Masterarbeit

Analyse hoher Männergesangsstimmen anhand des Obertonspektrums

Erstellt von Felix Frederik Koneberg
Matrikelnummer 2093808

Institut Selbstständiger Funktionsbereich für Phoniatrie und Pädaudiologie, Universitätsklinikum Düsseldorf
Gutachter Univ.-Prof. Dr. med. Wolfgang Angerstein
Prof. Dr. Mathias Getzlaff

Abgabe der Arbeit 21. April 2015

Erklärung

Hiermit versichere ich, dass ich diese Masterarbeit selbstständig verfasst und nicht vorher zur Erlangung eines akademischen Grades eingereicht habe. Ich habe dazu keine anderen als die angegebenen Quellen und Hilfsmittel verwendet. Die Zitate wurden kenntlich gemacht.

Düsseldorf, den 21. April 2015

_____

Felix Koneberg

Zusammenfassung

Das Ziel dieser Arbeit bestand darin, ein Spektrum an Methoden zur Charakterisierung verschiedener hoher Männergesangsstimmen zu erarbeiten. Die Charakterisierung erfolgte anhand des Schallspektrums mittels der Fast-Fourier-Transformation von ausschließlich kommerziell erhältlichen CD-Aufnahmen. Die zu analysierenden Gesangsstimmen lassen sich hierbei in die folgenden Populationen einteilen:

Unterscheidung nach Gesangstechnik: Hierbei unterscheidet man zwischen der Bruststimme und dem Falsett, wobei Letzteres häufiger von Männern bei hohem Gesang angewendet wird. Im Falsett schwingen die Stimmlippen nur mit reduzierter Masse und erzeugen so höhere Töne. Bei den Countertenören in der klassischen Musik singen die Sänger mit einer dieser Techniken oberhalb der Tenorlage. Aber auch in der Pop- und Rockmusik sind hoch singende Männer bekannt.

Medizinische Unterscheidung: Man unterscheidet zwischen sogenannten echten Kastraten und hormonellen Kastraten. Bei den echten Kastraten wurden im 16. bis 19. Jahrhundert jungen Knaben vor dem Einsetzen des Stimmbruchs die Hoden entfernt, sodass ihre Stimme anschließend nicht tiefer wurde. Bei den hormonellen Kastraten bleibt der Stimmbruch durch krankheitsbedingte Hormonstörungen aus.

Synthetisierte Kastratenstimme: Für den Film *Farinelli* wurden die Stimme eines männlichen Countertenors und die einer Sopranistin aufeinander abgeglichen, um den Eindruck einer Kastratenstimme zu erzeugen.

Frauenstimmen: Frauen haben physiologisch bedingt eine höhere Stimme. Der Unterschied zwischen Bruststimme und Falsett fällt weniger deutlich aus als bei Männern.

Nach einer intensiven Auseinandersetzung mit der Literatur zeigte sich, dass eine große Anzahl an Parametern beachtet werden muss, um präzise Unterscheidungen hinsichtlich Stimmgattung, Stimmklang und erweiterten Techniken wie Formantentuning vornehmen zu können. Da das Spektrum jedes Sängers sowohl von der gesungenen Lautstärke als auch von der Tonhöhe abhängt, kann eine triviale Auswahl an Spektralausschnitten vereinzelter Gesangspassagen hier keine präzise Antwort ergeben. Daher wurden aus verschiedenen Quellen unterschiedliche Methoden entnommen und mit selbst entwickelten Methoden kombiniert. Das selbst entwickelte Verfahren wird *Harmonischenextraktion* (HE) genannt und extrahiert aus dem zeitlichen Spektralverlauf gezielt die Obertöne des Sängers. Mit einer umfangreichen Fehleranalyse konnte im Rahmen der Arbeit gezeigt werden, dass die Methoden mit Einschränkungen auch unter Instrumentaleinfluss anwendbar sind. Durch die große Anzahl an recherchierten und erarbeiteten Methoden ist die vorliegende Arbeit sehr methodenorientiert, und die Zahl der analysierten Gesangsstimmen wurde auf zwei Beispiele (Jochen Kowalski und Russell Oberlin) limitiert, um den Umfang der Arbeit zu begrenzen.

Zu den Methoden zählen das Langzeit-gemittelte Spektrum (LTAS) sowie das tonhöhenkorrigierte Langzeit-gemittelte Spektrum (LTAS-T) für die Charakterisierung eines möglichen Sängerformanten. In den Spektren der zwei analysierten Sänger konnten eindeutige Unterschiede hinsichtlich der Position und der relativen Intensität des Sängerformanten festgestellt werden. Der Sängerformant von Jochen Kowalski lag über alle Vokale gemittelt mit 2687 Hz fast 600 Hz niedriger als der von Russell Oberlin mit 3250 Hz. Ähnliches zeigte sich auch für die Analyse der einzelnen Vokale.

Außerdem wurde das spektrale Gefälle, welches ein wichtiges Maß für den Klang einer Stimme darstellt, in Abhängigkeit von der Tonhöhe und der Intensität der ersten Harmonischen der Gesangsstimme bestimmt und in einer übersichtlichen Form dargestellt. Auch hier zeigten sich wesentliche Unterschiede zwischen den zwei Sängern. Oberlin weist für die niedrigen Tonhöhen und Intensitäten ein geringes spektrales Gefälle auf, welches mit zunehmender Frequenz und Intensität zunimmt. Das spektrale Gefälle Kowalskis ist über alle Tonhöhen deutlich größer.

Als weitere Methode wurden die Singing Power Ratio (SPR) und die Position des Singing Power Peaks (SPP) in Abhängigkeit von der Tonhöhe bestimmt. Die Maxima der SPR liegen für die beiden Sänger bei unterschiedlichen Tonhöhen (Kowalski: ca. 311 - 440 Hz, Oberlin: ca. 262 - 311 Hz). Da in den kommerziellen Aufnahmen die zur Analyse geeigneten Ausschnitte begrenzt sind, konnten nicht alle Vokale mit einer ausreichenden Gesamtdauer und für ausreichend viele Tonhöhen analysiert werden. Daher ist die Aussagekraft dieser Beobachtung begrenzt und sollte noch in einer größeren Analyse bestätigt werden.

Zuletzt konnte mittels Codierung durch lineare Prädiktion (LPC) die Anwendung von Formantentuning bei beiden Sängern nachgewiesen werden.

Die Methoden des LTAS-T, des spektralen Gefälles und der SPR verwenden die extrahierten Harmonischen der oben beschriebenen Harmonischenextraktion.

Im Rahmen der Arbeit konnte eine geeignete Methodenzusammenstellung für die Analyse von Gesangsstimmen erarbeitet und anhand zweier Analysen angewendet werden. Im weiteren Verlauf des Projekts kann mit diesen Methoden nun eine größere Anzahl an Sängern analysiert werden, um die Charakteristika einzelner Sänger und Sängerpopulationen weiter zu erforschen.

Danksagung

Ich möchte mich an dieser Stelle herzlich bei Herrn Univ.-Prof. Dr. med. Wolfgang Angerstein für die intensive Betreuung und dafür, dass er mir dieses spannende Thema ermöglicht hat, bedanken.

Außerdem geht mein Dank an Herrn Bodo Maass für die individuelle Anpassung des Overtone Analyzers und die hilfreichen Tipps.

Inhaltsverzeichnis

Abkürzungsverzeichnis

Abb.	Abbildung
Aufl.	Auflage
bzw.	beziehungsweise
ca.	circa
d.h.	das heißt
dB	Dezibel
dB/Okt	Dezibel pro Oktave
DFT	Diskrete Fourier-Transformation
et al.	und Mitarbeiter (lat. et alii)
f. / ff.	folgende (Seite) / folgende (Seiten)
FFT	Schnelle Fourier-Transformation (engl. Fast Fourier-Transform)
FT	Fourier-Transformation
ggf.	gegebenenfalls
HE	Harmonischenextraktion
IPA	Internationales Phonetisches Alphabet
LPC	Codierung durch lineare Prädiktion (engl. linear predictive coding)
LTAS	Langzeit-gemitteltes Spektrum (engl. long term average spectrum)
LTAS-T	Tonhöhenkorrigiertes Langzeit-gemitteltes Spektrum (siehe auch: LTAS)
M.	Musculus
OA	Overtone Analyzer
p. / pp.	page (Seite) / pages (Seiten)
SPP	Singing Power Peak
SPR	Singing Power Ratio
u.a.	unter anderem
usw.	und so weiter
v. d.	van den
z.B.	zum Beispiel

1 Historischer Überblick über die hohe Singstimme des Mannes

Gesang ist seit Jahrtausenden ein wichtiges Element der Gesellschaft. Obwohl es nicht unbedingt intuitiv ist, die von Natur aus tiefere Männerstimme für hohen Gesang einzusetzen, so lässt sich dies schon sehr lange beobachten. Sogar Papst Isidor von Sevilla schrieb in seinem Diktum, dass „die perfekte Stimme [...] ‚hoch, süß und klar' sein müsse" (Herr, 2013, S. 13). Dabei waren diese hohen Männerstimmen auch nicht immer „natürlich". So gab es beispielsweise im alten China des 17. Jahrhunderts Eunuchen am kaiserlichen Hof (Clapton, 2004, S. 1) und im alten Rom des 2. Jahrhundert v. Chr. kastrierte Sklaven zur gesanglichen Unterhaltung (Fritz, 1994, S. 43). Das Interesse an hohen Singstimmen ist bis heute sehr groß und viele der heutigen aufgeführten Stücke stammen aus der Barockzeit. Kastraten waren bis dahin Einzelfälle, doch ab dem 16. Jahrhundert häufte sich ihr Vorkommen. Es ist daher interessant, sich die Geschichte der hohen Männersingstimmen ab dem 15. Jahrhundert genauer anzusehen.

1.1 Knaben, Falsettisten und Countertenöre

Mit dem steigenden Anspruch an Kunst und Kultur im europäischen Raum des 15. und 16. Jahrhunderts stieg nach Herr (2013, S. 25 ff.) und Fritz (1994, S. 47 ff.) auch der musikalische Anspruch stetig. Besonders auf den Bühnen, in den Opern und in den Kirchenchören waren hohe Gesangsstimmen mehr und mehr gefragt. Frauen war es jedoch verboten, in den Kirchen zu singen, denn der erste Paulusbrief an die Korinther, Kapitel 16, Vers 34 besagt: „Let women be silent in the assemblies, for it is not permitted to them to speak" (Barbier, 1996, S. 19 ff.), sodass hier vor allem Knaben und Falsettisten eingesetzt wurden. Allerdings war die Stimme eines Knaben nicht so kraftvoll wie die eines Erwachsenen, und bereits vor Vollendung der Gesangsausbildung setzte der Stimmbruch ein, sodass die Knaben fortan nicht wie gewohnt hoch singen konnten. Als Falsettisten werden Sänger bezeichnet, welche entweder ihren Tonumfang (*Ambitus*) nach oben durch Wechsel von der *Bruststimme*[1] (*Modalstimme*) in die *Kopfstimme*[2] (*Falsettstimme*) erweitern oder komplett in Letzterer singen (Herr, 2013, S. 13 ff.). Somit können auch Männer Alt- und Sopranstimmen übernehmen. Doch es ist auch für manche Sänger (z.B. Russell Oberlin) möglich, in der Bruststimme über die Tenorlage hinaus zu kommen. Diese hohen Männerstimmen werden *Hautes Contres* oder *Altinos* genannt und singen nach der *Amerikanischen Schule*. Männliche Sänger, die im Alt oder Sopran singen, werden auch als *Countertenöre*[3] oder *Kontratenöre* bezeichnet. Der in

[1] Die Definition der Brust- und Kopfstimme ist in Abschnitt 2.5.3 zu finden.

[2] In Herr (2013, S. 445-450) werden Uneinigkeiten über die Anzahl an Stimmregistern beschrieben. Dabei besonders, ob das Kopfregister mit dem Falsettregister gleichzusetzen ist. Um hier schwer zu beschreibende Stimmklangeigenschaften zu vermeiden, wird im Folgenden die Kopfstimme mit der Falsettstimme gleichgesetzt. Ardran und David (1967) zeigten, dass die manchmal unterschiedenen Falsett- und Kopfregister physiologisch die gleiche Kehlkopfbewegung erzeugen.

[3] Es gibt, wie in Herr (2013, S. 450-454) und auch Giles (1994, S. xx-xxiii) zu sehen, eine große Debatte über die Betitelung der hohen männlichen Singstimme. Die Argumente, ob ein Sänger nun als Falsettist

der *Europäischen Schule* häufige Einsatz des Falsetts hat jedoch zur Folge, dass sich das Obertonspektrum und somit der Stimmklang ändert (siehe Abschnitt 2). Daher wurde das Falsettieren „in frühneuzeitlichen Gesangslehren häufig abgelehnt“ (Herr et al., 2012, S. 9).

1.2 Kastratengesang im 16. bis 19. Jahrhundert

Im Verlauf des 16. Jahrhunderts kam es besonders in Süditalien dazu, dass mehr und mehr *Kastratensänger* eingesetzt wurden. Die hier herrschende Armut begünstigte diese Entwicklung: Viele arme Familien ließen ihre jungen Söhne, in der Hoffnung auf eine Gesangskarriere mit guter Ausbildung und Verpflegung, kastrieren. Den jungen Knaben wurden hierbei vor dem Einsetzen des Stimmbruchs die Samenleiter durchtrennt oder die Hoden entfernt, sodass ihre Sopran- oder Altstimme durch das Ausbleiben des Testosteronschubes erhalten blieb, während ihr Körper weiterhin wuchs. Dies waren keine legalen Operationen, zudem äußerst gefährlich (Überlebenschance 10-80 % nach Barbier, 1996, S. 11). Als Ausreden dienten oftmals erfundene Geschichten, dass die Hoden des Kindes von einem wilden Tier zertrümmert wurden oder es sich unglücklich verletzt hatte (Heidecker, 2007, S. 9). Das Tragische war, dass die Kastration keine Garantie für eine Sängerkarriere bot. Die Begabten erhielten allerdings ausgezeichnete Gesangsausbildungen, die üblicherweise bis zu 10 Jahre dauerten (Barbier, 1996, S. 32). Durch ihre Körpergröße hatten die Kastraten den Brustumfang eines erwachsenen Mannes und dementsprechend eine deutlich ausdauerndere und kräftigere Modalstimme (Fritz, 1994, S. 78). Es war einigen von ihnen möglich, über mehr als drei Oktaven hinweg zu singen und die Töne dabei mit dem sogenannten *messa di voce*[4] extrem genau und bis zu einer Minute lang an- und abschwellen zu lassen (Fritz, 1994, S. 75 f.). Auch wiesen ihre Singstimmen einen ganz eigenen Charakter auf. So wurde das Stimmtimbre, das von der Präsenz der Obertöne geprägt ist, von Arthur Schopenhauer als „übernatürlich schön [...]“ (Kesting, 1995, S. 10) bezeichnet: „His beautiful supernatural voice cannot be compared with that of any woman's voice: there cannot exist a finer and fuller timbre, and with that silver purity, he acquires an idescribable power“ (aus Barbier, 1995, S. 229, zitiert nach Angus Heriot). Auch bei Fritz (1994, S. 75) findet sich die Aussage wieder, dass die Stimmen weder weiblich noch kindlich klangen und folglich etwas Einzigartiges darstellen. Für die in Barockopern dargestellten Götter und Helden waren Kastraten somit die perfekte Besetzung (Fritz, 1994, S. 75 ff.) und den Falsettisten gesanglich überlegen. Obwohl der Großteil der Literatur die Kastraten als übernatürliche Sänger beschreibt, gibt es auch hieran Zweifel. Baum (2012, S. 114 f.) beispielsweise bezweifelt, dass die oft genannte Zahl tausender kastrierter Kinder pro Jahr (u.a. in Heidecker, 2007, S. 9) und die außergewöhnliche gesangliche Befähigung der Wahrheit entsprechen. Auch Clapton (2004, S. 11) schreibt, dass nicht alle Kastraten großartige Sänger waren und oft die Neuartigkeit ihrer Stimmen für ihren Ruhm sorgte. Dennoch

oder Countertenor bezeichnet wird, beziehen sich dabei auf Eigenschaften wie Höreindruck oder verwendete Stimmregisterkombinationen. Der Begriff Countertenor bezeichnet im Folgenden Sänger, die durch Perfektionierung der Falsettstimme im Tenor, Alt oder gar Sopran singen. Darunter fallen somit beispielsweise Alfred Deller, als reiner Falsettist, aber auch Russell Oberlin, welcher mit seiner Bruststimme singt.

[4] An- und Abschwellen der Lautstärke (*Crescendo* und *Decrescendo*) eines lang gehaltenen Tones, wobei die Tonhöhe und der Vokal konstant bleiben (Titze et al., 1999).

(a) **(b)**

Abbildung 1.1: a) Carlo Broschi, auch Farinelli genannt, und b) Alessandro Moreschi (aus Clapton, 2004, S. 22 & 133)

kann man die Kastraten aufgrund ihrer Lebensweise und ihres Ansehens als „Popstars“ der damaligen Zeit beschreiben (Clapton, 2004, S. 14). Der wohl Bekannteste unter ihnen war Carlo Broschi (Abb. 1.1a), auch bekannt unter seinem Künstlernamen Farinelli (1705 - 1782).

Der Höhepunkt des Kastratengesangs lag um 1680 - 1780, danach nahm das Interesse langsam wieder ab. Gründe hierfür finden sich in der zunehmenden Kritik an der Kastration in der Aufklärung sowie in einem Wandel des Musikgeschmacks (Herr, 2013, S. 13, 414 ff.) und der Werte. Der Tenor wurde fortan als passender für die Heldenrolle empfunden. Im Chor der Sixtinischen Kapelle in Rom wurden Sopran- und Altstimme dennoch bis 1911 durch Kastraten besetzt (Fritz, 1994, S. 47 ff.). Die Einsicht, dass das Ritual des Kastrierens zur Produktion von Sängerkastraten unmenschlich war, führte letztendlich zu einem Verbot der Kastration, ausgesprochen von Papst Leo XIII. (Barbier, 1995, S. 238), welches nun, im Gegensatz zu vorherigen Verboten, beachtet wurde.

Einer der letzten Kastratensänger und auch Leiter des päpstlichen Chors war Alessandro Moreschi (1858 - 1922). Von ihm entstanden von 1902 bis 1904 die heute einzigen erhaltenen Gesangsaufnahmen eines echten Kastraten. Aufgrund der Aufnahmetechnik mittels Wachsschallplatte können diese Aufnahmen jedoch kein exaktes Bild einer echten Kastratenstimme vermitteln. Hinzu kommen weitere Faktoren, die Moreschi selbst betreffen, wie beispielsweise, dass er auf liturgische Musik spezialisiert war (Fuchs et al., 2000; Köwer, 2007).

1.3 Das 20. und 21. Jahrhundert

Nachdem die Kastraten an Bedeutung verloren hatten, etablierten sich im 20. und 21. Jahrhundert verschiedenste Kontratenöre wie Russell Oberlin, Alfred Deller, Jochen Kowalski, Andreas Scholl oder Philippe Jaroussky. Sie alle unterscheiden sich in ihrem Stimmtimbre und anderen gesanglichen Charakteristika. Anhand dieser Vielfalt verleihen Komponisten ihren Stücken gezielt einen bestimmten Ausdruck. Besonders bei Kontratenören ist diese Stimmcharakteristik wichtig, wobei hier oftmals der Registerwechsel beim Übergang von der Modalstimme ins Falsett entscheidend ist (Herr, 2012, S. 189). Jeder Sänger weist in seinem Ambitus bestimmte Bereiche auf, in denen seine Stimme mehr bzw. weniger klangliches Volumen erreicht (Herr, 2013, S. 475). Bekannt sind falsettierende Sänger jedoch nicht nur aus der klassischen Musik, sondern auch aus der Pop- und Rockmusik (siehe Abschnitt 1.3.2).

1.3.1 Die Rückkehr der Countertenöre

Während in vielen Teilen Europas vom 16. bis ins 20. Jahrhundert Kastraten in den Kirchenchören zu hören waren, galt dies nicht für England. Dort waren die Falsettisten als geachtete Sänger die ganze Zeit über erhalten geblieben (Herr, 2013, S. 444; Herr, 2012, S. 182). Der Kontratenor Alfred Deller (1912 - 1971) erlangte durch eines seiner Solokonzerte 1948 aufgrund seines einzigartigen Stimmklangs („Singularity of voice“ nach Herr, 2012, S. 183) große Bekanntheit und war somit einer der wichtigen Sänger des 20. Jahrhunderts, die den Kontratenören zur heutigen Beliebtheit verhalfen. Herr (2013, S. 460 ff.) entnimmt aus Dellers Biografie, dass dieser als Knabe im Kirchensopran gesungen hatte, und sein Stimmbruch sehr spät und schwach ausfiel. Danach sang er im Alt weiter, obwohl er von seiner Sprechstimme im Bariton einzuordnen war. Dellers Vermutung war, dass er bereits als Knabe falsettiert hatte und das nach dem Stimmbruch genau so weiterführte, sodass es für ihn „natürlich“ war, so zu singen. Sein Ambitus (ausschließlich Falsett) wird von Herr (2013, S. 472) mit g-ges^2 angegeben. Er selbst legte nach eigener Aussage weniger Wert darauf, die höchste Tonhöhe zu erreichen, sondern vielmehr die Resonanz seines Mitteltonbereichs stark zu nutzen. Zwei weitere wichtige Sänger der damaligen Zeit waren der Amerikaner Russell Oberlin (*1928) und der Deutsche Jochen Kowalski (*1954). Beide setzten im Vergleich zu Deller auf ein ausgeprägteres Vibrato, um ihrem Gesang einen anderen Charakter zu verleihen. Kowalskis Tief- und Mittellage war Dellers klanglich unterlegen; er setzte mehr auf die hohen Töne (Umfang: f-f^2). Oberlins Stimme ist von Natur aus höher, wodurch er angeblich mit seiner Bruststimme bis f^2 gelangt (Herr, 2013, S. 472 ff.).

Die Countertenöre für klassische Musik der heutigen Zeit legen im Gegensatz zu Deller vermehrt Wert auf die höheren Tonbereiche. Ihre Gesangsausbildung zielt darauf ab, die Schwächen der Falsettstimme durch gute Technik auszugleichen. Bekannte Sänger sind u.a. Andreas

Scholl, Philippe Jaroussky, Kai Wessel und David Daniels.

Im Film „Farinelli“ (1994) von Gérard Corbiau wurde eine Kastratenstimme aus den Gesangsaufnahmen des Countertenors Derek Lee Ragin und der Koloratursopranistin Ewa Malas-Godlewska erzeugt. Die genaue Vorgehensweise kann in Depalle et al. (1995) nachgelesen werden, vereinfachen lässt sie sich jedoch auf die folgenden Schritte: Die Musikstücke wurden auf die zwei Sänger entsprechend ihrer Stimmumfänge aufgeteilt und aufgenommen. Anschließend erfolgte ein nicht-triviales Angleichen der Stimmtimbres mittels eines sogenannten *Phase Vocoders*[5] und einer Änderung der Spektralhüllkurven. Zusätzlich wurden drei Noten, welche zu schwer zu singen waren, sogar synthetisiert. Aus dem Zusammensetzen der nun gleich klingenden Aufnahmen resultierte eine künstliche Kastratenstimme. Diese wurde zwar klanglich so angepasst, dass sie den historischen Erzählungen entspricht, jedoch hatten auch die Film- und Musikproduzenten einen Einfluss auf das Endergebnis (Depalle et al., 1995). Deshalb kann man nicht davon ausgehen, dass dies der Stimme eines echten Kastraten entspricht. Es handelt sich lediglich um eine Annäherung.

Heutzutage wird kein Kind mehr zu Gesangszwecken kastriert, sodass man nur anhand Moreschis letzter Aufnahme vage erahnen kann, wie Farinelli wirklich klang. Es gibt jedoch Sänger, die beispielsweise durch das Kallmann-Syndrom (Kallmann et al., 1944) seit ihrer Jugend einen gestörten Hormonhaushalt (Testosteronmangel) und dadurch auch einen seit ihrer Kindheit nahezu unveränderten Kehlkopf haben. Bekannte Vertreter sind Paulo Abel do Nascimento, Radu Marian und der Jazzsänger Jimmy Scott.

1.3.2 Falsettisten aus Pop und Rock

Doch nicht nur in der Klassik wird falsettiert, viele bekannte Pop- und Rockstars singen in hohen Lagen. Einer der bekanntesten war Michael Jackson (1958 - 2009). Charakteristisch für ihn waren der hohe Gesang und der Einsatz hoher Schreie, sowie seine einzigartige Selbstdarstellung und Inszenierung. Jacksons Stimme wird bei Herr (2013, S. 496 ff.) jedoch als nicht besonders gut ausgebildet und somit ohne Kraft beschrieben, weshalb auf seinen Konzerten und bei seinen Aufnahmen digitale Effekte zum Einsatz kamen, um diese Defizite zu kompensieren. Auch Sänger wie die Bee Gees, David Bowie, Nomi oder Freddy Mercury prägten die Musik mit dem Einsatz der hohen Gesangsstimme. Mercurys Stimme wird mit bis zu 4 Oktaven Umfang angegeben (MonstersAndCritics.com, undatiert), mit Sicherheit wird er aber nicht eine konstante Stimmqualität über den gesamten Ambitus entwickelt haben. Genau wie Deller hatten Jackson und Mercury nie Gesangsunterricht genommen.

[5]Ermöglicht die Manipulation der Amplituden- und Phaseninformationen eines zeitlichen Signals im Frequenzraum (Abschnitt 2.6). Durch Rücktransformation erhält man ein zeitliches Signal mit anderer Klangcharakteristik.

1.3.3 Motivation der Arbeit

Der Hauptfokus der vorliegenden Arbeit liegt in der Erstellung eines breiten Methodenspektrums (vgl. Abschnitt 3), welches in darauffolgenden Arbeiten zum Vergleich und zur Unterscheidung auf einzelne Sänger und Sängerpopulationen angewendet werden kann. So soll nach Möglichkeit das, was die Autoren der in diesem Kapitel angesprochenen Literatur in Worte fassen, auch physikalisch messbar und darstellbar sein. Die Populationen sind wie folgt aufgeteilt:

Unterscheidung nach Gesangstechnik Hierunter fallen die Countertenöre in der klassischen Musik der Amerikanischen und der Englischen Schule, aber auch Sänger mit hoher Stimme in der Pop- und Rockmusik.

Medizinische Unterscheidung In diese Population fallen der echte Kastrat Moreschi und der hormonelle Radu Marian.

Synthetisierte Kastratenstimme Die für den Film *Farinelli* erzeugte Kastratenstimme.

Frauenstimmen Diese Population soll einen Vergleich zwischen hohen Männer- und Frauenstimmen ermöglichen.

Zwei der in diesem Abschnitt angesprochenen Stimmen werden in der vorliegenden Masterarbeit exemplarisch untersucht: Russell Oberlin (vgl. Abschnitt 4.2.1) und Jochen Kowalski (vgl. Abschnitt 4.2.2).

2 Physik der Tonbildung und der Singstimme

In diesem Kapitel sollen die grundlegenden physikalischen Hintergründe von Tönen, Obertönen, Klangfarben sowie der Sprech- und Singstimme erläutert werden, welche die Basis für die Analyse von Stimmen darstellen. Weiter wird auf die Fourier-Transformation und ihre Möglichkeiten eingegangen sowie das Verfahren der Codierung durch lineare Prädiktion erläutert.

2.1 Musik-physikalische Begriffe

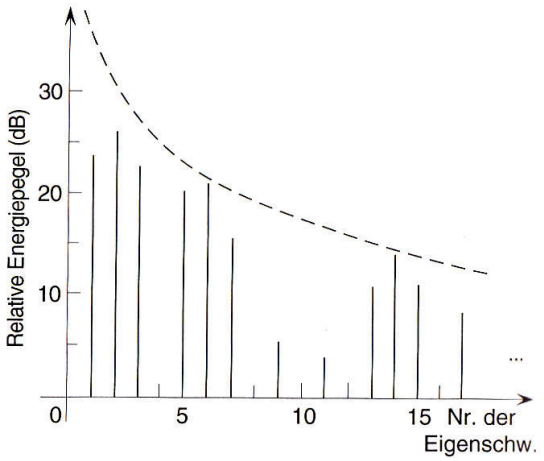


Abbildung 2.1: Beispielhaftes Spektrum an Harmonischen eines Instrumentes für einen gespielten Ton (aus Hall, 2008, S. 198)

Vergleicht man zwei Instrumente (z.B. ein Klavier und eine Klarinette), die einen Ton der gleichen Tonhöhe spielen, so unterscheidet man beide anhand ihres Klanges. Man sagt auch, sie haben unterschiedliche *Klangfarben* (*Timbres*). Selbiges gilt für Sprecher bzw. Sänger. Wenn man einen Ton hört, handelt es sich dabei nicht nur um eine einzelne Frequenz, sondern um Überlagerungen verschiedener Frequenzen mit jeweils unterschiedlichen Amplituden, also um ein Klangspektrum[6], welches das Timbre ausmacht. Bei den meisten Instrumenten wird man am Spektrum erkennen, dass die vorhandenen Frequenzen Vielfache einer Grundfrequenz f_0 sind. Diese durch unterschiedliche Eigenschwingungen erzeugten Vielfachen werden *Teiltöne* genannt. Handelt es sich um ganzzahlige Vielfache, so spricht man von *Harmonischen*, *harmonischen Teiltönen* oder auch *Obertönen*. Dabei ist zu beachten, dass die Teiltöne und Harmonischen H_N ab der Grundfrequenz gezählt werden ($H_1 = f_0,\ H_2 = 2f_0,\ H_3 = 3f_0,\ \ldots$), die Obertöne jedoch ab dem ersten Vielfachen[7] ($f_1 = 2f_0,\ f_2 = 3f_0,\ f_3 = 4f_0,\ \ldots$). Nichtharmonische (nicht ganzzahlige) Teiltöne kommen beispielsweise in Glocken und anderen

[6] In Abschnitt 2.6 wird beschrieben, wie man das Frequenzspektrum durch Fourier-Transformation erhält.

[7] Hier scheint es unterschiedliche Ansichten zu geben. Bei Hall (2008) und Rossing et al. (2014) beinhalten die Harmonischen die Grundfrequenz, bei Nair (1999) ist die erste Harmonische gleich dem ersten Oberton.

perkussiven Instrumenten vor, während die meisten nicht-perkussiven Instrumente Harmonische oder zumindest annähernd Harmonische erzeugen. Ein beispielhaftes harmonisches Spektrum an Eigenschwingungen ist in Abbildung 2.1 zu sehen.

Beim Hören kann man einem Ton eine *Tonhöhe* zuordnen. Die dabei wahrgenommene Tonhöhe ist meist die Frequenz des Grundtones. Erstaunlicherweise funktioniert dies auch dann, wenn der Grundton (oder sogar noch weitere Teiltöne) sehr schwach oder gar nicht vorhanden ist (Rossing et al., 2014, S. 130 f.). Das Gehirn ergänzt die fehlenden Teiltöne selbstständig. Man kennt dies vom Telefonieren: Übertragen wird nur das Frequenzband von 300 Hz bis 3400 Hz (Elektronik-Kompendium.de, Website), wodurch der Grundton der Stimme (etwa 100-200 Hz) nicht beim Gegenüber ankommt. Dennoch hört sich die Stimme dadurch nicht höher oder tiefer an als eine Stimme mit Grundton.

2.2 Übliche Größendarstellung in der Akustik

Energie E, Leistung P und Intensität I einer beliebigen Schwingung sind proportional zum Quadrat der Schwingungsamplitude A:

$$E \propto A^2 \qquad P \propto A^2 \qquad I \propto A^2 \tag{2.1}$$

In der Akustik müssen häufig Schwingungsamplituden A mehrerer Größenordnungen verglichen werden. Daher wird hier üblicherweise die Schallpegel-Skala mit der logarithmischen Einheit Dezibel (0.1 Bel = 1 Dezibel) eingesetzt, welche immer Bezug auf einen Referenzwert A_{ref} (bzw. E_{ref} P_{ref}, I_{ref}) nimmt. Es gilt dann beispielsweise für den Schallintensitätspegel:

$$I[\mathrm{dB}] = 10 \cdot \log \left| \frac{A^2}{A_{ref}^2} \right| \ \mathrm{dB} = 20 \cdot \log \left| \frac{A}{A_{ref}} \right| \ \mathrm{dB} = 10 \cdot \log \left| \frac{I}{I_{ref}} \right| \ \mathrm{dB} \tag{2.2}$$

So bedeutet ein Unterschied von -6 dB eine Halbierung der Amplitude (-6 dB $\approx 20\log(0.5)$), -20 dB jedoch bereits eine Reduktion auf ein Hundertstel (-20 dB $= 20\log(0.01)$). Auch eine Intensität von 0 dB bedeutet lediglich, dass die Amplitude genau den Betragswert 1 hat (0 dB $= 20\log(0)$). Der Referenzwert ist abhängig von der Anwendung. Wenn die Amplitude beispielsweise der Schalldruckamplitude p in Luft entspricht, so ist ein gängiger Referenzwert $p_{ref} = 20$ µPa (Hall, 2008, S. 94). Dies bezieht sich auf die absolute Hörschwelle des Menschen für einen Ton mit 1000 Hz. Dadurch erhält man den Schalldruckpegel $L_p = 20\log(p/p_{ref})$.

In der vorliegenden Arbeit wird mit Audiodateien ohne bekannten Referenzwert gearbeitet. Wie die Referenzwerte dennoch gewählt werden, wird in Abschnitt 3.3 erläutert.

2.3 Entstehung des Stimmschalls

Beim Ausatmen verursacht die Atemmuskulatur in der Lunge einen Überdruck und somit einen Luftstrom, der am Ende der Luftröhre durch den Kehlkopf (*Larynx*) und dann weiter durch den Vokaltrakt bis hin zu Mund und Nase gelangt und dort austritt (vgl. Abb. 2.2 links).

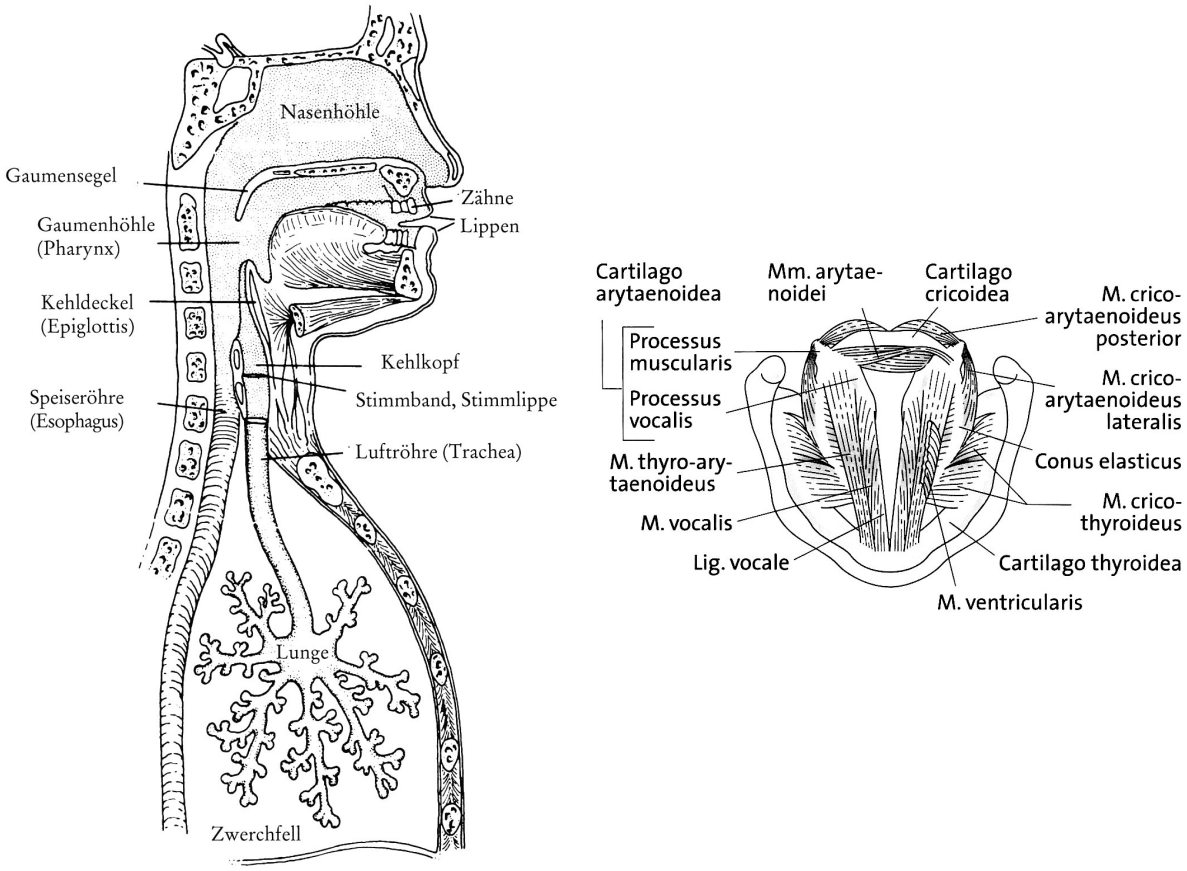


Abbildung 2.2: Links: Anatomische Übersicht über den stimmbildenden Bereich (aus Hall, 2008, S. 298)Rechts: Ansicht auf die Larynxmuskulatur von oben. Zu erkennen ist die V-förmige Glottis im offenen Zustand (aus Nawka & Wirth, 2008, S. 37).

Der Kehlkopf ist ein durch verschiedene Knorpel und Muskeln geformtes Organ und verfügt über zwei Stimmlippen, welche an der inneren Wand des Kehlkopfes anliegen und die V-förmige Stimmritze (*Glottis*) bilden (Abb. 2.2 rechts). Durch Anspannung der Kehlkopfmuskeln kann die Glottis geschlossen werden. Ab einem bestimmten Überdruck in der Lunge öffnet sich die Stimmritze, sodass die Luft hindurchströmt. Die Verengung der Glottis durch die dicht aneinander liegenden Stimmlippen erhöht die Geschwindigkeit der durchströmenden Luft und der Druck fällt, sodass ein Sog entsteht (Bernoulli-Effekt). Das Zusammenspiel dieser Bernoulli-Kraft und der angespannten Muskeln schließt die Stimmritze wieder. Hierdurch wird eine kurze Unterbrechung des Luftstroms, ähnlich der Scheidenkante einer Blockflöte oder dem Rohrblatt einer Klarinette, verursacht. Geschieht dieses Öffnen und Schließen schnell genug und periodisch, entsteht der *primäre Stimmschall* bestehend aus Grundfrequenz und Obertönen (Abb. 2.3). In der Literatur findet sich für den Intensitätsabfall der Obertöne als theoretischer Richtwert 12 dB pro Oktave (Sundberg, 1987, S. 64f.). Titze (1994, S. 120) teilt zusätzlich die Steigung in klangliche Eigenschaften ein. So ist der Stimmklang bei einem Abfall von 6 dB/Okt als metallisch („brassy“, Titze, 1994, S. 120) und bei 18 dB/Okt als flötenartig („fluty“, Titze, 1994, S. 120) zu beschreiben. Der vom subglottischen Druck

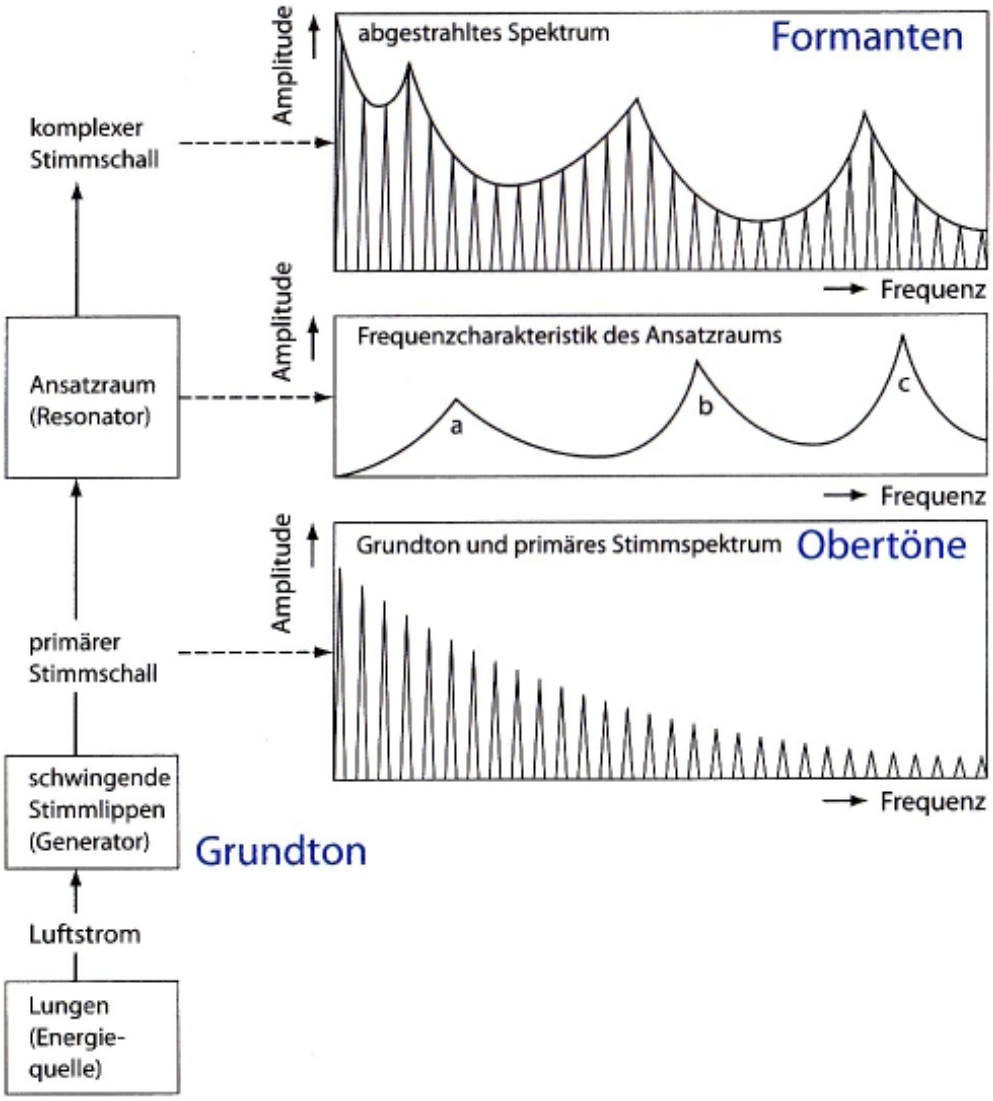


Abbildung 2.3: Schema der Stimmschallentstehung (aus Schneider-Stickler & Bigenzahn, 2013, S. 31)

abhängige Offenquotient der Glottis („Verhältnis von offenem und geschlossenem Anteil innerhalb einer Schwingungsperiode“ nach Schutte & Seidner, 2006, S. 80) beeinflusst dabei das entstehende Obertonspektrum wesentlich (Schutte & Seidner, 2006, S. 80). Abbildung 2.4 zeigt die Luftdurchflussmenge während der Phonation für drei Lautstärken und die dazugehörigen Spektren des primären Stimmschalls. Dabei ist die Glottis geschlossen, wenn der Luftstrom gleich Null ist. Besonders für laute Phonation (vgl. Abb. 2.4c) fällt auf, dass die Glottis länger geschlossen ist (fast 1/3 der Periodendauer) als bei normaler (2.4b). Für die leise Phonation ist sie nie vollständig geschlossen (2.4a).

