger}$	$0{,}89^{\dagger}$	$6{,}28$
Knötchen			$1{,}06^{\dagger}$	$0{,}92^{\dagger}$	$0{,}97^{\dagger}$	$6{,}31$
Polyp				$0{,}19^{\dagger}$	$0{,}20^{\dagger}$	$5{,}29$
Zyste					$0{,}29^{\dagger}$	$5{,}40$
Reinke-Ödeme						$5{,}44$

t-Test für ungleiche Varianzen der Irregularitätskomponente

	Kontaktg.	Knötchen	Polyp	Zyste	Reinke-Ö.	Aphonie
normal	$-0{,}20^{\dagger}$	$-0{,}11^{\dagger}$	$-1{,}16$	$-1{,}00^{\dagger}$	$-1{,}08$	$-5{,}93$
Kontaktgranulom		$0{,}08^{\dagger}$	$-0{,}96$	$-0{,}80^{\dagger}$	$-0{,}89^{\dagger}$	$-5{,}73$
Knötchen			$-1{,}05^{\dagger}$	$-0{,}89^{\dagger}$	$-0{,}97^{\dagger}$	$-5{,}81$
Polyp				$0{,}16^{\dagger}$	$0{,}08^{\dagger}$	$-4{,}77$
Zyste					$-0{,}09^{\dagger}$	$-4{,}93$
Reinke-Ödeme						$-4{,}84$

Die statistische Auswertung ergibt, dass sich die Gruppen der Zysten und der Polypen signifikant (mindestens einer der drei Tests in Tabelle 13.2 ist signifikant) von den Normalstimmen unterscheiden. Die Gruppe der Kontaktgranulome unterscheidet sich signifikant von den Polypen. Alle gutartigen Gruppen zeigen in beiden Komponenten signifikant kleinere Werte als die aphone Gruppe.

13.2. Stimmlippenpolypen: prä- und post-operativ

Aus der Gruppe der Patienten mit Polypen wurden die neun Patienten ausgewählt, von denen Stimmaufnahmen prä- und postoperativ (nach Wundheilung) vorhanden waren. Als Vergleich wurde auch hier die normale und die aphone Gruppe herangezogen. In

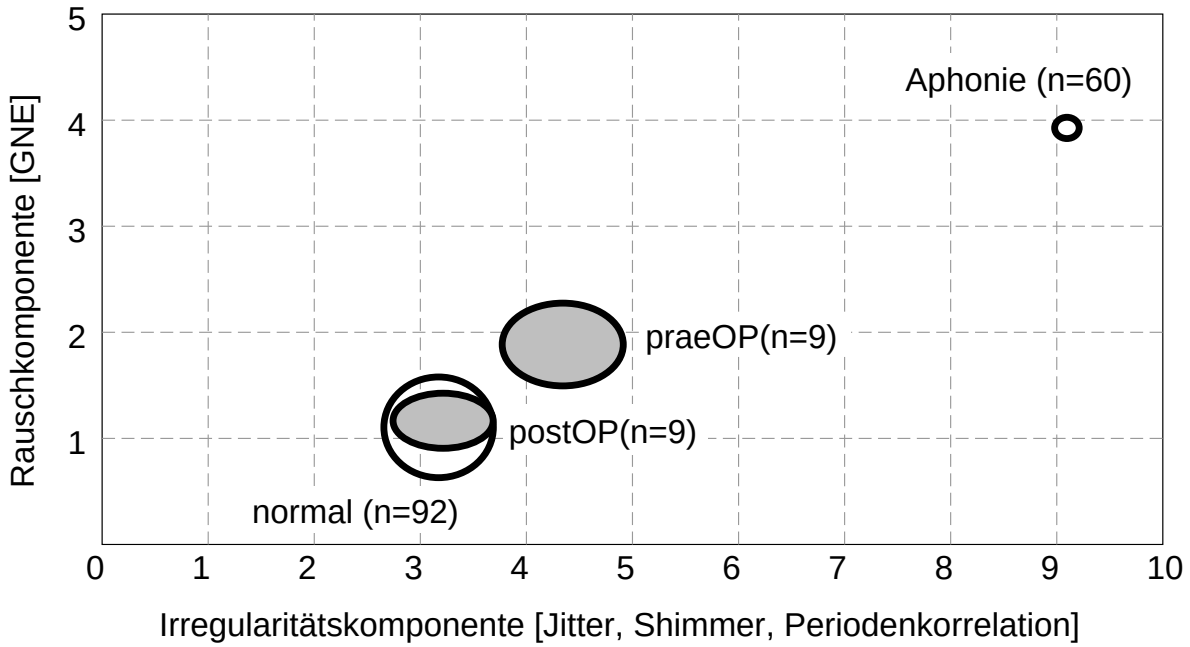


Abbildung 13.2.: Sprecher mit Polyp auf der Stimmlippe: Sprecher mit normaler Phonation, Sprecher mit Polyp präoperativ, Sprecher mit Polyp postoperativ und aphone Sprecher (simulierte Aphonie durch Flüstern)

Abbildung 13.2 und Tabelle 13.3 sind die Mittelwerte und Standardabweichungen zu sehen. Die Mittelwerte der Irregularitätskomponente und der Rauschkomponente bei der präoperativen Gruppe liegen etwas über denen der Polypengruppe im vorherigen Abschnitt. Die Gruppe der postoperativen überlappt stark mit den Normalstimmen. Die statistische Auswertung (Tabelle 13.4) bestätigt den Eindruck der Abbildung 13.2: Die präoperative Gruppe zeigt gegenüber der Normalgruppe signifikant erhöhte Irregularitäts- und Rauschwerte (in allen drei Tests). Die postoperative Gruppe lässt sich statistisch nicht mehr von der Normalgruppe trennen. Prä- und postoperative Gruppe liegen signifikant unter den Werten der aphonen Gruppe. Für leichtgradige Stimmstörungen

Tabelle 13.3.: Gruppe Polypen, prä und post OP im Heiserkeits-Diagramm

		Irregulatitätskomponente		Rauschkomponente	
	n	Mittelwert	Standardabw.	Mittelwert	Standardabw.
normal	92	3,17	0,52	1,10	0,47
Polyp prä	9	4,34	0,57	1,88	0,39
Polyp post	9	3,22	0,47	1,17	0,26
Aphonie	60	9,10	0,12	3,93	0,10

ist somit teilweise eine Differenzierung gegenüber den Normalstimmen möglich. Präpostoperative Veränderungen (Verbesserungen) konnten für die Polypengruppe gezeigt werden.

Tabelle 13.4.: Statistische Tests für Polypen, prä und post OP. †: Unterschied nicht signifikant.

Zweidimensionaler Kolmolgorov-Smirnov Test			
	Polyp prä	Polyp post	Aphonie
normal	1,41	0,07†	6,56
Polyp prä		1,34	5,17
Polyp post			6,50

t-Test für ungleiche Varianzen der Rauschkomponente			
	Polyp prä	Polyp post	Aphonie
normal	−0,78	−0,06†	−2,82
Polyp prä		0,72	−2,04
Polyp post			−2,76

t-Test für ungleiche Varianzen der Irregularitätskomponente			
	Polyp prä	Polyp post	Aphonie
normal	−1,17	−0,04†	−5,92
Polyp prä		1,13	−4,75
Polyp post			−5,88

13.3. Gruppen mit Lähmungen

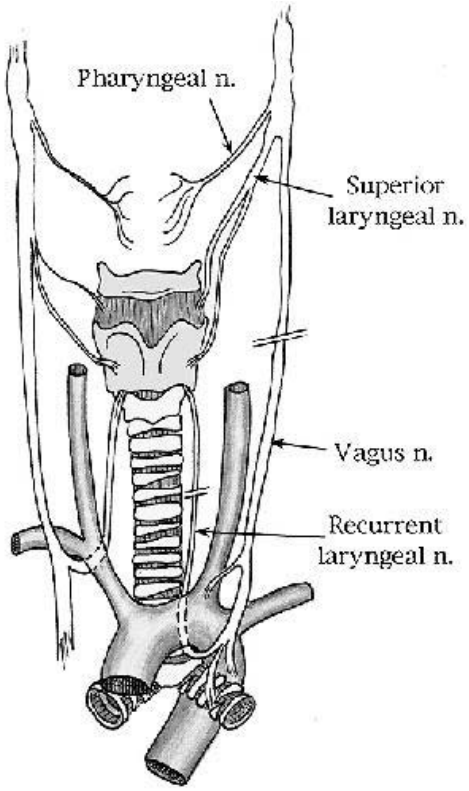


Abbildung 13.3.: Darstellung der Kehlkopfnerven (mit freundlicher Genehmigung von C. Mascher 1998).

In diesem Abschnitt werden Gruppen von Patienten analysiert, bei denen die Beweglichkeit der Stimmlippen oder anderer Kehlkopfmuskeln beeinträchtigt ist. In Abbildung 13.3 sind die Kehlkopfnerven dargestellt. Der wichtigste Nerv ist der Vagusnerv, der als Teil des vegetativen Nervensystems unter anderem zur Steuerung von Kehlkopf und Atmung zuständig ist. Von diesem zweigt der „Zurückläufer“, der Recurrens ab, der wiederum einen Teilbereich des Kehlkopfes, darunter den Stimmlippenmuskel, innerviert. Eine Verletzung des Vagus oder des Recurrens durch Unfall oder z.B. bei operativer Kropfentfernung führt zu einer Beeinträchtigung der Stimme. Fällt der Recurrens aus, so führt dies letztlich zur Erschlaffung der entsprechenden Stimmlippe. Die Regulierbarkeit der Stimmlippenspannung ist nicht mehr gegeben. Dadurch kommt keine beidseitige, synchrone Schwingung beider Stimmlippen mehr zustande. Dies führt zu einer behauchten Phonation mit schlechtem Stimmlippenschluss. Ist der Vagus geschädigt, so ist der Mechanismus zum Wechsel zwischen Phonationsstellung (Stimmlippen geschlossen) und Atemstellung (Stimmlippen sind V-förmig geöffnet) beeinträchtigt, auf der gelähmten Seite verharrt die Stimmlippe in einer bestimmten Position. Die genaue Position kann dabei mehr der Phonations- oder mehr der Atemstellung entsprechen, Typisch ist je-

doch eine Stellung in Richtung der Atemstellung. Bei der Phonation müsste die gesunde Stimmlippe nun weit über die Medianstellung hinaus auf die gelähmte Stimmlippe zu bewegt werden, um einen Stimmlippenschluss zu erzielen. Je nachdem, wie weit die gelähmte Stimmlippe lateral liegt, gelingt dies besser oder schlechter. Auf jeden Fall ist der Stimmlippenschluss gegenüber der Normalstimme deutlich beeinträchtigt.

Die Gruppen Fixation und Stillstand bezeichnen eine ungewisse Diagnose. Bei der Fixation wird ein Stimlippenstillstand beobachtet, wobei jedoch eine Lähmung ausgeschlossen werden kann. Beim Stillstand kann die Art der Lähmung nicht eindeutig bestimmt werden. Bei diesen Gruppen ist der Stimmlippenschluss uneinheitlich, es treten alle Zwischenstufen zwischen komplettem Schluss und Aphonie mit komplett fehlendem Schluss auf.

Folgende Gruppen mit Beeinträchtigungen der Stimmlippenbeweglichkeit wurden untersucht: Patienten mit einer Schädigung des Recurrens (n=24), Patienten mit einer Fixation der Stimmlippe (n=5), Patienten mit einem Stimmlippenstillstand (n=6) sowie Patienten mit einer Vaguslähmung (n=7). Bei der Gruppe mit Vaguslähmung wurde noch die Teilgruppe der Patienten, die keinen Stimmlippenschluss erzielen konnten, gesondert analysiert, denn zwei der Patienten mit Vaguslähmung waren, eher untypisch, in der Lage durch die Bewegung der gesunden Stimmlippe über den Median hinaus einen Stimmlippenschluss zu erreichen. Die Patienten der gesonderten Gruppe „Vagus (ohne)“ zeigten hingegen bei der videostroboskopischen Untersuchung keinen Stimmlippenschluss.

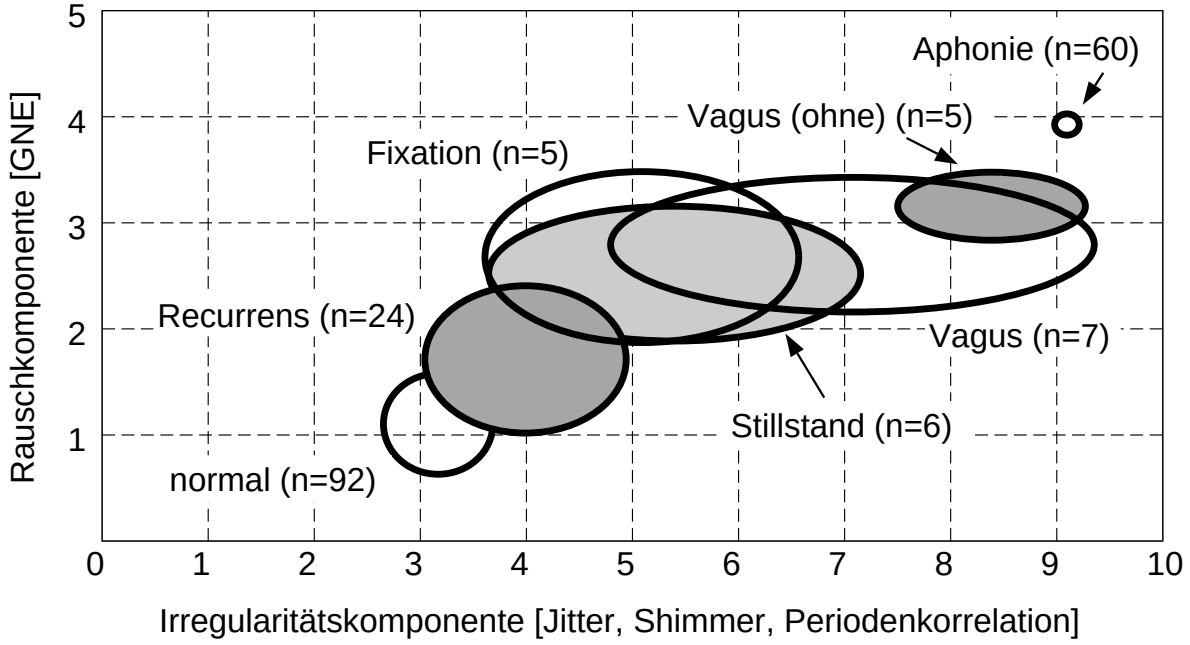


Abbildung 13.4.: Gruppen mit verschiedenen Lähmungsarten im Heiserkeits-Diagramm: Sprecher mit normaler und aphone Sprecher (simulierte Aphonie durch Flüstern)

Tabelle 13.5.: Lähmungsgruppen im Heiserkeits-Diagramm

		Irregulatitätskomponente		Rauschkomponente	
	n	Mittelwert	Standardabw.	Mittelwert	Standardabw.
normal	92	3,1678	0,5157	1,1044	0,4748
Stillstand	6	5,3977	1,7534	2,5210	0,6356
Recurrens	24	3,9915	0,9491	1,7125	0,6943
Vagus	7	7,0737	2,2796	2,7938	0,6346
Vagus (ohne)	5	8,3854	0,8848	3,1565	0,3207
Fixation	5	5,0882	1,4790	2,6767	0,8058
Aphonie	60	9,0958	0,1158	3,9266	0,1003

Die Lage der Gruppen entsprechend den Mittelwerten und Standardabweichungen im Heiserkeits-Diagramm ist in Abbildung 13.4 gezeigt. Die Werte selbst sind in Tabelle 13.5 aufgelistet. Als Referenz sind wieder Normalstimmen und aphone Stimmen beigefügt.

Die Gruppe der Recurrenslähmungen liegt der Normalgruppe am nächsten. Die Lage der Recurrensgruppe ist mit den Positionen der Gruppen mit Zysten und Polypen vergleichbar. Sowohl die Irregularität als auch der Rauschanteil sind höher als normal. Der höhere Rauschanteil kann als schlechterer Schluss gedeutet werden. Die höhere Irregularität weist darauf hin, dass eine periodische Schwingung bei einem asymmetrischen Schwingungssystem schwerer zu erreichen ist. Eine weitere Ursache der erhöhten Irregularität kann jedoch von dem Einfluss herrühren, den Rauschen auf die Irregularitätskomponente hat: Wird z.B. einem streng periodischen Sinussignal zunehmend Rauschen hinzugefügt, so wird die Energie pro Periode zunehmend variieren und die mittlere Periodenkorrelation sinken.

Die Gruppe der Vaguslähmungen liegt demgegenüber deutlich in die Richtung der aphonen Stimmen verschoben. Betrachtet man nur die Gruppe der Vaguslähmungen ohne Stimmlippenschluss, so ist diese Verschiebung in Richtung aphoner Stimmen noch deutlicher. Der Stimmlippenschluss dieser Gruppe ist sehr schlecht, das Signal wird von Rauschen dominiert.

Die Gruppen Fixation und Stillstand liegen im mittleren Bereich des Heiserkeits-Diagramms und streuen relativ stark. Dieses Ergebnis entspricht den undifferenzierten Phonationsmechanismen dieser Gruppen.

Tabelle 13.6.: Statistische Tests für Lähmungsgruppen. †: Unterschied nicht signifikant.

Zweidimensionaler Kolmolgorov-Smirnov Test						
	Stillstand	Recurrens	Vagus	Vagus (o.)	Fixation	Aphonie
normal	−1,42	−0,61	−1,69	−2,05	−1,57	−2,82
Stillstand		0,81†	−0,27†	−0,64†	−0,16†	−1,41
Recurrens			−1,08†	−1,44†	−0,96†	−2,21
Vagus				−0,36†	0,12†	−1,13
Vagus (ohne)					0,48†	−0,77
Fixation						−1,25

t-Test für ungleiche Varianzen der Rauschkomponente						
	Stillstand	Recurrens	Vagus	Vagus (o.)	Fixation	Aphonie
normal	2,64	1,02	4,26	5,61	2,48†	6,57
Stillstand		1,62†	1,70†	3,05†	0,35†	3,96
Recurrens			3,27†	4,63	1,46†	5,56
Vagus				1,36†	1,99†	2,32
Vagus (ohne)					3,33†	1,05
Fixation						4,20

t-Test für ungleiche Varianzen der Irregularitätskomponente						
	Stillstand	Recurrens	Vagus	Vagus (o.)	Fixation	Aphonie
normal	−2,23	−0,82	−3,91	−5,22	−1,92†	−5,93
Stillstand		1,41†	−1,68†	−2,99†	0,31†	−3,70
Recurrens			−3,08†	−4,39	−1,10†	−5,10
Vagus				−1,31†	1,99†	−2,02†
Vagus (ohne)					3,30†	−0,71†
Fixation						−4,01

13.4. Verschiedene Phonationsmechanismen

Die folgenden Gruppenanalysen beziehen sich alle auf Patienten bei denen ein bösartiger glottaler Tumor entfernt wurde [27, 30, 32, 67, 87, 158]. Je nach Art des Tumors musste die Gewebestruktur mehr oder weniger stark verändert werden. Diese verschiedenen postoperativen Zustände der Kehlkopfstruktur ermöglichen nach der Wundheilung unterschiedliche Arten der Ersatzphonation. Hier wird von Ersatzphonationen gesprochen, da eine Tumorentfernung immer Narben nach sich zieht, so dass die spätere Phonation immer von der normalen, physiologischen Phonation abweicht.

Die postoperativen Ersatzphonationsmechanismen können in vier Klassen eingeteilt werden: glottische Ersatzphonation mit Schwingung der operierten Stimmlippe, pseudoglottische Ersatzphonation ohne Schwingung der operierten Stimmlippe, ventrikuläre Ersatzphonation (oder Taschenfaltenstimme) und die ary-epiglottische Ersatzphonation. Auf den folgenden Abbildungen sind laryngoskopische Bilder dieser Phonationsmechanismen dargestellt.

Bei dem Patienten in Abbildung 13.5 ist im oberen Bild, im oberen Drittel der rechten Stimmlippe die Operationsnarbe zu erkennen. In der Atmungsstellung im oberen Bild bilden die Stimmlippen ein „V“, zwischen diesem blickt man in die Luftröhre. Die Spitze des V's zeigt, vom Patienten aus gesehen, nach vorne. Zur Phonation werden die Aryknorpel, die im unteren Bild am oberen Rand (etwas unscharf) zu erkennen sind, aufeinanderzubewegt. Die Stimmlippen werden dadurch zusammengebracht. Ein Glottisverschluss bei der Phonation ist möglich. In der stroboskopischen Videoaufnahme dieses Patienten ist zu erkennen, dass beide Stimmlippen, die gesunde und die operierte, schwingen.

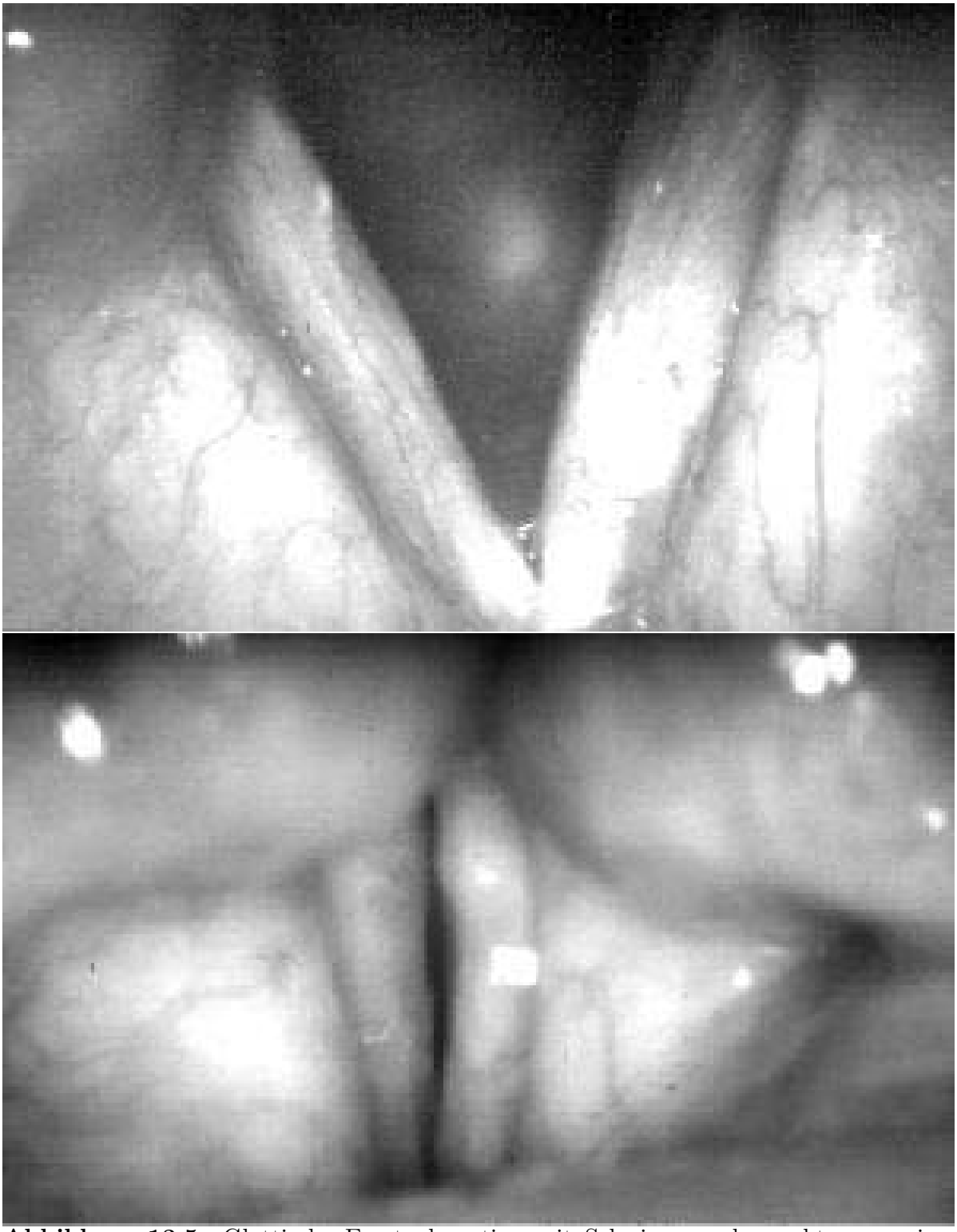


Abbildung 13.5.: Glottische Ersatzphonation mit Schwingung der rechten, operierten Stimmlippe (in der fotografischen Aufsicht links). Oben: Atmungsstellung. Unten: Phonationsstellung

Die linke Stimmlippe des Patienten, dessen laryngoskopisches Kehlkopfbild in Abbildung 13.6 gezeigt wird, zeigt deutliche Narbenstrukturen von der Operation. Diese Stimmlippe ist weitgehend steif. Bei der Phonation (unten in der Abbildung) schwingt nur die gesunde Stimmlippe. Der Stimmlippenschluss ist entsprechend unzureichend.

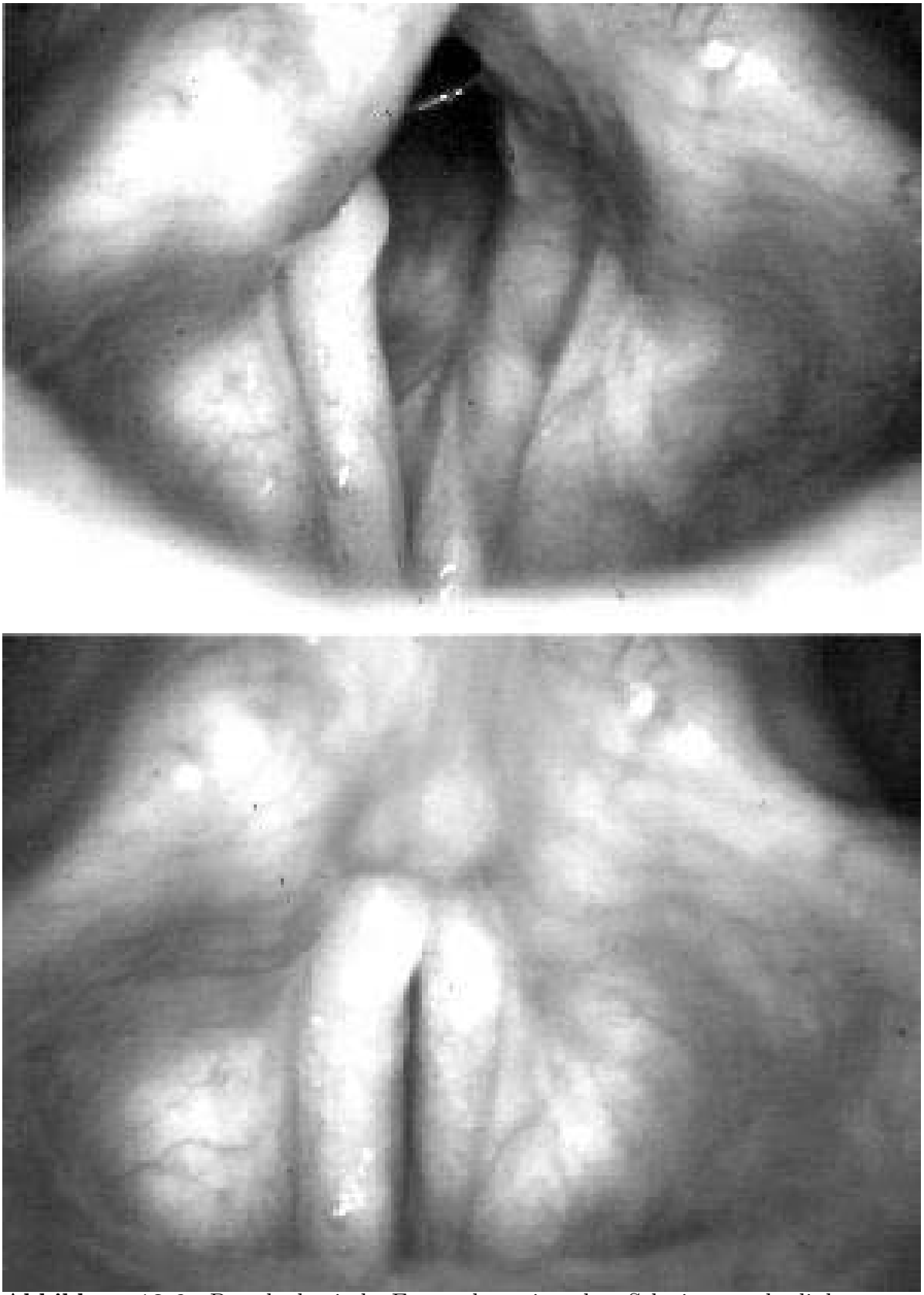


Abbildung 13.6.: Pseudoglottische Ersatzphonation ohne Schwingung der linken, operierten Stimmlippe (in der fotografischen Aufsicht rechts). Oben: Atmungsstellung. Unten: Phonationsstellung

Im Gegensatz zu den beiden vorigen Patienten wurde die Phonationsebene bei dem Patienten in Abbildung 13.7 nach oben zu den Taschenfalten verlagert. Die Taschenfalten sind im oberen Bild als leichte Vorwölbungen oder Hügel seitlich oberhalb der Stimmlippen zu erkennen. Im unteren Bild ist deutlich der Verschluss durch die Taschenfalten zu sehen.

Die Taschenfalten bilden wie die Stimmlippen eine symmetrische Struktur, die zur Phonation verwendet werden kann. Die Taschenfaltenstimme klingt im Allgemeinen tiefer (durch die größere Masse der Taschenfalten) und rauher als die Stimmlippenphonation (prominente Beispiele für den Einsatz der Taschenfalten in der Musik sind Louis Armstrong und Joe Cocker). Entwicklungsbiologisch bilden die Taschenfalten und die Stimmlippen ein Atmungsdoppelventil, dessen Primärfunktion es ist, das Eindringen von Fremdkörpern in die Lunge zu verhindern. Die Stimmfunktion ist eine Sekundärfunktion zu der Verschlussfunktion und sehr viel jünger.

Dank dieses Doppelventils steht den Patienten nach Entfernung einer oder beider Stimmlippen ein Ersatzsystem für die Phonation zur Verfügung, das eine akzeptable Stimmqualität erreicht.

Im oberen Bild ist unten der Kehldeckel, die Epiglottis, zu erkennen.

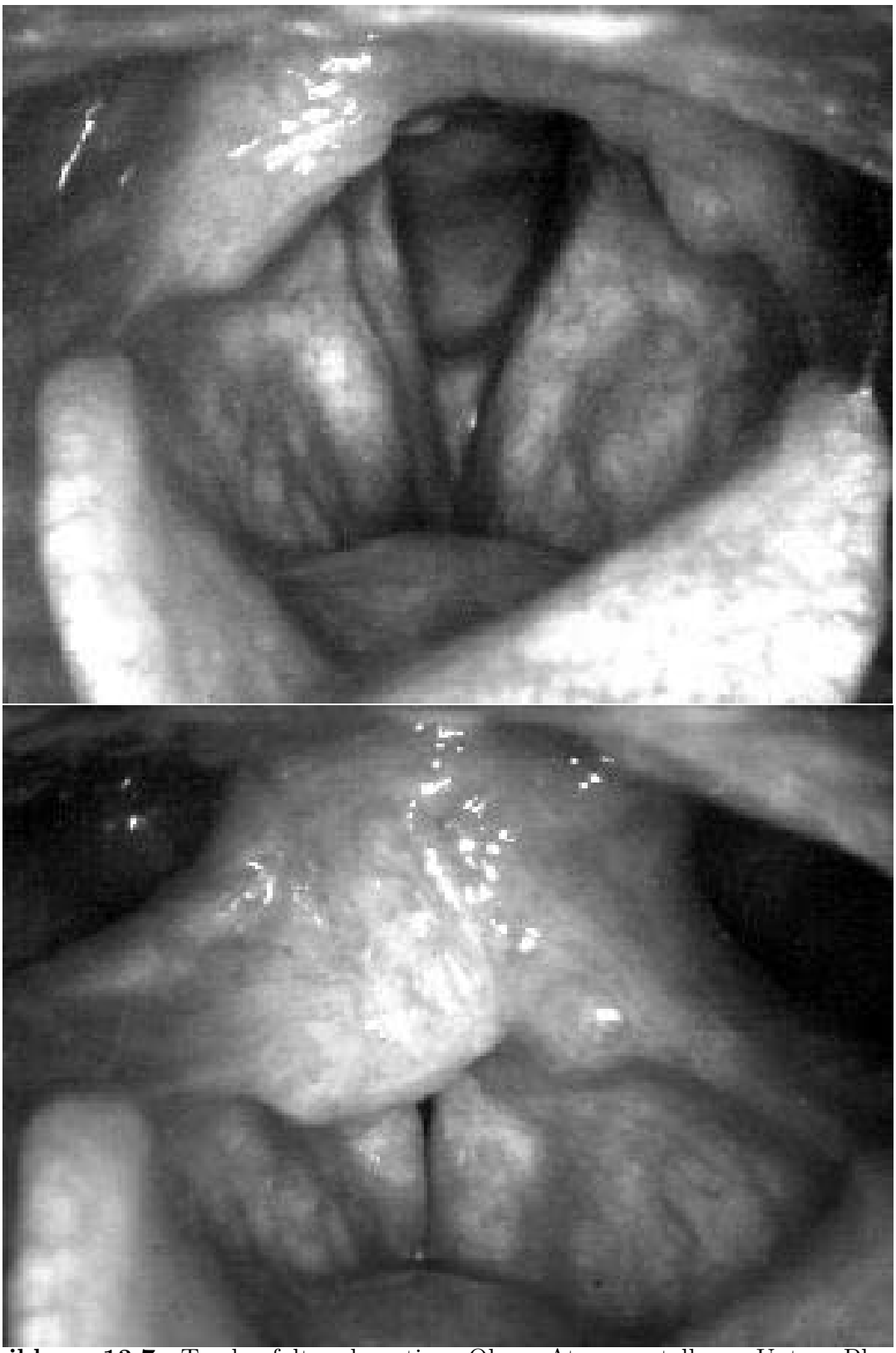


Abbildung 13.7.: Taschenfaltenphonation. Oben: Atmungsstellung. Unten: Phonationsstellung

Wenn die Stimmlippen und die Taschenfalten entfernt werden müssen, wie bei dem Patienten in Abbildung 13.8 zu sehen, so kann es zur Ausprägung der aryepiglottischen Ersatzphonation kommen. Bei diesem Patienten wird der linke Aryknorpel (im oberen Bild rechts oben) zur Phonation in Richtung der Epiglottis geführt (unteres Bild). Es entsteht eine sehr unregelmäßige Schwingung, die auch im stroboskopischen Aufnahmen nicht genau zu orten ist, da der Aryknorpel die Sicht beschränkt. Bei der Phonation ist deutlich ein relativ starker Rauschanteil hörbar. Diese Phonationsart ist in funktioneller Hinsicht relativ weit von der normalen Stimmgebung entfernt.

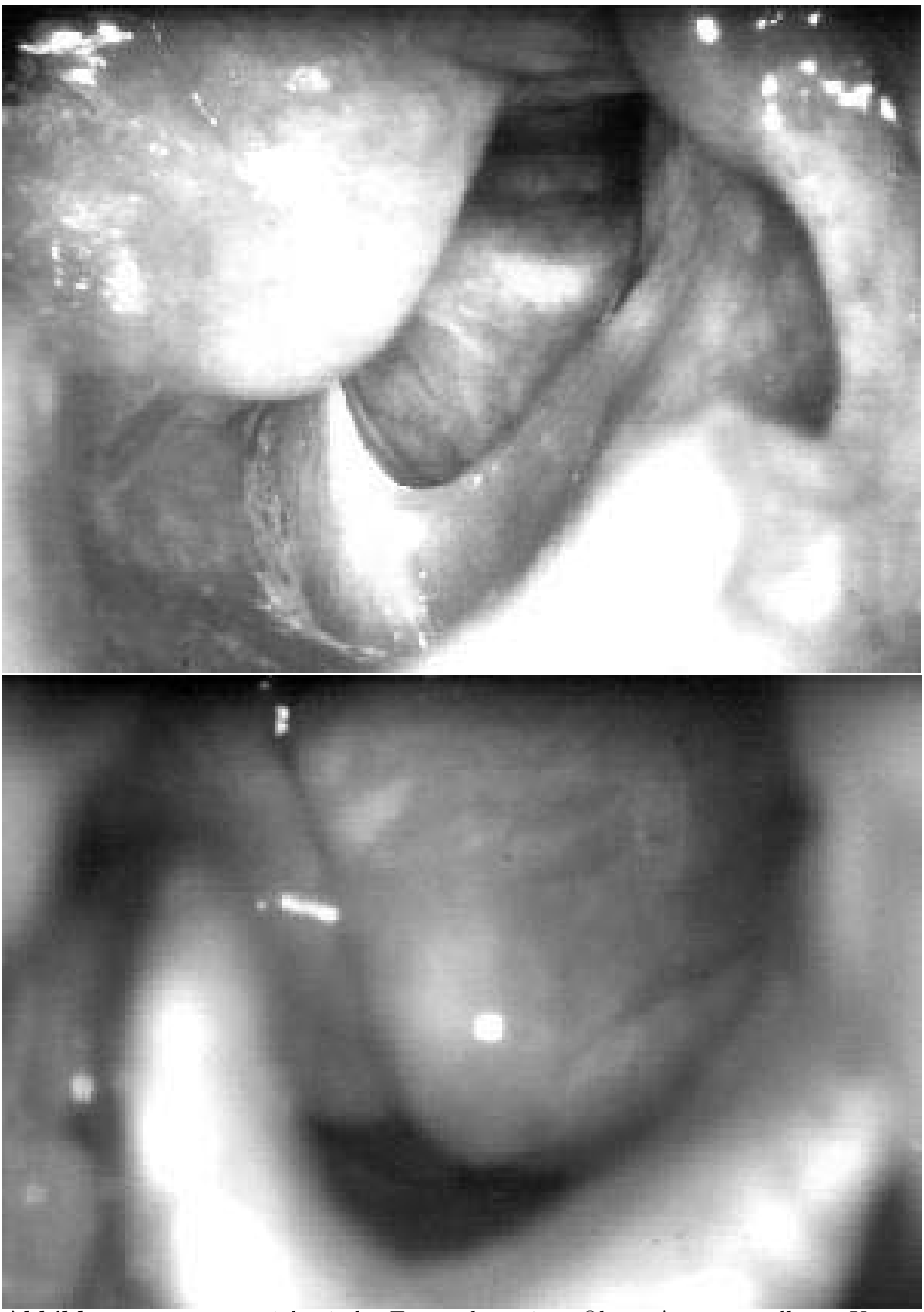


Abbildung 13.8.: aryepiglottische Ersatzphonation. Oben: Atmungsstellung. Unten: Phonationsstellung

Im Folgenden werden die Irregularitäts- und Rauschwerte von Patientengruppen mit den eben beschriebenen Phonationsmechanismen untersucht. Als Vergleich dienen wieder die normale und die aphone Gruppe.

Die Mittelwerte und Standardabweichungen der Irregularitäts und Rauschkomponenten folgender Gruppen sind in der Abbildung 13.9 und der Tabelle 13.7 aufgeführt: normal mit 92 Sprechern, glottische Phonation mit Schwingung der operierten Stimmlippe mit 15 Patienten, glottische Phonation ohne Schwingung der operierten Stimmlippe mit 9 Patienten, ventrikuläre Ersatzphonation (oder Taschenfaltenstimme) mit 10 Patienten, aryepiglottische Ersatzphonation mit 6 Patienten und die Gruppe der aphonen Stimmen mit 30 Aufnahmen.

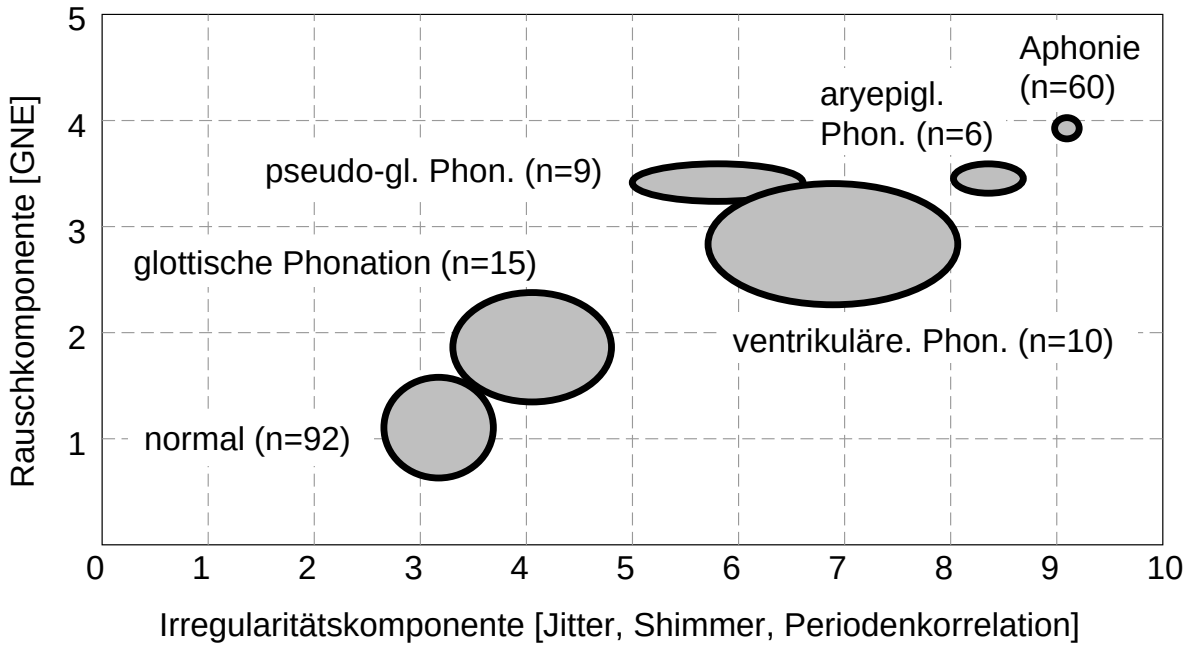


Abbildung 13.9.: Gruppen mit verschiedenen Phonationsmechanismen im Heiserkeits-Diagramm: Sprecher mit normaler Phonation, mit glottischer Ersatzphonation, mit pseudo-glottischer Ersatzphonation, ventrikulärer Ersatzphonation, aryepiglottischer Ersatzphonation und aphone Sprecher (Simulierte Aphonie durch Flüstern)

Tabelle 13.7.: Krebsgruppen mit verschiedenen Phonationsmechanismen im Heiserkeits-Diagramm

		Irregulatitätskomponente		Rauschkomponente	
	n	Mittelwert	Standabw.	Mittelwert	Standabw.
normal	92	3,17	0,52	1,10	0,47
gl. EP	15	4,05	0,75	1,86	0,52
pseudo-gl. EP	9	5,80	0,80	3,41	0,18
ventrik. EP	10	6,89	1,18	2,83	0,57
ary-epigl. EP	6	8,36	0,33	3,45	0,14
Aphonie	60	9,10	0,12	3,93	0,10

Tabelle 13.8.: Statistische Tests für Krebsgruppen. †: Unterschied nicht signifikant.

Zweidimensionaler Kolmolgorov-Smirnov Test

	gl. EP	pseudo-gl. EP	ventrik. EP	ary-epigl. EP	Aphonie
normal	1,16	3,50	4,10	5,69	6,56
gl. EP		2,34	3,00	4,59	5,45
pseudo-gl. EP			1,23	2,55	3,33
ventrik. EP				1,59	2,46
ary-epigl. EP					0,88

t-Test für ungleiche Varianzen der Rauschkomponente

	gl. EP	pseudo-gl. EP	ventrik. EP	ary-epigl. EP	Aphonie
normal	−0,76	−2,31	−1,73	−2,35	−2,82
gl. EP		−1,55	−0,97	−1,59	−2,06
pseudo-gl. EP			0,58†	−0,04†	−0,51
ventrik. EP				−0,62†	−1,09
ary-epigl. EP					−0,47

t-Test für ungleiche Varianzen der Irregularitätskomponente

	gl. EP	pseudo-gl. EP	ventrik. EP	ary-epigl. EP	Aphonie
normal	−0,88	−2,63	−3,72	−5,18	−5,92
gl. EP		−1,75	−2,84	−4,30	−5,04
pseudo-gl. EP			−1,09†	−2,55	−3,29
ventrik. EP				−1,47†	−2,21
ary-epigl. EP					−0,74

Die Verteilungen der Gruppen im Heiserkeits-Diagramm sind alle signifikant verschieden (Tabelle 13.8). Die akustische Stimmgütebeschreibung mit dem Heiserkeits-Diagramm belegt also, dass alle Ersatzphonationen eine Beeinträchtigung der Stimmgüte gegenüber der Normalgruppe zeigen. Weiterhin zeigt sich eine Staffelung der Ersatzphonationsmechanismen, die dem Grad der strukturellen Veränderungen des Kehlkopfes entspricht: Die glottische Ersatzphonation mit Schwingung der operierten Stimmlippe liegt der normalen Stimmgebung am nächsten. Dieser folgt die pseudo-glottische Ersatzphonation ohne Schwingung der operierten Stimmlippe bei deutlich erhöhter Rauschkomponente und die ventrikuläre Ersatzphonation bei deutlich erhöhter Irregularität. Noch weiter in Richtung der aphonen Stimmen liegt die aryepiglottische Ersatzphonation, die sich jedoch noch signifikant von der Aphonie unterscheidet.

14. Korrelation von akustischen und (subjektiven) perzeptiven Stimmgütemessgrößen

Bei der perzeptiven Stimmbewertung geht es darum, den subjektiven Höreindruck der Stimmqualität in möglichst geeigneter Form zu quantifizieren. Bei Untersuchungen auf diesem Gebiet erfolgt die Stimmbewertung meist durch mehrere Spezialisten, wie z.B Logopäden, Phoniater, Linguisten oder Sprachpathologen.

Es gibt bei der perzeptiven Stimmbeurteilung im Wesentlichen zwei verschiedene Beurteilungsstrategien, wobei wiederum verschiedene Verfahren zur Analyse dieser Beurteilungen angewandt werden. Diese Beurteilungsmethoden sind: Erstens die Beurteilung einer bestimmten Stimmqualität anhand einer vorgegebenen Skala, die die Quantität der jeweiligen Stimmeigenschaft ausdrückt, z.B. die Beurteilung der Heiserkeit einer Stimme auf einer Skala von 1 bis 4. Zweitens kann man die Ähnlichkeiten von je zwei Stimmen anhand einer bestimmten Stimmqualität oder des Gesamteindruckes der Stimmqualität auswerten.

Für Beurteilungen der ersten Art haben die sogenannte GRBAS-Skala, die von der Japanischen Gesellschaft für Logopädie vorgeschlagen wurde, und Untermengen dieser Skala Verbreitung in der Literatur gefunden. Dabei bedeuten die einzelnen Buchstaben: (G) grade, Gesamtgrad der Störung, (R) roughness, Rauhigkeit der Stimme, (B) breathiness, Behauchtheit, (A) asthenia Schwächung, (S) strained, Grad der Anspannung. Diese Qualitäten werden auf einer 4-Punkte-Skala bewertet: (0) normal, (1) leicht, (2) moderat, (3) extrem.

Eine Skala mit nur 3 Stimmqualitäten ist in Deutschland verbreitet, die HBR-Skala. Dabei steht H für Hoarseness, die anderen beiden Buchstaben entsprechen der Bedeutung in der GRBAS-Skala: B breathiness und R roughness.

Die perzeptive Stimmbewertung durch Experten ist der Maßstab, an dem sich die technischen Methoden zur Stimmbeurteilung messen müssen, denn es geht hierbei gerade darum, den subjektiven Höreindruck in Zahlen und Messgrößen festzuhalten. Eine wichtige Voraussetzung dafür ist, dass verschiedene Experten untereinander konsistent und dass einzelne Sprecher bei einem Test-Retest-Versuch selbstkonsistent ihren Höreindruck von einer Stimme quantifizieren können.

