

Julia Merrill

Stimmen – schön schrecklich oder schrecklich schön?

Beschreibung, Bewertung und Wirkung
des vokalen Ausdrucks in der Musik



Julia Merrill

Stimmen – schön schrecklich oder schrecklich schön?
Beschreibung, Bewertung und Wirkung des vokalen
Ausdrucks in der Musik

Die vorliegende Arbeit ist Teil der Habilitationsschrift, die vom Fachbereich Humanwissenschaften der Universität Kassel als Habilitation zur Erlangung des akademischen Grades Doctor philosophiae habitatus (Dr. phil. habil.) angenommen wurde. Die vorliegende Fassung wurde gekürzt.

Erster Gutachter: Prof. Dr. Jan Hemming, Universität Kassel
Zweiter Gutachter: Prof. Dr. Clemens Wöllner, Universität Hamburg
Tag der öffentlichen Probevorlesung: 11. Juli 2018



Das e-book ist lizenziert unter einer Creative Commons Namensnennung - 4.0 International Lizenz.

Bibliografische Information der Deutschen Nationalbibliothek
Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <http://dnb.dnb.de> abrufbar.

ISBN 978-3-7376-5087-8
DOI: <http://dx.medra.org/10.19211/KUP9783737650878>
URN: <https://nbn-resolving.org/urn:nbn:de:0002-450874>

© 2019, kassel university press GmbH, Kassel
www.upress.uni-kassel.de

Covergestaltung: Felix Bernouilly
Printed in Germany

Vorbemerkung und Danksagung

Bei der vorliegenden Arbeit handelt es sich um einen Auszug aus der Habilitationsschrift, die am 15.01.2018 dem Fachbereich 01 Humanwissenschaften der Universität Kassel vorgelegt wurde. Weitere Teile der Schrift bestanden aus Studien, die an anderer Stelle publiziert wurden (Merrill, 2017; Merrill & Larrouy-Maestri, 2017; Merrill, 2019). Diese sind allerdings nicht zum Verständnis der vorliegenden Arbeit vonnöten.

Die Untersuchungen entstanden sowohl an der Universität Kassel als auch am Max-Planck-Institut für empirische Ästhetik, Frankfurt/Main. Deshalb gilt mein Dank zuerst den Mitarbeitern am MPIEA, durch die eine solch umfangreiche Datenerhebung überhaupt erst möglich gemacht wurde: Ganz besonders hervorzuheben ist Taren Ackermann, die mich in allen Teilen der Interview-Studie unterstützt und beraten hat sowie Sandro Wiesmann, der an der Durchführung und Transkription der Interviews sowie der Gruppentestung wesentlich mitgewirkt hat. Melanie Wald-Fuhrmann danke ich für die Freiheit, meinen eigenen Ideen nachgehen zu können und dafür die Infrastruktur des MPI nutzen zu dürfen. Für die sprechwissenschaftliche Expertise geht mein Dank an Lutz-Christian Anders und Clara Finke. Für wertvolle Kommentare zur Verbesserung des Manuskriptes danke ich Taren Ackermann, Jan Hemming und Clemens Wöllner, außerdem für das Korrekturlesen Gunhild Groh und Julius Reich. Auch möchte ich an dieser Stelle den vielen Studienteilnehmern danken – ohne sie gäbe es diese Forschung nicht.

Zuletzt bleibt mir noch, Jan Hemming zu danken, denn er hat mir die Möglichkeit gegeben, an der Universität Kassel zu habilitieren. Er hat dafür gesorgt, dass ich mich auf meine Arbeit konzentrieren konnte und ich konnte mir seiner Unterstützung immer gewiss sein.