Die Tonhöhe des Primärschalls ist abhängig vom egressiven Luftdruck, von der Länge der Stimmlippen sowie deren Masse und Spannung (der genaue Vorgang im Kehlkopf wird in Abschnitt 2.5.3 erläutert). Bei Männern sind die Stimmlippen allgemein länger und dicker

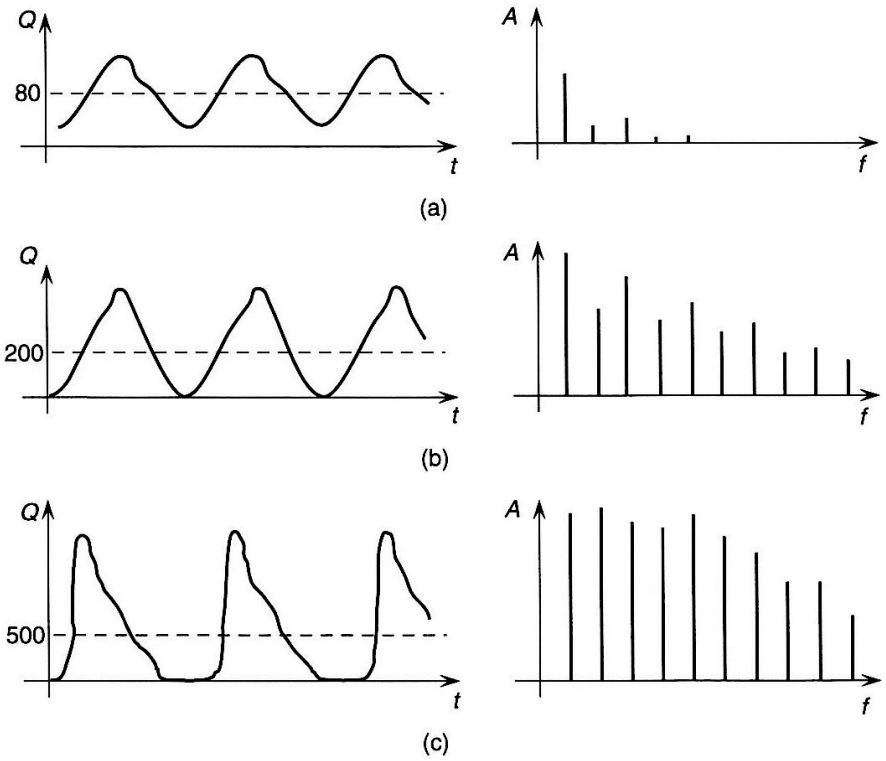


Abbildung 2.4: Links: Luftdurchflussmenge Q $[cm^3/s]$ als Indikator für das Öffnungs- und Schließverhalten der Glottis bei verschiedenen Lautstärken: a) leise, b) normal, c) laut. Während bei leiser Phonation kein Glottisschluss stattfindet, nimmt dieser bei lauter Phonation einen deutlichen Teil der Schwingungsperiode ein. Rechts: Entsprechende Spektren des primären Stimmschalls. Durch längeren Glottisschluss fällt das Spektrum obertonreicher aus (aus Hall, 2008, S. 306).

als bei Frauen und Kindern, sodass der Grundton in normaler Gesprächslage bei Männern etwa bei 100 Hz und bei Frauen etwa doppelt so hoch liegt (Nawka & Wirth, 2008, S. 101; Hollien, 2013).

Anschließend gelangt der Primärschall in den Vokaltrakt (auch Ansatzraum genannt), der die Funktion eines Resonators, also eines Frequenzfilters, auf den primären Stimmschall hat (Abb. 2.3). Denselben Effekt hat man bei einer Flöte (offener Zylinder), bei der die Länge maßgebend für die Resonanzen ist und somit bestimmt, welche Töne in der Flöte verstärkt und damit gespielt werden. Der Vokaltrakt ist beim erwachsenen Mann etwa 17-18 cm, bei Frauen etwa 14-15 cm lang (Högberg, 1995) und schließt den Rachen-, Mund- und Nasenraum ein. Der enorme Vorteil gegenüber dem starren Röhrenresonator eines Instruments ist hierbei, dass sich durch bewusste Verformung des Vokaltraktes (Position von Zunge, Gaumensegel, Lippen, Kiefer, Rachenwänden, ...) gezielt charakteristische Resonanzfrequenzen erzeugen lassen, die den Primärschall dann gemäß der aktuellen Vokaltraktform filtern (Abb. 2.3 Maxima a, b und c). Der fast ausschließlich (d.h. mit Ausnahme von geringen Schallemissionen aus Schädelknochen und Brustwänden) aus Mund und Nase austretende Schall wird *komplexer Stimmschall* genannt. Nach der *source-filter-Theorie* von Fant (1970) wird der primäre Stimmschall (source) durch den Vokaltrakt (filter) verändert und so der kom-

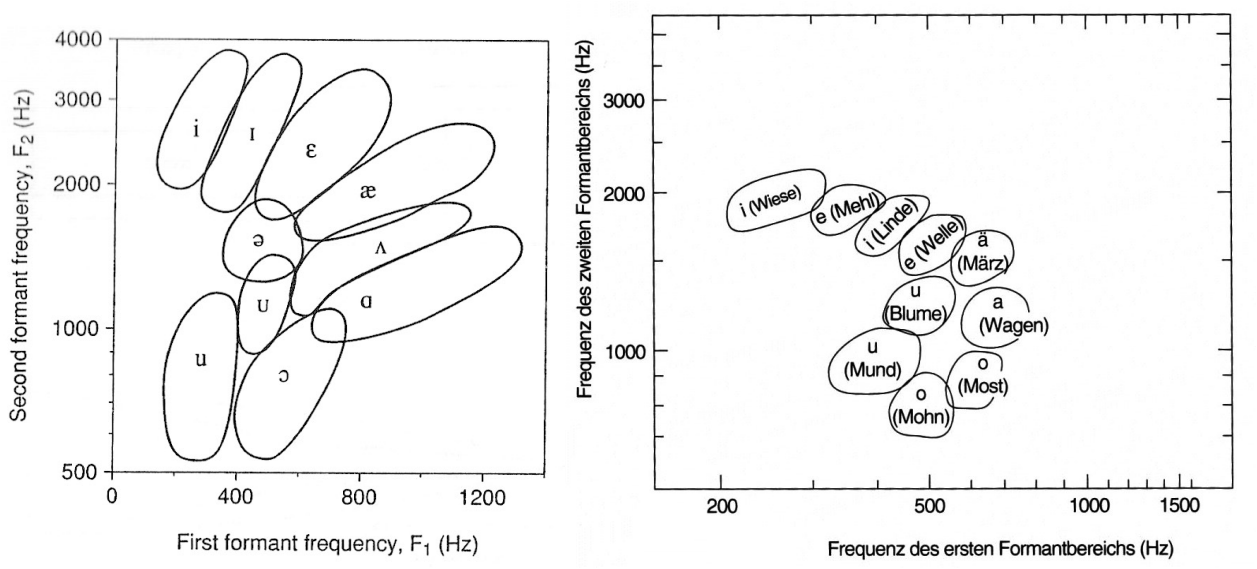


Abbildung 2.5: Typische Bereiche für Formanten verschiedener Vokale, empirisch ermittelt durch Mittelung vieler Sprecher. Links sind Formanten englischer Sprecher, rechts deutscher zu sehen (aus Titze, 1994, S. 149 und Hall, 2008, S. 309).

plexe Stimmschall erzeugt. Ihre Theorie und die Anwendung werden nochmals genauer in Abschnitt 2.7 beschrieben. Die Frequenzen, bei denen der Primärschall maximal verstärkt wird, nennt man *Formanten* oder *Formantbereiche* (Abb. 2.3). Formanten sind somit essenziell für die Erzeugung der Stimme, insbesondere der Vokale, und für das Stimmtimbre des Sprechers. Zur Bildung von Vokalen sind hauptsächlich die ersten zwei Formanten (F1, F2) ausschlaggebend, die höheren Formanten F3, F4, F5 usw. sind wichtig für die Tragfähigkeit[8] der Stimme und für das Stimmtimbre.

Dabei hängt F1 hauptsächlich von der Öffnung des Unterkiefers, F2 von der Form des Zungenhauptkörpers und F3 von der Zungenspitze ab (Hall, 2008, S. 307). Aufgrund der anatomischen Unterschiede jedes Sprechers und dessen aktueller Tagesform und auch Herkunft liegen Formanten bestimmter Vokale nicht immer bei derselben Frequenz, sondern mehr in einem bestimmten Frequenzbereich, wie es in Abbildung 2.5 dargestellt ist. Tendenziell sind Männer in dieser Darstellung in den jeweiligen Bereichen unten links, Frauen mittig und Kinder oben rechts verortet.

2.4 Entstehung der Formanten

Man kann das Entstehen der Formanten erklären, indem man sich den Vokaltrakt zunächst als einfachen Zylinder der Länge $L = 17\ cm$ mit einem offenen Ende (Mund) und einem geschlossenen Ende (Anfang des Vokaltraktes) vorstellt. Hierbei werden Frequenzen verstärkt,

[8]Nach Klingholz (2000) : „Tragfähigkeit[:] wie Durchschlagskraft oder Durchdringungsfähigkeit akustisch nicht definierter Begriff, der die Reichweite der Stimme und ihr Vermögen, Musikinstrumente zu übertönen, charakterisieren soll“. Pahn et al. (2001) geben (ausdrücklich für die Sprechstimme) eine quantitative Möglichkeit der Bewertung der Durchdringungsfähigkeit an.

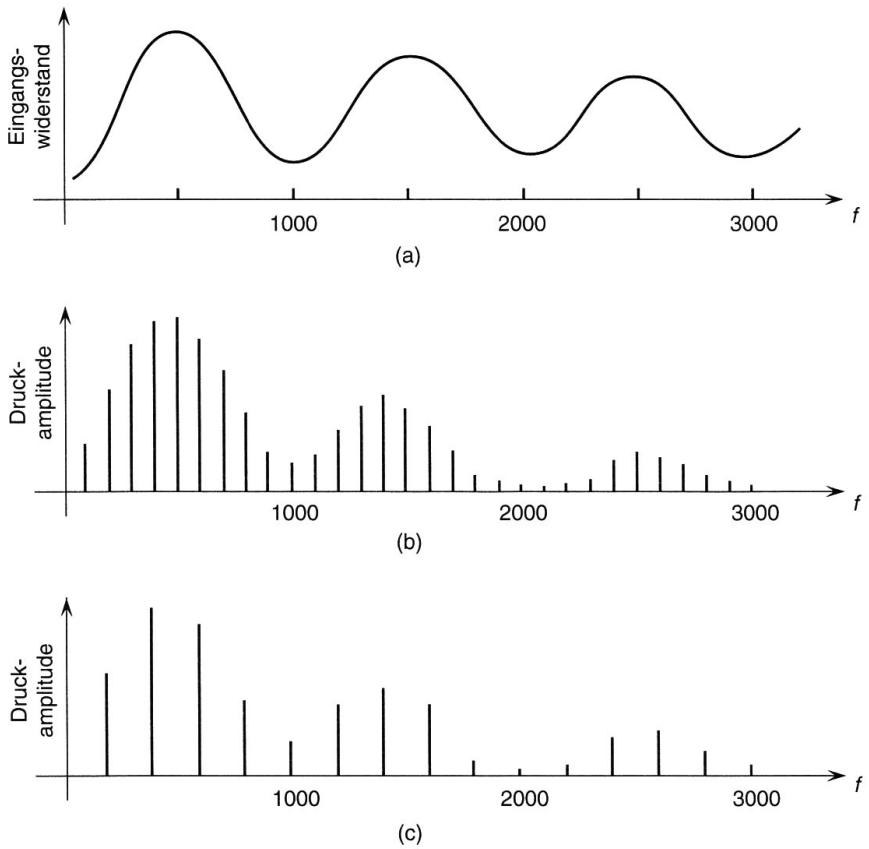


Abbildung 2.6: a) Filterfunktion des Vokaltraktes ausgehend von einem zylinderförmigen Modell mit 17 cm Länge. b) Spektrum an Teiltönen nach Verlassen des Ansatzrohres. f_1 hat eine Frequenz von 100 Hz, es wird von einer mittleren Lautstärke wie in Abbildung 2.4b) ausgegangen (jedoch ohne die ungeraden Teiltöne abzuschwächen). c) Selbiges mit einer Grundfrequenz von 200 Hz (aus Hall, 2008, S. 307).

für die gilt:

$$F_N = \frac{Nc}{4L} \qquad \text{mit } N = 1, 3, 5, \ldots$$

Mit einer Schallgeschwindigkeit von $c = 340$ m/s in Luft erhält man so $F_1 = 500$ Hz, $F_3 = 1500$ Hz und $F_5 = 2500$ Hz. Untersuchungen zeigen, dass die ersten drei Formanten bei Bildung des Vokals /e/ typischerweise um 500 Hz, 1800 Hz und 2500 Hz (Rossing et al., 2014, S. 290) liegen und damit bereits ziemlich nah an den Ergebnissen des einfachen Modells sind.

Die Resonanzspitzen (a, b, c in Abb. 2.3) sind jedoch breit und nicht so schmalbandig wie sie es beispielsweise bei einer Röhre aus Holz sind, denn das weiche Gewebe der Schleimhaut absorbiert die Schallenergie (Hall, 2008, S. 304). Dadurch fallen die Resonanzspitzen flacher und breiter aus und Obertöne zwischen den Formanten sind im Spektrum immer noch zu erkennen. Für die halboffene Röhre ergibt sich die in Abbildung 2.6a dargestellte Resonanzkurve. Formanten sind somit relativ breite Frequenzbereiche (vgl. Abb. 2.6b,c), wobei sich

die angegebene Frequenz auf ihr Maximum bezieht. Ebenfalls ist in der Abbildung zu erkennen, dass die Formanten weitestgehend unabhängig von der Grundfrequenz sind. Erhöht man den Grundton von 100 Hz auf 200 Hz, so verdoppeln sich die Abstände der Teiltöne, die Formaten bleiben jedoch an derselben Position. Anders ist dies beispielsweise, wenn man Helium einatmet, denn hier ändert sich die Schallgeschwindigkeit (etwa $c \approx 1000\ m/s$) im Ansatzraum, und die Formanten verschieben sich zu höheren Frequenzen. Weitere Modelle zur Berechnung von Formanten verwenden beispielsweise zwei Zylinder mit unterschiedlichen Durchmessern (vgl. Abb. 2.7) und simulieren dadurch Eng- und Weitstellen im Vokaltrakt.

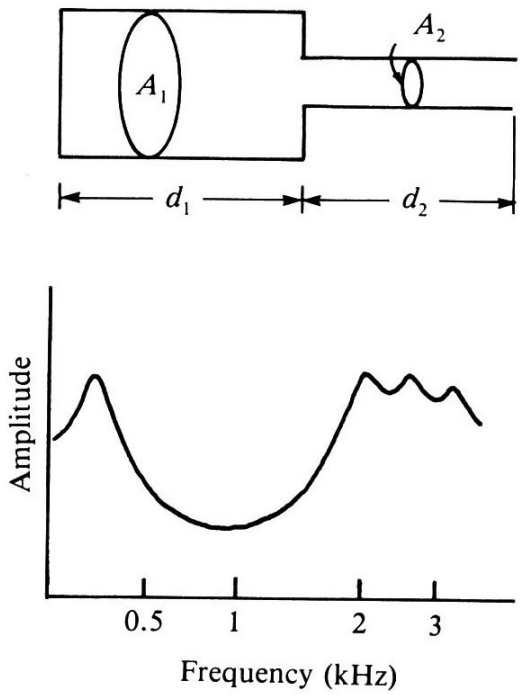


Abbildung 2.7: Weiteres Modell bestehend aus einer Röhre mit unterschiedlichen Radien zur Berechnung von Formanten (hier zum Vokal /i/) (aus Rossing et al., 2014, S. 352)

2.5 Gesang

Singen und Sprechen unterscheiden sich physikalisch und physiologisch auf den ersten Blick nicht, weisen bei genauerer Betrachtung jedoch wesentliche Unterschiede auf.

2.5.1 Sängerformanten

In Oper und klassischer Musik singen die Sänger meist ohne Mikrofon und sind dennoch hörbar, obwohl sie gegen ein ganzes Orchester „ansingen“. Diese enorme Tragfähigkeit der Stimme kommt durch eine Vergrößerung der Resonanzräume beim Tiefertreten des Kehlkopfes zustande. Vergleicht man Singen und Sprechen, so ist bei professionellen Sängern der Kehlkopf niedriger, der Kiefer ist weiter geöffnet, und die Zungenspitze sowie die Lippen sind bei bestimmten Vokalen weiter vorn (Sundberg, 1974). Besonders durch den tieferen Kehlkopf entsteht im Rachen ein weiterer kleiner Resonanzraum (Sundberg, 1974), der vor allem

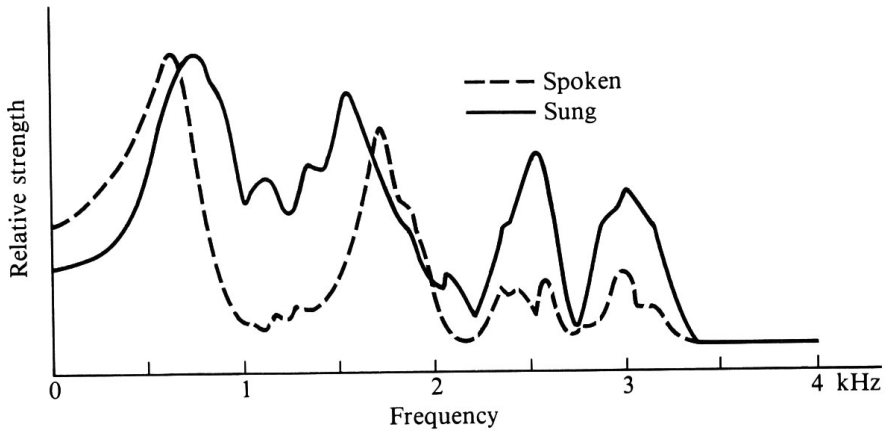


Abbildung 2.8: Spektrum eines von einem professionellen Sänger gesprochenen und gesungenen Vokals /ae/. F1 und F2 sind leicht verschoben, F3 und F4 sind beim Singen deutlich verstärkt (aus Rossing et al., 2014, S. 379).

Frequenzen von 2500 Hz bis 4000 Hz verstärkt und sich somit im Bereich von F3 und F4, oder auch von F4 und F5 befindet (Nair, 1999, S. 202, Sundberg, 1974). Dies ist der sogenannte Sängerformant, den man deutlich erkennt, wenn man das Langzeit-gemittelte Spektrum[9] eines gesprochenen und eines professionell gesungenen Vokals aus Abbildung 2.8 vergleicht. F3 und F4 sind deutlich verstärkt, was zu einer erhöhten „Brillanz“[10] und Tragfähigkeit der Stimme führt und somit dem Sänger ermöglicht, unverstärkt neben einem vollen Orchester zu singen (Abb. 2.9). Das Orchester weist den höchsten Schallpegel im Bereich des Grundtons auf. Der Schallpegel der Obertöne nimmt mit zunehmender Frequenz stark ab. Hört man sich in einer Tonaufnahme den Sängerformant isoliert an, so klingt dieser nicht spektakulär. Dennoch reichen die etwa 10 dB Verstärkung, damit das Gehirn die Stimme als wesentlich präsenter und klarer wahrnimmt (Nollmeyer, 2013, S. 133 ff. mit dazugehörigen Videos).

Interessant ist, dass der Frequenzbereich des Sängerformanten nahe der Resonanzfrequenz des menschlichen Gehörgangs liegt, weshalb das menschliche Gehör in diesem Bereich besonders empfindlich ist (Rossing et al., 2014, S. 380). Nach Rossing et al. (2014, S. 377) zeigt sich der Sängerformant bei guten männlichen Sängern besonders im Modalregister, bei Sopranstimmen und Falsettisten ist er jedoch weniger stark ausgeprägt.

2.5.2 Formantentuning

Professionelle Sänger erhöhen mit dem sogenannten *Formantentuning* oder *Formantenschieben* die Lautstärke ihrer Stimme. Denn je höher der Grundton liegt, desto weiter liegen die Obertöne auseinander. Somit fallen die Teiltöne nicht zwangsläufig in die Formantenbereiche, wodurch „Stimmpotenzial“ verschenkt wird. Im Sopran und bei Kontratenören ist dieser Effekt besonders markant, denn wenn die Grundfrequenz oberhalb des ersten Formanten

[9] Siehe Abschnitt 2.6.5

[10] Wie auch Tragfähigkeit ist Brillanz nicht akustisch definiert, sondern mehr ein Höreindruck für einen klaren Ton.

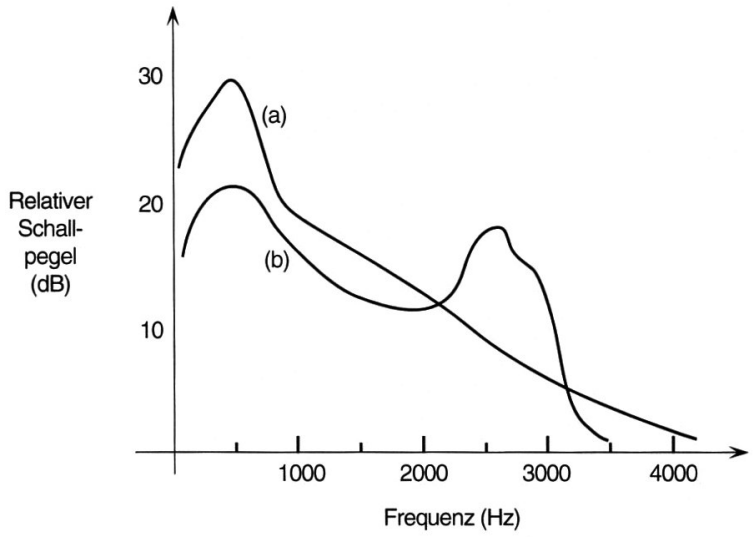


Abbildung 2.9: Über einen längeren Zeitraum gemitteltes Spektrum a) eines Orchesters und b) eines professionellen Operntenors (aus Hall, 2008, S. 314)

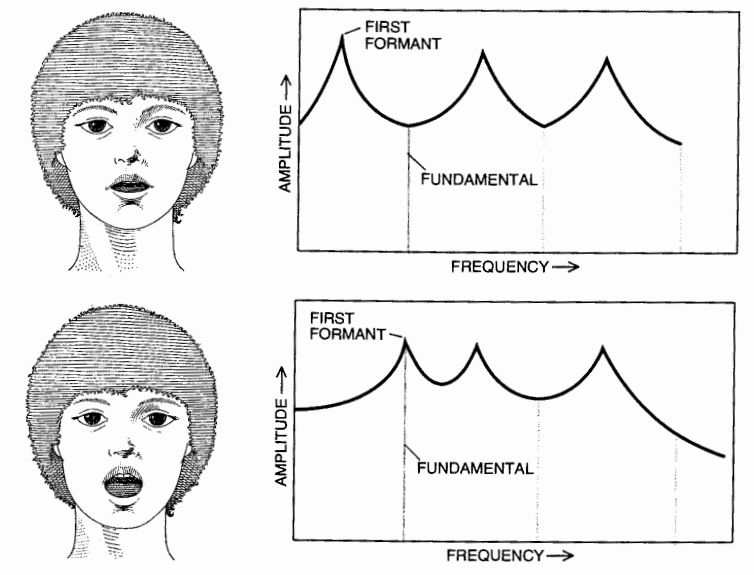


Abbildung 2.10: Verschiebung des ersten Formanten auf die Grundfrequenz durch Absenken des Unterkiefers (aus Sundberg, 1977)

liegt, kommt von diesem keinerlei Verstärkung (Rossing et al., 2014, S. 382), was in Abbildung 2.10 dargestellt ist. Ein professioneller Sänger kann nun, zum Beispiel durch Öffnen des Unterkiefers, seinen Vokaltrakt so anpassen, dass der erste Formant zur Grundfrequenz „hinverschoben“ und diese damit maximal verstärkt wird. Verschoben werden meist nur F1 oder F2 (und zwar in den höheren Frequenzbereich), wobei eine „Verschiebung nach unten“ (d.h. in tiefere Frequenzbereiche) ebenfalls möglich ist (Rossing et al., 2014, S. 379 f.). Durch das Formantentuning ändert sich zwangsläufig auch der Klang gesungener Vokale. Dennoch werden die gesungenen Vokale in den meisten Fällen trotzdem verstanden, da die Klangver-

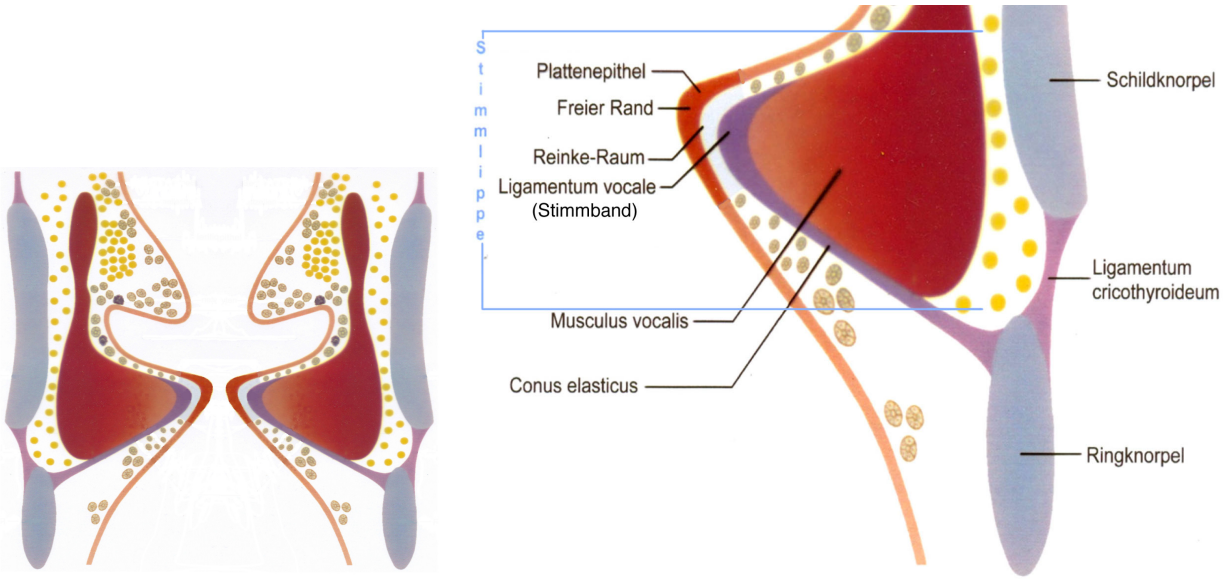


Abbildung 2.11: Links: Frontalschnitt des Kehlkopfes (ohne Epiglottis). Rechts: Ausschnitt des Schichtaufbaus einer Stimmlippe (nach Lehmann et al., 1981, S. 15)

änderungen nicht allzu groß sind und man Rückschlüsse aus dem Kontext ziehen kann (im Gegensatz zu isoliert gesungenen Vokalen). In den höchsten Sopranlagen kommt es allerdings vor, dass Vokale durch Formantenschieben so stark verändert werden, dass der Ausgangsvokal nicht mehr erkennbar ist. Aus diesem Grund ist nach Rossing et al. (2014, S. 380) der unverständliche Text solcher Gesangspassagen entweder unwichtig oder wird nochmals in einer tieferen Tonlage wiederholt.

2.5.3 Stimmregister

Ein Stimmregister ist nach Schürenberg (1989, S. 22) definiert als „eine Reihe aufeinanderfolgender, nach einem ähnlichen physiologischen Prinzip gebildeter Töne“, bzw. nach Nadoleczny (1923, S. 45) als „eine Reihe von aufeinanderfolgenden gleichartigen Stimmklängen, die das musikalisch geübte Ohr von einer anderen sich daran anschließenden Reihe ebenfalls unter sich gleichartiger Klänge an bestimmten Stellen abgrenzen kann. Ihr gleichartiger Klang ist durch ein bestimmtes konstantes Verhalten der Obertöne bedingt.“

Man differenziert dabei nach Schürenberg (1989, S. 22) zwischen Brustregister (auch Vollregister oder Modalregister) und Kopfregister (auch Randregister). Um die beiden Register zu unterscheiden, ist es notwendig, sich die physiologischen Vorgänge im Kehlkopf bei Verwendung des Brust- und Kopfregisters genauer anzusehen: Die Stimmlippen, wie in Abbildung 2.11 dargestellt, weisen einen Schichtaufbau auf. Die Schleimhautschicht (auch *Cover* genannt) ist durch eine Verschiebeschicht (Reinke-Raum) von der Unterlage (auch *Body* genannt, bestehend aus Stimmband und dem M. vocalis) getrennt. Strömt Luft durch den Kehlkopf, so können durch muskuläre Vorspannung und den Bernoulli-Effekt die Stimmlippen zum Schwingen gebracht werden. Dabei können Body und Cover gegeneinander in Schwingung versetzt werden. Im Vollregister schwingen beide, Body und Cover, im Randregister lediglich

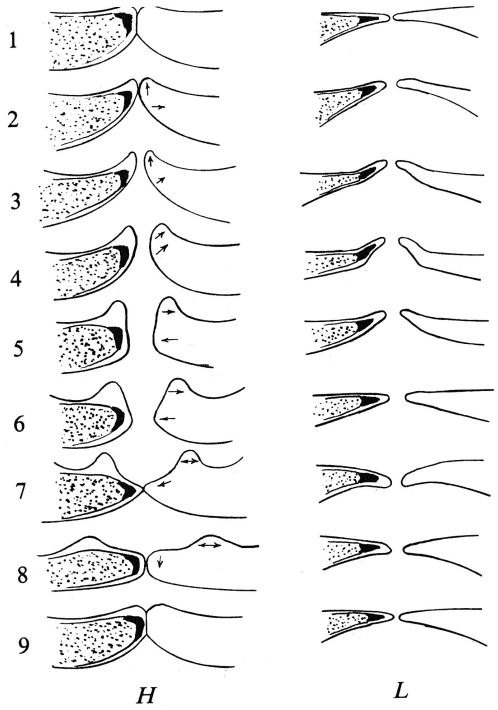


Abbildung 2.12: Seitenansicht des Schwingungsvorgangs der Stimmlippen bei Verwendung der Bruststimme (H) und der Kopfstimme (L) (aus Rossing et al., 2014, S. 387)

der freie Rand der Schleimhaut. Die erzeugte Grundfrequenz des primären Stimmschalls wird dabei höher, je höher die Spannung der Stimmlippen, eingestellt durch den M. anticus (M. cricothyreoideus) und den M. vocalis, und je geringer die effektive schwingende Masse der Stimmlippen ist. Die grobe Vorspannung erfolgt durch den M. anticus (Approximierung von Schild- und Ringknorpel durch Kippung und Anhebung des Ringknorpels gegen den fixierten Schildknorpel), während der M. vocalis die Feinabstimmung übernimmt. Das Einzigartige am M. vocalis ist, dass er durch die zopfartige Verwringung seiner Muskelstränge die Spannung erhöhen kann, ohne sich wesentlich zu verkürzen (weitgehend isometrische Kontraktion).

Im Vollregister schwingt somit viel Masse und erzeugt einen tiefen Ton, der sowohl durch Anspannung des M. vocalis als auch durch Anspannung des M. anticus erhöht werden kann. Man spürt durch Vibrationen auf der Brust, dass dort bei tiefen Tönen die größte Resonanz stattfindet, weshalb man auch von der Bruststimme spricht. Während des Schwingungsvorgangs (Abb. 2.12 links) tritt in der Regel ein vollständiger Glottisschluss auf. Dabei ergeben sich nach Nawka & Wirth (2008, S. 95) und Rossing et al. (2014, S. 388 f.) deutlich ausgeprägte Obertöne (Abb. 2.4c).

Dieser Zusammenhang zwischen Glottisschluss und der Ausprägung der Obertöne wird bei Titze (1994, S. 115 ff.) so erklärt: Das Schließverhalten der Glottis definiert die Wellenform der Schallwellen durch die Veränderung des Luftstromes. Durch diesen Zusammenhang tendieren die erzeugten Schallwellen bei längeren geschlossenen Phasen zu komplexeren Formen (vgl. Abb. 2.4c). Der Zusammenhang mit den dadurch entstehenden höheren Frequenzkompo-

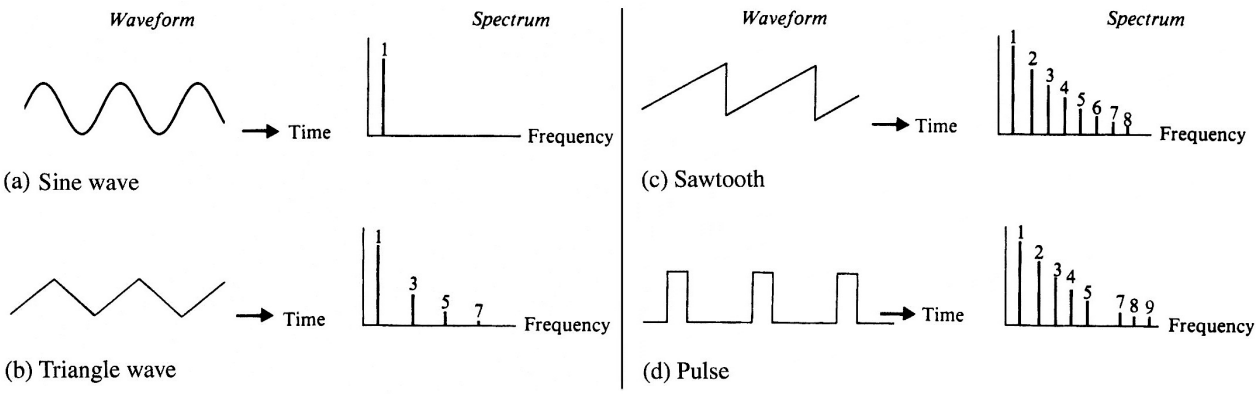


Abbildung 2.13: Spektren verschiedener Wellenformen (nach Rossing et al., 2014S. 137)

nenten lässt sich mit Abbildung 2.13 erklären. Eine einfache Sinusschwingung enthält genau eine Frequenzkomponente (vgl. Abb. 2.13a), während komplexere Formen als Überlagerung mehrerer Sinusschwingungen unterschiedlicher Amplitude und Frequenz beschrieben werden können (vgl. Abb. 2.13b, c, d). Somit erzeugt ein kleinerer Offenquotient komplexere Wellenformen und stärkere Obertöne.

Im Gegensatz zum Vollregister ist beim Randregister nur wenig schwingende Masse vorhanden, und der M. vocalis stärker angespannt (Abb. 2.12). Durch die geringe Masse kann die Grundfrequenz erhöht werden. Der Glottisschluss ist jetzt meist nicht mehr vollständig gegeben, was wiederum Auswirkungen auf die Modulation des egressiven Luftstroms und somit den Primärschall hat. Während der Grundton nun stark betont wird, fallen nach Nawka & Wirth (2008, S. 96 f.) und Rossing et al. (2014, S. 386 ff.) die Obertöne deutlich schwächer aus.

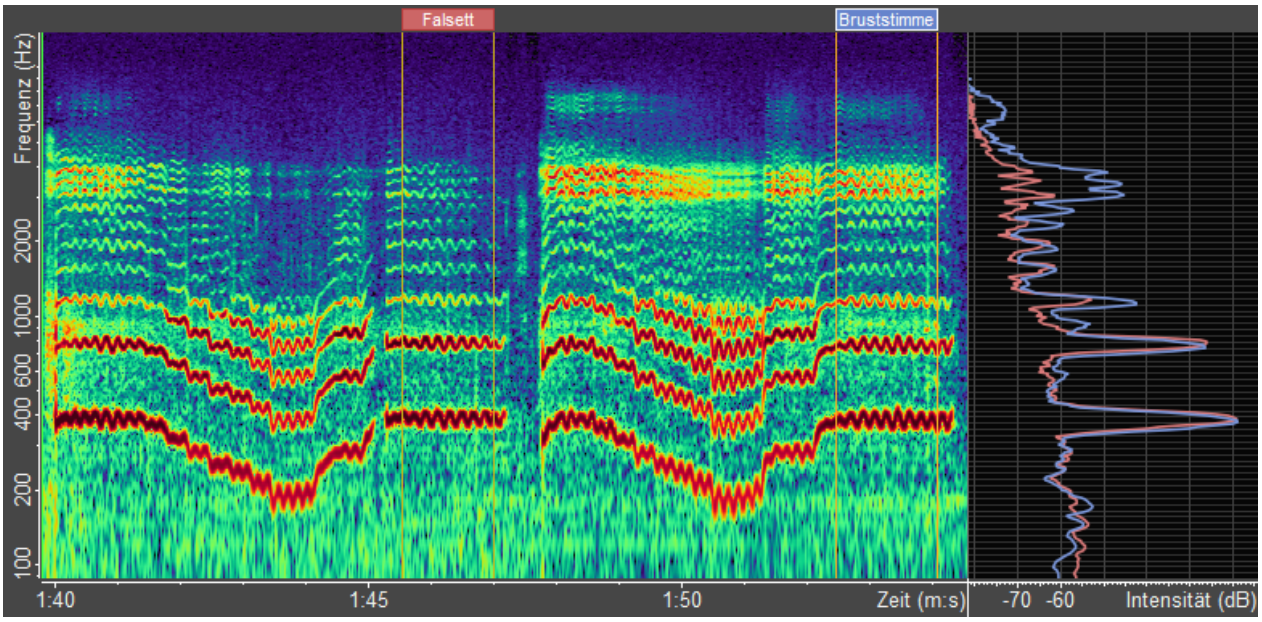


Abbildung 2.14: Mit dem Overtone Analyzer erstelltes Spektrogramm einer Vokalise, gesungen von Russell Oberlin, zunächst im Falsett und dann in Modalstimme. Rechts zu sehen sind die zeitlich gemittelten Spektren in den jeweiligen Registern. Rot entspricht der Falsettstimme, Blau der Modalstimme.

In einem Videointerview[11] mit Russel Oberlin singt dieser eine Vokalise einmal im Falsett und einmal in der Modalstimme. Das Spektrogramm[12] der Vokalisen und die beiden zeitlich gemittelten Spektren sind in Abbildung 2.14 zu sehen. Man erkennt in der spektralen Mittelung deutlich, dass sich die Spektren[13] bis etwa 2000 Hz kaum unterscheiden. Ab 3000 Hz zeichnet sich ein sehr klarer Sängerformant für die Bruststimme ab. Auch im Spektrogramm ist der durchgehende Sängerformant um 3000 - 4000 Hz deutlich zu erkennen. Da Herr Oberlin jedoch fast ausschließlich in der Modalstimme singt, kann nicht aufgrund dieses Bildes allein darauf geschlossen werden, dass der fehlende Sängerformant allgemein für das Falsett gilt. Eventuell erreicht ein Sänger wie Alfred Deller, der ausschließlich falsettiert, durch eine ausgereifte Falsett-Technik einen Ausgleich des fehlenden Sängerformanten.

[11] Youtube.com, Titel des Videos: „Russell Oberlin explica o que é um contratenor“, `https://www.youtube.com/watch?v=2YgrPBTRjMk`, zuletzt aufgerufen am: 13.04.2015.

[12] Spektrogramme werden in Kapitel 2.6.4 erklärt.

[13] Siehe 2.6.5

2.6 Fourier-Transformation (FT)

Um die Obertöne von Stimmen analysieren und darstellen zu können, muss das zeitliche Signal der Aufnahme mit der *Fourier-Transformation (FT)* in den Frequenzraum transformiert werden. Ein kontinuierliches (zunächst mathematisches) Signal $x(t)$ in Abhängigkeit der Zeit lässt sich durch das Fourier-Integral

$$X(f) = \int_{-\infty}^{+\infty} x(t) \exp(-2\pi i f t) dt \tag{2.3}$$

in den Frequenzraum (in Abhängigkeit der Frequenz $f = \omega/2\pi$) transformieren (Meyer, 2000, S. 27). Die Fourier-Transformierte $X(f)$ ist komplex und stellt ein kontinuierliches Frequenzspektrum dar. Für eine einfache Cosinusschwingung $cos(2\pi f_0 x)$ ergäben sich im Frequenzraum beispielsweise zwei δ-Peaks an den Stellen $\pm f_0$. Der komplexe Wert von $X(f)$ ließe sich auch über eine Phase definieren, was jedoch in dieser Arbeit keine Anwendung findet. Die Energie eines Signals kann sowohl über den Frequenzraum als auch über den Zeitraum bestimmt werden. Hier gilt das Parsevalsche Theorem:

$$\int_{-\infty}^{+\infty} |x(t)|^2 dt = \int_{-\infty}^{+\infty} |X(f)|^2 df \tag{2.4}$$

Dabei ist $|X(f)|^2$ die *Energiedichtefunktion*, welche die Energie des Signals im Frequenzintervall f bis $f + df$ beschreibt (Hoffmann, 1998, S. 137).

2.6.1 Diskrete Fourier-Transformation

Reale Signale (wie z.B. ein Audiosignal) sind jedoch nicht kontinuierlich messbar, daher muss die mathematische Fourier-Transformation angepasst werden. Signale werden mit einer Abtastrate f_s abgetastet und sind dadurch *diskret*. Dies kann mit Abbildung 2.15a verdeutlicht werden:

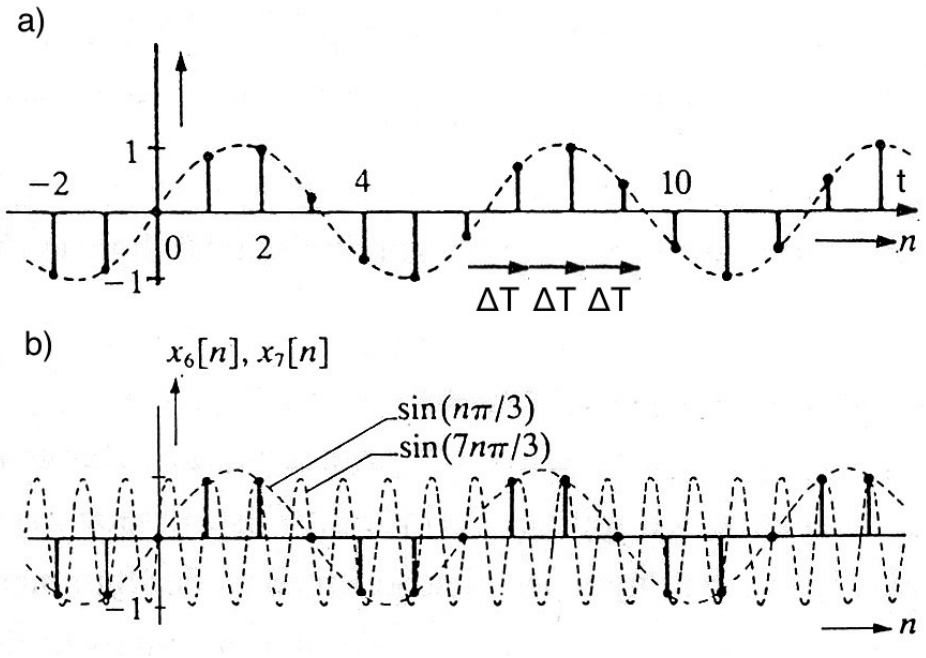


Abbildung 2.15: a) Kontinuierliches (gestrichelt) und diskretes (senkrechte Striche mit Punkten) Signal einer Sinuskurve. b) Diskretes Signal, welches zu zwei unterschiedlichen kontinuierlichen Sinuskurven gehören könnte (nach v. d. Enden & Verhoeckx, 1990, S. 57).

Hier deutet die gestrichelte Sinuskurve ein kontinuierliches Signal an, während die tatsächlich erfassten Amplitudenwerte nur die Punkte im Abtastinterval $T = 1/f_s$ sind. Das diskrete Signal ist also

$$x[n] = x(nT) \qquad n = ..., -2, -1, 0, 1, 2, ... \tag{2.5}$$

Ebenfalls zu beachten ist, dass die Zeitabschnitte nicht zwangsläufig genau auf das Ende der Periode fallen. Wird die Abtastrate nicht genügend hoch gewählt, gehen Informationen verloren, denn der Wert der Amplitude zwischen den Punkten ist dann ungewiss, sodass der wahre Verlauf anders aussehen könnte. Dies ist in Abbildung 2.15b zu sehen. Für dieselben Abtastpunkte kann beispielsweise noch eine Sinuskurve mit höherer Frequenz untergebracht werden. Bei CD-Aufnahmen ist die Abtastrate üblicherweise 44100 Hz.