In der Literatur wurde an vielen Stellen der Zusammenhang zwischen akustischer und perzeptiver Stimmgütebeurteilung sowie die Verlässlichkeit der perzeptiven Beurteilung

diskutiert [3, 13, 18, 19, 21, 33, 35, 36, 40–42, 49, 57, 61–65, 92, 93, 108, 109, 127, 149]. Dabei stellte sich zum Beispiel in einigen Studien heraus, dass Jitter signifikant mit der Behauchung korreliert ist. Aus einem signaltheoretischen Blickwinkel ist diese Ergebnis sehr unbefriedigend. Denn Behauchung wird im Allgemeinen mit meist höherfrequenten Rauschanteilen assoziiert. Jitter jedoch soll ein Maß für Periodenlängenschwankungen sein und deshalb eher mit der perzeptiven Rauhigkeit korrelieren.

Für diese Studie über den Zusammenhang von akustischen und perzeptiven Maßen der Stimmgüte standen 120 Aufnahmen pathologischer Stimmen, die im Rahmen einer multizentrischen Studie in Berlin aufgenommen wurden, zur Verfügung. Hierbei handelte es sich um die DAT-Aufnahmen der Vokale [a:] die jeweils zweimal hintereinander auf dem Band aufgezeichnet waren. Da diese Aufnahmen einen tieffrequenten Störsignalanteil enthielten (der aber perzeptuell nicht wahrnehmbar war), wurden die Aufnahmen mit einem 70Hz Hochpassfilter vorverarbeitet. Bei visueller Signaldarstellung war der Störanteil nach der Filterung nicht mehr zu erkennen [157].

Die folgende Studie lässt sich am einfachsten mit einer Behauptung motivieren, die experimentell geprüft werden soll:

Akustische Stimmgütemessgrößen sollen einen Zusammenhang zu perzeptiven Eigenschaften der Stimme haben. Insbesondere sollen Stimmgütemessgrößen gefunden werden, die spezifisch für die perzeptive Rauhigkeit und spezifisch für Behauchung sind.

Im Folgenden sollen deshalb, wenn möglich, spezifische Stimmgütemessgrößen aus einer Menge von Messgrößen selektiert werden.

Die Voraussetzung für eine solche Studie ist Datenmaterial, das einerseits die perzeptive Beurteilung und andererseits die akustischen Messgrößen enthält.

Die perzeptive Beurteilungen wurden für 120 Vokale [a:] von 20 Laien (Anfängerinnen der Logopädenausbildung) und 8 Experten (Logopäden, Phoniatern) erhoben. Die Bewerter sollten auf einem Testbogen für jeden der 120 Vokale den Grad der Rauhigkeit (R), der Behauchung (B) und der Heiserkeit (H) auf einer diskreten Skala von 0 (nicht vorhanden) bis 3 (extrem) bewerten. Dabei wurde an die Beurteilungen die Randbedingung gestellt, dass der Heiserkeitswert H nicht kleiner sein darf als die Behauchung und die Rauhigkeit. Diese Bedingung wurde nicht immer von allen Bewertern eingehalten. Die Vokale wurden bei zwei Sitzungen von einem DAT-Band über Lautsprecher jeweils einer Gruppe von Bewertern vorgespielt.

Die 120 Vokale enthielten drei Vokale doppelt, anhand derer die Zuverlässigkeit der Bewerter beurteilt wurde. Die Daten eines Beurteilers wurden von der weiteren Untersuchung ausgeschlossen, wenn die Summe der Abweichungen von R, B und H bei einem dieser drei Vokale größer als eins war. Diese Bedingung traf für 7 Laien und einen Experten zu, so dass für die weitere Analyse die Beurteilungen von 7 Experten und 13 Laien zur Verfügung standen.

Da es in dieser Arbeit nicht primär um die perzeptive Beurteilung von Stimmen gehen soll, werden hier keine Ergebnisse über die Vergleichbarkeit der Bewertungen zwischen den Beurteilern dargestellt. Diese ist in anderen Studien bereits untersucht worden (s.o.). Für das Folgende ist nur wichtig, dass für drei Gruppen (a) alle Bewerter, n=20, (e) Experten, n=7 und (l) Laien, n=13 jeweils der Medianwert der RBH-Beurteilungen gebildet wurde und als Basis für die weiteren Untersuchungen dient.

Aus den Audioaufnahmen der Vokale wurden 36 akustische Messgrößen berechnet. Diese Messgrößen wurden alle schon weiter oben beschrieben.

Tabelle 14.1.: Korrelationen zwischen den perzeptiven Beurteilungsgrößen und den akustischen Irregularitätsmessgrößen (* bedeutet, die "Korrelation ist nicht signifikant bei einem multiplen Signifikanzniveau von $p < 0,05$ für die Bonferoni-Holm Korrektur", † bedeutet, dass hier nur 117 der 120 Patienten erfolgreich analysiert werden konnten.)

	alle			Experten			Laien		
	R	B	H	R	B	H	R	B	H
p	0,09*	−0,23*	−0,10*	0,24*	−0,17*	−0,09*	0,07*	−0,23*	−0,09*
MWC	0,80	0,73	0,80	0,64	0,74	0,78	0,77	0,74	0,78
J2	0,79	0,69	0,77	0,69	0,65	0,75	0,78	0,68	0,75
J3	0,78	0,70	0,77	0,67	0,66	0,75	0,77	0,69	0,76
J5	0,79	0,69	0,77	0,68	0,65	0,75	0,78	0,68	0,75
J7	0,79	0,68	0,76	0,70	0,64	0,75	0,79	0,67	0,75
J11	0,79	0,66	0,75	0,71	0,63	0,74	0,79	0,65	0,74
J15	0,79	0,64	0,74	0,71	0,61	0,73	0,78	0,63	0,73
S2	0,79	0,66	0,74	0,67	0,64	0,72	0,75	0,66	0,73
S3	0,77	0,69	0,75	0,62	0,67	0,72	0,74	0,69	0,75
S5	0,79	0,69	0,76	0,66	0,67	0,74	0,76	0,68	0,74
S7	0,80	0,67	0,74	0,68	0,65	0,73	0,76	0,67	0,73
S11	0,80	0,64	0,73	0,70	0,63	0,72	0,76	0,64	0,72
S15	0,79	0,62	0,70	0,71	0,61	0,70	0,75	0,61	0,70
hd-irreg	0,81	0,74	0,80	0,67	0,74	0,79	0,78	0,74	0,79
hd-ed	0,80	0,80	0,84	0,64	0,80	0,83	0,79	0,80	0,82
Fo†	−0,04*	0,30*	0,16*	−0,18*	0,24*	0,17*	−0,01*	0,31*	0,16*
Jita†	0,56	0,62	0,65	0,43	0,61	0,64	0,57	0,61	0,63
Jitt†	0,58	0,71	0,69	0,38	0,68	0,69	0,59	0,70	0,67
RAP†	0,58	0,71	0,69	0,38	0,69	0,69	0,59	0,70	0,67
PPQ†	0,58	0,70	0,68	0,39	0,68	0,69	0,59	0,69	0,66
ShdB†	0,69	0,64	0,72	0,58	0,61	0,73	0,66	0,64	0,70
Shim†	0,70	0,66	0,73	0,59	0,63	0,74	0,67	0,66	0,71
APQ†	0,69	0,62	0,70	0,59	0,59	0,71	0,66	0,62	0,68

In den Tabellen 14.1 und 14.2 sind die Korrelationswerte (Pearson's r) zwischen den Medianwerten der perzeptiven RBH-Größen und den akustischen Messgrößen zusammengefasst. Bis auf die mittlere Periodenlänge p (mit dem Waveform Matching berechnet), der mittleren Grundfrequenz F_0 (mit MDVP berechnet) und dem Soft Phonation Index SPI sind *alle akustischen Messgrößen mit allen perzeptiven Größen signifikant korreliert*. Die drei Ausnahmen hingegen korrelieren mit keiner perzeptiven Größe bei allen Gruppen. Die fehlende Korrelation zur Periodenlänge bzw. zur Grundfrequenz lässt sich so deuten, dass die Bewerter in der Lage sind, von dem Tonhöheneindruck abzusehen und sich nur auf die RBH-Auffälligkeiten zu konzentrieren. Die fehlende Korrelation zum SPI zeigt, dass das Verhältnis von harmonischer Energie im Bereich 70-1600Hz zum höherfrequenten Rauschanteil (1600-4500Hz) kein entscheidender Hinweis auf eine der perzeptiven RBH- Größen ist.

Tabelle 14.2.: Korrelationen zwischen den perzeptiven Beurteilungsgrößen und den akustischen Messgrößen für den Rauschanteil (* bedeutet, die Korrelation ist nicht signifikant bei einem multiplen Signifikanzniveau von $p < 0,05$ für die Bonferoni-Holm Korrektur, † bedeutet, dass hier nur 117 der 120 Patienten erfolgreich analysiert werden konnten.)

	alle			Experten			Laien		
	R	B	H	R	B	H	R	B	H
GNE1	−0,64	−0,83	−0,79	−0,45	−0,82	−0,78	−0,65	−0,83	−0,76
GNE2	−0,57	−0,82	−0,76	−0,37	−0,82	−0,74	−0,61	−0,82	−0,73
GNE3	−0,54	−0,79	−0,73	−0,34	−0,80	−0,72	−0,58	−0,77	−0,71
NNE1	0,78	0,65	0,74	0,66	0,66	0,74	0,74	0,66	0,73
NNE2	0,77	0,64	0,73	0,66	0,65	0,73	0,73	0,65	0,72
NNE3	0,72	0,80	0,82	0,53	0,77	0,75	0,73	0,79	0,82
CHNR1	−0,75	−0,78	−0,81	−0,57	−0,78	−0,78	−0,74	−0,78	−0,79
CHNR2	−0,74	−0,77	−0,80	−0,56	−0,78	−0,77	−0,72	−0,77	−0,78
CHNR3	−0,70	−0,83	−0,83	−0,48	−0,81	−0,76	−0,72	−0,83	−0,82
NHR†	0,54	0,51	0,57	0,43	0,49	0,61	0,53	0,50	0,55
VTI†	0,51	0,50	0,52	0,36	0,53	0,54	0,49	0,50	0,50
SPI†	0,04*	0,01*	0,04*	0,03*	−0,04*	−0,04*	0,03*	−0,03*	0,06*

Insgesamt ist dieses Ergebnis zunächst einmal sehr erfreulich, zeigt es doch, dass sich fast alle implementierten akustischen Messgrößen eignen um zumindest Teilaspekte der wahrgenommenen Stimmstörungen richtig zu beschreiben. Damit ist für die praktische Anwendung eine wichtige Grundvoraussetzung an die Interpretierbarkeit der akustischen Messgrößen erfüllt.

Andererseits ist dieses Ergebnis auch schon in anderen Studien gefunden worden und die Aussage erlaubt keinerlei Bewertung hinsichtlich der vorangestellten Behauptung, dass akustische Messgrößen spezifischen perzeptiven Stimmeigenschaften entsprechen sollten.

Als Ursache für die unspezifische Korrelation fast aller Messgrößen zu allen perzeptiven Größen kommen mindestens zwei Gesichtspunkte in Frage: Erstens könnten alle Messgrößen auf akustische Signaleigenschaften reagieren, die bei allen drei perzeptiven Größen R, B und H auftreten. Zweitens können die den drei perzeptiven Größen entsprechenden Signaleigenschaften zwar verschieden sein, jedoch bei den hier gewählten 120 Vokalen oft zusammen auftreten. Demnach wäre das Datenmaterial bezüglich dieser Eigenschafte schon nicht unkorreliert.

Tabelle 14.3.: Besetzung der R und B Bewertung (Median aller Bewerter)

	B=0	B=1	B=2	B=3
R=0	18	16	0	0
R=1	11	22	12	3
R=2	0	6	8	11
R=3	0	0	5	8

Tabelle 14.4.: Korrelationen zwischen den perzeptiven Beurteilungsgrößen

	B	H
alle:		
R	0,71	0,84
B		0,91
Experten:		
R	0,48	0,69
B		0,82
Laien:		
R	0,74	0,87
B		0,89

Ein Blick auf die Tabellen 14.3 und 14.4 bestätigt zunächst einmal den zweiten Verdacht: Tabelle 14.3 zeigt, dass bestimmte R-Werte sehr häufig mit B-Werten gepaart sind, die den gleichen Wert haben oder sich höchstens um einen Skalenpunkt unterscheiden. Die Wertepaare R=0 und B=2,3, R=2 und B=0 sowie R=3 und B=0,1 treten gar nicht auf. Tabelle 14.4 bestätigt zusätzlich, dass bei allen Gruppen die R-, B- und H-Bewertungen sehr hoch miteinander korreliert sind. Der kleinste Korrelationswert und somit die beste Unabhängigkeit findet sich bei den Experten zwischen B und R. Sie sind vielleicht besser in der Lage, auf spezielle Stimmeigenschaften zu achten und sich nicht so sehr von dem Gesamteindruck beeinflussen zu lassen.

An diesen hohen Korrelationswerten kann man aber auch ablesen, dass die perzeptiven Größen Behauchtheit und Rauhigkeit oft zusammen auftreten (bei H war das wegen der Randbedingung, dass H mindestens so groß sein sollte wie B und R, zu erwarten). Deswegen stammt ein Teil der Korrelationen der akustischen Messgrößen untereinander aus der vorhandenen Korrelation im Datenmaterial.

Um diesem Problem zu begegnen, wurde deshalb in einem ersten Schritt jeweils die Bewertung für Rauhigkeit (bzw. Behauchung) vom linearen Behauchungstrend (bzw. Rauhigkeitstrend) befreit und erst dann mit den akustischen Messgrößen korreliert (Tabelle 14.5) und in einem zweiten Schritt wurde die Rauhigkeit (bzw. Behauchung) vom linearen Behauchungs- und Heiserkeitstrend (bzw. Rauhigkeits- und Heiserkeitstrend) befreit und dann korreliert (Tabelle 14.6).

Dadurch sinken zwar zwangsläufig die Korrelationswerte; falls jedoch die perzeptiven Größen einen, von den anderen perzeptiven Größen unabhängigen Anteil an der Varianz des gegebenen Datenmaterials aufweisen, sollte noch eine Restkorrelation zu den akustischen Messgrößen bestehen bleiben.

In Tabelle 14.5 sind die Zeilen derjenigen akustischen Messgrößen hervorgehoben, die bei allen Gruppen (a,e,l) nur mit *einem* der beiden perzeptiven Maße (Behauchung, Rauhigkeit) signifikant korreliert sind. Diese Parameter können vom Blickwinkel dieser Untersuchung her als spezifische Parameter betrachtet werden, zumindest was die Beziehung zu perzeptiven Messgrößen anbelangt. Je deutlicher hier der Unterschied in den Korrelationswerten zu Behauchung und Rauhigkeit ausfällt, umso besser ist die Spezifität der entsprechenden Messgröße.

Am deutlichsten ist dies bei den drei GNEs der Fall, bei denen die Korrelationen zu der, von der rauhigkeitsbefreiten Behauchungsbewertung (0,51 bis 0,73) deutlich über denen zu der behauchungsbefreiten Rauhigkeit (0,01 bis 0,07) liegen. Diese drei Messgrößen können nach dieser Untersuchung mit gutem Recht als Messgrößen bezeichnet werden, die einen wichtigen Aspekt mit der perzeptiven Behauchung gemeinsam haben und nur wenig mit dem Rauhigkeitseindruck der Stimme. Das Design des GNE und die Tests mit synthetischen Signalen, legen die Vermutung nahe, dass dieser gemeinsame Aspekt der im Signal enthaltene Rauschanteil ist.

Ähnlich spezifisch, aber mit deutlich geringerer Differenz der Korrelationswerte zu Behauchung und Rauhigkeit folgen dem GNE die CHNRs und der NNE 3. Diese Messgrößen sind zwar nicht signifikant, aber doch in der zu erwartenden Richtung mit der Rauhigkeit korreliert: Je höher die Rauhigkeit, desto höher NNE bzw. desto niedriger CHNR. Dies entspricht den Erwartungen, dass Rauhigkeit mit höherem Jitter und Shimmer einhergeht. Denn die Versuche mit synthetischen Signalen hatten ja schon gezeigt, dass CHNR und NNE deutlich stärker durch Jitter und Shimmer beeinflusst werden als die GNEs.

Somit wäre auch aufgrund der Korrelationsmessungen zu perzeptiven Größen dem GNE der Vorzug gegenüber NNE und CHNR zu geben, wenn auf hohe Spezifität Wert gelegt wird.

Der Vergleich der akustischen Maße zeigte, dass einige Maße spezifisch für bestimmte perzeptive Größen sind in dem Sinne, dass sie bei einer entsprechenden Trendbefreiung der perzeptiven Größen jeweils nur mit einer perzeptiven Größe signifikant korreliert sind.

Die Maße zur Messung des Rauschanteils GNE (1,2,3), CHNR (1,2,3) und NNE (nur NNE3) sind spezifisch für die Behauchung. Insbesondere bei den GNEs sind die Korrelationen zu dem Rauhigkeitsresiduum sehr gering und die zu dem Behauchungsresiduum relativ hoch. Deshalb entspricht der GNE am ehesten der Behauptung am Anfang des Kapitels, dass nämlich akustische Maße spezifisch zu perzeptiven Maßen sein sollten (hier spezifisch für Behauchung). Die Maße der MDVP Software NHR und VTI sind nicht spezifisch.

Die akustischen Jitter- und Shimmer-Maße des MDVP sind ebenfalls nicht spezifisch: Die Jitter- und Shimmer-Maße, die bei dieser Software auf ereignisbasierter Periodenlängenberechnung beruhen, sind offensichtlich rauschanfällig, denn auch nach der

Trendbefreiung von der Rauhigkeitsbewertung sind diese Maße des MDVP signifikant mit dem Behauchungsresiduum (der Expertengruppe) korreliert. Anders sieht es bei den Jitter- und Shimmer-Maßen aus, die auf dem Waveform Matching basieren: J11 und J15 sowie S11 und S15 sind im obigen Sinne spezifisch für Rauhigkeit. Bei Jitter und Shimmer sind gerade die Perturbationsmaße spezifisch, welche lokal über relativ viele Perioden mitteln (11 und 15). Dies ist ein Hinweis darauf, dass der perzeptive Rauhigkeitseindruck nicht mit den schnellsten Fluktuationen zusammenhängt.

Das vierte akustische Maß das neben GNE, Jitter und Shimmer eingeht, nämlich MWC ist im oberen Sinne nicht spezifisch.

Insgesamt sind also drei der vier akustischen Maße des Heiserkeits-Diagramms spezifisch: GNE, der allein in die Rauschkomponente eingeht ist spezifisch für Behauchung, Jitter und Shimmer (mit Waveform-Matching-Verfahren), die zwei Drittel der Irregularitätskomponente bestimmen, sind rauhigkeitsspezifisch.

Nach zusätzlicher Befreiung vom Heiserkeitstrend bleiben nur noch drei signifikante Korrelationswerte (Tabelle 14.6). Damit können keine Schlüsse auf die Spezifität der Maße gezogen werden.

Tabelle 14.5.: Korrelationen ausgewählter Messgrößen mit den perzeptiven Beurteilungen Rauhigkeit und Behauchung. Hier wurde vor der Korrelationsberechnung der lineare Trend der Behauchung aus der beurteilten Rauhigkeit und der lineare Trend der Rauhigkeit aus der Behauchung rausgerechnet.(* bedeutet, dass die Korrelation nicht signifikant ist bei einem multiplen Signifikanzniveau von $p < 0,05$ für die Bonferonie-Holm Korrektur für 717 Tests.)

	Rauhigkeitsresiduum			Behauchungsresiduum		
	r (alle)	r (Expert.)	r (Laien)	r (alle)	r (Expert.)	r (Laien)
Irr.	0,40	0,36	0,35	0,24*	0,48	0,24*
MWC	0,39	0,33*	0,33*	0,24*	0,49	0,25*
J2	0,43	0,43	0,41	0,17*	0,36	0,15*
J3	0,40	0,40	0,38	0,20*	0,39	0,18*
J5	0,43	0,43	0,41	0,18*	0,36	0,15*
J7	0,44	0,45	0,43	0,16*	0,35	0,13*
J11	**0,46**	**0,46**	**0,45**	0,13*	0,33*	0,09*
J15	**0,47**	**0,47**	**0,47**	0,11*	0,31*	0,07*
S2	0,45	0,41	0,38	0,15*	0,36	0,15*
S3	0,40	0,34	0,33*	0,20*	0,43	0,21*
S5	0,43	0,38	0,37	0,17*	0,40	0,18*
S7	0,46	0,42	0,39	0,14*	0,37	0,15*
S11	**0,49**	**0,45**	**0,42**	0,10*	0,34*	0,11*
S15	**0,50**	**0,48**	**0,44**	0,08*	0,30*	0,08*
GNE1	−0,07*	−0,07*	−0,06*	**−0,53**	**−0,69**	**−0,51**
GNE2	0,02*	0,02*	0,01*	**−0,59**	**−0,73**	**−0,55**
GNE3	0,04*	0,05*	−0,01*	**−0,58**	**−0,73**	**−0,51**
NNE1	0,45	0,39	0,38	0,14*	0,39	0,16*
NNE2	0,45	0,39	0,37	0,13*	0,38	0,15*
NNE3	0,21*	0,18*	0,20*	**0,41**	**0,59**	**0,38**
CHNR1	−0,28*	−0,22*	−0,23*	**−0,34**	**−0,58**	**−0,35**
CHNR2	−0,28*	−0,21*	−0,22*	**−0,34**	**−0,58**	**−0,35**
CHNR3	−0,15*	−0,10*	−0,15*	**−0,48**	**−0,66**	**−0,44**
Jita	0,21*	0,19*	0,22*	0,30*	0,45	0,25*
Jitt	0,21*	0,19*	0,19*	0,34*	0,52	0,31*
RAP	0,20*	0,19*	0,19*	0,35	0,52	0,32*
PPQ	0,21*	0,18*	0,19*	0,34*	0,51	0,31*
ShdB	0,38	0,37	0,32*	0,18*	0,39	0,19*
Shim	0,39	0,39	0,34*	0,21*	0,42	0,20*
APQ	0,40	0,41	0,35	0,19*	0,39	0,18*
NHR	0,32*	0,28*	0,29*	0,14*	0,34	0,13*
VTI	0,25*	0,22*	0,23*	0,16*	0,34*	0,14*

Tabelle 14.6.: Korrelationen ausgewählter Messgrößen mit den perzeptiven Beurteilungen Rauhigkeit und Behauchung. Hier wurde vor der Korrelationsberechnung der lineare Trend der Behauchung *und der Heiserkeit* aus der beurteilten Rauhigkeit und der lineare Trend der Rauhigkeit *und der Heiserkeit* aus der Behauchung herausgerechnet.(* bedeutet, dass die Korrelation nicht signifikant ist bei einem multiplen Signifikanzniveau von $p < 0,05$ für die Bonferoni-Holm Korrektur für 717 Tests.)

	Rauhigkeitsresiduum			Behauchungsresiduum		
	r (alle)	r (Expert.)	r (Laien)	r (alle)	r (Expert.)	r (Laien)
Irr.	0,27*	0,20*	0,21*	0,09*	0,19*	0,11*
MWC	0,24*	0,19*	0,19*	0,06*	0,22*	0,11*
J2	0,27*	0,26*	0,26*	0,03*	0,11*	0,05*
J3	0,25*	0,23*	0,24*	0,05*	0,13*	0,07*
J5	0,27*	0,25*	0,26*	0,03*	0,11*	0,05*
J7	0,28*	0,26*	0,27*	0,02*	0,10*	0,04*
J11	0,29*	0,28*	0,29*	0,00*	0,09*	0,02*
J15	0,30*	0,29*	0,30*	−0,01*	0,08*	0,00*
S2	0,31*	0,26*	0,24*	0,05*	0,14*	0,06*
S3	0,27*	0,21*	0,19*	0,07*	0,18*	0,08*
S5	0,30*	0,23*	0,23*	0,05*	0,15*	0,08*
S7	0,32*	0,26*	0,25*	0,05*	0,15*	0,06*
S11	**0,35**	0,30*	0,27*	0,03*	0,14*	0,03*
S15	**0,36**	0,33*	0,29*	0,02*	0,12*	0,01*
GNE1	−0,03*	0,06*	−0,01*	−0,27*	−0,30*	−0,32*
GNE2	0,04*	0,12*	0,02*	−0,30*	−0,33*	**−0,36**
GNE3	0,07*	0,15*	0,03*	−0,27*	−0,33*	−0,31*
NNE1	0,28*	0,24*	0,22*	0,00*	0,15*	0,05*
NNE2	0,28*	0,24*	0,22*	0,00*	0,15*	0,04*
NNE3	0,09*	0,07*	0,04*	0,15*	0,28*	0,15*
CHNR1	−0,17*	−0,10*	−0,13*	−0,13*	−0,27*	−0,19*
CHNR2	−0,16*	−0,10*	−0,11*	−0,13*	−0,27*	−0,19*
CHNR3	−0,06*	−0,01*	−0,04*	−0,20*	−0,33*	−0,22*
Jita	0,07*	0,05*	0,09*	0,05*	0,15*	0,09*
Jitt	0,17*	0,06*	0,15*	0,21*	0,21*	0,22*
RAP	0,17*	0,06*	0,14*	0,21*	0,21*	0,22*
PPQ	0,16*	0,05*	0,15*	0,20*	0,20*	0,22*
ShdB	0,24*	0,20*	0,19*	0,04*	0,12*	0,08*
Shim	0,24*	0,21*	0,20*	0,05*	0,13*	0,08*
APQ	0,26*	0,23*	0,20*	0,05*	0,12*	0,06*
NHR	0,22*	0,11*	0,22*	0,05*	0,07*	0,09*
VTI	0,19*	0,14*	0,19*	0,09*	0,17*	0,12*

15. Reduzierung des Aufnahmeumfanges?

Zur Berechnung des Heiserkeits-Diagramms ist die Aufnahme der Vokale $[\varepsilon:]_1$, [a:], [e:], [i:], [o:], [u:], $[\varepsilon:]_2$ jeweils in den Tonlagen normal (1), tief, hoch, normal (2), also von 28 Vokalen vorgesehen. Diese Protokoll deckt einen großen Teil der stimmhaften Artikulationsvarianten ab und verlangt vom Patienten außerdem eine Tonhöhenvariation, wie sie etwa in der Umgangssprache auftritt. Hierdurch ist es möglich ein umfassendes Bild der Stimmgebung zu erhalten. Am Heiserkeits-Diagramm einiger Patienten sieht man zum Teil deutliche Unterschiede, z.B. bei verschiedenen Tonhöhen. Hier kann es sein, dass eine Stimmbehinderung nur bei ganz bestimmten Anforderungen auftritt, oder umgekehrt, dass bestimmte Vokale oder Tonhöhen besonders gut gemeistert werden. Aus dieser Perspektive hat unsere Forschungsgruppe sehr gute Erfahrungen mit dem umfangreichen Protokoll gemacht. Darüber hinaus steigt die statistische Aussagekraft der Analyse, je mehr Stimmmaterial zur Berechnung hinzugezogen wird. Ausreißer machen sich so weniger störend bemerkbar.

Wenn ein solches Verfahren wie das Heiserkeits-Diagramm außerhalb der Forschung, etwa in phoniatrischen oder logopädischen Praxen, eingesetzt werden soll, so ist neben der Genauigkeit und Zuverlässigkeit der Messmethode die Zeit, die für die Analyse benötigt wird, ein wichtiger Faktor. Deshalb stellt sich die Frage, welche Auswirkung eine Reduzierung des Aufnahmeumfanges auf die Analyse der Stimmgüte mit dem Heiserkeits-Diagramm hat [81].

Um diese Frage zu beantworten, wird davon ausgegangen, dass der Mittelwert der Irregularitätskomponente und der Rauschkomponente, die aus allen 28 Vokalen berechnet wurden, die „richtigen“ Analysewerte sind. Nimmt man nun z.B. nur einen einzigen Vokal von den 28 Vokalen heraus, so wird sich der Mittelwert aller analysierten Segmente von diesem einen Vokal im Allgemeinen von dem Mittelwert über alle Vokale unterscheiden. Ein erster Ansatz zur Beurteilung der Qualität der Analyse mit nur einem Vokal könnte deshalb sein, von vielen Patienten den Mittelwert über alle Vokale sowie von dem untersuchten Vokal zu berechnen und dann als Fehlermaß z.B. den RMS-Wert der Differenzen zu betrachten.

Es wurde jedoch schon gezeigt, dass sich die einzelnen Vokale systematisch um den Mittelwert im Heiserkeits-Diagramm verteilen; z.B. würde ein [u:] im Durchschnitt höhere Rauschwerte und niedrigere Irregularitäten aufweisen. Diesen Trend sollte man herausrechnen, bevor man die Ergebnisse eines Vokals mit dem Mittelwert aller Vokale

vergleicht.

Da ein Vokal sicher nicht ausreicht, um die „richtigen“ Analysewerte zu liefern, sollen im Folgenden auch Kombinationen von zwei und drei Vokalen untersucht werden. Wir beschränken uns hier auf die Kombination von Vokalen bei normaler Tonhöhe. Erstens bleibt so die Anzahl der Kombinationsmöglichkeiten handhabbar und zweitens kann man bei allen Patienten die normale Tonhöhe aufnehmen. Einige Patienten haben große Schwierigkeiten einen tiefen oder hohen Ton zu erzeugen, teilweise wegen einer Stimmerkrankung, teilweise aber auch aus mangelnder Übung.

Um den Trend der Vokale herauszurechnen, gibt es viele Möglichkeiten. Die einfachste Möglichkeit ist, aus einer Linearkombination der Mittelwerte für einen, zwei oder drei Vokale den Mittelwert zu approximieren (mehrdimensionale lineare Regression). Der Vorteil ist, dass hier die Lösung bei einem gegebenen Satz von Aufnahmen eindeutig ist und sich geschlossen bestimmen lässt. Der Nachteil könnte der lineare Ansatz sein. Denn es ist zunächst nur eine Vermutung, dass sich der Mittelwert aller Vokale bei allen Tonhöhen linear aus nur wenigen Vokalen approximieren lässt. Deshalb wurde hier neben der Linearkombination eine Methode verwendet, die auch nichtlineare Zusammenhänge zwischen Eingangsgrößen und Ausgangsgrößen approximieren kann, nämlich ein neuronales Netz. Hier kommt ein Backpropagation-Netz zum Einsatz, dessen Topologie in Abbildung 10.1 bereits vorgestellt wurde. Auf die Auswahl der Netzwerkparameter wird im Folgenden Abschnitt eingegangen.

15.1. Die Netzwerkparameter

Die Netzwerkparameter müssen bei einem Backpropagation-Netzwerk dem jeweiligen Problem angepasst werden, es gibt keine allgemein gültigen Werte.

Dazu wurden Vorabtests mit der Vokalkombination [ɛ:], [e:], [o:] durchgeführt. Von 170 Patienten waren vollständige Aufnahmen mit 28 Vokalen und mindesten 2s Signal pro Vokal (entsprechend 7 Analysewerten) vorhanden. Die Analyseergebnisse der ersten zwei Sekunden (Irregularitätskomponente und Rauschkomponente) wurden für jeden Vokal gemittelt.

Von den 170 Patienten wurden jeweils folgende Mittelwerte für die Irregularitätskomponente und die Rauschkomponente berechnet:

- Die Mittelwerte über die ersten zwei Sekunden jedes normalen Vokals (6 pro Patient)
- die Mittelwerte über die erste Sekunde aller Vokale und der Tonlagen normal (1), tief, hoch
- die Mittelwerte über die zweite Sekunde aller Vokale und der Tonlagen normal (1), tief, hoch
- die Mittelwerte über alle Vokale der Tonhöhe normal (1)
- die Mittelwerte über alle Vokale und alle Tonhöhen (Referenzwert)

Die Wiederholung des Vokals [ɛ:] bei jeder Tonhöhe und die Vokalreihe normal (2) wurden nicht berücksichtigt, damit jeder Vokal und jede Tonhöhe gleich häufig vertreten war.

Um die Leistungsfähigkeit eines neuronalen Netzes zu beurteilen, muss ein Trainings- und eine Erkennungsdatensatz vorhanden sein. Deshalb wurde im Folgenden mit 100 Datensätzen trainiert und mit 70 Datensätzen die Erkennung und Erkennungsfehlerberechnung durchgeführt.

Damit nicht eine zufällige sonderbare Kombination von Trainings- und Datenmengen die Ergebnisse verfälscht, wurde stets mit mehreren zufälligen Kombinationen von 100 Trainings- und 70 Erkennungsdatensätzen gearbeitet. Bei den Voruntersuchungen zu den Netzwerkparametern waren dies 5 Zufallskombinationen, bei der späteren Analyse jeweils 50.

Das Backpropagation-Netzwerk bietet noch die Freiheit, die Aktivierungsfunktion bei der versteckten und bei der Ausgangsschicht zu wählen. Hier haben Vorversuche gezeigt, dass bei der versteckten Schicht eine sigmoide Aktivierungsfunktion und bei der Ausgabeschicht eine lineare Funktion sinnvoll ist.

Die Voruntersuchungen wurden wie gesagt mit je 5 zufälligen Kombinationen von Trainings- und Erkennungsdaten durchgeführt. Als Eingabe dienten stets die Mittelwerte pro Patient der Vokale [ɛ:], [e:], [o:]. Die zu lernende Ausgabe war der Mittelwert über alle Vokale und Tonhöhen. Der mittlere Fehler ist die durchschnittliche Differenz zwischen dem wirklichen und dem vom Netzwerk geschätzten Mittelwert über alle Vokale und Tonhöhen der Erkennungsdaten. Der Fehler wurde jeweils getrennt für die Irregularitäts- und die Rauschkomponente berechnet. Generell liegt der Fehler der Irregularitätskomponente über dem der Rauschkomponente. Da die Irregularitätskomponente auch einen etwa doppelt so großen Wertebereich besitzt, ist dies nicht verwunderlich.

In den folgenden Abbildungen wird stets ein Netzparameter variiert und alle anderen festgehalten. Da man am Anfang natürlich noch nicht weiß, welche Werte für die jeweils anderen Parameter am besten sind, sind in den Abbildungen nicht die ersten Tests gezeigt, sondern Tests, bei denen die festgehaltenen Parameter schon in der Nähe der optimalen Werte liegen. Die Aufzählung der fixierten Parameter wird deshalb hier unterlassen.

In Abbildung 15.1 sieht man, dass eine zu große Zahl versteckter Zellen dazu führt, dass das Netz die Trainingsdaten „auswendig lernt“ und nicht mehr in der Lage ist auf die Erkennungsdaten zu verallgemeinern: Bei mehr als zehn versteckten Zellen steigt der mittlere Fehler deutlich an. Im Folgenden wird daher mit 6 versteckten Zellen gearbeitet. In dem unteren Teil der Abbildung ist zu sehen, dass eine Erhöhung der Lernschritte über 10000 keine Verbesserung mehr bringt. Deshalb werden von nun an stets 10000 Lernschritte berechnet.

Die Gewichtsmatrizen des Netzwerks werden mit Zufallszahlen initialisiert, die in einem bestimmten Bereich gleichverteilt sind. Die Größe dieses Bereiches ist nicht sehr kritisch (Abbildung 15.2 oben). Im Folgenden wird 0,001 verwendet. Der Trägheitsparameter darf nicht zu klein gewählt werden (Abbildung 15.2 unten), hier wird der Wert 0,8 benutzt.

Für den Lernparameter wurde noch untersucht, ob bei einer größeren Anzahl der

Lernschritte vielleicht ein kleinerer Lernparameter zu insgesamt besseren Ergebnissen führt, als ein größerer Lernparameter bei weniger Lernschritten. Dass dies nicht so ist, ist in der Abbildung 15.3 zu erkennen.

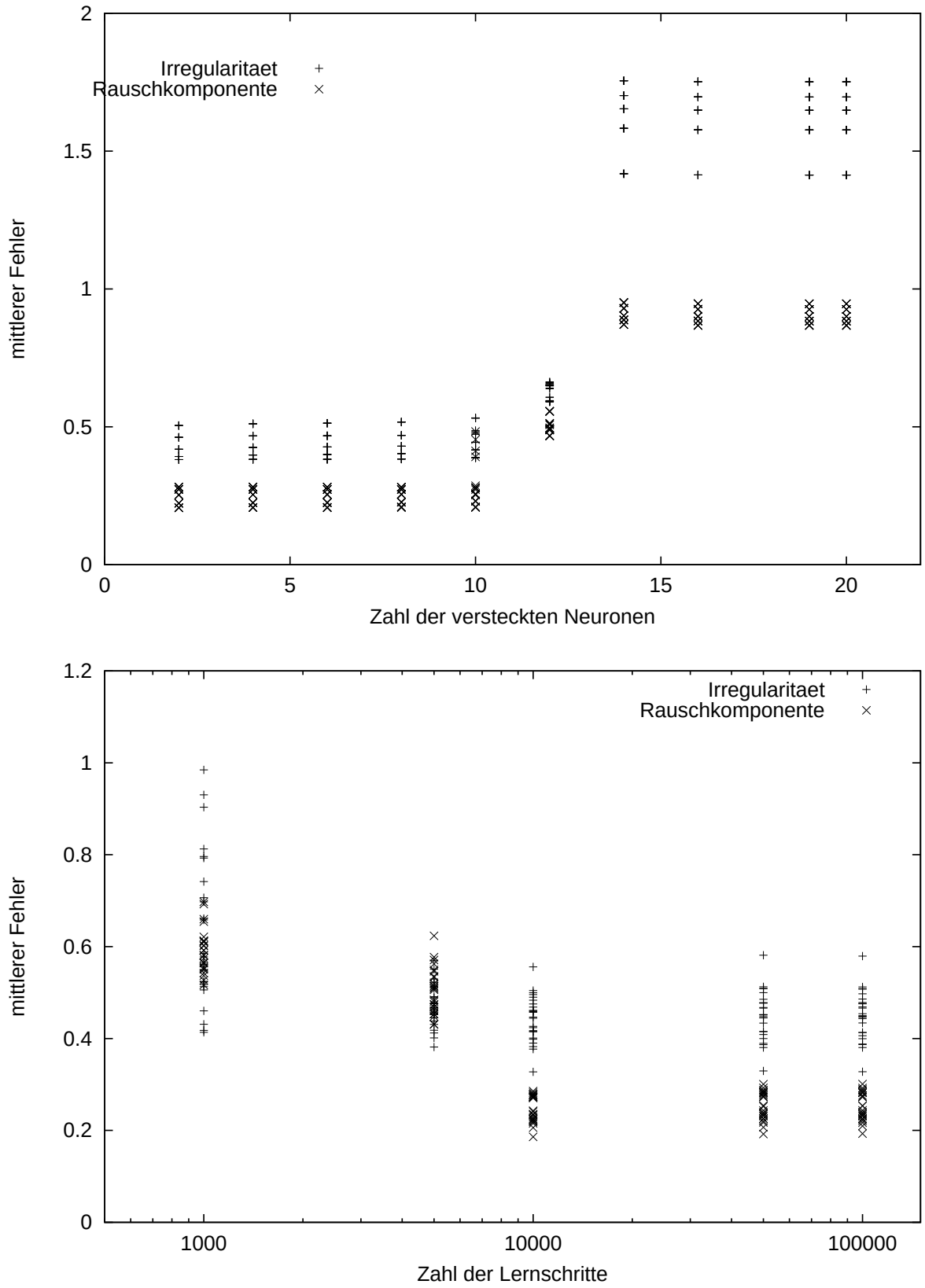


Abbildung 15.1.: Verschiedene Netzparameter

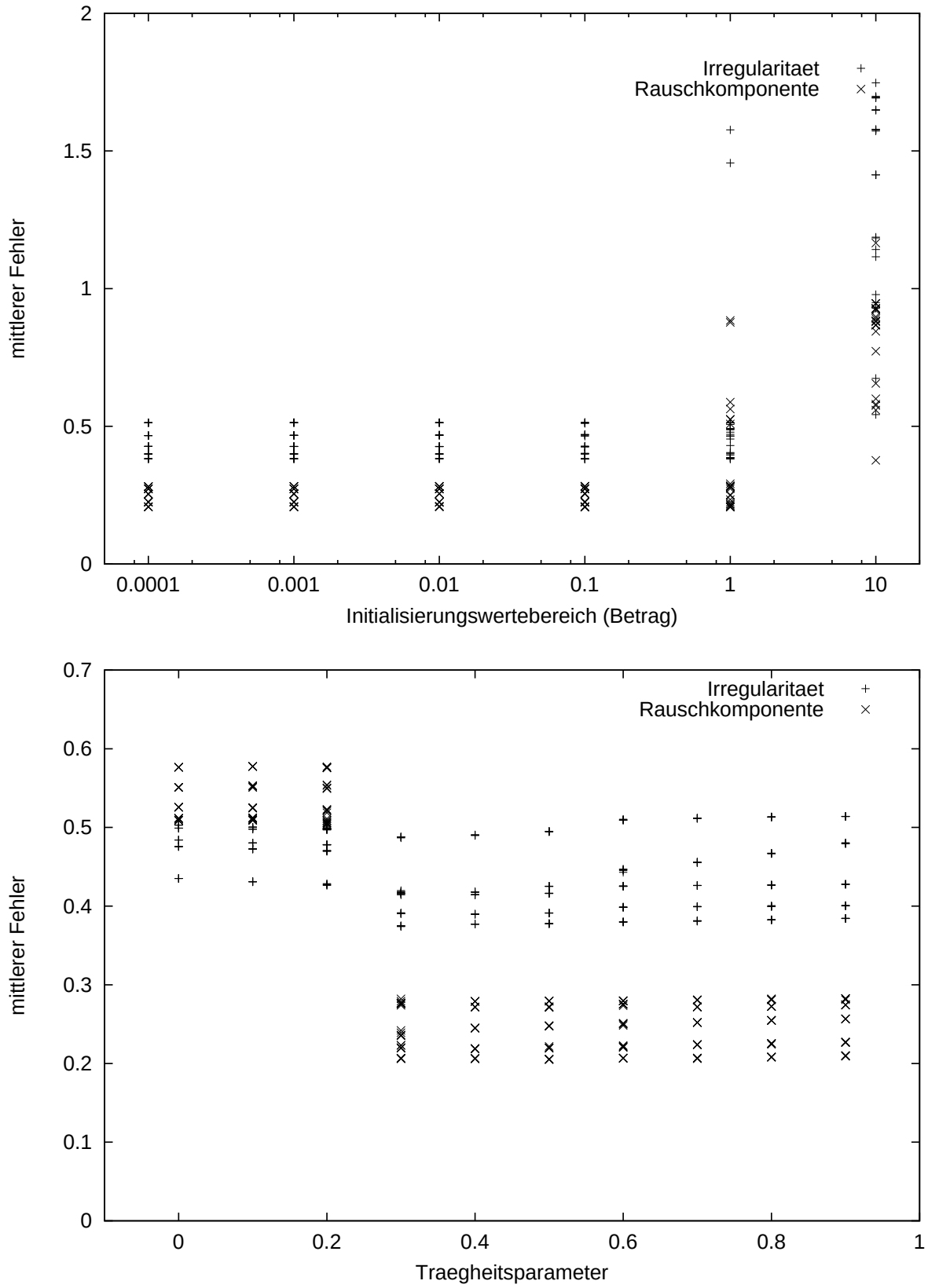


Abbildung 15.2.: Verschiedene Netzparameter

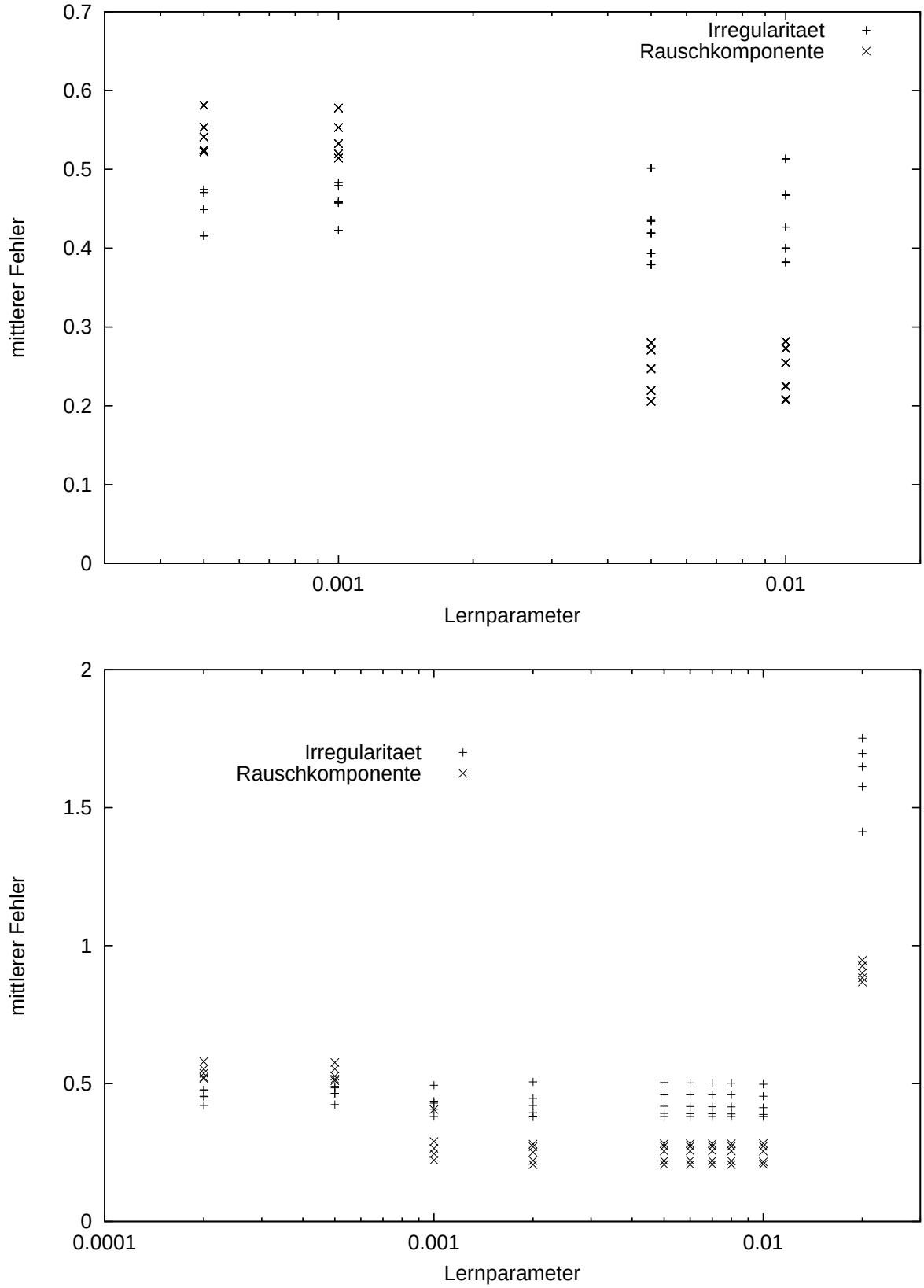


Abbildung 15.3.: Oben: 10000 Lernschritte, unten: 100000 Lernschritte

15.2. Vergleich von linearer Regression und neuronalem Netz

Um die Ergebnisse der linearer Regression und des neuronalen Netzes zu vergleichen, wurde auch bei der linearen Regression mit 50 zufälligen Kombinationen von Trainings- und Erkennungsmenge gearbeitet: Jeweils 100 Datensätze dienten zur Bestimmung der Koeffizienten der Regression. Mit den verbleibenden 70 Datensätzen wurde der Fehler berechnet.