Inhaltsverzeichnis

1	Einleitung.....	1
2	Theoretische und empirische Erkenntnisse zur Stimme.....	5
2.1	Kommunikative Funktionen der Stimme.....	5
2.1.1	Die ‚normale‘ Stimme und ihre Phonation.....	6
2.1.2	Das ‚normal-gestört‘-Kontinuum.....	10
2.1.3	Merkmale der gestörten Stimme.....	12
2.2	Evaluation von Stimmen und stimmlichen Funktionen.....	13
2.2.1	Klinik.....	13
2.2.2	Phonetik und Sprechwissenschaft.....	15
2.2.3	Reliabilität bei der Bewertung vokaler Merkmale.....	21
2.3	Singstimmen.....	22
2.3.1	Untersuchungen zur Singstimme in der populären Musik.....	24
2.4	Stimme und Emotion.....	26
2.4.1	‚Perceived and felt emotions‘ und ästhetische Emotionen.....	28
2.5	Stimme und Ästhetik.....	29
2.5.1	Die schöne und die gute Stimme.....	30
2.5.2	Zur Ästhetik der heiseren Stimme.....	33
2.5.3	Präferenz und die ideale Stimme.....	36
2.6	Zusammenfassung und Ausblick auf die aktuelle Studienreihe.....	41
2.6.1	Auswahl der Merkmale, Skalen und Musikstile.....	42
3	Prolog: Zuordnung von Stimmen und ihren Beschreibungen.....	47
3.1	Methoden.....	47
3.2	Auswertung und Ergebnisse.....	49
3.2.1	Offene Beschreibungen zum klassischen Sänger.....	50
3.2.2	Offene Beschreibungen zum Gospel-Sänger.....	52
3.2.3	Zuordnungstest.....	54
3.3	Diskussion.....	56
4	Beschreibung und Ablehnung vokaler Merkmale von Singstimmen – Interview-Studie.....	59
4.1	Methoden.....	59
4.1.1	Auswahl der Forschungsmethode.....	59
4.1.2	Teilnehmer.....	61
4.1.3	Ablauf des Interviews und Interviewleitfaden.....	63
4.1.4	Fragebogen.....	64
4.1.5	Expertenbefragung.....	66
4.1.6	Analyse.....	67
4.2	Ergebnisse.....	69
4.2.1	Einfluss von Hintergrundwissen.....	69
4.2.2	Gründe für die Ablehnung und Reaktionen auf die Stimmen.....	72
4.2.3	Abgelehnte Eigenschaften und Stimmbeschreibungen.....	76
4.2.4	Auswertung der Merkmale.....	79
4.2.5	Schlussfolgerungen für die Fragebogenentwicklung.....	87
4.2.6	Schlussbemerkungen.....	89

5	Präferenzen vokaler Merkmale und Fragebogenentwicklung – Gruppen-Studie	91
5.1	Methoden.....	91
5.1.1	Auswahl der Musiktitel.....	91
5.1.2	Fragebogen.....	92
5.1.3	Teilnehmer und Durchführung.....	93
5.2	Ergebnisse	95
5.2.1	Gefallen	95
5.2.2	Abhängigkeit der Bewertungen von der Position der Stimme im Testverlauf.....	101
5.2.3	Zusammengehörigkeit von Merkmalen	101
5.2.4	Einigkeit und Übereinstimmung mit den Expertenratings.....	103
5.2.5	Berichtete Schwierigkeiten bei der Bewertung	104
5.2.6	Schlussfolgerungen und Items für den neuen Fragebogen	106
5.2.7	Beschreibungsmuster und Diskriminanz der Sänger.....	108
6	Fragebogenentwicklung und ästhetische Emotionen – Online-Studie	113
6.1	Methoden.....	113
6.1.1	Auswahl der Stimmen und Items	113
6.1.2	Ablauf	114
6.1.3	Teilnehmer	116
6.2	Ergebnisse	117
6.2.1	Deskriptive Statistik der Merkmale	117
6.2.2	Expertise	119
6.2.3	Einheitlichkeit der Beurteilungen.....	119
6.2.4	Klassifikation der Sänger	123
6.2.5	Zusammenhänge von Merkmalen und Emotionen.....	128
6.2.6	Gefallen	129
6.2.7	Andere Einflüsse.....	135
6.2.8	Benötigte Stichprobengröße	139
7	Gesamtbetrachtung der Ergebnisse.....	143
7.1	EVEEx – Ein neues Inventar zur Beschreibung von Stimmen in der (populären) Musik.....	145
7.2	Ästhetische Emotionen und Werturteile	149
7.3	Fazit und Ausblick.....	152
8	Anhang	155
8.1	Studie Kapitel 3	155
8.1.1	Texte.....	155
8.1.2	Tabellen	157
8.2	Interview-Studie Kapitel 4.....	160
8.2.1	Anweisungen für die Experten	162
8.3	Gruppen-Studie Kapitel 5.....	165
8.3.1	Abschlussbefragung	165
8.3.2	Übersicht der Kommentare zu den Schwierigkeiten mit einzelnen Merkmalen	166
8.4	Online-Studie Kapitel 6	167
9	Literaturverzeichnis.....	173