Mit diskreten Signalen wird die Fourier-Transformation nach Meyer (2000, S. 137) 2.3 zunächst zur *Fourier-Transformation für Abtastsignale* (FTA)*:*

$$X_n(f) = \sum_{n=-\infty}^{+\infty} x(nT) \exp(-2\pi i n f T) = \sum_{n=-\infty}^{+\infty} x[n] \exp(-2\pi i n f T) \tag{2.6}$$

Es ist jedoch noch nicht praktikabel, von $-\infty$ bis $+\infty$ zu summieren. Deshalb wird zunächst davon ausgegangen, dass das Signal periodisch nach N Abtastwerten sei (und N sei gerade). Dann ist es ausreichend, sich genau dieses „Zeitfenster“ der Länge NT Sekunden herauszunehmen (vgl. Abb. 2.16a), und man erhält nach Meyer (2000, S. 147):

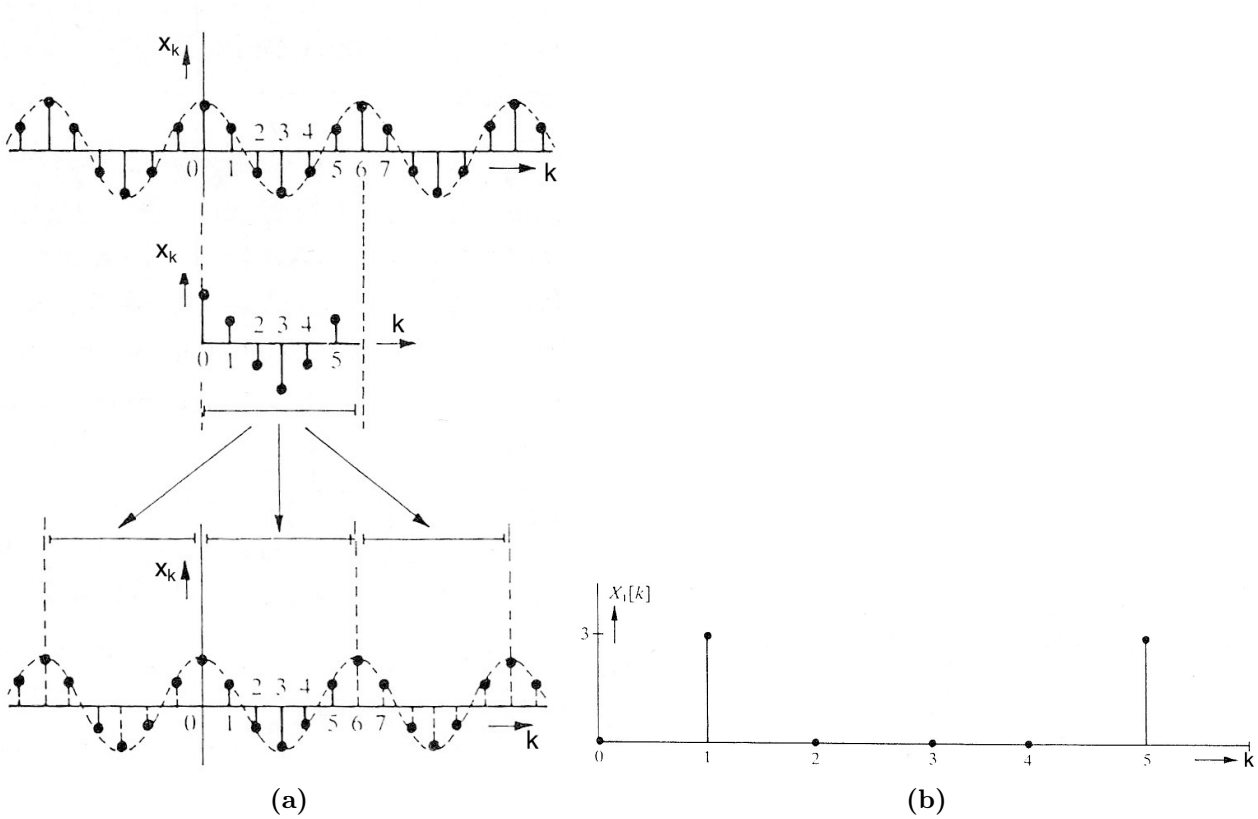


Abbildung 2.16: a) Verwendung eines periodischen Signalausschnittes der Länge N = 6 aus einem diskreten Signal. b) Das Spektrum zum Signal in a) (aus v. d. Enden & Verhoeckx, 1990, S. 131 & 133).

$$X_n(f) = \sum_{n=0}^{N-1} x[n] \exp(-2\pi i n f T) \tag{2.7}$$

Aus den N Abtastwerten von $x[n]$ ergeben sich N komplexe spektrale Amplitudenwerte auf der Frequenzachse

$$X[k] = X_n(k/NT) \qquad \text{mit } k = 0, 1, ..., N-1 \tag{2.8}$$

im gleichmäßigen Frequenzintervall von

$$1/(NT) \tag{2.9}$$

was gleichzeitig der maximalen Frequenzauflösung entspricht, und dadurch die *Diskrete Fourier-Transformation (DFT)* ergibt:

$$X[k] = \sum_{n=0}^{N-1} x[n] \exp\left(-2\pi i \frac{kn}{N}\right) \tag{2.10}$$

wobei

$$f = k/(NT) \tag{2.11}$$

gilt. Dabei wird anschaulich der gewählte Ausschnitt periodisch fortgesetzt. Dies ermöglicht die Transformation beliebiger Signalausschnitte, was in Abschnitt 2.6.3 erläutert wird. Auch hier gilt das Parsevalsche Theorem:

$$\sum_{n=0}^{N-1} |x[n]|^2 = \frac{1}{N} \sum_{k=0}^{N-1} |X[k]|^2 \tag{2.12}$$

Hieraus ergibt sich wieder die Energie jeder Frequenzkomponente. Für die Berechnung der Leistung einer Frequenzkomponente oder ihrer Intensität muss entsprechend noch durch die Dauer bzw. die Dauer und eine Bezugsfläche geteilt werden. Bei entsprechend einheitlicher Berechnung sind die spektralen Amplitudenwerte als Energie, Leistung und Intensität lediglich unterschiedlich skaliert und daher auf einer Dezibel-Skala um einen konstanten Dezibel-Wert zueinander verschoben. Welche Werte für die Spektralanalysen der vorliegenden Arbeit verwendet werden, wird in Abschnitt 3.3 erläutert.

Noch zu erwähnen ist der Begriff der *schnellen Fourier-Transformation* (engl. *Fast-Fourier-Transform,* kurz *FFT*), welcher einen schnellen Algorithmus zur Berechnung großer DFTs bezeichnet. Insbesondere der von Cooley & Tukey (1965) beschriebene Algorithmus, mit N als Zweierpotenz (z.B. 1024, 2048, 4096, usw.), ist besonders effizient. Wenn im Folgenden von FFT die Rede ist, dann ist damit die Berechnung durch ein entsprechendes Programm gemeint, während DFT der mathematischen Beschreibung dient.

2.6.2 Frequenzbereiche der DFT

Bei genauerer Betrachtung des Spektrums aus Abbildung 2.16b stellt man fest, dass das Spektrum am Punkt $k = N/2$ gespiegelt werden könnte, es fehlt lediglich der Punkt $X[k = N]$. Dieser ist jedoch identisch mit $X[k = 0]$, was aus der Periodizität des Spektrums der FTA für reale $x[n]$ folgt (Meyer, 2000, S. 147). Das heißt, dass man anstatt $k = 0...N-1$ auch $k = -N/2... + N/2 - 1$ bzw. $k = -N/2 + 1...N/2$ setzen kann und mit 2.6.1 somit negative und positive Frequenzen f erhält (vgl. dazu das am Anfang des Kapitels erwähnte Beispiel der Fourier-Transformation einer Cosinusfunktion). Die negativen Frequenzen haben für die Interpretation von Klang-Frequenzen keine Bedeutung, da sie keine weiteren Informationen enthalten. Die „äußerste“ und damit höchste Frequenz heißt *Nyquist-Frequenz* und ist durch die Abtastrate f_s folgendermaßen definiert:

$$f_N = \frac{f_s}{2} \tag{2.13}$$

Das Spektrum kann maximal bis zu dieser Frequenz analysiert werden.

2.6.3 Fensterfunktionen

Durch die Anwendung der DFT auf beliebige Signale entsteht jedoch ein nicht zu vernachlässigender Nebeneffekt, der sogenannte *Leckeffekt,* welcher anhand von Abbildung 2.17a erklärbar ist: Hier wurde N so gewählt, dass der Ausschnitt bei periodischer Fortsetzung einen Sprung (vgl. rote Markierung in 2.17a) aufweist und das Spektrum somit verfälscht wird (vgl. Abb. 2.17b). Hier ist zu sehen, dass selbst für Frequenzintervalle, die eigentlich keine Frequenzen enthalten sollten, die Amplituden nicht null sind. In der Praxis ist dies unvermeidbar, da es nahezu unmöglich ist, immer perfekt die Periode zu treffen. Jedoch kann der Effekt mit geeigneten *Fensterfunktionen* verringert werden, welche den Signalausschnitt zu den Enden hin gegen Null abschwächen. Tatsächlich entspricht bereits das Anwenden der Fourier-Transformation für Abtastsignale auf einen kurzen Signalabschnitt der Anwendung eines Rechteckfensters auf das Signal, wodurch das Signal „abgehackt“ wird. Anhand von Abbildung 2.17b wird nun gezeigt, wie ein besseres Fenster den Leckeffekt reduziert:

a Vom Signal $x[n]$ wird für $n = 0, 1, ..., 15$

b ein Ausschnitt der Länge $N = 16$ “herausgenommen”.

c Das Fenster hat ebenfalls die Länge $N = 16$ Punkte und flacht an den Enden gegen 0 ab.

d Das Fenster wird mit dem Signalausschnitt multipliziert, wodurch sich

e ein beinahe-periodisches Signal ergibt,

f dessen Spektrum im mittleren Bereich wieder beinahe Null ist.

Wie man erkennen kann, ist der Leckeffekt nicht verschwunden und auch die Amplituden der Hauptausschläge (für $k = 2, 3, 13, 14$) haben sich verringert. Jedoch sind die restlichen Amplituden deutlich geringer und Frequenzen mit kleiner Amplitude können besser erkannt werden. Mathematisch kann gezeigt werden, dass die Fourier-Transformation zweier multiplizierter Funktionen der Faltung der beiden einzelnen Fourier-Transformationen entspricht (Hoffmann, 1998, S. 183):

$$FT(x \cdot w) = FT(x) * FT(w) \tag{2.14}$$

Hierbei ist x wieder das Signal und w die Fensterfunktion. Aus der FT der Fensterfunktion kann man also bereits Informationen über die Auswirkung auf das Spektrum und den Leckeffekt erhalten.

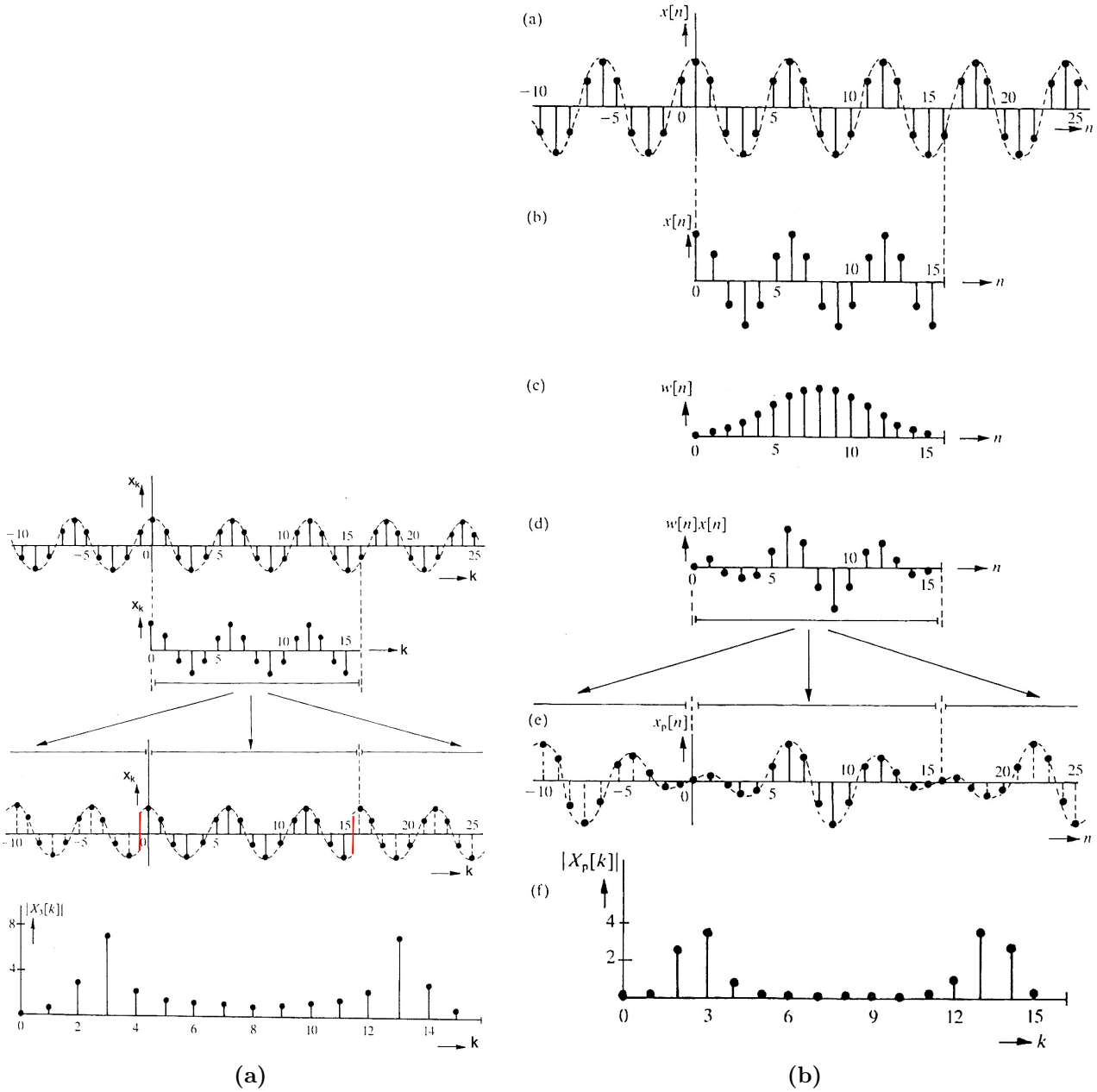


Abbildung 2.17: a) Durch einen Sprung (rote Markierung) in der periodischen Fortsetzung des ausgewählten Signalausschnitts $x[n]$ der Länge N (hier: $N = 16$) entstehen im Spektrum Frequenzanteile, die im eigentlichen Signal nicht vorhanden sind. Dies ist der Leckeffekt. b) Durch die Fensterung des Signalausschnitts mit einer Fensterfunktion $w[n]$ ist die periodische Fortsetzung besser möglich, was den Leckeffekt reduziert, jedoch nicht vollkommen eliminiert (nach v. d. Enden & Verhoeckx, 1990, S. 131 & 134).

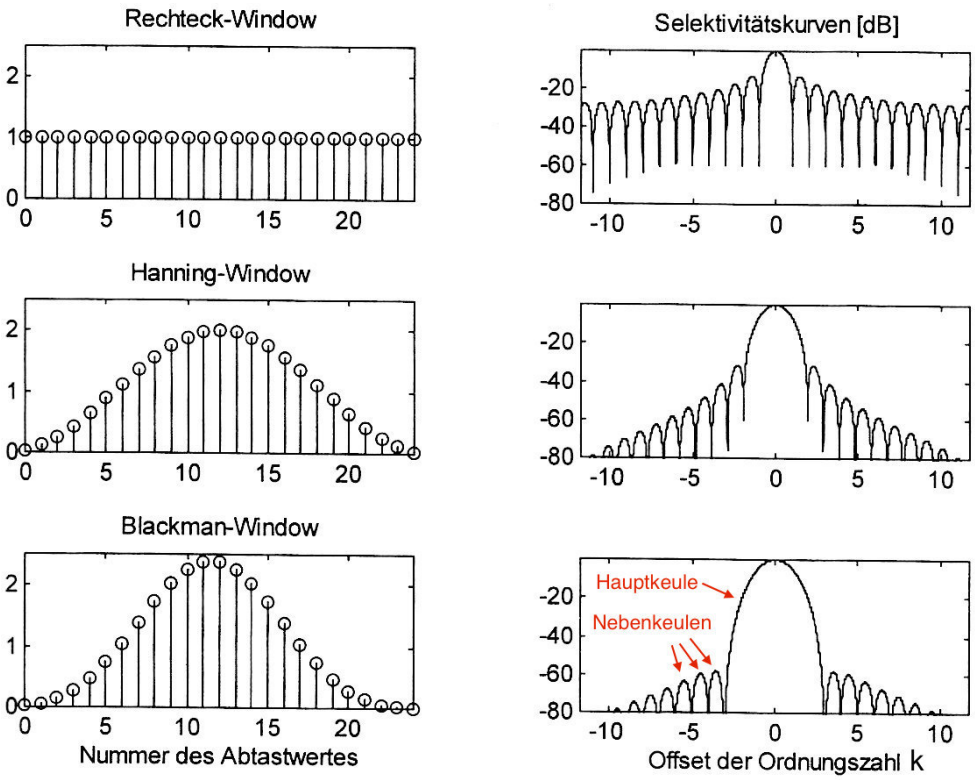


Abbildung 2.18: Verschiedene Fenster und ihre normierten Intensitäten im Frequenzraum (Ordnungszahl k). Das Rechteckfenster hat eine sehr schmale Hauptkeule, die Nebenkeulen sind allerdings sehr stark und fallen wenig ab. Beim Hanning- und Blackman-Fenster ist die Hauptkeule breiter, die Nebenkeulen fallen dafür stärker ab (nach Meyer, 2000, S. 165).

Abbildung 2.18 zeigt drei Fenster mit den normierten Intensitäten I ihrer Transformationen (im Frequenzraum kontinuierlich, aber auf die diskreten 30 Punkte im Frequenzraum zentriert verteilt), berechnet durch

$$I(f) = 20 \log \left(\frac{|W(f)|}{|W(0)|} \right) \text{ dB}$$

Die Maxima werden Keulen genannt. Charakteristische Größen sind beispielsweise die Breite der Hauptkeule, die Höhe der größten Nebenkeule, der Abfall der Intensität zu den Seiten hin, aber auch maximal mögliche Amplitudenverluste (Harris, 1978 und Nuttall, 1981). Zwar kann mit dem Rechteckfenster (Abb. 2.18, oben) durch die geringe Hauptkeulenbreite sehr genau im Frequenzraum aufgelöst werden, möchte man jedoch benachbarte Frequenzen mit geringerer Intensität untersuchen, so sind diese möglicherweise durch die hohen Nebenkeulen nicht mehr erkennbar. Zudem fallen die Nebenkeulen sehr langsam ab und verhindern dadurch möglicherweise auch die Untersuchung von Frequenzen, die weit von der Hauptkeule entfernt sind. Das Hanning- und das Blackman-Fenster in derselben Abbildung haben eine etwas breitere Hauptkeule, die Nebenkeulen fallen jedoch stärker ab, sodass Frequenzkomponenten geringerer Intensität eher erkennbar sind. Die Wahl des Fensters ist demnach je nach Anwendung unterschiedlich und kann nicht pauschal getroffen werden. Die Wahl der

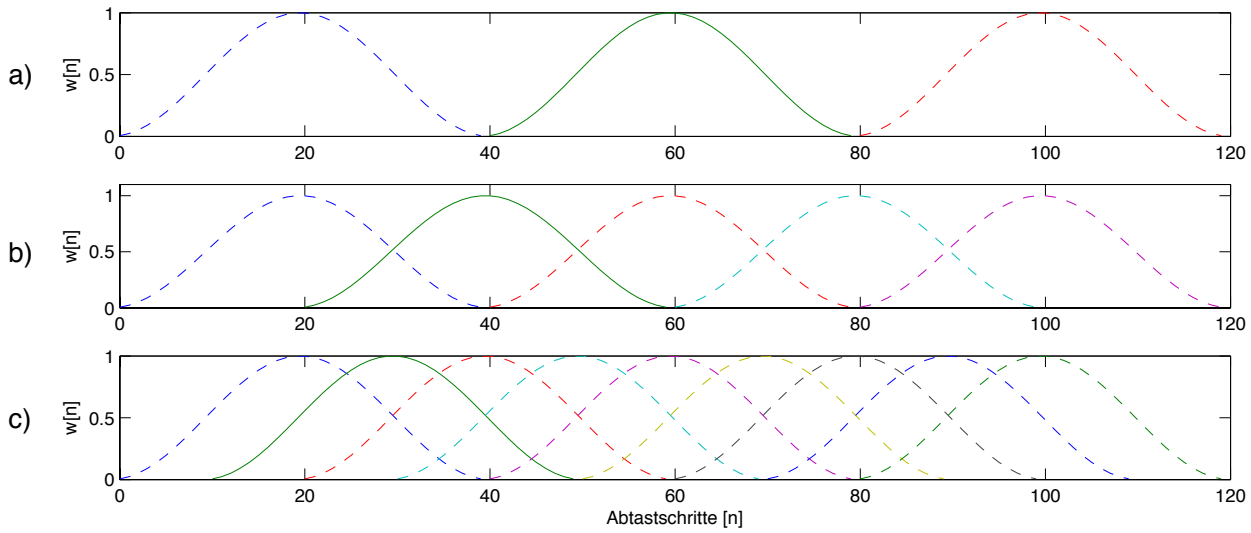


Abbildung 2.19: Zur Veranschaulichung der Überlappung von Fenstern über ein Signal. a) Keine Überlappung der Zeitfensterung, b) 50% Überlappung, c) 75% Überlappung.

Fenstereinstellungen für die spätere Analyse wird in Kapitel 3.6.1 erläutert.

2.6.4 Spektrogramme

Aus den bisher angesprochenen Spektren der DFT erhält man keine Informationen über den zeitlichen Verlauf der Frequenzen. Man kann jedoch mehrere aufeinanderfolgende Zeitausschnitte wählen und für diese jeweils separat die DFT berechnen. Dies wird in Abbildung 2.19a veranschaulicht, in der ein Zeitfenster nach dem anderen über die Abtastwerte gelegt wird. Dadurch erhält man allerdings nur verschwindend geringe Informationen über die Frequenzen an den Enden der Fenster, was durch sogenanntes *Überlappen* (engl. *overlap*) korrigiert werden kann. Anstatt den Signalausschnitt immer um eine Fensterlänge zu verschieben, wird der Ausschnitt um einen gewissen Prozentsatz der Fensterlänge verschoben. Abbildung 2.19b und c zeigen 50% und 75% Überlappung, Werte, die in der Praxis üblicherweise verwendet werden (Harris, 1978). Als visuelle Darstellungsform eignet sich hierbei besonders das *Spektrogramm*, für das in Abbildung 2.20 vier Beispiele zu sehen sind. Für die Fensterlängen sind 512 und 2048 Punkte, für die Überlappung 0 bzw. 75% dargestellt. Dabei wird auf der x-Achse die Zeit (quantisiert durch die Fensterabstände), auf der y-Achse die Frequenz und in Farben ($\widehat{=}$ z-Achse) die Intensität der Frequenzkomponenten in dB abgebildet. Der Sänger singt dabei von 0.2 s bis 2 s einen Ton mit einer Grundfrequenz von etwa 400 Hz, variiert ihn aber leicht durch ein Frequenzvibrato. In den oberen Abbildungen (Auflösung $1/NT = 44100/512 \approx 86$ Hz) erkennt man das Vibrato höchstens an den Obertönen, die ersten drei Harmonischen werden jedoch nur schlecht aufgelöst. Unten links sind die Harmonischen deutlich erkennbar (Auflösung $1/NT = 44100/2048 \approx 21$ Hz), der zeitliche Verlauf ist jedoch durch die wenigen berechneten Zeitpunkte sehr grob. Eine Erhöhung der Überlappung auf 75% (unten rechts) korrigiert dies, sodass man das Vibrato nun gut sehen

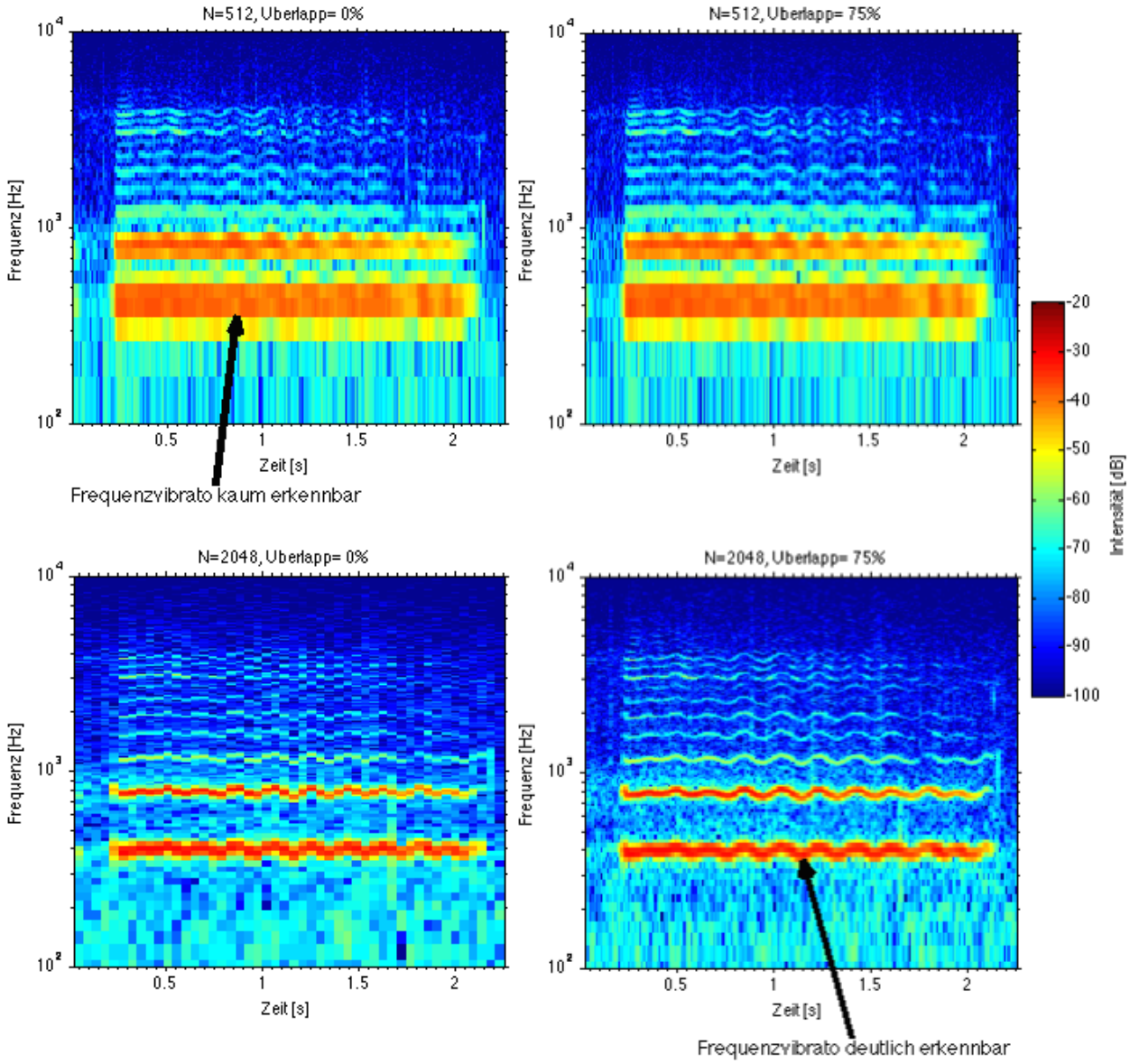


Abbildung 2.20: Spektrogramme einer gesungenen Passage mit unterschiedlicher Fensterlänge N (512 und 2048 Punkte) und Überlapp (0% und 75%) bei $f_s = 44100$ Hz.

kann. Bei dieser Abtastrate von 44100 Hz ist also die Fensterlänge von $N = 2048$ besser für eine klare Darstellung geeignet als die kürzere ($N = 512$), da hier effektiv mehr Perioden und somit Abtastwerte ins Analysefenster fallen. Die Auflösung kann jedoch in der Regel nicht beliebig gesteigert werden, da sich das Signal meist zeitlich ändert (z.B. hört der Sänger auf zu singen). Durch die sogenannte *Nullpolsterung* (engl. *zero-padding*) kann die Auflösung aber interpoliert werden, was feinere Frequenzabstufungen erlaubt. Dabei wird der eigentliche zeitgefensterte Signalausschnitt beibehalten und dann bis zur gewünschten Länge mit Nullen aufgefüllt, sodass N größer wird. Feinere Strukturen können dadurch nicht aufgelöst werden, es ergeben sich lediglich genauere Abstufungen der Intensitäten. Dies alles zeigt, wie

sorgfältig die Parameter aufeinander abgestimmt werden müssen.

Die Spektrogramm-Darstellung kann man sich genauer als dreidimensionales Höhenprofil wie bei einer Landkarte vorstellen (x = Zeit, y = Frequenz, z = Farbe = Intensität). Daher lässt sich ein Spektrogramm auch als dreidimensionale Abbildung darstellen, sodass hohe Intensitäten Bergen entsprechen und geringe den Tälern (vgl. Abb. 2.21).

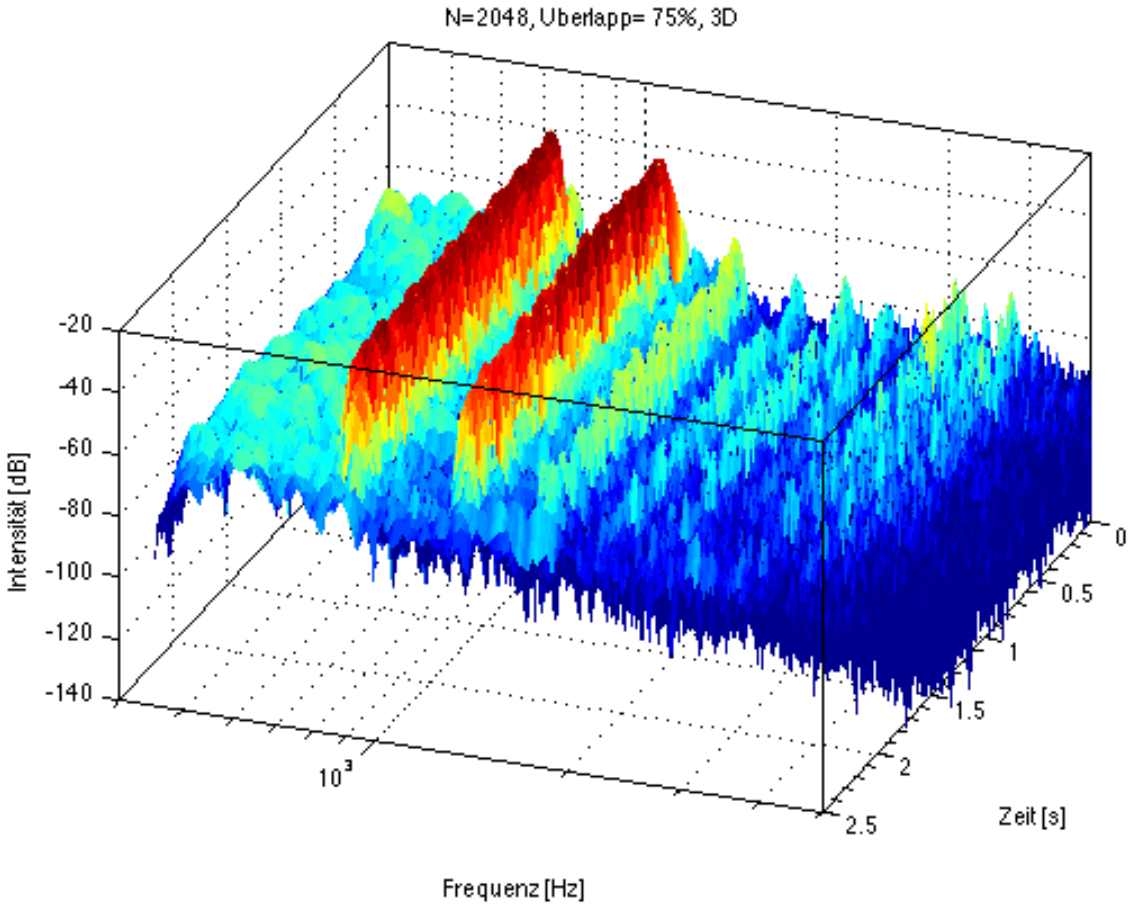


Abbildung 2.21: Die 3D Darstellung des unteren rechten Spektrogramms aus Abbildung 2.20.

2.6.5 Langzeit-gemittelte Spektren (LTAS)

Zur Bestimmung eines *Langzeit-gemittelten Spektrums* (engl. *long term average spectrum*, kurz *LTAS*) berechnet man, ähnlich wie für das Spektrogramm, viele einzelne Spektren über kurze Zeitabschnitte mit gleicher Fensterlänge. Anschließend werden die Spektren frequenzkomponentenweise gemittelt. Will man das Betragsquadrat der k-ten Frequenzkomponente über M Abschnitte mitteln, so berechnet man entsprechend

$$\frac{|X_1[k]|^2 + |X_2[k]|^2 + ... + |X_{M-1}[k]|^2 + |X_M[k]|^2}{M}$$

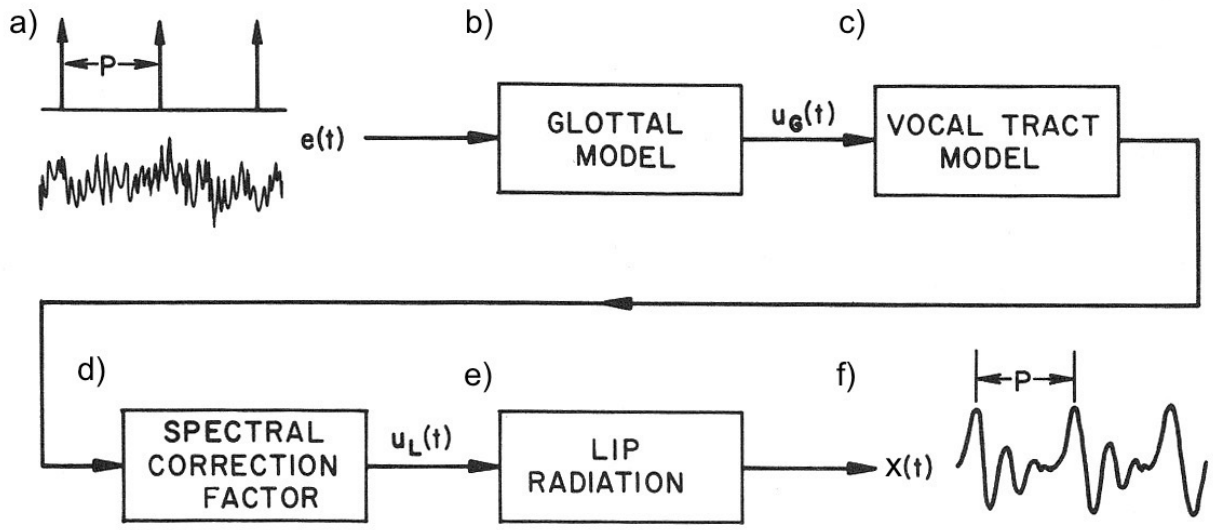


Abbildung 2.22: Das lineare Modell der Spracherzeugung geht a) von einem Eingangssignal $e(t)$ (Dirac-Stoß-Folge mit Zeitabstand P oder zufälliges Rauschen) aus. Der erste Filter b) modelliert die Wellenform des Volumenstroms an der Glottis $u_G(t)$. Der zweite c) und dritte d) Filter modellieren die Formanten und die spektrale Korrektur. Ein letzter Filter e) erzeugt aus der Wellenform des Volumenstroms an den Lippen $u_L(t)$ die austretende Schallwelle $x(t)$ mit Grundschwingung $F_0 = 1/P$ (f). (nach Markel & Gray, 1976, S. 6)

Eine Frequenzauflösung von 125 Hz ist in der Regel für die Langzeitanalyse von Stimmen ausreichend, da hier nicht gezielt einzelne Harmonische betrachtet werden (z.B. Sundberg, 2001 und Boersma & Kovacic, 2006). Nach einer Mittelung über 20-30 s ist das LTAS in der Regel für einen Sänger stabil in seiner Form (Sundberg, 2001) und zeigt charakteristische Merkmale, wie z.B. den Sängerformanten[14].

2.7 Das lineare Modell der Spracherzeugung

Zur Analyse der Gesangsstimmen wird in der vorliegenden Arbeit ein Verfahren namens *Codierung durch lineare Prädiktion* (engl. *linear predictive coding,* kurz: *LPC*) verwendet, mit dem unter anderem Formantenpositionen bestimmt werden (Abschnitt 3.6). Zum Verständnis der LPC (Abschnitt 2.7.1) ist das auf Gunnar Fant (1970[15]) zurückgehende *lineare Modell der Spracherzeugung* (*linear speech production model*) essenziell. Eine ausführliche Erklärung des LPC-Verfahrens ist im begrenzten Rahmen dieser Masterarbeit nicht möglich, da es nur einen kleinen Anteil der Spektralanalysen einnimmt. Beim linearen Modell der Spracherzeugung handelt es sich vereinfacht gesprochen um eine Anordnung hintereinandergereihter Filter[16], die ein bekanntes Eingangssignal $e(t)$ so verändern, dass sich ein Ausgangssignal (Sprachsignal) $x(t)$ ergibt, was nun anhand von Abbildung 2.22 gezeigt wird.

In Teil a) der Abbildung ist das zeitliche Eingangssignal $e(t)$, für das entweder eine Dirac-

[14]Siehe Abschnitt 2.5.1

[15]Die erste Auflage erschien bereits 1960.

[16]Filter ermöglichen beispielsweise die Veränderung der spektralen Amplituden eines Signals.

Stoß-Folge im zeitlichen Abstand P (für Vokale) oder zufälliges Rauschen (für Frikative[17]) angenommen wird (Markel & Gray, 1976, S. 5), zu sehen. Das Frequenzspektrum der Dirac-Stoß-Folge (oder des Rauschens) fällt aufgrund seiner sehr komplexen Wellenform nicht ab (spektrales Gefälle = 0 dB/Okt), was eine wichtige Voraussetzung zur späteren Formanten-Analyse ist. Durch einen Tiefpassfilter 2. Ordnung (Abb. 2.22b) mit einer Grenzfrequenz[18] von 100 Hz, angewendet auf $e(t)$, wird die durch die Glottis modulierte Luftstromwellenform $u_G(t)$ nachempfunden (vgl. Abb. 2.4c in Abschnitt 2.3). Die Intensität des Spektrums davon fällt ab 100 Hz mit 12 dB/Oktave ab und entspricht damit dem des theoretischen primären Stimmschalls (vgl. Abb. 2.3 in Abschnitt 2.3). Durch mehrere weitere Filter 2. Ordnung (Abb. 2.22c) wird das Verhalten des Vokaltrakts nachempfunden, sodass jedem Formanten ein Resonanzmaximum in einer bestimmten Frequenz entspricht. Der spektrale Korrekturfaktor (Abb. 2.22d) hebt die Intensität der tiefen Frequenzen an und modelliert dadurch die Luftstromwellenform an den Lippen $u_L(t)$. Im letzten Filterschritt (Abb. 2.22e) wird die Wellenform des Luftstroms dann in eine akustische Druckwelle umgewandelt, was dem Sprachsignal $x(t)$ mit der Grundfrequenz $F_0 = 1/P$ entspricht (Abb. 2.22f).

Die Wirkung von Filtern wird üblicherweise mit der *z-Transformation* dargestellt. Die z-Transformation lautet nach Meyer (2000, S. 179):

$$X(z) = \sum_{n=-\infty}^{\infty} x[n] z^{-n} \qquad z = e^{(\sigma + i\omega)T} = e^{sT} \tag{2.15}$$

Sie transformiert ein zeitdiskretes Signal $x[n]$ in den z-Bereich und ist für

$$z = e^{(0+i\omega)T} = e^{i\omega T} = e^{i2\pi fT} \tag{2.16}$$

gleich der Fourier-Transformation für Abtastsignale (2.6) (Meyer, 2000, S. 181). Die Wirkung des Filters (im Zeitbereich als *Impulsantwort* $h[n]$ bezeichnet) auf das Eingangssignal $e[n]$ und das dadurch resultierende Ausgangssignal $x[n]$ wird mit der z-Transformation als einfache Division

$$H(z) = \frac{X(z)}{E(z)} \tag{2.17}$$

dargestellt. $H(z)$ ist in diesem Fall die *Übertragungsfunktion* des linearen Systems. Für das lineare Modell der Spracherzeugung gilt in diesem Fall nach Markel & Gray (1976, S. 6):

$$H(z) = G(z)V(z)L(z) = \frac{X(z)}{E(z)} \tag{2.18}$$

[17] Reibelaute durch Luftverwirbelungen, wie beispielsweise /f/ oder /s/.

[18] Ab dieser Frequenz werden die Amplituden der Frequenzanteile des Signals verändert.

Die einzelnen Filter[19] sind das Glottismodell $G(z)$, das Vokaltraktmodell $V(z)$ und das Lippenmodell $L(z)$. In $V(z)$ wird eine Anzahl von K Formanten jeweils mit Frequenzlage und Bandbreite durch Filterkoeffizienten festgelegt. Nun lässt sich zeigen (Markel & Gray, 1976, S. 8), dass als Annäherung für die verschiedenen Filter[20]

$$A(z) = \sum_{i=0}^{M} a_i z^{-i} \approx \frac{1}{G(z)V(z)L(z)} \qquad \text{mit } a_0 = 1 \text{ und } M \geq 2K+1 \tag{2.19}$$

verwendet werden kann. Das hierbei eingeführte $A(z)$ ist ein sogenannter *all-zero Filter,* während $1/A(z)$ ein *all-pole Filter* ist. a_i sind dabei konstante Filterkoeffizienten, die entsprechend passend den Filtern $G(z)$, $V(z)$ und $L(z)$ gewählt werden. $A(z)$ wird in diesem Fall auch als *inverser Filter* bezeichnet. Formel (2.18) kann dadurch als

$$X(z) = E(z)\frac{1}{A(z)} \qquad \textit{Synthese-Modell} \tag{2.20}$$

geschrieben werden, wodurch ein erzeugtes Sprachsignal $x[n]$ (z.B. ein Vokal) in z-Darstellung $X(z)$ beschrieben wird. Für die Analyse eines Signals gilt dann durch einfache Umformung:

$$E(z) = X(z)A(z) \qquad \textit{Analyse-Modell} \tag{2.21}$$

Abbildung 2.23 zeigt die Zeit-Signale[21] $x(t)$, $e(t)$ und die logarithmierten Spektren von $X(z)$, $A(z)$, $E(z)$ und $1/A(z)$ in der Form

$$LM(X) = 10\log|X(z)|^2 \qquad \text{mit } z = e^{i\theta} \tag{2.22}$$

wobei $\theta = \omega T = 2\pi f/f_s$ die normierte Frequenz ist. Die Spektren sind also die einer FTA über den positiven Frequenzbereich.

[19] Das Fehlen des spektralen Korrekturfaktors wird bei Markel & Gray (1976, S. 6) nicht weiter erwähnt. Vermutlich ist es in den anderen Komponenten enthalten.