Zum Vergleich wurde nun für jede Kombination von einem, zwei oder drei Vokalen der mittlere Fehler der linearen Regressionen (50) von dem mittleren Fehler der neuronalen Netze (ebenfalls 50) abgezogen. Diese Differenz ist in Abbildung 15.4 und in den Tabellen 15.1 und 15.2 dargestellt. Eine positive Differenz bedeutet, dass die lineare Regression einen kleineren Fehler liefert als das neuronale Netz. Mit vier Ausnahmen ist der Fehler der linearen Regression jeweils bei der Irregularitätskomponente und bei der Rauschkomponente geringer als bei dem neuronalen Netz. Gemessen an den Wertebereichen der Komponenten (etwa 4 Einheiten bei der Rauschkomponente und etwa 7 Einheiten bei der Irregularitätskomponente) und dem Betrag der mittleren Fehler (ca. 0,4 bis 0,6 bei der Irregularitätskomponente und ca. 0,2 bis 0,3 bei der Rauschkomponente) ist der Wert dieser Differenzen (ca. 0,02) beinahe zu vernachlässigen. Das neuronale Netz ist bei den einzelnen Vokale [o:] und [u:] und auch bei deren Kombination besser in der Vorhersage der Rauschkomponente.

Bei einigen Patienten zeigen diese Vokale Ausreißer gegenüber dem Mittelwert zu erhöhten Rauschwerten. Diese treten gerade dann auf, wenn die Patienten bei der Artikulation dieser Vokale die Lippen beinahe verschließen, so dass hier Rauschen entstehen kann. Bei anderen Patienten überwiegt auch bei diesen Vokalen die glottale Anregung. Das neuronale Netz kann diese zwei Artikulationsarten offensichtlich besser lernen und daraus auf den Mittelwert schließen als die lineare Regression.

Da aber alles in allem die lineare Regression kleinere Fehler zeigt und weil der Unterschied der Methoden relativ klein ist, werden die Ergebnisse nur noch für die lineare Regression besprochen.

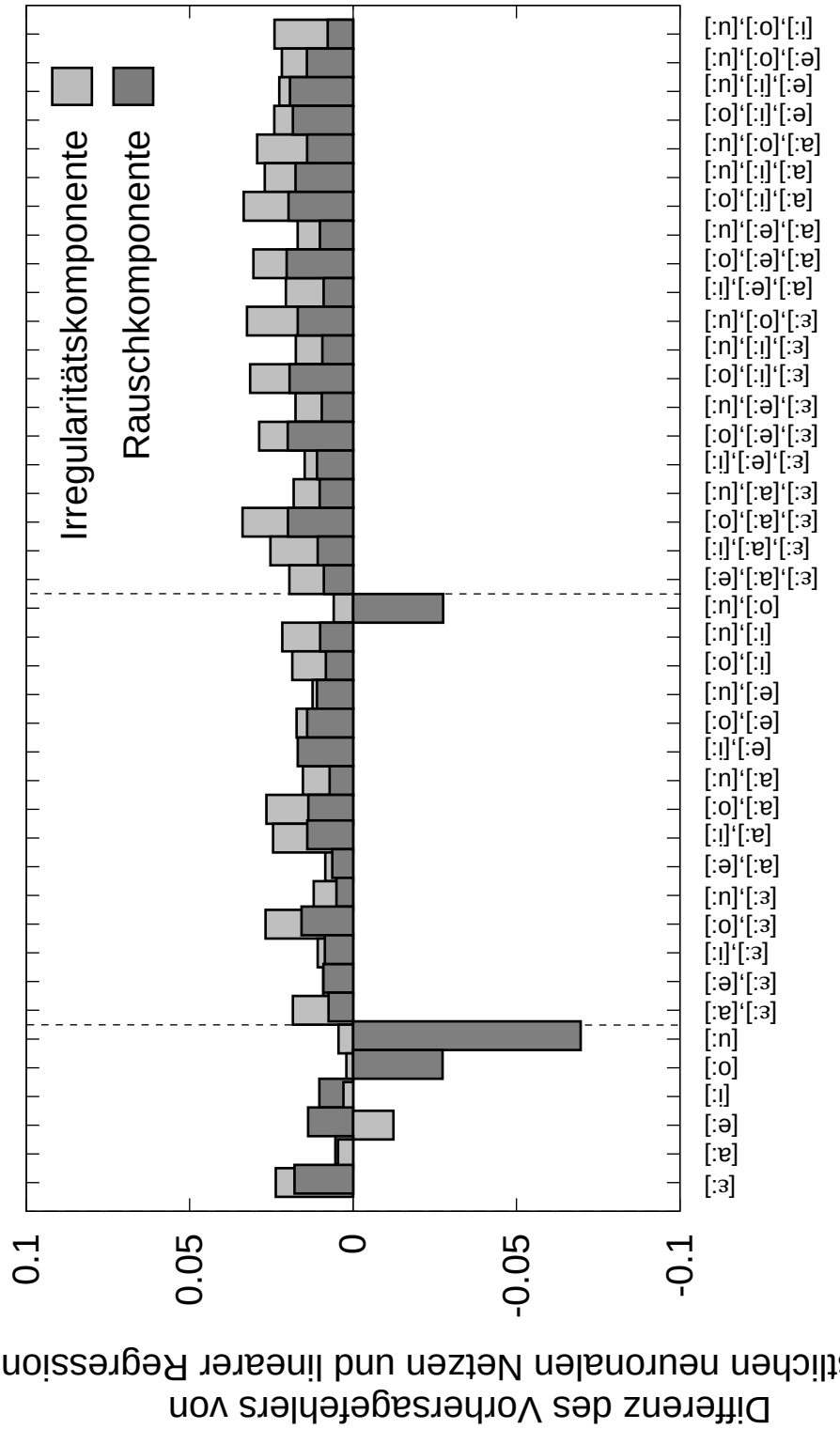


Abbildung 15.4.: Differenz der Vorhersagefehler durch lineare Regression und durch neuronale Netze

15.3. Ergebnisse der linearen Regression

Den besten Vokal und die beste Kombination von zwei und drei Vokalen kann man aus den Werten in den Tabellen 15.1 und 15.2 sowie den Abbildungen 15.5 und 15.6 ablesen. Wie zu erwarten, sinkt der Fehler mit zunehmender Anzahl der Vokale. Der Einzelvokal mit dem kleinsten Fehler ist das [o:]. Hier beträgt der RMS-Fehler 0,53 bei der Irregularitätskomponente und 0,38 bei der Rauschkomponente. 95% aller Fehlerbeträge sind kleiner als 1,1 bei der Irregularitätskomponente und 0,759 bei der Rauschkomponente.

Die Kombination von zwei Vokalen, die den kleinsten Fehler aufweist, ist [ɛ:] und[o:]. Hier beträgt der RMS-Fehler 0,44 bei der Irregularitätskomponente und 0,25 bei der Rauschkomponente. 95% aller Fehlerbeträge sind kleiner als 0,81 bei der Irregularitätskomponente und 0,52 bei der Rauschkomponente.

Die Fehler werden bei der Dreierkombination mit dem kleinsten Fehler nochmals geringer. Die Kombination [ɛ:], [i:] und [u:] hat zwar nicht den kleinsten RMS-Fehler bei der Irregularitätskomponente (0,424) und der Rauschkomponente (0,234), die Unterschiede zu den besseren Kombinationen sind aber minimal. Sie zeigt aber die kleinsten 95% Fehler mit 0,74 für die Irregularitätskomponente und 0,45 bei der Rauschkomponente. In der Anwendung ist der 95% Wert höher zu bewerten als der RMS-Wert, denn hier geht es auch um eine Minimierung der möglichen Ausreißer, die in dem 95% Wert besser erfasst werden.

Der Fehler verbessert sich nur noch unwesentlich, wenn statt der besten Dreierkombination alle 6 Vokale berücksichtigt werden: Der RMS-Wert liegt dann bei 0,35 für die Irregularitätskomponente und bei 0,20 für die Rauschkomponente. Die 95% Werte fallen auch nur noch leicht auf 0,71 bei der Irregularitätskomponente und 0,39 bei der Rauschkomponente.

Dagegen wird der Fehler noch einmal deutlich kleiner, wenn statt der 2 Sekunden nur die erste oder nur die zweite Sekunde aller Vokale ausgewertet wird. Ein Unterschied zwischen der ersten und der zweiten Sekunde ist nicht zu erkennen. Dies ist nicht trivial, denn im Allgemeinen ist der Stimme zu Beginn noch etwas ungleichmäßiger als zu einem späteren Zeitpunkt. Diese Unterschiede sind perzeptiv und akustisch bestimmbar (siehe de Krom 1994 und 1995 [18], [19]).

Tabelle 15.1.: Ergebnisse der Vorhersage von Irregularitäts- und Rauschkomponente von 6 Vokalen in je drei Tonlagen mit einem Vokal oder einer Kombination aus zwei Vokalen. Differenz: mittlerer Vorhersagefehler der linearen Regression subtrahiert von dem mittleren Vorhersagefehler der neuronalen Netze. RMS: Root Mean Square-Fehler der linearen Regression. 95%: Intervall, in dem 95% der Fehlerbeträge liegen.

	Differenz		RMS		95%	
Vokal(e)	Irreg. komp.	Rausch-komp.	Irreg. komp.	Rausch-komp.	Irreg. komp.	Rausch-komp.
[ɛ:]	0,024	0,018	0,651	0,336	1,442	0,723
[a:]	0,005	0,005	0,724	0,344	1,464	0,750
[e:]	−0,012	0,014	0,640	0,309	1,228	0,654
[i:]	0,003	0,010	0,684	0,323	1,410	0,636
[o:]	0,002	−0,027	0,525	0,381	**1,050**	**0,759**
[u:]	0,004	−0,070	0,734	0,587	1,488	1,013
[ɛ:] [a:]	0,018	0,007	0,610	0,302	1,203	0,612
[ɛ:] [e:]	0,009	0,009	0,538	0,279	1,196	0,625
[ɛ:] [i:]	0,011	0,009	0,537	0,260	1,153	0,523
[ɛ:] [o:]	0,027	0,016	0,442	0,250	**0,807**	**0,524**
[ɛ:] [u:]	0,012	0,005	0,466	0,293	0,857	0,600
[a:] [e:]	0,009	0,006	0,548	0,266	1,196	0,498
[a:] [i:]	0,025	0,014	0,532	0,244	1,161	0,498
[a:] [o:]	0,027	0,014	0,428	0,262	0,851	0,551
[a:] [u:]	0,015	0,007	0,503	0,300	1,128	0,573
[e:] [i:]	0,010	0,017	0,565	0,255	1,198	0,523
[e:] [o:]	0,017	0,014	0,448	0,250	0,874	0,587
[e:] [u:]	0,012	0,011	0,445	0,280	0,967	0,606
[i:] [o:]	0,019	0,008	0,457	0,270	0,846	0,484
[i:] [u:]	0,022	0,010	0,526	0,287	1,130	0,613
[o:] [u:]	0,006	−0,028	0,490	0,375	0,875	0,774

Tabelle 15.2.: Ergebnisse der Vorhersage von Irregularitäts und Rauschkomponente von 6 Vokalen in je drei Tonlagen aus einer Kombination aus drei Vokalen, aus der ersten Sekunde der 18 Vokale, aus der zweiten Sekunde und aus den 6 Vokalen in normaler Tonhöhe. Differenz: mittlerer Vorhersagefehler der linearen Regression subtrahiert von dem mittleren Vorhersagefehler der neuronalen Netze. RMS: Root Mean Square-Fehler der linearen Regression. 95%: Intervall in dem 95% der Fehlerbeträge liegen.

	Differenz		RMS		95%	
Vokal	Irreg. komp.	Rausch-komp.	Irreg. komp.	Rausch-komp.	Irreg. komp.	Rausch-komp.
[ɛ:] [a:] [e:]	0,020	0,009	0,519	0,261	1,132	0,553
[ɛ:] [a:] [i:]	0,025	0,011	0,506	0,237	1,053	0,481
[ɛ:] [a:] [o:]	0,034	0,020	0,416	0,236	0,827	0,485
[ɛ:] [a:] [u:]	0,018	0,010	0,444	0,266	0,962	0,515
[ɛ:] [e:] [i:]	0,015	0,011	0,501	0,241	1,134	0,526
[ɛ:] [e:] [o:]	0,029	0,020	0,415	0,228	0,834	0,572
[ɛ:] [e:] [u:]	0,018	0,010	0,393	0,252	0,883	0,553
[ɛ:] [i:] [o:]	0,031	0,019	0,414	0,222	0,818	0,454
[ɛ:] [i:] [u:]	0,018	0,009	0,424	0,234	**0,738**	**0,448**
[ɛ:] [o:] [u:]	0,032	0,017	0,402	0,246	0,771	0,523
[a:] [e:] [i:]	0,021	0,009	0,494	0,226	1,102	0,488
[a:] [e:] [o:]	0,031	0,020	0,402	0,223	0,892	0,474
[a:] [e:] [u:]	0,017	0,010	0,398	0,242	0,828	0,432
[a:] [i:] [o:]	0,033	0,020	0,394	0,216	0,790	0,406
[a:] [i:] [u:]	0,027	0,018	0,425	0,220	0,873	0,452
[a:] [o:] [u:]	0,029	0,014	0,398	0,256	0,860	0,554
[e:] [i:] [o:]	0,024	0,018	0,429	0,225	0,879	0,470
[e:] [i:] [u:]	0,023	0,019	0,427	0,234	0,882	0,514
[e:] [o:] [u:]	0,022	0,014	0,399	0,246	0,873	0,555
[i:] [o:] [u:]	0,024	0,008	0,431	0,262	0,913	0,500
1. Sekunde			0,152	0,081	0,295	0,172
2. Sekunde			0,135	0,072	0,279	0,159
[ɛ:] [a:] [e:] [i:] [o:] [u:]			0,354	0,199	0,707	0,393

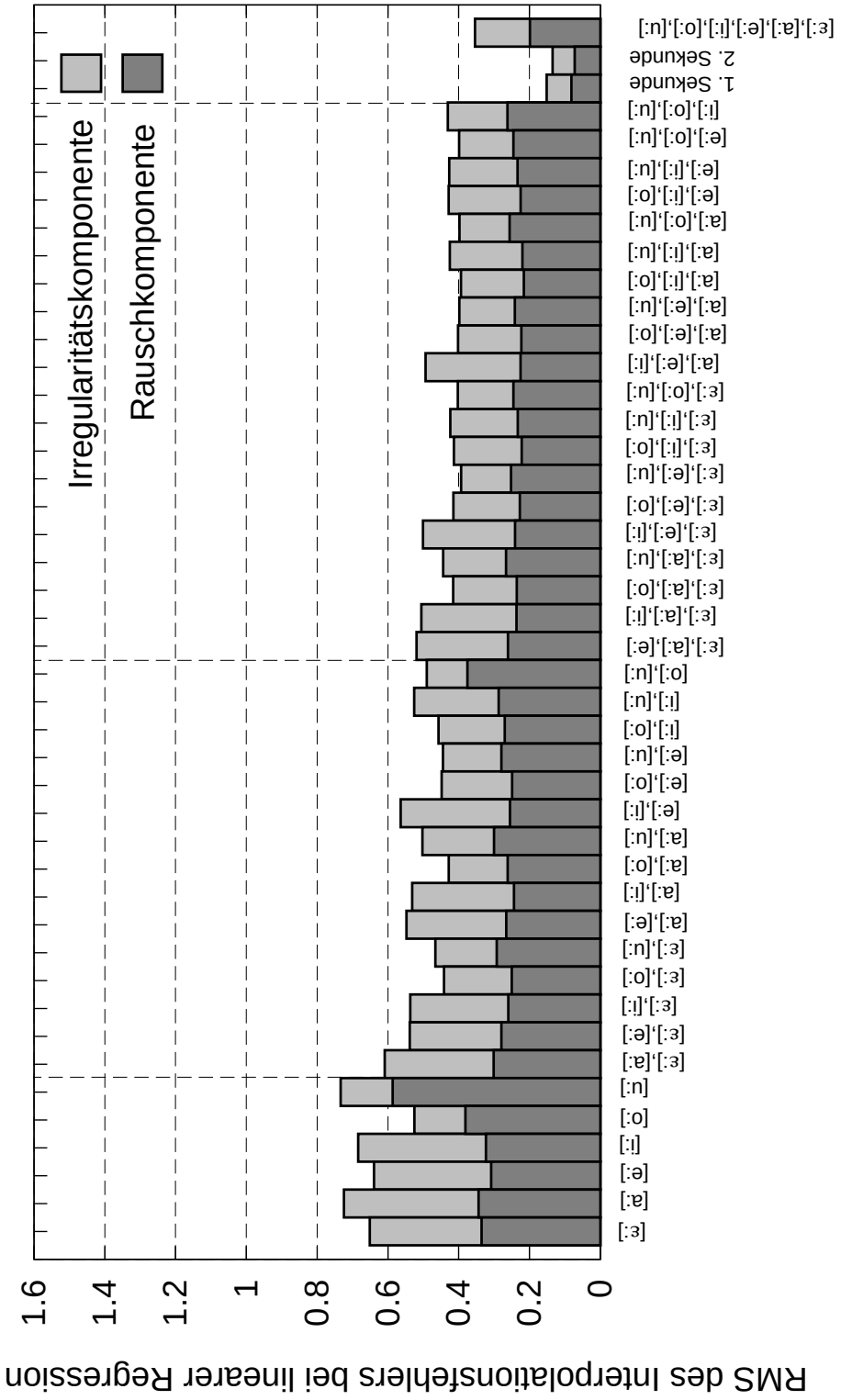


Abbildung 15.5.: Standardabweichung des Vorhersagefehlers bei linearer Regression

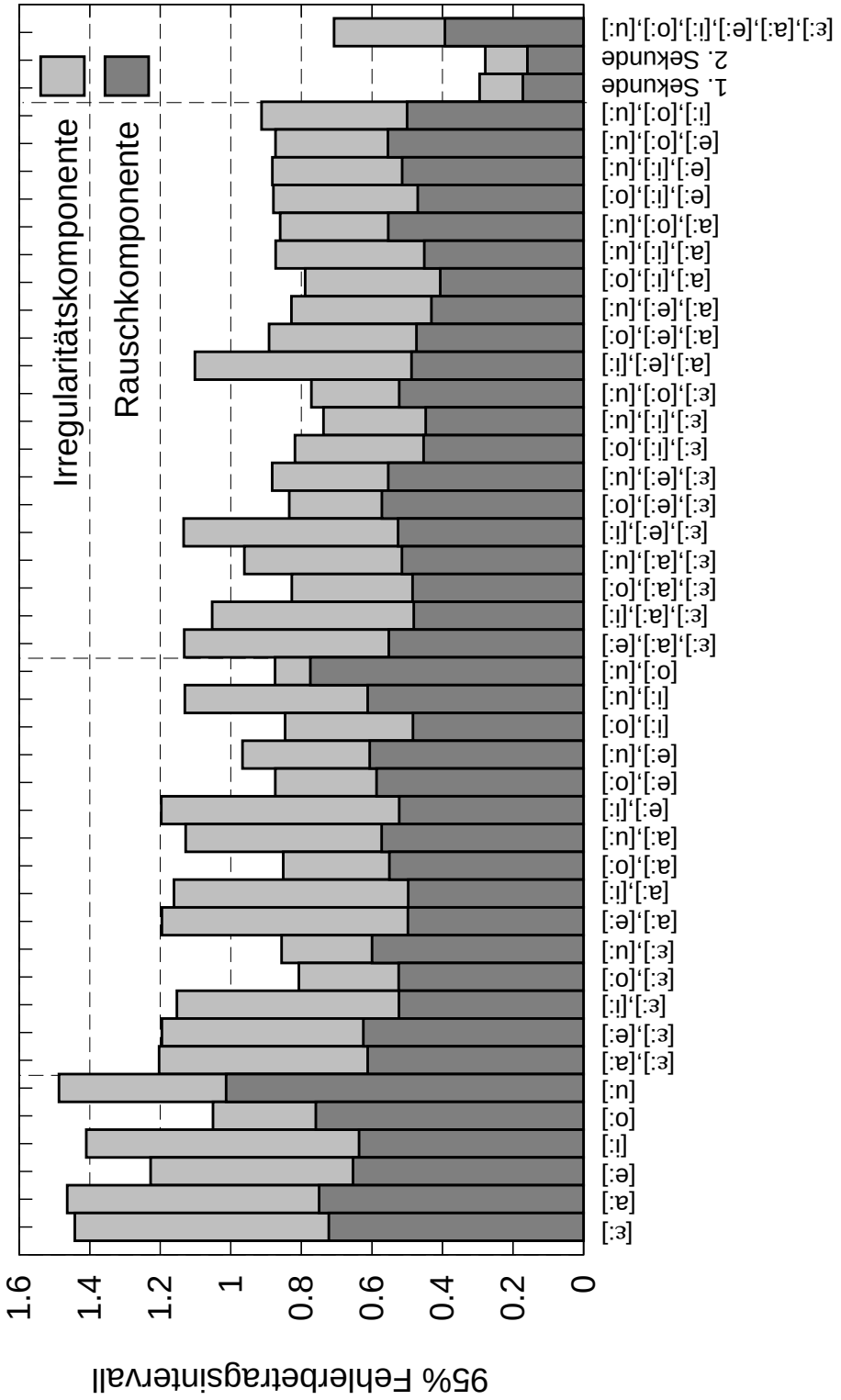


Abbildung 15.6.: Intervalle, in denen 95% der Vorhersagefehler bei linearer Regression liegen.

16. Patientenkatalog

Im Anhang D wird ein umfangreicher Katalog von Heiserkeits-Diagrammen beigefügt. Die Diagramme stammen von Patienten aus der Abteilung für Phoniatrie und Pädaudiologie der Universitätsklinik Göttingen. Bei den hier ausgewählten Beispielen waren jeweils mehrere Aufnahmen vorhanden, an denen sich ein zeitlicher Verlauf der Stimmgüteentwicklung ablesen lässt. Dieser Katalog wird beigefügt, weil sich bei den einzelnen Patienten viele interessante Einzelphänomene zeigen, die nicht sinnvoll in einer Gruppenanalyse zusammengefaßt werden können. Andererseits kann man durch die Fallbeispiele ein Gefühl für die Möglichkeiten und Grenzen der Beschreibung der Stimmgüte mit dem Heiserkeits-Diagramm bekommen. Der Katalog soll dazu beitragen, dem potentiellen Anwender des Heiserkeits-Diagramms Fallbeispiele vorzuführen, mit denen eigene Analyseergebnisse verglichen werden können.

An jedem Aufnahmetag wurden von dem Patienten die Vokale $[\varepsilon:]_1$, [a:], [e:], [i:], [o:], [u:] und $[\varepsilon:]_2$ in den Tonhöhen normal (1), tief, hoch und normal (2) aufgenommen. Zwischen „hoch“ und normal (2) wurde der Text „Nordwind und Sonne“ gelesen. Dieser Text stellt einerseits eine Stimmbelastung dar und wird andererseits auch zur Erforschung der akustischen Stimmgütemaßen bei gesprochener Sprache genutzt.

Die Heiserkeits-Diagramme enthalten rechts oben die Angaben: Patientennummer, Geschlecht des Patienten (m=männlich, w=weiblich) und das Geburtsjahr des Patienten. Die Rausch- und Irregularitätskomponenten werden jeweils in 0,5s langen Fenstern bei 0,25s Fenstervorschub für jeden Vokal und jede Tonhöhe nach Gleichung 7.10 berechnet. Die einzelnen Komponenten werden jeweils über alle analysierten Fenster gemittelt und die Standardabweichung berechnet. Für jeden Aufnahmetag wird eine Ellipse in das Heiserkeits-Diagramm gezeichnet, deren Mittelpunkt dem Mittelwert und deren Halbachsen der Standardabweichung der beiden Komponenten entspricht (so dass die Höhe und Breite der Ellipsen zwei Standardabweichungen darstellt). Neben dem Datum der Aufnahme ist z.B. mit $n = 288$ die gesamte Anzahl der analysierten Fenster für dieses Datum angegeben. Sehr kleine Werte deuten hier an, dass der Patient nicht in der Lage war, die Vokale lange zu halten oder nicht das gesammte Aufnahmeprotokoll leisten konnte. Die Farbe der Symbole neben dem Datum entspricht der Farbe der Ellipse. Neben jedem Datum sind noch die numerischen Werte der Rausch- und Irregularitätskomponente abgedruckt.

Der hellgrau schattierte Bereich links unten im Heiserkeits-Diagramm bezeichnet eine Fläche in der 95% aller Datenpunkte von 52 analysierten Normalstimmen gelegen haben. Der Bereich dient zur groben Einordnung der Analyseergebnisse.

17. Zusammenfassung und Ausblicke

Ein akustisches Maß zur Bestimmung des Rauschanteils bei der Stimmanregung der GNE (Glottal to Noise Excitation Ratio) wurde beschrieben. Das Maß beruht auf Korrelationen der Hilberteinhüllenden des inversgefilterten Mikrofonsignals in verschiedenen Frequenzbändern. Der GNE ist als Maximum der Korrelationswerte zwischen allen Einhüllenden der Kanäle definiert, deren Mittenfrequenzen mindestens eine halbe Frequenzbandbreite Abstand besitzen.

Der GNE wurde mit den akustischen Maßen NNE (Normalized Noise Energy) und CHNR (Cepstral Harmonics to Noise Ratio) verglichen. Alle Maße zeigten die gewünschte Fähigkeit, den Rauschanteil in einem Signal zu messen. Darüber hinaus wurde jedoch nur der GNE (bei 3000Hz Bandbreite) weder durch Jitter noch durch Shimmer beeinflusst. Die beiden anderen Maße zeigten erhebliche Jitterabhängigkeit. Schon bei Jitterwerten, die einer Normalstimme entsprechen, zeigten diese Maße die gleichen Werte wie bei 15-20dB Rauschanteil im Signal.

Die Beziehungen zwischen Perturbationsmaßen und Maßen für den Rauschanteil wurden durch eine Analyse der Messergebnisse an Stichproben pathologischer Stimmen untersucht. Unter anderem wurden Perturbationsmaße mit verschiedener lokaler Reichweite (2-15 Perioden) verglichen. Diese Maße sind untereinander hoch korreliert, so dass eine Beschränkung auf ein Jitter- und ein Shimmer-Maß sinnvoll erscheint.

Eine Vergleich der Informationsentropie zeigte, dass die beste Dreierkombination aus Jitter, Shimmer und dem mittleren Periodenkorrelationswert MWC (Mean Waveform Matching Coefficient) die Kombination {J3, S15, MWC} ist.

Die Berechnung der zusätzlichen normierten Information stellte unter den Rauschmaßen den GNE mit 3000Hz Bandbreite der Frequenzbandbreite als beste Ergänzung zu diesen drei Maßen heraus.

Eine Hauptachsentransformation ergab, dass der aus GNE3, J3, S15, MWC gebildete akustische Stimmgüteraum im Wesentlichen zweidimensional ist und in einem Diagramm (dem Heiserkeits-Diagramm) dargestellt werden kann. Im Vergleich zu den anderen Rauschmaßen hatten die Datenräume mit GNE als viertem Maß eine deutlich höhere Varianz in der zweiten Hauptrichtung. Das bedeutet, dass GNE weniger von Jitter, Shimmer und MWC (linear) abhängig ist als die anderen Rauschmaße.

Die Achsen des Heiserkeits-Diagramms, Irregularitätskomponente und Rauschkomponente, wurden definiert: Die Irregularitätskomponente besteht aus einer Summe der normalisierten Jitter-, Shimmer- und MWC-Werte. Die Rauschkomponente ist eine Lineartransformation des GNE.

Beim Heiserkeits-Diagramm erfolgt die Berechnung der Periodenlängen zur Bestimmung von Jitter und Shimmer mit dem Waveform-Matching-Verfahren. Am Beispiel von Flüsterstimmen wurde gezeigt, dass bei reinen Rauschsignalen stets sehr hohe Jitter- und Shimmerwerte gemessen werden. Die mittlere Periodenlängenkorrelation ist klein, so dass insgesamt sehr hohe Irregularitätskomponenten gemessen werden (die höchsten im Vergleich mit allen anderen pathologischen Stimmen, ca. 1500, die bisher untersucht wurden).

Im zweiten Teil wurde die offene Frage geklärt, wie sich das Waveform Matching Verfahren bei sehr kleinen und sehr großen Jitter- und Shimmerwerten verhält. Bei einer Abtastfrequenz von 44100Hz ist bei Grundfrequenzen bis ca. 600Hz die Genauigkeit des Verfahrens hoch genug, um praktisch vorkommende (kleinen) Jitter- und Shimmerwerte zu messen. Bei großen Jitter- und Shimmerwerten (größer als ca. 15%) treten wechselseitige Einflüsse von Jitter und Shimmer auf: Bei reinem Jitter über 15% wird erhöhter Shimmer gemessen, entsprechend für den Shimmer.

Messungen an synthetischen Signalen zeigten, dass sich durch den Vokaltrakt abhängig von der Grundfrequenz Jitter und Shimmer gegenseitig beeinflussen. Insbesondere wird durch Jitter im Glottissignal (grundfrequenzabhängig) Shimmer im Ausgangssignal induziert. Es konnte ein theoretisches Modell zur Beschreibung dieser Wechselwirkung hergeleitet werden, dass nur auf der Impulsantwort des Vokaltraktes und dem Zeitverlauf der Glottisfunktion beruht.

Es wurde festgestellt, dass die Vokale ein charakteristisches Muster im Heiserkeits-Diagramm bilden: die Vokale [u:] und [o:] liegen bei erhöhten Rauschkomponenten und verringerten Irregularitätskomponenten. [ε:] und [a:] zeigen demgegenüber einen höheren Irregularitätswert und eine kleinere Rauschkomponente. [e:] und [i:] ordnen sich dazwischen ein.

Die Analyse der Verteilungen von Gruppen mit spezifischen Pathologien zeigte, dass sich Gruppen mit verschiedenen Phonationsmechanismen im Heiserkeits-Diagramm signifikant voneinander und von Normalstimmen und aphonen Stimmen unterscheiden. Die Anordnung der Gruppen im Heiserkeits-Diagramm entsprach dabei ausnahmslos den Erwartungen aufgrund der Schwingungseigenschaften der entsprechenden Gewebestrukturen bei der Stimmgebung: Die Gruppen lagen im Heiserkeits-Diagramm umso näher an der Gruppe der Normalstimmen, je eher die Stimmgebung der normalen physiologischen Stimmgebung entsprach.

Korrelationen mit perzeptiven Maßen (Behauchung und Rauhigkeit) zeigten, dass GNE spezifisch mit Behauchung und Jitter und Shimmer (mit Periodenlängenberechnung nach dem Waveform-Matching-Verfahren) spezifisch mit Rauhigkeit korreliert sind. Im Gegensatz dazu wurde gezeigt, dass die Jitter-, Shimmer- und Rauschmaße der kommerziellen MDVP Software nicht spezifisch sind.

Im Anhang werden die Heiserkeits-Diagramme von 48 Patienten gezeigt, in denen die Stimmgüteentwicklung dokumentiert wird. Hier werden unter anderem Therapieerfolge und die Auswirkungen von Operationen oder Lähmungen auf die Stimmgütedarstellung sichtbar. Die Tatsache, dass die Kliniker bisher (nach ca. 1500 Auswertungen) mit der Darstellung der Stimmgüte durch das Heiserkeits-Diagramm zufrieden sind und dass es zu keinen Widersprüchen zwischen der akustischen Stimmgütebeschreibung mit dem

Heiserkeits-Diagramm und der Einschätzung der Kliniker gekommen ist, ist für den Autor eine wichtige Bestätigung der Tauglichkeit der hier entwickelten Methode.

Zu den Perspektiven des Heiserkeits-Diagramms ist anzumerken, dass mittlerweile eine Version des Heiserkeits-Diagramms kurz vor der Fertigstellung ist, die am Oldenburger Hörzentrum implementiert wird und auf einem modernen Windows-PC lauffähig ist. Gespräche mit einer schwedischen Softwarefirma (Nyala Software) lassen auf eine baldige Integration des Heiserkeits-Diagramm in das professionelle Signalverarbeitungspaket „Swell“ hoffen. In der Diplomarbeit von Jan Lessing wurde bereits die prinzipielle Übertragbarkeit des Heiserkeits-Diagramms von Vokalen auf fließende Sprache nachgewiesen, so dass in naher Zukunft auch die Güte der Sprechstimme der Patienten mit akustischen Methoden bewertet werden kann [**?**, **?**].

A. Rekursiver Filter zweiter Ordnung

A.1. Definition in der z-Ebene

Die Übertragungsfunktion des rekursiven Filters zweiter Ordnung sei wie folgt definiert

$$H(z) = \frac{r\sin(\gamma)z}{z^2 - 2r\cos(\gamma)z + r^2}\,. \tag{A.1}$$

Die einzige Nullstelle ist bei $z_0 = 0$. Die beiden Polstellen $z_{\infty_{1/2}}$ liegen bei:

$$\begin{aligned} z_{\infty_{1/2}} &= r\cos\gamma \pm \sqrt{r^2\cos^2\gamma - r^2} \\ &= re^{\pm i\gamma} \end{aligned}$$

Damit folgt für den Betrag der Übertragungsfunktion:

$$\begin{aligned} |H(z)| &= \frac{r\sin(\gamma)|z|}{|z - z_{\infty_1}| \cdot |z - z_{\infty_2}|} \\ |H(z = e^{i\omega\tau})| = |H(\omega)| &= \frac{r\sin\gamma}{|e^{i\omega\tau} - re^{i\gamma}| \cdot |e^{i\omega\tau} - re^{-i\gamma}|} =: \frac{r\sin\gamma}{|N(\omega)|} \end{aligned}$$

,

wobei die Abtastperiode $\tau = \frac{1}{f_s}$ der Kehrwert der Abtastfrequenz f_s ist. Das Quadrat des Nenners $N(\omega)$ läßt sich wie folgt in ein Polynom zweiten Grades in $\cos(\omega\tau)$ auflösen ($t = \omega\tau$ im Folgenden):

$$\begin{aligned} |N(t)|^2 &= |e^{i\omega\tau} - re^{i\gamma}|^2 \cdot |e^{i\omega\tau} - re^{-i\gamma}|^2 \\ &= [4r^2]\cos^2 t - [4r(1+r^2)\cos\gamma]\cos t + [(1+r^2)^2 - 4r^2\sin^2\gamma]. \end{aligned}$$

Damit ist

$$|H(\omega)| = \frac{r\sin\gamma}{\sqrt{[4r^2]\cos^2\omega\tau - [4r(1+r^2)\cos\gamma]\cos\omega\tau + [(1+r^2)^2 - 4r^2\sin^2\gamma]}}$$

Die Resonanzstelle liegt dort, wo $|H(\omega)|$ maximal, d.h. der Nenner minimal wird. Die notwendige Bedingung ist, dass die Ableitung des Terms in der Wurzel (im Folgenden $N(y)$) des Nenners verschwindet. Dazu wird im Nenner folgende Variablensubstitution

vorgenommen: $\cos(2\pi f\tau) = \cos(\omega\tau) = y$. (Rückwärts: $f = \frac{1}{2\pi\tau}\arccos(y)$.) Wegen des zyklischen Verhaltens der Variablen $z = e^{i\omega\tau}$ und der Spiegelsymmetrie zur reellen Achse reicht die Betrachtung auf dem Intervall $\omega\tau \in [0, \pi]$ entsprechend $y \in [-1, 1]$ aus.

$$\begin{aligned} N(y) &= [4r^2]y^2 - [4r(1+r^2)\cos\gamma]y + [(1+r^2)^2 - 4r^2\sin^2\gamma] \\ \left.\frac{dN(y)}{dy}\right|_{y=y_R} &= [8r^2]y - [4r(1+r^2)\cos\gamma] = 0 \Leftrightarrow \\ y_R &= \frac{(1+r^2)\cos\gamma}{2r} \end{aligned}$$

so dass die Resonanzfrequenz $f_R = \frac{1}{2\pi\tau}\arccos y_R$ mit $\gamma = 2\pi f_\gamma \tau$

$$\boxed{f_R = \frac{1}{2\pi\tau}\arccos(\frac{(1+r^2)\cos 2\pi f_\gamma\tau}{2r})} \tag{A.2}$$

ist.

Die quadrierte Übertragungsfunktion an der Stelle y_R ist:

$$\begin{aligned} |H(y_R)|^2 &= \frac{r^2\sin^2\gamma}{[4r^2]\frac{(1+r^2)^2\cos^2\gamma}{4r^2} - [4r(1+r^2)\cos\gamma]\frac{(1+r^2)\cos\gamma}{2r} + [(1+r^2)^2 - 4r^2\sin^2\gamma]} \\ &= \frac{r^2}{(1-r^2)^2} \end{aligned}$$

Also ist

$$\boxed{|H(f_R)| = \frac{r}{1-r^2}}. \tag{A.3}$$

Die Bandbreite des Resonanzfilters ergibt sich aus der Forderung, dass bei den Grenzfrequenzen $f_{G_{1/2}}$ die Leistung $|H(f_{G_{1/2}})|^2$ gleich der Hälfte der Leistung im Resonanzmaximum $\frac{1}{2}|H(f_R)|^2$ ist. Die Bandbreite ist dann die Differenz der Grenzfrequenzen.

$$\begin{aligned} |H(f_{G_{1/2}})|^2| &= \frac{r^2\sin^2\gamma}{[4r^2]y_G^2 - [4r(1+r^2)\cos\gamma]y_G + [(1+r^2)^2 - 4r^2\sin^2\gamma]} \\ &= \frac{r^2}{2(1-r^2)^2} \end{aligned}$$

oder

$$\begin{aligned} &4r^2y_G^2 - 4r(1+r^2)\cos\gamma y_G + (1+r^2)^2 - 4r^2\sin^2\gamma = 2(1-r^2)^2\sin^2\gamma \\ &4r^2y_G^2 - 4r(1+r^2)\cos\gamma y_G + (1+r^2)^2 - 2[(1-r^2)^2 + 2r^2]\sin^2\gamma = 0 \\ &y_G^2 - \frac{(1+r^2)\cos\gamma}{r}y_G + \frac{(1+r^2)^2 - 2(1+r^4)\sin^2\gamma}{4r^2} = 0 \end{aligned}$$

$$
\begin{aligned}
y_{G_{1/2}} &= \frac{(1+r^2)\cos\gamma}{2r} \pm \sqrt{\frac{(1+r^2)^2\cos^2\gamma}{4r^2} - \frac{(1+r^2)^2 + 2(1+r^4)\sin^2\gamma}{4r^2}} \\
&= y_R \pm \frac{(1-r^2)\sin\gamma}{2r}
\end{aligned}
$$

$$
\boxed{f_{G_{1/2}} = \frac{1}{2\pi\tau}\arccos(y_{G_{1/2}})} \tag{A.4}
$$

In der Abbildung A.1 ist die Resonanzfrequenz f_R in Abhängigkeit von der Polfrequenz f_γ für den Bereich 0 bis 1000Hz bei einer Abtastfrequenz $f_s = 44100$Hz dargestellt.

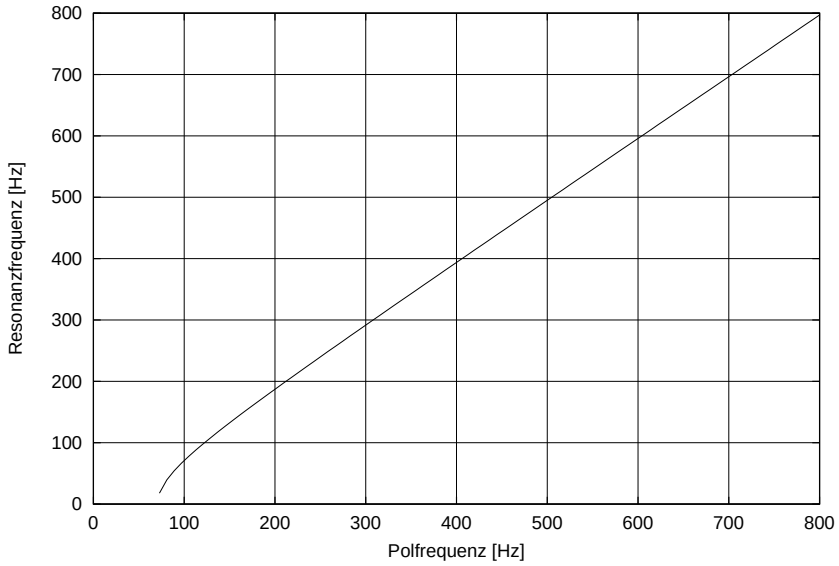


Abbildung A.1.: Resonanzfrequenz f_R in Abhängigkeit von der Polfrequenz f_γ

B. Hilberteinhüllende von zwei aufeinander folgenden Deltafunktionen

Zwei Deltafunktionen

$$f(t) = \delta(t + \frac{\Delta t}{2}) + \delta(t - \frac{\Delta t}{2}) \tag{B.1}$$

werden im Frequenzbereich mit zwei Hanning-Fenstern $h_a(\omega)$ der Breite $2a$ bei b und $-b$, $a \leq b$, bandpassgefiltert.

Die Hilberteinhüllende, also der Betrag des analytischen Signals ergibt sich zu

$$|\sigma(t)| = \sqrt{H_a(t - \frac{\Delta t}{2})^2 + H_a(t + \frac{\Delta t}{2})^2 - 2H_a(t - \frac{\Delta t}{2})H_a(t + \frac{\Delta t}{2}) \sin^2 b\frac{\Delta t}{2}} \tag{B.2}$$

mit $H_a(t) \equiv \frac{\pi \sin at}{2t(\pi^2 - a^2t^2)}$.

Die Abb. B.1 zeigt die Hilberteinhüllenden von einem Kanal bei $b_1 = 2\pi * 500\text{Hz}$ und einem bei $b_2 = 2\pi * 1000\text{Hz}$. Die Bandbreite beträgt $a = 2\pi * 500\text{Hz}$. Die beiden Deltapulse liegen $\Delta t = 1\text{ms}$ auseinander. Die Parameter sind so gewählt, dass der $\sin^2$ aus (B.2) bei b_1 gerade 1 und bei b_2 0 ist. Man erkennt, dass die Einhüllenden völlig verschieden aussehen, deshalb ist die Korrelation zwischen den Einhüllenden klein. Dies ist der ungünstigste Fall, der eintreten kann, wenn die Fensterbreite zu klein gewählt wurde, so dass die Pulse zu schnell aufeinander folgen. Das führt dazu, dass die Korrelation sinkt, obwohl es sich um eine Pulsanregung handelt.

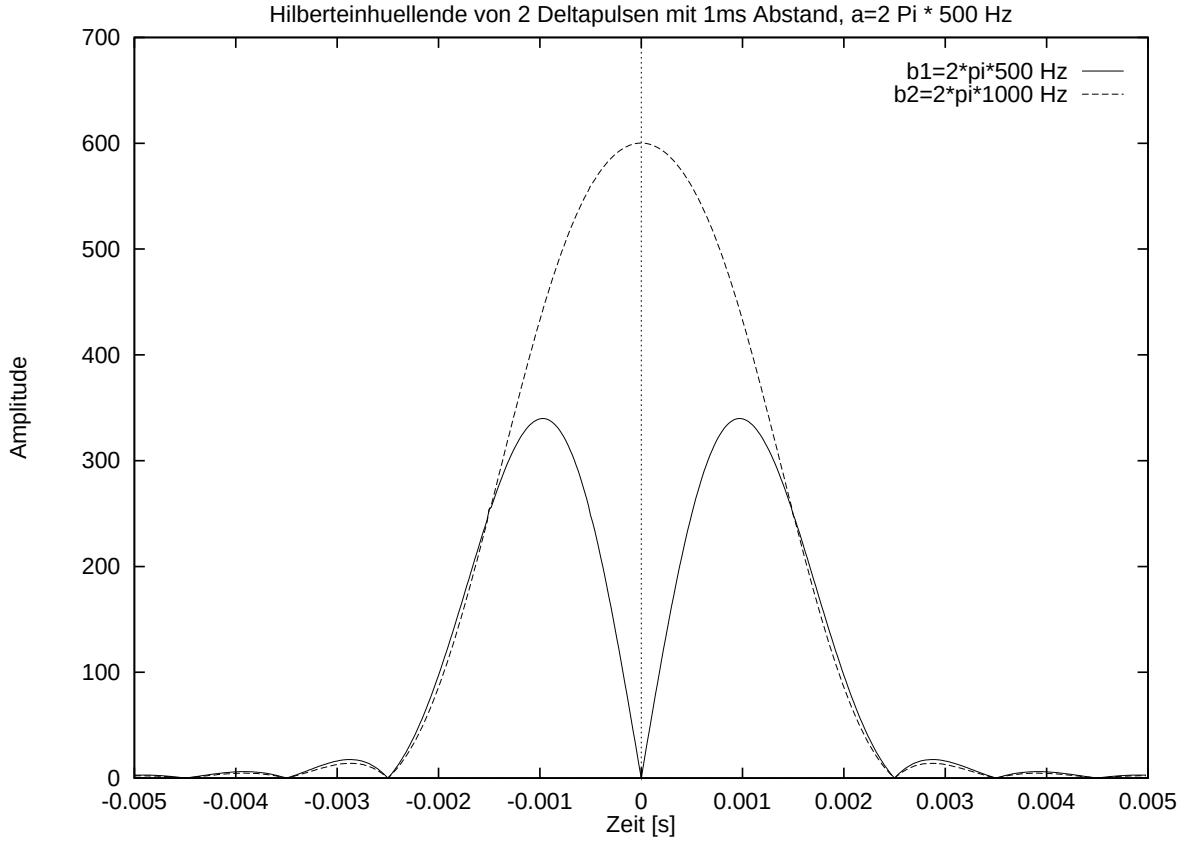


Abbildung B.1.: Bei zwei dicht aufeinander folgenden Deltapulsen sinkt die Korrelation durch Interferenzeffekte.