1 Einleitung

Spätestens durch den Bestseller *The Human Voice* (Karpf, 2006b; dt. *Frauen reden anders, Männer auch*, Karpf, 2006a) wurden faszinierende Eigenschaften und Funktionen der Stimme in der alltäglichen Kommunikation einer breiten Öffentlichkeit vor Augen geführt – vom Alter zum Geschlecht, vom Milieu zu kulturellen Unterschieden. Es ist allerdings überraschend, auf wie wenige seriöse Quellen die Wissenschaftsjournalistin, trotz offensichtlich sorgfältiger und langer Recherche, zurückgreifen konnte. Dies kann mehrere Ursachen haben. Ein sehr wichtiger Grund scheint zu sein, dass die bisherige Forschung zur Stimme nur in begrenzten Kontexten wie der Klinik (Phoniatrie) oder beschränkt auf Basisemotionen stattfindet. Zwischenmenschliche Feinheiten lassen sich damit nicht abbilden, die aber gerade so populäre Themen wie *Warum Männer nicht zuhören und Frauen schlecht einparken* (Pease & Pease, 2001) publikumswirksam bedienen würden. Dies wiederum muss die interdisziplinäre Forschung zur Stimme nicht berücksichtigen – sie zielt auf wissenschaftliche bzw. medizinisch-therapeutische Ergebnisse ab, die Auswirkungen auf eine Behandlung/Therapie oder auf Erkenntnisse darüber haben, wie stimmbezogene Kommunikation funktioniert oder wie die Stimme in ästhetischen Zusammenhängen eingesetzt wird. Dies wiederum erfordert Grundlagenforschung zur Frage, wie die Stimme Freude, Angst und Trauer ausdrückt und mit welchen physiologischen oder neurophysiologischen Korrelaten dies zusammenhängt, um daraus beispielsweise Rückschlüsse auf die Funktion einzelner Gehirnregionen zu ziehen oder Beschreibungsebenen für die phonetische Forschung zu formulieren. Die meisten wissenschaftlichen Disziplinen betrachten primär die Sprechstimme. Das Singen wird häufig eher vergleichend herangezogen (z.B. in neurowissenschaftlichen Untersuchungen zum Singen und Sprechen), aber auch in der klinischen Forschung hinsichtlich seiner Funktionen und Störungen betrachtet. Auch die Pädagogik beschäftigt sich mit beiden, natürlich vor allem in Bezug auf die Praxis des Stimmtrainings, und die künstlerische Forschung versucht den Gesang im darstellerischen Kontext zu untersuchen.

Ein wesentliches Ziel vieler Ansätze ist die Beschreibung der Stimme, um sie in einem gewissen Kontext für einen gewissen Zweck zu evaluieren, von der klinischen Diagnostik bis zur Bewertung an Musikhochschulen. Stimmen lassen sich auf unterschiedlichen Ebenen beschreiben: Zum einen können dies Klangeigenschaften oder artikulatorische Merkmale sein, zum anderen aber auch der Ausdruck des Sängers und die im Hörer ausgelösten Emotionen – denn besonders der Gesang, als das direkte Ausdrucksmedium des Menschen, kann den Hörer berühren und emotional begeistern (z.B. Pfeleiderer, Hähnel, Horn & Bielefeldt, 2015). Auch können Persönlichkeitseigenschaften genauso wie der emotionale Zustand des Sprechers/Sängers am Klang der Stimme erkannt werden. Hier werden gerne Begriffe der Sinnesmodalitäten gewählt (Nessel, 1960) wie zum Beispiel ‚hell, warm, sanft, rau und schwarz‘ (wie der ‚schwarze Bass‘, der eine besonders dunkle und tiefe Stimme hat). Die Vielfalt an