[20] Auf die Ausformulierung von $G(z)$, $V(z)$, und $L(z)$ wird an dieser Stelle verzichtet. Sie kann bei Bedarf bei Markel & Gray (1976, S. 7 & 8) nachgeschlagen werden.

[21] Hier im Bild als quasi-kontinuierlich mit $x(t)$ durch entsprechend kleine Zeitschritte dargestellt.

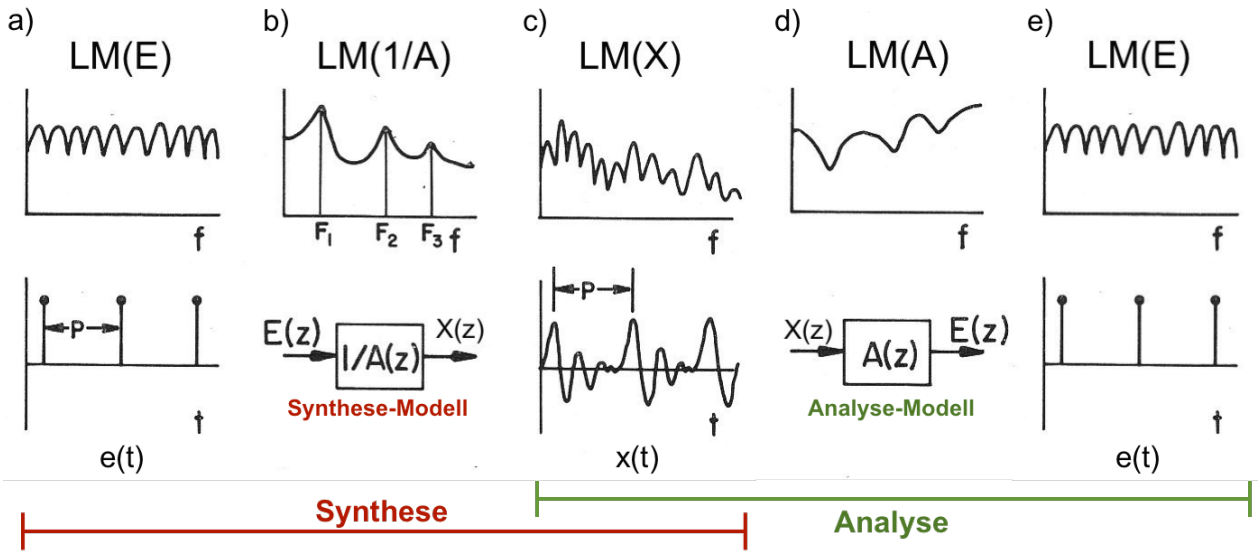


Abbildung 2.23: Oben sind jeweils die Betragsquadrate der Frequenzspektren in dB (mit LM gekennzeichnet) und unten der zeitliche Signalverlauf (hier mit t statt n) bzw. die Modellveranschaulichung für den Filterprozess in z-Darstellung zu sehen. a)-c) zeigen das Synthesemodell der Spracherzeugung: Die Dirac-Stoß-Folge $e(t)$ wird durch den Filter $1/A(z)$ (u.a. mit Koeffizienten für die Formanten $F_1, F_2, , F_3$) zu einem c) Sprachsignal $x(t)$. Es gilt: $LM(X) = LM(E) + LM(1/A)$. c)-e) zeigen das Analysemodell: c) Das zu analysierende Signal $x(t)$ ergibt nach d) Anwendung des (entsprechend gewählten) inversen Filters $A(z)$ e) das Ursprungssignal $e(t)$. Hier gilt: $LM(E) = LM(X) + LM(A) = LM(X) - LM(1/A)$ (nach Markel & Gray, 1976, S. 9).

Hiermit lässt sich das Analyse- und Synthese-Modell gut veranschaulichen: Bei der Synthese (Abb. 2.23a, b, c) eines gewünschten Sprachsignals $x(t)$ mit Spektrum $X(z)$ liegt das bekannte Ausgangssignal $e(t)$ und der Filter $1/A$ vor, bei dem die Filterkoeffizienten so angepasst werden, dass das entsprechende Formantenspektrum in $LM(1/A)$ (F_1, F_2, F_3 in Abb. 2.23b) entsteht. Bei der Analyse (Abb. 2.23c, d, e) sind das zeitliche Stimmsignal $x(t)$, das Eingangssignal des Modells $e(t)$ sowie deren Spektren $LM(X)$ und $LM(E)$ bekannt. Die Filterkoeffizienten von $A(z)$ werden nun so bestimmt, dass das Spektrum des Signals $LM(X)$ nach der Filterung das bekannte Spektrum $LM(E)$ (mit 0 dB/Okt Gefälle) ergibt. Die dadurch bestimmten Filterkoeffizienten a_i lassen dann eine Bestimmung der Formanten zu. Es fällt auf, dass $LM(1/A) = -LM(A)$ ist, wodurch die Bezeichnung *inverser Filter* deutlich wird.

2.7.1 Codierung durch lineare Prädiktion (LPC)

Prädiktoren oder auch *Vorhersagefilter* M-ter Ordnung schätzen den Signalverlauf eines abgetasteten Signals $x[n]$ von M vorher bekannten Werten aus dem Signal durch M *Prädiktorkoeffizienten* ab (Hoffmann, 1998, S. 261). Mögliche Anwendungen hierbei sind die Datenkompression von Sprachsignalen oder eben die Formantenanalyse, wozu das Signal $x[n]$ bekannt sein muss. Dabei ist das vorhergesagte Signal $\hat{x}[n]$ eine Linearkombination aus den Werten $x[n-1]$, ..., $x[n-M]$ und den Prädiktorkoeffizienten b_i

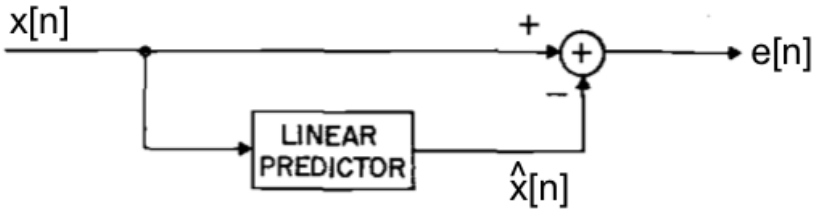


Abbildung 2.24: Modell des linearen Prädiktors: Aus dem bekannten Signal $x[n]$ wird durch den linearen Prädiktorfilter ein vorhergesagtes Signal $\hat{x}[n]$ erstellt. Die Differenz $x[n] - \hat{x}[n]$ ergibt den Prädiktionsfehler $e[n]$, welcher durch die Wahl von Prädiktorkoeffizienten minimiert wird (nach Rabiner et al., 1977).

$$\hat{x}[n] = \sum_{i=1}^{M} b_i x[n-i] \tag{2.23}$$

und man definiert den *Prädiktionsfehler* nach Markel & Gray (1976, S. 10) als

$$e[n] = x[n] - \hat{x}[n] \tag{2.24}$$

welcher durch die optimale Wahl an Prädiktorkoeffizienten minimiert werden soll. Das Schema der linearen Prädiktion ist in Abbildung 2.24 zu sehen. Da Sprachsignale nicht stationär sind (der Sänger ändert den Vokal), können hier nicht beliebig viele Werte vorausgesagt werden. Es wird nur ein Signalabschnitt der Länge N betrachtet, wofür M Prädiktorkoeffizienten mit $M < N$ verwendet werden.

Um zurück zum linearen Modell der Spracherzeugung zu kommen, wird nach Markel & Gray (1976, S. 10) das Eingangssignal $e[n]$ (welches bisher Rauschen oder eine Dirac-Stoßfolge war) nun als Prädiktionsfehler interpretiert. Das Frequenzspektrum des Fehlersignals fällt dabei flach (d.h. ohne Gefälle) aus (Markel & Gray, 1976, S. 144 ff.), weshalb das Modell der linearen Spracherzeugung zur Formantenbestimmung funktioniert. Die Prädiktorkoeffizienten werden als die negativen Filterkoeffizienten $b_i = -a_i$ des linearen Prädiktorfilters

$$F(z) = -\sum_{i=1}^{M} a_i z^{-i} \overset{(2.19)}{=} 1 - A(z) \tag{2.25}$$

festgelegt.

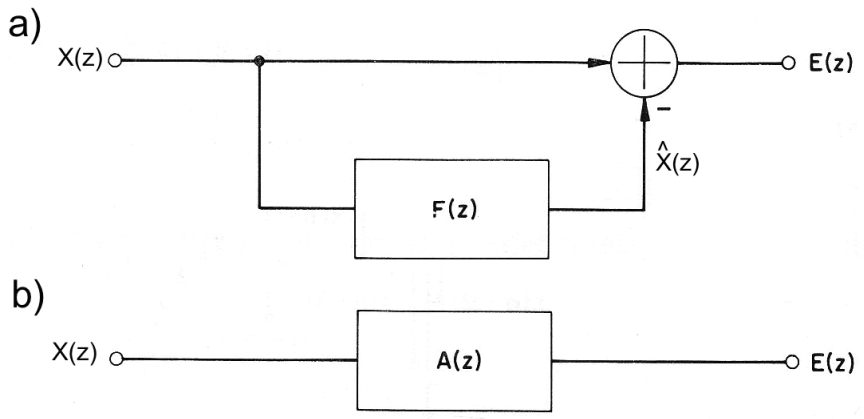


Abbildung 2.25: z-Darstellung des Modells der linearen Prädiktion, a) analog zu Abbildung 2.24 und b) als inverser Filter, analog zu Abbildung 2.23d (nach Markel & Gray, 1976, S. 12).

Das Schema der linearen Prädiktion in Abbildung 2.23 lässt sich ebenfalls in der z-Transformation darstellen, sodass mit dem Analyse-Modell (2.21) und (2.25)

$$\begin{aligned} E(z) =& X(z)[1 - F(z)] \\ E(z) =& X(z)A(z) \end{aligned} \tag{2.26}$$

gilt, was in Abbildung 2.25 dargestellt ist. Hierbei sind $X(z)$, $\hat{X}(z)$ und $E(z)$ die zugehörigen z-Transformierten von $x[n]$, $\hat{x}[n]$ und $e[n]$ sowie $A(z)$ der inverse Filter.

Unter der Bedingung, dass der Fehler $e[n]$ minimal werden soll, ergibt sich ein lineares Gleichungssystem mit M Gleichungen, das durch verschiedene Methoden gelöst werden kann (z.B. Kovarianz-Methode und Autokorrelations-Methode, Markel & Gray, 1976, S. 14 ff.), deren Erläuterung jedoch an dieser Stelle zu umfangreich wäre. Für die LPC-Analysen der vorliegenden Arbeit wird die in der Sprachanalyse-Software *Praat*[22] implementierte Methode, die bei Press et al. (2007) nachgeschlagen werden kann, verwendet.

Das LPC-Verfahren wird dadurch eingeschränkt, dass die Anzahl M der Filterkoeffizienten die Anzahl K der entdeckten Formanten bestimmt und vorher abgeschätzt werden muss. Für einen längeren Vokaltrakt werden mehr Formanten erwartet und somit mehr Koeffizienten benötigt (Markel & Gray, 1976, S. 154). Auch die Abtastrate muss berücksichtigt werden, da eine größere Abtastrate ein breiteres Frequenzspektrum bedeutet und damit mehr Koeffizienten benötigt werden, um das gesamte Spektrum zu beschreiben. Hoffmann (1998, S. 273) empfiehlt,

$$M = f_s/\text{kHz} + 4 \text{ oder } 5 \tag{2.27}$$

für Vokale zu wählen.

[22]Entwickelt von Boersma, Paul & Weenink, David, `http://www.praat.org/`

3 Vorgehensweise und Methoden

Bei der Untersuchung der Gesangsstimmenpopulationen (vgl. Abschnitt 1.3.3) sollen besonders die folgenden Punkte beantwortet werden, um einen Vergleich einzelner Stimmen innerhalb der Populationen und populationsübergreifend zu ermöglichen:

1. Existiert ein Sängerformant? Wenn ja, wo ist dieser lokalisiert und wie groß ist seine relative Intensität?
2. In welche Stimmgattung (Tenor, Alt, Sopran) lassen sich die jeweiligen Sänger einordnen?
3. Gibt es weitere charakteristische spektrale Merkmale für bestimmte Populationen (Countertenor-Falsett (Europäische Schule), Countertenor-Bruststimme (Amerikanische Schule), „Pop"-Stimme, Kastratenstimme, Sopranstimme einer Frau)?
4. Verwenden die Sänger die Technik des Formantentunings?

Zur Anwendung der verschiedenen Methoden standen für die vorliegende Arbeit keine Sänger zur Verfügung, sondern lediglich kommerziell erhältliche CDs. Für zukünftige, an die vorliegende Arbeit anknüpfende Arbeiten ist es eventuell notwendig, zusätzlich Aufnahmen von Video-Portalen wie etwa Youtube.com zu verwenden. Dadurch kann die Anzahl der zur Verfügung stehenden Lieder erheblich gesteigert werden, weshalb im Folgenden ebenfalls die mögliche Datenkompression berücksichtigt wird. In diesem Kapitel soll zunächst genauer auf die Problemstellung dieser Ausgangssituation eingegangen werden. Anschließend sollen die erarbeiteten Möglichkeiten zur Problembewältigung dargestellt werden. Entsprechend den vielfältigen Fragestellungen und den multiplen Einflussfaktoren auf die Stimme werden einige Methoden benötigt, um zu aussagekräftigen Ergebnissen zu gelangen. Sämtliche Arbeitsabläufe wurden mittels selbst geschriebener Matlab[23]-Skripte bzw. eines selbst geschriebenen Praat-Skriptes durchgeführt. Auf eine vollständige Beschreibung der Skripte durch Auflisten der einzelnen Programmzeilen wird jedoch verzichtet, da dies der Umfang der Masterarbeit nicht erlauben würde. Des Weiteren kam der Overtone Analyzer[24] (OA) zur Auffindung geeigneter Gesangspassagen zum Einsatz, was im Folgenden erläutert wird.

3.1 Problematiken der Aufnahmebedingungen

Bei der Analyse bestimmter Sängergattungen (beispielsweise Seidner et al., 1983, Boersma & Kovacic, 2006, Fuchs et al., 2000, Weiss et al. 2001, Deme, 2014) werden meist entsprechende Sänger zur Aufnahme ins Tonstudio eingeladen. Dadurch können die Rahmenbedingungen wie Mikrofontyp, Mikrofonabstand, Raumakustik und Audioformat konstant gehalten sowie weitere Messgeräte, wie z.B. ein Elektroglottograph (EGG), einbezogen werden. Ein EGG nimmt mit Strom- oder Widerstandsmessungen (nichtinvasiv) Stimmlippenbewegungen auf

[23] The MathWorks GmbH, `http://www.mathworks.com`
[24] Sygyt Software, `https://www.sygyt.com/`

(Nawka & Wirth, 2008, S. 149), mit denen das Audiosignal abgeglichen werden kann. Außerdem ist es möglich, den Schallpegel der Gesangsstimme aufzunehmen, sodass ein eindeutiger Referenzwert für die Intensität der Aufnahme festgelegt werden kann. Damit lässt sich eindeutig sagen, ob der Sänger gerade tatsächlich laut singt, oder ob nur die Aufnahme durch eine andere Abmischung des Toningenieurs lauter erscheint. Hat man mehrere Sänger bei sich im Studio, so kann man außerdem alle dasselbe Stück oder dieselben Vokale singen lassen. Erschwerend kommt bei kommerziellen CD-Aufnahmen noch die Instrumentalbegleitung hinzu, welche manche Analysemethoden (z.B. LPC) außer Kraft setzt.

Über all diese Parameter kann in der vorliegenden Arbeit nicht bestimmt werden. Daher wird versucht, unter den gegebenen Bedingungen dennoch möglichst viele Informationen zu erhalten. Auch in der Literatur (z.B. Sundberg, 2001 und Fuchs et al., 2000) wird gelegentlich auf kommerzielle Aufnahmen zurückgegriffen, daher werden diese hier ebenfalls als Anhaltspunkt für die im Folgenden vorgestellten Methoden dienen.

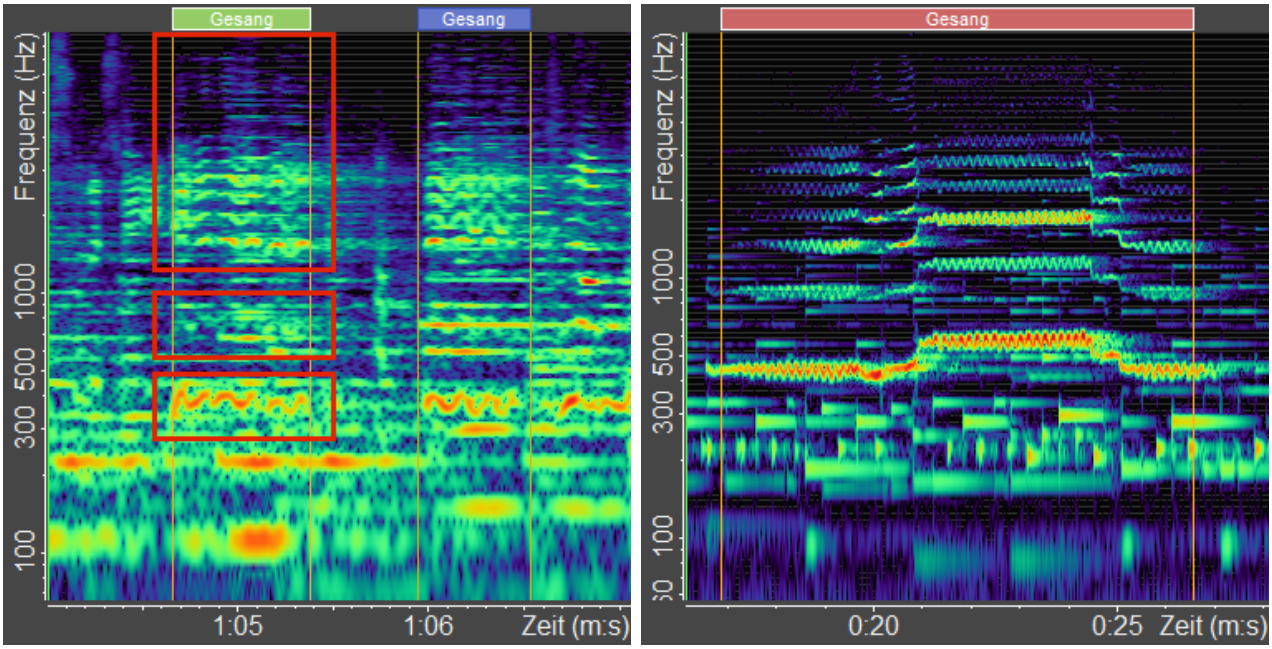
3.2 Auswahlkriterien der Sänger und Musikstücke

Es eignet sich bei weitem nicht jede Aufnahme eines Sängers zur Analyse. Von vornherein können nur Vokale mit einer Dauer von 0.01 s und länger untersucht werden, da die festgelegte Frequenzauflösung von der Länge des Analysefensters abhängt. Die Einbeziehung solch kurzer Ausschnitte macht aber auch nur dann Sinn, wenn der Sänger bereits etwa 0.1 s vorher angefangen hat den Ton zu singen, da sonst möglicherweise Einflüsse durch Einschwingvorgänge hinzukommen.

In vielen Liedern sind die Instrumente so laut, dass im Spektrogramm die Harmonischen des Sängers teilweise oder fast vollkommen in denen des Orchesters untergehen (vgl. Abb. 3.1a). Deshalb müssen zunächst für den jeweiligen Sänger alle vorhandenen Aufnahmen nach analysierbaren Passagen untersucht werden. Im Optimalfall finden sich dafür Passagen mit reinem Solo-Gesang, was aber selten vorkommt. Daher werden auch Passagen, in denen die Harmonischen des Sängers noch deutlich von der Instrumentalbegleitung unterscheidbar sind, verwendet. Abbildung 3.1b zeigt hierzu eine nahezu optimale Stelle, in der der Grundton und alle Obertöne des Sängers deutlich sichtbar sind. Die Idee dabei ist, dass diese deutlichen Obertöne noch verwendbare Informationen über die Stimme enthalten, denn das menschliche Gehör kann auch bei Instrumentalbegleitung noch Sänger A von Sänger B unterscheiden sowie verschiedene klangliche Merkmale an der Stimme feststellen. Diese einzelnen Harmonischen sollen daher aus dem Klangspektrum des Orchesters extrahiert werden, was im folgenden Kapitel weiter erklärt wird.

Für eine möglichst eindeutige Auswahl der Vokale werden die Wörter der Liedtexte (welche nicht immer in der selben Sprache sind) mit dem Online-Wörterbuch von PONS[25] und dessen Übersetzung in das Internationale Phonetische Alphabet (IPA) auf deren Aussprache überprüft. Besonders wird auf den vorderen Vokal /i/ und die hinteren Vokale /o/ und /u/ geachtet, da diese Extreme in der Artikulation darstellen. Teilweise sind jedoch andere Vokale

[25]PONS GmbH, Stuttgart, `http://de.pons.com`



(a) **(b)**

Abbildung 3.1: a) Nicht zur Analyse geeignete Passage, da die meisten Obertöne des Gesangs (rote Markierung) durch die der Instrumente zu stark überlagert sind und b) gut analysierbare Passage mit deutlich erkennbaren Obertönen des Sängers.

wie /a/ oder /e/ häufiger vorhanden und werden daher auch verwendet. Es sei aber darauf hingewiesen, dass es nahezu unmöglich ist, exakt diese Vokale in kommerziellen Aufnahmen zu finden, da bereits personen- und sprachbedingt die Vokale unterschiedlich ausfallen können (vgl. Abb. 2.5). Durch diese Auswahlkriterien kommt es durchaus dazu, dass manche Sänger, besonders wenn die Auswahl an Liedern nicht groß genug ist, überhaupt nicht untersucht werden können.

3.3 Berechnete Spektralamplituden und Normalisierung der Lautstärke

Da zu den Aufnahmen keinerlei Referenzwerte für die tatsächliche Lautstärke vorliegen, muss überlegt werden, wie sich dennoch der Vergleich mehrerer Aufnahmen ermöglichen lässt. Ein einfaches Normalisieren aller Lieder durch Einstellen der Maximalamplitude auf einen festen Wert sorgt besonders dann für falsche Lautstärkenverhältnisse, wenn an einer Stelle im Stück das Orchester in voller Besetzung spielend den Lautstärkepegel enorm anhebt. Deshalb wird das Betragsquadrat der stärksten extrahierten Harmonischen aus dem jeweiligen Lied als Marke für 0 dB gelten. Bei Album-internen Aufnahmen wird ggf. die stärkste Harmonische des gesamten Albums verwendet, da die Stücke in der Regel in ihrer Lautstärke aufeinander abgestimmt sind. Zusätzlich wird bei der Liedauswahl darauf geachtet, ob es sich tendenziell eher um eine ruhige Ballade oder eine lautstark dramatisierte Oper handelt. So kann dies im Vergleich solch unterschiedlicher Lieder bewusst berücksichtigt werden.

Obwohl mit Matlab die Betragsquadrate der spektralen Amplituden berechnet werden, wird im Text und in den Abbildungen dennoch meist von „Intensität“ die Rede sein. Das ist legitim, da das Betragsquadrat der Spektralamplitude, die Energie, die Leistung und die Intensität in den hier vorgenommenen Berechnungen proportional zueinander sind. Denn jeder Zeitpunkt im Spektrogramm wird in einheitlichen Zeitabständen zum nächsten berechnet. Auch fallen die nicht bekannten Einheiten der Audiodateien durch den entsprechenden Referenzwert in der Dezibelskala weg. Somit ist beispielsweise der Oberton mit der höchsten „Energie“ auch der mit der höchsten „Intensität“.

Im Gegensatz zu den in Matlab berechneten Spektralwerten normalisiert der Overtone Analyzer (OA) die dargestellten Intensitäten so, dass ein Sinussignal von 1000 Hz bei einer Amplitude von 1 (Einheiten sind bei Audiodaten nicht vorhanden) genau 0 dB entspricht. Dies sei für die Betrachtung der mit dem OA erstellten Bilder der vorliegenden Arbeit im Hinterkopf zu behalten.

3.4 Harmonischenextraktion (HE)

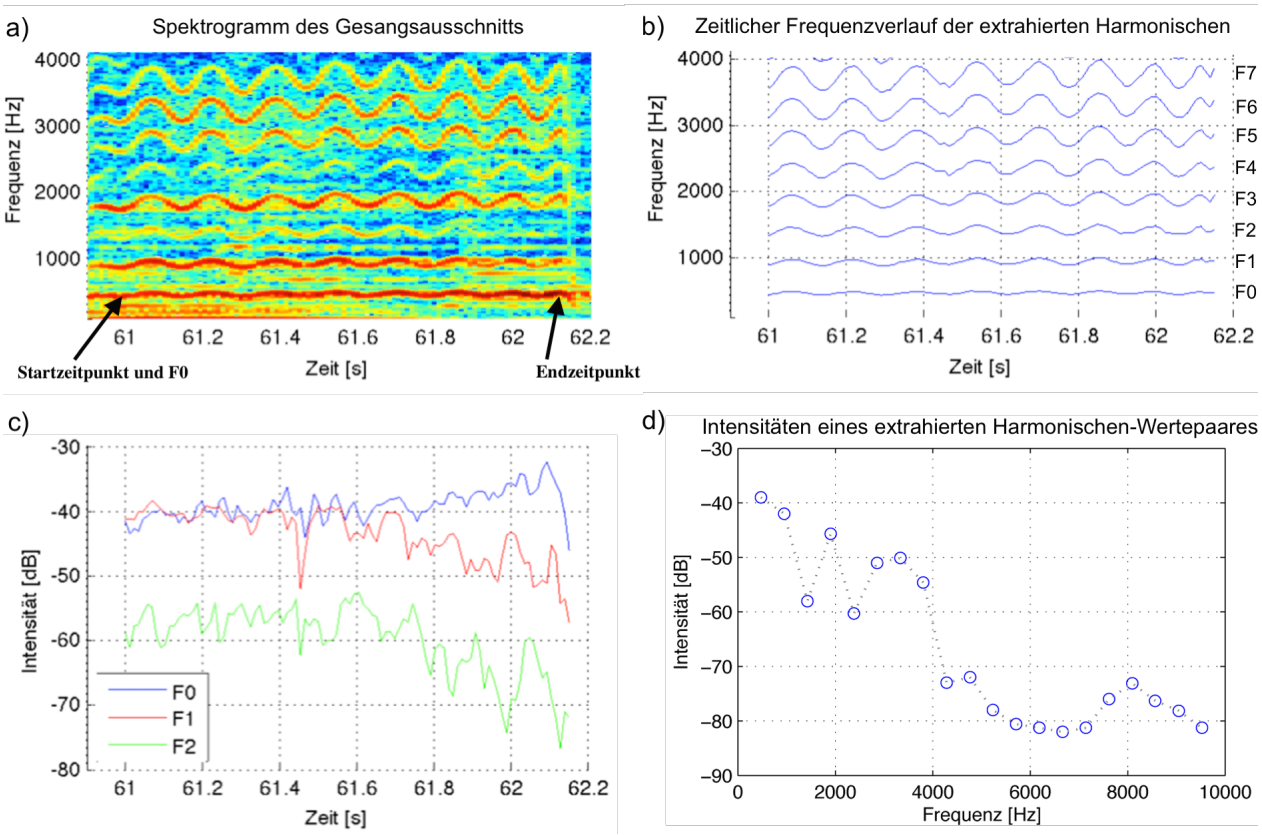


Abbildung 3.2: a) Spektrogrammausschnitt eines Liedes, aus dem die Harmonischen von $t_0 = 61$ s bis $t_e \approx 62,18$ s extrahiert werden. F_0 liegt bei ca. 480 Hz. b) Frequenzverlauf aller Harmonischen nach der Extraktion. c) Intensitätsverlauf der ersten drei Harmonischen über die Zeit. d) Intensitäten eines beliebig herausgegriffenen Harmonischen-Wertepaares.

Da die Anzahl an verwendbaren Gesangspassagen ohne Instrumentalbegleitung relativ gering und nicht für alle Sänger im gleichen Maße ausfällt, wurde mit Matlab ein Skript entwickelt, welches aus dem Spektrogrammausschnitt einer Audiodatei (vgl. Abb. 3.2a) die ersten 20 Harmonischen des Sängers mit den Informationen zu Frequenz und Intensität extrahiert (Abb. 3.2b). Im weiteren Verlauf wird dieser Prozess *Harmonischenextraktion* (*HE*) genannt. Das Skript erkennt jedoch nicht von selbst die Obertöne der Stimme. Vielmehr muss zunächst manuell für jede einzelne verwendbare Stelle eines Liedes der Start- und Endzeitpunkt und die ungefähre Tonhöhe $\tilde{F}_0$ zum Startzeitpunkt t_0 angegeben werden. Damit sucht das Skript bei jeder Markierung zunächst zum Startzeitpunkt den größten Spektralintensitätswert $I_0(t_0)$ nahe der angegebenen ungefähren Tonhöhe $\tilde{F}_0(t_0)$ (Abb. 3.2a). Mit diesem Maximalwert $I_0(t_0)$ wird dann die tatsächliche Grundfrequenz $F(t_0)$ festgelegt und abgespeichert. Anschließend werden die maximalen Intensitäten $I_N(t_0)$ in der Nähe von $\tilde{F}_1(t_0) = 2F_0(t_0)$, $\tilde{F}_2(t_0) = 3F_0(t_0)$, ... , $\tilde{F}_{19}(t_0) = 20F_0(t_0)$ gesucht, sodass auch hier wieder die tatsächlichen Frequenzen der maximalen Intensitäten $F_N(t_0)$ für $N = 2...20$ abgespeichert werden können. Dann passiert dasselbe für den darauffolgenden Zeitschritt $t_0 + \Delta t$ im Spektrogramm (ein Zeitschritt entspricht einem Bildpunkt in x-Richtung), so lange, bis der Endzeitpunkt t_e (Abb. 3.2a) erreicht ist. Es folgt genau dieser Prozess für alle angegebenen Markierungen im gesamten Lied.

Abbildung 3.2b zeigt den Frequenzverlauf der extrahierten Harmonischen, Abbildung 3.2c die Intensitätsverläufe der Grundschwingung I_{F_0} und der ersten beiden Obertöne I_{F_1} und I_{F_2} desselben Ausschnitts. In Abbildung 3.2d sind die Intensitäten aller Harmonischen zu einem beliebig herausgegriffenen Zeitpunkt aus der markierten Passage zu sehen.

Zusätzlich muss markiert werden, um welchen Vokal es sich handelt, und ob es sich um eine reine Gesangspassage, eine reine Instrumentalpassage oder eine Mischung handelt. Diese Markierungen werden entsprechend den Auswahlkriterien manuell mit dem Overtone Analyzer erstellt, wovon ein Screenshot in Abbildung 3.3 zu sehen ist:

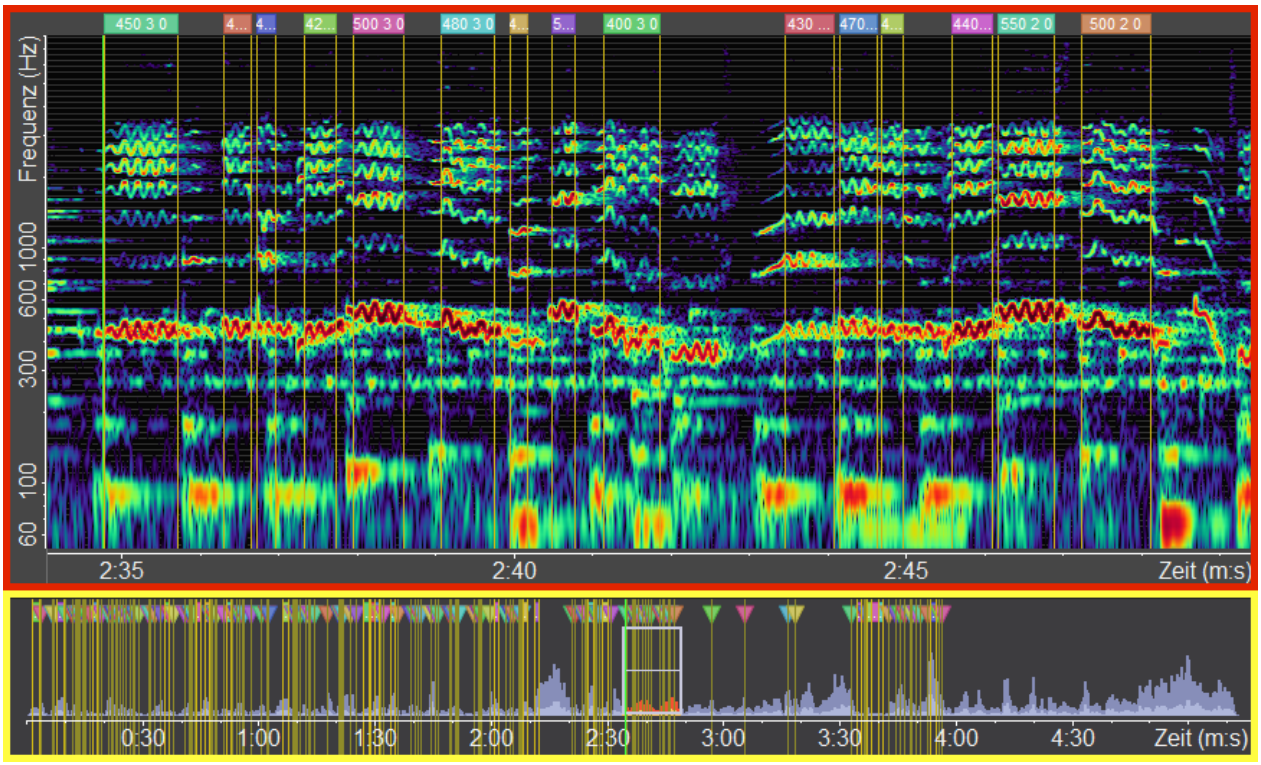


Abbildung 3.3: Screenshot des Overtone Analyzers. Rot umrahmt: Spektrogrammansicht mit markierten Passagen. Die letzte (braune) Markierung „500 2 0“ bedeutet beispielsweise, dass die grobe Grundfrequenz $\tilde{F} = 500$ Hz ist, es sich um den Vokal /e/ („2“) und um keine Solopassage („0“) handelt. Gelb umrahmt: Gesamtübersicht des gesamten Liedes. Jeder senkrechte gelbe Strich ist dabei ein markierter Bereich.

Im roten Bereich ist das Spektrogramm eines beispielhaften Liedausschnittes mit 15 Markierungen zu sehen. Der gelbe Rahmen zeigt den Intensitätsverlauf des gesamten Liedes und alle 164 von Hand ausgewählten und beschrifteten Markierungen in Form von senkrechten gelben Strichen. Dabei ist ein Lied mit einer solch großen Anzahl von 164 verwendbaren Passagen sehr selten. Vom Entwickler des OA, Bodo Maass, wurde mir eine speziell erweiterte Programmversion des OA mit einer Exportfunktion zur Verfügung gestellt, sodass die Informationen in den Markierungen über Zeitpunkte, ungefähre Tonhöhe, Vokale und Art der Passage in Matlab als CSV-Daten eingelesen werden können.

Der Vorteil dieser Methode gegenüber der Verwendung aller Frequenzkomponenten ist, dass der Instrumentaleinfluss reduziert wird und sich daraus verschiedene weitere Methoden realisieren lassen, die in den folgenden Abschnitten erläutert werden. Außerdem werden Einflüsse von Hall und Echo reduziert, da immer nur genau die 20 stärksten Harmonischen extrahiert werden.

3.4.1 Beschreibung des Stimmklangs durch die spektrale Steigung

Um die Möglichkeit zu erhalten, einen Sänger in verschiedenen Tonhöhen und bei unterschiedlichen Intensitäten klanglich zu charakterisieren und dies gleichzeitig übersichtlich dar-

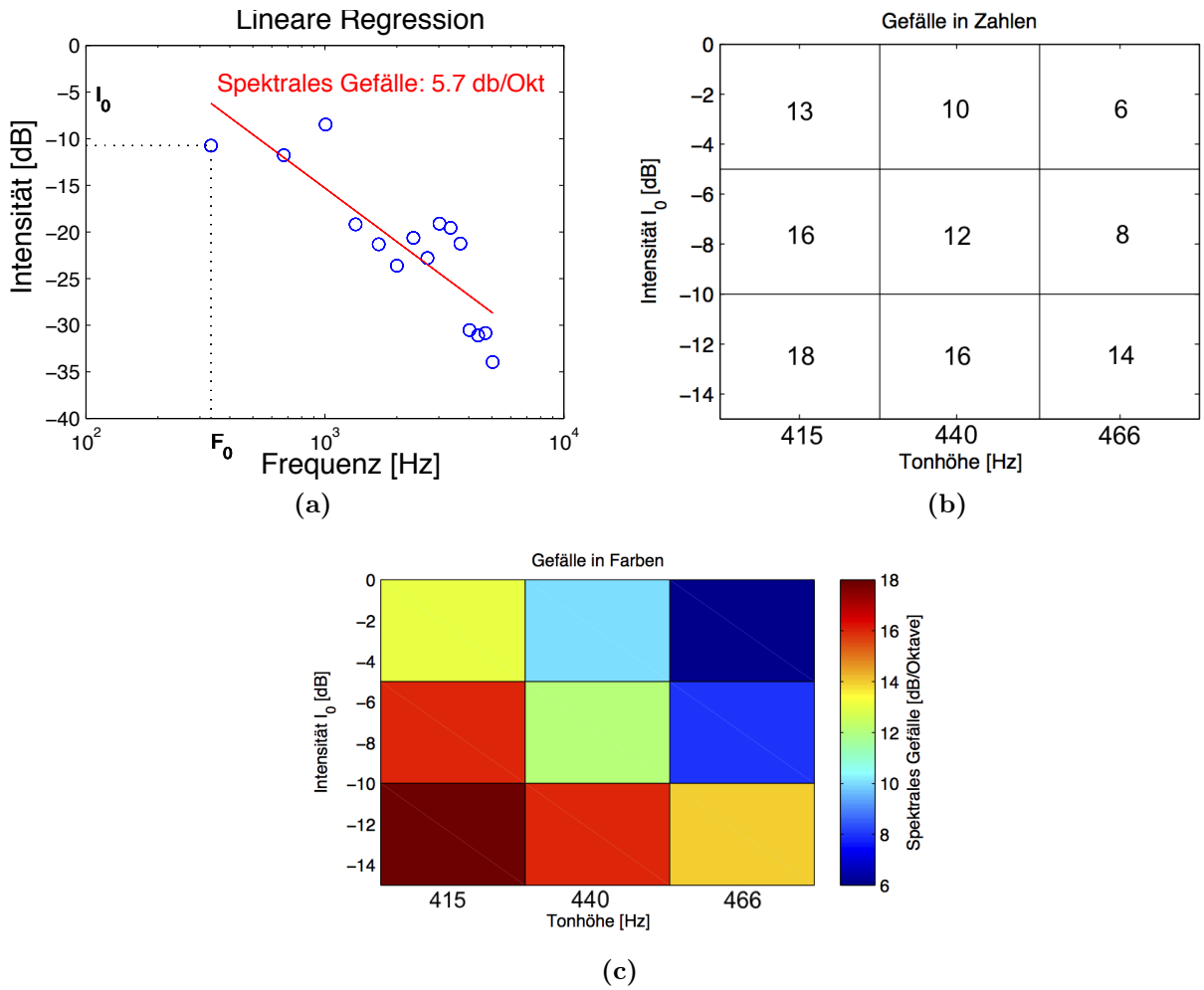


Abbildung 3.4: a) Lineare Regression an einem einzelnen Harmonischen-Wertepaar mit $F_0 \approx 333.8$ Hz und $I_0 \approx -10.7$ dB und Gefälle $S = 5.7$ dB/Okt. b) Beispielhafte (kleine) Matrix an gemittelten Werten für das spektrale Gefälle (in dB/Okt) für drei Noten und drei Intensitätsabstufungen. c) Die Werte aus b) nun als Farben dargestellt.

zustellen, wird für die spektrale Steigung (vgl. Abschnitt 2.3) eine Darstellungsform ähnlich der eines Spektrogramms gewählt (x = Tonhöhe, y = Intensität, Farbe = Steigung). Dabei soll ausdrücklich darauf hingewiesen werden, dass sich der theoretische Wert der spektralen Gefälle (6 dB/Okt = „flötenartig“, 18 dB/Okt = „metallisch“) auf den primären und nicht auf den komplexen Stimmschall bezieht. Da das Gefälle aber einen wichtigen Punkt bei der Entstehung des komplexen Stimmschalls darstellt, ist es dennoch sinnvoll diesen Wert zu berücksichtigen.

Jedes einzelne extrahierte Harmonischen-Wertepaar wird mit Matlab in doppelt-logarithmischer Auftragung (x & y) einer linearen Regression unterzogen (da in dieser Darstellung ein Gefälle von beispielsweise 6 dB/Okt eine einfache Gerade darstellt) und die Gefälle werden S in dB/Okt berechnet (Abb. 3.4a). Im darauffolgenden Schritt sollen diese

Gefälle zum einen nach der zugehörigen Grundfrequenz F_0 und zum anderen nach der Intensität I_0 der Grundschwingung geordnet werden. Die Frequenzen sollen hierbei den Tonhöhen der Noten in der gleichstufigen Stimmung mit dem Kammerton von $A_4 = 440$ Hz (Schreibweise der Noten nach Young, 1939) entsprechen. Diese Noten liegen auf den Frequenzen

$$f_i = 440 \cdot 2^{i/12} \text{ Hz} \tag{3.1}$$

mit $i =..., -1, 0, 1, ...$ (Wolfe, undatiert). Die Frequenz-Grenzen der Einteilung jeder Note erhält man für $i =..., -1.5, -0.5, 0.5, 1.5, ...$. Die Intensitäten werden in 5-dB-Schritten gruppiert (0 bis −5 dB, −5 bis −10 dB, ...). Für das in Abbildung 3.4a dargestellte Harmonischenpaar erhält man somit die Werte: $F_0 \approx 333.8$ Hz, $I_0 \approx -10.7$ dB, $S = 5.7$ dB/Okt, die entsprechend der Note E4 (329 Hz) und der zweiten Intensitätsabstufung (−5 bis −10 dB) einzuordnen sind. Für genügend analysierte Stellen erhält man somit viele Tonhöhen und Intensitäten, zu denen sich jeweils ein Steigungswert berechnen lässt. Die Steigungen können dann für jede Gruppierung gemittelt werden.

Abbildung 3.4b zeigt eine Matrixeinteilung (x = Tonhöhe, y = Intensität) für drei Tonhöhen- und drei Intensitätsgruppen mit erfundenen Spektralgefällen. Für viele Tonhöhen und viele Intensitäten wird diese Matrix schnell sehr groß und unübersichtlich, daher ist eine farbliche Darstellung in diesem Fall wesentlich besser, was in Abbildung 3.4c gezeigt ist. Starke spektrale Gefälle sind rot und schwache blau dargestellt.

Für die Regression werden lediglich Werte bis 5000 Hz verwendet, da etwa ab hier bei verschiedenen Audioformaten das spektrale Gefälle, durch die Datenkompression bedingt, unterschiedlich ausfällt. Abbildung 3.5 zeigt einen Liedausschnitt einmal in unkomprimierter CD-Qualität, und einmal in komprimierter Form von Youtube. Bis etwa 5000 Hz bleiben die Spektren noch vergleichbar.