C. Spektrale Konsequenzen des Shimmers bei endlichen, diskreten Signalen

Die DFT, wie sie hier benutzt wird, stammt aus Stearns (S. 95). Die diskrete Fourier-Transformation (DFT) einer Folge mit N Werten f_n^N ist wie folgt definiert:

$$F_m^N = \sum_{n=0}^{N-1} f_n^N d^N(m,n), \quad m = 0, 1, \ldots, N-1 \tag{C.1}$$

mit

$$d^N(m,n) = \frac{1}{\sqrt{N}} e^{-i2\pi \frac{mn}{N}}, \quad m, n = 0, 1, \ldots, N-1. \tag{C.2}$$

Entsprechend die diskrete Fourier-Rücktranformation (IDFT):

$$f_n^N = \sum_{n=0}^{N-1} F_m^N D^N(m,n), \quad n = 0, 1, \ldots, N-1 \tag{C.3}$$

mit

$$D^N(m,n) = \frac{1}{\sqrt{N}} e^{i2\pi \frac{mn}{N}}, \quad m, n = 0, 1, \ldots, N-1. \tag{C.4}$$

C.1. Signal der Periodenlänge $T = N/M$

Im Folgenden sei eine Folge g_n^T mit der Länge $T = N/M$, wobei $T, N, M \in \mathbb{N}$ sind. Das Signal f_n^N enthalte M Wiederholungen der Folge g_n^T. Aus der Definition von $d^N(m,n)$ nach (C.2) ergibt sich folgende Beziehung zwischen den Matrixelementen einer N- und einer T-dimensionalen DFT:

$$d^T(m,n) = \frac{1}{\sqrt{T}} e^{-i2\pi \frac{mn}{T}} = \sqrt{\frac{M}{N}} e^{-i2\pi \frac{mnM}{N}} = \sqrt{M} d^N(Mm, n). \tag{C.5}$$

Damit folgt die diskrete Fouriertransformierte F_m^N an der Stelle $m = kM$ aus der Fouriertransformierten G_k^T von g_n^T:

$$\begin{aligned} F_{m=kM}^N &= \sum_{n=0}^{N-1} f_n^N d^N(kM,n) = M \sum_{n=0}^{T-1} f_n^N d^N(kM,n) \\ &= M \sum_{n=0}^{T-1} f_n^N d^T(k,n) \frac{1}{\sqrt{M}} = \sqrt{M} G_k^T. \end{aligned} \tag{C.6}$$

Sei oBdA $\sum_{n=0}^{T-1} |g_n^T|^2 = 1$ und damit auch (ohne Beweis) $\sum_{m=0}^{T-1} |G_m^T|^2 = 1$. Dann folgt für die spektrale Energie an den Stellen $m = kM$, $k \in \mathbb{N}$:

$$\sum_{k=0}^{T-1} |F_{kM}^N|^2 = \sum_{k=0}^{T-1} M |G_k^T|^2 = M. \tag{C.7}$$

Da aber auch

$$\sum_{n=0}^{N-1} |f_n^N|^2 = M \sum_{n=0}^{T-1} |g_n^T|^2 = M = \sum_{m=0}^{N-1} |F_m^N|^2 \tag{C.8}$$

ist, folgt für $m \neq kM$:

$$\sum_{m=0; m \neq kM}^{N-1} |F_m^N|^2 = 0 \tag{C.9}$$

und damit

$$F_m^N = 0 \quad , m \neq kM. \tag{C.10}$$

C.2. Diskretes Rechteckfenster

Ein diskretes Rechteckfenster, das aus insgesamt N Samples besteht und bei dem die ersten T Samples den Wert 1 haben, sei wie folgt definiert:

$$\boxed{r_n^{T,N} = \begin{cases} 1 & : \quad 0 \leq n < T \\ 0 & : \quad T \leq n < N \end{cases} , \quad n = 0, 1, \ldots, N-1.} \tag{C.11}$$

Die diskrete Fouriertransformierte $R_m^{T,N}$ ergibt sich unter Verwendung der geometrischen Reihe:

$$R_m^{T,N} = \sum_{n=0}^{N-1} r_n^{T,N} \frac{1}{\sqrt{N}} e^{-i2\pi \frac{mn}{N}} \tag{C.12}$$

$$= \frac{1}{\sqrt{N}} \sum_{n=0}^{T-1} (e^{-i2\pi \frac{m}{N}})^n \tag{C.13}$$

$$R_m^{T,N} = \begin{cases} \frac{1}{\sqrt{N}} \frac{1-e^{-i2\pi \frac{mT}{N}}}{1-e^{-i2\pi \frac{m}{N}}} & : \quad m = 1, 2, \ldots, N-1 \\ \frac{T}{\sqrt{N}} & : \quad m = 0 \end{cases} \tag{C.14}$$

Das Betragsquadrat der DFT folgt sogleich, wenn man die Gleichung

$$\begin{aligned} |1 - e^{i\varphi}|^2 &= |1 - \cos\varphi - i\sin\varphi|^2 && \text{(C.15)} \\ &= (1 - \cos\varphi)^2 + \sin^2\varphi && \text{(C.16)} \\ &= 1 - 2\cos\varphi + \cos^2\varphi + \sin^2\varphi && \text{(C.17)} \\ &= 2 - 2\cos\varphi && \text{(C.18)} \end{aligned}$$

benutzt:

$$|R_m^{T,N}|^2 = \begin{cases} \frac{1}{N} \frac{1-\cos(2\pi \frac{mT}{N})}{1-\cos(2\pi \frac{m}{N})} & : \quad m = 1, 2, \ldots, N-1 \\ \frac{T^2}{N} & : \quad m = 0 \end{cases} \tag{C.19}$$

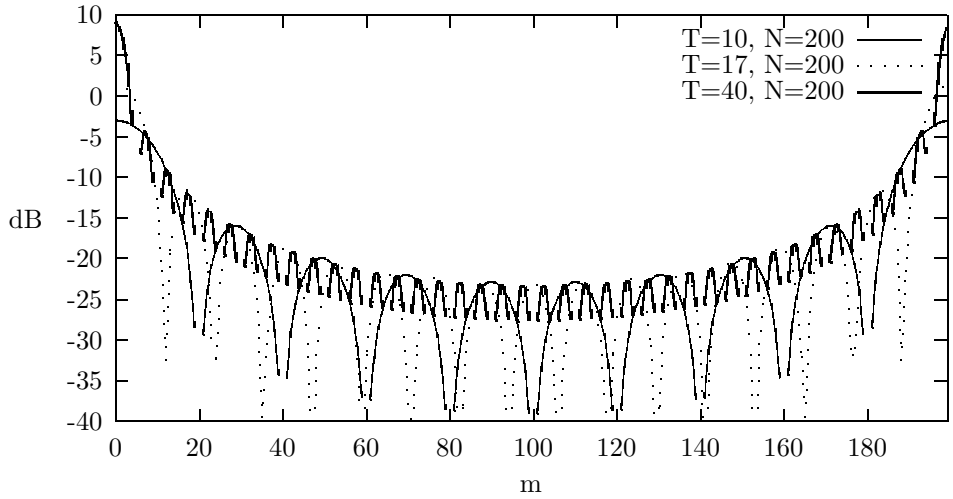


Abbildung C.1.: Spektrum (DFT) eines Rechteckfensters von 0 bis $T-1$.

C.3. Shimmer in einem diskreten, periodischen Signal

Wir betrachten ein Signal f_n^N der Länge N, bei dem sich das Signal g_n^T der Länge T genau M mal wiederholt. Dieses Signal soll eine zufällige Amplitudenvariation (Shimmer) S enthalten. Die Variation wird dadurch erreicht, daß jede Periode mit $1+S\theta_k^M$ multipliziert wird, wobei θ_k^M eine z.B. gleich- oder normalverteilte Folge von M Zufallszahlen ist (im

Bereich [-1,1] bzw. mit der Standardabweichung 1). Das resultierende Signal h_n^N lässt sich wie folgt schreiben:

$$h_n^N = f_n^N[1 + S\sum_{k=0}^{M-1}\theta_k^M\delta(n-kT) * r_n^{T,N}] \tag{C.20}$$

; dabei steht '$*$' für die diskrete Faltung. Mit der Definition:

$$x_n^N = \sum_{k=0}^{M-1}\theta_k^M\delta(n-kT) \tag{C.21}$$

und der Bezeichnung X_m^N für die DFT von x_n^N folgt die DFT von h_n^N:

$$H_m^N = F_m^N + SF_m^N * (X_m^N \cdot R_m^{T,N}). \tag{C.22}$$

Die DFT von x_n^N lautet:

$$\begin{aligned} X_m^N &= \frac{1}{\sqrt{N}}\sum_{n=0}^{N-1}\sum_{k=0}^{M-1}\theta_k^M\delta(n-kT)e^{-i2\pi\frac{mn}{N}} && \text{(C.23)}\\ &= \frac{1}{\sqrt{N}}\sum_{k=0}^{M-1}\theta_k^M e^{-i2\pi\frac{mkT}{N}} && \text{(C.24)}\\ &= \frac{1}{\sqrt{N}}\sum_{k=0}^{M-1}\theta_k^M e^{-i2\pi\frac{mk}{M}} && \text{(C.25)}\\ &= \frac{\sqrt{M}}{\sqrt{N}}\Theta_m^M = \frac{1}{\sqrt{T}}\Theta_m^M && \text{(C.26)} \end{aligned}$$

und damit folgt

$$H_m^N = F_m^N + SF_m^N * (\frac{1}{\sqrt{T}}\Theta_m^M \cdot R_m^{T,N}). \tag{C.27}$$

D. Patientenkatalog

Zur Beschreibung des Kataloges siehe auch Kapitel 16.

D.1. Tumorpatienten

D.1.1. Glottische Ersatzphonation nach Tumorentfernung

Patient 1

12/95	(rot) Zustand nach partieller Stimmlippenentfernung beidseitig an der vorderen Komissur. Außerdem partielle Taschenfaltenresektion. **Hoher Rauschanteil**, moderate Irregularität.
2/96	(violett) deutliche **Verringerung der Rauschkomponente**.
3/96	(dunkelblau) Zustand nach Rezidiv-Operation 2/96. Postoperativ **sehr hohe Irregularität und sehr hoher Rauschanteil.**
5/96-6/96	(hellblau-gelb) deutliche **Verringerung der Irregularität** während Stimmtherapie.
7/96-6/99	**Verringerung des Rauschanteils** während der Stimmtherapie. **Stabilisierung** der Stimmqualität bei normaler Irregularität und leicht erhöhtem Rauschanteil.

Bei diesem Patienten verringerten sich die Komponenten des Diagramms vom Zeitpunkt der Rezidivoperation (3/96) bis 9/96 kontinuierlich. Aus der Darstellung drängt sich die Annahme auf, dass sich die Stimmqualität des Patienten in diesem Zeitraum stetig verbessert hat (dies entspricht dem Urteil der betreuenden Logopädin, ist jedoch nicht mit einer unabhängigen Messmethode nachprüfbar gewesen). Die Plausibilität der Annahme wird deutlicher, wenn man die Gegenthese formuliert: Die Stimmqualität des Patienten hat sich diskontinuierlich verändert, aber mangelnde Darstellungsgüte des Diagramms führt zufällig dazu, dass sich die Analyseergebnisse in diesem Zeitraum kontinuierlich verbessern. Nichtsdestotrotz handelt es sich bei der Annahme, dass die Darstellung der Stimmgüte im Heiserkeits-Diagramm auch der objektiven (wenn es auch sehr schwer ist, hier „objektiv“ zu definieren) Entwicklung der Stimmgüte entspricht, um eine Plausibilitätsannahme.

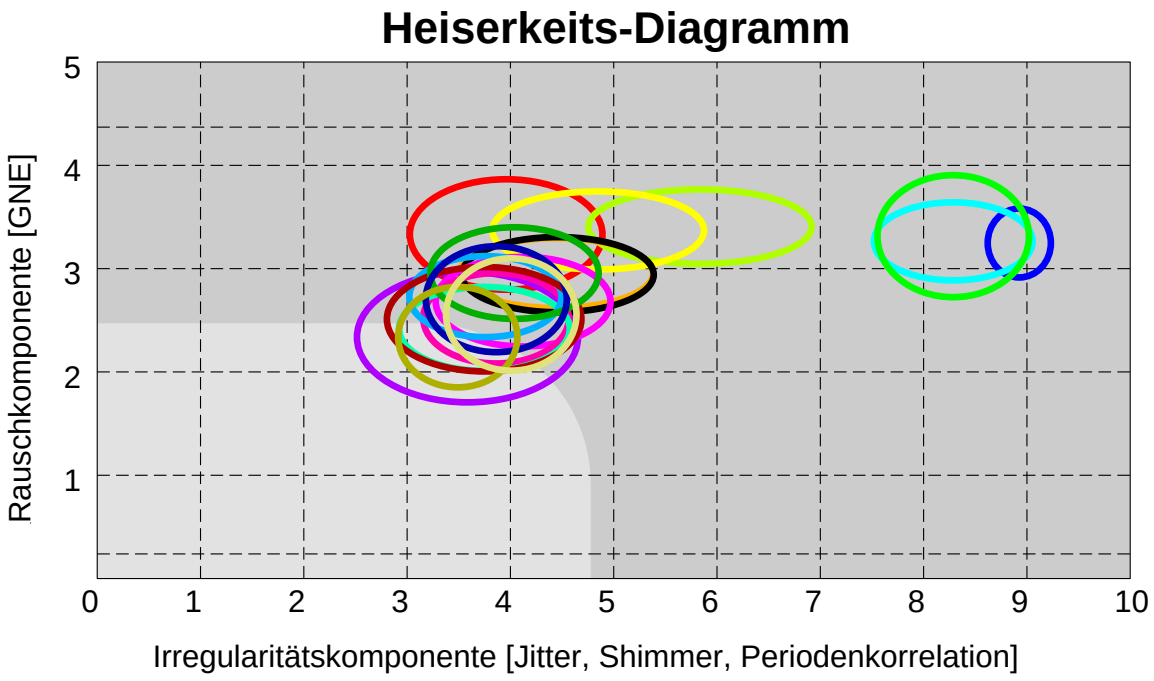
Die Lage der Verteilungen sind für je zwei Aufnahmen signifikant verschieden, obwohl die Mittelpunkte der Ellipsen zum Teil eng beieinanderliegen. Die signifikante Trennung ist durch die hohe Anzahl der analysierten Segmente pro Aufnahme möglich, denn die

Trennschärfe steigt mit der Anzahl der Datenpunkte. Da selbst die relativ kleinen Unterschiede bei diesem Patienten signifikant sind, wurden für die folgenden Patienten nicht extra statistische Tests durchgeführt.

Heiserkeits-Diagramm | Patient 1, m, 42

Datum: 15.06.99 Zeit: 18:37

	Irreg.komp.	Rauschkomp.
06.12.95 n=145	4.0 (0.93)	3.3 (0.53)
15.02.96 n=411	3.6 (1.07)	2.3 (0.63)
13.03.96 n=109	8.9 (0.31)	3.2 (0.33)
24.05.96 n=376	8.3 (0.77)	3.3 (0.38)
31.05.96 n=311	8.3 (0.73)	3.3 (0.59)
12.06.96 n=264	5.8 (1.08)	3.4 (0.36)
26.06.96 n=269	4.8 (1.03)	3.4 (0.38)
11.07.96 n=255	4.4 (0.94)	3.0 (0.34)
26.07.96 n=302	4.5 (0.93)	2.9 (0.36)
07.08.96 n=323	4.1 (0.85)	2.7 (0.43)
23.08.96 n=255	3.9 (0.70)	2.5 (0.44)
04.09.96 n=253	3.7 (0.83)	2.4 (0.41)
10.12.96 n=298	3.8 (0.74)	2.7 (0.39)
24.06.97 n=209	3.7 (0.94)	2.5 (0.51)
15.01.98 n=164	4.0 (0.81)	3.0 (0.44)
08.07.98 n=280	3.9 (0.68)	2.7 (0.51)
09.12.98 n=234	3.5 (0.58)	2.3 (0.49)
09.06.99 n=163	4.0 (0.63)	2.6 (0.54)



Drittes Physikalisches Institut und Abt. Phoniatrie und Pädaudiologie Universität Göttingen

15. Juni 1999

Patient 2

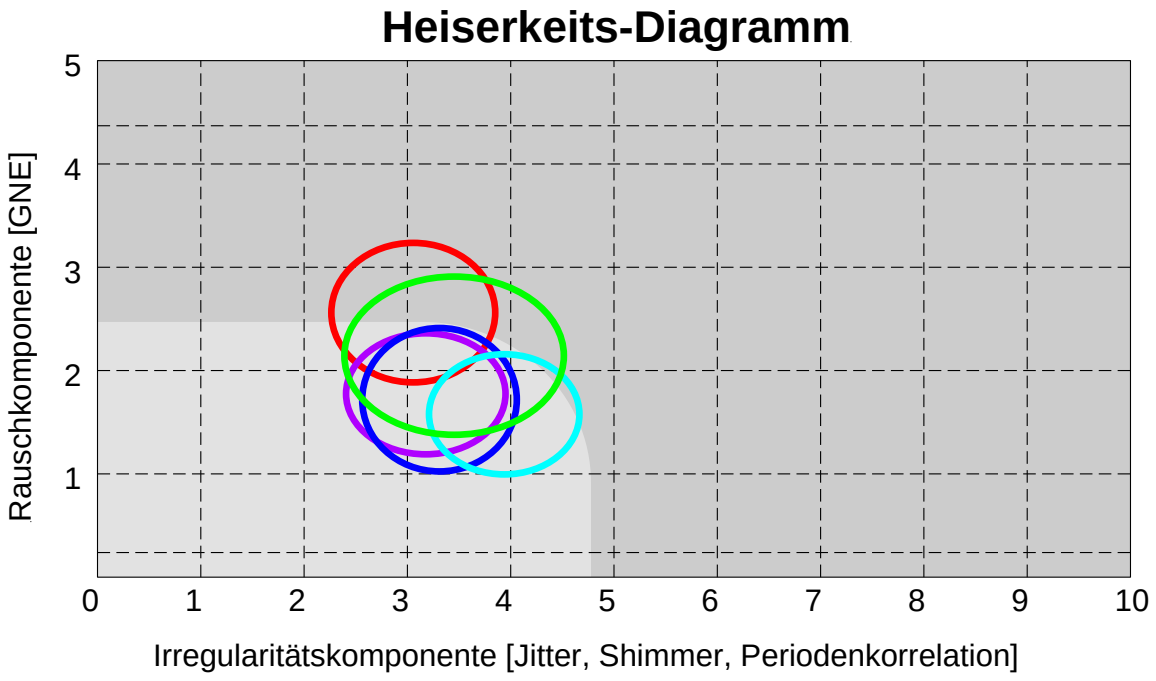
7/96-11/96 Zustand nach partieller Stimmlippenentfernung rechts: (rot-hellblau) leichte **Verringerung der Rauschkomponente** bei leichter Erhöhung der Irregularitätskomponente.

4/99 Kontrolluntersuchung: (grün) im Verlauf leichte Zunahme der Rauschkomponente, leichte Verringerung der Irregularitätskomponente.

Heiserkeits-Diagramm | Patient 2, m, 34

Datum: 30.04.99 Zeit: 17:30

	Irreg.komp.	Rauschkomp.
31.07.96 n=211	3.1 (0.79)	2.6 (0.67)
15.08.96 n=338	3.2 (0.77)	1.8 (0.59)
23.08.96 n=284	3.3 (0.75)	1.7 (0.69)
25.11.96 n=96	3.9 (0.73)	1.6 (0.58)
26.04.99 n=209	3.5 (1.06)	2.1 (0.77)



Drittes Physikalisches Institut und Abt. Phoniatrie und Pädaudiologie Universität Göttingen 30. April 1999

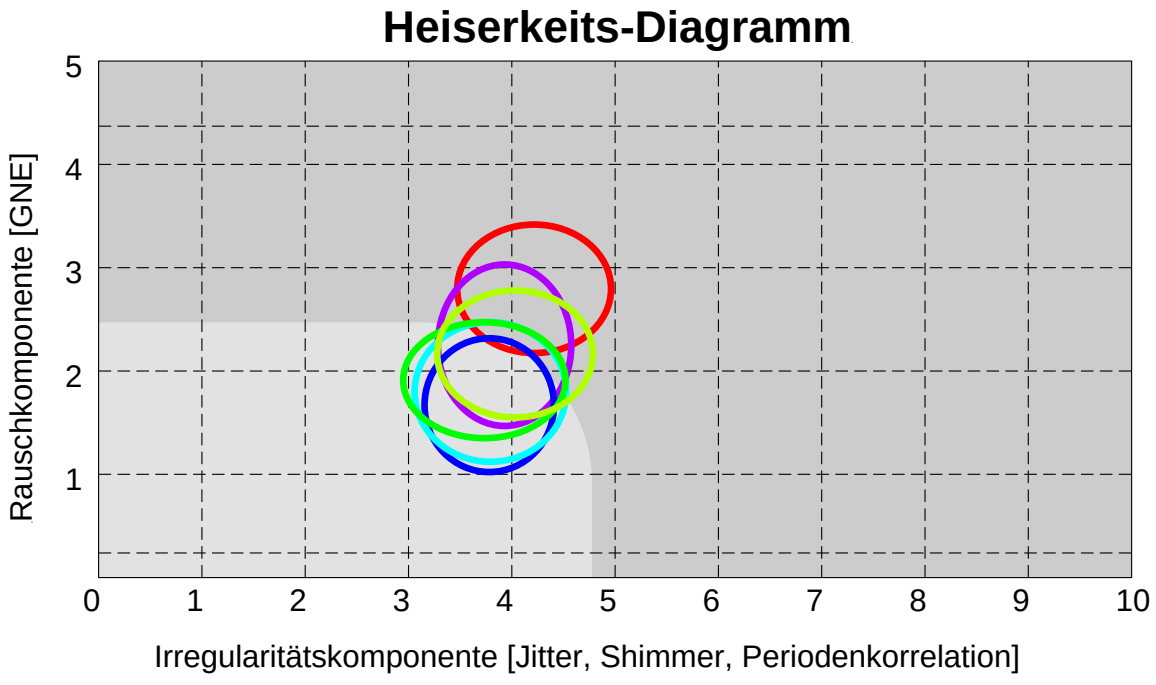
Patient 3

1/96-11/98 Zustand nach Rezidiv-Operation 12/95 (rot). Im Verlauf **deutliche Verringerung der Rauschkomponente**. Korrektur-Operation 9/98. **Stabilisierung** der Stimmqualität bis 1/99

Heiserkeits-Diagramm | Patient 3, m, 27

Datum: 09.11.99 Zeit: 16:21

	Irreg.komp.	Rauschkomp.
24.01.96 n=306	4.2 (0.74)	2.8 (0.62)
19.04.96 n=288	3.9 (0.65)	2.2 (0.78)
09.11.98 n=329	3.8 (0.63)	1.7 (0.65)
14.12.98 n=483	3.8 (0.73)	1.8 (0.68)
18.01.99 n=488	3.7 (0.79)	1.9 (0.56)
13.10.99 n=324	4.0 (0.75)	2.2 (0.61)



Drittes Physikalisches Institut und Abt. Phoniatrie und Pädaudiologie Universität Göttingen 9. Nov. 1999

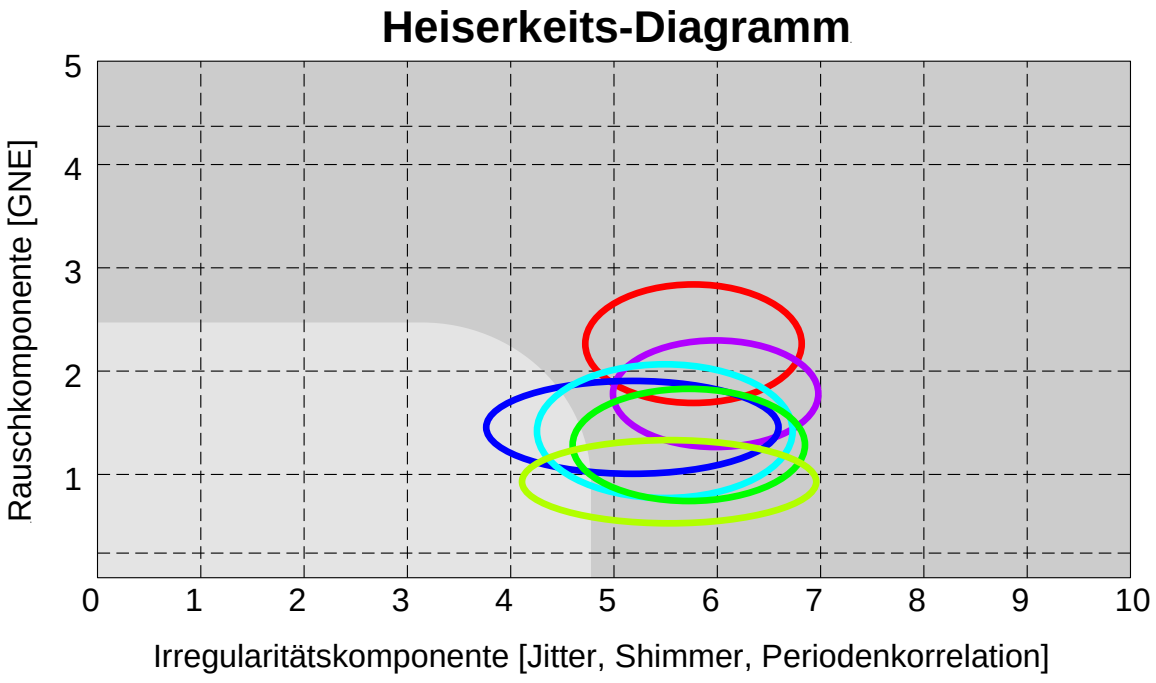
Patient 4

11/95-4/99 Zustand nach partieller Stimmlippenentfernung rechts und Stimmlippen-Dekortikation links 8/93. Im Verlauf deutliche Verringerung der Rauschkomponente

Heiserkeits-Diagramm | Patient 4, m, 33

Datum: 03.05.99 Zeit: 10:58

	Irreg.komp.	Rauschkomp.
29.11.95 n=256	5.8 (1.05)	2.3 (0.57)
17.04.96 n=254	6.0 (0.99)	1.8 (0.52)
23.10.96 n=212	5.2 (1.42)	1.5 (0.45)
16.10.97 n=237	5.5 (1.24)	1.4 (0.65)
11.05.98 n=157	5.7 (1.12)	1.3 (0.54)
28.04.99 n=79	5.5 (1.42)	0.9 (0.40)



Drittes Physikalisches Institut und Abt. Phoniatrie und Pädaudiologie Universität Göttingen 3. Mai 1999

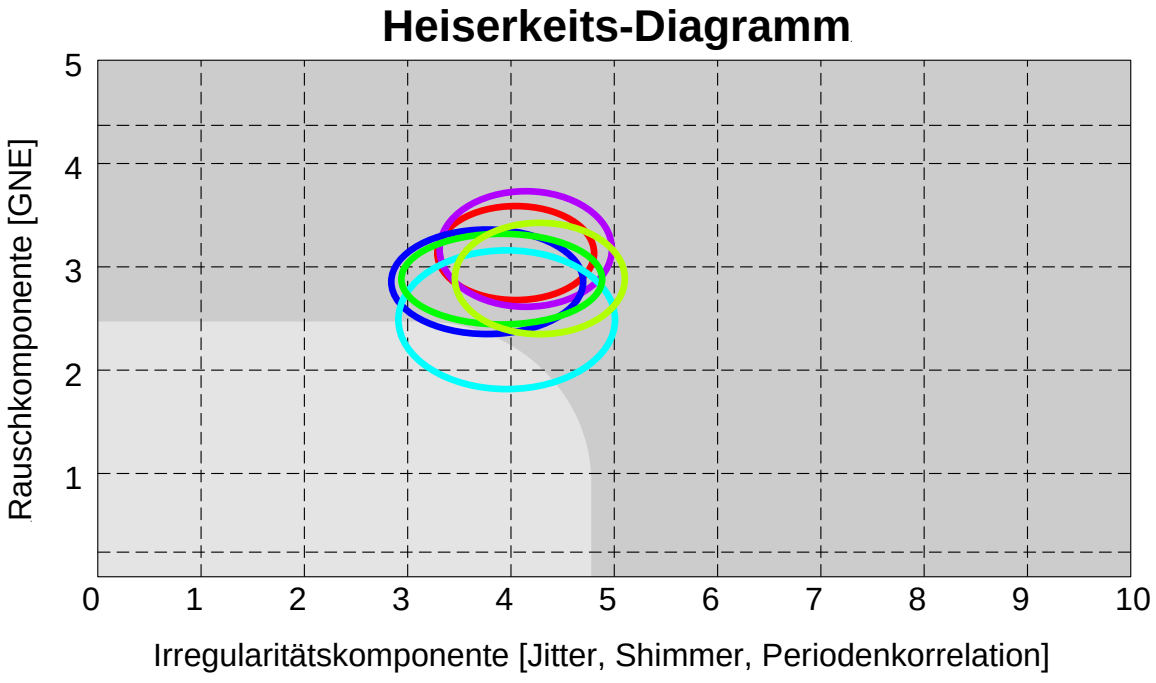
Patientin 5

2/97-12/97 Zustand nach partieller Stimmlippenentfernung rechts. **Stabile Lage** bei leicht erhöhter Rauschkomponente, im Mittel leichte Verringerung der Rauschkomponente

Heiserkeits-Diagramm | Patientin 5, w, 43

Datum: 17.06.98 Zeit: 16:40

	Irreg.komp.	Rauschkomp.
12.02.97 n=317	4.0 (0.76)	3.1 (0.45)
20.08.97 n=220	4.1 (0.83)	3.2 (0.56)
26.08.97 n=222	3.8 (0.93)	2.9 (0.51)
04.09.97 n=186	4.0 (1.05)	2.5 (0.67)
05.09.97 n=153	3.9 (0.97)	2.9 (0.44)
10.12.97 n=332	4.3 (0.82)	2.9 (0.54)



Drittes Physikalisches Institut und Abt. Phoniatrie und Pädaudiologie Universität Göttingen

17. Juni 1998

Patient 6

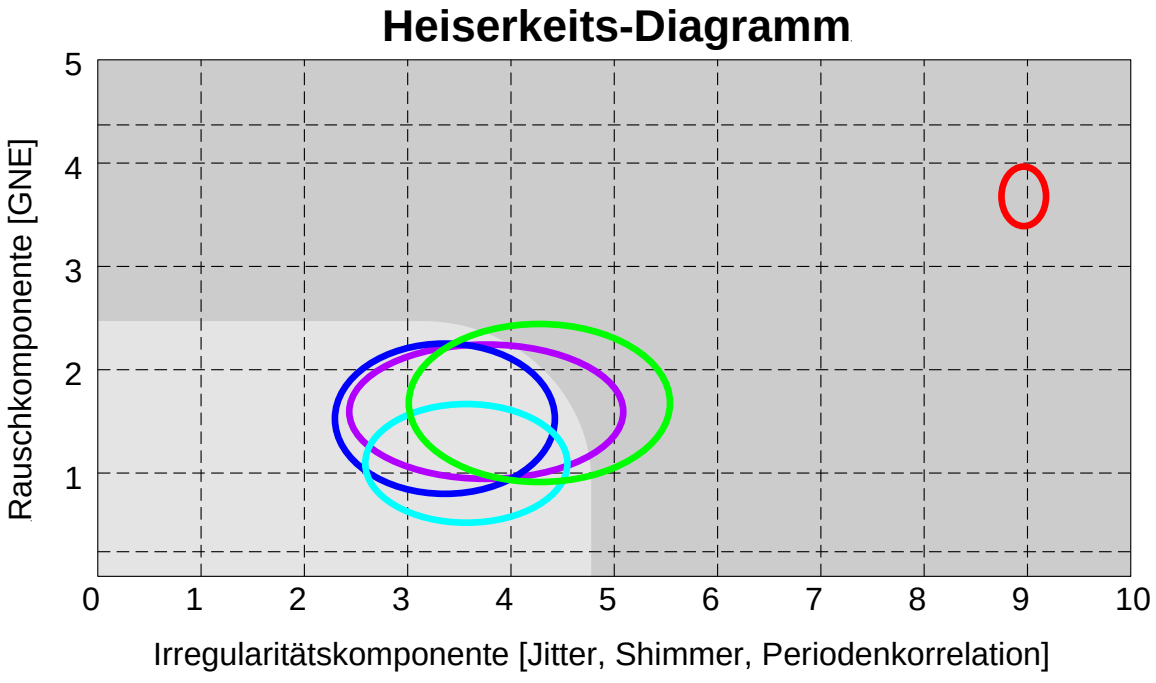
11/95 (rot) **Postoperativer Zustand nach partieller Stimmlippenentfernung rechts.**

1/97-11/97 Zustand nach partieller Taschenfaltenresektion 7/96. **Stabile Lage am Rande des Normalbereichs.**

Heiserkeits-Diagramm | Patient 6, m, 34

Datum: 29.06.99 Zeit: 17:09

	Irreg.komp.	Rauschkomp.
30.11.95 n=55	9.0 (0.22)	3.7 (0.29)
21.01.97 n=241	3.8 (1.33)	1.6 (0.65)
04.02.97 n=226	3.4 (1.06)	1.5 (0.73)
17.02.97 n=291	3.6 (0.98)	1.1 (0.57)
05.11.97 n=98	4.3 (1.26)	1.7 (0.76)



Drittes Physikalisches Institut und Abt. Phoniatrie und Pädaudiologie Universität Göttingen

29. Juni 1999

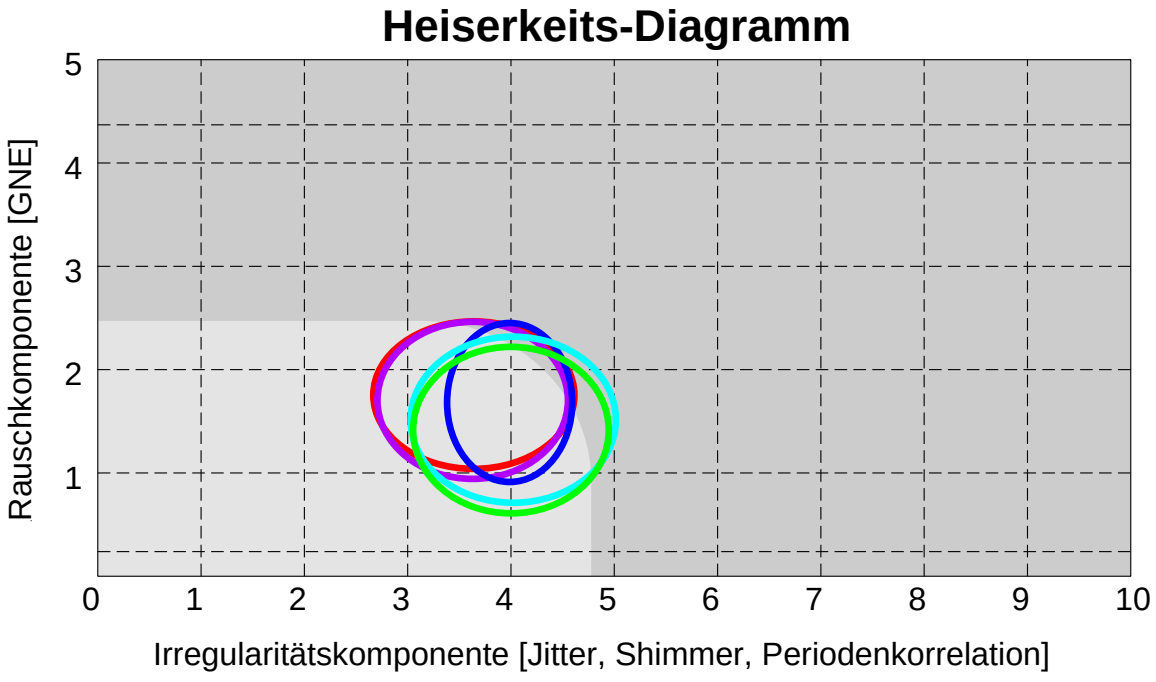
Patient 7

5/96-5/98 Zustand nach partieller Stimmlippenentfernung links 9/94. **Stabile Lage im Normalbereich**

Heiserkeits-Diagramm | Patient 7, m, 33

Datum: 17.06.98 Zeit: 16:29

	Irreg.komp.	Rauschkomp.
08.05.96 n=361	3.6 (0.97)	1.8 (0.71)
13.11.96 n=443	3.6 (0.92)	1.7 (0.76)
14.05.97 n=181	4.0 (0.61)	1.7 (0.77)
12.11.97 n=158	4.0 (0.99)	1.5 (0.80)
13.05.98 n=337	4.0 (0.95)	1.4 (0.81)



Drittes Physikalisches Institut und Abt. Phoniatrie und Pädaudiologie Universität Göttingen 17. Juni 1998

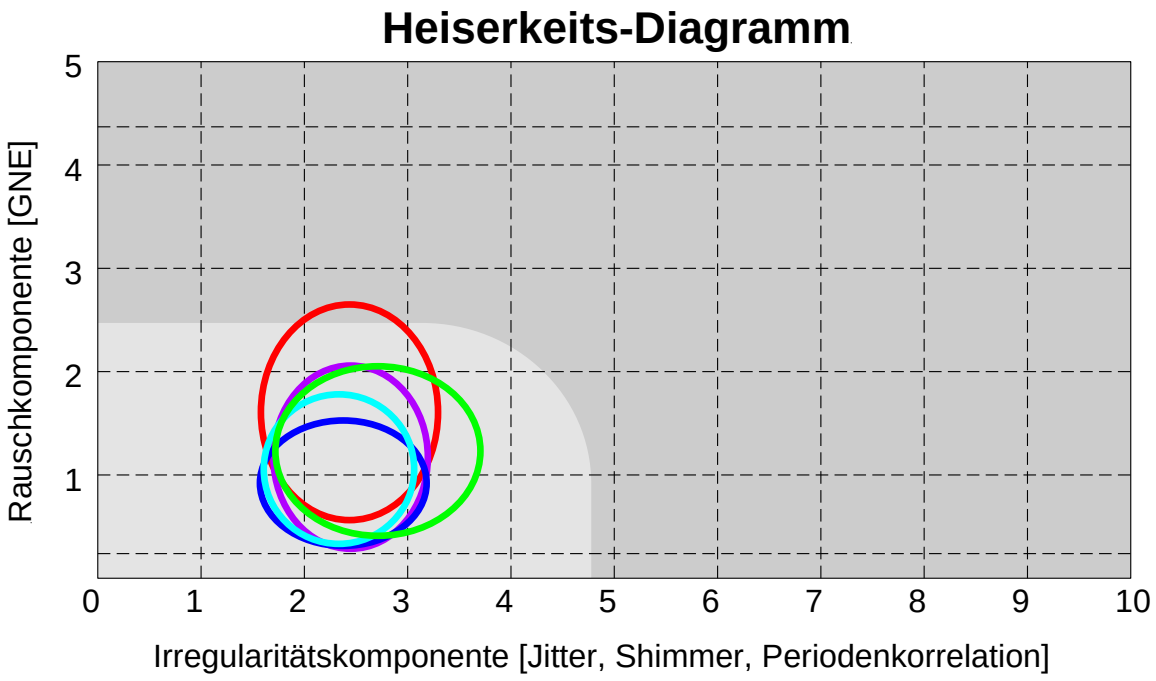
Patientin 8

9/96-7/97 Zustand nach partieller Stimmlippenentfernung links. **Stabile Lage im Normalbereich**, geringe Irregularität.

Heiserkeits-Diagramm | Patientin 8, w, 64

Datum: 17.06.98 Zeit: 16:31

	Irreg.komp.	Rauschkomp.
16.09.96 n=289	2.4 (0.86)	1.6 (1.04)
08.10.96 n=364	2.4 (0.75)	1.2 (0.89)
08.11.96 n=397	2.4 (0.81)	0.9 (0.61)
17.12.96 n=421	2.3 (0.73)	1.1 (0.72)
04.07.97 n=451	2.7 (0.99)	1.2 (0.82)



Drittes Physikalisches Institut und Abt. Phoniatrie und Pädaudiologie Universität Göttingen 17. Juni 1998

Patient 9

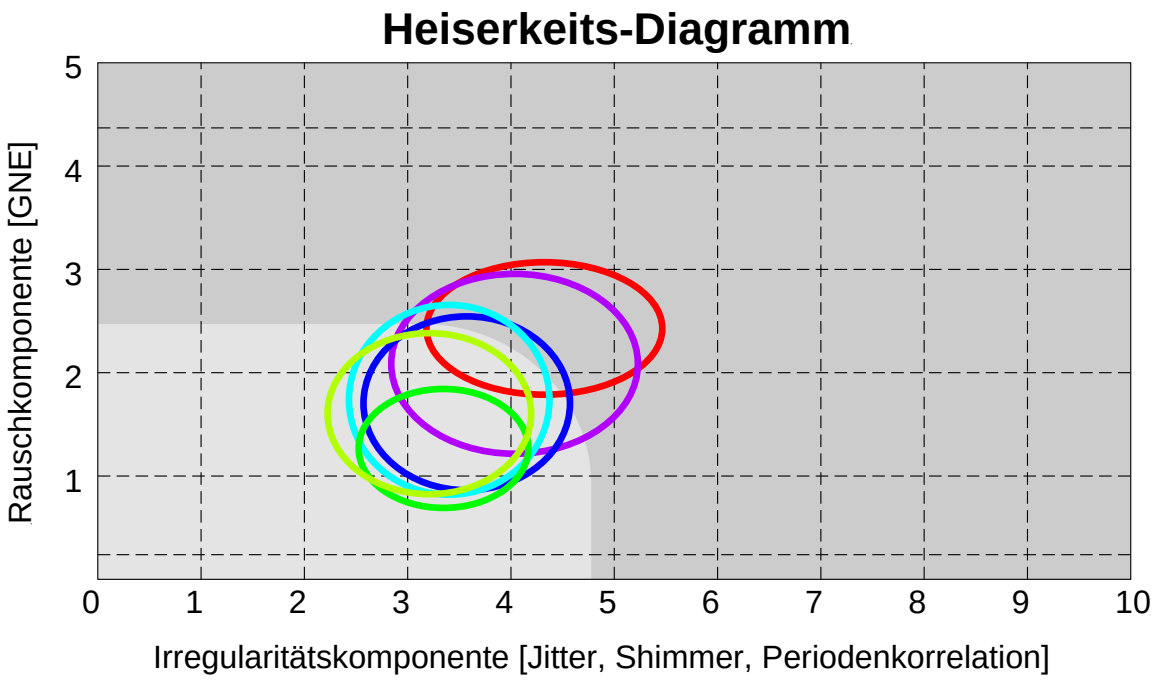
6/98-11/98 Zustand nach partieller Stimmlippenentfernung rechts. **Verringerung von Rausch- und Irregularitätskomponente** während Stimmtherapie. Abschluss der Stimmrehabilitation 11/98

3/99-7/99 **Stabilisierung** im Normalbereich.

Heiserkeits-Diagramm | Patient 9, m, 26

Datum: 22.07.99 Zeit: 16:38

	Irreg.komp.	Rauschkomp.
10.06.98 n=95	4.3 (1.14)	2.4 (0.64)
28.10.98 n=163	4.0 (1.19)	2.1 (0.87)
17.11.98 n=134	3.6 (1.00)	1.7 (0.84)
24.03.99 n=221	3.4 (0.97)	1.7 (0.92)
21.05.99 n=315	3.3 (0.82)	1.3 (0.58)
12.07.99 n=191	3.2 (0.99)	1.6 (0.78)



Drittes Physikalisches Institut und Abt. Phoniatrie und Pädaudiologie Universität Göttingen 22. Juli 1999

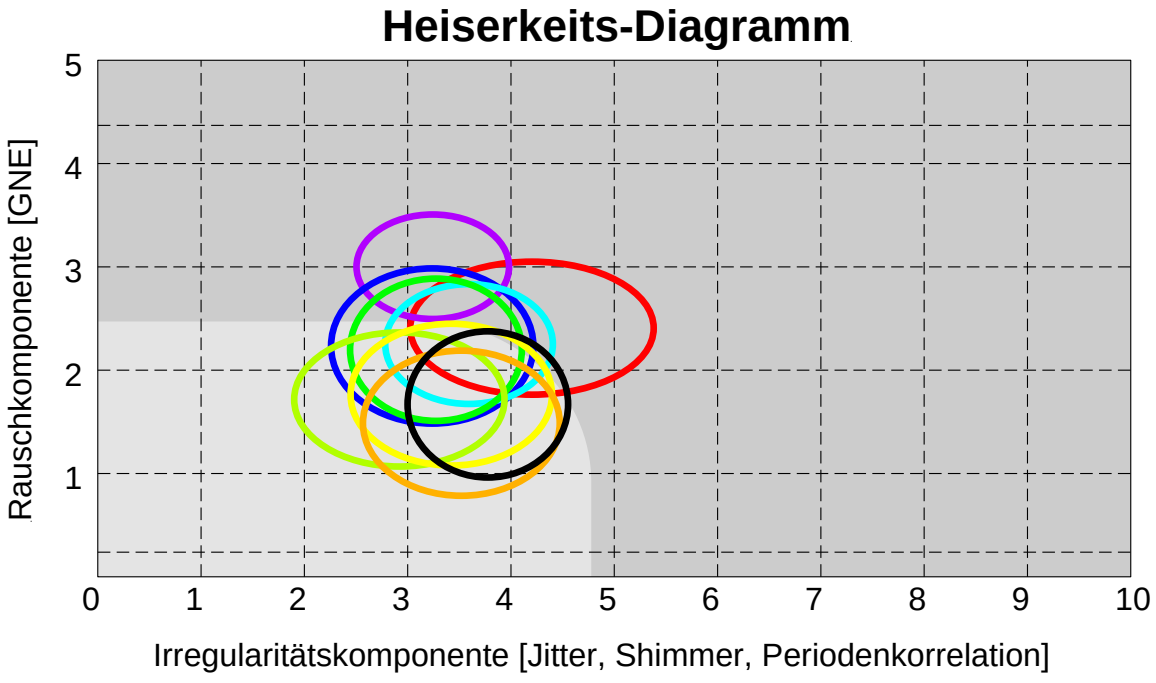
Patient 10

2/96	(rot) Stimmlippentumor links **präoperativ**
10/96	(violett) Zustand nach partieller Stimmlippenentfernung links 3/96. **Erhöhte Rauschkomponente**
11,12/96	Verringerung der Rauschkomponente bei Therapie
7/97-3/99	**Stabilisierung** im Normalbereich.

Heiserkeits-Diagramm | Patient 10, m, 41

Datum: 22.03.99 Zeit: 16:57

	Irreg.komp.	Rauschkomp.
21.02.96 n=400	4.2 (1.18)	2.4 (0.64)
29.10.96 n=389	3.2 (0.74)	3.0 (0.51)
14.11.96 n=304	3.2 (0.98)	2.2 (0.75)
04.12.96 n=233	3.6 (0.81)	2.3 (0.58)
18.12.96 n=409	3.3 (0.83)	2.2 (0.69)
09.07.97 n=273	2.9 (1.02)	1.7 (0.65)
08.10.97 n=385	3.4 (0.97)	1.8 (0.69)
28.01.98 n=391	3.5 (0.95)	1.5 (0.70)
03.03.99 n=215	3.8 (0.78)	1.7 (0.71)



Drittes Physikalisches Institut und Abt. Phoniatrie und Pädaudiologie Universität Göttingen

22. Maerz 1999

Patient 11

6/97 (rot) Zustand nach partieller Stimmlippenentfernung links. **Erhöhte Rausch- und Irregularitätskomponente**

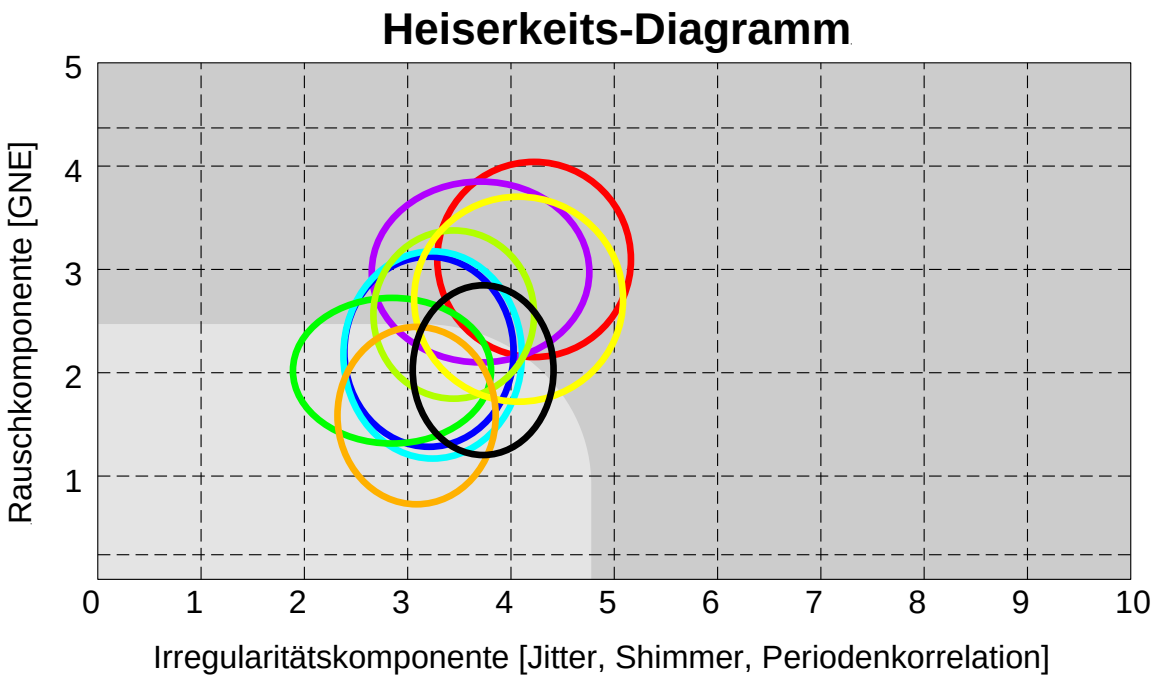
6,7/97 Verringerung der Rausch- und Irregularitätskomponente bei Therapie. Zwei Aufnahmen am 14.07.97: **vor und nach Betäubung** zur Laryngoskopie. Die Betäubung hat keinen wesentlichen Einfluss auf die Analyse.