Wissenschaftsdisziplinen, die an der Stimme und der professionellen Beschreibung ihrer Klangeigenschaften interessiert sind, führte zu einer Reihe unterschiedlicher Vokabulare. Je nach Expertise wird ein anderer Schwerpunkt gesetzt und andere Begriffe genutzt, z.B. in der Stimmheilkunde, im Schauspiel oder in der Gesangspädagogik, die sich nicht nur auf den klassischen Gesang, sondern mittlerweile auch auf poplarmusikalische Stile beziehen, innerhalb derer sich wiederum weitere Unterschiede finden. Während Experten eines Faches sich meistens über die Begrifflichkeiten und deren Bedeutung einig sind, wird die Verständigung fächerübergreifend schwieriger. Dies betrifft auch Laien, die ihre eigenen Begriffe kreieren und den fachspezifischen Termini teilweise nicht folgen können, da ihnen nicht nur die Bedeutung unbekannt ist, sondern auch ihr Gehör nicht hinsichtlich der benötigten Diskriminationsfähigkeit geschult ist.

Während Beschreibungen vokaler Merkmale, wie ‚hoch-tief‘ oder ‚hell-dunkel‘, präzise Eigenschaften von Stimmen schildern, verweisen andere Beschreibungen von Stimmen eher auf Konzepte wie die schöne oder die ideale Stimme und das Gefallen oder das Nicht-Gefallen. Den Eindruck der schönen Stimme, zum Beispiel, konstituieren nicht nur einzelne Merkmale, sondern u.a. Erwartungen, Hörgewohnheiten, etwaige Schlüsselerlebnisse, Kontext und persönlicher Geschmack. Diese können auf der Ebene einzelner Stimmen oder auch ganzer Gesangsstile untersucht werden, indem die Gründe für diese Einstellungen sowie die damit assoziierten vokalen Merkmale abgefragt werden. Dieses Vorgehen entspricht musikwissenschaftlicher Forschung, die sich mit den Gründen und Funktionen des Musikgeschmacks (hier verstanden als eine Einstellung) und der Präferenz (hier verstanden als ein Vorzugsurteil) von Musikstilen sowie in ersten Ansätzen auch mit der Ablehnung von Musikstilen beschäftigt (Ackermann, 2014, 2018; Bryson, 1996, S. 894f.; Greasley, Lamont & Sloboda, 2013, S. 406).

Die Untersuchungen der vorliegenden Arbeit sind genau in diesem Spannungsfeld aus Sprechwissenschaft, Phonetik, Phoniatrie und Musikwissenschaft zu verorten und nutzen Theorien und Ergebnisse aller dieser Disziplinen, um sich dem Komplex der *Beschreibung, Bewertung und Wirkung des vokalen Ausdrucks in der Musik* mit einem breiten methodischen Spektrum zu nähern. In der vorliegenden Studienreihe werden die Beschreibungen vokaler Merkmale von Stimm-Experten und Nicht-Experten sowie deren ästhetische Bewertung verschiedener Stimmen überwiegend aus der populären Musik (hier verstanden als ein Oberbegriff verschiedener Musikstile und -genres in Abgrenzung zur ‚Kunstmusik‘ des 16.-21. Jahrhunderts der westlich-abendländischen Tradition) untersucht. Neben der Untersuchung von Präferenzen und Disferenzen (Ackermann, 2014) vokaler Merkmale in verschiedenen Gesangsstilen von Stimm-Experten und Nicht-Experten ist es das Ziel der Arbeit, ein Inventar zur Beschreibung von Stimmen in der Musik zu entwickeln, welches auch von Nicht-Experten zur charakteristischen Beschreibung von Singstimmen genutzt werden kann.

Mit Studie 1 soll einer der Leitfragen der Arbeit nachgegangen werden: Inwiefern ist das Schreiben über Stimmen verständlich, und lässt sich ein Sänger charakteristisch beschreiben? Ein Zuordnungstest von Musikkritiken und den ihnen zugehörigen Aufnahmen der Stimmen soll erste Anhaltspunkte geben. Als Beispiel dienen drei klassische Tenöre, die eine Arie einer veristischen Oper interpretieren, sowie drei Frauenstimmen mit verschiedenen Gospel-Songs. Befragt wurden Musikstudierende, da bei ihnen eine höhere Expertise im Umgang mit Stimmen vorausgesetzt werden kann.