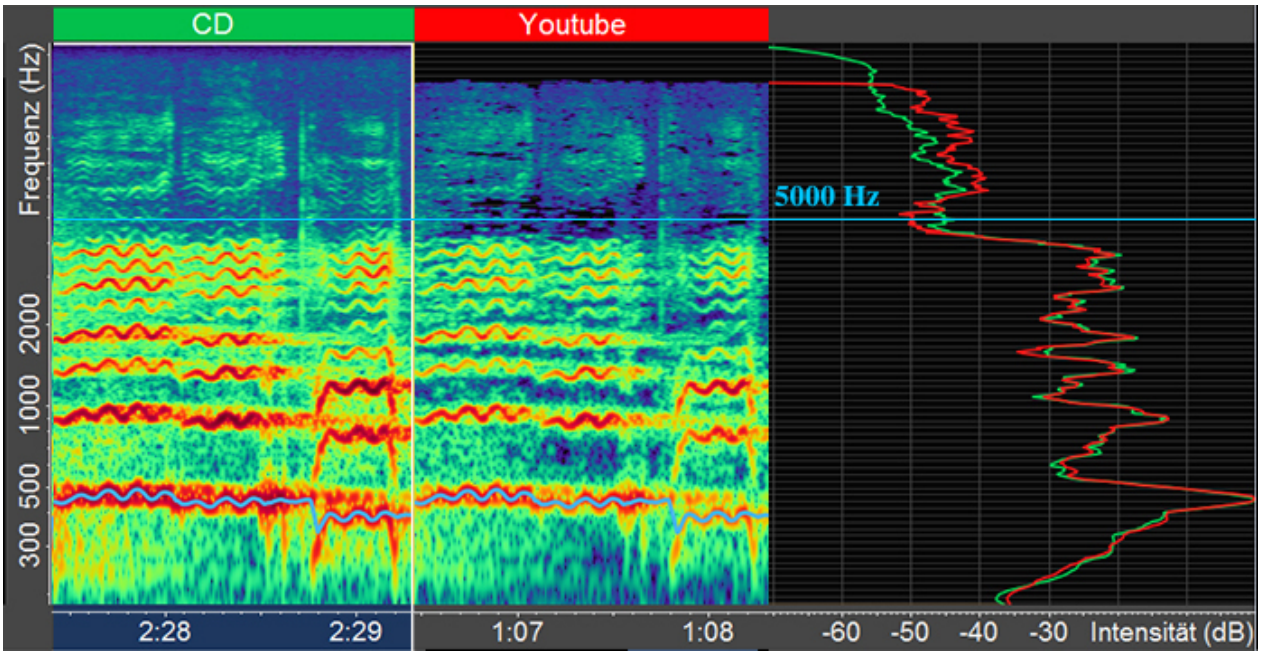


Abbildung 3.5: Spektrogramm (links) und LTAS (rechts) mit unkomprimierter CD-Qualität (grün) und dem selben Lied auf Youtube (rot). Bis etwa 5000 Hz sind die Spektren noch nahezu identisch, danach treten deutliche Unterschiede auf.

3.5 Analysemethoden für den Sängerformant

3.5.1 LTAS und tonhöhenkorrigiertes LTAS

Aus dem Langzeitspektrum (LTAS, vgl. Abschnitt 2.6.5) lassen sich bei der Mittelung möglichst vieler Gesangsausschnitte (mit und ohne Instrumentalbegleitung) Informationen über den Sängerformant finden. Möglicherweise gibt es hier je nach Gesangstechnik markante Unterschiede. Die Frequenzauflösung wird auf 125 Hz eingestellt. Zusätzlich wird in der Auswertung eine veränderte Form des LTAS zum Einsatz kommen:

Bei Boersma & Kovacic (2006) wird die Methode des sogenannten *tonhöhenkorrigierten LTAS* beschrieben. Das Langzeitspektrum hängt hier weniger von der gesungenen Tonhöhe ab (z.B. stark ausgeprägte Maxima durch einzelne Harmonische) als das beim normalen LTAS der Fall ist und lässt dadurch eine bessere Aussage über die Formanten und das Spektrum des Primärschalls zu. Die Methode aus der genannten Quelle wurde mit Hilfe eines Matlab-Skripts umgesetzt und zur Anwendung mit der HE leicht modifiziert. Mit Abbildung 3.6 lässt sich die Berechnung anschaulich erklären:

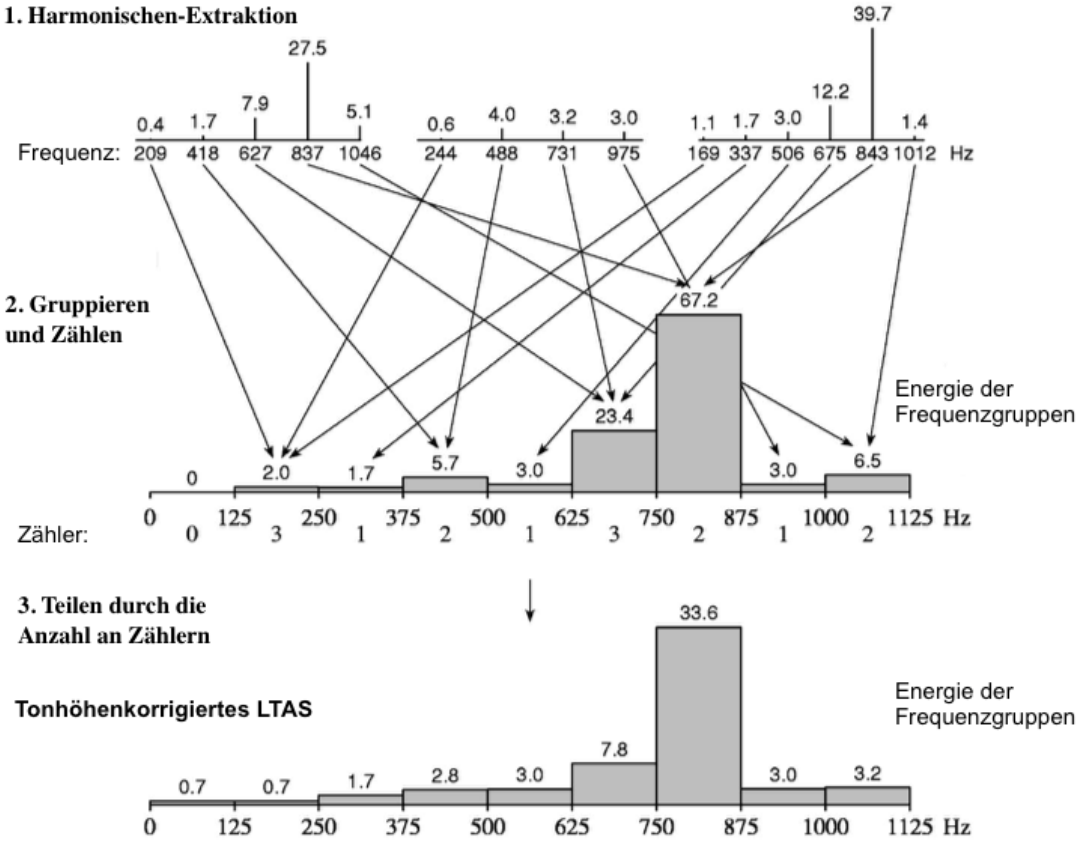


Abbildung 3.6: Schematischer Ablauf zur Erstellung des tonhöhenkorrigierten LTAS. 1. Von allen extrahierten Harmonischenpaaren sind Frequenz und Energie bekannt. 2. Alle Harmonischen werden nach ihrer Frequenz gruppiert, die Energien der Gruppen werden addiert. 3. Die Energien jeder Gruppe werden durch die Anzahl an Harmonischen in der Gruppe geteilt. Leere Gruppen werden linear interpoliert (nach Boersma & Kovacic, 2006).

1. Im ersten Schritt werden die Harmonischen mittels HE extrahiert. Jede Harmonische ist durch einen Energiewert[26] und eine Frequenz festgelegt.

2. Im zweiten Schritt werden 125 Hz breite Frequenzgruppierungen (1. Gruppe: 0 - 125 Hz, 2. Gruppe 125 - 250 Hz, usw.) gebildet. Jede Harmonische wird anhand ihrer Frequenz in eine dieser Gruppen eingeordnet, wobei die Energien addiert werden.

3. Schritt 3 besteht darin, die Summe der Energien jeder Gruppe durch die Anzahl der Harmonischen zu teilen, die in diese Gruppe gefallen sind. Dadurch werden die Energien in den Frequenzgruppen, in die viele Harmonische fallen, im Verhältnis abgeschwächt.

[26]Bei Boersma & Kovacic (2006) wird die „Energie“ verwendet, was in den Berechnungen der vorliegenden Arbeit dem Betragsquadrat der Spektralamplitude dividiert durch das Betragsquadrat der stärksten extrahierten Spektralamplitude entspricht.

Im normalen LTAS hingegen wird jede Frequenzgruppe durch die gleiche Zeit geteilt und dadurch nur skaliert, wodurch stärkere Abhängigkeiten von der Tonhöhe entstehen. Sollte eine Gruppe keine Harmonische enthalten, so werden die Energiewerte linear interpoliert. Wären diese Werte ..., 3, 4, **0**, 6, 2, ... dann würden sie durch die Interpolation also zu ..., 3, 4, **5**, 6, 2, ... ergänzt.

In der Originalmethode nach Boersma & Kovacic (2006) werden am Schluss noch Korrekturfaktoren hinzumultipliziert. Dies ist für die vorliegende Arbeit jedoch irrelevant, da die tatsächlichen Energien nicht bekannt sind (vgl. Kapitel 3.3) und das tonhöhenkorrigierte Langzeitspektrum mit seinem stärksten Energiewert auf 1 (0 dB) normalisiert wird. Außerdem sind noch zwei weitere Unterschiede zur Originalmethode nach Boersma & Kovacic (2006) zu nennen: Zum einen erhalten im Original leere Frequenzgruppen an den äußeren Frequenzenden denselben Energiewert wie die nächste noch gefüllte Gruppe (in Abbildung 3.6 ist das im dritten Schritt für die Frequenzgruppe von 0 - 125 Hz geschehen). Zum anderen werden bei Boersma & Kovacic (2006) ausschließlich reine Gesangspassagen von Sängerprobanden verwendet.

Die Vorteile des tonhöhenkorrigierten LTAS sind also eine geringere Tonhöhenabhängigkeit und möglicherweise ein geringerer Einfluss durch die Instrumente. In Abschnitt 4.1 wird Letzterer nochmals untersucht.

Im Folgenden wird, mit dem Ziel der eindeutigen Begriffstrennung, das tonhöhenkorrigierte langzeitgemittelte Spektrum als LTAS-T bezeichnet. Sowohl das LTAS, als auch das LTAS-T werden mit dem Matlab-Skript für alle Gesangspassagen berechnet, die für die HE markiert wurden (vgl. Abschnitt 3.4).

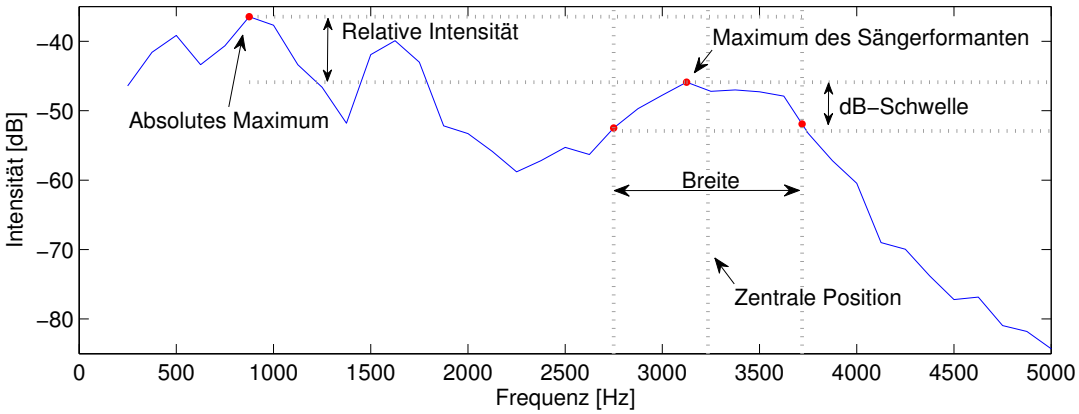


Abbildung 3.7: Bestimmung der Sängerformantenposition aus einem LTAS bzw. LTAS-T. Ausgehend von der Frequenz der maximalen Intensität des Sängerformanten zwischen 2000 und 5000 Hz werden die zwei nächsten Frequenzen (höher und niedriger) bestimmt, bei denen der Intensitätswert um einen bestimmten dB-Schwellenwert abgesunken ist. Die zentrale Position des Sängerformanten liegt dann zwischen den äußeren Punkten. Zudem wird die relative Intensität des Sängerformanten aus der Maximalintensität des Sängerformanten und dem absoluten Maximum bestimmt.

3.5.2 Bestimmung der Sängerformanten-Position

In der Literatur (z.B. Sundberg, 2001) wird die Position des Sängerformanten gewöhnlich aus einem LTAS oder LTAS-T gewonnen. Dabei wird der Maximalwert des Spektrums im Bereich von 2000 - 5000 Hz bestimmt (vgl. Abb. 3.7). Von diesem Maximum aus wird jeweils in die positive und die negative Frequenzrichtung der nächste Wert gesucht, der um einen bestimmten dB-Schwellenwert niedriger ist als das Maximum. Die Frequenz genau zwischen diesen äußeren Punkten ist dann die zentrale Position des Sängerformanten. Der Schwellenwert wird in der Literatur zwischen 3 dB (Sundberg, 2001), 10 dB (Kovacic et al., 2003) und 15 dB (Seidner, 1983) angesetzt, was davon abhängt, ob man Sänger im Studio oder kommerzielle Aufnahmen untersucht. Besonders bei Instrumentaleinfluss sticht der Sängerformant im Langzeitspektrum weniger hervor, sodass 15 dB als Schwellenwert nicht verwendbar sind. Für diese Arbeit wird ein geringer Wert von 3 dB gewählt, da bei Sundberg (2001) ebenfalls kommerzielle Aufnahmen untersucht wurden. Genau zwischen den damit ermittelten Grenzwerten befindet sich das Zentrum des Sängerformanten, das sich mit der Stimmgattung zu verschieben scheint (Sundberg, 2001 und Seidner, 1983). Um ein Maß für die Stärke des Sängerformanten zu erhalten, wird zudem das absolute Intensitätsmaximum des Spektrums bestimmt. Bildet man die Differenz aus dem absoluten Maximum und dem Maximum des Sängerformanten (in der Dezibel-Skala), erhält man die relative Intensität des Sängerformanten.

3.5.3 Singing Power Ratio (SPR)

Omori et al. beschrieben 1996 eine einfache Methode, um objektiv die Qualität einer Stimme mit der sogenannten *Singing Power Ratio* (*SPR*) zu bestimmen. Hierfür wird aus einem Kurzzeitspektrum jeweils der stärkste Oberton in den Bereichen 0 - 2000 Hz und 2000 - 4000

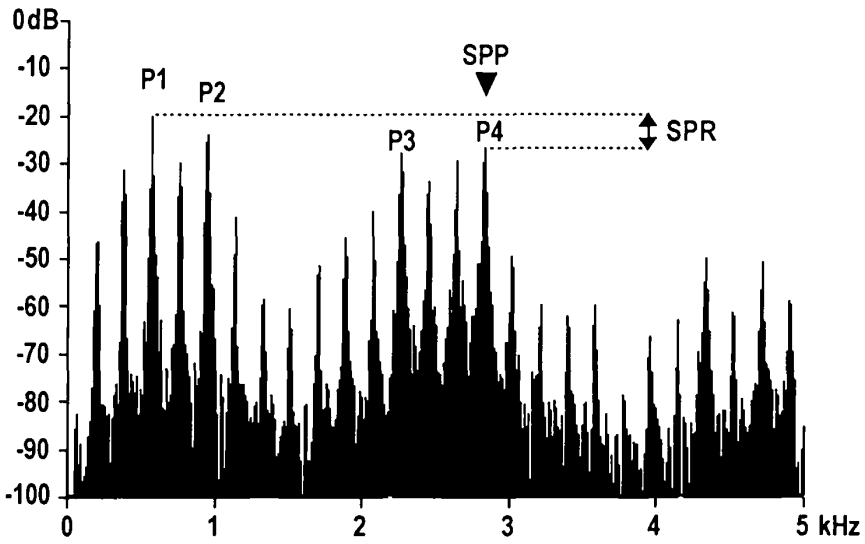


Abbildung 3.8: Zur Bestimmung der SPR: Der stärkste Oberton im Bereich von 2000 - 4000 Hz (hier P1) und zwischen 2000 - 4000 Hz (hier P4, oder auch mit SPP bezeichnet) wird bestimmt. In der Dezibelskala entspricht die SPR genau der Differenz der Intensitäten (aus Omori et al., 1996).

Hz bestimmt (im Fall der vorliegenden Arbeit mittels eines Matlab-Skripts für jedes einzelne extrahierte Harmonischen-Wertepaar) und daraus ein Intensitätsverhältnis gebildet, was in der Dezibelskala einer einfachen Differenz entspricht (vgl. Abb. 3.8). Das Maximum zwischen 2000 und 4000 Hz wird als *Singing Power Peak* (*SPP*) bezeichnet. Dieser sollte erwartungsgemäß nahe dem Sängerformanten liegen. Dabei wurde bei Omori et al. (1996) ein Vokal festgelegt und vom Probanden „at a comfortable pitch and intensity“ (Omori et al., 1996, S.229) gesungen. Es wird also keinerlei Angabe zur individuellen Lautstärke oder Tonhöhe gemacht. Als Modifizierung soll in dieser Arbeit die SPR zusätzlich in Abhängigkeit von der Tonhöhe dargestellt werden. Daher wird jeder SPR-Wert, der aus einem Harmonischen-Wertepaar hervorgeht, der Tonhöhe F_0 des ersten Grundtons zugewiesen. Die Frequenzabstände sind ebenfalls wie in Abschnitt 3.4.1 nach den Notenwerten der gleichstufigen Stimmung (Kammerton 440 Hz) festgelegt. Gleichzeitig wird die Position des SPP gespeichert. So kann zusätzlich für jede Tonhöhe eine gemittelte Lage des SPP bestimmt werden. Möglicherweise zeigt sich eine Verbindung zur Position des Sängerformanten, die, wie in Abschnitt 3.5.2 beschrieben, aus dem LTAS(-T) bestimmt wird. Bei Omori et al. (1996) wird die SPR immer in Dezibel angegeben und auch weitere Berechnungen mit ihr (wie z.B. Mittelwert und Standardabweichung) werden ebenfalls in Dezibel durchgeführt. Um vergleichbare Ergebnisse zu erhalten, werden in der vorliegenden Arbeit ebenfalls nur die Dezibelwerte verwendet.

Allerdings muss angemerkt werden, dass die Anwendung der SPR nur in wenigen Quellen zu finden ist. Die wissenschaftliche Aussagekraft könnte daher in Frage gestellt werden. Die Methode ergänzt sich jedoch sehr gut mit der HE-Methode und lässt möglicherweise Aussagen über die Sängerformantenintensität und auch -position zu, da so für jeden Vokal und für jeden der vielen Zeitschritte mit dem in Matlab geschriebenen Skript ein SPR-Wert gebildet werden

kann.

3.6 Methode zur Untersuchung auf Formantentuning

Boersma & Kovacic (2006) beschreiben eine Möglichkeit zur Feststellung, ob Formantentuning angewendet wird. Sie basiert auf der Bestimmung der Formantenlage mittels LPC (Abschnitt 2.7.1). Mit Praat wurde nach den Angaben bei Boersma & Kovacic (2006) ein Skript geschrieben, welches die manuell ausgewählten Solopassagen eines Musikstückes im Frequenzbereich bis 5000 Hz auf die Lage der ersten fünf Formanten und die aktuelle Tonhöhe F_0 untersucht. Dabei werden alle 25 ms vier Formantenpositionen und ein F_0-Wert bestimmt. Der fünfte Formant wird vermutlich nicht verwendet, da es nicht immer fünf Formantenwerte gibt, was aber bei Boersma & Kovacic (2006) nicht weiter erläutert wird. Außerdem werden die Werte den angegebenen Vokalen entsprechend markiert, um sie später separat betrachten zu können. Ein beispielhafter Datensatz ist in Abbildung 3.9a zu sehen, wobei zusätzlich die theoretischen Positionen der Harmonischen H_1 ... H_N aufgetragen sind. Da es algorithmusbedingt falsch ermittelte F_0- und Formanten-Werte gibt, werden, wie bei Boersma & Kovacic (2006) beschrieben, nur die F_0-Werte zwischen der 5. und 95. Perzentile der Gesamt-F_0-Verteilung verwendet. Auch werden die Formantenwerte nur einbezogen, sofern alle vier Formanten dieses Messpaares innerhalb der 5. und 95. Perzentile ihrer jeweiligen Formantengesamtverteilung liegen. Dadurch schränkt sich der Datensatz weiter ein (Abb. 3.9b). Abschließend werden alle Werte in Tonhöhenschritten von 1 Hz gruppiert, wodurch ein Mittelwert (farbige Linie) für jeden der vier Formanten in jeder Tonhöhe und die Standardabweichung (senkrechte schwarze Linien) gebildet werden kann (Abb. 3.9c). Sofern die Formantenwerte korrekt sind, kann Formantentuning daran erkannt werden, dass die Werte genau auf den eingezeichneten Linien der Harmonischen liegen. Die Methode funktioniert jedoch ausschließlich bei reinen Solopassagen, da das LPC-Verfahren von reinen Sprachsignalen ausgeht und somit jegliche Instrumentalisierung die Formantenlage verfälscht.

3.6.1 Wahl der Einstellparameter für die FFT

Verschiedene Parameter wie Fenster, Fensterlänge und Überlappung wurden variiert und mittels Spektrogrammansicht verglichen. Dabei sollen die Obertöne besonders in den Frequenzen ab 200 Hz ausreichend scharf sein und viele Zeitschritte erreicht werden. Besonders die von Harris (1978) und Nuttall (1981) empfohlenen Fenster wurden betrachtet. Die Entscheidung fiel bei einer Abtastrate von 44100 Hz auf das in Matlab implementierte Blackman-Nuttall-Fenster[27] mit einer Länge von 4096 Punkten, welche durch Nullpolsterung auf 8192 Punkte erweitert wird, und einer Überlappung von 90%. Das Nuttall-Fenster hat eine relativ breite Hauptkeule, aber die erste Nebenkeule liegt bei etwa -93.6 dB, wodurch die schwächeren Intensitäten in den höheren Frequenzbereichen noch deutlicher erkennbar bleiben. Durch die hohe Überlappung können viele Harmonischen-Wertepaare in den kurzen Markierungen extrahiert werden. Eine weitere Erhöhung der Fensterlänge führt besonders beim Vibrato dazu,

[27] Genaue Beschreibung in der Matlab-Dokumentation : `http://de.mathworks.com/help/signal/ref/nuttallwin.html`, zuletzt aufgerufen am: 02.04.2015.

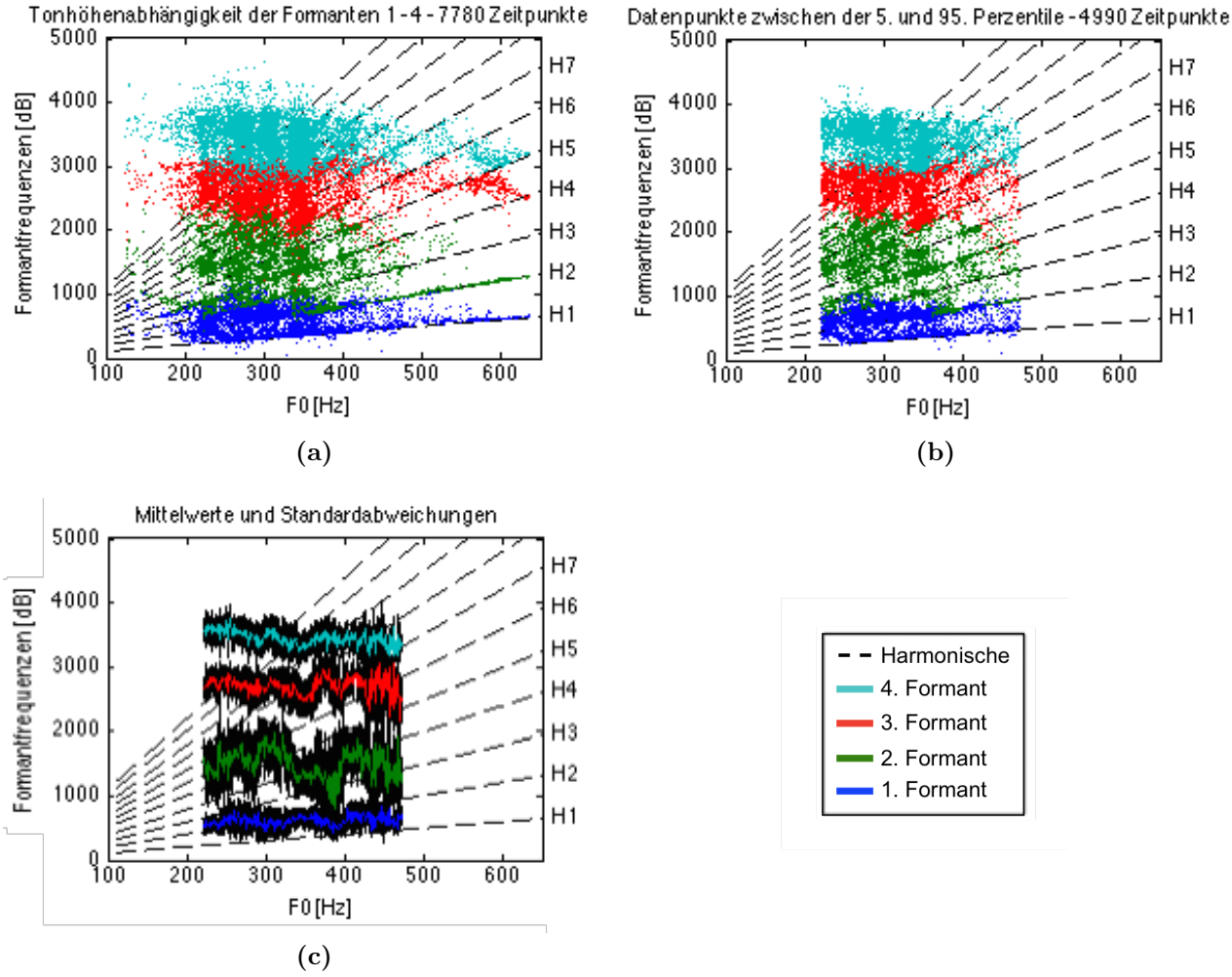


Abbildung 3.9: a) Aus 7780 Zeitpunkten erfasste Lagen der Formanten 1 bis 4 für unterschiedliche Tonhöhen F_0 und die Lage der Harmonischen. b) Von den Werten aus a) werden lediglich die Wertegruppen zwischen der 5. und 95. Perzentile in ihrer jeweiligen Verteilung gezeigt. c) Die Werte aus b) in 1-Hz-Schritten gruppiert. In Farbe ist der Mittelwert, in Schwarz die Standardabweichung zu sehen. Sollte der Sänger Formantentuning anwenden, so liegen die Formanten auf den Harmonischen.

dass es zu unscharf dargestellten Obertönen kommt, eine Verringerung bietet jedoch eine zu geringe Auflösung. Durch die Nullpolsterung sind die Frequenzen feiner abgestuft, ohne die tatsächliche Auflösung zu ändern. Die im OA erstellten Spektrogramm-Abbildungen werden im Folgenden ebenfalls mit diesen Einstellungen erstellt.

4 Ergebnisse der Spektralanalysen

In diesem Kapitel wird zunächst abgeschätzt, welchen Einfluss Instrumente auf die in Abschnitt 3 beschriebenen Analysemethoden haben können. Anschließend folgen die exemplarischen Analysen dieser Methoden anhand des Sängers Jochen Kowalski und Russell Oberlin.

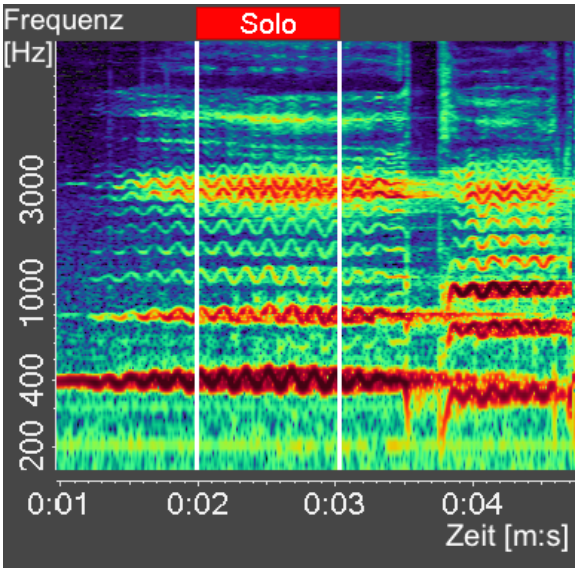
4.1 Abschätzung des Fehlers durch Instrumentaleinfluss

Um abzuschätzen, welchen Einfluss Instrumente auf die Obertöne der Stimme haben, wurde eine Aufnahme von Jochen Kowalski (Gesamtdauer: ca. 5 Minuten), die zur HE (vlg. Abschnitt 3.4) markierten verwendbaren Passagen von fast 90 s reinem Gesang enthält, mit Instrumenten unterlegt. In diesen 90 s sind die Vokale wie folgt aufgeteilt: /a/ ca. 25 s, /e/ ca. 19 s, /i/ ca. 25 s, /o/ ca. 18 s und /u/ ca. 2 s. Die orchestrale Instrumentalpassage mit einer Dauer von etwa 6 s wurde aus demselben Lied von gesangsfreien Stellen entnommen und mit der Software Audacity 2.0.6 [28] so oft aneinander gereiht, bis die Gesamtdauer der aneinandergereihten Instrumentalpassage der Gesamtdauer des ursprünglichen Liedes von 5 Minuten entsprach, und als eigene Spur im unkomprimierten WAV-Format exportiert. Anschließend wurde diese neue „Instrumentalspur“ mit der Originalspur ohne Änderung der Lautstärken der einzelnen Spuren abgemischt und ebenfalls exportiert. Somit sind in dieser Abmischung die oben genannten 90 s Gesang von Instrumenten begleitet und können absolut identisch zum Original mit reinem Gesang analysiert werden. Dadurch ist es möglich, gezielt die Spektren von Gesang und Instrumenten einzeln sowie in ihrer Abmischung zu betrachten und so zu erkennen, welche Einflüsse die Instrumentalbegleitung auf die Analysemethoden hat. Die Auswahl der Instrumentalspur wurde so gewählt, dass die resultierende Abmischung gerade nicht mehr den Auswahlkriterien (vgl. Abschnitt 3.2) entspricht.

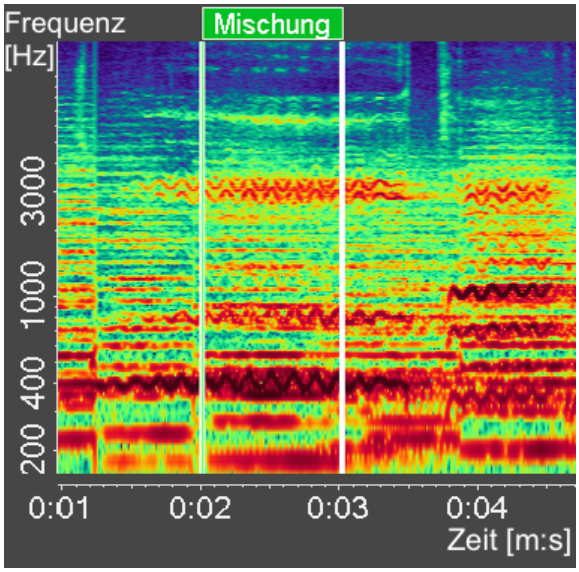
Einfluss der Instrumentalbegleitung auf das Spektrum eines einsekündigen Zeitausschnitts:

Die Abbildungen 4.1a & b zeigen denselben Ausschnitt des Liedes einmal ohne und einmal mit Instrumentalbegleitung, wobei jeweils eine Markierung von 2 - 3 s gesetzt wurde. Man erkennt deutlich, dass die Obertöne mit geringerer Intensität (ca. 1000 - 2000 Hz in der Markierung) durch die Instrumente fast völlig verdeckt werden während die mit hoher Intensität (ca. 400 Hz und 3000 Hz) deutlich erkennbar bleiben. In Abbildung 4.2 sind die über die Dauer von 1 s gemittelten Spektren dieser markierten Bereiche zusammen mit dem gemittelten Spektrum aus der korrespondierenden Instrumentalspur zu sehen. Für diesen zeitlich kurzen Ausschnitt erkennt man gut, dass das abgemischte Spektrum bei fast allen Frequenzen den höchsten Intensitäten eines der beiden ursprünglichen Spektren (Solo-Gesang, Instrumental) entspricht. Die Obertöne der reinen Stimme werden zwischen 1000 und 2000 Hz fast völlig durch die lauten Instrumente verdeckt, dafür stechen die bei ca. 3000 Hz und 6000 - 8000 Hz deutlich hervor. Auch die ersten Harmonischen der Gesangsstimme um 400 Hz und 800 Hz sind noch erkennbar, die Maxima der Mischung werden in diesem Fall breiter durch die Instrumente.

[28] `http://audacity.sourceforge.net`, zuletzt aufgerufen am: 26.03.2015



(a) Ausschnitt aus der reinen Gesangsspur



(b) Gleicher Ausschnitt wie in a) mit Instrumenten

Abbildung 4.1: Kurzer Liedausschnitt als Spektrogramm a) ohne und b) mit Instrumentalbegleitung. In der markierten Stelle in a) sind die Obertöne des Gesangs gut zu erkennen, in der gleichen Stelle in b) sind sie von etwa 1000 - 2000 Hz nicht mehr erkennbar.

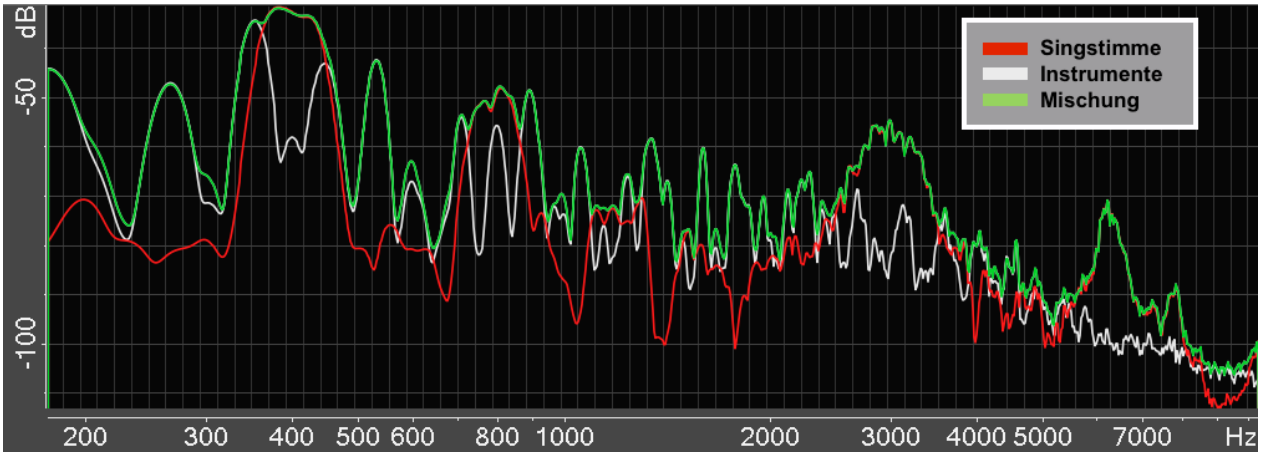
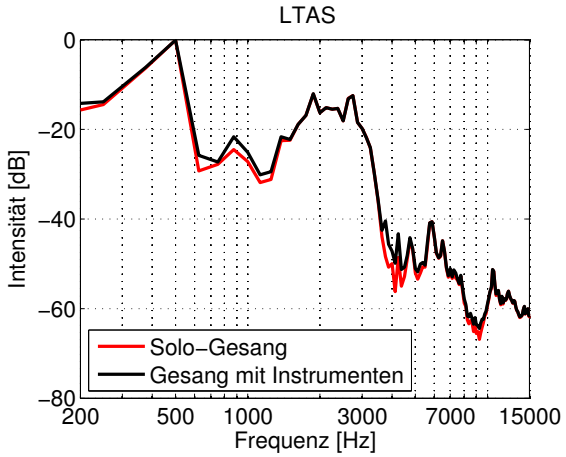
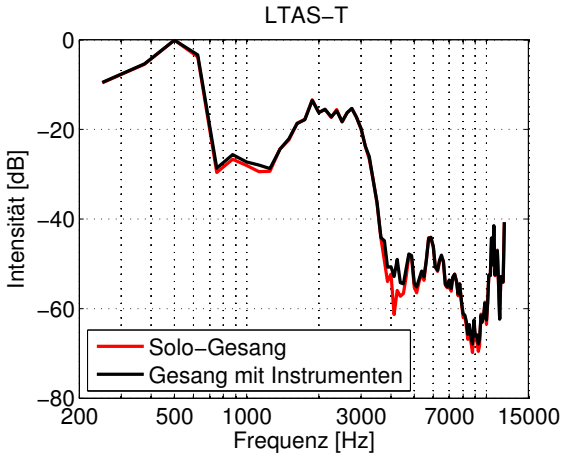


Abbildung 4.2: Spektrum (x=Frequenz, y=Intensität) der markierten Ausschnitte (gemittelt über 1 s) aus Abbildung 4.1 a) und b) sowie der reinen Instrumentalspur zu diesem Zeitpunkt. Das Spektrum der Abmischung entspricht bei fast allen Frequenzen dem jeweils stärkeren der Einzelspektren, daher sind diese meist nicht erkennbar.



(a) LTAS des Vokals /i/ über das gesamte Lied jeweils mit und ohne Instrumente.



f

(b) LTAS-T des Vokals /i/ über das gesamte Lied jeweils mit und ohne Instrumente.

Abbildung 4.3: a) LTAS und b) LTAS-T des Vokals /i/ (Gesamtdauer: ca. 25 s) jeweils mit und ohne Instrumentalbegleitung. Die Maximalintensität jeder Kurve ist zur besseren Vergleichbarkeit auf 0 dB angehoben. Die Intensitäten bleiben über den gesamten angezeigten Frequenzbereich beinahe gleich. Lediglich an wenigen Stellen (z.B. 600 - 1000 Hz und 3500 - 4500 Hz) werden die geringen Intensitäten der reinen Gesangsstimme erhöht.

Einfluss der Instrumentalbegleitung auf das LTAS und das LTAS-T:

Um den Einfluss der Instrumentalbegleitung auf das Langzeit-gemittelte Spektrum zu bestimmen, sind in Abbildung 4.3a & b das LTAS und das LTAS-T für den Vokal /i/[29] jeweils ohne bzw. mit Instrumentalbegleitung dargestellt. Das heißt, dass im gesamten Lied alle Stellen des Vokals /i/, die den in Abschnitt 3.2 erläuterten Auswahlkriterien entsprechen und somit zur Analyse markiert wurden, in das LTAS bzw. LTAS-T einberechnet werden. Weder das LTAS noch das LTAS-T haben sich durch die hier gewählte Instrumentalbegleitung wesentlich verändert. An wenigen Stellen, an denen das Gesangsstück geringere Intensitäten aufweist (z.B. 600 - 1000 Hz und 3500 - 4500 Hz), werden die Intensitäten der Gesangsstimme erhöht. Das LTAS-T zeigt oberhalb von 10000 Hz deutliche Intensitätsschwankungen (für die Solo- und Misch-Version) im Vergleich zum LTAS und bricht dann ab. Das abrupte Ende resultiert daher, dass nur der Grundton und die nächsten 19 Obertöne extrahiert werden, die dann zum LTAS-T beitragen, und somit nicht alle Frequenzen bis zur Nyquist-Frequenz von 22050 Hz (bei einer Abtastrate der CDs von 44100 Hz) enthalten sind (hierfür müsste der Sänger mit einer Grundfrequenz von 22050/20 Hz $\approx$ 1100 Hz singen). In Abbildung 4.4 sind alle extrahierten Harmonischen, die zum LTAS-T aus Abbildung 4.3b beitragen, für die Solo-Stimme und die Misch-Version (Stimme mit Instrumentalbegleitung) dargestellt. Hier kann man zum einen gut die starke Intensitätsschwankung ab 10000 Hz erklären, da nur

[29]Das LTAS und das LTAS-T für den Vokal /i/ wurden exemplarisch ausgewählt. Bei den LTAS/LTAS-T der restlichen Vokale (/a/, /e/, /o/, /u/) lassen sich keine nennenswerten Unterschiede und damit Erkenntnisse in Bezug auf den Instrumentaleinfluss erkennen.

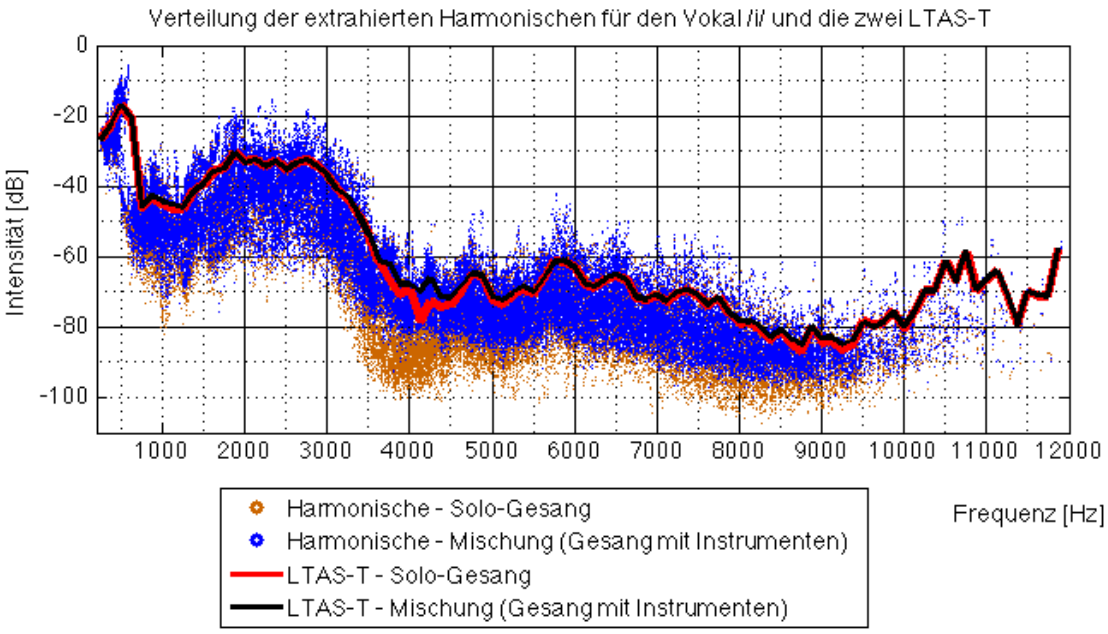


Abbildung 4.4: Nochmals das in Abbildung 4.3b dargestellte LTAS-T mit und ohne Instrumente bei nicht-logarithmischer Frequenzachse. Zusätzlich sind alle extrahierten Harmonischen (mit und ohne Instrumentalbegleitung), die für die beiden LTAS-T verwendet wurden, als einzelne Punkte dargestellt. Die Intensitäten beider Datensätze sind jeweils an der Harmonischen mit der höchsten Intensität normiert (vgl. Abschnitt 3.3). In der Verteilung der Harmonischen mit Instrumentalbegleitung (blau) sind weniger Harmonische mit geringer Intensität zu erkennen, was besonders im Bereich von etwa 3500 Hz bis 5000 Hz deutlich wird. Unter -80 dB sind dort fast ausschließlich die schwachen Obertöne des reinen Gesangs (orange) zu sehen. Somit ändert sich der gemittelte Intensitätsverlauf über den Frequenzbereich (LTAS-T) kaum und die LTAS-T-Kurven verlaufen nahezu deckungsgleich. Ab etwa 10000 Hz wird die Anzahl an extrahierten Harmonischen deutlich geringer, was zu starken Intensitätsschwankungen führt.

noch wenige Punkte zum LTAS-T beitragen und somit ein einzelner extrahierter Wert möglicherweise die Intensität der dortigen Frequenzgruppe darstellt. Zum anderen ist deutlich zu erkennen, dass der Intensitätsverlauf des LTAS-T auch durch die Instrumente erhalten bleibt, die schwachen Intensitäten der reinen Gesangsversion (orange Punkte besonders bei 3500 Hz bis 5000 Hz und 7500 Hz bis - 9000 Hz) jedoch wegfallen. Dadurch, dass die stärkste extrahierte Harmonische als Referenzwert (vgl. Abschnitt 3.3) verwendet wird, bleiben auch die mittleren Intensitäten vergleichbar und die beiden LTAS-T-Kurven verlaufen nahezu deckungsgleich.