4/98-2/99 **leichte Variation** am Rande des Normalbereichs.

Heiserkeits-Diagramm | Patient 11, m, 54

Datum: 02.02.99 Zeit: 23:23

	Irreg.komp.	Rauschkomp.
02.06.97 n=315	4.2 (0.94)	3.1 (0.94)
13.06.97 n=312	3.7 (1.05)	3.0 (0.88)
14.07.97 n=609	3.2 (0.82)	2.2 (0.92)
14.07.97 n=664	3.2 (0.86)	2.2 (1.00)
24.07.97 n=531	2.8 (0.96)	2.0 (0.70)
01.09.97 n=478	3.4 (0.78)	2.6 (0.81)
05.09.97 n=495	4.1 (1.01)	2.7 (0.99)
22.04.98 n=644	3.1 (0.77)	1.6 (0.86)
02.02.99 n=790	3.7 (0.68)	2.0 (0.82)



Drittes Physikalisches Institut und Abt. Phoniatrie und Pädaudiologie Universität Göttingen 2. Feb. 1999

D.1.2. Glottische Ersatzphonation nach Tumorentfernung ohne Schwingung der operierten Stimmlippe

Patient 12

2,8/96 (rot,violett) Zustand nach partieller Stimmlippenentfernung links 10+11/93. **Erhöhte Rausch- und Irregularitätskomponente**

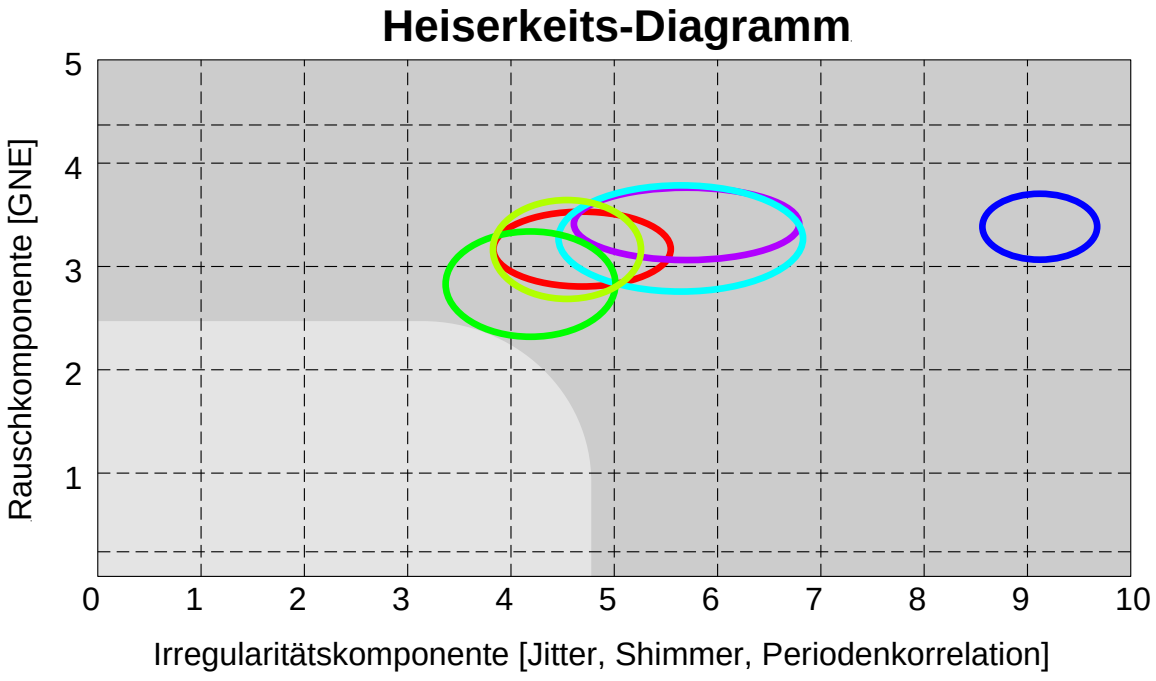
1/99 Zustand nach partieller Stimmlippenentfernung rechts (10/98) und links (11/98). Abschluss Wundheilung

5-9/99 Deutliche**Verringerung der Rausch- und Irregularitätskomponente** während Stimmrehabilitation. Relativ hohe Rauschkomponente (9/99).

Heiserkeits-Diagramm | Patient 12, m, 31

Datum: 23.11.99 Zeit: 02:17

	Irreg.komp.	Rauschkomp.
06.02.96 n=245	4.7 (0.86)	3.2 (0.36)
07.08.96 n=189	5.7 (1.09)	3.4 (0.35)
13.01.99 n=71	9.1 (0.56)	3.4 (0.32)
11.05.99 n=226	5.6 (1.18)	3.3 (0.51)
04.06.99 n=241	4.2 (0.82)	2.8 (0.51)
13.09.99 n=271	4.5 (0.72)	3.2 (0.48)



Drittes Physikalisches Institut und Abt. Phoniatrie und Pädaudiologie Universität Göttingen

23. Nov. 1999

Patient 13

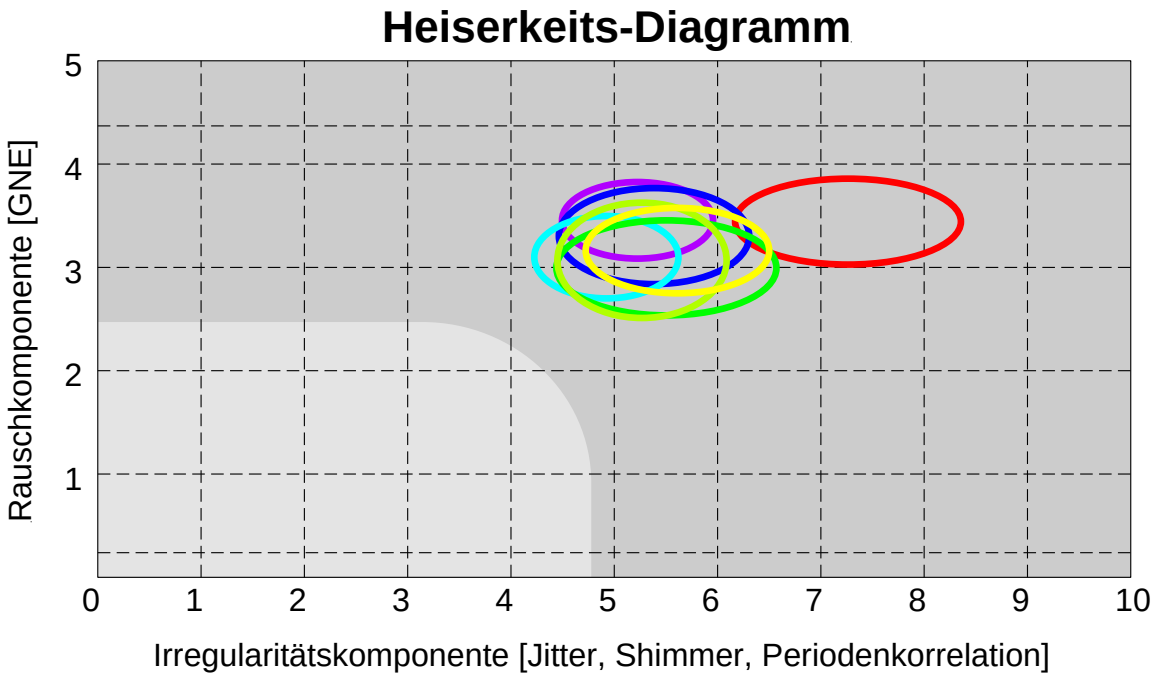
1/97 (rot) Zustand nach subtotaler Stimmlippenentfernung links 11/96. **Sehr hohe Rausch- und Irregularitätskomponente.**

7/97-3/99 Verringerung der Irregularitätskomponente. **Stabilisierung** bei hoher Rausch- und erhöhter Irregularitätskomponente.

Heiserkeits-Diagramm | Patient 13, m, 34

Datum: 22.03.99 Zeit: 16:08

	Irreg.komp.	Rauschkomp.
22.01.97 n=126	7.3 (1.09)	3.4 (0.42)
17.07.97 n=222	5.2 (0.73)	3.5 (0.37)
24.07.97 n=242	5.4 (0.92)	3.3 (0.46)
05.11.97 n=269	4.9 (0.70)	3.1 (0.40)
23.03.98 n=138	5.5 (1.06)	3.0 (0.46)
22.07.98 n=150	5.3 (0.82)	3.1 (0.56)
17.03.99 n=131	5.6 (0.89)	3.2 (0.41)



Drittes Physikalisches Institut und Abt. Phoniatrie und Pädaudiologie Universität Göttingen 22. Maerz 1999

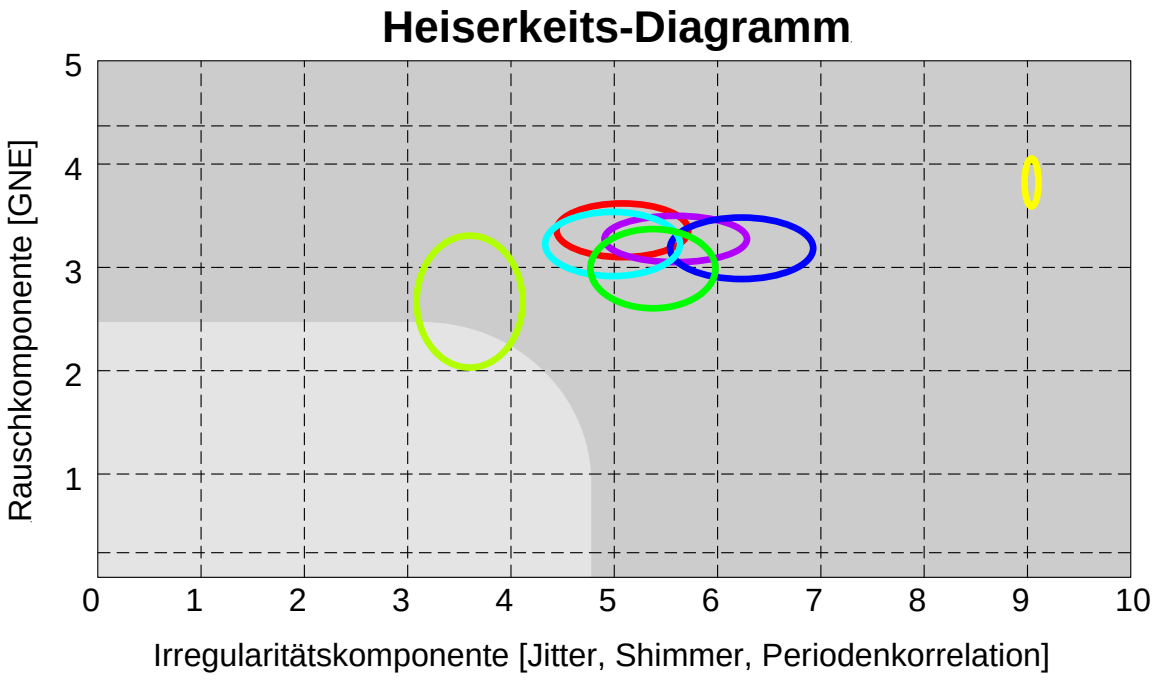
Patientin 14

11/95-8/96	Zustand nach partieller Stimmlippenentfernung (T3 Tumor) links 6/95. Relativ stabil bei **hoher Rausch- und Irregularitätskomponente.**
1/97	**Deutliche Verringerung der Irregularitätskomponente.**
9/97	Aphonie bei Tumorrezidiv.

Heiserkeits-Diagramm | Patientin 14, w, 48

Datum: 04.06.99 Zeit: 20:35

	Irreg.komp.	Rauschkomp.
24.11.95 n=190	5.1 (0.64)	3.4 (0.26)
30.11.95 n=129	5.6 (0.69)	3.3 (0.22)
28.02.96 n=86	6.2 (0.69)	3.2 (0.30)
22.04.96 n=152	5.0 (0.65)	3.2 (0.31)
07.08.96 n=175	5.4 (0.60)	3.0 (0.38)
20.01.97 n=304	3.6 (0.51)	2.7 (0.64)
17.09.97 n=129	9.0 (0.07)	3.8 (0.23)



Drittes Physikalisches Institut und Abt. Phoniatrie und Pädaudiologie Universität Göttingen 4. Juni 1999

D.1.3. Taschenfaltenstimme - ventrikuläre Ersatzphonation

Patient 15

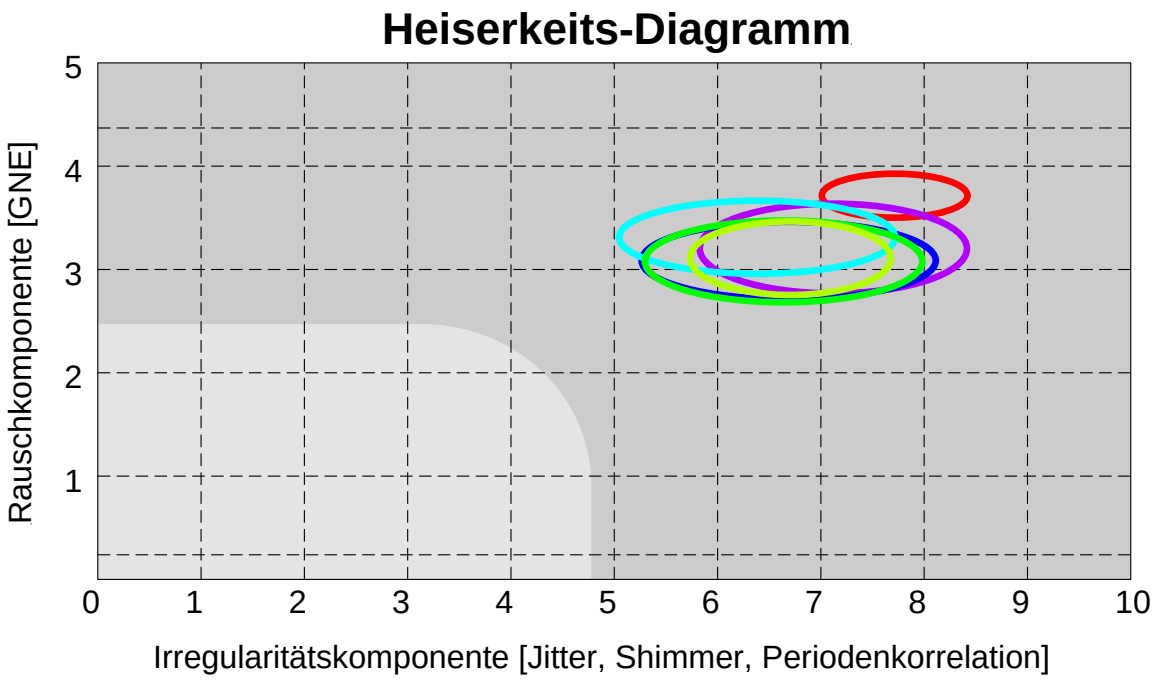
6/97 (rot) Zustand nach subtotaler Stimmlippenentfernung links 3/97. **Sehr hohe Rausch- und Irregularitätskomponente**

3/98-4/99 **Verringerung der Rausch- und Irregularitätskomponente**, Stabilisierung bei hoher Rausch- und Irregularitätskomponente.

Heiserkeits-Diagramm | Patient 15, m, 36

Datum: 09.11.99 Zeit: 23:07

	Irreg.komp.	Rauschkomp.
18.06.97 n=117	7.7 (0.70)	3.7 (0.21)
03.03.98 n=71	7.1 (1.29)	3.2 (0.43)
31.03.98 n=185	6.7 (1.42)	3.1 (0.38)
17.06.98 n=145	6.4 (1.34)	3.3 (0.35)
28.04.99 n=91	6.6 (1.34)	3.1 (0.40)
03.11.99 n=92	6.7 (0.97)	3.1 (0.36)



Drittes Physikalisches Institut und Abt. Phoniatrie und Pädaudiologie Universität Göttingen 9. Nov. 1999

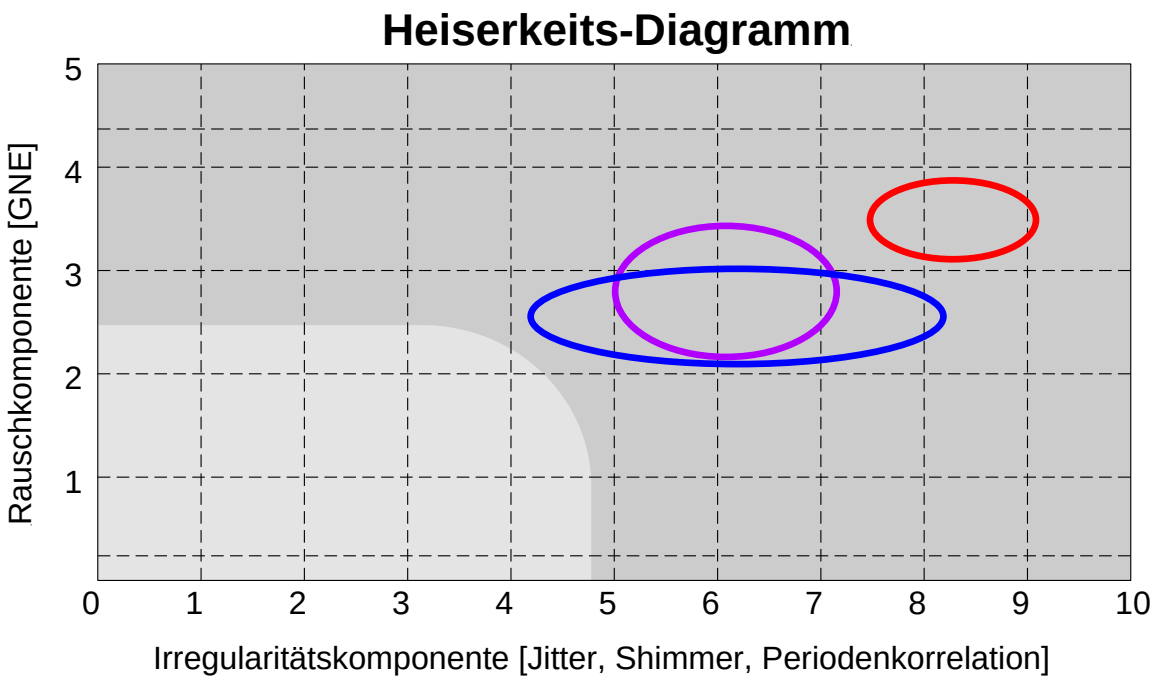
Patient 16

6/98 (rot) Zustand nach kompletter Stimmlippenentfernung links 4/96. **Sehr hohe Rausch- und Irregularitätskomponente.** Beginnende ventrikuläre Ersatzphonation. Wundheilungsphase.
8/98 **Verringerung der Rausch- und Irregularitätskomponente**
2/99 weitere leichte Verringerung der Rauschkomponente. **Starke Streuung der Irregularitätskomponente.**

Heiserkeits-Diagramm | Patient 16, m, 41

Datum: 14.05.99 Zeit: 14:18

	Irreg.komp.	Rauschkomp.
10.06.98 n=62	8.3 (0.80)	3.5 (0.38)
13.08.98 n=72	6.1 (1.07)	2.8 (0.63)
22.02.99 n=138	6.2 (2.00)	2.6 (0.46)



Drittes Physikalisches Institut und Abt. Phoniatrie und Pädaudiologie Universität Göttingen 14. Mai 1999

Patient 17

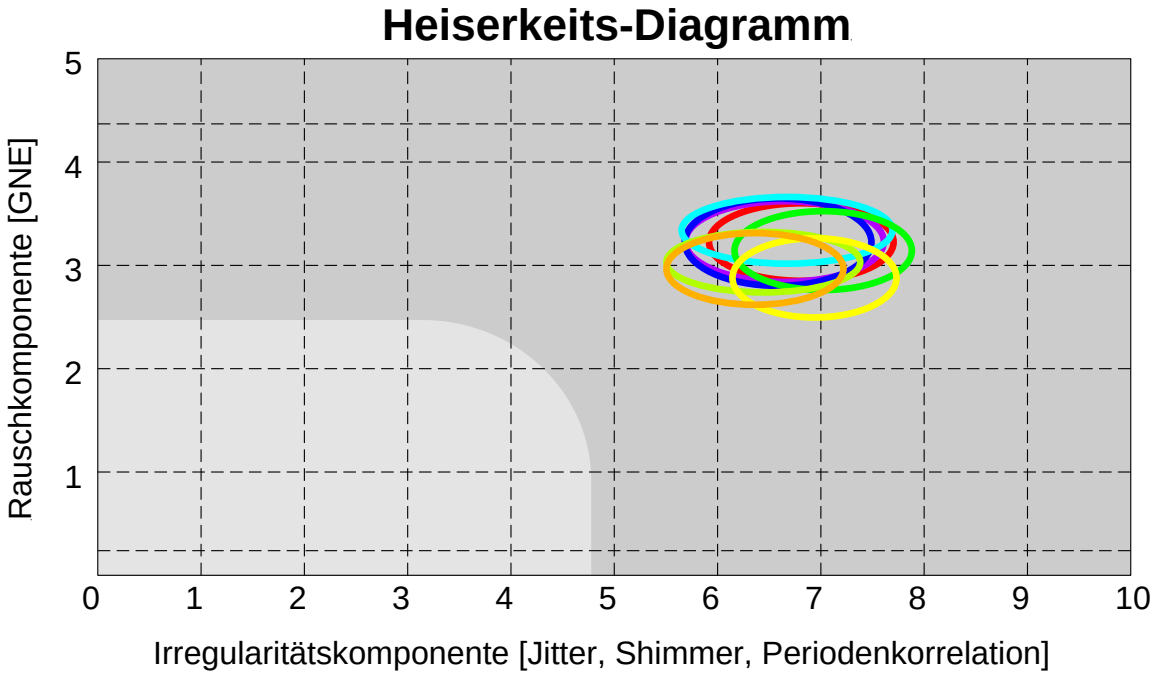
2/97 (rot) Zustand nach subtotaler Stimmlippenentfernung rechts 1/97 und nach Kontrolloperation 3/97 **hohe Rausch- und Irregularitätskomponente.**

7/97-9/99 **Stabile Lage** bei hoher Rausch- und Irregularitätskomponente.

Heiserkeits-Diagramm | Patient 17, m, 48

Datum: 15.11.99 Zeit: 09:35

	Irreg.komp.	Rauschkomp.
02.07.97 n=219	6.8 (0.89)	3.2 (0.38)
25.07.97 n=281	6.6 (0.96)	3.2 (0.39)
31.07.97 n=320	6.6 (0.90)	3.2 (0.42)
05.08.97 n=207	6.7 (1.02)	3.3 (0.32)
19.11.97 n=240	7.0 (0.86)	3.1 (0.38)
14.10.98 n=116	6.4 (0.94)	3.0 (0.29)
24.02.99 n=141	6.9 (0.79)	2.9 (0.38)
06.09.99 n=194	6.4 (0.86)	3.0 (0.35)



Drittes Physikalisches Institut und Abt. Phoniatrie und Pädaudiologie Universität Göttingen

15. Nov. 1999

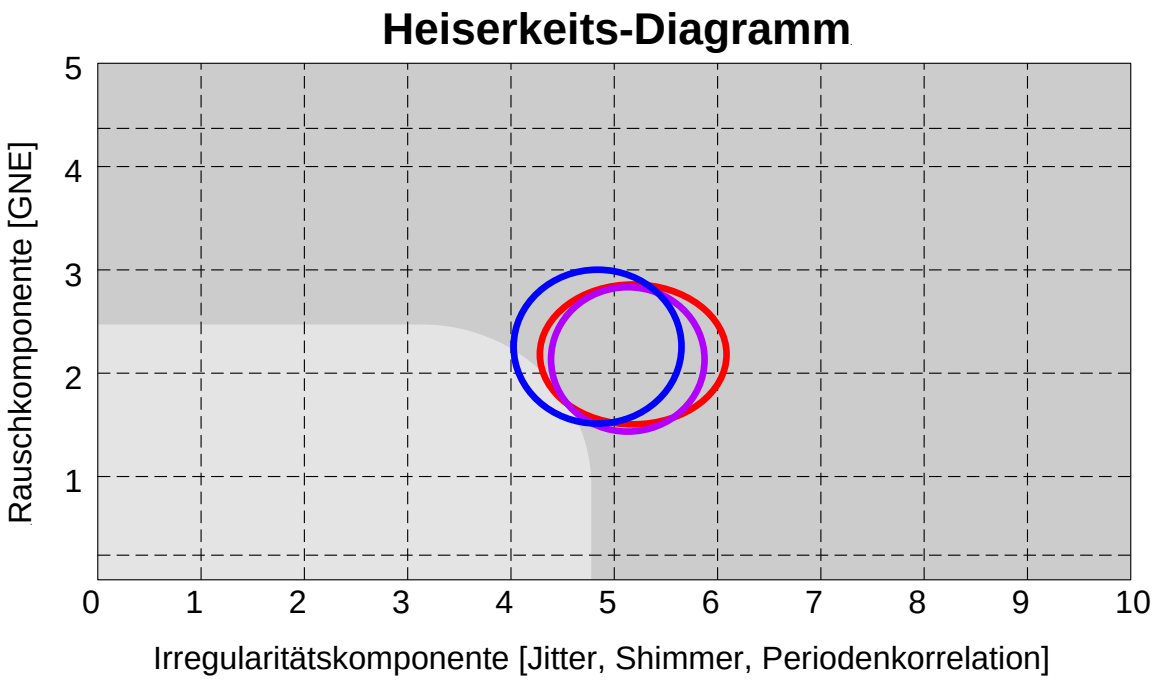
Patientin 18

5/96 Zustand nach kompletter Stimmlippenentfernung links.
8/96 **Stabile Lage** bei erhöhter Rausch- und erhöhter Irregularitätskomponente (Irregularität relativ gering für Taschenfaltenstimme).

Heiserkeits-Diagramm | Patientin 18, w, 53

Datum: 17.06.98 Zeit: 16:41

	Irreg.komp.	Rauschkomp.
30.05.96 n=223	5.2 (0.90)	2.2 (0.68)
14.08.96 n=288	5.1 (0.74)	2.1 (0.70)
28.08.96 n=225	4.8 (0.81)	2.3 (0.74)



Drittes Physikalisches Institut und Abt. Phoniatrie und Pädaudiologie Universität Göttingen 17. Juni 1998

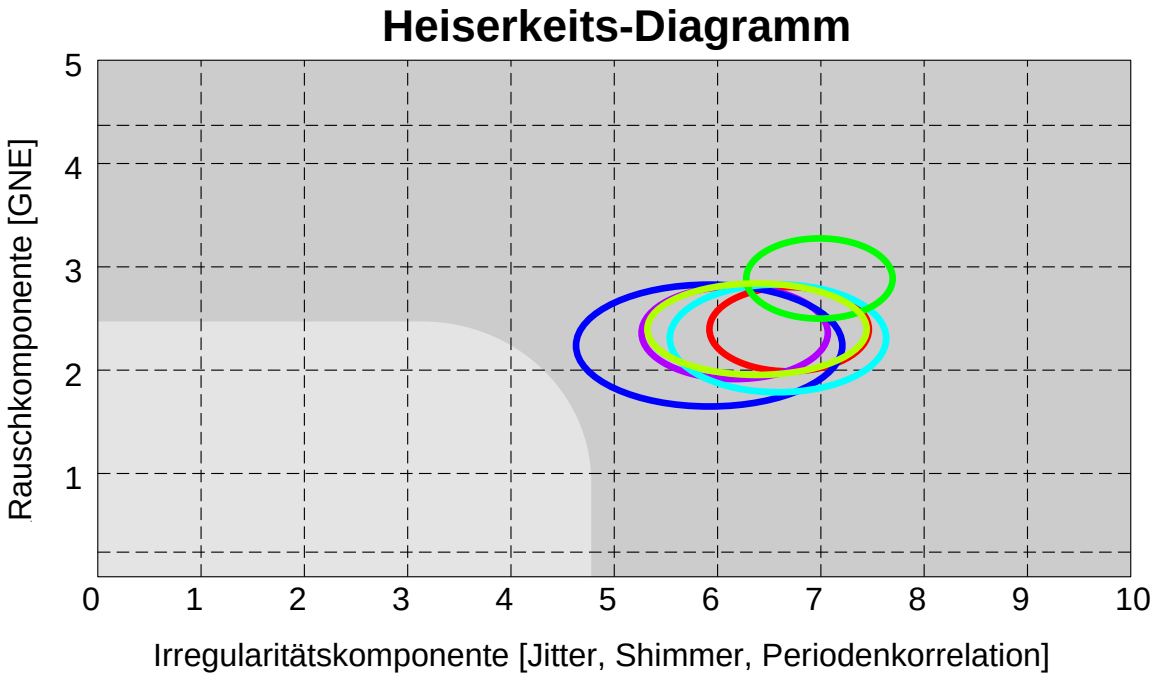
Patient 19

5/96	Zustand nach kompletter Stimmlippenentfernung links 10/93.
11/96-11/98	**Stabile Lage** bei erhöhter Rausch- und erhöhter Irregularitätskomponente. Leicht erhöhte Rauschkomponente 6/98

Heiserkeits-Diagramm | Patient 19, m, 40

Datum: 28.07.99 Zeit: 21:14

	Irreg.komp.	Rauschkomp.
15.05.96 n=163	6.7 (0.77)	2.4 (0.41)
13.11.96 n=123	6.2 (0.90)	2.4 (0.45)
05.11.97 n=144	5.9 (1.29)	2.2 (0.59)
06.05.98 n=81	6.6 (1.05)	2.3 (0.52)
23.06.98 n=151	7.0 (0.71)	2.9 (0.39)
11.11.98 n=49	6.4 (1.06)	2.4 (0.44)



Drittes Physikalisches Institut und Abt. Phoniatrie und Pädaudiologie Universität Göttingen 28. Juli 1999

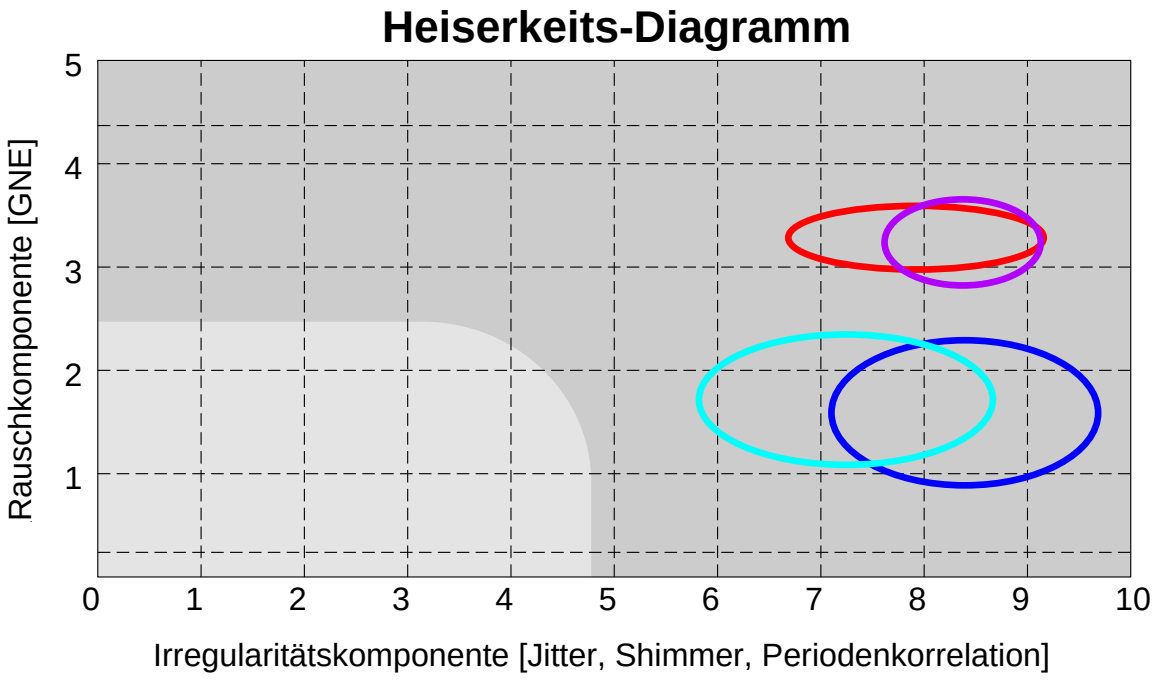
Patient 20

11/97 Zustand nach kompletter Stimmlippenentfernung links und Bestrahlung. Vor Stimmrehabilitation.
5/98 **Sehr hohe Rausch- und Irregularitätskomponente.**
5/98 Korrektur-Operation und Bestrahlung. **Geringe Rauschkomponente hohe Irregularität.**
5/99 **Verringerung der Irregularitätskomponente** bei gleichbleibend geringer Rauschkomponente.

Heiserkeits-Diagramm | Patient 20, m, 45

Datum: 17.05.99 Zeit: 16:37

	Irreg.komp.	Rauschkomp.
24.11.97 n=360	7.9 (1.24)	3.3 (0.31)
04.05.98 n=91	8.4 (0.76)	3.2 (0.42)
15.05.98 n=195	8.4 (1.29)	1.6 (0.70)
17.05.99 n=273	7.2 (1.42)	1.7 (0.63)

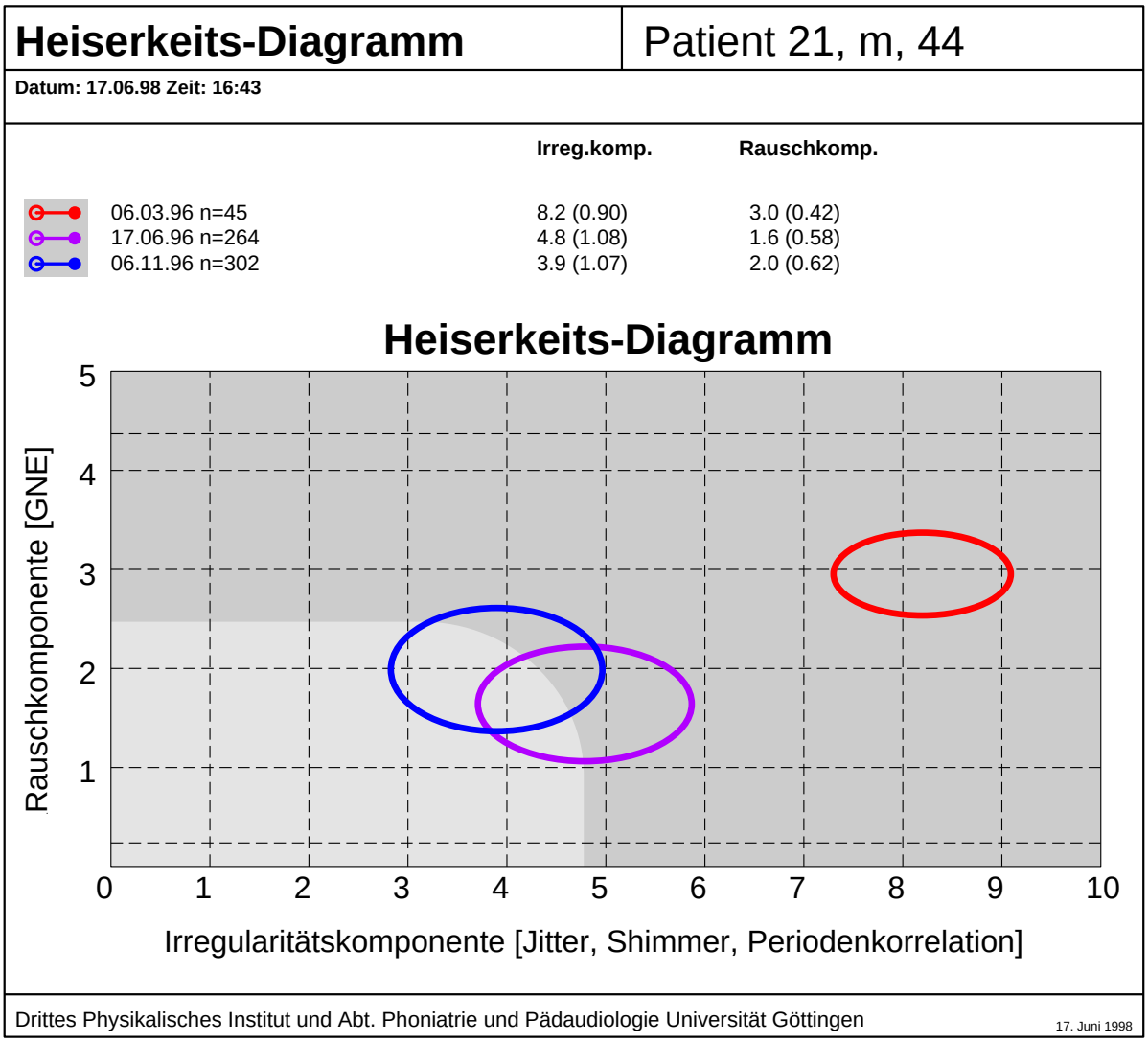


Drittes Physikalisches Institut und Abt. Phoniatrie und Pädaudiologie Universität Göttingen 17. Mai 1999

Patient 21

3/96 Zustand nach partieller Stimmlippenentfernung rechts 2/96. **Spontane postoperative Taschenfaltenphonation.** Sehr hohe Rausch- und Irregularitätskomponente.

6/96-11/96 **Wechsel zur glottischen Ersatzphonation.** Rausch- und Irregularitätskomponente am Rande des Normalbereichs.



Patient 22

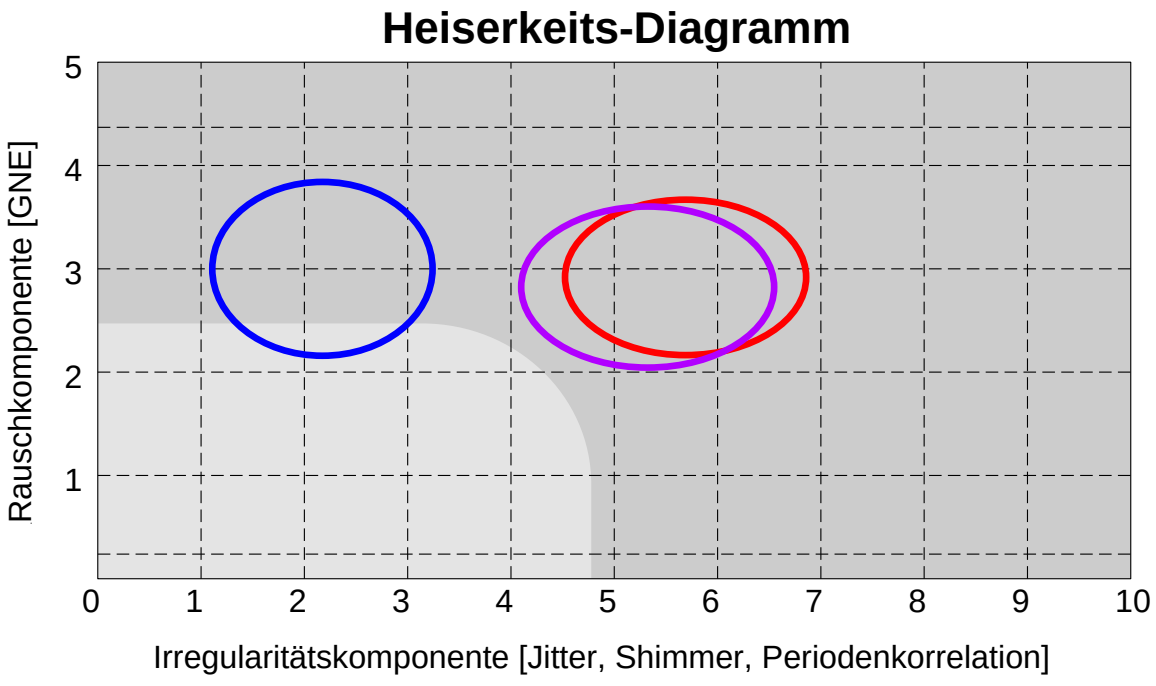
6/96 Taschenfaltenstimme nach Operation wegen chronischer hyperplastischer Laryngitis. Hohe Rausch- und Irregularitätskomponente.

7/96 **Taschenfaltenstimme** (violett) und **Übungsstimme** bei 600Hz Grundfrequenz (dunkelblau) bei deutlich verringerter Irregularität.

Heiserkeits-Diagramm | Patient 22, m, 76

Datum: 17.06.98 Zeit: 16:28

	Irreg.komp.	Rauschkomp.
04.06.96 n=599	5.7 (1.17)	2.9 (0.75)
30.07.96 n=357	5.3 (1.22)	2.8 (0.78)
30.07.96 n=399	2.2 (1.07)	3.0 (0.84)



Drittes Physikalisches Institut und Abt. Phoniatrie und Pädaudiologie Universität Göttingen

17. Juni 1998

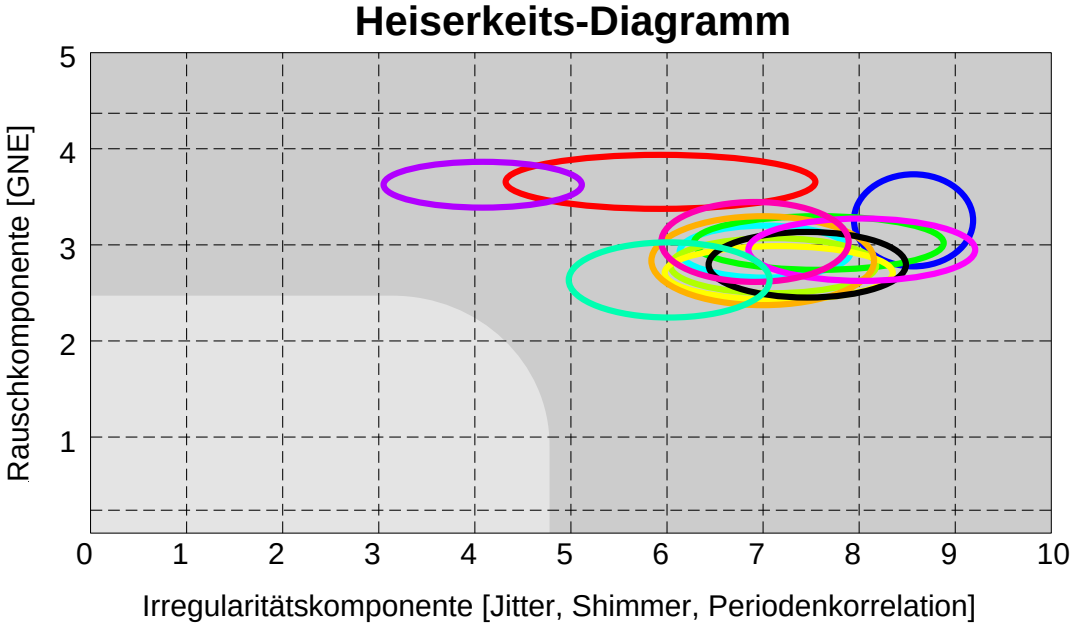
Patient 23

4/96	Zustand nach Bestrahlung der rechten Stimmlippe (85) und partieller Stimmlippenentfernung links 11/92.
8/96	Glottische Ersatzphonation ohne Schwingung der operierten Stimmlippe. Hohe Rauschkomponente. Hohe Sprechtonlage (350Hz).
3/97-7/98	Zustand nach Kontrolloperation 9/96. **Umstellung der Phonationsebene von Glottis auf Taschenfaltenstimme.** Stimmtherapie.
10/98	Zustand nach Kontrolloperation 10/98, kein Rezidiv.
6/99	Taschenfaltenstimme bei relativ geringer Irregularität. Sprechtonlage 120Hz.

Heiserkeits-Diagramm | Patient 23, m, 19

Datum: 29.06.99 Zeit: 13:03

	Irreg.komp.	Rauschkomp.
16.04.96 n=433	5.9 (1.61)	3.7 (0.28)
07.08.96 n=509	4.1 (1.03)	3.6 (0.24)
10.03.97 n=119	8.6 (0.62)	3.3 (0.48)
21.04.97 n=194	7.0 (0.89)	2.9 (0.27)
16.05.97 n=442	7.6 (1.30)	3.0 (0.28)
15.07.97 n=227	7.1 (1.07)	2.8 (0.29)
05.11.97 n=460	7.2 (1.19)	2.7 (0.28)
09.03.98 n=518	7.0 (1.16)	2.8 (0.46)
06.07.98 n=195	7.5 (1.03)	2.8 (0.34)
06.10.98 n=791	8.0 (1.18)	2.9 (0.32)
25.01.99 n=354	6.9 (0.97)	3.0 (0.42)
10.06.99 n=182	6.0 (1.05)	2.6 (0.39)



Drittes Physikalisches Institut und Abt. Phoniatrie und Pädaudiologie Universität Göttingen 29. Juni 1999

D.2. Ary-epiglottische Ersatzphonationen

Patient 24

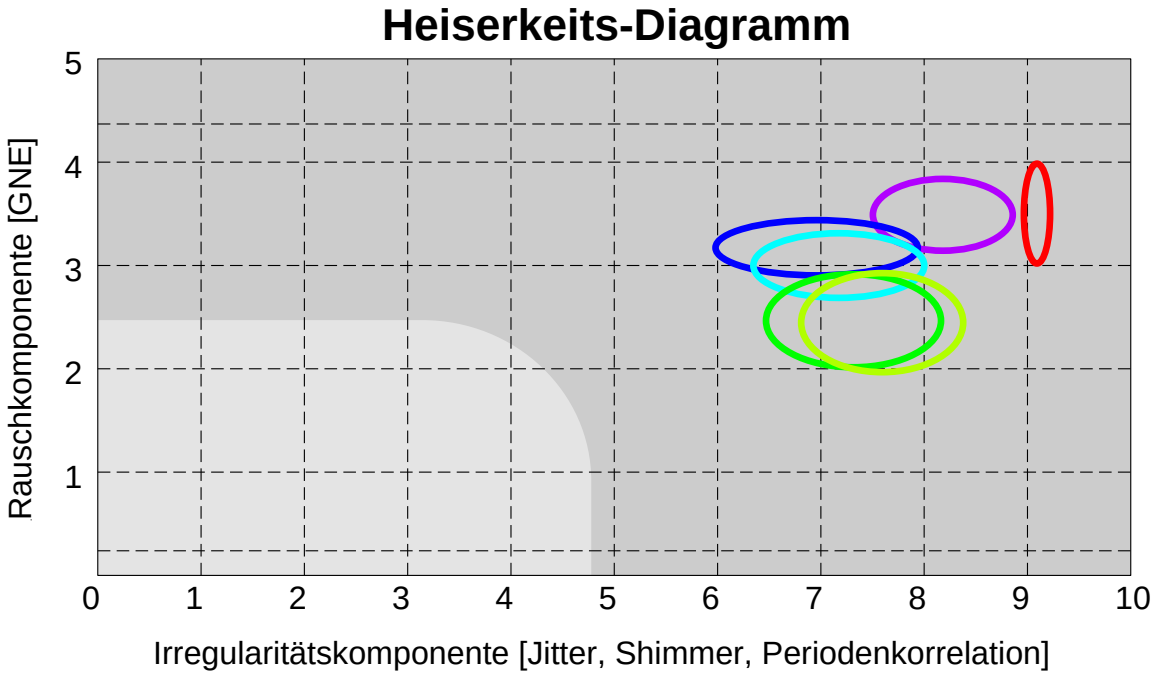
3/96 **Aphonie** (sehr hohe Rausch- und Irregularitätskomponente) nach Kehlkopfteilresektion bei Therapiebeginn.

3/96-11/96 Stetige, deutliche **Abnahme der Rauschkomponente** im Therapieverlauf.