Grundlage der Arbeit sind die Beschreibungen und Bewertungen von Singstimmen von Nicht-Experten. Als Nicht-Experten wurden für diese Studien Hörer ausgewählt, die es nicht explizit gelernt haben, sich über Stimmen auszutauschen, d.h. sie haben sich keine Begriffe aktiv dafür angeeignet. Als Ausgangspunkt dient ein Fragebogen, der von Experten für Experten (überwiegend aus der Sprechwissenschaft) erstellt wurde. Dieser wird jedoch systematisch den Fähigkeiten und dem Verstehen der Nicht-Experten angepasst, um zu einem Inventar zu gelangen, welches Experten verschiedener Ausrichtungen (Singen und Sprechen) und Personen ohne besonderen Hintergrund gleichermaßen zur Beschreibung stimmlich-artikulatorischer Eigenschaften von Sängern populärer Musik verwenden können. Über drei Studien hinweg wurde dieses Inventar unter Einsatz qualitativer und quantitativer Methoden entwickelt. Es ist zu beachten, dass die Zielsetzung nicht die Erstellung eines standardisierten Inventars war. Solche Inventare finden sich vor allem in der klinischen Stimm-diagnostik und erfordern eine Vereinheitlichung sowohl der Durchführung von Tests als auch der Auswertung und Interpretation der Daten. Das hier entwickelte Inventar hingegen soll die Möglichkeit zur Anwendung in verschiedenen Situationen und für verschiedene Gegenstände geben und somit auf verschiedene Fragestellungen anzupassen sein.

Auch die ästhetische Bewertung von Stimmen hat einen Schwerpunkt eingenommen, wobei nicht nur das Gefallen, sondern auch konkret die Ablehnung untersucht werden. Es werden sowohl die Gründe für die Ablehnung (vgl. Ackermann, 2018) einer Stimme als auch die stimmlichen Merkmale, die zum Gefallen oder Nicht-Gefallen einer Stimme führen können, erörtert. Durch den qualitativen Ansatz konnten Anmerkungen und Kritik der Teilnehmer berücksichtigt werden, wodurch sich, durch das Bedürfnis vieler Teilnehmer, ein zusätzlicher Fokus auf die emotionale Wirkung von Stimmen entwickelte.

Es sei an dieser Stelle anzumerken, dass der Begriff ‚Stimme‘ in der vorliegenden Arbeit oftmals nicht nur klangliche, sondern auch artikulatorische und andere Ausdrucksmerkmale beinhaltet (siehe Kapitel 2.1). Außerdem sei darauf hingewiesen, dass aus Gründen der besseren Lesbarkeit auf die gleichzeitige Verwendung männlicher und weiblicher Sprachformen verzichtet wird. Sämtliche Personenbezeichnungen gelten gleichermaßen für alle Geschlechter.