Einfluss der Instrumentalbegleitung auf Position und Intensität des Sängerformanten:

Zwar ist der Einfluss der Instrumente auf das LTAS/LTAS-T wie bisher gezeigt relativ gering, es kann aber dennoch zu Änderungen am Sängerformanten kommen, was mit Abbildung 4.5 genauer erläutert werden soll. Dabei wurden für die reine Gesangsversion und die Abmischung die zentrale Position, die Position des Maximums und die 3-dB-Breite des Sängerformanten anhand der in Abschnitt 3.5.2 erläuterten Methode für jeden Vokal bestimmt. „Alle“ bedeutet, dass das LTAS/LTAS-T über alle markierten Vokale berechnet wurde und dann aus diesen die eben aufgezählten Werte berechnet wurden. Bei „/a/“ (bzw. „/e/“, „/i/“, „/o/“ und „/u/“) wurden diese nur aus dem LTAS/LTAS-T des Vokals /a/ (bzw. /e/, /i/, /o/ und /u/) berechnet. In der Abbildung bezeichnen die Kreuze jeweils diese zentrale Position, die senkrechten Balken die 3-dB-Breite und der ausgefüllte rote Kreis die Lage des dazugehörigen Intensitätsmaximums. Wie die geringe Veränderung am LTAS/LTAS-T (vgl. Abb. 4.3) bereits vermuten lässt, ist der Einfluss der Instrumente auf die ermittelte zentrale Position, die Position des Maximums und die Breite des Sängerformanten meist gering. Lediglich bei den Vokalen /a/ und /i/ zeigen sich im LTAS-T Änderungen: Während sich beim Vokal /a/ nur die Position des Maximums leicht ändert, ist beim /i/ sowohl die Breite als auch die zentrale Position deutlich verändert. Abbildung 4.6 zeigt zur Erklärung der stark veränderten zentralen Position und Breite des Sängerformanten einen Ausschnitt aus Abbildung 4.3b von 1500 Hz bis 3000 Hz. Das Maximum befindet sich für beide LTAS-T (mit und ohne Instrumentaleinfluss) bei derselben Frequenz. In Richtung der niedrigen Frequenzen nimmt die Intensität bis zum ersten Minimum ab und dann wieder zu. Im Fall der reinen Gesangsversion (rote Kurve), ist der Intensitätsunterschied vom Maximum zum ersten Minimum 3.03 dB, im Fall der mit Instrumenten begleiteten Version (schwarz) beträgt der Unterschied aber lediglich 2.87 dB (was selbst in dieser Abbildung noch nicht zu erkennen ist). Somit wird erst bei 1625 Hz die 3-dB-Schwelle gefunden, was sich somit auf die Position und die Breite auswirkt.

Zusätzlich wurden die relativen Intensitäten der Sängerformanten für beide Versionen des Liedes für jeden Vokal einmal im LTAS und einmal im LTAS-T nach der in Abschnitt 3.5.2 beschrieben Methode berechnet und in Abbildung 4.7a dargestellt. Man erkennt, dass sich die Intensitäten nur wenig unterscheiden. Je größer die relative Intensität, desto stärker ist der Sängerformant ausgebaut. Der Vokal /u/ weist sowohl im LTAS als auch im LTAS-T eine sehr geringe relative Intensität auf. Dies kann durch die kurze Gesamtdauer von nur 2 s begründet werden, da hier keine erkennbare statistische Verteilung vorliegt und wahrscheinlich die gesungene Lautstärke in den 2 s eher gering ist. Für eine generelle Aussage müssen mehr Passagen mit dem gesungen Vokal /u/ untersucht werden. Zur genaueren Betrachtung der Unterschiede der Intensitäten mit und ohne Instrumente wurde von den relativen Intensitäten (in dB) die Differenz $\Delta I = I_{mischung} - I_{solo}$ der instrumental begleiteten Version $I_{mischung}$ und der reinen Gesangsversion I_{solo} gebildet. Beim LTAS liegen die Unterschiede zwischen $\Delta I_{ltas,min} \approx -0.060$ dB und $\Delta I_{ltas,max} \approx 0.048$ dB, beim LTAS-T fallen diese etwas größer aus: $\Delta I_{ltas,min} \approx -0.61$ dB bis $\Delta I_{ltas,max} \approx 0.65$ dB. Die bestimmten relativen Intensitäten

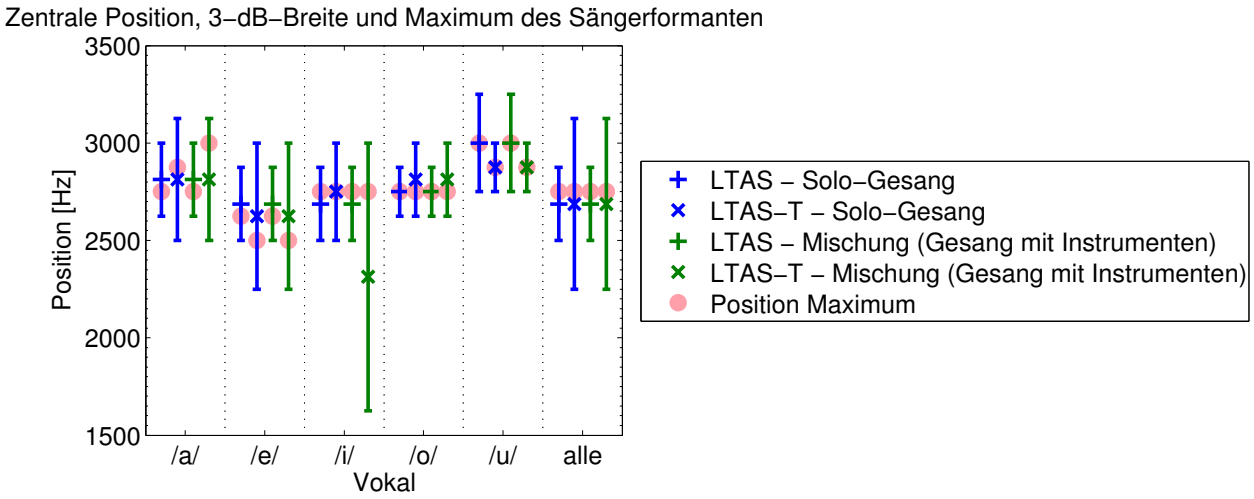


Abbildung 4.5: Für die reine Gesangsversion und die Abmischung mit Instrumenten sind für jeden Vokal jeweils die aus dem LTAS und dem LTAS-T ermittelten zentralen Positionen (Kreuze) des Sängerformanten, die 3-dB-Breiten (senkrechte Balken) und die Positionen der Maximalintensitäten (rote Kreise) der Sängerformanten dargestellt (vgl. Abschnitt 3.5.2). Beim Vokal /a/ ändert sich die Position des Maximums im LTAS-T durch die Instrumente, beim Vokal /i/ kommt es zu einer deutlichen Verschiebung der zentralen Position und der 3-dB-Breite.

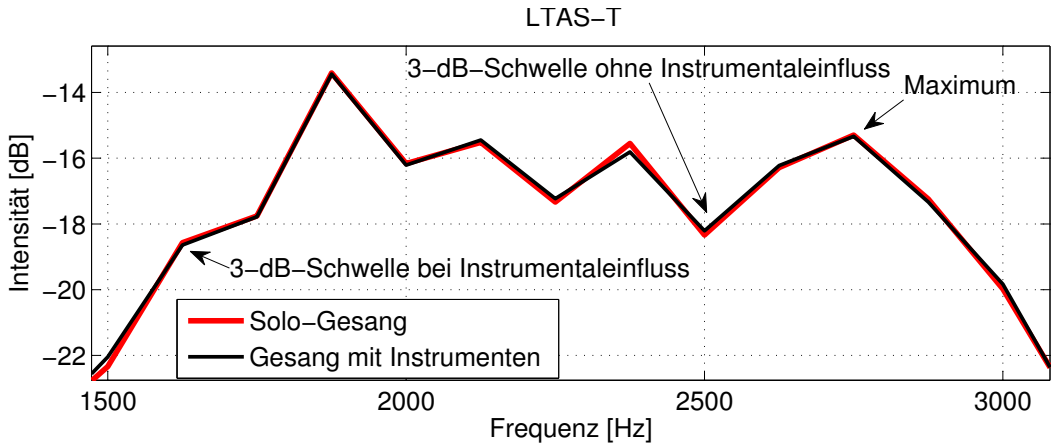
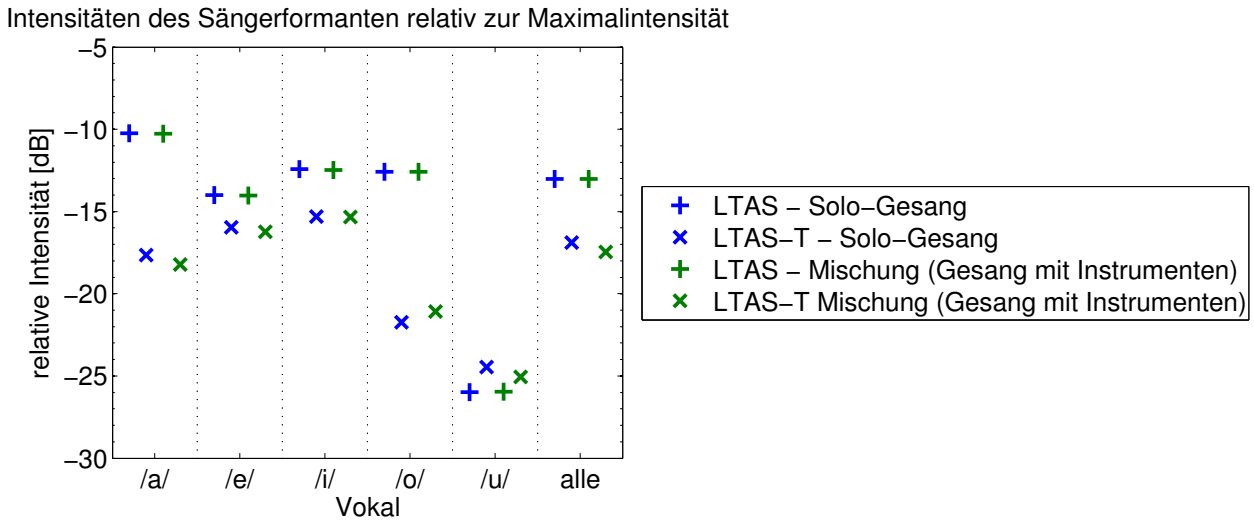
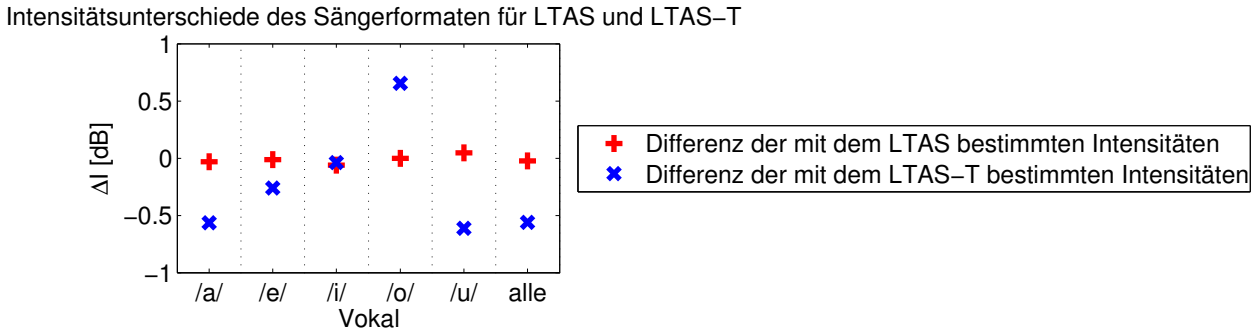


Abbildung 4.6: Für den Vokal /i/ in Abbildung 4.5 ändert sich die zentrale Position und die Breite des Sängerformanten im LTAS-T unter Instrumenteneinfluss deutlich. Hier ist ein Ausschnitt von Abbildung 4.3b zu sehen, mit dem diese Veränderung erklärbar ist. Ausgehend von der maximalen Intensität des LTAS-T zwischen 2000 Hz und 4000 Hz wird die 3-dB-Schwelle bestimmt. Beim LTAS-T ohne Instrumente ist diese bei 2500 Hz mit 3.03 dB Differenz erreicht. Beim LTAS-T unter Instrumentaleinfluss sind es bei 2500 Hz nur 2.87 dB Differenz, sodass die 3dB-Schwelle erst bei 1625 Hz erreicht wird.



(a)



(b)

Abbildung 4.7: a) Relative Intensitäten des Sängerformanten bestimmt nach der in Abschnitt 3.5.2 beschriebenen Methode aus dem LTAS und LTAS-T der einzelnen Vokale jeweils mit und ohne Instrumentaleinfluss. b) Die Differenzen der jeweiligen relativen Intensitäten (in dB) $\Delta I = I_{mischung} - I_{solo}$ liegen zwischen $\Delta I_{ltas,min} \approx -0.060$ dB und $\Delta I_{ltas,max} \approx 0.048$ dB für das LTAS und zwischen $\Delta I_{ltas,min} \approx -0.61$ dB und $\Delta I_{ltas,max} \approx 0.65$ dB für das LTAS-T.

des Sängerformanten im LTAS-T unter Instrumentaleinfluss sind für alle Vokale (außer für den Vokal /o/) geringer als für die reine Gesangsversion, liegen aber dennoch in akzeptablen Größenordnungen, während beim LTAS kaum ein Unterschied festzustellen ist. Ob dies generell für den Vokal /o/ gilt, kann an dieser Stelle nicht gesagt werden, da dies sowohl von der Lautstärke der Instrumente als auch von der Konfiguration des Vokaltraktes des Sängers und dessen Lautstärke abhängt. Auch hier müssen für eine detaillierte Antwort zusätzliche Analysen durchgeführt werden.

Eine weitere Beobachtung zu den Intensitäten in Abbildung 4.7a ist, dass die relativen Intensitäten im LTAS größer sind als im LTAS-T (außer beim eben angesprochenen Vokal /u/, welcher eine nur sehr geringe Dauer aufweist). Dies ist wahrscheinlich mit der reduzierten Tonhöhenabhängigkeit zu begründen. Abbildung 4.8 zeigt das LTAS im Vergleich mit dem

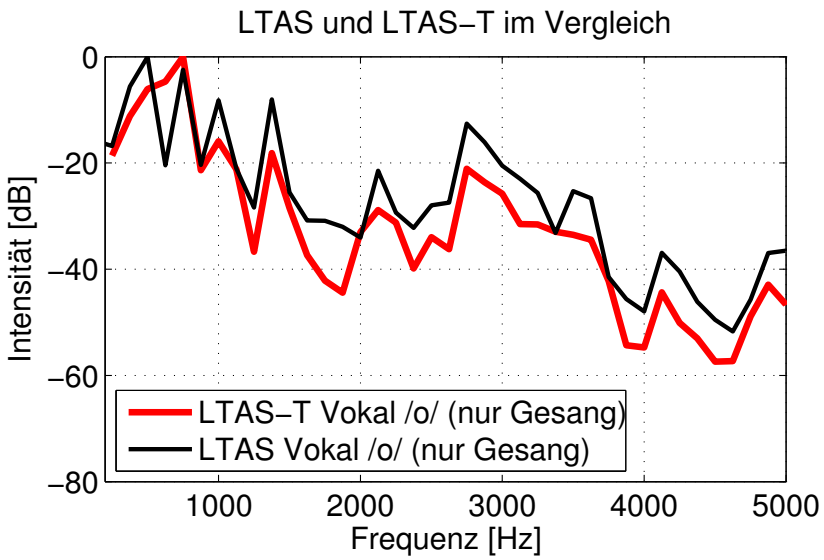


Abbildung 4.8: LTAS und LTAS-T für den Vokal /o/ (nur Gesang). Im direkten Vergleich erkennt man hier die größere Tonhöhenabhängigkeit des LTAS: Das LTAS weist von 0 Hz bis 1000 Hz drei sehr spitz zulaufende Maxima auf, während beim LTAS-T im selben Bereich drei weniger spitze Maxima auftreten. Die scharfen Spitzen werden wie in Boersma & Kovacic (2006) beschrieben direkt durch die ersten Harmonischen verursacht, wenn der Sänger besonders oft auf der selben Tonhöhe singt. Im LTAS-T verschwindet diese Abhängigkeit nicht vollkommen, sie wird jedoch reduziert. Die dadurch breiter, aber flacher ausfallenden Maxima sind wahrscheinlich der Grund, warum die relative Intensität geringer wird.

LTAS-T (beide ohne Instrumente) für den Vokal /o/. Während im LTAS die Maxima relativ spitz zulaufen, sind die entsprechenden Maxima im LTAS-T etwas abgeflachter. In der Abbildung ist dies an den Maxima im Bereich von 0 Hz bis 1000 Hz oder bei 2200 Hz zu erkennen. Diese spitzen Maxima werden durch die Harmonischen der Singstimme verursacht. Wenn die Singstimme vorwiegend in einem bestimmten Tonhöhenbereich (z.B. 400 Hz bis 500 Hz) singt und weniger häufig auch in höheren (z.B 500 Hz bis 800 Hz) oder tieferen Tonhöhen (z.B 200 Hz bis 400 Hz), dann werden beim LTAS die weniger häufig gesungenen Tonhöhen über dieselbe Zeit gemittelt wie die häufig gesungenen Tonhöhen. Dadurch erhalten sie im Verhältnis weniger Gewichtung. Beim LTAS-T wird jede Tonhöhe separat mit ihrem tatsächlichen Vorkommen gewichtet (vgl. Abschnitt 3.5.1) und führt so zu einer geringeren Tonhöhenabhängigkeit. Die somit flacher zulaufenden Maxima sind daher der Grund für die geringere relative Intensität des Sängerformanten im Vergleich zum LTAS (vgl. Abb. 4.7).

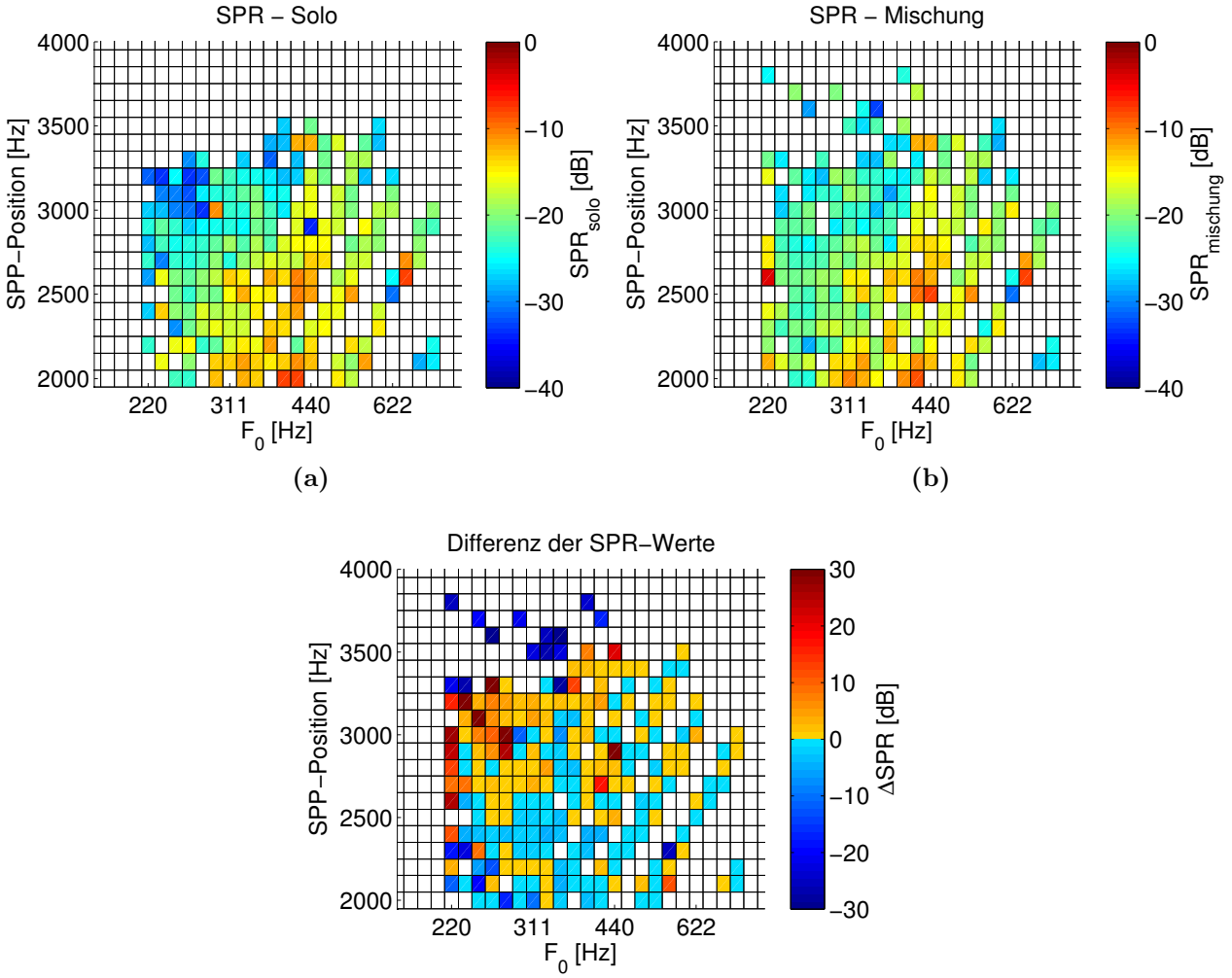


Abbildung 4.9: Die SPR über alle Vokale ist in Abhängigkeit von der Tonhöhe und von der Position des SPP a) ohne und b) mit Instrumenten dargestellt. Im direkten Vergleich erkennt man besonders, dass sich etwa im Bereich von $F_0 = 220$ Hz bis 311 Hz und der SPP-Position von etwa 3000 bis 3250 Hz beim reinen Gesang ein deutlicher Unterschied der SPR abzeichnet. Durch die Instrumentalbegleitung sind in b) oberhalb der SPP-Position von 3500 Hz neue SPR-Werte hinzu gekommen. In diesem Frequenzbereich waren die Obertöne der Instrumente (an den entsprechenden Stellen im Lied) von höherer Intensität als die der Singstimme und führen so zu diesen neu erkannten SPP-Positionswerten. In c) ist zur Verdeutlichung der Unterschiede von a) und b) die Differenz der SPR-Werte (in dB) gebildet ($\Delta SPR = SPR_{mischung}[\text{dB}] - SPR_{solo}[\text{dB}]$) und in derselben Form dargestellt. Besonders für die SPP-Positionen unterhalb von 2700 Hz tendiert die Differenz ins Negative. Besonders für Tonhöhen unter 440 Hz ist die Differenz größtenteils positiv. Eine positive Differenz bedeutet, dass die berechnete SPR unter Instrumentaleinfluss größer wird (umgekehrt für negative Differenzen). An den extremen Farbwerten (dunkelrot und dunkelblau) sind die neu erkannten Positionen deutlich zu erkennen.

Einfluss der Instrumentalbegleitung auf die Singing Power Ratio:

Die Abbildungen 4.9a & b zeigen die SPR in Abhängigkeit von der gesungenen Tonhöhe und von der Lage des Singing Power Peak für den reinen Gesang und für die Mischung mit Instrumenten (die Berechnung wurde in Abschnitt 3.5.3 erläutert, die grafische Darstellung ist analog zu der in Abschnitt 3.4.1 besprochenen Methode). Zunächst wurde für die Berechnung nicht zwischen den Vokalen unterschieden. Man erkennt im Vergleich der Abbildungen 4.9a und b, dass die SPR durch die Instrumente teilweise stark unterschiedlich ausfällt (z.B. für $F_0 \approx 220$ Hz bis etwa 311 Hz und den SPP-Position 3000 - 3300 Hz) und auch an einigen Stellen (z.B für die SPP-Position oberhalb von 3500 Hz) komplett neue SPP-Positionen registriert werden. Ansonsten fallen die Farbdifferenzen gering aus. Um dies besser zu veranschaulichen, ist in Abbildung 4.9c die Differenz der SPR-Werte (in dB) aus den Abbildungen 4.9a & b dargestellt ($\Delta SPR = SPR_{mischung} - SPR_{solo}$). Hieran erkennt man deutlich, dass die SPR durch die Instrumente teilweise größer (positive ΔSPR) und teilweise kleiner (negative ΔSPR) ausfällt, als sie bei reinem Gesang ist. Der Mittelwert und die Standardabweichung für alle Differenzen und Tonhöhen betragen $\Delta\overline{SPR} = (-0.35 \pm 9.66)$ dB. Der mittlere SPR-Wert ist also bei der Version mit Instrumenten nur wenig geringer. Dass sich die Position des SPP ändert, ist an der dunkelblauen bzw. dunkelroten Färbung zu erkennen. Da diese veränderten Positionen in der Differenz weit vom Mittelwert entfernt sind, wirken sie sich in diesem Fall besonders auf die relativ große Standardabweichung aus. Um den Einfluss der Instrumente auf die SPR und die SPP-Position in Abhängigkeit von der Tonhöhe zu untersuchen, wurden die ΔSPR-Werte aus Abbildung 4.9c für jede Tonhöhe gemittelt und nach Vokalen aufgeteilt (vgl. Abb. 4.10a). In den hohen Tonhöhen zeigen sich die Unterschiede eher gering, während sie für die tiefen teilweise stark ausgeprägt sind. Besonders der Vokal /o/, aber auch der Vokal /a/ sind hier zu erkennen. Zudem wurde aus den SPP-Positionswerten für jede Tonhöhe die mittlere SPP-Position $p(F_0)$ mit und ohne Instrumentaleinfluss bestimmt und daraus die Differenz $\Delta p = p_{mischung} - p_{solo}$ gebildet (vgl. Abb. 4.10b). Auch hier ist ein geringerer Einfluss der Instrumente auf die SPP-Position für hohe Tonhöhen zu erkennen, und sowohl der Vokal /o/ als auch /a/ zeigen die größten Änderungen in der SPP-Position.

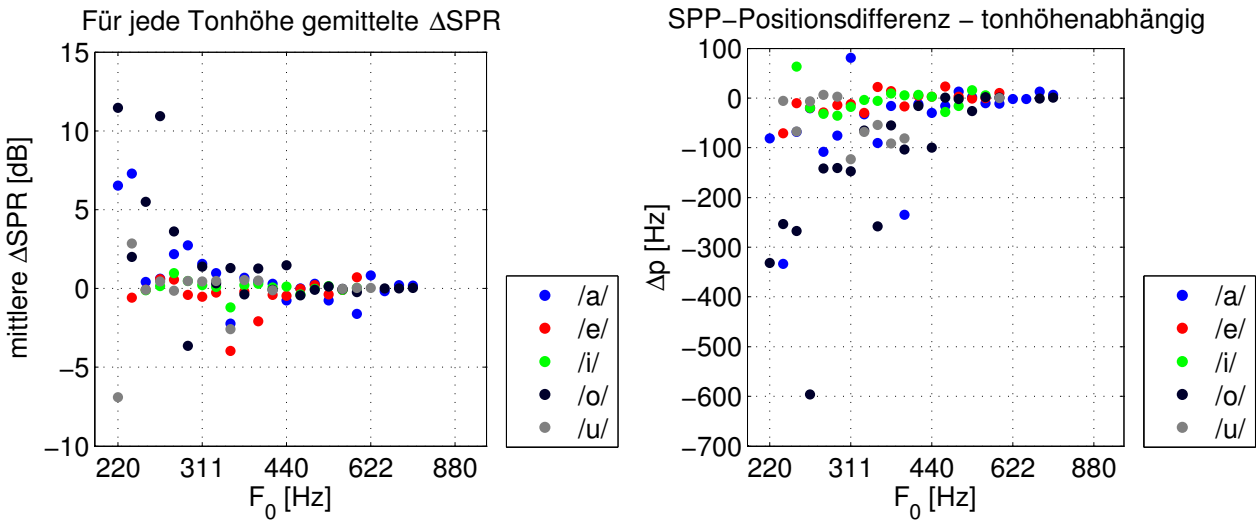


Abbildung 4.10: a) Für jede Tonhöhe wurden die ΔSPR aus Abbildung 4.9c gemittelt und auf die einzelnen Vokale aufgeteilt. Besonders bei den tiefen Tönen ist der Instrumenteneinfluss deutlicher. Für den Vokal /o/ ist hier die Abweichung besonders groß. b) Für jede Tonhöhe und jeden Vokal wurde die mittlere Position p des SPP ohne und mit Instrumentaleinfluss berechnet. Die Differenz zwischen beiden SPP-Positionen $\Delta p = p_{mischung} - p_{solo}$ ist für jeden Vokal aufgetragen. Auch hier zeigt sich vor allem bei den tieferen Tönen eine stärkere Beeinflussung der SPP-Position durch die Instrumentalbegleitung. Diese Abweichung ist auch hier für den Vokal /o/ besonders groß.

Einfluss der Instrumentalbegleitung auf das spektrale Gefälle:

Das spektrale Gefälle S wurde nach der in Abschnitt 3.4.1 erläuterten Methode aus allen Vokalen bestimmt. Abbildung 4.11a, b & c zeigen die Verteilungen ohne und mit Instrumentalbegleitung sowie die Differenz analog zum eben besprochenen Einfluss auf die SPR. Die Differenzwerte in Abbildung 4.11c sind fast ausschließlich negativ (blau), und nehmen besonders für geringe Tonhöhen und geringere Intensitäten weiter ab (dunklerer Blauton). Das heißt, dass die bestimmte Steigung durch den Instrumentaleinfluss fast immer geringer ausfällt. Die minimalen und maximalen Unterschiede des Gefälles sind -14.6 dB/Okt und 13.9 dB/Okt. Diese Werte lassen sich durch stellenweise veränderte I_0- und F_0-Werte aufgrund des Instrumentaleinflusses erklären: Wenn die Werte von I_0 bzw. F_0 beim reinen Gesang nahe den definierten Grenzen der Notenwerte bzw. der Intensitätsabstufungen (vgl. Abschnitt 3.4.1) sind, sorgen bereits geringe Änderungen durch die Instrumente dafür, dass die Werte des Gefälles S in eine andere Gruppierung eingeordnet werden und somit die Differenz zwischen reinem Gesang und Mischversion (Gesang und Instrumente) entsprechend groß ausfällt. Da, wie in Abbildung 4.2 gezeigt, die Harmonischen der Gesangsstimme mit höherer Intensität trotz Instrumentaleinfluss nahezu unverändert bleiben, führt ein laut gesungener Ton (mit größerem I_0 und tendenziell höheren Intensitäten der Obertöne) zu einem geringeren Einfluss der Instrumente auf das Gefälle. Der gemittelte Gefälleunterschied und die Standardabweichung betragen $\Delta\bar{S} = (-1.44 \pm 3.66)$ dB/Okt. Somit fällt das ermittelte

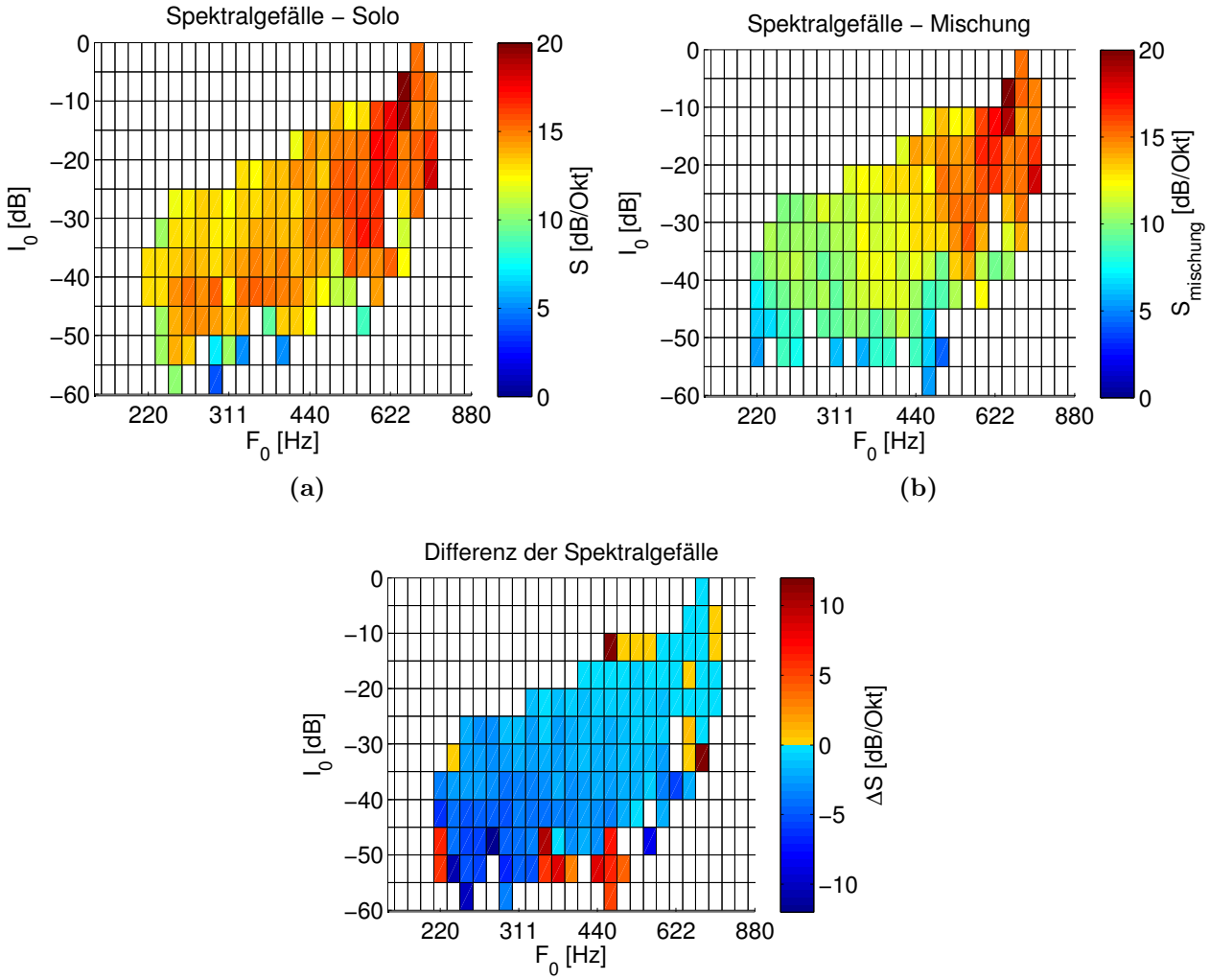


Abbildung 4.11: Das spektrale Gefälle S über alle Vokale ist in Abhängigkeit von der Tonhöhe F_0 und von der Intensität der Grundfrequenz I_0 a) ohne und b) mit Instrumentalbegleitung dargestellt. In c) wurde zur Verdeutlichung der Unterschiede von a) und b) die Differenz der Gefälle gebildet ($\Delta S = S_{mischung} - S_{solo}$) und in derselben Form dargestellt.

Gefälle eines mit Instrumenten begleiteten Liedes meist niedriger aus als bei reinem Gesang ohne Instrumente.

Da die Instrumente so gewählt wurden, dass die untersuchten Gesangspassagen durch den Einfluss der Instrumente teilweise gerade nicht mehr den in Abschnitt 3.2 besprochenen Auswahlkriterien entsprechen, ist die in diesem Kapitel besprochene Abschätzung der Instrumente bezüglich der Methoden und deren Messwerte (Form des Kurz- und Langzeitspektrums, Positionen und relative Intensitäten des Sängerformanten, SPR und SPP-Position sowie spektrales Gefälle) eine Abschätzung des größten Fehlers. Im Kurzzeitspektrum über 1 s wurde erkannt, dass ggf. die Harmonischen der Gesangsstimme von denen der Instrumente über-

deckt werden (vgl. Abbildungen 4.1 und 4.2). Im LTAS und LTAS-T (vgl. Abb. 4.3 und 4.4) sowie in den daraus berechneten Messwerten (relative Intensität, zentrale Position und Maximalposition des Sängerformanten) konnten geringe Veränderungen festgestellt werden (vgl. Abb. 4.5, 4.6 und 4.7). Bei der SPR und der SPP-Position ergaben sich besonders für den Vokal /o/, aber auch für den Vokal /a/ speziell bei Tonhöhen unter 440 Hz klare Unterschiede durch den Instrumentaleinfluss im Vergleich zur reinen Gesangsstimme (vgl. Abb. 4.10a, b).

Dies sind allerdings keine generell gültigen Erkenntnisse. Es ist wahrscheinlich, dass andere Instrumentenkombinationen mit unterschiedlichen Lautstärken andere Einflüsse auf die hier besprochenen Parameter (Form des Kurz- und Langzeitspektrums, Positionen und relative Intensitäten des Sängerformanten, SPR und SPP-Position sowie spektrales Gefälle) haben. Bereits durch die quasi endlosen Kombinationsmöglichkeiten an Instrumenten und Lautstärken kann dies in der vorliegenden Arbeit nicht erfasst werden.

4.2 Beispielhafte Analyse zweier Sänger

4.2.1 Spektralanalyse von Jochen Kowalski

Als erstes Beispiel für die Anwendung der in Abschnitt 3 erläuterten Methoden soll der Falsettist (vgl. Abschnitt 1.1) Jochen Kowalski (vgl. Abschnitt 1.3 und 1.3.1) besprochen werden. Dabei wird das bereits in Abschnitt 4.1 verwendete Lied verwendet (Gesamtdauer der analysierten Vokale: /a/ ca. 25 s, /e/ ca. 19 s, /i/ ca. 25 s, /o/ ca. 18 s und /u/ ca. 2 s). Nur für die LPC-Analyse zur Feststellung von Formantentuning (erläutert in Abschnitt 3.6) waren mehrere Lieder notwendig, um ausreichend viele Daten zu erhalten.

Spektralcharakteristika im Langzeitspektrum

Abbildung 4.12 zeigt das LTAS und LTAS-T von Jochen Kowalski. Dadurch, dass hier alle Vokale (Gesamtdauer ca. 90 s) gemittelt wurden, weisen die Spektren im Vergleich zu denen des Vokals /i/ (Abb. 4.3 in Abschnitt 4.1) im Bereich von 600 Hz bis 1200 Hz keinen solch starken Intensitätsabfall von etwa 20 dB auf. Dies ist durch die unterschiedlichen Formantenlagen der Vokale zu erklären (vgl. Abb. 2.5), weshalb sich in Abbildung 4.12 der Sängerformant lediglich als schwache (5 dB bzw. 8 dB) Intensitätserhebung bei 2800 Hz zeigt. In Abbildung 4.5a (Abschnitt 4.1) sind die einzelnen ermittelten zentralen Positionen $p_{fs,z}$, die Maximalpositionen $p_{fs,max}$ und die 3-dB-Breiten $b_{fs,3db}$ des Sängerformanten für die Vokale einzeln und gemittelt dargestellt. Die genauen Zahlenwerte sind in Tabelle 4.1 angegeben. Zusätzlich wurden hier die relativen Intensitäten des Sängerformanten aus Abbildung 4.7a in Abschnitt 4.1 hinzugefügt. Wie bereits in Abschnitt 4.1 erläutert, beziehen sich die bestimmten Positions- und Intensitätswerte (unter „/a/“, „/e/“, „/i/“, „/o/“, „/u/“, „alle“) jeweils auf das entsprechende LTAS/LTAS-T, aus dem sie entnommen wurden: entweder über einen einzelnen Vokal (wie z.B. /i/ in Abb. 4.3) oder über alle Vokale insgesamt gemittelt (Abb. 4.12). Alle Maximal- und Zentralpositionswerte liegen für das LTAS und das LTAS-T zwischen von 2500 Hz und 3000 Hz und gemittelt über alle Vokale bei 2750 Hz (Maximum) und 2687.5 Hz (zentrale Position). Lediglich die Breiten und die relativen Intensitäten unterscheiden sich beim LTAS ($b_{fs,max,ltas} = 375$ Hz, $I_{fs,ltas} = -13.01$ dB) und dem

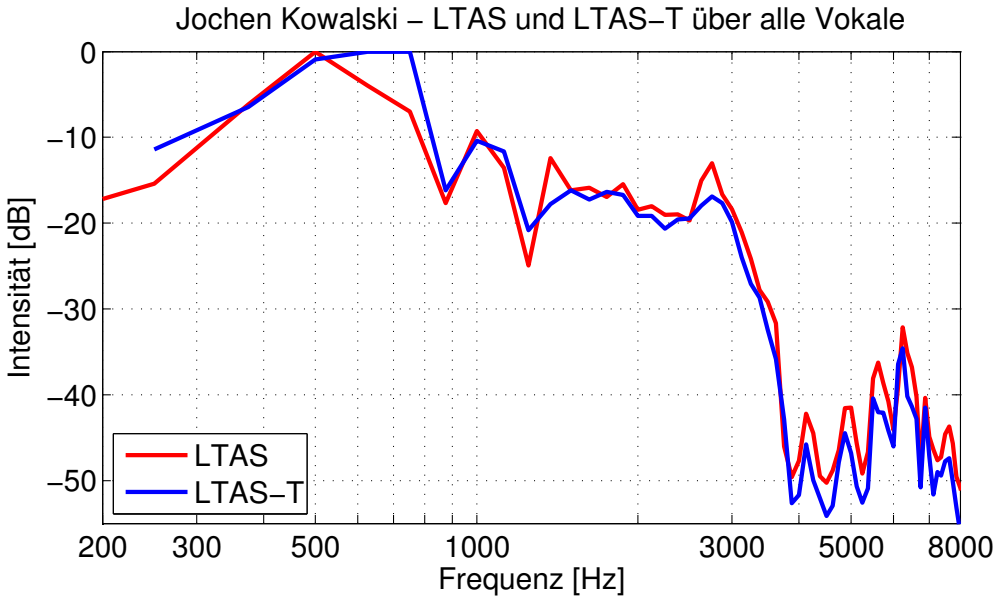


Abbildung 4.12: LTAS und LTAS-T über alle Vokale von Jochen Kowalski (ca. 90 s Gesamtdauer). Der Sängerformant bei ca. 2800 Hz hebt sich optisch nicht so stark hervor wie im Vergleich des LTAS/LTAS-T für den einzelnen Vokal /i/ (vgl. Abb. 4.3 in Abschnitt 4.1). Dies ist durch die verschiedenen Formantenpositionen der jeweiligen Vokale zu begründen. Ein weiteres Intensitätsmaximum kann bei 6250 Hz beobachtet werden.

Tabelle 4.1: Werte der zentralen Position $p_{fs,z}$, der 3-dB-Breite $b_{fs,3db}$, der Maximumsposition $p_{fs,max}$ und der relativen Intensitäten I_{fs} des Sängerformanten für Jochen Kowalski für alle Vokale separat und zusammen. Die Werte sind in den Abbildungen 4.5 und 4.7a in Abschnitt 4.1 grafisch präsentiert.