Heiserkeits-Diagramm | Patient 24, m, 52

Datum: 17.06.98 Zeit: 16:30

	Irreg.komp.	Rauschkomp.
19.03.96 n=118	9.1 (0.13)	3.5 (0.48)
22.03.96 n=55	8.2 (0.68)	3.5 (0.35)
07.05.96 n=201	7.0 (0.98)	3.2 (0.27)
05.06.96 n=296	7.2 (0.83)	3.0 (0.31)
28.10.96 n=503	7.3 (0.85)	2.5 (0.45)
01.11.96 n=457	7.6 (0.78)	2.4 (0.48)



Drittes Physikalisches Institut und Abt. Phoniatrie und Pädaudiologie Universität Göttingen

17. Juni 1998

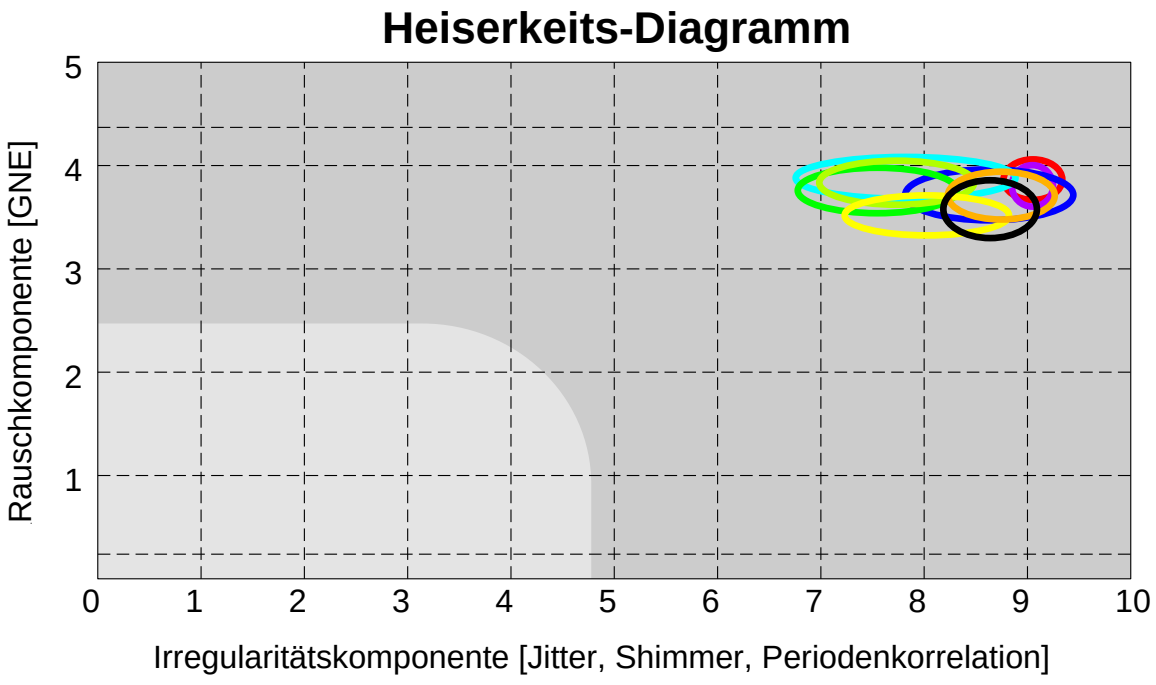
Patient 25

12/95	(rot, violett) **Aphonie.** Zustand nach Teilresektionen der Stimmlippen und Taschenfalten 8+9/95.
2/96-9/96	**Verringerung der Irregularität.**
2/97	Zustand nach Rezidiv- Operation 2/97, vordere Komissur links.
5/97-11/99	Zustand nach Kontrolloperation 4/97.

Heiserkeits-Diagramm | Patient 25, m, 43

Datum: 17.11.99 Zeit: 17:35

	Irreg.komp.	Rauschkomp.
04.12.95 n=265	9.1 (0.28)	3.9 (0.20)
20.12.95 n=123	9.0 (0.20)	3.8 (0.20)
02.02.96 n=142	8.6 (0.81)	3.7 (0.25)
01.03.96 n=217	7.8 (1.06)	3.9 (0.20)
16.04.96 n=247	7.5 (0.77)	3.8 (0.22)
02.09.96 n=172	7.7 (0.75)	3.8 (0.21)
24.02.97 n=209	8.0 (0.80)	3.5 (0.19)
15.05.97 n=137	8.7 (0.52)	3.7 (0.23)
04.11.99 n=165	8.6 (0.45)	3.6 (0.28)



Drittes Physikalisches Institut und Abt. Phoniatrie und Pädaudiologie Universität Göttingen 17. Nov. 1999

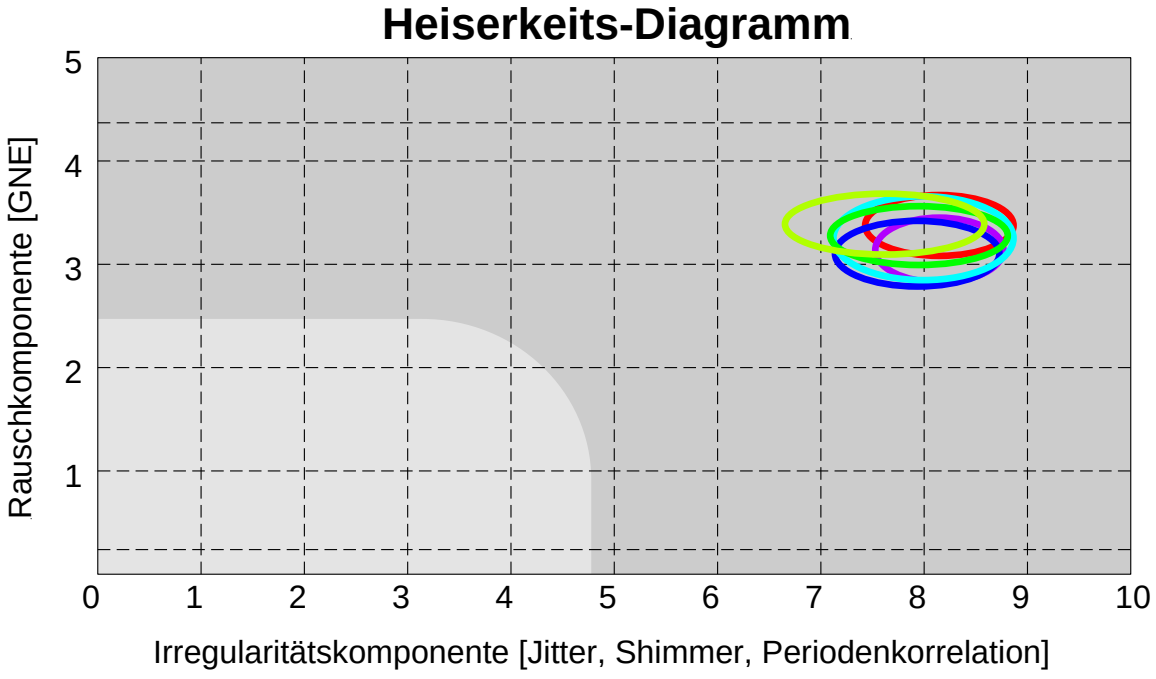
Patient 26

4/96-6/98 Zustand nach endolaryngealer Teilresektion rechts. **Stabile Lage** bei hoher Rausch und Irregularitätskomponente.

Heiserkeits-Diagramm | Patient 26, m, 49

Datum: 09.11.99 Zeit: 22:31

	Irreg.komp.	Rauschkomp.
17.04.96 n=219	8.1 (0.72)	3.4 (0.29)
23.10.96 n=236	8.1 (0.62)	3.1 (0.31)
21.05.97 n=224	7.9 (0.80)	3.1 (0.32)
19.11.97 n=261	8.0 (0.87)	3.3 (0.41)
17.06.98 n=169	7.9 (0.86)	3.3 (0.28)
13.10.99 n=316	7.6 (0.96)	3.4 (0.29)



Drittes Physikalisches Institut und Abt. Phoniatrie und Pädaudiologie Universität Göttingen 9. Nov. 1999

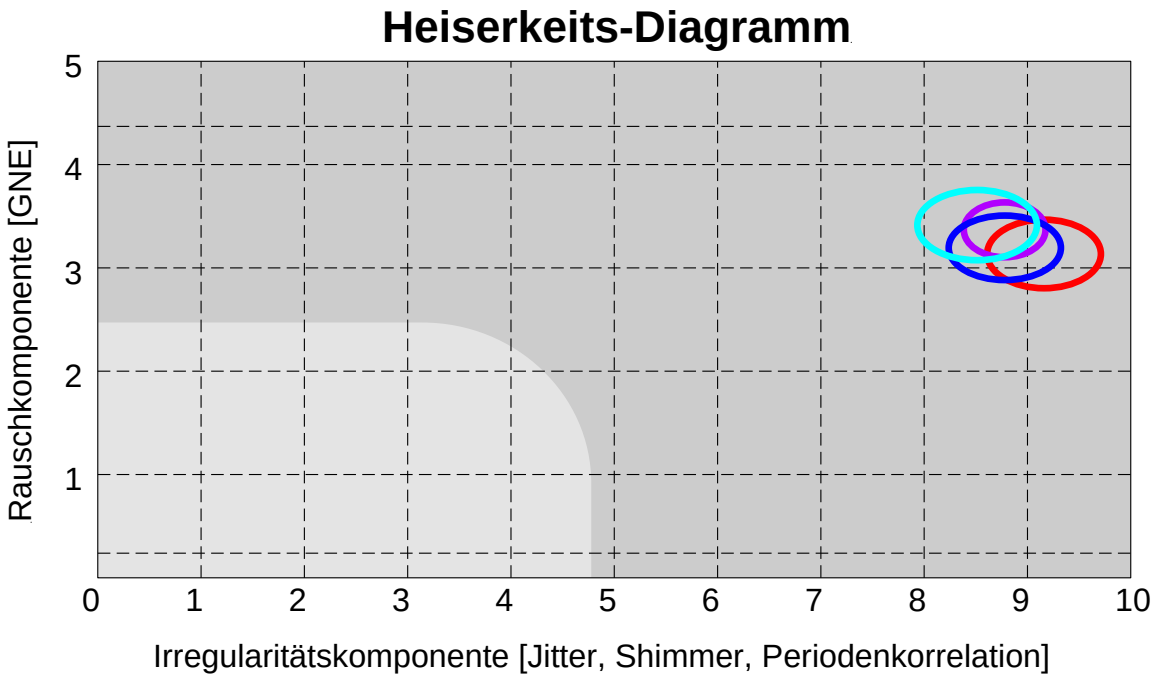
Patient 27

2/96-9/96 Zustand nach partieller Stimmbandentfernung links. **Leichte Verringerung der Irregularität bei leichtem Anstieg der Rauschkomponente**. Allgemein sehr hohe Rausch- und Irregularitätskomponente.

Heiserkeits-Diagramm | Patient 27, m, 25

Datum: 17.06.98 Zeit: 16:41

	Irreg.komp.	Rauschkomp.
05.02.96 n=279	9.2 (0.55)	3.1 (0.33)
27.02.96 n=87	8.8 (0.40)	3.4 (0.27)
12.03.96 n=247	8.8 (0.54)	3.2 (0.31)
30.09.96 n=101	8.5 (0.58)	3.4 (0.34)



Drittes Physikalisches Institut und Abt. Phoniatrie und Pädaudiologie Universität Göttingen

17. Juni 1998

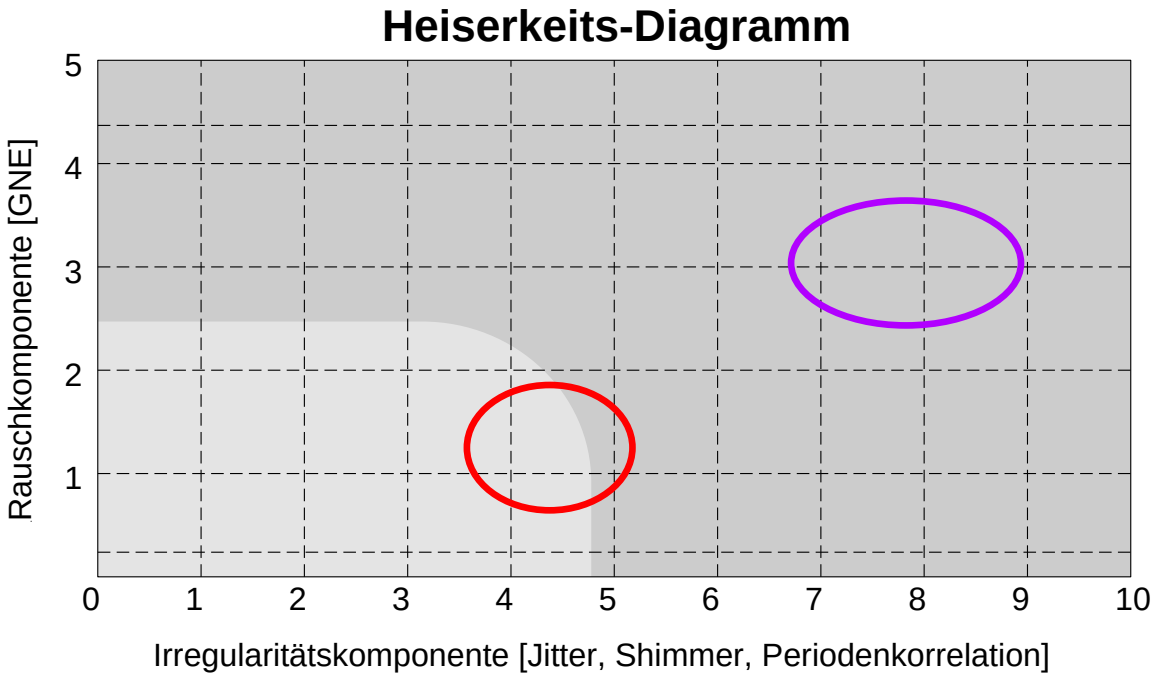
Patient 28

3/96 Zustand bei T1b- Tumor auf der rechten Stimmlippe.
6/97 Zustand nach beidseitiger Stimmbandentfernung, Taschenfaltenresektion rechts, postoperativ.

Heiserkeits-Diagramm | Patient 28, m, 27

Datum: 17.06.98 Zeit: 16:30

	Irreg.komp.	Rauschkomp.
12.03.96 n=65	4.4 (0.80)	1.3 (0.61)
18.06.97 n=200	7.8 (1.11)	3.0 (0.60)

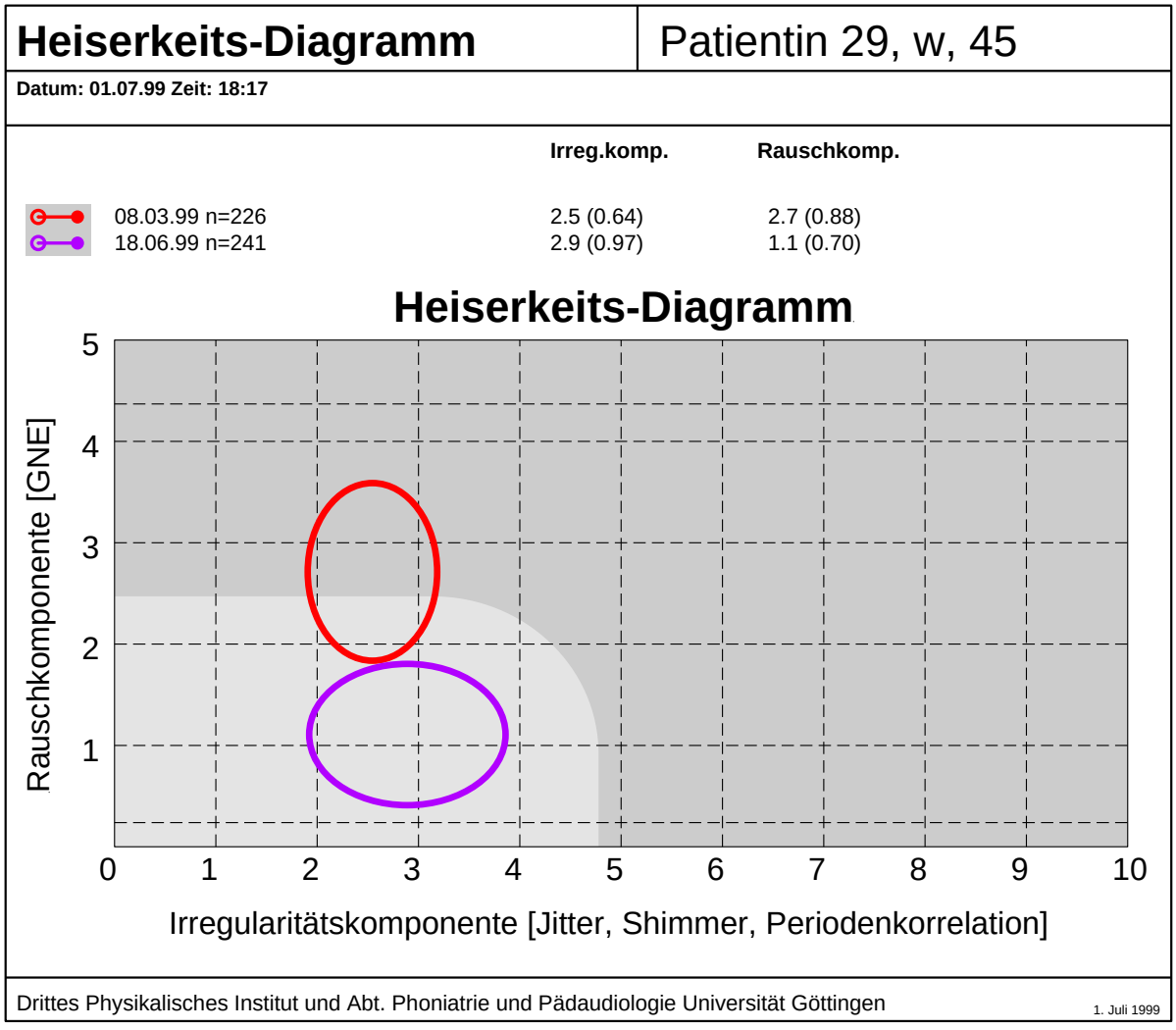


Drittes Physikalisches Institut und Abt. Phoniatrie und Pädaudiologie Universität Göttingen 17. Juni 1998

D.3. Funktionelle Stimmstörungen - hypofunktionelle Dysphonie

Patientin 29

3/99 Verdacht auf psychosomatische Dysphonie. **Erhöhte Rauschkomponente**
6/99 Abschluss Stimmtherapie. **Normalbefund**.



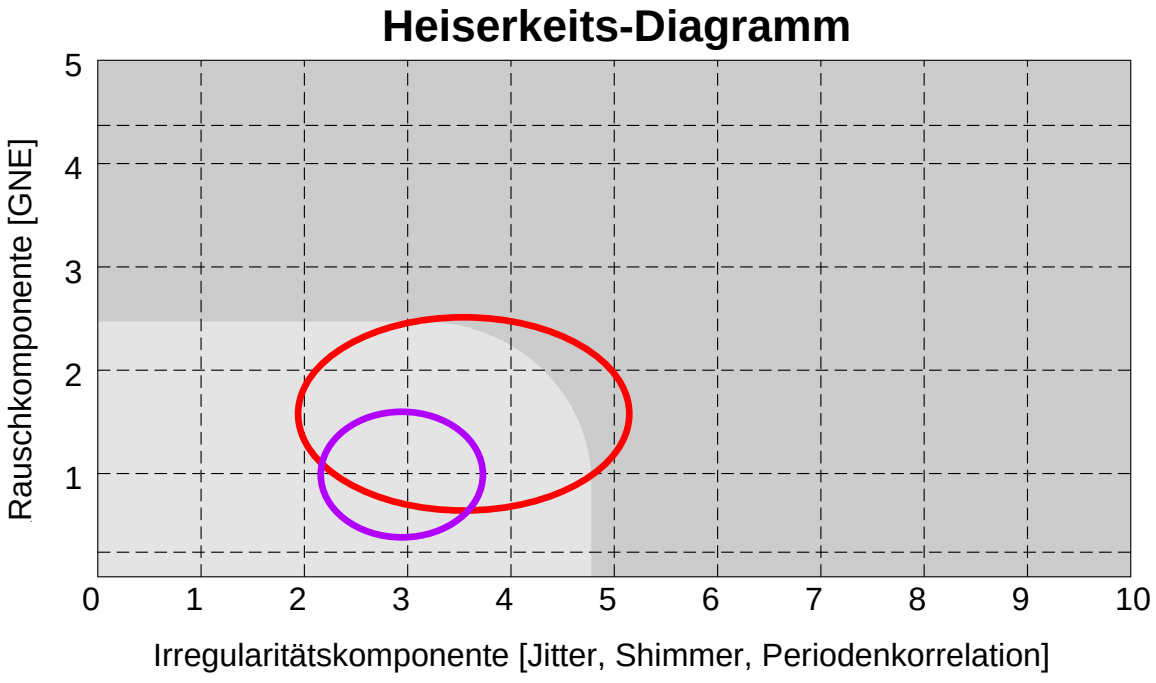
Patientin 30

1/97 Hypofunktionelle Dysphonie. Lage im Normalbereich. **Starke Variation**.
2/97 **Normalbefund** nach Stimmtherapie.

Heiserkeits-Diagramm | Patientin 30, w, 68

Datum: 17.06.98 Zeit: 16:36

	Irreg.komp.	Rauschkomp.
28.01.97 n=308	3.5 (1.60)	1.6 (0.94)
13.02.97 n=318	2.9 (0.79)	1.0 (0.61)



Drittes Physikalisches Institut und Abt. Phoniatrie und Pädaudiologie Universität Göttingen

17. Juni 1998

Patient 31

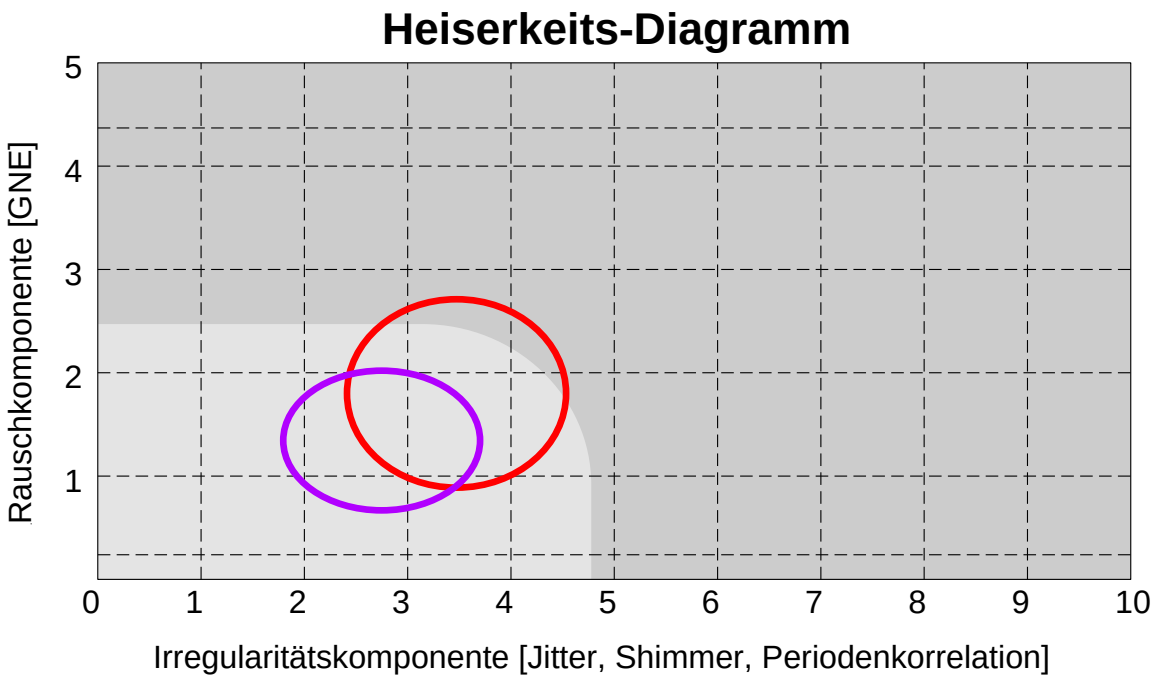
3/98 Hypofunktionelle Dysphonie. Beginn Stimmtherapie. Hochgradige Innenohr- Schwerhörigkeit.

3/98 Zustand nach hypofunktioneller Dysphonie. **Verringerte Rausch- und Irregularitätskomponente.**

Heiserkeits-Diagramm | Patient 31, m, 66

Datum: 17.06.98 Zeit: 16:52

	Irreg.komp.	Rauschkomp.
09.03.98 n=374	3.5 (1.06)	1.8 (0.91)
31.03.98 n=109	2.7 (0.95)	1.3 (0.68)



Drittes Physikalisches Institut und Abt. Phoniatrie und Pädaudiologie Universität Göttingen

17. Juni 1998

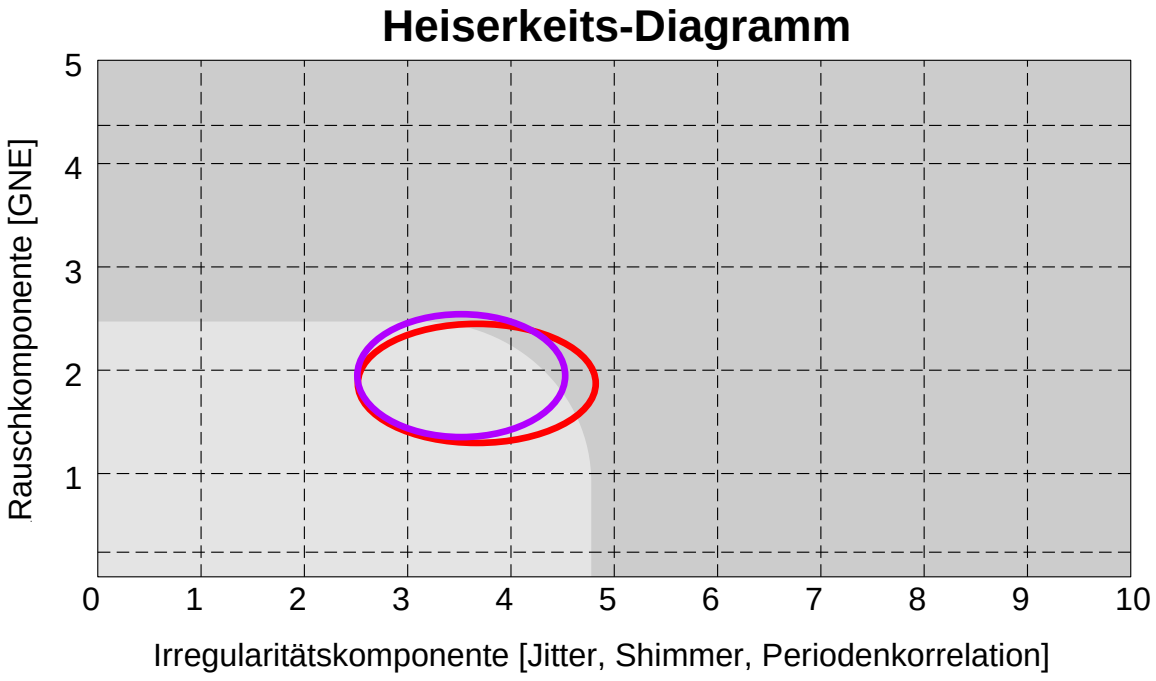
Patient 32

12/98 Hypofunktionelle Dysphonie.
12/98 Abschluss Stimmrehabilitation. **Nahezu unverändert** am Rande des Normalbereichs.

Heiserkeits-Diagramm | Patient 32, m, 67

Datum: 29.12.98 Zeit: 04:57

	Irreg.komp.	Rauschkomp.
01.12.98 n=378	3.7 (1.15)	1.9 (0.58)
22.12.98 n=430	3.5 (1.01)	1.9 (0.59)



Drittes Physikalisches Institut und Abt. Phoniatrie und Pädaudiologie Universität Göttingen 29. Dez. 1998

Patient 33

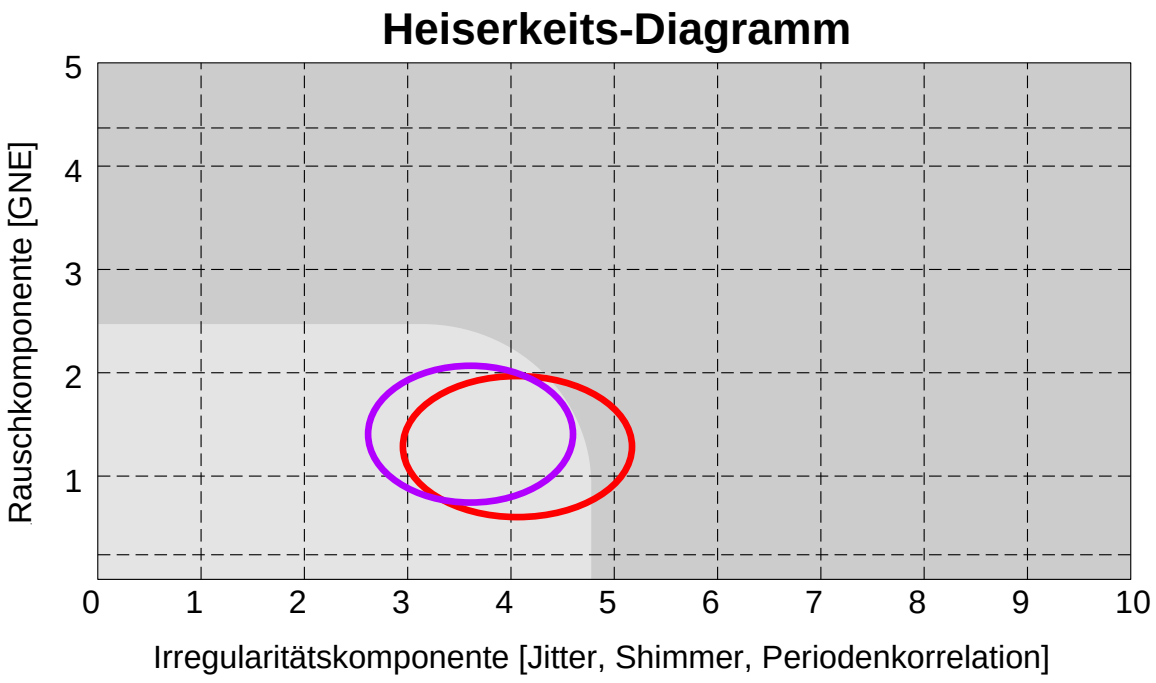
3/98 Hypofunktionelle Dysphonie. Beginn Stimmtherapie. **Erhöhte Irregularität.**

3/98 Abschluss Stimmrehabilitation. **Leichte Verringerung der Irregularitätskomponente.**

Heiserkeits-Diagramm | Patient 33, m, 27

Datum: 26.02.99 Zeit: 17:39

	Irreg.komp.	Rauschkomp.
09.03.98 n=169	4.1 (1.11)	1.3 (0.68)
24.03.98 n=119	3.6 (0.99)	1.4 (0.66)



Drittes Physikalisches Institut und Abt. Phoniatrie und Pädaudiologie Universität Göttingen 26. Feb. 1999

D.4. Patienten mit Zysten auf den Stimmbändern

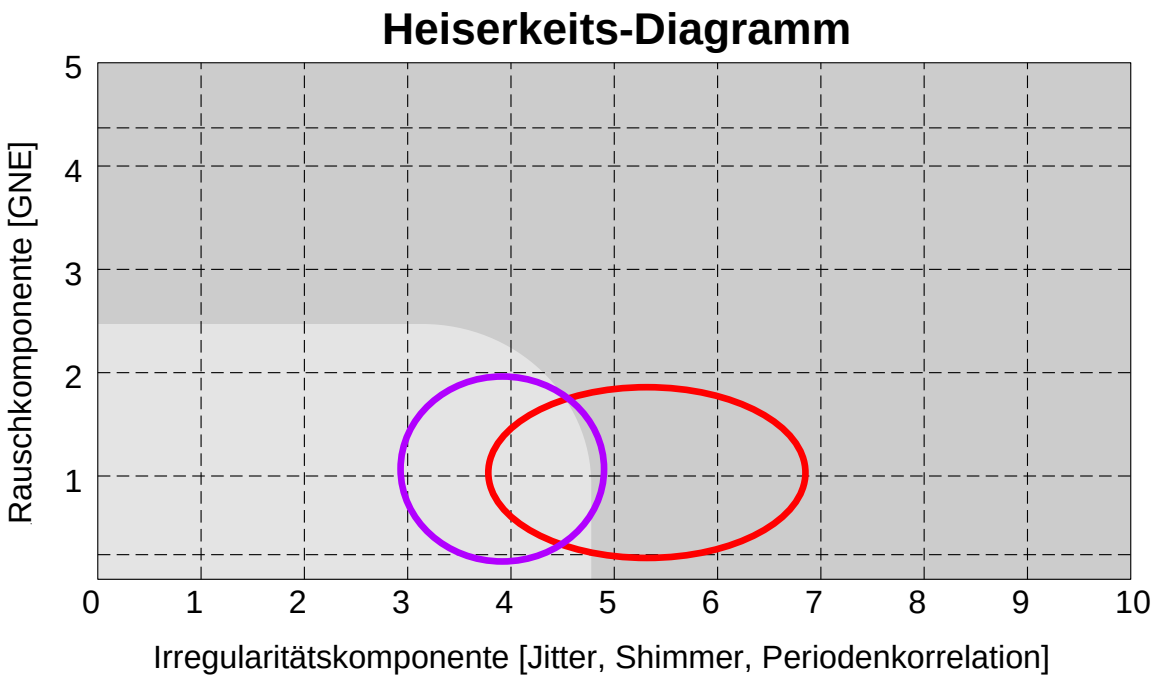
Patientin 34

8/96 Stimmlippenzyste links. **Erhöhte Irregularität.**
4/98 Zustand nach Entfernung der Zyste 11/97 und Wundheilung. **Deutliche Verringerung der Irregularitätskomponente.**

Heiserkeits-Diagramm | Patientin 34, w, 43

Datum: 17.06.98 Zeit: 16:32

	Irreg.komp.	Rauschkomp.
30.08.96 n=212	5.3 (1.54)	1.0 (0.83)
01.04.98 n=115	3.9 (0.98)	1.1 (0.90)



Drittes Physikalisches Institut und Abt. Phoniatrie und Pädaudiologie Universität Göttingen 17. Juni 1998

Patientin 35

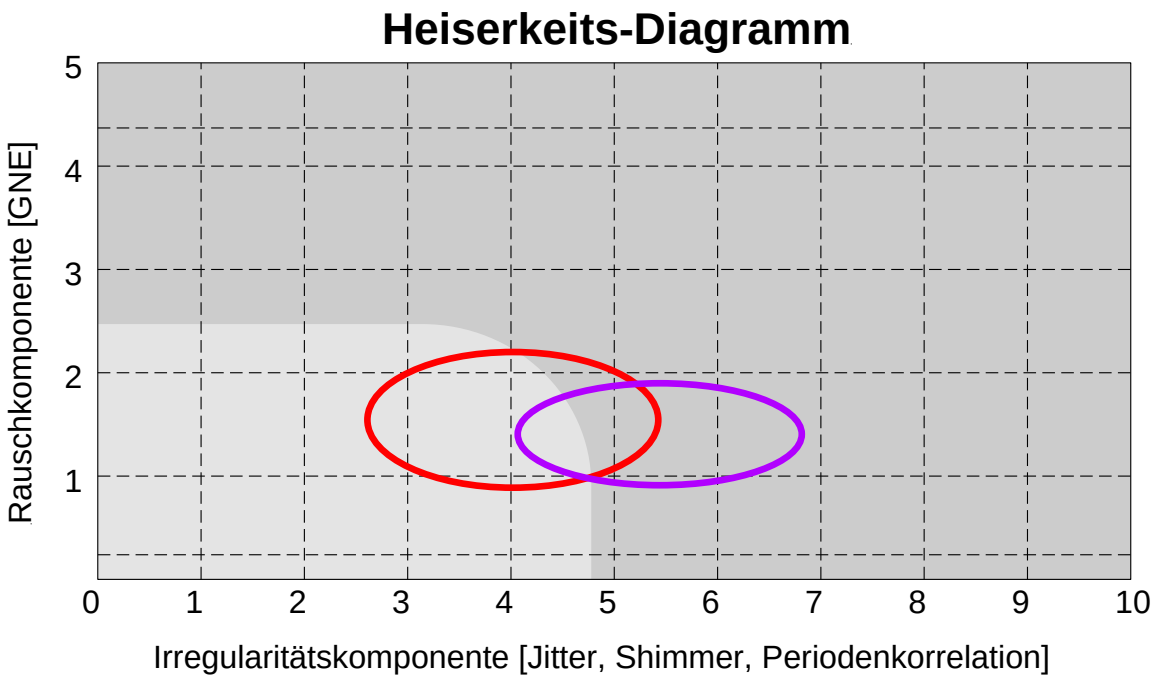
8/96 Intravokale Stimmlippenzyste rechts. **Lage am Rande des Normalbereichs**

3/97 Zustand nach Entfernung der Zyste 12/96. **Deutliche Erhöhung der Irregularitätskomponente.**

Heiserkeits-Diagramm | Patientin 35, w, 60

Datum: 17.06.98 Zeit: 16:46

	Irreg.komp.	Rauschkomp.
06.08.96 n=186	4.0 (1.41)	1.5 (0.66)
04.03.97 n=202	5.4 (1.38)	1.4 (0.49)



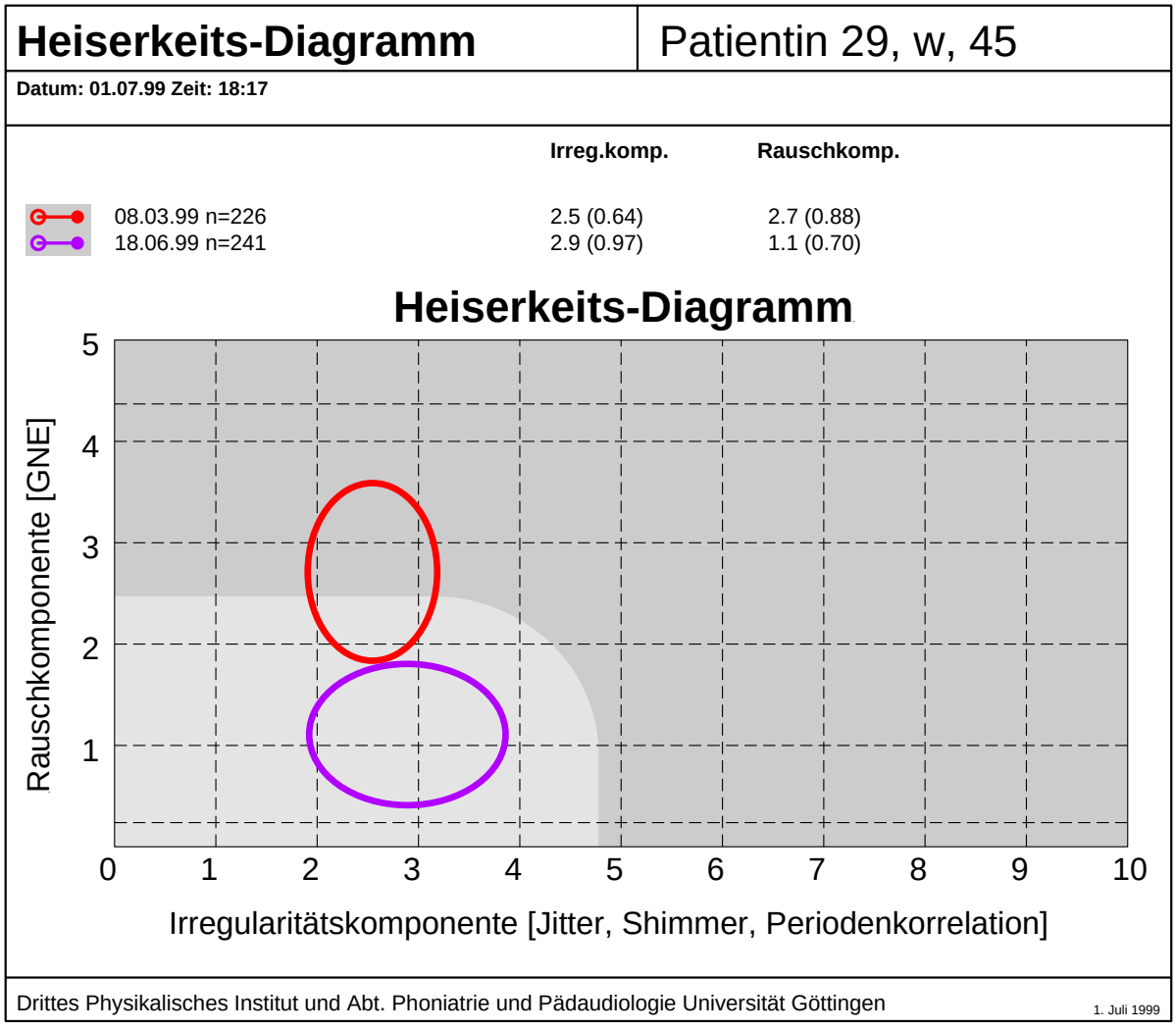
Drittes Physikalisches Institut und Abt. Phoniatrie und Pädaudiologie Universität Göttingen

17. Juni 1998

D.3. Funktionelle Stimmstörungen - hypofunktionelle Dysphonie

Patientin 29

3/99 Verdacht auf psychosomatische Dysphonie. **Erhöhte Rauschkomponente**
6/99 Abschluss Stimmtherapie. **Normalbefund**.



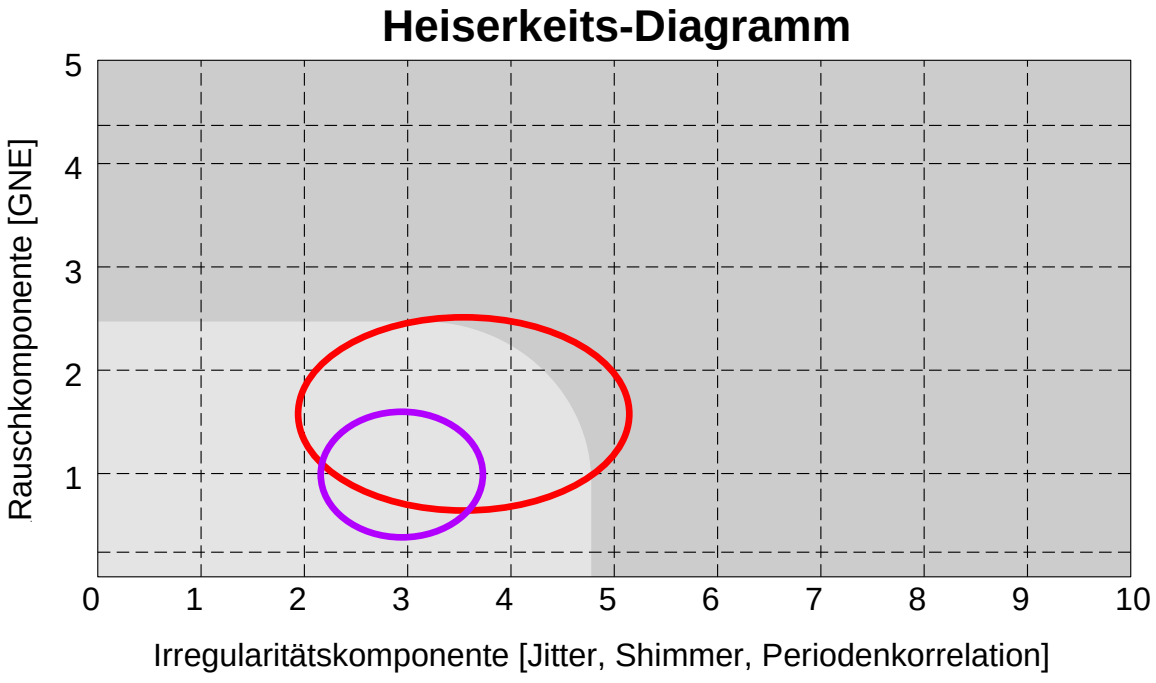
Patientin 30

1/97 Hypofunktionelle Dysphonie. Lage im Normalbereich. **Starke Variation**.
2/97 **Normalbefund** nach Stimmtherapie.

Heiserkeits-Diagramm | Patientin 30, w, 68

Datum: 17.06.98 Zeit: 16:36

	Irreg.komp.	Rauschkomp.
28.01.97 n=308	3.5 (1.60)	1.6 (0.94)
13.02.97 n=318	2.9 (0.79)	1.0 (0.61)



Drittes Physikalisches Institut und Abt. Phoniatrie und Pädaudiologie Universität Göttingen

17. Juni 1998

Patient 31

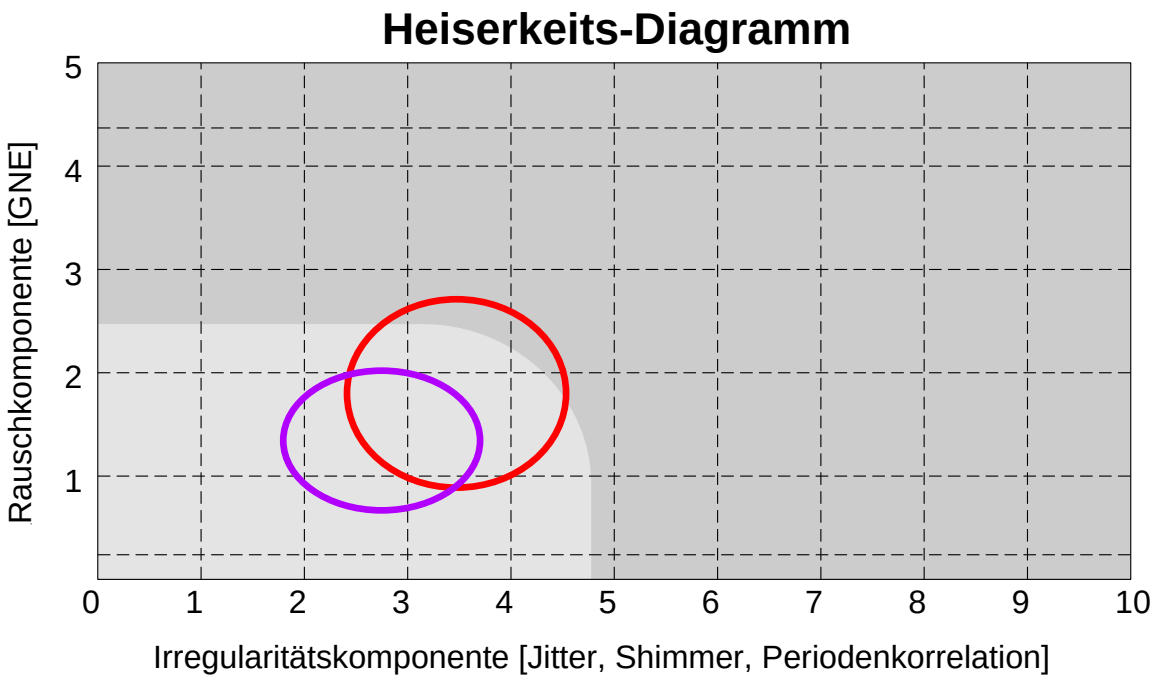
3/98 Hypofunktionelle Dysphonie. Beginn Stimmtherapie. Hochgradige Innenohr- Schwerhörigkeit.