Vokal	$p_{fs,z}$ [Hz]	$b_{fs,3db}$ [Hz]	$p_{fs,max}$ [Hz]	I_{fs} [dB]
/a/ LTAS	2812.5	375	2750	-10.23
/a/ LTAS-T	2812.5	625	2875	-17.66
/e/ LTAS	2687.5	375	2625	-14.01
/e/ LTAS-T	2625	750	2500	-15.97
/i/ LTAS	2687.5	375	2750	-12.41
/i/ LTAS-T	2750	500	2750	-15.30
/o/ LTAS	2750	250	2750	-12.58
/o/ LTAS-T	2812.5	375	2750	-21.73
/u/ LTAS	3000	500	3000	-25.99
/u/ LTAS-T	2875	250	2875	-24.45
alle LTAS	2687.5	375	2750	-13.01
alle LTAS-T	2687.5	875	2750	-16.89

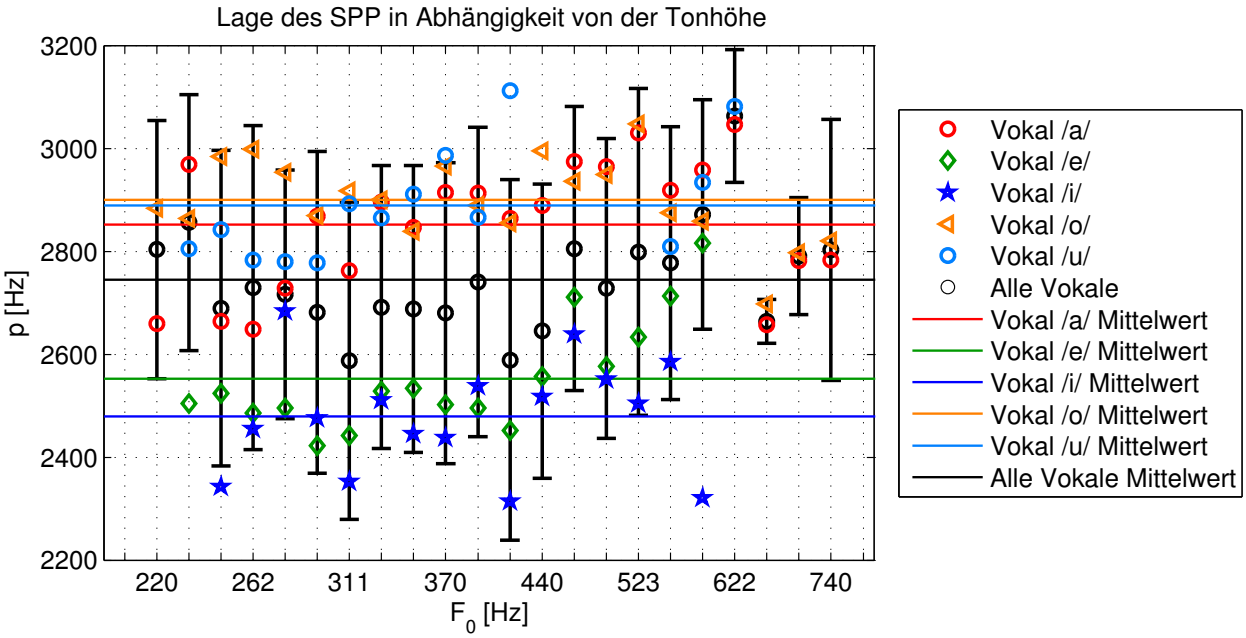


Abbildung 4.13: Berechnete SPP-Positionen (vgl. Abschnitt 3.5.3) für jeden Vokal und jede Tonhöhe. Für die SPP-Position über „alle Vokale“ sind die senkrechten Balken mit ±1 Standardabweichung eingefügt. Wie auch in der Bestimmung der zentralen Position bzw. der Position des Maximums des Sängerformanten, zeichnet sich eine eindeutige Verteilung der einzelnen Vokale über den Frequenzbereich ab (vgl. Abb. 4.5 bzw. Tabelle 4.1).

LTAS-T ($b_{fs,max,ltas-t} = 875$ Hz, $I_{fs,ltas-t} = -16.89$ dB), was bereits in Abschnitt 4.1 durch die Tonhöhenkorrektur im LTAS-T begründet wurde.

Zusätzlich kann in Abbildung 4.12 eine Erhöhung der spektralen Intensität bei 6250 Hz beobachtet werden. Die relativen Intensitäten dieses Maximums betragen −32.16 dB für das LTAS und −34.58 dB für das LTAS-T. Dieses Maximum ist einem der höheren Formanten zuzuschreiben, es kann jedoch anhand dieser einzelnen Beispielanalyse nicht darauf geschlossen werden, ob dies generell bei Kowalski beobachtbar ist.

Singing Power Ratio

Für alle Vokale wurde die Position p des SPP in Abhängigkeit von der Tonhöhe berechnet (vgl. Abschnitt 3.5.3) und in Abbildung 4.13 dargestellt. Zusätzlich sind die Mittelwerte der SPP-Positionen $\bar{p}$ jedes einzelnen Vokals und der Mittelwert aller Vokale mit ±1 Standardabweichung zu sehen. Es ist eindeutig erkennbar, dass die Verteilung der Positionen und die mittlere Position stark vom Vokal abhängen, wie dies auch bei der Berechnung der zentralen Position der Sängerformanten der Fall ist (vgl. Abb. 4.5 bzw. Tabelle 4.1). Die mittlere SPP-Position der Vokale /i/ und /e/ ist deutlich niedriger als die der restlichen Vokale. Die Standardabweichung aller gemittelten Vokale ist entsprechend der großen Streuung der SPP-Positionswerte relativ hoch. Eine eindeutige Tendenz ist nicht zu erkennen. Die SPP-Positionen könnten mit der Tonhöhe zunehmen, die letzten drei Tonhöhen sprechen aber

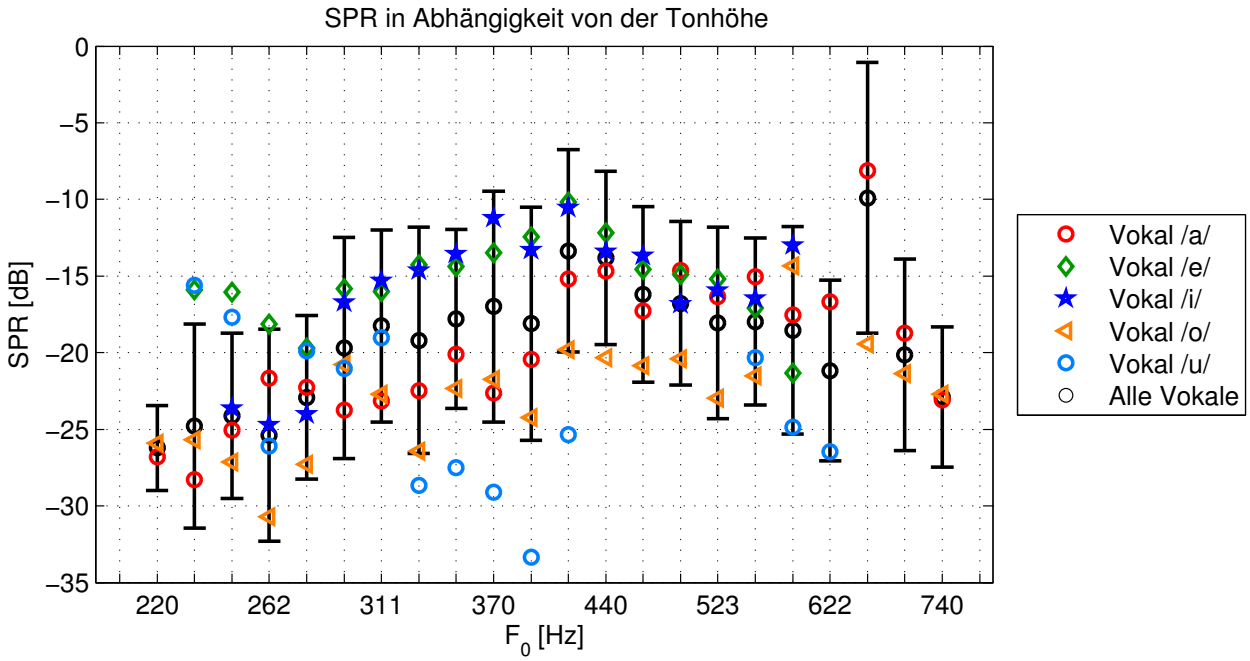


Abbildung 4.14: Berechnete SPR (vgl. Abschnitt 3.5.3) für jeden Vokal und jede Tonhöhe. Von den niedrigen in die höheren Tonhöhen ist bis etwa 1 Halbton unter 440 Hz eine Zunahme der SPR für alle Vokale zu erkennen, bis das mittlere Maximum von −13.36 dB erreicht ist. Danach wird sie wieder geringer. Der Vokal /u/ hat nur eine Gesamtdauer von 2 s, was die starke Abweichung zwischen 311 Hz und 440 Hz erklärt.

eher dagegen, da diese für alle Vokale sehr dicht an der mittleren Verteilung liegen.

Zusätzlich wurden die SPR ebenfalls in Abhängigkeit von der Tonhöhe berechnet (vgl. Abschnitt 3.5.3). Abbildung 4.14 zeigt die berechneten SPR-Werte. Auch hier wird die Standardabweichung der Mittelung aller Vokale durch die senkrechten Balken dargestellt. Mit Ausnahme der Ausreißer des Vokals /u/ (welcher nur 2 s Gesamtdauer aufweist) zwischen 311 Hz und 440 Hz ergibt sich ein tonhöhenabhäniger Verlauf der SPR, welcher zunächst in den niedrigen Tonhöhen sein Minimum erreicht und dann bis einen Halbton unter 440 Hz auf −13.36 dB ansteigt. Für höhere Tonhöhen fällt die SPR wieder ab. Die starke Abweichung einen Halbton über 622 Hz lässt sich mit der geringen Anzahl an SPR-Werten begründen, die zu den jeweiligen Mittelwerten der Vokale beitragen. Das Maximum im SPR-Verlauf entlang der Tonhöhe könnte darauf hindeuten, dass Kowalski in diesem Tonhöhenbereich von etwa 350 Hz bis etwa 500 Hz einen stärkeren Sängerformanten hat. In Tabelle 4.2 sind alle SPP-Positionen, sowie die mittlere SPR über alle Tonhöhen aufgelistet:

Tabelle 4.2: Mittelwerte der bestimmten SPP-Positionen $\bar{p}$ und SPR-Werte $\overline{SPR}$ sowie deren Standardabweichungen für alle Vokale von Jochen Kowalski. Es zeigen sich vergleichbare Größenordnungen im Vergleich zu den Werten in Tabelle 4.1 sowohl für die SPP-Positionen mit den Positionen des Sängerformanten als auch für die mittlere SPR im Vergleich mit den relativen Intensitäten aus dem LTAS-T.

	$\bar{p}$ [Hz]	σ_p [Hz]	$\overline{SPR}$ [dB]	σ_{SPR} [dB]
Vokal /a/	2851	206	−19.72	5.25
Vokal /e/	2552	199	−15.39	4.78
Vokal /i/	2479	259	−16.06	4.57
Vokal /o/	2899	175	−22.80	5.50
Vokal /u/	2888	135	−23.92	3.90
Alle Vokale	2745	268	−19.20	6.31

Es zeigen sich bei mittleren SPR-Werten ähnliche Größenordnungen, wie dies bei den relativen Intensitäten des Sängerformanten der Fall ist (vgl. Tabelle 4.1). Besonders die dem LTAS-T entnommenen Intensitätswerte ähneln denen der SPR, was auf die ähnlichen Verfahren zurückgeht. Man könnte die gemittelte SPR auch als LTAS-T mit nur 2 statt 20 Harmonischen ansehen.

Spektrales Gefälle und Einteilung in die Stimmgattung

Das in Abbildung 4.11a dargestellte spektrale Gefälle Kowalskis wurde für einen einfacheren Vergleich des nun Folgenden nochmals in Abbildung 4.15a dargestellt. Es wurden Grundfrequenzen F_0 im Frequenzbereich von 220 Hz bis 740 Hz und im Intensitätsbereich von 0 dB bis -60 dB extrahiert. Das minimale Gefälle liegt bei $F_{0,min} = 293$ Hz im Bereich $I_{0,min} = -60 \text{ - } -55$ dB und beträgt $S_{min} = 3.78$ dB/Okt, das maximale bei $F_{0,max} = 660$ Hz und $I_{0,max} = -10 \text{ - } -5$ dB und beträgt $S_{max} = 20.11$ dB/Okt. Dabei zählen zu diesen gemittelten Spektralgefällen nur $N_{min} = 4$ bzw. $N_{max} = 2$ Werte, weshalb diese nicht sonderlich aussagekräftig sind. Um bessere Aussagen über das Gefälle für den gesamten F_0-I_0-Bereich machen zu können, sind in Abbildung 4.15b & c die Anzahl an verwendeten Gefälle-Werten bzw. die Standardabweichungen des Gefälles σ_S dargestellt. In der Verteilung der gemessenen Anzahl an Gefälle-Werten (Abb. 4.15b) ist deutlich zu erkennen, dass etwa zwischen 300 Hz und 600 Hz und von −45 dB bis −15 dB die meisten Gefälle-Werte extrahiert wurden. Für die meisten F_0-I_0-Gruppen am Rande der Verteilung ist die Anzahl wesentlich geringer (dunkelblaue Stellen) und reicht, wie bereits erwähnt, bis zu gerade einmal zwei Werten. Dies hat mehrere Gründe: Zum einen konnten die am besten geeigneten Gesangspassagen (nach den in Abschnitt 3.2 besprochenen Auswahlkriterien) in diesem Frequenzbereich gefunden werden. Zum anderen hängt es vom Lied ab, wie viele hoch und tief gesungene Passagen darin vorkommen. Betrachtet man die Standardabweichung (Abb. 4.15c), so liegt diese für die meisten F_0-I_0-Gruppen (und auch im Bereich der meisten extrahierten Werte, vgl. dazu die Maxima in 4.15b) zwischen 2.5 dB/Okt und 1 dB/Okt. Mögliche Einflussfaktoren auf das Gefälle bzw. die Standardabweichung sind der starke Hall in der Aufnahme, die Aufnahmebedingungen oder eventuell auch der Sänger selbst. Ersteres hat besonders dann einen größeren Einfluss

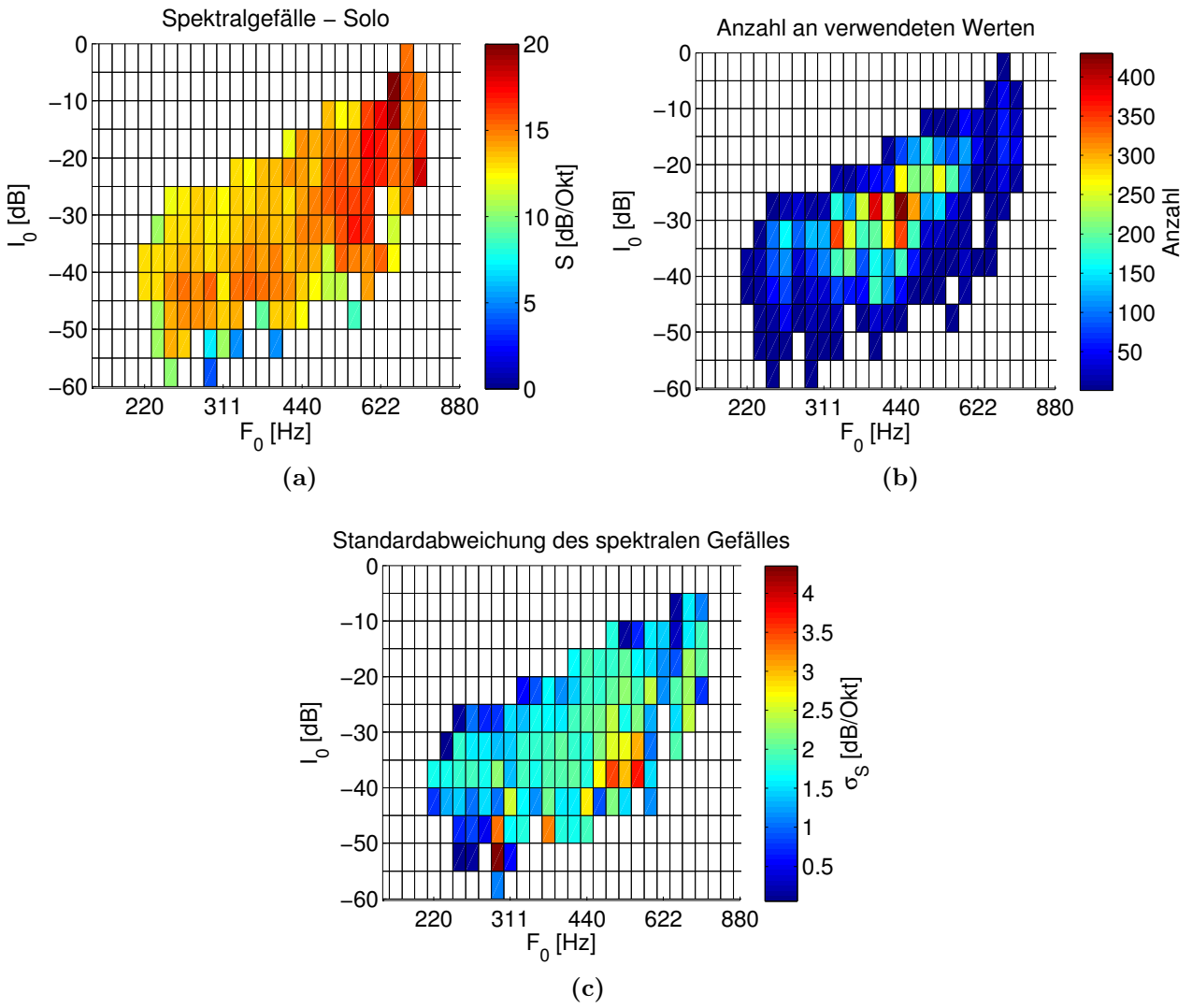


Abbildung 4.15: a) Spektrales Gefälle (vgl. Methode in Abschnitt 3.4.1) für ein Lied von Jochen Kowalski. Im niedrigen Tonhöhen- und Intensitätsbereich (ca. 220 Hz bis 440 Hz und −45 dB bis −60 dB) liegt das Gefälle größtenteils unterhalb von 15 dB/Okt, in den hohen Bereichen (ca. ab 600 Hz und −30 dB bis 0 dB) darüber. b) Die Anzahl der für a) verwendeten Spektralwerte. Bei den mittleren Frequenzen (ca. 300 Hz bis 600 Hz) und Intensitäten (−45 dB bis −15 dB) wurden wesentlich mehr Gefälle-Werte extrahiert. c) Die Standardabweichungen der gemittelten Gefälle-Werte für jede F_0-I_0-Gruppe liegt, wie am Farbton zu erkennen ist, meist unterhalb von 2.5 dB/Okt. Besonders für die F_0-I_0-Gruppen mit einer hohen Anzahl an bestimmten Gefälle-Werten (an den hellblauen bis dunkelroten Farben in b) zu erkennen) weist dies auf einen geringen Fehler im Mittelwert hin.

auf das Gefälle, wenn der Sänger die Tonhöhe ändert und gleichzeitig die gesungene Lautstärke reduziert. Dadurch sind die Frequenzanteile des Halls (also des zuvor Gesungenen) möglicherweise noch stärker als einige Harmonischen der Singstimme, was zu fehlerhafter Harmonischenextraktion führt und somit das Gefälle verfälscht. Außerdem ist nicht bekannt, ob der Sänger während der Aufnahme immer frontal in das Mikrofon gesungen hat oder ob

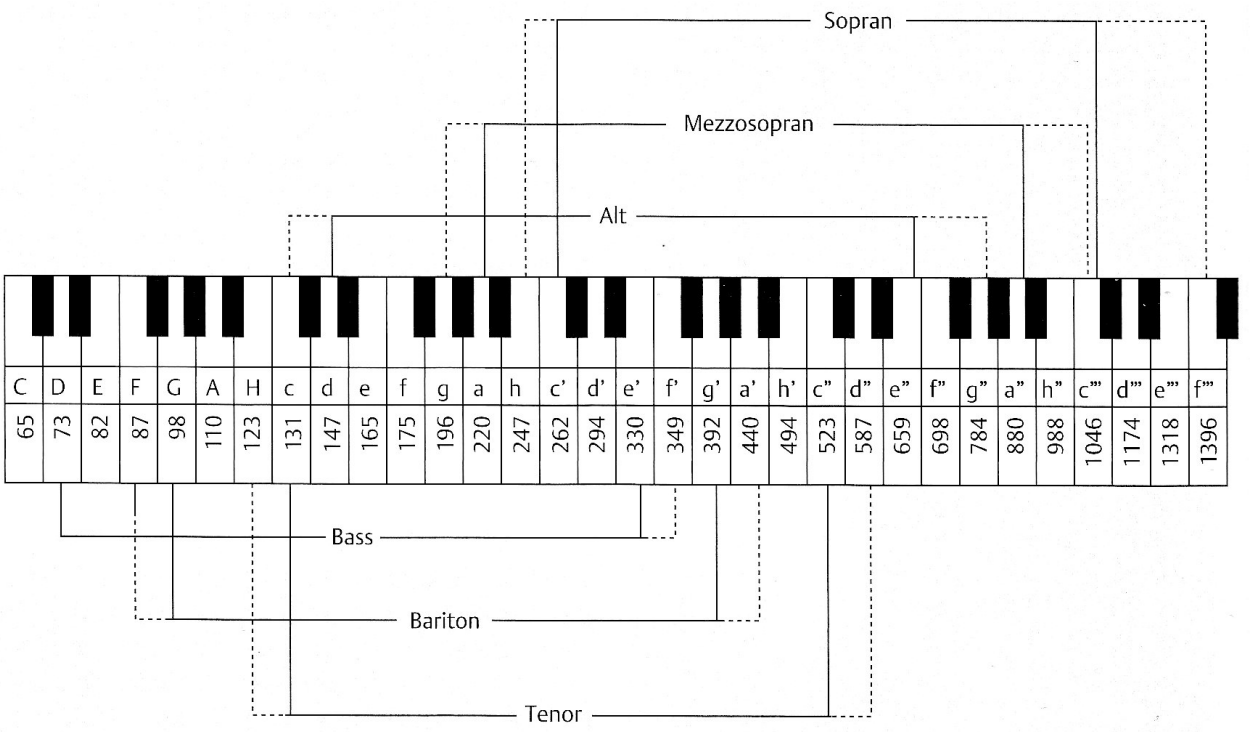


Abbildung 4.16: Aufteilung der Stimmgattungen nach Noten bzw. Frequenzen (aus Seidner & Wendler, 2006, S. 98)

diese Abweichungen vielleicht sogar normal für die menschliche Stimme sind.

Davon ausgehend, dass das hier analysierte Lied auf den Ambitus von Kowalski abgestimmt ist, wird mit den Abbildungen 4.15a & b ebenfalls der Ambitus oder sogar eine für Kowalski angenehm zu singende Tonhöhe dargestellt. Anhand der F_0-Grenzen in den Abbildungen 4.15a, b oder c und der in Abbildung 4.16 dargestellten Aufteilung der Stimmgattungen kann man nun Kowalskis Stimmgattung bestimmen. Für Kowalski wurden die gesungenen Grundfrequenzen von 220 Hz (Note A_3) bis 740 Hz ($F_5^{\#}$) festgestellt, weshalb man ihn in den hohen Alt- oder den niedrigen Mezzosopran-Bereich einordnen könnte. Da hier nur ein einziges Lied und nicht die gesamte Diskographie untersucht wurde, kann an dieser Stelle nicht gesagt werden, ob dies dem gesamten Ambitus Kowalskis entspricht. Wenn man das spektrale Gefälle in Abbildung 4.15a klanglich interpretiert, so tendiert seine Stimme für die höheren Intensitäten und Tonhöhen in Richtung „flötenartig“ (siehe Abschnitt 2.3) mit stärkerem Gefälle. Auch hier sind weitere Analysen notwendig, um dies zu bestätigen.

Formantentuning

Aus fünf Liedern einer kommerziellen CD konnte nur ausreichend Datenmaterial für die drei Vokale /e/ (14.7 s), /i/ (27.7 s) und /o/ (25.7 s) gewonnen werden, für die restlichen Vokale war die Methodik durch die Beschränkung der Perzentilen (vgl. Abschnitt 3.6) zu streng. Die Abbildungen 4.17a, b zeigen bei Kowalski gute Übereinstimmungen des ersten Formanten mit der ersten Harmonischen im gesamten Tonhöhenbereich (ca. 280 Hz bis 540 Hz bzw. bis

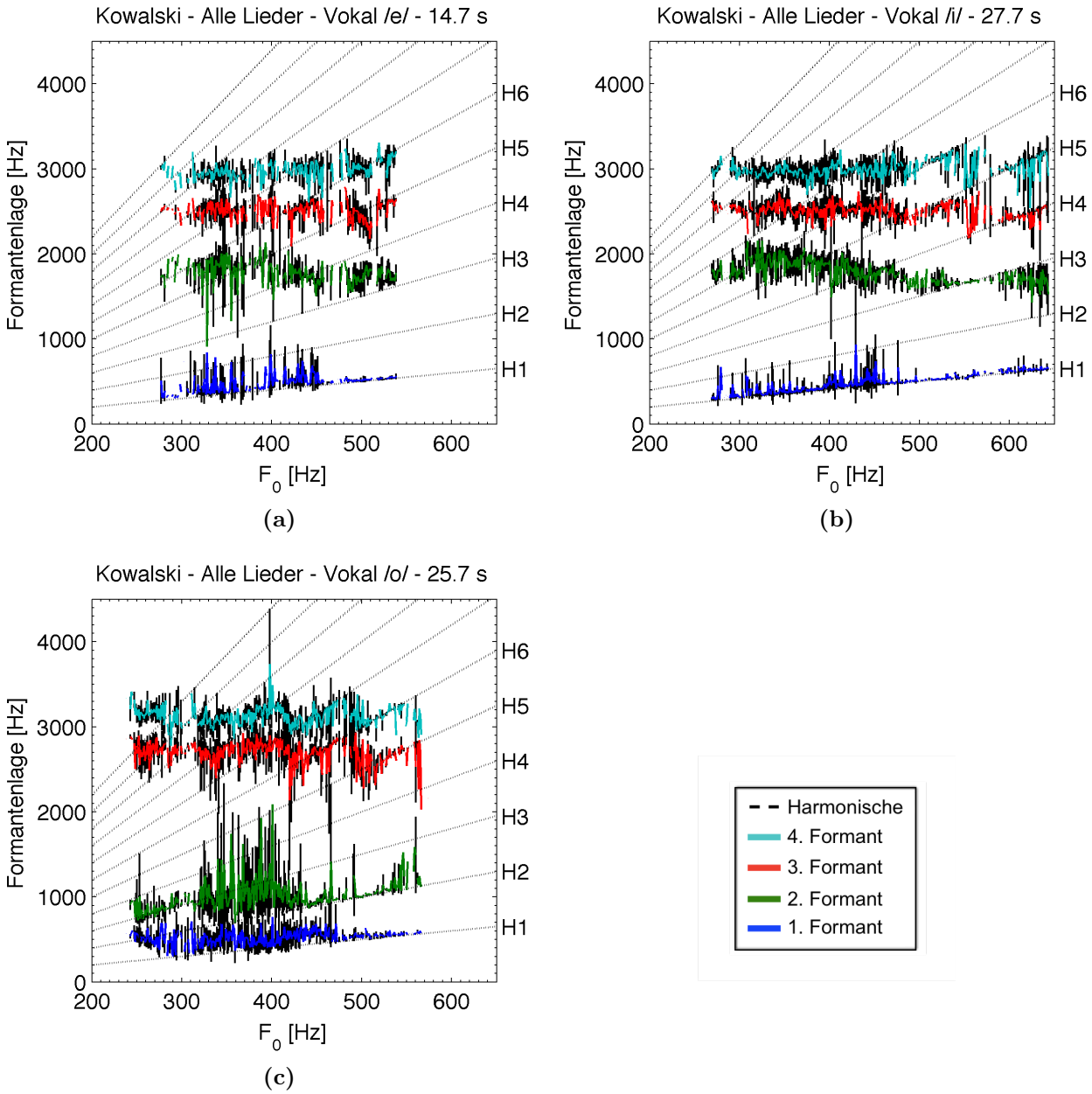


Abbildung 4.17: Mit LPC (vgl. Abschnitt 3.6) ermittelte Formantenlagen der ersten 4 Formanten in Abhängigkeit von der Tonhöhe. a) und b) zeigen für den ersten Formanten sehr gute Übereinstimmungen mit dem ersten Harmonischen. In c) kann dies für den ersten Formanten F1 nur oberhalb der Tonhöhe von 450 Hz beobachtet werden. F2 liegt mit starken Schwankungen auf der zweiten (für F_0 ca. von 450 Hz bis 650 Hz) oder dritten (für F_0 ca. von 250 Hz bis 400 Hz) Harmonischen. In a), b) und c) liegt der dritte Formant gemittelt über alle Tonhöhen sehr gut auf den im LTAS/LTAS-T bestimmten Positionen des Sängerformanten (vgl. Tabelle 4.1) bzw. im Bereich der mittleren SPP-Positionen (vgl. Tabelle 4.2).

650 Hz). Bei Vokal /o/ (Abb. 4.17c) kann dies nicht für alle Tonhöhen festgestellt werden. Oberhalb von 450 Hz liegen die ersten zwei Formanten F1 und F2 zwar auf den Harmoni-

schen, dafür gibt es unterhalb von 450 Hz starke Abweichungen der Formantenlagen. Der Mittelwert der Formantenposition von F1 liegt etwa von 280 Hz bis 450 Hz fast ausschließlich zwischen den Harmonischen und weist relativ hohe Schwankungen von ±95 Hz auf. F2 liegt von 250 Hz bis 400 Hz im Mittel meist auf der zweiten Harmonischen, die Schwankungen liegen aber bei ±500 Hz. Auszuschließen ist Formantentuning in diesem Fall jedoch nicht, das Abweichen von den Harmonischen könnte auch durch die ungeeigneten Aufnahmebedingungen bedingt sein. Hierbei ist ebenfalls besonders der Hall zu erwähnen, welcher in kommerziellen Aufnahmen oft vorhanden ist und bei schnellen Tonhöhenänderungen starke Intensitäten bei Frequenzen verursacht, die nicht mehr von den tatsächlichen Harmonischen des Sängers zum aktuellen Zeitpunkt verursacht werden, sondern von früheren Zeitpunkten stammen. Eine weitere Beobachtung ist, dass der dritte Formant (Mittelwert für die Vokale: /e/ 2510 Hz, /i/ 2500 Hz und /o/ 2690 Hz) meist den ermittelten Sängerformanten (vgl. Tabelle 4.1) bzw. den mittleren SPP-Positionen (vgl. Tabelle 4.2) entspricht. Dies spricht für die Konsistenz der drei unterschiedlichen Methoden (LPC-Analyse, LTAS/LTAS-T und SPP-Positionsbestimmung).

4.2.2 Spektralanalyse von Russell Oberlin

Als zweites Beispiel für die entwickelten Methoden (vgl. Kapitel 3) wurde der Haute Contre (siehe Abschnitt 1.1) Russell Oberlin (vgl. Abschnitt 1.3 und 1.3.1) ausgewählt. Für alle Methoden mit Ausnahme des LPC-Verfahrens zur Analyse auf Formantentuning wurden aus einem Lied ca. 62 s den Auswahlkriterien (vgl. Abschnitt 3.2) entsprechende Ausschnitte verwendet. Von den 62 s sind etwa 23 s ohne Instrumentalbegleitung, die Gesamtdauern für jeden Vokal sind gerundet: /a/ 14 s, /e/ 26 s, /i/ 5 s und /o/ 17 s. Für den Vokal /u/ gab es in diesem Lied keine den Auswahlkriterien entsprechenden Passagen. In den restlichen zur Verfügung stehenden Liedern wurde jedoch keines mit besseren Voraussetzungen zur Analyse gefunden.

Spektralcharakteristika im Langzeitspektrum

Abbildung 4.18 zeigt das LTAS und das LTAS-T aller Vokale für dieses Lied. Tabelle 4.3 fasst die Werte der Positionen und Breiten des Sängerformanten nach Abschnitt 3.5.2 für alle Vokale in Zahlen. Auch hier zeigen sich im LTAS spitzer zulaufende Maxima im Frequenzbereich bis ca. 1200 Hz. Von dort ist der Intensitätsverlauf für das LTAS und das LTAS-T beinahe gleich. Dies spiegelt wiederum die reduzierte Tonhöhenabhängigkeit des LTAS-T wider. Das als Sängerformant identifizierbare Maximum von etwa 3000 Hz bis 4000 Hz ist deutlich zu erkennen. Oberhalb von 4000 Hz lassen sich keine besonderen Maxima im Intensitätsverlauf erkennen. Die Informationen über die Breiten, Positionen und relative Intensität des Sängerformanten nach Abschnitt 3.5.2 sind in Abbildung 4.19a & b dargestellt:

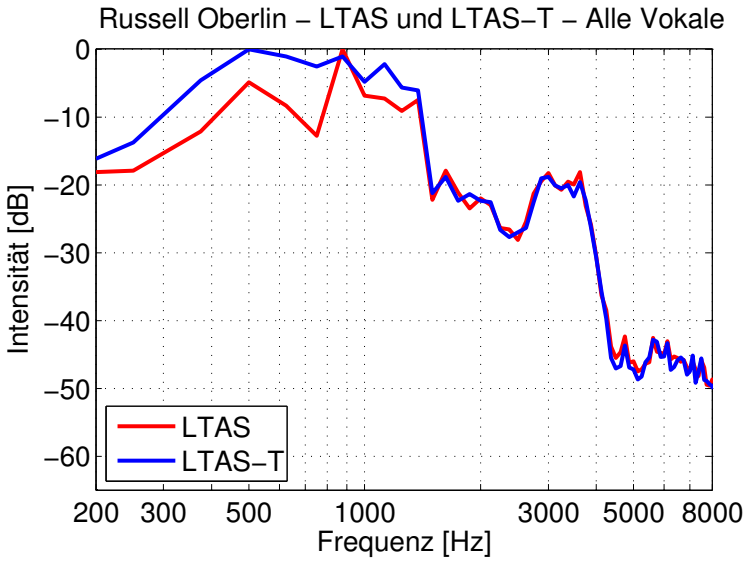


Abbildung 4.18: LTAS und LTAS-T von Russel Oberlin über alle Vokale (Gesamtdauer: ca. 62 s). Von etwa 2000 Hz bis 4000 Hz ist eine deutliche Intensitätserhöhung sowohl im LTAS, als auch im LTAS-T zu erkennen, welche dem Sängerformanten zuzuordnen ist. Die Intensität des LTAS-T ist im Frequenzbereich von etwa 200 Hz bis 1200 Hz meist höher und zeigt weniger spitze Maxima (z.B. 500 Hz, 900 Hz) wie das LTAS. Von 1200 Hz etwa verlaufen die Intensitäten für das LTAS und das LTAS-T sehr ähnlich. Im Frequenzbereich von 5000 Hz bis 8000 Hz ist keine Erhebung zu erkennen.

Tabelle 4.3: Werte der zentralen Position $p_{fs,z}$, der 3-dB-Breite $b_{fs,3db}$, der Maximumposition $p_{fs,max}$ und der relativen Intensitäten I_{fs} des Sängerformanten für Russell Oberlin für alle Vokale separat und zusammen. Die Werte sind in den Abbildungen 4.5 und 4.7a in Abschnitt 4.1 grafisch dargestellt.

Vokal	$p_{fs,z}$ [Hz]	$b_{fs,3db}$ [Hz]	$p_{fs,max}$ [Hz]	I_{fs} [dB]
/a/ LTAS	3250	1000	3375	-18.09
/a/ LTAS-T	3187.5	625	3125	-19.25
/e/ LTAS	2875	500	2750	-13.98
/e/ LTAS-T	3187.5	1125	2875	-17.59
/i/ LTAS	2875	250	2875	-16.35
/i/ LTAS-T	2125	250	2125	-22.49
/o/ LTAS	3625	250	3625	-19.80
/o/ LTAS-T	3625	250	3625	-22.45
alle LTAS	3250	1000	3625	-18.13
alle LTAS-T	3250	1000	3000	-18.85

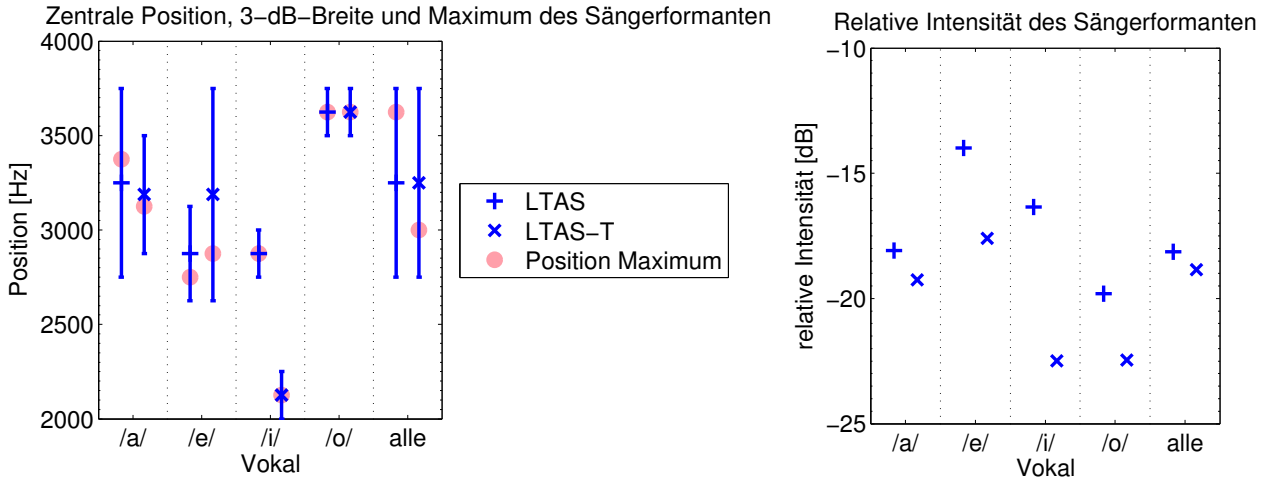


Abbildung 4.19: a) Zentrale Position (blaue Kreuze), Maximumsposition (rote Kreise) und 3-dB-Breite (senkrechte Striche) des Sängerformanten für alle Vokale. Die Abweichungen des Vokals /i/ lassen sich durch die geringe Dauer (ca. 5 s) erklären, der starke Unterschied in der Breite bei Vokal /e/ ist ähnlich zu begründen wie die Änderung der Breite des Vokals /i/ in Abbildung 4.6 (Abschnitt 4.1). Hierbei wurde die 3-dB-Schwelle abermals knapp verfehlt, was in diesem Fall mit der größeren Tonhöhenabhängigkeit des LTAS in Verbindung steht. b) Relative Intensitäten des Sängerformanten für alle Vokale. Abermals ist die Intensitätsabweichung für /i/ mit der kurzen Zeit zu begründen, während bei Vokal /e/ die größere Intensität des LTAS durch die Tonhöhenabhängigkeit verursacht wird, wenn die gesungenen Tonhöhen für Vokal /e/ meist gleich sind.

Es ergeben sich für die unterschiedlichen Vokale auch bei Russell Oberlin verschiedene Positionen des Sängerformanten (Abb. 4.19a) und verschiedene Intensitäten (Abb. 4.19a). Beim Vokal /e/ wurde die 3-dB-Breite durch eine geringere Intensitätserhebung im Vergleich zum LTAS erst bei 3750 Hz erreicht. Die Position des Vokals /i/ (ähnlich der in Abbildung 4.6, Abschnitt 4.1 erläuterten Situation durch die Instrumente) ist wegen der kurzen Zeit der analysierten Ausschnitte wenig aussagekräftig. Ob diese Intensitäten und Positionen für Oberlin typisch sind, kann an dieser Stelle nicht gesagt werden.

Singing Power Ratio

Die Lage des SPP in Abhängigkeit von der Tonhöhe (vgl. Abschnitt 3.5.3) für Russell Oberlin ist in 4.20 zu sehen. Auch in diesem Fall geben die senkrechten Striche in der Abbildung jeweils die Standardabweichung aller gemittelten Positionen in beide Frequenzrichtungen an. Die Vokale zeigen wie im soeben besprochenen LTAS und den daraus extrahierten Positionen Frequenzbereiche, in denen der jeweils entsprechende SPP tendenziell liegt. Der Vokal /o/ ist wie auch in Abbildung 4.19 am höchsten, während der Vokal /i/ die niedrigste Frequenz einnimmt, jedoch gleichzeitig wieder nur 5 s Gesamtdauer aufweist. Die mittlere Frequenzlage der SPP aller Vokale liegt mit 3032 Hz etwas tiefer als die im LTAS/LTAS-T bestimmten Positionen (3250 Hz). Es konnte nicht für jeden Vokal der gleiche Tonhöhenbereich erfasst werden. Für den Vokal /e/ konnte der gesamte Bereich von 139 Hz bis 698 Hz durch ein Solo

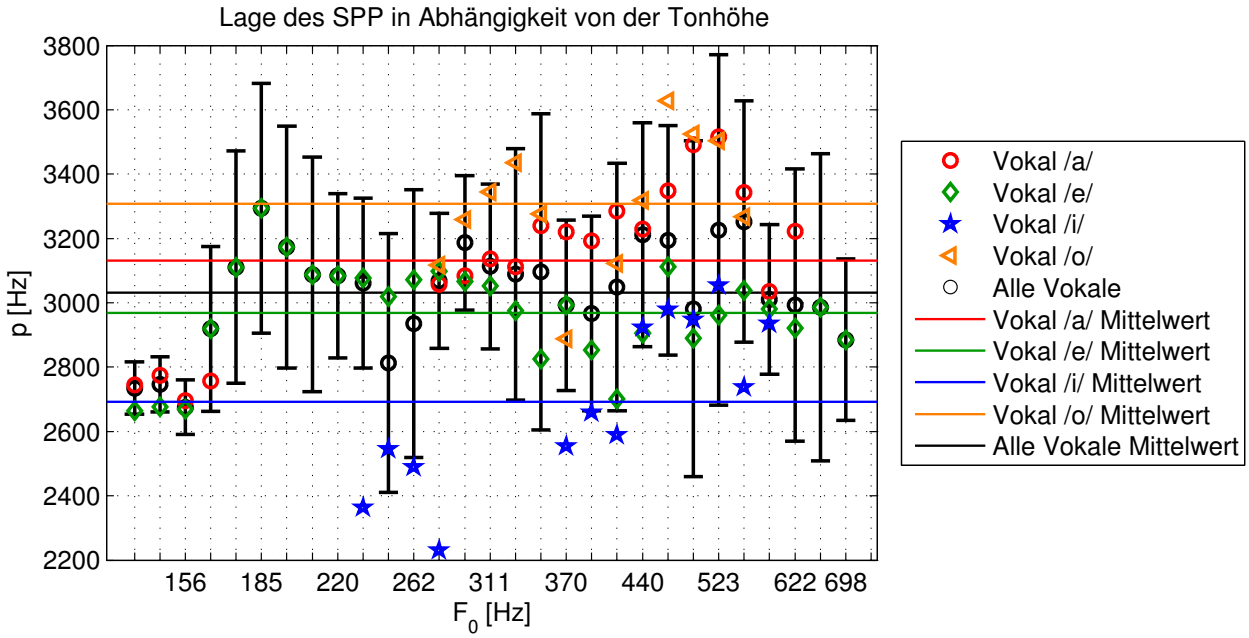


Abbildung 4.20: Berechnete SPP-Positionen (siehe Abschnitt 3.5.3) für jeden Vokal und jede Tonhöhe. Für die SPP-Position über „alle Vokale“ sind die senkrechten Balken mit ±1 Standardabweichung eingefügt. Wie auch in der Bestimmung der zentralen Position bzw. der Position des Maximums des Sängerformanten, zeichnet sich eine eindeutige Verteilung der einzelnen Vokale über den Frequenzbereich ab (vgl. Abb. 4.19). Die Verteilung des Vokals /i/ ist durch die geringe Dauer von 5 s nicht ausreichend bewertbar. Für Vokal /e/ konnten die meisten Tonhöhen extrahiert werden.

des Sängers erfasst werden.