3/98 Zustand nach hypofunktioneller Dysphonie. **Verringerte Rausch- und Irregularitätskomponente.**

Heiserkeits-Diagramm | Patient 31, m, 66

Datum: 17.06.98 Zeit: 16:52

	Irreg.komp.	Rauschkomp.
09.03.98 n=374	3.5 (1.06)	1.8 (0.91)
31.03.98 n=109	2.7 (0.95)	1.3 (0.68)



Drittes Physikalisches Institut und Abt. Phoniatrie und Pädaudiologie Universität Göttingen

17. Juni 1998

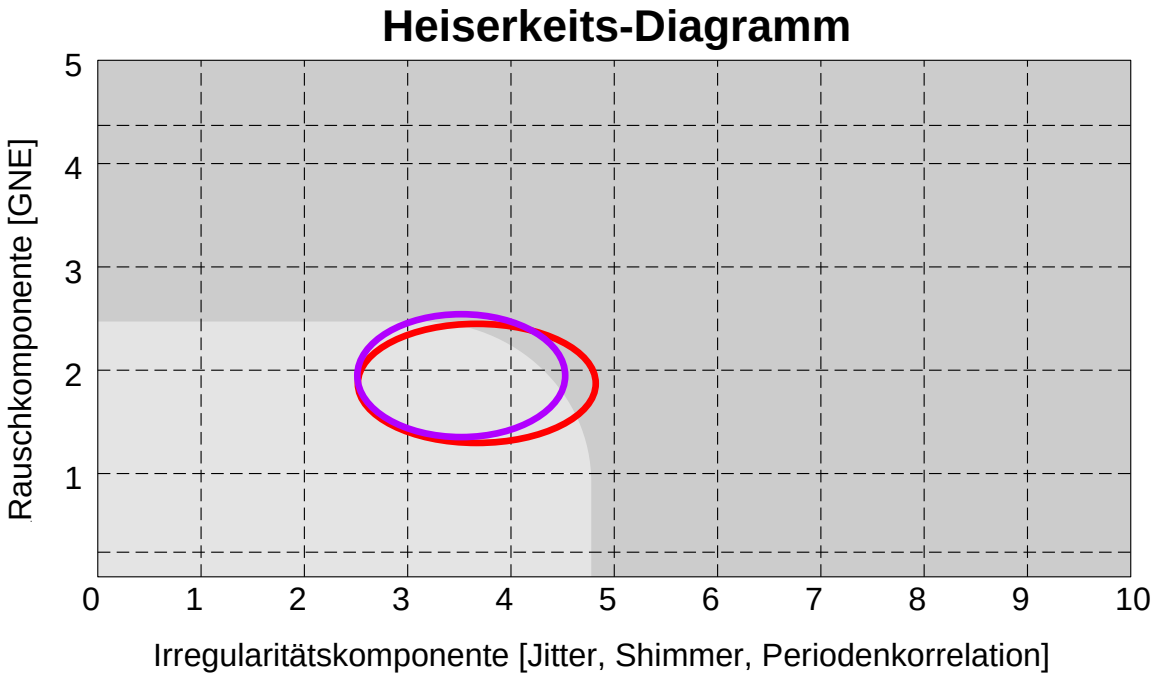
Patient 32

12/98 Hypofunktionelle Dysphonie.
12/98 Abschluss Stimmrehabilitation. **Nahezu unverändert** am Rande des Normalbereichs.

Heiserkeits-Diagramm | Patient 32, m, 67

Datum: 29.12.98 Zeit: 04:57

	Irreg.komp.	Rauschkomp.
01.12.98 n=378	3.7 (1.15)	1.9 (0.58)
22.12.98 n=430	3.5 (1.01)	1.9 (0.59)



Drittes Physikalisches Institut und Abt. Phoniatrie und Pädaudiologie Universität Göttingen 29. Dez. 1998

Patient 33

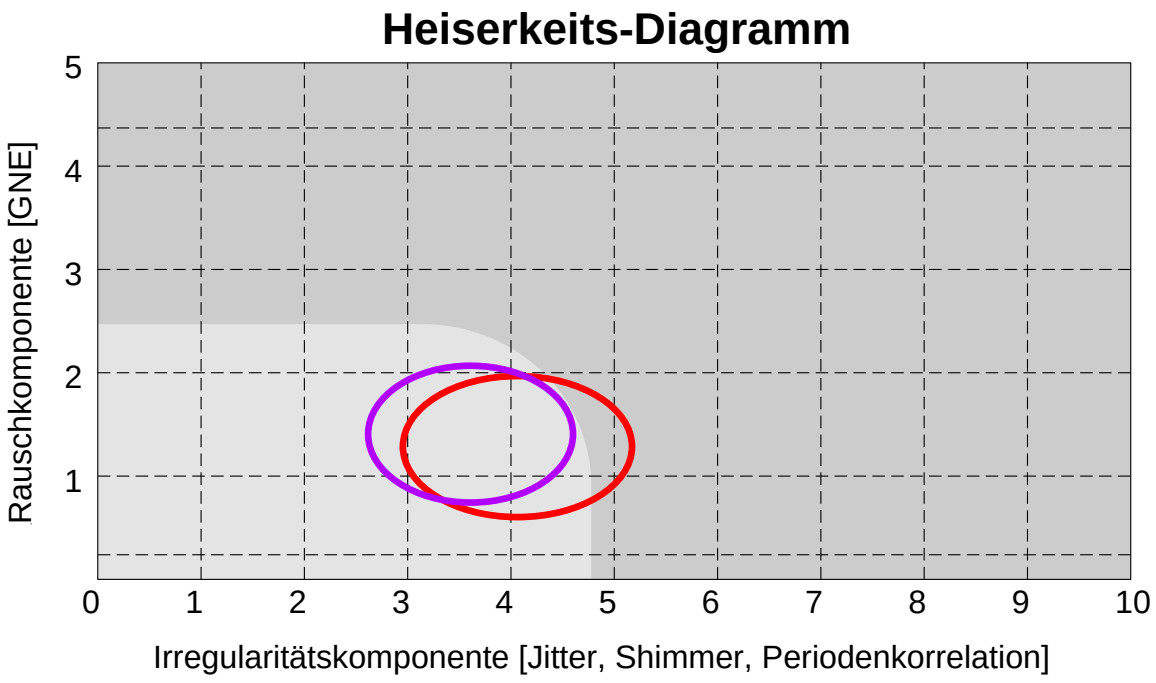
3/98 Hypofunktionelle Dysphonie. Beginn Stimmtherapie. **Erhöhte Irregularität.**

3/98 Abschluss Stimmrehabilitation. **Leichte Verringerung der Irregularitätskomponente.**

Heiserkeits-Diagramm | Patient 33, m, 27

Datum: 26.02.99 Zeit: 17:39

	Irreg.komp.	Rauschkomp.
09.03.98 n=169	4.1 (1.11)	1.3 (0.68)
24.03.98 n=119	3.6 (0.99)	1.4 (0.66)



Drittes Physikalisches Institut und Abt. Phoniatrie und Pädaudiologie Universität Göttingen 26. Feb. 1999

D.4. Patienten mit Zysten auf den Stimmbändern

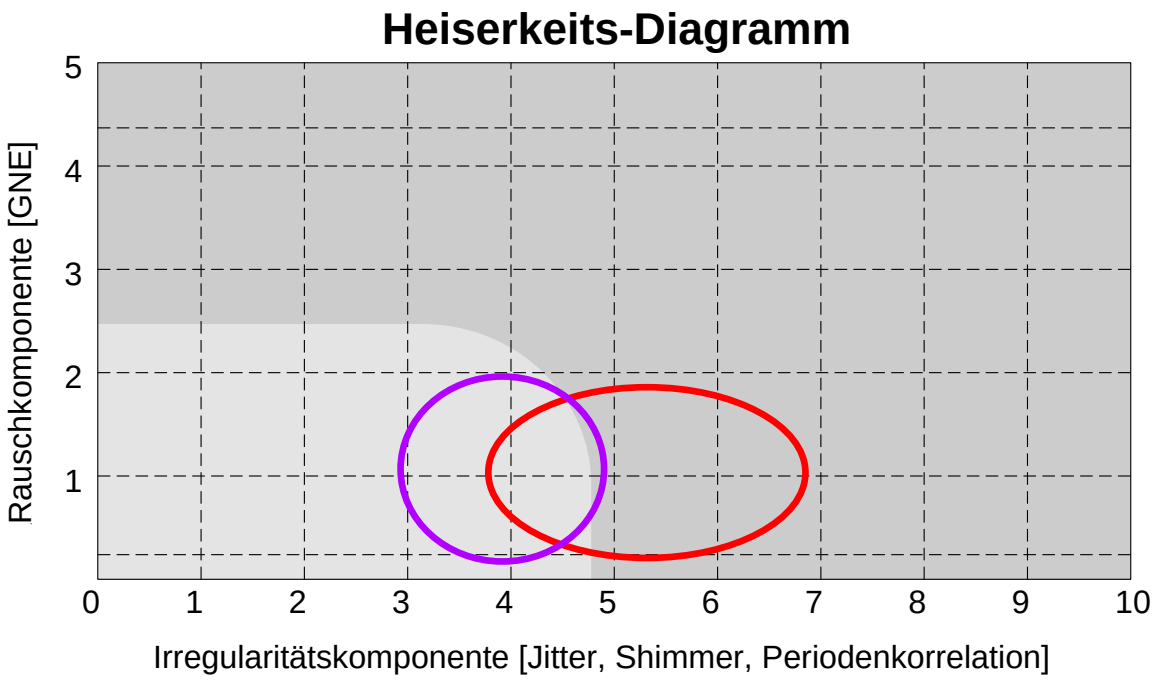
Patientin 34

8/96 Stimmlippenzyste links. **Erhöhte Irregularität.**
4/98 Zustand nach Entfernung der Zyste 11/97 und Wundheilung. **Deutliche Verringerung der Irregularitätskomponente.**

Heiserkeits-Diagramm | Patientin 34, w, 43

Datum: 17.06.98 Zeit: 16:32

	Irreg.komp.	Rauschkomp.
30.08.96 n=212	5.3 (1.54)	1.0 (0.83)
01.04.98 n=115	3.9 (0.98)	1.1 (0.90)



Drittes Physikalisches Institut und Abt. Phoniatrie und Pädaudiologie Universität Göttingen 17. Juni 1998

Patientin 35

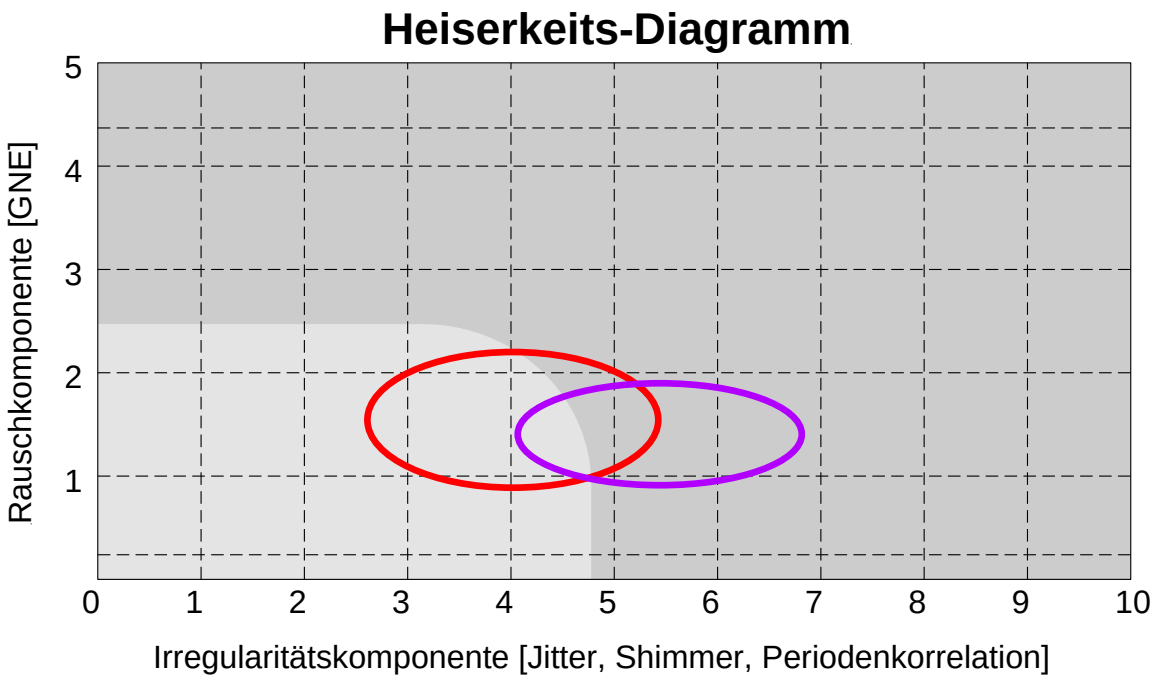
8/96 Intravokale Stimmlippenzyste rechts. **Lage am Rande des Normalbereichs**

3/97 Zustand nach Entfernung der Zyste 12/96. **Deutliche Erhöhung der Irregularitätskomponente.**

Heiserkeits-Diagramm | Patientin 35, w, 60

Datum: 17.06.98 Zeit: 16:46

	Irreg.komp.	Rauschkomp.
06.08.96 n=186	4.0 (1.41)	1.5 (0.66)
04.03.97 n=202	5.4 (1.38)	1.4 (0.49)



Drittes Physikalisches Institut und Abt. Phoniatrie und Pädaudiologie Universität Göttingen

17. Juni 1998

Patientin 36

6/96 Intravokale Stimmlippenzyste rechts. Kontaktreaktion links. **Lage am Rande des Normalbereichs**

9/96 Zustand nach Entfernung der Zyste 7/96. **Leichte Erhöhung der Irregularitätskomponente.**

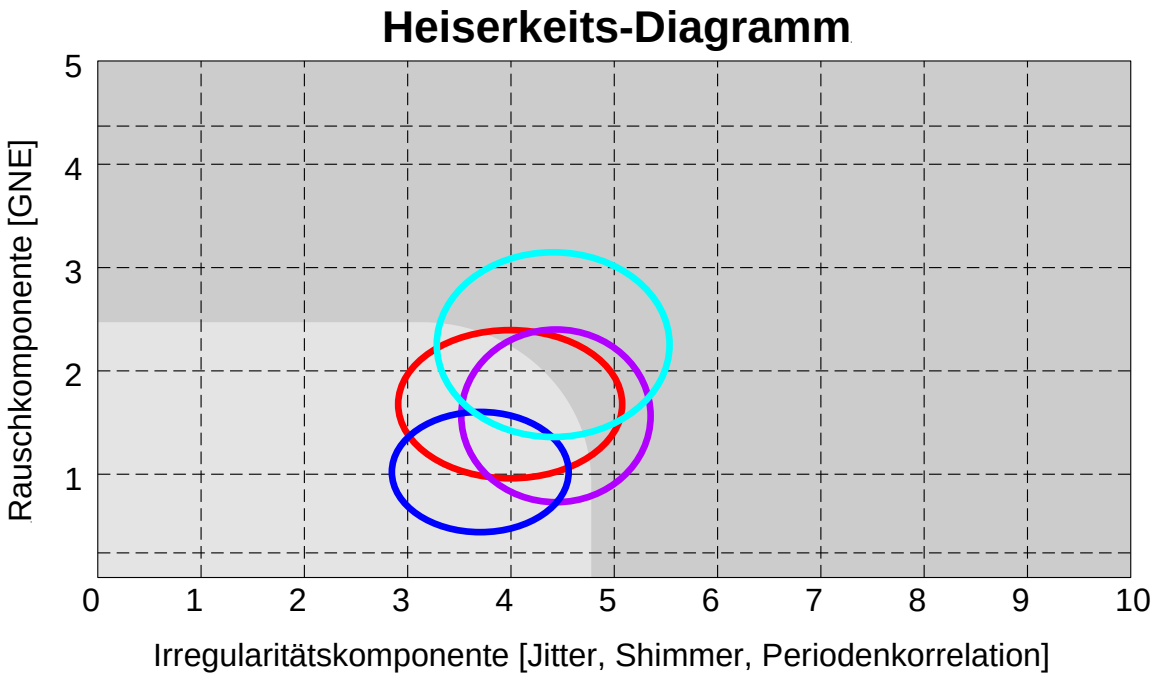
11/96 Verzögerte Wundheilung. **Deutliche Verringerung der Rausch- und Irregularitätskomponente.**

7/97 Verdacht auf Rezidiv der intravokalen Zyste rechts. **Deutliche Erhöhung der Rausch- und Irregularitätskomponente.**

Heiserkeits-Diagramm | Patientin 36, w, 77

Datum: 17.06.98 Zeit: 16:47

	Irreg.komp.	Rauschkomp.
11.06.96 n=211	4.0 (1.08)	1.7 (0.72)
23.09.96 n=105	4.4 (0.92)	1.6 (0.84)
28.11.96 n=242	3.7 (0.86)	1.0 (0.58)
16.07.97 n=292	4.4 (1.13)	2.3 (0.89)



Drittes Physikalisches Institut und Abt. Phoniatrie und Pädaudiologie Universität Göttingen

17. Juni 1998

D.5. Patientin mit Reinke-Ödem

Patientin 37

9/96 Reinke-Ödeme auf beiden Stimmlippen. **Sehr hohe Rausch- und hohe Irregularitätskomponente.**

10/96 Zustand nach Entfernung der Ödeme 9/96 und Wundheilung. **Deutliche Verringerung der Rausch- und Irregularitätskomponente.**

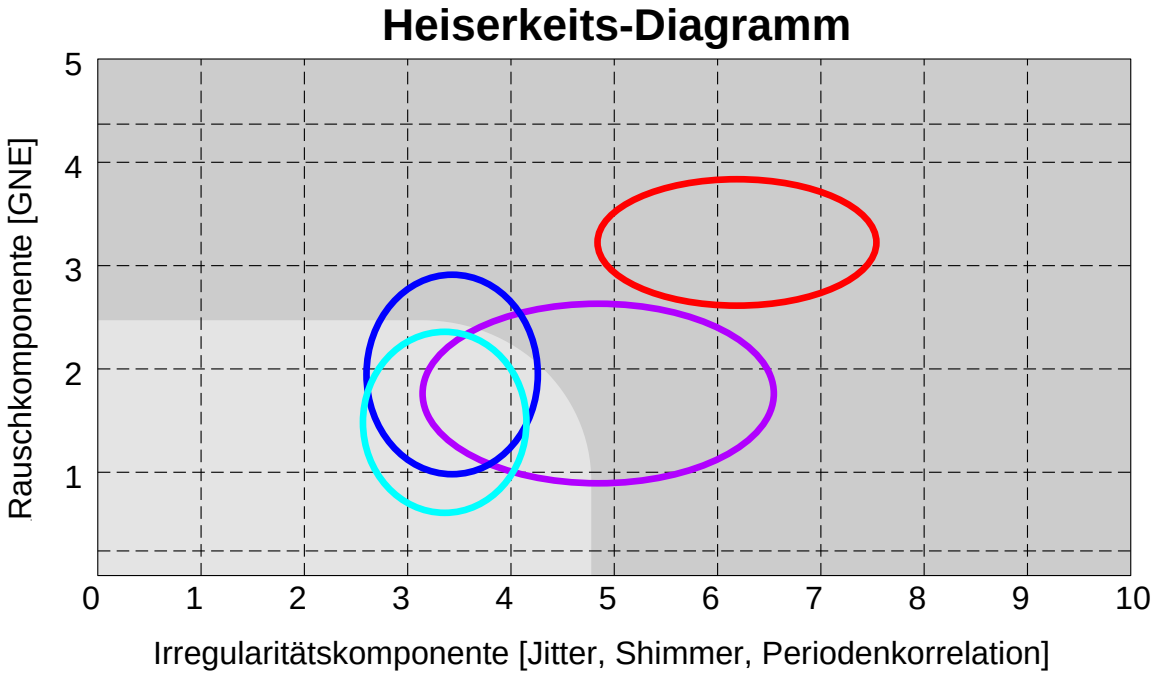
12/96 **Nochmals deutliche Verringerung der Irregularitätskomponente sowie deren Variation.**

2/97 Leichte Verringerung der Rauschkomponente. **Jetzt Lage im Normalbereich.**

Heiserkeits-Diagramm | Patientin 37, w, 34

Datum: 17.06.98 Zeit: 16:32

	Irreg.komp.	Rauschkomp.
12.09.96 n=109	6.2 (1.35)	3.2 (0.61)
21.10.96 n=194	4.8 (1.70)	1.8 (0.87)
13.12.96 n=508	3.4 (0.83)	1.9 (0.97)
18.02.97 n=398	3.4 (0.79)	1.5 (0.88)



Drittes Physikalisches Institut und Abt. Phoniatrie und Pädaudiologie Universität Göttingen

17. Juni 1998

D.6. Patientin mit Stimmlippenknötchen

Patientin 38

3/98 Zustand bei hypofunktioneller Dysphonie und Stimmlippenknötchen. **erhöhte Rausch- und Irregularitätskomponente.**

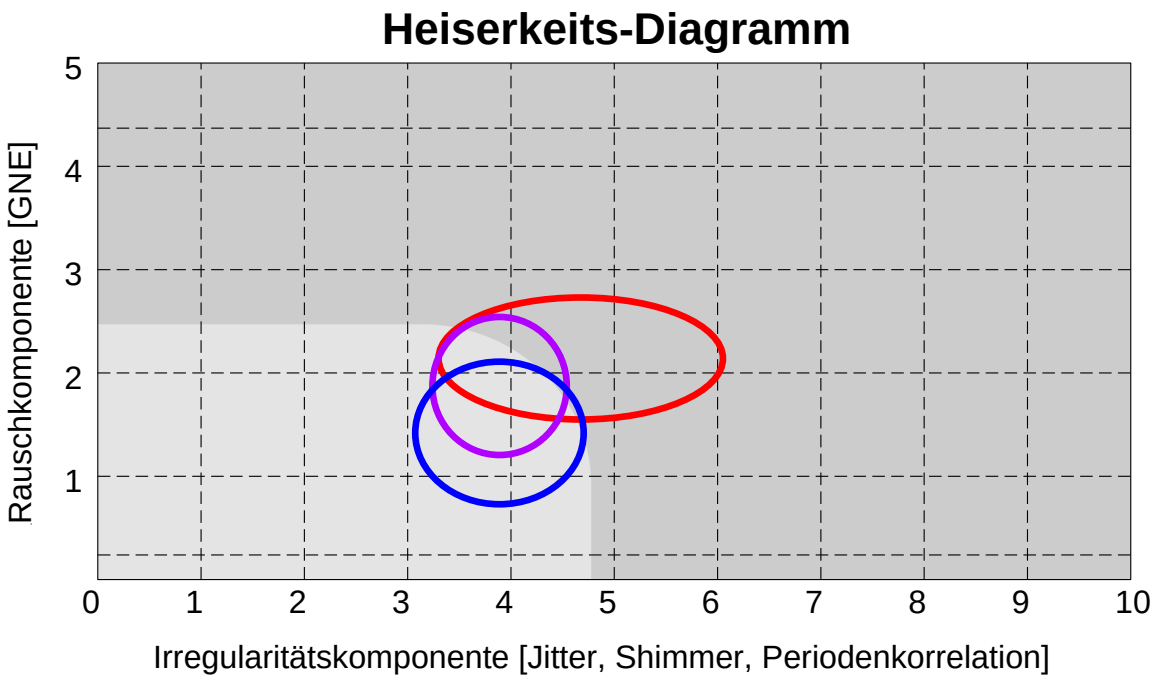
3/98 Abschluss Stimmrehabilitation. **Deutliche Verringerung der Irregularitätskomponente und deren Variabilität. Leichte Verringerung der Rauschkomponente.**

6/98 Normale Kehlkopf"—funktion, evtl. Knötchenansätze. **Nochmalige leichte Verringerung der Rauschkomponente, Lage im Normalbereich.**

Heiserkeits-Diagramm | Patientin 38, w, 78

Datum: 17.06.98 Zeit: 16:35

	Irreg.komp.	Rauschkomp.
10.03.98 n=103	4.7 (1.38)	2.1 (0.59)
25.03.98 n=94	3.9 (0.65)	1.9 (0.67)
03.06.98 n=159	3.9 (0.82)	1.4 (0.69)



Drittes Physikalisches Institut und Abt. Phoniatrie und Pädaudiologie Universität Göttingen

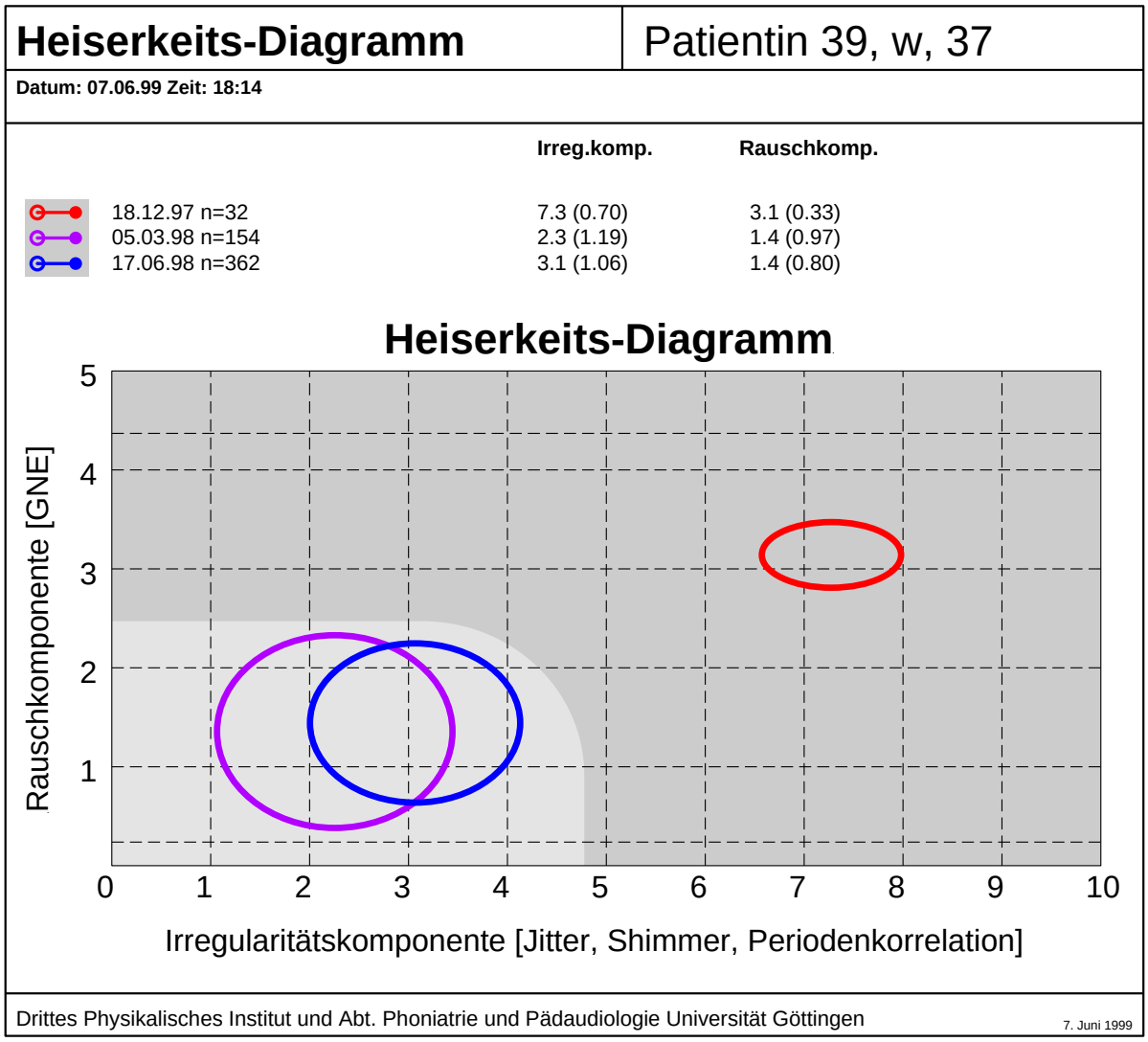
17. Juni 1998

D.7. Patientin mit Lähmung des Recurrens-Nerven

Bei den hier gezeigten Patienten mit Lähmungen des Nervus recurrens oder des Vagusnerven tritt relativ häufig eine Wiederbeweglichkeit der Stimmlippe auf. Dieser Verlauf kann im Heiserkeits-Diagramm gut verfolgt werden: Die Lähmung geht meistens mit einem hohen Rauschanteil und hoher Irregularität einher. Mit der Lähmung verschwindet dieses Muster. Die Stimmen liegen wieder im oder nahe beim Normalbereich des Heiserkeits-Diagramms. Auch durch Medianverlagerung der Stimmlippe wird oft eine Verbesserung der Stimmqualität erzielt, die auch im Heiserkeits-Diagramm sichtbar wird.

Patientin 39

12/97 Recurrenslähmung rechts. Stimmlippenstellung paramedian. Zustand nach Kropfentfernung 12/97. **Hohe Rausch- und Irregularitätskomponente.**
3/98 Beinahe komplette, spontane Remission. **Lage im Normalbereich.**
6/98 Normale Kehlkopf"—funktion.



Patientin 40

8/98 Recurrenslähmung beidseitig. Zwei Phonationsebenen: glottische Phonation und Taschenfaltenphonation. **Sehr hohe Rausch- und Irregularitätskomponente.**

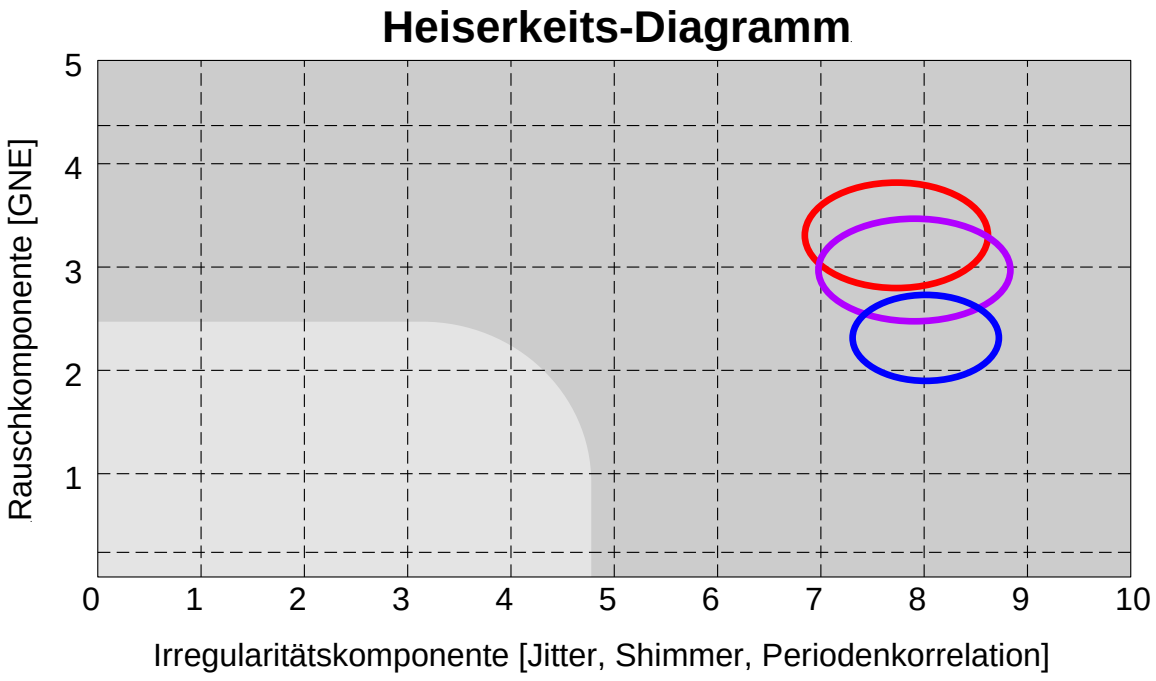
3/99 Zustand nach Glottiserweiterung. Beginn der Stimmrehabilitation. **Leichte Verringerung der Rauschkomponente.**

4/99 **Taschenfaltenphonation.** Abschluss Stimmrehabilitation. **Nochmalige Verringerung der Rauschkomponente.**

Heiserkeits-Diagramm | Patientin 40, w, 38

Datum: 04.05.99 Zeit: 05:10

	Irreg.komp.	Rauschkomp.
05.08.98 n=274	7.7 (0.89)	3.3 (0.51)
29.03.99 n=262	7.9 (0.93)	3.0 (0.50)
30.04.99 n=210	8.0 (0.71)	2.3 (0.42)



Drittes Physikalisches Institut und Abt. Phoniatrie und Pädaudiologie Universität Göttingen

4. Mai 1999

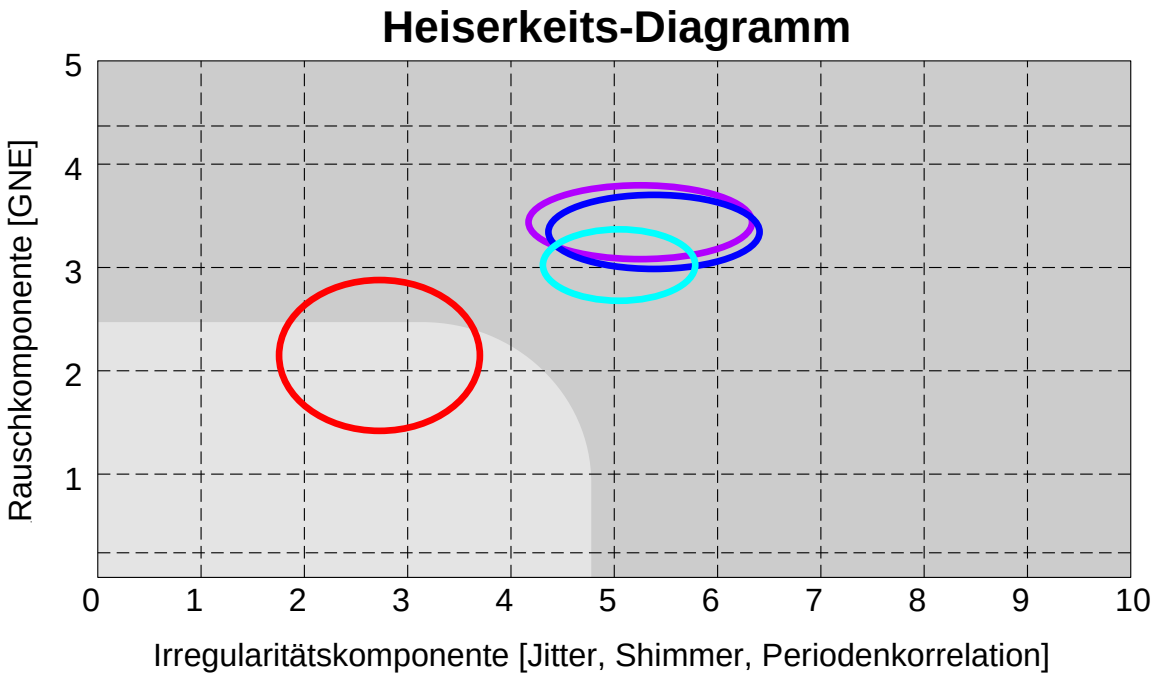
Patientin 41

8/98 Recurrenslähmung links. **Erhöhte Rauschkomponente.**
1/99 Zustand nach Glottiserweiterung rechts 11/98. Restmobilität links. **Sehr hohe Rausch- und hohe Irregularitätskomponente.**
4/99 Kontrolle nach Abschluss der Wundheilung. **Lage unverändert.**
7/99 Kontrolle nach Abschluss der Stimmrehabilitation. **Leichte Verringerung der Rauschkomponente.**

Heiserkeits-Diagramm | Patientin 41, w, 47

Datum: 19.07.99 Zeit: 18:25

	Irreg.komp.	Rauschkomp.
17.08.98 n=194	2.7 (0.97)	2.1 (0.73)
15.01.99 n=125	5.3 (1.08)	3.4 (0.36)
15.04.99 n=131	5.4 (1.02)	3.3 (0.36)
08.07.99 n=213	5.0 (0.74)	3.0 (0.35)



Drittes Physikalisches Institut und Abt. Phoniatrie und Pädaudiologie Universität Göttingen 19. Juli 1999

Patientin 42

8/98 Recurrenslähmung links. Zentrale Dysphonie bei Multipler Sklerose. **Sehr hohe Irregularitäts- und Rauschkomponente.**

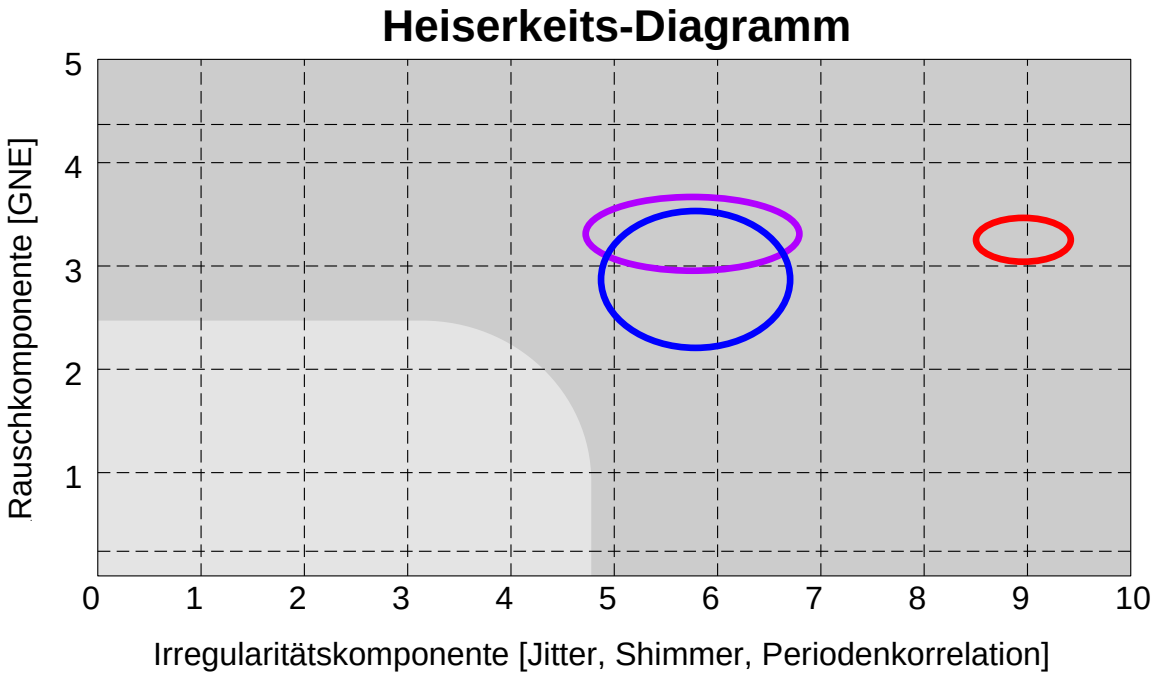
8/98 Zustand nach Medianverlagerung. **Deutliche Verringerung der Irregularität.**

10/98 **Zusätzliche leichte Verringerung der Rauschkomponente.**

Heiserkeits-Diagramm | Patientin 42, w, 48

Datum: 27.10.98 Zeit: 16:04

	Irreg.komp.	Rauschkomp.
17.08.98 n=20	9.0 (0.46)	3.3 (0.21)
20.08.98 n=76	5.8 (1.03)	3.3 (0.36)
09.10.98 n=36	5.8 (0.92)	2.9 (0.66)

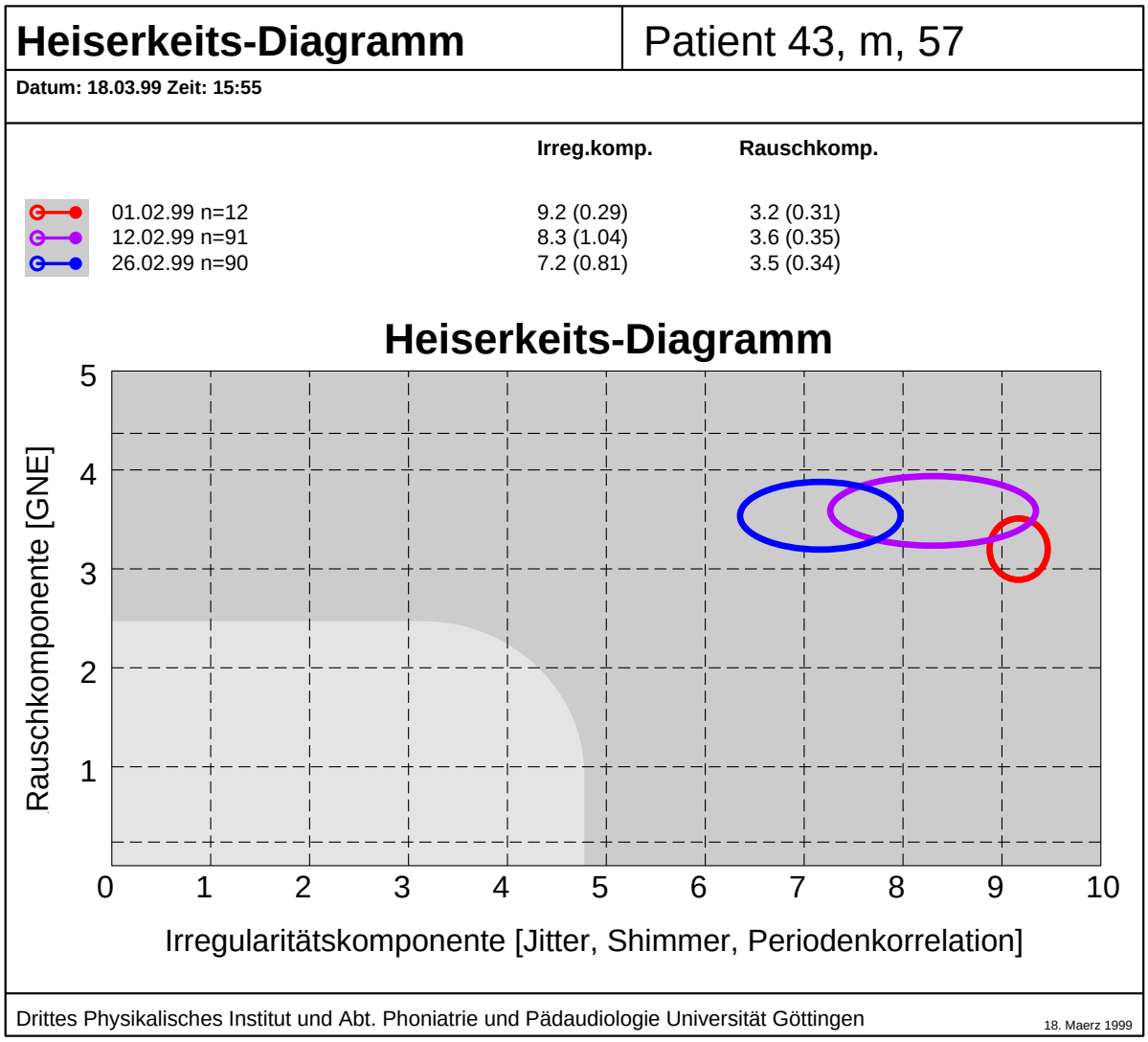


Drittes Physikalisches Institut und Abt. Phoniatrie und Pädaudiologie Universität Göttingen

27. Okt. 1998

Patient 43

2/99 (Seit längerem) Recurrenslähmung beidseitig. **Sehr hohe Irregularität, hohe Rauschkomponente.**
2/99 Kontrolle während der Stimmrehabilitation. **Verringerung der Irregularität, leichte Erhöhung der Rauschkomponente.**
2/99 **Weitere deutliche Verringerung der Irregularität.**



Patient 44

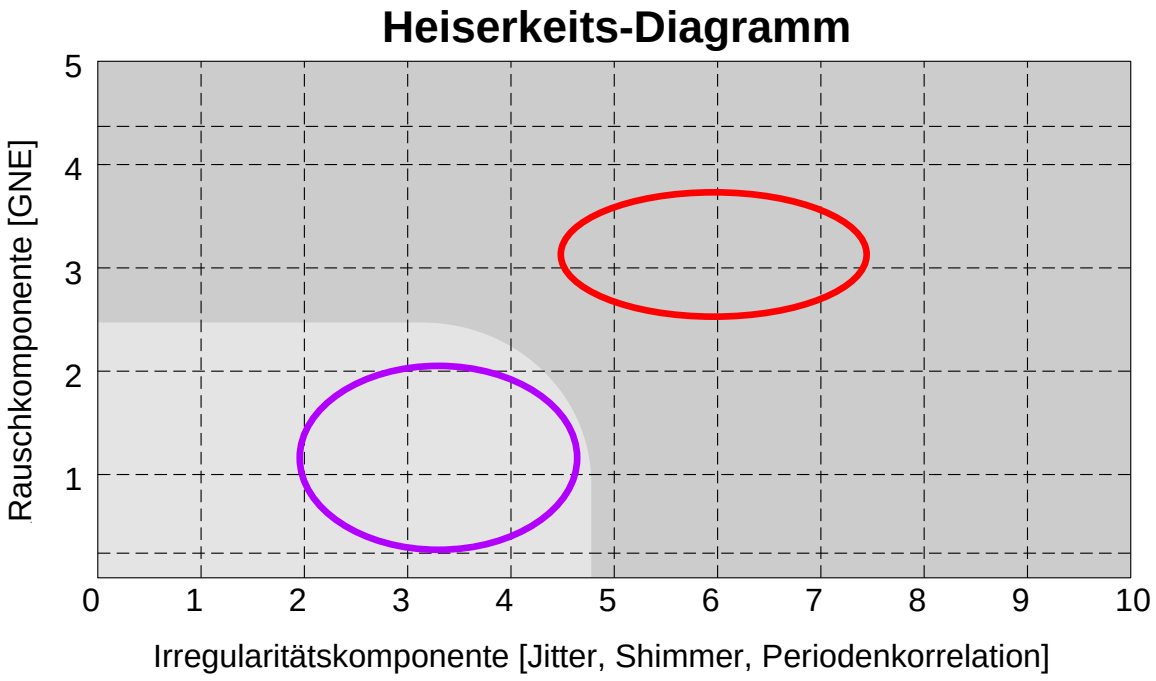
8/96 Recurrenslähmung rechts. Zustand nach Kropfentfernung rechts. **Sehr hohe Rausch- und erhöhte Irregularitätskomponente.**

11/96 Zustand nach Brustkorberöffnung. **Lage im Normalbereich.**

Heiserkeits-Diagramm | Patient 44, m, 49

Datum: 17.06.98 Zeit: 16:44

	Irreg.komp.	Rauschkomp.
30.08.96 n=255	6.0 (1.48)	3.1 (0.60)
08.11.96 n=257	3.3 (1.34)	1.2 (0.89)



Drittes Physikalisches Institut und Abt. Phoniatrie und Pädaudiologie Universität Göttingen 17. Juni 1998

D.8. Patienten mit Lähmung des Nervus Vagus

Patientin 45

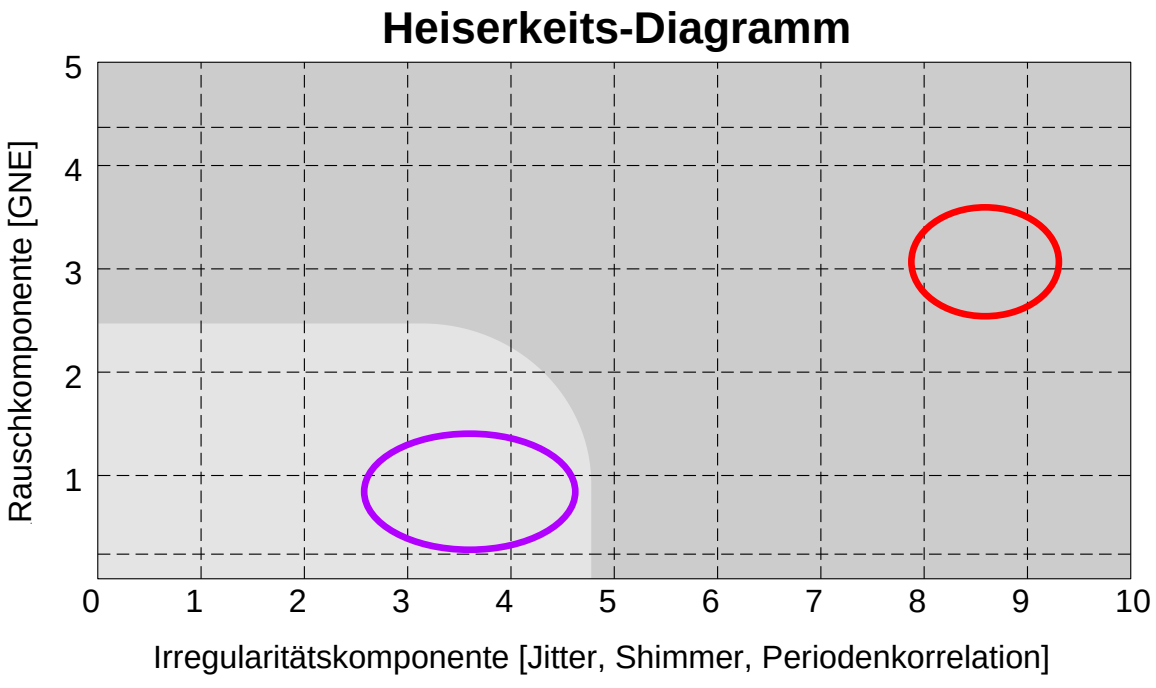
9/96 Stimmlippenstillstand links. Evtl. Vaguslähmung links. **Hohe Rausch- und sehr hohe Irregularitätskomponente.**