Der nächste Schritt ist die Berechnung der SPR in Abhängigkeit von der Tonhöhe (vgl. Abb. 4.21). Analog 4.20 sind nicht alle Vokale für alle Tonhöhen vorhanden. Für den Vokal /o/ nimmt die SPR (ausgehend von den niedrigen Tonhöhen) zunächst zu und weist etwa im Tonhöhenbereich von 262 Hz bis 311 Hz ein Maximum auf, bevor sie wieder abnimmt. Der Vokal /i/ folgt möglicherweise diesem SPR-Verlauf, für die Vokale /a/ und /o/ ist dies nicht zu erkennen. Der SPR-Verlauf des Vokals /a/ über die Tonhöhe weist ein kleines Maximum zwei Halbtöne[30] oberhalb der 311 Hz auf. In Tabelle 4.4 sind die mittleren SPR-Werte und SPP-Positionen aufgeführt:

[30] Jeder Schritt auf der x-Achse stellt einen Halbton dar (vgl. Abschnitt 3.4.1)

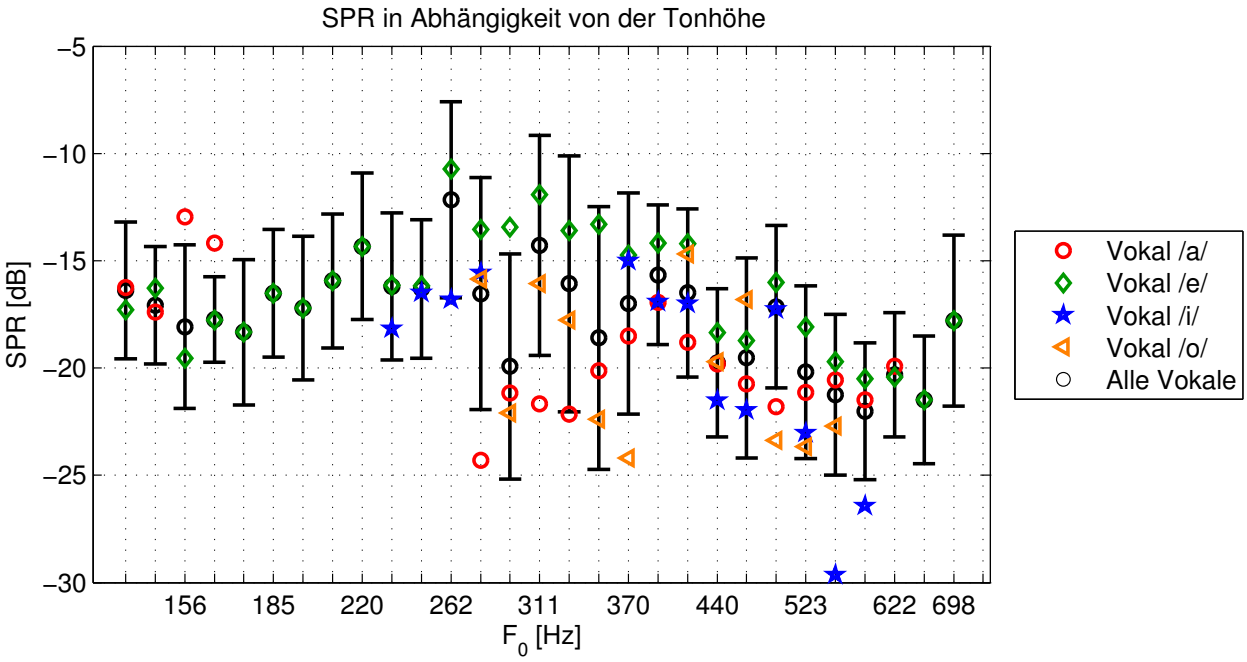


Abbildung 4.21: Die SPR in Abhängigkeit von der Tonhöhe. Nicht alle Vokale wurden im Lied für alle Tonhöhen gefunden. Vokal /e/ ist durch ein Solo des Sängers über den gesamten Tonhöhenbereich erfasst worden. Die SPR für Vokal /e/ hat ein Maximum zwischen 262 Hz und 311 Hz. Vokal /a/ könnte ein Maximum zwischen 370 Hz und 440 Hz haben, es könnte sich aber auch um eine zufällige Verteilung handeln.

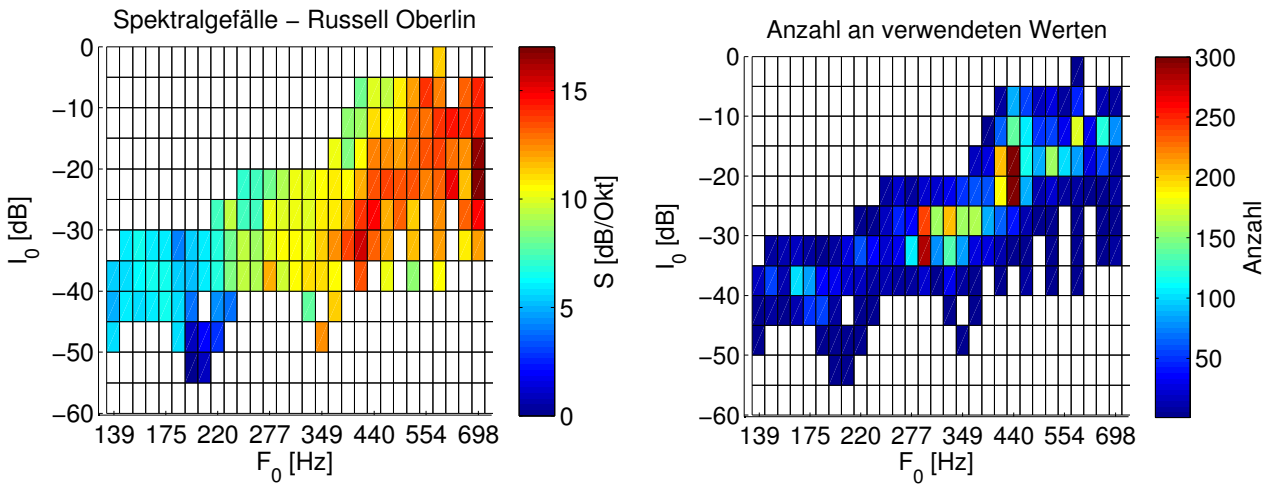
Tabelle 4.4: Mittelwerte der bestimmten SPP-Positionen $\bar{p}$ und SPR-Werte $\overline{SPR}$ sowie deren Standardabweichung σ für alle Vokale.

	$\bar{p}$ [Hz]	σ_p [Hz]	$\overline{SPR}$ [dB]	σ_{SPR} [dB]
Vokal /a/	3131	208	−19.47	3.14
Vokal /e/	2969	335	−16.42	3.66
Vokal /i/	2693	351	−19.68	3.32
Vokal /o/	3307	225	−19.94	3.45
Alle Vokale	3032	345	−17.60	3.99

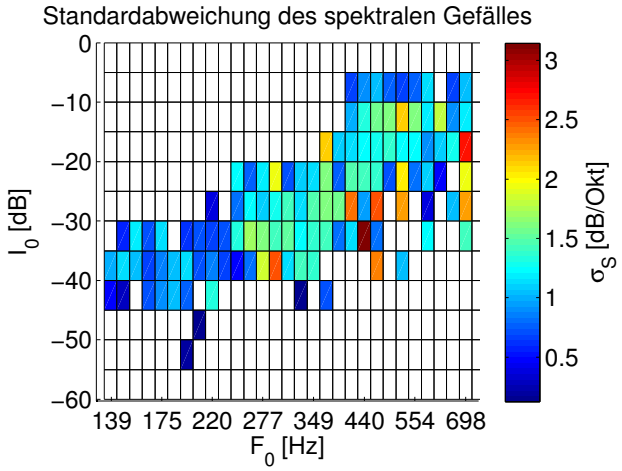
Für den Vokal /o/ und die Mittelung über alle Vokale zeigt sich eine ähnliche Intensität wie die der relativen Intensitäten des Sängerformanten im LTAS, für den Vokal /a/ ist die SPR näher am Intensitätswert aus dem LTAS-T (vlg. 4.4 und 4.3).

Spektrales Gefälle und Einteilung in die Stimmgattung

Die Betrachtung des spektralen Gefälles (vgl. Abb. 4.22a) für Russell Oberlin lässt direkt eine starke Abhängigkeit des Gefälles von der Tonhöhe und der Intensität des ersten Obertons erkennen. Unterhalb von 220 Hz kann die Stimme somit als „metallisch“ beschrieben werden (vgl. Abschnitt 2.3), und ab etwa 400 Hz aufwärts als eher „flötenartig“. Für höhere Intensitäten beträgt das Gefälle der Stimme bis etwa drei Halbtöne oberhalb von 440 Hz stellenweise (gelbe und grüne Bereiche) weniger als 12 dB/Okt. Abbildung 4.22b zeigt die



(a) (b)



(c)

Abbildung 4.22: a) Spektrales Gefälle für alle Tonhöhen (nach Abschnitt 3.4), b) die Anzahl der dafür verwendeten Gefälle-Werte und c) die sich durch die Mittelung ergebenen Standardabweichungen. In a) zeichnet sich ein eindeutiger Verlauf des Gefälles in Abhängigkeit von der Tonhöhe ab: Unterhalb von 220 Hz ist das Gefälle unter 7 dB/Okt, zwischen 220 Hz und 349 Hz etwa bei 12 dB/Okt und oberhalb von 440 Hz zwischen 12 dB/Okt und 17 dB/Okt. Besonders für die hohen Intensitäten im Frequenzbereich von etwa 400 Hz bis 500 Hz ist das Gefälle jedoch etwas geringer. Die Standardabweichung in b) liegt größtenteils unter 2 dB/Okt (grün und blau), was besonders in den Bereichen mit einer großer Anzahl an Gefälle-Werten (vgl. b) einen geringeren Fehler des mittleren Gefälles in c) bedeutet.

Anzahl an Gefälle-Werten für die Berechnung des Gefälles in Abbildung 4.22a. Hier ist klar zu sehen, dass die meisten Harmonischen zwischen ca. 250 Hz und 500 Hz extrahiert wurden. Die Standardabweichung (Abb. 4.22c) bewegt sich größtenteils unter 2 dB/Okt. Teilweise fällt sie auch deutlich kleiner aus (dunkelblaue Bereiche), was aber aufgrund der oftmals geringen Anzahl an Gefälle-Werten (Abb. 4.22a) wenig aussagekräftig ist. Mit den Grenzen

der extrahierten Tonhöhen in den Abbildungen 4.22a & b sowie Abbildung 4.16 in Abschnitt 4.2.1 lässt sich Russell Oberlin in die Stimmgattung des hohen Alt einordnen. Der tiefste aus dem Lied extrahierte Ton liegt bei 139 Hz ($C_2^{\#}$), der höchste bei 698 Hz (F_5). Auch hier muss erwähnt werden, dass nur ein einziges Lied und nicht die gesamte Diskographie Oberlins untersucht wurde.

Formantentuning

Da das LPC-Verfahren nur auf reine Gesangspassagen angewendet werden kann, wurden drei Lieder verwendet, um ausreichend Daten zu erhalten. Dabei konnten die Vokale /a/ (Dauer: 11.5 s), /e/ (Dauer: 12.2 s), /i/ (Dauer: 13.3 s) und /o/ (Dauer: 13.1 s) analysiert werden (vgl. Abb. 4.23). Für den Vokal /e/ (Abb. 4.23a) ist kein Formantentuning erkennbar. Dies liegt möglicherweise daran, dass nahezu die gesamten 12.2 s aus einem virtuos gesungenen Solo stammen. Der starke Hall der CD-Aufnahme zusammen mit den schnellen Tonhöhenänderungen hat hierbei Auswirkungen auf die LPC-Analyse. Eine andere Erklärung ist, dass Oberlin bei diesem rasanten Tempo die Formanten nicht schnell genug auf die Harmonischen schieben kann. Die restlichen Abbildungen 4.23b, c, & d deuten darauf hin, dass er wahrscheinlich unter ruhigeren Gesangsbedingungen Formantentuning anwendet: In Abbildung 4.23b liegen die ersten beiden Formanten (F1 und F2) für fast alle Tonhöhen auf einer der ersten drei Harmonischen (H1, H2 und H3). Auffällig ist, dass auch F3 und F4 auf den höheren Harmonischen verlaufen. Über Formantentuning mit diesen Formanten konnten bei der Literaturrecherche zur vorliegenden Arbeit keine Informationen gefunden werden. Auch in den Abbildungen 4.23c & d verläuft der erste Formant beinahe ausnahmslos auf H2 (Abb. 4.23c) bzw. H1 (Abb. 4.23d), was vermuten lässt, dass Russel Oberlin Formantentuning einsetzt.

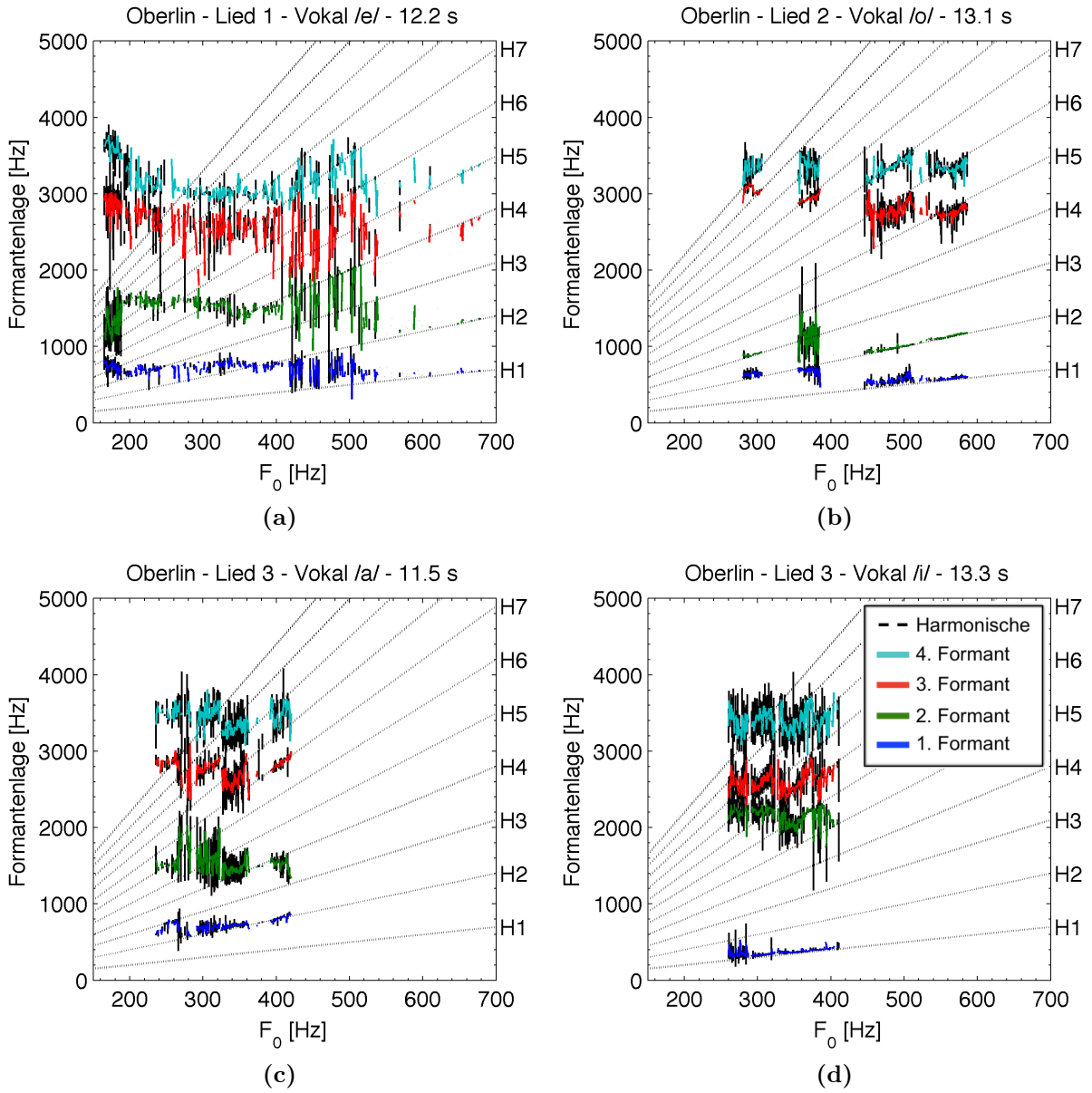


Abbildung 4.23: Mit der LPC-Methode (vgl. Abschnitt 3.6) ermittelte Formantenlagen der ersten vier Formanten in Abhängigkeit von der Tonhöhe. a) Die Positionswerte der Formanten entstammen einem Solo, bei dem Russel Oberlin den gesamten Tonhöhenbereich von ca. 160 Hz bis 700 Hz abdeckt. Es zeigt sich keine Übereinstimmung der ersten beiden Formanten mit den Harmonischen. b) Die ersten zwei Formanten F1 und F2 liegen auf den Harmonischen H1, H2 oder auch H3. Auffällig ist, dass auch die Formanten F3 und F4 in fast allen Tonhöhen direkt auf den Harmonischen liegen. c) Der erste Formant liegt von ca. 300 Hz bis 400 Hz auf der zweiten Harmonischen. d) Von 250 Hz bis 400 Hz liegt F1 sehr genau auf H1.

5 Diskussion und Ausblick

Die in der vorliegenden Arbeit ausgewählte und teilweise selbst entwickelte Kombination von Methoden (vgl. Kapitel 3) wurde zusammengestellt, um möglichst viele Eigenschaften einer Gesangsstimme physikalisch erfassen und darstellen zu können. Es konnte in der Literatur keine *einzelne* Methode gefunden werden, mit der Sänger der gleichen Population (vgl. Kapitel 1.3.3) und Stimmgattung unterscheidbar sind. Da im Rahmen der Masterarbeit lediglich kommerzielle CD-Aufnahmen von Sängern verfügbar waren, welche meist Instrumentalbegleitung beinhalten, wurde die sogenannte Harmonischenextraktion (vgl. Abschnitt 3.4) entwickelt, welche mit Ausnahme des LTAS und des LPC als Basis für alle der im Folgenden diskutierten Methoden dient und in begrenztem Maß die Extraktion der Harmonischen aus instrumental begleiteten Aufnahmen ermöglicht.

Das LTAS (z.B. Seidner et al., 1983, Sundberg, 2001, Kovacic & Boersma, 2003) bzw. LTAS-T (Boersma & Kovacic, 2005) ist eine sehr weit verbreitete Methode zur Analyse von Gesangsstimmen. Damit können spektrale Charakteristika wie der Sängerformant (Sundberg, 1974) und auch der Stimmklang mittels spektraler Steigung bestimmt und beurteilt werden. In Seidner et al. (1983) und Sundberg (2001) wurden unterschiedliche Stimmen untersucht und konnten anhand der Position des Sängerformanten in Stimmgattungen eingeordnet werden. Für die Unterscheidung verschiedener Sänger derselben Stimmgattung ist dies möglicherweise nicht ausreichend, weshalb die weiteren Methoden in der vorliegenden Arbeit ausgewählt wurden. Beim direkten Vergleich der Langzeitspektren von Kowalski (vgl. Abb. 4.18) und Oberlin (vgl. Abb. 4.12) ist zu erkennen, dass sich bei Oberlin ein vom Rest des Spektrums deutlicher abgehobener Sängerformant von ca. 3000 Hz bis 4000 Hz abzeichnet und dieser für das LTAS und LTAS-T nahezu identisch ist. Die unterschiedliche Ausprägung des Sängerformanten bei Kowalski und Oberlin könnte durch die unterschiedlichen Gesamtdauern der einzelnen Vokale erklärbar sein. Für den Vokal /i/ (bei Kowalski 25 s, bei Oberlin nur 5 s) liegt nach Abbildung 2.5 in Abschnitt 2.3 der zweite Formant zwischen 2000 Hz und 4000 Hz, was somit die Intensitäten dieser Frequenzbereiche im LTAS/LTAS-T deutlich anhebt. Ab 4000 Hz sind im Spektrum von Oberlin (vgl. Abb. 4.12) keine weiteren Auffälligkeiten zu erkennen, während sich im Spektrum von Kowalski (vgl. Abb. 4.18) bei 6250 Hz ein weiterer Formant (Erhöhung der Intensität) abzeichnet. Ein möglicher Grund hierfür sind Unterschiede in den Vokaltrakten. Die zwei in der vorliegenden Arbeit untersuchten Beispiele reichen allerdings nicht aus, um hierüber ein abschließendes Urteil zu fällen. Bei der Betrachtung der Sängerformantenpositionen (Abb. 4.5 bzw. Tabelle 4.1 bei Kowalski bzw. Abb. 4.19a und Tabelle 4.3 bei Oberlin) liegt Oberlins Sängerformant mit 3250 Hz um 600 Hz höher als der von Kowalski (2687.5 Hz). Im Vergleich mit den Erkenntnissen in Sundberg (2001) ist Oberlin demnach im Alt und Kowalski eher im Tenor einzuordnen. Nach den Ergebnissen in Seidner et al. (1983) ist aber auch Kowalski im Alt einzuordnen, was besser mit dem für Kowalski gefundenen Ambitus (A_3 mit 220 Hz bis $F_5^{\#}$ mit 740 Hz) übereinstimmt. Zudem deckt sich diese Beobachtung besser mit dem in Abschnitt 1.3.1 erwähnten Ambitus nach Herr (2013, S. 472 ff.). Die einzelnen Vokale zeigen bei beiden Sängern unterschiedliche Positionen, wobei

diese bei Oberlin stärker variieren. Von dieser Stichprobe allein kann allerdings nicht abgeleitet werden, ob dies typisch für Oberlin ist. Dass es tatsächlich Unterschiede für die einzelnen Vokale gibt, steht in Übereinstimmung mit den Ergebnissen von Seidner et al. (1983). Die relative Intensität des Sängerformanten ist bei Kowalski im Mittel mit −13.01 dB (LTAS) und −16.89 dB (LTAS-T) größer als bei Oberlin mit −18.13 dB (LTAS) und −18.85 dB (LTAS-T). Nach Seidner et al. (1983) könnten auch anhand dieses Kriteriums beide in die Stimmgattung Alt eingeordnet werden.

Wie in Abschnitt 4.1 gezeigt werden konnte, waren durch die Instrumentaleinflüsse die Positionen und die Breite des Sängerformanten mit Ausnahme des Vokals /i/ gering. Für diesen Vokal haben sich die zentrale Position und die 3-dB-Breite deutlich verändert (vgl. Abb. 4.5). Es ist anzunehmen, dass eine größere dB-Schwelle in diesem Fall einheitlichere Ergebnisse erzielt hätte. Eine 10-dB-Schwelle wie bei Kovacic et al. (2003) oder eine 15-dB-Schwelle wie bei Seidner et al. (1983) hätte dann allerdings im LTAS/LTAS-T über alle Vokale im Fall von Kowalski (vgl. 4.12) und Oberlin (vgl. 4.18) kein Ergebnis erbracht, da der Frequenzbereich unterhalb des Maximums des Sängerformanten vollständig oberhalb der 15-dB-Schwelle liegt. Größere dB-Schwellen sind somit besser für die gezielte Betrachtung einzelner Vokale geeignet. Die relativen Intensitäten des Sängerformanten änderten sich in der instrumental begleiteten Version (Gesang und Instrumente) durch den Einfluss der Instrumente (vgl. Abb. 4.7a, b) nur gering. Es soll nochmals erwähnt werden, dass andere Instrumentenzusammensetzungen oder andere Lautstärken mit Sicherheit andere Einflüsse auf das Langzeitspektrum haben. Sofern die zu analysierenden Passagen den Auswahlkriterien (vgl. Abschnitt3.2) entsprechen, bietet das LTAS/LTAS-T eine Möglichkeit zur groben Einteilung in eine Stimmgattung anhand der entsprechenden Referenzquellen (Sundberg, 2001, bzw. Seidner et al., 1983).

Die bei Nordenberg & Sundberg (2004) beschriebene Abhängigkeit des Langzeitspektrums von der gesungenen Lautstärke wird durch die Methode des spektralen Gefälles in Abhängigkeit von der Tonhöhe F_0 und der Intensität des Grundtons I_0 (vgl. Abschnitt 3.4.1) berücksichtigt. Die Methode zeigt bei Jochen Kowalski (vgl. Abb. 4.15a) und Russell Oberlin (vgl. 4.22a) einen deutlichen Zusammenhang zwischen Tonhöhe, Intensität des Grundtons und spektralem Gefälle. Bei beiden nimmt das Gefälle mit zunehmender Tonhöhe zu. Für Russell Oberlin ist zu erkennen, dass eine höhere gesungene Intensität das Gefälle reduziert. In beiden Fällen konnte allerdings für die verschiedenen Tonhöhen und Intensitäten keine einheitliche Anzahl an Gefälle-Werten bestimmt werden (vgl. 4.15b bzw. 4.22b), sodass auch mit zufälligen Streuungen der Gefälle-Werte gerechnet werden muss. Der Einfluss der Instrumente auf das spektrale Gefälle (vgl. Abb. 4.11a, b, c in Abschnitt 4.1) ist für die niedrigen Tonhöhen (unterhalb von 500 Hz) und für geringere Intensitäten des ersten Obertons (etwa unter −25 dB) besonders deutlich zu erkennen. Mit Abbildung 4.2 lässt sich die Verringerung des spektralen Gefälles anschaulich erklären: Während die Harmonischen des Sängers bei etwa 400 Hz, 800 Hz und 2800 Hz bis 3200 Hz in der Mischung (Gesang und Instrumente) dominieren und somit nach wie vor von der Harmonischenextraktion korrekt erfasst werden, sind die dazwischen liegenden Harmonischen des Sängers durch die Instrumente deutlich verdeckt.

Die HE erfasst in der Mischung trotzdem an den erwarteten Frequenzen der Obertöne (z.B. 1200 Hz, 1600 Hz, usw.) des Sängers die größte Intensität, wodurch sich das berechnete Gefälle verringert. Sollte die Methode der Harmonischenextraktion in einer Folgearbeit verwendet werden, so wäre eine Funktion zum Ausschluss bestimmter Frequenzbereiche (in denen die Instrumente zu hohe Intensitäten aufweisen) eine mögliche Erweiterung. Außerdem könnte dann verglichen werden, ob das spektrale Gefälle von Kowalski und Oberlin in verschiedenen Aufnahmen den in der vorliegenden Arbeit gefundenen Ergebnissen entspricht und ob sich Unterschiede innerhalb der Countertenöre oder zwischen verschiedenen Sängerpopulationen (z.B. Countertenöre, chirurgische / hormonelle / genetische Kastraten, Frauenstimmen) zeigen. Solche Darstellungen konnten während der Recherchen zur vorliegenden Arbeit nicht in der Literatur gefunden werden.

Die von Omori et al. (1996) beschriebene Singing Power Ratio ist nach den dortigen Angaben ein Maß für die Qualität der Singstimme. In dieser Studie waren die Tonhöhe und Intensität jedem Sänger selbst überlassen, was bei kommerziellen CD-Aufnahmen nicht der Fall ist. Daher wurde die SPR in der vorliegenden Arbeit in Abhängigkeit von der Tonhöhe berechnet. Bei der Analyse von Kowalski (vgl. Abschnitt 4.2.1, Abb. 4.14) zeigte sich durchaus eine Abhängigkeit der SPR aller Vokale von der Tonhöhe (ausgenommen /u/, von welchem nur wenige Sekunden analysiert werden konnten) mit einem Maximum der SPR zwischen etwa 370 Hz und 440 Hz. Bei Russel Oberlin liegt für den Vokal /e/ ein schwach ausgeprägtes Maximum der SPR zwischen 262 Hz und 311 Hz (vgl. Abschnitt 4.2.2, Abb. 4.21). Für eine eindeutige Aussage sind jedoch weitere Auswertungen notwendig. Bei den Positionen des Singing Power Peak in Abhängigkeit von der Tonhöhe zeigte sich für beide Beispiele (Jochen Kowalski, vgl. Abb. 4.13 und Russel Oberlin, vgl. Abb. 4.20) keine erkennbare Tonhöhenabhängigkeit. Dafür ergeben sich für beide Sänger deutliche Unterschiede in der SPP-Position (für Oberlin im Mittel 3032 Hz, für Kowalski 2745 Hz), die tendenziell mit den Positionen des Sängerformanten aus dem LTAS/LTAS-T vergleichbar sind. Omori et al. (1996) konnten eine statistisch signifikante Unterscheidung verschiedener Sängergattungen durch die SPP-Position aufzeigen. Dort wurde der Vokal /a/ verwendet. Ein Vergleich der hier bestimmten Positionen des SPP des Vokals /a/ mit den mittleren Werten in Omori et al. (1996) führt zu einer Einordnung von Kowalski als Mezzosopran und von Oberlin als Sopran. Dies stimmt mit den Erkenntnissen von Seidner et al. (1983) und der einfachen Einteilung anhand der registrierten Tonhöhen nicht überein. Bei Omori et al. (1996) zeigt sich für der Lage des SPP eine statistisch signifikant höhere Frequenz für Sängerinnen im Sopranbereich gegenüber Sängerinnen aus dem Mezzosopranbereich. Für die in dieser Studie untersuchten Bariton- und Tenorsänger ist der Mittelwert der SPP-Position jedoch fast gleich. Innerhalb jeder einzelnen Stimmgattung liegen die Extremwerte der SPP-Positionen um bis zu 1000 Hz auseinander, daher ist die SPP-Position kein präzises Maß für die Stimmkategorisierung. Sowohl die Singing Power Ratio als auch die Position des SPP wiesen unter Instrumentaleinfluss (vgl. Abschnitt 4.1) Unterschiede zwischen der Version mit reinem Gesang und der mit Instrumentalbegleitung auf (4.9a, b, c). Besonders die Werte der Vokale /a/ und /o/ zeigen

bei Tonhöhen unter 440 Hz die größten Abweichungen (vgl. Abb. 4.10a, b). Ob das ein sängerspezifisches Phänomen ist, kann durch dieses einzelne Beispiel nicht gesagt werden und muss gegebenenfalls in einer Folgearbeit untersucht werden. Sollte die Abhängigkeit des Maximums der SPR von der Tonhöhe für verschiedene Sänger tatsächlich unterschiedlich sein, so könnte dies, sofern die Ergebnisse bei Omori et al. (1996) aussagekräftig sind, ein Hinweis auf die (klanglich) optimale Tonhöhe eines Sängers sein. Wie in Abschnitt 1.3 angesprochen, weist nach Herr (2013, S. 475) jeder Sänger in seinem Ambitus bestimmte Bereiche auf, in denen seine Stimme mehr bzw. weniger klangliches Volumen erreicht. Erkenntnisse zukünftiger SPR-Analysen nach der in der vorliegenden Arbeit angewendeten Methode könnten mit den Meinungen fachkundiger Musikwissenschaftler verglichen werden, um die Aussagekraft der SPR weiter zu verdeutlichen.

Das nach Boersma & Kovacic (2006) angewendete LPC-Analyseverfahren zur Bestimmung von Formantentuning (vgl. Abschnitt 3.6) konnte in keiner anderen Quelle gefunden werden. Zwar werden bei Fuchs et al. (2000) auch verschiedene Sänger (sowohl kommerzielle Aufnahmen als auch Aufnahmen im phoniatrischen Labor) auf Formantentuning hin untersucht. Eine genaue Beschreibung der Methode wird jedoch nicht angegeben. Für die zwei in der vorliegenden Arbeit untersuchten Beispiele (vgl. Abschnitt 4.2.1, Abb. 4.17a, b, c und Abschnitt 4.2.2, Abb. 4.23a, b, c, d) konnte das Formantentuning gut nachgewiesen werden. Wie bereits angesprochen, ist der starke Hall in den Aufnahmen jedoch eine mögliche Fehlerquelle, die bei kommerziellen Aufnahmen nicht umgangen werden kann.

Ein letzter Punkt, der in der vorliegenden Arbeit nicht beantwortet werden kann, ist die Frage, inwiefern kommerzielle CD-Aufnahmen mit denen für Studienzwecke im Tonlabor entstandenen Aufnahmen vergleichbar sind. Hier gibt es eine enorme Vielzahl an Parametern (Raumakustik, Mikrofoncharakteristik, Mikrofonabstand zum Sänger, Ausrichtung des Sängers zum Mikrofon, Abmischung durch den Toningenieur, ...), von denen alle einen Einfluss auf die endgültige Aufnahme ausüben. Für die Beantwortung dieser Frage müsste der Gesang des Sängers einer kommerziell erhältlichen CD unter kontrollierten Bedingungen in einem Tonstudio aufgenommen und mit der CD verglichen werden.

Abschließend ergeben sich somit viele neue Ansatzpunkte für zukünftige Arbeiten, in denen die hier beschriebene Methodenkombination auf die verschiedenen Sängerpopulationen (vgl. Abschnitt 1.3.3) angewendet werden kann.

Literatur

1. Ardran, G. M., David, W.
   The alto or countertenor voice. *Music and Letters*, 1967;48(1):17-22.

2. Barbier, P.
   Farinelli. Der Kastrat der Könige. Die Biographie. ECON Verlag, Düsseldorf, 1995.

3. Barbier, P.
   The world of the castrati. The history of an extraordinary operatic phenomenon. Souvenir Press, London, 1996.

4. Baum, H.
   Die Sängerkastraten der Barockzeit. Ibidem Verlag, Stuttgart, 2012.

5. Boersma, P., Kovacic, G.
   Spectral characteristics of three styles of croatian folk singing. *The Journal of the Acoustical Society of America, 2006;119(3):1805-1816.*

6. Borch, D. Z., Sundberg, J.
   Spectral distribution of solo voice and accompaniment in pop music. *Logopedics Phonatrics Vocology*, 2002;43(1):31-35.

7. Clapton, N.
   Moreschi. The last castrato. Publishing Limited, London, 2004.

8. Cooley, J. W., Tukey, J. W.
   An algorithm for the machine calculation of complex Fourier series. *Mathematics of Computation, 1965;19(90):297-301.*

9. Deme, A.
   Formant strategies of professional female singers at high fundamental frequencies. *ISSP Cologne*, 2014.

10. Depalle, P., Garcia, G., Rodet, X.
    The recreation of a castrato voice, Farinelli's voice. *IEEE ASSP Workshop on Applications of Signal Processing to Audio and Acoustics*, 1995:242-245.

11. Elektronik-Kompendium.de
    `http://www.elektronik-kompendium.de/sites/kom/0211193.htm`, Website, undatiert, zuletzt aufgerufen am: 15.07.2014.

12. Enden, A. W. M. van den, & Verhoeckx, N. A. M.
    Digitale Signalverarbeitung. Vieweg Verlag, Braunschweig / Wiesbaden, 1990.

13. Fant, C. G. M.
On the predictability of formant levels and spectrum envelopes from formant frequencies. *For Roman Jakobson, 1956:109-120.*

14. Fant, C. G. M.
Acoustic theory of speech production: with calculations based on X-ray studies of Russian articulations. Walter de Gruyter, Paris, 2. Aufl., 1970.

15. Fritz, H.
Kastratengesang. Hormonelle, konstitutionelle und pädagogische Aspekte. Schneider Verlag, Tutzing, 1994.

16. Fuchs, M., Kleinke, A., Behrendt, W., Geissner, H.
Computergestützte Analyse einer Kastratenstimme im Vergleich zum männlichen Altus, zur Frauen- und Knabenstimme. In: Geissner, H. K. [Hrsg.]: *Stimmen hören. 2. Stuttgarter Stimmtage. Akademie für gesprochenes Wort.* Röhrig Universitätsverlag, St. Ingbert, 2000, S. 167-172.

17. Giles, P.
The history and technique of the counter-tenor: a study of the male high voice family. SCOLAR PRESS, Hants, 1994.

18. Hall, D. E.
Musikalische Akustik. Ein Handbuch. Schott Music Verlag, Mainz, 2008.

19. Harris, F. J.
On the use of windows for harmonic analysis with the discrete fourier transform. *Proceedings of the IEEE,* 1978;66(1):51-83.

20. Heidecker, E.
Die Geschichte der Gesangskastraten unter besonderer Berücksichtigung Farinellis. GRIN Verlag, München, 2008.

21. Herr, C.
Alfred Deller und die Grundlagen des Countergesangs im 20. Jahrhundert. In: Herr, C., Jacobshagen, A., Wessel, K.: *Der Countertenor. Die männliche Falsettstimme vom Mittelalter zur Gegenwart.* Schott Music Verlag, Mainz, 2012, S. 181-196.

22. Herr, C.
Gesang gegen die 'Ordnung der Natur'? Kastraten und Falsettisten in der Musikgeschichte. Bärenreiter Verlag, Kassel, 2013.

23. Herr, C., Jacobshagen, A., Wessel, K.
Der Countertenor. Die männliche Falsettstimme vom Mittelalter zur Gegenwart. Schott Music Verlag, Mainz, 2012.

24. Hoffmann, R.
Signalanalyse und -erkennung. Springer Verlag, Heidelberg, 1998.

25. Högberg, J.
From sagittal distance to area function and male to female scaling of the vocal tract. *Speech Transmission Laboratory, Quarterly Progress and Status Report*, 1995;36(4):11-54.

26. Hollien, H.
Vocal fold dynamics for frequency change. *Journal of Voice*, 2013;28(4):305-405.

27. Kallmann, F. J., Schönfeld, W. A., Barrera, S. E.
The genetic aspects of primary eunuchoidism. *American Journal of Mental Deficiency*, 1944;48(3):203-236.

28. Kesting, J.
Die Stimme als Kunstwerk. In: Barbier, P.: *Farinelli. Der Kastrat der Könige*. ECON Verlag, Düsseldorf, 1995, S. 9-16 (Einleitung).

29. Klingholz, F.
Medizinischer Leitfaden für Sänger. Online verfügbar: `http://www.xinxii.com/gratis/124789dir1328286681.pdf`, zuletzt aufgerufen am: 16.07.2014, Libri Books on Demand, Hamburg, 2000.

30. Kovacic, G., Boersma, P., Domitrovic, H.
Long-term average spectra in professional folk singing voices: A comparison from the Klapa and Dozivacki styles. *Proceedings Institute of Phonetic Sciences, Univ. of Amsterdam, 2003;25:53-64.*

31. Köwer, M.
Alessandro Moreschi. Sind seine Tonaufnahmen stellvertretend für den Kastratengesang? GRIN Verlag, München, 2007.

32. Lehmann, W., Pidoux, J.-M., Widmann, J.-J.
Larynx. Inpharmazam Medical-Pubclications, Cadempino, 1981.

33. Markel, J. E., Gray, A. H.
Linear prediction of speech. Springer Verlag, New York, 1976.

34. MonstersAndCritics.com
`http://www.monstersandcritics.com/people/Freddie-Mercury/biography/`, Website, undatiert, zuletzt aufgerufen am: 06.05.2014.

35. Meyer, M.
Signalverarbeitung. Analoge und digitale Signale, Systeme und Filter. Vieweg Verlag, Braunschweig / Wiesbaden, 2. Aufl., 2000.

36. Nadoleczny, M.
Untersuchungen über den Kunstgesang. 1. Atem-und Kehlkopfbewegungen. Springer Verlag, Berlin, 1923.

37. Nair, G.
Voice - tradition and technology. A state-of-the-art studio. Singular Publishing Group, San Diego, 1999.

38. Nawka, T., Wirth, G.
Stimmstörungen. Deutscher Ärzteverlag, Köln, 5. Aufl., 2008.

39. Nollmeyer, O.
VoxVisionEar. Stimmarbeit mit dem interaktiven Sonagramm. Ilmenau / Göttingen, 2013.

40. Nordenberg, M., Sundberg, J.
Effect on LTAS of vocal loudness variation. *Logopedics Phonatrics Vocology*, 2004;29(4):183-191.

41. Nuttall, A. H.
Some windows with very good sidelobe behavior. *IEEE Transactions on Acoustics, Speech and Signal Processing,* 1981;29(1):84-91.

42. Omori, K., Kacker, A., Carroll, L. M., Riley, W. D., Blaugrund, S. M.
Singing power ratio: quantitative evaluation of singing voice quality. *Journal of Voice*, 1996;10(3):228-235.

43. Pahn, R., Dahl, R., Pahn, E.
Beziehung zwischen Messung der stimmlichen Durchdringungsfähigkeit, Stimmstatus nach Pahn und ausgewählten Parametern des Stimmanalyseprogramms MDVP (Kay). *Folia Phoniatrica et Logopaedica*, 2001;53(6):308-316.

44. Press, W. H., Teukolsky, W. T., Vetterling, W. T., Flannery, B. P.
Numerical recipes: the art of scientific computing. Cambridge University Press, Cambridge, 3. Auflage, 2007.

45. Rabiner, L., Atal, B. S., Sambur, M.
LPC prediction error–Analysis of its variation with the position of the analysis frame. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 1977;25(5):434-442.

46. Rossing, T. D., Moore, R. F., Wheeler, P. A.
The science of sound. Pearson Education, Edinburgh, 3. Aufl., 2014.

47. Schneider-Stickler, B., Bigenzahn, W.
Stimmdiagnostik. Ein Leitfaden für die Praxis. Springer Verlag, Wien, 2. Aufl., 2013.

48. Schürenberg, B.
Leistungen des Atem- und Stimmapparates. In: Kittel, G.: *Phoniatrie und Pädaudiologie*. Deutscher Ärzteverlag, Köln, 1989, S. 21-23 (Kap. 2).

49. Schutte, H. K., Seidner, W.
Physiologische Grundlagen. In: Wendler, J., Seidner, W., Eysholdt, U.: *Lehrbuch der Phoniatrie und Pädaudiologie*. 4. Auflage, Thieme Verlag , Stuttgart, 2006, S. 71-90 (Kap. 7).

50. Seidner, W., Schutte, H., Wendler, J., Rauhut, A.
Dependence of the high singing formant on pitch and vowel in different voice types. *Proceedings of the Stockholm Music Acoustics Conference,* 1983;46(1):261-268.

51. Seidner, W., Wendler, J.
Sprech- und Singstimme. In: Wendler, J., Seidner, W., Eysholdt, U.: *Lehrbuch der Phoniatrie und Pädaudiologie*. 4. Auflage, Thieme Verlag , Stuttgart, 2006, S. 96-104 (Kap. 9).

52. Sundberg, J.
Articulatory interpretation of the singing formant. *The Journal of the Acoustical Society of America*, 1974;55(4):838-844.

53. Sundberg, J.
The acoustics of the singing voice. *Scientific American*, 1977;236(3):104-116.

54. Sundberg, J.
The science of the singing voice. Northern Illinois University Press, Illinois, 1987.

55. Sundberg, J.
Level and center frequency of the singer's formant. Journal of voice, 2001;15(2):176-186.

56. Titze, I.R.
Principles of voice production. Prentice-Hall, New Jersey, 1994.

57. Titze, I. R., Long, R., Shirley, G. I.
Messa di voce: an investigation of the symmetry of crescendo and decrescendo in a singing exercise. *Journal of the Acoustical Society of America*, 1999;105(5):2933-2940.

58. Weiss, Rudolf, W. S. Brown, Moris, J.
Singer's formant in sopranos: fact or fiction?. *Journal of Voice*, 2001;15(4):457-468.

59. Wolfe, J.
Homepage des Instituts für „Music Acoustics“ der Universität New South Wales: *Note names, MIDI numbers and frequencies*. `http://newt.phys.unsw.edu.au/jw/notes.html`, Website, undatiert, zuletzt aufgerufen am: 15.04.2015.

60. Young, R. W.
Terminology for logarithmic frequency units. *The Journal of the Acoustical Society of America*, 1939;11(1):166.