1/97 Zustand nach Vaguslähmung. Jetzt Leukoplakie links. **Lage im Normalbereich.**

Heiserkeits-Diagramm | Patientin 45, w, 51

Datum: 17.06.98 Zeit: 16:47

	Irreg.komp.	Rauschkomp.
06.09.96 n=60	8.6 (0.71)	3.1 (0.53)
09.01.97 n=299	3.6 (1.02)	0.8 (0.56)



Drittes Physikalisches Institut und Abt. Phoniatrie und Pädaudiologie Universität Göttingen

17. Juni 1998

Patientin 46

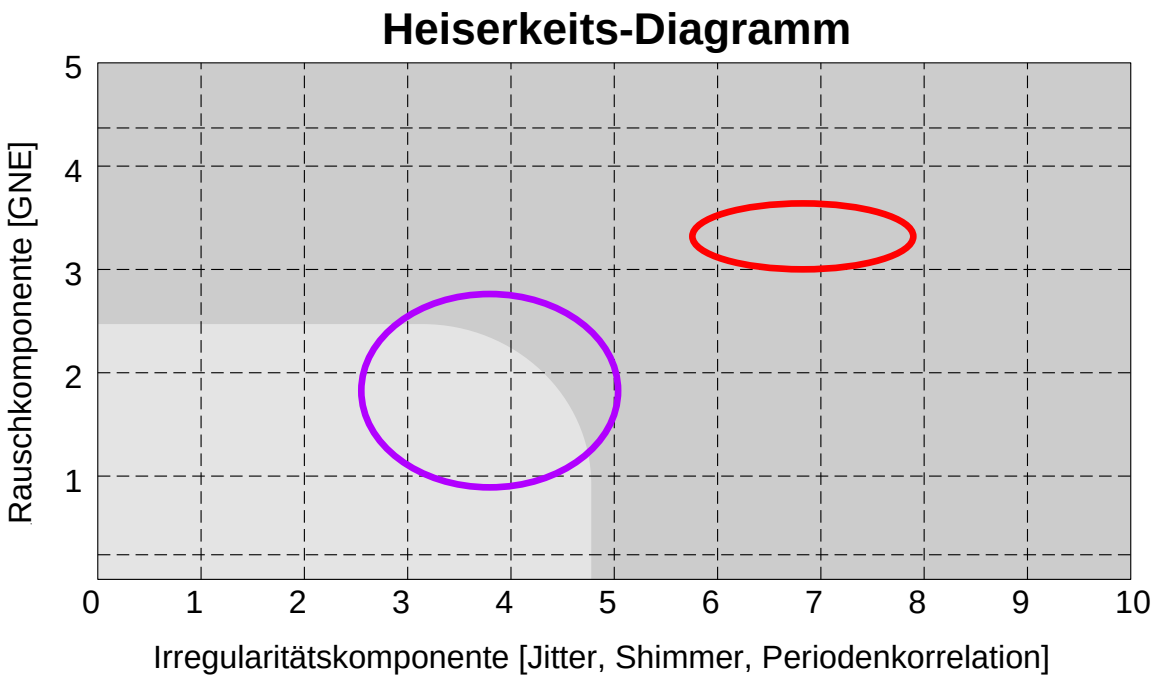
4/98 Vaguslähmung links. **Hohe Rausch- und Irregularitätskomponente.**

9/98 Normalbefund. Zustand nach Vaguslähmung. **Lage am Rande des Normalbereichs.**

Heiserkeits-Diagramm | Patientin 46, w, 39

Datum: 23.10.98 Zeit: 13:37

	Irreg.komp.	Rauschkomp.
23.04.98 n=262	6.8 (1.07)	3.3 (0.32)
29.09.98 n=275	3.8 (1.24)	1.8 (0.94)



Drittes Physikalisches Institut und Abt. Phoniatrie und Pädaudiologie Universität Göttingen 23. Okt. 1998

Patientin 47

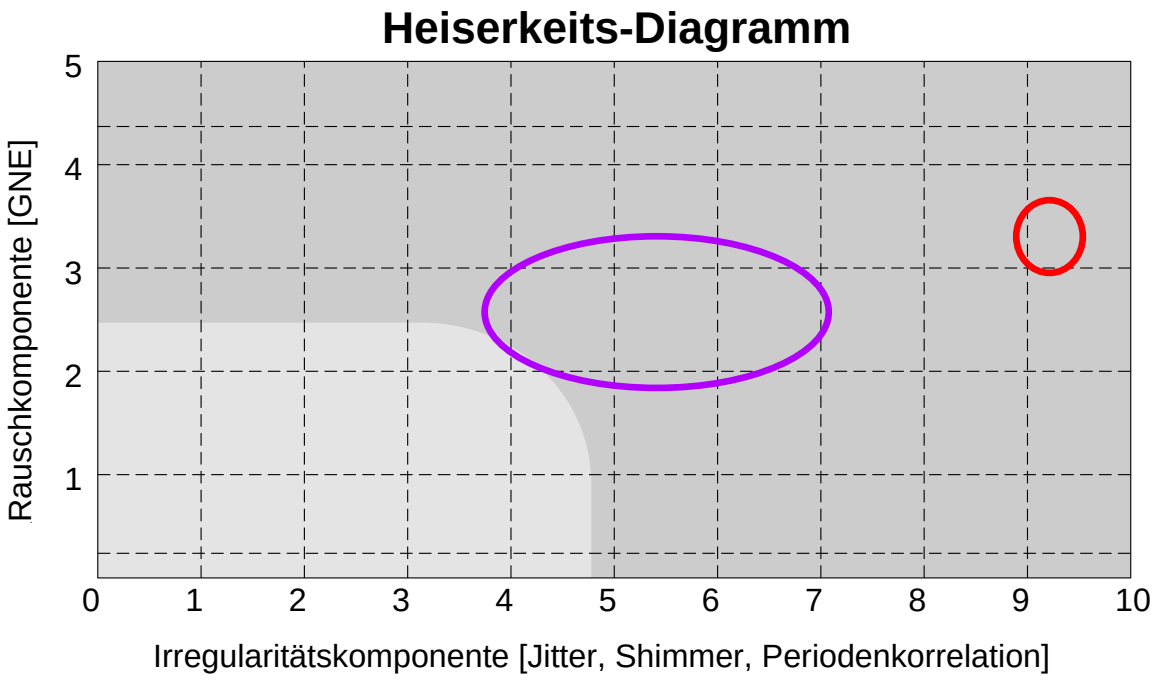
1/96 Vaguslähmung rechts. **Sehr hohe Rausch- und Irregularitätskomponente.**

11/96 Zustand nach Kehlkopferöffnung und Medianverlagerung der rechten Stimmlippe. **Deutliche Verringerung der Rausch- und Irregularitätskomponente.**

Heiserkeits-Diagramm | Patientin 47, w, 24

Datum: 17.06.98 Zeit: 16:29

	Irreg.komp.	Rauschkomp.
10.01.96 n=77	9.2 (0.32)	3.3 (0.35)
20.11.96 n=228	5.4 (1.67)	2.6 (0.73)



Drittes Physikalisches Institut und Abt. Phoniatrie und Pädaudiologie Universität Göttingen 17. Juni 1998

Patientin 48

7/97 Stimmlippenfixation paramedian und Stimmlippenhämatom links. Verdacht auf Intubationstrauma. Zustand nach Kropfentfernung 6/97. **Leicht erhöhter Rauschanteil. Relativ hohe Varianz der Rauschkomponente.**

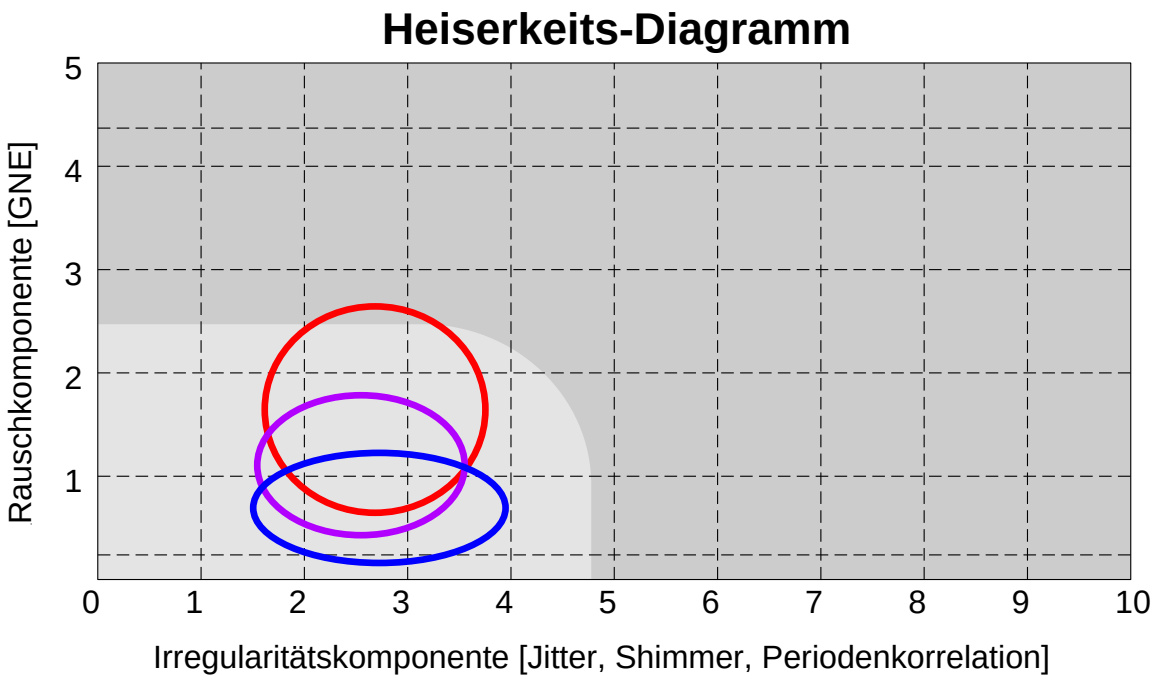
7/97 **Leichte Verringerung der Rauschkomponente und deren Variation.**

10/97 Wiederbeweglichkeit der linken Stimmlippe. **Weitere leichte Verringerung der Rauschkomponente und deren Variation.**

Heiserkeits-Diagramm | Patientin 48, w, 50

Datum: 17.06.98 Zeit: 16:38

	Irreg.komp.	Rauschkomp.
01.07.97 n=383	2.7 (1.07)	1.6 (1.00)
09.07.97 n=421	2.5 (1.00)	1.1 (0.68)
09.10.97 n=359	2.7 (1.22)	0.7 (0.53)



Drittes Physikalisches Institut und Abt. Phoniatrie und Pädaudiologie Universität Göttingen

17. Juni 1998

Literaturverzeichnis

[1] Hans Joachim Arndt and Adelheid Schäfer. Der Weiten-Längen-Quotient der Glottis als Maß für die Amplitudengröße. *Folia Phoniatrica*, 46:265–270, 1994.

[2] Anders Askenfelt, Jan Gauffin, and Johan Sundberg. A comparison of contact microphone and electroglottograph for the measurement of vocal fundamental frequency. *J. Speech Hear. Res.*, 23(2):258–273, 1980.

[3] Anders G. Askenfelt and Britta Hammarberg. Speech waveform perturbation analysis: A perceptual-acoustical comparison of seven measures. *J. Speech Hear. Res.*, 29:50–64, 1986.

[4] B. S. Atal and M. R. Schroeder. Predictive coding of speech signals. In *Proceedings of the 6th International Congress on Acoustics*, pages C–13 – C–16, Tokyo, Japan, August 1968.

[5] Bishnu S. Atal and Manfred R. Schroeder. Adaptive predictive coding of speech signals. *Bell Systems Technical Journal*, 49:1973–1986, 1970.

[6] Bishnu S. Atal and Manfred R. Schroeder. Predictive coding of speech signals and subjective error criteria. *IEEE Trans. Acoust., Speech, Sig. Process.*, ASSP-27(3):247–254, 1979.

[7] Thomas Baer, Anders Löfqvist, and Nancy S. McGarr. Laryngeal vibrations: a comparison between high-speed filming and glottographic techniques. *J. Acoust. Soc. Am.*, 73(4):1304–1308, 1983.

[8] R. J. Baken and Robert F. Orlikoff. Changes in vocal fundamental frequency at the segmental level: control during voiced fricatives. *J. Speech Hear. Res.*, 31:207–211, June 1988.

[9] R.J. Baken and Robert F. Orlikoff. The effect of articulation on fundamental frequency in singers and speakers. *Journal of Voice*, 1(1):68–76, 1987.

[10] Ronald J. Baken. Electroglottography. *Journal of Voice*, 6(2):98–110, 1992.

[11] Germana Banci, Simonetta Monini, Alessandro Falaschi, and Nicolo de Sario. Vocal fold disorder evaluation by digital speech analysis. *Journal of Phonetics*, 14:495–499, 1986.

[12] Michael Blomgren, Yang Chen, Manwa L. Ng, and Harvey R. Gilbert. Acoustic, aerodynamic, physiologic, and perceptual properties of modal and vocal fry registers. *J. Acoust. Soc. Am.*, 103:2649–2658, May 1998.

[13] Marc S. De Bodt, Floris L. Wuyts, Paul H. Van de Heyning, and Christophe Croux. Test-retest study of the GRBAS scale: Influence of experience and professional background on perceptual rating of voice quality. *Journal of Voice*, 11(1):74–80, 1997.

[14] Daniel E. Callan, Ray D. Kent, Nelson Roy, and Stephen M. Tsko. Self-organizing map for the classification of normal and disordered female voices. *J. Speech Lang. Hear. Res.*, 42:355–366, April 1999.

[15] D.G. Childers and C.K. Lee. Vocal quality factors: Analysis, synthesis, and perception. *J. Acoust. Soc. Am.*, 90(5):2394–2410, 1991.

[16] C.Kiese-Himmel, D. Michaelis, and E. Kruse. Vergleich von Kontaktgranulom-Patienten und stimmgesunden Personen in laryngealen, stimmfunktionalen und akustischen Variablen. In *Aktuelle phoniatrisch-pädaudiologische Aspekte 1999*, pages 16–18, 1996.

[17] Guus de Krom. A cepstrum-based technique for determining a harmonics-to-noise ratio in speech signals. *J. Speech Hear. Res.*, 36:224–266, April 1993.

[18] Guus de Krom. Consistency and reliability of voice quality ratings for different types of speech fragments. *J. Speech Hear. Res.*, 37:985–1000, October 1994.

[19] Guus de Krom. Some spectral correlates of pathological breathy and rough voice quality for different types of vowel fragments. *J. Speech Hear. Res.*, 38:794–811, August 1995.

[20] Dimitar D. Deliyski. Acoustic model and evaluation of pathological voice production. In *Eurospeech '93*, volume 3, pages 1969–1972, September 1993.

[21] L. Eskenazi, D.G. Childers, and D.M. Hicks. Acoustic correlates of vocal quality. *J. Speech Hear. Res.*, 33:298–306, 1990.

[22] U. Eysholdt, M. Tigges, T. Wittenberg, and U. Pröschel. Direct evaluation of high-speed recordings of vocal fold vibrations. *Folia Phoniatrica*, 48:163–170, 1996.

[23] Andrew M. Fraser and Harry L. Swinney. Independent coordinates for strange attractors from mutual information. *Physical Review A*, 33(2):1134–1140, February 1986.

[24] B. Fritzell, J. Gauffin, B. Hammarberg, I. Karlsson, and J. Sundberg. Measuring insufficient vocal fold closure during phonation. *STL-QPSR*, 4:50–59, 1983.

[25] B. Fritzell, B. Hammarberg, J. Gauffin, and J. Sundberg. Acoustic inverse filtering of breathy phonation. *XIX Congress of the International Association of Logopaedics and Phoniatrics*, 1983.

[26] Matthias Fröhlich, Dirk Michaelis, and Eberhard Kruse. Image sequences as necessary supplement to a pathological voice data base. In G. de Krom, editor, *Proceedings of VOICEDATA98*, pages 64–69, Utrecht, 1998. Utrecht Institute of Linguistics OTS.

[27] Matthias Fröhlich, Dirk Michaelis, and Eberhard Kruse. Objektive Beschreibung der Stimmgüte unter Verwendung des Heiserkeits-Diagramms. *HNO*, 46:684–689, 1998.

[28] Matthias Fröhlich, Dirk Michaelis, and Hans Werner Strube. Acoustic "breathiness measures" in the description of pathologic voices. In *Proceedings ICASSP 98*, volume 2, pages 937–940, Seattle, WA, 1998.

[29] Matthias Fröhlich, Dirk Michaelis, Hans Werner Strube, and Eberhard Kruse. Acoustic voice quality description: Case studies for different regions of the hoarseness diagram. In Thomas Wittenberg, Patrick Mergell, Monika Tigges, and Ulrich Eysholdt, editors, *Advances in Quantitative Laryngoscopy, 2nd 'Round Table'*, pages 143–150, Erlangen, 1997. Dept. Phoniatrics.

[30] Matthias Fröhlich, Dirk Michaelis, Hans Werner Strube, and Eberhard Kruse. Stimmgütebeschreibung mit Hilfe des Heiserkeits-Diagramms: Untersuchung verschiedener pathologischer Gruppen. In M. Gross, editor, *Aktuelle phoniatrisch-pädaudiologische Aspekte 1997/98*, volume 5, pages 42–48, Heidelberg, 1998. Median Verlag.

[31] Matthias Fröhlich, Dirk Michaelis, Hans Werner Strube, and Eberhard Kruse. Akustische Stimmqualität unter verschiedenen Rahmenbedingungen. In M. Gross, editor, *Aktuelle phoniatrisch-pädaudiologische Aspekte 1998/99*, volume 6, pages 34–39, Heidelberg, 1999. Median Verlag.

[32] Matthias Fröhlich, Dirk Michaelis, Hans Werner Strube, and Eberhard Kruse. Voice quality assessment by means of the hoarseness diagram. Angenommen zur Veröffentlichung bei J. Speech Lang. Hear. Res., 2000.

[33] Bruce R. Gerratt, Jody Kreiman, Norma Antonanzas-Barroso, and Gerald S. Berke. Comparing internal and external standards in voice quality judgments. *J. Speech Hear. Res.*, 36:14–20, 1993.

[34] Tomoyuki Haji, Satoshi Horiguchi, Thomas Baer, and Wilbur J. Gould. Frequency and amplitude perturbation analysis of electroglottograph during sustained phonation. *J. Acoust. Soc. Am.*, 80(1):58–62, July 1986.

[35] B. Hammarberg, B. Fritzell, J. Gauffin, J. Sundberg, and L. Wedin. Perceptual and acoustic correlates of abnormal voice qualities. *Acta Otolaryngol. (Stockh.)*, 90:441–451, 1980.

[36] Britta Hammarberg, Björn Fritzell, Jan Gauffin, and Johan Sundberg. Acoustic and perceptual analysis of vocal dysfunction. *Journal of Phonetics*, 14:533–547, 1986.

[37] Stellan Hertegård and Jan Gauffin. Glottal area and vibratory patterns studied with simultaneous stroboscopy, flow glottography, and electroglottography. *J. Speech Hear. Res.*, 38:85–100, 1995.

[38] Wolfgang Hess and Helge Indefrey. Accurate time-domain pitch determination of speech signals by means of a laryngograph. *Speech Communication*, 6:55–68, 1987.

[39] Wolfgang J. Hess. Determination of glottal excitation cycles in running speech. *Phonetica*, 52:196–204, 1995.

[40] James Hillenbrand. Perception of aperiodicities in synthetically generated voices. *J. Acoust. Soc. Am.*, 83(6):2361–2371, June 1988.

[41] James Hillenbrand, Ronald A. Cleveland, and Robert L. Erickson. Acoustic correlates of breathy vocal quality. *J. Speech Hear. Res.*, 37:769–778, August 1994.

[42] James Hillenbrand and Robert A. Houde. Acoustic correlates of breathy vocal quality: Dysphonic voices and continuous speech. *J. Speech Hear. Res.*, 39:311 – 321, April 1996.

[43] Minoru Hirano. Objective evaluation of the human voice: Clinical aspects. *Folia Phoniatr*, 41:89–144, 1989.

[44] Minoru Hirano, Seishi Hibi, Tetsuji Yoshida, Yoshio Hirada, Hideki Kasuya, and Yoshinobu Kikuchi. Acoustic analysis of pathological voice. *Acta Otolaryngol*, 105:432–438, 1988.

[45] Nobuaki Hiraoka, Yasuhiro Kitazoe, Hisashi Ueta, Shinzo Tanaka, and Masahiro Tanabe. Harmonic-intensity analysis of normal and hoarse voices. *J. Acoust. Soc. Am.*, 76(6):1648–1651, December 1984.

[46] Sture Holm. A simple sequentially rejective multiple test procedure. *Scand. J. Statist.*, 6:65–70, 1979.

[47] Satoshi Horigushi, Tomoyuki Haji, Thomas Baer, and Wilbur J. Gould. *Laryngeal function in Phonation and respiration*, chapter Comparison of electroglottographic and acoustic waveform perturbation measures, pages 509–518. San Diego: College Hill Press, 1987.

[48] Tzu-Yu Hsiao, Nancy Pearl Solomon, Erich S. Luschei, and Ingo R. Titze. Modulation of fundamental frequency by laryngeal muscles during vibrato. *Journal of Voice*, 8(3):224–229, 1994.

[49] Satoshi Imaizumi. Acoustic measure of roughness in pathological voice. *Journal of Phonetics*, 14:457–462, 1986.

[50] K. Ishizaka and J.L. Flanagan. Synthesis of voiced sounds from a two-mass model of the vocal cords. *Bell Sys. Tech. J.*, 51(6):1233–1268, July-August 1972.

[51] Hideki Kasuya, Yasuo Endo, and Sokol Saliu. Novel acoustic measurements of jitter and shimmer characteristics from pathological voice. In *Eurospeech '93*, volume 3, pages 1973–1976, September 1993.

[52] Hideki Kasuya, Kanji Masubuchi, Satoshi Ebihara, and Hajime Yoshida. Preliminary experiments on voice screening. *Journal of Phonetics*, 14:463–468, 1986.

[53] Hideki Kasuya, Shigehi Ogawa, and Yoshinobu Kikuchi. An adaptive comb filtering method as applied to acoustic analyses of pathological voice. In *ICASSP 86*, pages 669–672, 1986.

[54] Hideki Kasuya, Shigeki Ogawa, and Yoshinobu Kikuchi. An acoustic analysis of pathological voice and its application to the evaluation of laryngeal pathology. *Speech Communication*, 5:171–181, 1986.

[55] Hideki Kasuya, Shigeki Ogawa, Kazuhiko Mashima, and Satoshi Ebihara. Normalized noise energy as an acoustic measure to evaluate pathologic voice. *J. Acoust. Soc. Am.*, 80(5):1329–1334, November 1986.

[56] Shigeru Kiritani, Hajime Hirose, and Hiroshi Imagawa. High-speed digital image analysis of vocal cord vibration in diplophonia. *Speech Communication*, 13:23–32, 1993.

[57] Dennis H. Klatt and Laura C. Klatt. Analysis, synthesis, and perception of voice quality variations among female and male talkers. *J. Acoust. Soc. Am.*, 87(2):820–857, 1990.

[58] F. Klingholz. The measurement of the signal-to-noise ratio (SNR) in continuous speech. *Speech Communication*, 6:15–26, 1987.

[59] James B. Kobler, Robert E. Hillman, Steven M. Zeitels, and Jeff Kuo. Assessment of vocal function using simultaneous aerodynamic and calibrated videostroboscopic measures. *Ann. Otol. Rhinol. Laryngol.*, 107:477–485, 1998.

[60] M.N. Kotby, I. R. Titze, M.M. Saleh, and D.A. Berry. Fundamental frequency stability in functional dysphonia. *Acta Otolaryngol. (Stockh.)*, 113:439–444, 1993.

[61] Jody Kreiman, Bruce R. Gerrat, and Gerald S. Berke. The multidimensional nature of pathologic vocal quality. *J. Acoust. Soc. Am.*, 96(3):1291 – 1302, September 1994.

[62] Jody Kreiman and Bruce R. Gerratt. The perceptual structure of pathologic voice quality. *J. Acoust. Soc. Am.*, 100(3):1787–1795, 1996.

[63] Jody Kreiman and Bruce R. Gerratt. Validity of rating scale measures of voice quality. *J. Acoust. Soc. Am.*, 140(3):1598–1608, 1998.

[64] Jody Kreiman, Bruce R. Gerratt, Gail B. Kempster, Andrew Erman, and Gerald S. Berke. Perceptual evaluation of voice quality: Review, tutorial, and a framework for future research. *J. Speech Hear. Res.*, 36:21–40, 1993.

[65] Jody Kreiman, Bruce R. Gerratt, and Kristin Precoda. Listener experience and perception of voice quality. *J. Speech Hear. Res.*, 33:103–115, 1990.

[66] B. J. Kröger. Zur Auswirkung der Glottis-Sprechtrakt-Kopplung auf die Stimmreinheit. *Sprache-Stimme-Gehör*, 15(4):139–142, 1991.

[67] E. Kruse, D. Michaelis, P. Zwirner, and E. Bender. Stimmfunktionelle Qualitätssicherung in der kurativen Mikrochirurgie der Larynxmalignome auf der Basis der „Laryngealen Doppelventilfunktion". *HNO*, 45:712–718, 1997.

[68] Eberhard Kruse, Matthias Fröhlich, and Dirk Michaelis. Phonatory conditions and acoustic analysis of pathologic voices: Is there a correspondence ? In *Program and Abstract Book, 24th IALP congress*, page 127, Amsterdam, 1998.

[69] John Laver, Steven Hiller, Janet Mackenzie, and Edmund Rooney. An acoustic screening system for the detection of larygeal pathology. *Journal of Phonetics*, 14:517–524, 1986.

[70] Lea Leinonen, Tapio Hiltunen, Maija-Liisa Laakso, Heikki Rihkanen, and Håkan Poppius. Categorization of voice disorders with six perceptual dimensions. *Folia Phoniatr. Logop.*, 49:9–20, 1997.

[71] Philip Lieberman. Perturbation in vocal pitch. *J. Acoust. Soc. Am.*, 33(5):597–603, May 1961.

[72] Philip Lieberman. Some acoustic measures of the fundamental periodicity of normal and pathologic larynges. *J. Acoust. Soc. Am.*, 35(3):344–353, March 1963.

[73] J. Liljencrants. *Speech Synthesis with a Reflection-Type Line Analog*. PhD thesis, Royal Institute of Technology, Stockholm, Sweden, 1985.

[74] J. D. Markel and A. H. Gray. *Linear Prediction of Speech*. Springer-Verlag Berlin Heidelberg New York, 1976.

[75] David Martin, James Fitch, and Virginia Wolfe. Pathologic voice type and the acoustic prediction of severity. *J. Speech Hear. Res.*, 38:765–771, 1995.

[76] Yoav Medan, Eyal Yair, and Dan Chazan. Super resolution pitch determination of speech signals. *IEEE Trans. Sig. Process.*, 39(1):40–48, 1991.

[77] Patrick Mergell, Hanspeter Herzel, Thomas Wittenberg, Monika Tigges, and Ulrich Eysholdt. Phonation onset: Vocal fold modeling and high-speed glottography. *J. Acoust. Soc. Am.*, 104(1):464–470, 1998.

[78] D. Michaelis and H. W. Strube. Informationstheoretischer Ansatz zur Selektion und Kombination akustischer Stimmgüteparameter. In *Fortschritte der Akustik DAGA 1997*, 1997.

[79] D. Michaelis, H. W. Strube, and E. Kruse. Multidimensionale Analyse akustischer Stimmgüteparameter. In *Aktuelle phoniatrisch-pädaudiologische Aspekte 1996*, pages 16–18, 1996.

[80] Dirk Michaelis, Matthias Fröhlich, and Hans Werner Strube. Selection and combination of acoustic features for the description of pathologic voices. *J. Acoust. Soc. Am.*, 103(3):1628–1639, 1998.

[81] Dirk Michaelis, Matthias Fröhlich, Hans Werner Strube, and Eberhard Kruse. Reliabilität akustischer Stimmgütebeschreibung bei reduziertem Umfang der Stimmaufnahmen. In M. Gross, editor, *Aktuelle phoniatrisch-pädaudiologische Aspekte 1997/98*, volume 5, pages 48–53, Heidelberg, 1998. Median Verlag.

[82] Dirk Michaelis, Matthias Fröhlich, Hans Werner Strube, Eberhard Kruse, Brad Story, and Ingo R. Titze. Some simulations concerning jitter and shimmer measurement. In T. Lehmann, C. Palm, K. Spitzer, and T. Tolxdorff, editors, *Advances in Quantitative Laryngoscopy, Voice and Speech Research. Proceedings of the 3rd International Workshop*, pages 71–80, Aachen, June 19-20, 1998 1998. RWTH University of Technology.

[83] Dirk Michaelis, Matthias Fröhlich, Hans Werner Strube, Eberhard Kruse, Brad Story, and Ingo R. Titze. Grenzen der Jitter– und Shimmer–Messung pathologischer Stimmen mit dem unüberwachten Waveform–Matching Verfahren. In A. Sill, editor, *Fortschritte der Akustik – DAGA 98 (Zürich)*, pages 382–383. Oldenburg, 1998.

[84] Dirk Michaelis, Tino Gramss, and Hans Werner Strube. Glottal-to-noise excitation ratio – a new measure for describing pathological voices. *Acustica / acta acustica*, 83:700–706, 1997.

[85] Dirk Michaelis and Hans Werner Strube. Empirical study to test the independence of different acoustic voice parameters on a large voice database. In J.M. Pardo, E. Enriquez, J. Ortega, J. Ferreiros, J. Macias, and F.J. Valverde, editors, *Eurospeech '95*, volume 3, pages 1891–1894, September 1995.

[86] Dirk Michaelis and Hans Werner Strube. Orthogonale akustische Stimmgüteparameter zur Stimmtherapiedokumentation. In W. Arnold and S. Hirsekorn, editors, *Fortschritte der Akustik*, volume II, pages 1035–1038. Deutsche Gesellschaft für Akustik e.V., March 1995.

[87] Dirk Michaelis, Hans Werner Strube, and Eberhard Kruse. Reliabilität und Validität des Heiserkeits-Diagramms. In M. Gross and U. Eysholdt, editors, *Aktuelle phoniatrisch-pädaudiologische Aspekte 1996*, volume 4, pages 25–26, Heidelberg, 1997. Median Verlag.

[88] Dirk Michaelis, Hans Werner Strube, Petra Zwirner, and Eberhard Kruse. Frequenzkanalabhängige Korrelationen der Stimmschallanregung als akustisch-diagnostischer Stimmgüteparameter. In M. Gross, editor, *Aktuelle phoniatrisch-pädaudiologische Aspekte 1994*, volume 2, page 128, Heidelberg, 1994. Median Verlag.

[89] Paul Milenkovic. Least mean square measures of voice perturbation. *J. Speech Hear. Res.*, 30(4):529–538, 1987.

[90] Richard J. Morris and W.S. Brown, Jr. Comparison of various automatic means for measuring mean fundamental frequency. *Journal of Voice*, 10(2):159–165, 1996.

[91] B. Müller and J. Reinhard. *Neural Networks*. Springer-Verlag, 1990.

[92] Thomas Murry. Multidimensional classification of abnormal voice qualities. *J. Acoust. Soc. Am.*, 61(6):1631–1635, June 1977.

[93] Thomas Murry and Sadanandh Singh. Multidimensional analysis of male and female voices. *J. Acoust. Soc. Am.*, 68(5):1294–1300, 1980.

[94] Hirishi Muta and Thomas Baer. A pitch-synchronous analysis of hoarseness in running speech. *J. Acoust. Soc. Am.*, 84(4):1292–1301, October 1988.

[95] A. Michael Noll. Short-time spectrum and cepstrum techniques for vocal-pitch detection. *J. Acoust. Soc. Am.*, 36(2):296–302, February 1964.

[96] A. Michael Noll. Short-time spectrum and cepstrum techniques for vocal-pitch detection. *J. Acoust. Soc. Am.*, 36(2):296–302, February 1964.

[97] A. Michael Noll. Cepstrum pitch determination. *J. Acoust. Soc. Am.*, 41(2):293–309, 1967.

[98] A. Michael Noll and M.R. Schroeder. Short-time cepstrum pitch detection. *J. Acoust. Soc. Am.*, 36(2):1030, May 1964.

[99] Robert F. Orlikoff. Assessment of the dynamics of vocal fold contact from normal male subjects. *J. Speech Hear. Res.*, 34:1066–1072, 1991.

[100] Robert F. Orlikoff. Vocal stabiliy and vocal tract configuration: an acoustic and electroglottographic investigation. *Journal of Voice*, 9(2):173–181, 1995.

[101] Vijay Parsa and Donald G. Jamieson. A comparison of high precision F0 extraction algorithms for sustained vowels. *J. Speech Lang. Hear. Res.*, 42:112–126, 1999.

[102] Robert C. Peppard, Diane M. Bless, and Paul Milenkovic. Comparison of young adult singers and nonsingers with vocal nodules. *Journal of Voice*, 2(3):250–260, 1988.

[103] Jay F. Piccirillo, Colin Painter, Dennis Fuller, and John M. Fredrickson. Multivariate analysis of objective vocal function. *Ann. Otol. Rhinol. Laryngol.*, 107:107–112, 1998.

[104] Neal B. Pinto and Ingo R. Titze. Unification of perturbation measures in speech signals. *J. Acoust. Soc. Am.*, 87(3):1278–1289, March 1990.

[105] Fabrice Plante, Jocelyne Borel, Christian Berger-Vachon, and Isabelle Kauffmann. Acoustic detection of laryngeal diseases in children. In *Eurospeech '93*, volume 3, pages 1965–1972, September 1993.

[106] William H. Press, Brian P. Flannery, Saul A. Teukolsky, and William T. Vetterling. *Numerical Recipes in C*. Cambridge University Press, 1989.

[107] Yingyong Qi, Bernd Weinberg, Ning Bi, and Wolfgang J. Hess. Minimizing the effect of period determination on the computation of amplitude perturbation in voice. *J. Acoust. Soc. Am.*, 97(4):2525–2532, April 1995.

[108] C. Rose Rabinov, Jody Kreiman, Bruce R. Gerratt, and Steven Bielamowicz. Comparing reliability of perceptual ratings of roughness and acoustic measures of jitter. *J. Speech Hear. Res.*, 38:26–32, 1995.

[109] Linda A. Rammage, Robert C. Peppard, and Diane M. Bless. Aerodynamic, laryngoscopic, and perceptual-acoustic characteristics in dysphonic females with posterior glottal chinks: A retrospective study. *Journal of Voice*, 6(1):64 – 78, 1992.

[110] A.E. Rosenberg. Effect of glottal pulse shape on the quality of natural vowels. *J. Acoust. Soc. Am.*, 49(2):583–590, 1970.

[111] M. Rothenberg and S. Zahorian. Nonlinear inverse filtering technique for estimating the glottal-area waveform. *J. Acoust. Soc. Am.*, 61(4):1063–1071, 1977.

[112] Martin Rothenberg. A new inverse-filtering technique for deriving the glottal air flow waveform during voicing. *J. Acoust. Soc. Am.*, 53(6):1632–1645, 1973.

[113] Ronald C. Scherer, Wilbur J. Gould, Ingo R. Titze, Arlen D. Meyers, and Robert T. Sataloff. Preliminary evaluation of selected acoustic and glottographic measures for clinical phonatory function analysis. *Journal of Voice*, 2(3):230–244, 1988.

[114] Ronald C. Scherer, Vernon J. Vail, and Chwen Geng Guo. Required number of tokens to determine representative voice perturbation values. *J. Speech Hear. Res.*, 38:1260–1269, 1995.

[115] Stephan Schneider. Zum Einsatz fraktaler Verfahren in der Sprachkompression. In Dieter Mehnert, editor, *Elektronische Signalverarbeitung*, volume 16, pages 104 – 111. w.e.b. Universitätsverlag, 1999.

[116] J. Schoentgen and R. de Guchteneere. Auto-regressive linear models of jitter. In *Eurospeech '93*, volume 3, pages 2033–2036, September 1993.

[117] Jean Schoentgen. Quantitative evaluation of the discrimination performance of acoustic features in detecting laryngeal pathology. *Speech Communication*, 1:269–282, 1982.

[118] Jean Schoentgen and Raoul de Guchteneere. An algorithm for the measurement of jitter. *Speech Communication*, 10:533–538, 1991.

[119] Jean Schoentgen and Raoul de Guchteneere. Time series analysis of jitter. *Journal of Phonetics*, 23:189–201, 1995.

[120] Jean Schoentgen and Raoul de Guchteneere. Predictable and random components of jitter. *Speech Comm.*, 21:255–272, 1997.

[121] M. R. Schroeder. Period histogram and product spectrum: New methods for fundamental-frequency measurement. *J. Acoust. Soc. Am.*, 43(4):829–834, January 1968.

[122] M. R. Schroeder and B. S. Atal. Generalized short-time power spectra and autocorrelation function. *J. Acoust. Soc. Am.*, 34(11):1679–1683, November 1962.

[123] Manfred R. Schroeder. *Fractals, Chaos, Power Laws: Minutes from an Infinite Paradise*. W. H. Freeman and Company, 1991.

[124] Manfred R. Schroeder. *Computer Speech. Recognition, Compression, Synthesis*, volume 35 of *Springer Series in Information Sciences*. Springer Verlag, Berlin, 1999.

[125] Maria Södersten, Alf Håkansson, and Britta Hammarberg. Comparison between automatic and manual inverse filtering procedures for healthy female voices. *Log. Phon. Vocol.*, 24:26–38, 1999.

[126] Maria Södersten, Stellan Hertegård, and Britta Hammarberg. Glottal closure, transglottal airflow, and voice quality in healthy middle-aged women. *Journal of Voice*, 9(2):182 – 197, 1995.

[127] Maria Södersten, Stellan Hertegård, and Britta Hammarberg. Glottal closure, transglottal airflow, and voice quality in healthy middle-aged women. *Journal of Voice*, 9(2):182–97, 1995.

[128] Man Mohan Sondhi. New methods of pitch extraction. *IEEE Transactions on Audio and Electroacoustics*, AU-16(2):262–266, June 1968.

[129] M.M. Sondhi and J. Schroeter. A hybrid time-frequency domain articulatory speech synthesizer. *IEEE Trans. Acoust., Speech, Sig. Process.*, ASSP-35(7):955–967, July 1987.

[130] Brad H. Story. *Physiologically-based speech simulation using an enhanced wave-reflection model of the vocal tract*. PhD thesis, University of Iowa, 1995.

[131] Brad H. Story, Ingo R. Titze, and Eric A. Hoffman. Vocal tract area functions for an adult female speaker based on volumetric imaging. *J. Acoust. Soc. Am.*, 14(1):471–487, 1998.

[132] Hans Werner Strube. Determination of the instant of glottal closure from the speech wave. *J. Acoust. Soc. Am.*, 56(5):1625–1629, November 1974.

[133] Hans Werner Strube, Dirk Michaelis, and Matthias Fröhlich. Akustische Sprachparameter zur Bewertung glottaler Pathologien. In D. Mehnert, editor, *Elektronische Sprachsignalverarbeitung*, number 13 in Studientexte zur Sprachkommunikation, pages 52–58, Dresden, 1996.

[134] H.W. Strube. Time-varying wave digital filters for modeling analog systems. *IEEE Trans. Acoust., Speech, Sig. Process.*, ASSP-30(6):864–868, December 1982.

[135] Arend M. Sulter and Hero P. Wit. Glottal volume velocity waveform characteristics in subjects with and without vocal training, related to gender, sound intensity, fundamental frequency, and age. *J. Acoust. Soc. Am.*, 100(5):3360–3373, 1996.

[136] Johan Sundberg. Vocal fold vibration patterns and modes of phonation. *Folia Phoniatrica*, 47:218–228, 1995.

[137] Ingo R. Titze. Interpretation of the electroglottographic signal. *Journal of Voice*, 4(1):1–9, 1990.

[138] Ingo R. Titze. Toward standards in acoustic analysis of voice. *Journal of Voice*, 8(1):1–7, 1994.

[139] Ingo R. Titze, Thomas Baer, Donld Cooper, and Ronald Scherer. Automated extraction of glottographic waveform parameters and regression to acoustic and physiologic variables. In D.M. Bless and J.H. Abbs, editors, *Vocal Fold Physiology: Contemporary Research and Clinical Issues*, chapter 11, pages 146–154. College-Hill Press, San Diego, 1983.

[140] Ingo R. Titze, Yoshiyuki Horii, and Ronald C. Scherer. Some technical considerations in voice perturbation measurements. *J. Speech Hear. Res.*, 30:250–260, 1987.

[141] Ingo R. Titze and Haixiang Liang. Comparison of f0 extraction methods for high precision voice perturbation measurement. *NCVS Status and Progress Report*, 3:97–115, December 1992.

[142] Ingo R. Titze and Haixiang Liang. Comparison of F_0 extraction methods for high-precision voice perturbation measurements. *J. Speech Hear. Res.*, 36:1120–1133, December 1993.

[143] Ingo R. Titze and William S. Winholtz. The effect of microphone type and placement on voice perturbation measurement. *NCVS Status and Progress Report*, 3:117–134, December 1992.

[144] Ingo R. Titze and William S. Winholtz. Effect of microphone type and placement on voice perturbation measurements. *J. Speech Hear. Res.*, 36:1177–1190, December 1993.

[145] Ingo R. Titze, Darrell Wong, Martin A. Milder, Susan R. Hensley, and Lorraine O. Ramig. Comparison between clinician-assisted and fully automated procedures for obtaining voice range profiles. *J. Speech Hear. Res.*, 38:526–535, June 1995.

[146] I.R. Titze. Parameterization of the glottal area, glottal flow, and vocal fold contact area. *J. Acoust. Soc. Am.*, 75(2):570–580, 1984.

[147] I.R. Titze and H. Liang. Comparison of F_0 extraction methods for high precision voice perturbation measurements. *NCVS Status and Progress Report*, 3:97–115, 1992.

[148] J. Verstraete, G. Forrez, P. Mertens, and F. Debruyne. The effect of sustained phonation at high and low pitch on vocal jitter and shimmer. *Folia Phoniatrica*, 45:223–228, 1993.

[149] J. Wendler, A. Rauhut, and H. Krüger. Classification of voice qualities. *Journal of Phonetics*, 14:483–488, 1986.

[150] William S. Winholtz and Ingo R. Titze. Miniature head mount microphone for acoustic analysis. *NCVS Status and Progress Report*, 5:49–52, November 1993.

[151] William S. Winholtz and Ingo R. Titze. Miniature head mount microphone for acoustic analysis. *J. Speech Hear. Res.*, 40:894–899, 1997.

[152] Virginia Wolfe, James Fitch, and Richard Cornell. Acoustic prediction of severity in commonly occurring voice problems. *J. Speech Hear. Res.*, 38:273–279, April 1995.

[153] Virginia Wolfe, James Fitch, and David Martin. Acoustic measures of dysphonic severity across and within voice types. *Folia Phoniatr Logop*, 49:292–299, 1997.

[154] Peak Woo. Quantification of videostroboscopic findings - measurements of the normal glottal cycle. *Laryngoscope*, 106:1–27, March 1996.

[155] Naoaki Yanagihara. Significance of harmonic changes and noise components in hoarseness. *J. Speech Hear. Res.*, 10:531–541, 1967.

[156] Eiji Yumoto, Wilbur J. Gould, and Thomas Baer. Harmonic-to-noise ratio as an index of the degree of hoarseness. *J. Acoust. Soc. Am.*, 71(6):1544–1550, June 1982.

[157] Petra Zwirner, Dirk Michaelis, Matthias Fröhlich, Hans Werner Strube, and Eberhard Kruse. Korrelationen zwischen perzeptueller Beurteilung von Stimmen nach dem RBH-System und akustischen Parametern. In M. Gross, editor, *Aktuelle phoniatrisch-pädaudiologische Aspekte 1997/98*, volume 5, pages 63–67, Heidelberg, 1998. Median Verlag.

[158] Petra Zwirner, Dirk Michaelis, and Eberhard Kruse. Akustische Stimmanalysen zur Dokumentation der Stimmrehabilitation nach laserchirurgischer Larynxkarzinomresektion. *HNO*, 44:514–520, 1996.

5. Danksagung

An erster Stelle möchte ich mich bei Prof. Manfred R. Schroeder für die Betreuung dieser Doktorarbeit bedanken. Ihm gebührt der Dank dafür, dass an der Göttinger Universität am Fachbereich Physik die Möglichkeit zur physikalischen Sprachforschung entstanden ist. An seinem Vorbild haben sich der Autor und viele andere Wissenschaftler orientiert, und daran gearbeitet, die Methoden der experimentellen und der mathematischen Physik sowie der reinen Mathematik auf die Arbeitsgebiete der Akustik, der Sprach- und Gehörforschung zu übertragen und damit neue Impulse in diesen Arbeitsgebieten zu setzen.

Daneben gilt mein besonderer Dank Dr. Hans Werner Strube, der stets ein offenes Ohr für Detailfragen auf dem Gebiet der Signalverarbeitung und andere Fragen der Sprachforschung sowie für die vielen kleinen und großen Probleme mit dem Arbeitsgerät Computer hat.

Prof. Eberhard Kruse, der Leiter der Abteilung Phoniatrie und Pädaudiologie der Universitätsklinik Göttingen, hat die Initiative für die Zusammenarbeit der Phoniatrie und Pädaudiologie mit dem Dritten Physikalischen Institut ergriffen. Durch seine Bemühungen wurde diese Arbeit und wird ein weiteres Projekt von der Deutschen Forschungsgemeinschaft unterstützt. Dafür und für die Aufmunterung in schwierigen Zeiten möchte ich ihm herzlich danken.

Ein großer Dank ist auch an Dr. Matthias Fröhlich gerichtet. Die hervorragende Zusammenarbeit mit ihm hat viel zu dem Gelingen des Projektes beigetragen. Er ist stets der erste Diskussionspartner bei fachlichen Schwierigkeiten gewesen.

Ein ebensolcher großer Dank geht auch an alle Mitarbeiter der Arbeitsgruppe. An Knut, Heiko, Jan, Tillmann, Jannis, Olaf, Joachim, Holger, Wotan, Kyrill, Hansjörg, Martin und an Tino der den Dank leider nicht mehr vernimmt. Alle haben diese Arbeit durch Diskussionen und Tips beim Umgang mit dem Computer sehr unterstützt und ein sehr angenehmes Arbeiten ermöglicht.

Ein ganz persönlicher und ganz herzlicher Dank ist an Astrid gerichtet, die mich in den schwierigsten Zeiten mit allen Kräften unterstützt hat.