2 Theoretische und empirische Erkenntnisse zur Stimme

Die Stimme ist ein wichtiger Bestandteil menschlicher Kommunikation. Sie nimmt dementsprechend eine wichtige Rolle in einer Vielzahl wissenschaftlicher Disziplinen ein, was sich auch in der folgenden Darstellung zum theoretischen und empirischen Kenntnisstand zur Stimme spiegelt. Es werden zunächst generelle Aspekte der Stimme in ihrer kommunikativen Funktion dargestellt, um dann zum besseren Verständnis der Evaluation von Stimmen spezifisch auf die pathologische (bzw. gestörte) und die normale Stimme einzugehen. Gerade dieser Unterschied wurde bislang am meisten untersucht und zeigt wertvolle Aspekte für das Hören und Beurteilen von Stimmen auf, die auch auf den künstlerischen Bereich bezogen werden können. Weiterhin soll ein Überblick über bestehende Inventare und Begriffe zur Stimmbeschreibung und -bewertung gegeben werden, der von der Klinik (Phoniatrie) und Phonetik über die Sprech- und Musikwissenschaft bis zur Emotionsforschung reicht. Es zeigen sich dabei sowohl Unterschiede als auch Überschneidungen in der Verwendung von Begriffen und dahinterstehenden Konzepten. So beschreibt die Phonetik eher akustische oder physiologisch-anatomische Verhältnisse der Stimme und des Sprechens, die Klinik solche Eigenschaften, die Störungen abbilden sollen, und die Musikwissenschaft nutzt Begriffe aus der musikalischen Praxis (z.B. Staccato und Glissando) oder der Gesangspädagogik und -forschung (z.B. Belcanto, Belting, Twang). Darüber hinaus wird der Zusammenhang zwischen dem vokalen Ausdruck und verschiedenen Emotionen betrachtet und in Bezug zum Gesang gesetzt, da vor allem die Singstimme im Fokus der Untersuchungen steht. Weiterhin wird die wenige Literatur dargestellt, die sich direkt mit ästhetischen Werturteilen zu Stimmen beschäftigt.

2.1 Kommunikative Funktionen der Stimme

Die Stimme macht Sprechen und Sprache hörbar und spielt eine große Rolle in der sozialen Interaktion, denn sie dient als ein mächtiger Träger der persönlichen Identität, des emotionalen Zustands, der Bildung und des Milieus des Sprechers. Sie trägt linguistische und paralinguistische Merkmale. Letztere umfassen neben der nonverbalen Kommunikation auch den Ausdruck von Emotionen in der Kommunikation. Die linguistischen Merkmale werden unterteilt in die Segmentalia (Phoneme) und die Suprasegmentalia. „Zu den suprasegmentalen Merkmalen gehören Sprechmelodie, Lautheit, Dauer, Sprechgeschwindigkeit, Sprechspannung, Pausen sowie (indexikalisch bedingte) Stimmqualität und Stimmausdruck (Timbre) und deren jeweilige Variation“ (Bose, Hirschfeld & Neuber, 2016, 38f.). Diese auch als Prosodie bezeichneten Merkmale wirken zusammen und haben neben einer strukturierenden Funktion auch die Aufgabe, die emotionalen Zustände und Gefühle des Sprechers wie Trauer, Freude und verschiedene Intentionen deutlich zu machen. Wichtig für die Dekodierung ist eine Kenntnis soziokultureller Merkmale von Seiten des Hörers, damit er diese Zeichen im Sinne des Sprechers deuten kann.

Auf der Suche nach vergleichenden Merkmalen und einer standardisierten Vergleichsebene wird immer wieder versucht, die ‚normale‘ Stimme zu definieren. Dies ist besonders im klinischen Bereich von Interesse, da hier besonders eine Standardisierung gewünscht wird (vgl. auch Goy, Fernandes, Pichora-Fuller & van Lieshout, 2013). Berg et al. (2017) untersuchten zum Beispiel die Sprechstimme in der Bevölkerung in Abhängigkeit vom sozioökonomischen Status sowie vom Lifestyle, um Standardwerte für die klinische Diagnostik zu erhalten. Die zur Normierung verwendeten Daten umfassten die Sprechstimme in verschiedenen Lautstärken und zeigten, dass die Sprechstimmlage bei lauterer Stimmproduktion steigt, allerdings nicht in der Rufstimme. Bei Männern steigt sie zusätzlich mit dem Alter an. Aktive Raucher haben hingegen eine tiefere Sprechstimmlage, die sich bei ehemaligen Rauchern jedoch nicht mehr findet. Hier werden Unterschiede zwischen der ‚normalen‘ und der gestörten Stimme deutlich, die im Folgenden weiter betrachtet werden.

2.1.1 Die ‚normale‘ Stimme und ihre Phonation

Für die folgende Betrachtung der Phonation (Stimmerzeugung) wird vor allem auf die Terminologie der Phoniatrie und Sprechwissenschaft zurückgegriffen (weitere Erklärungen finden sich z.B. bei Richter, 2014, Nawka & Wirth, 2008, oder mit starkem Bezug zur Sängerstimme bei Seidner & Wendler, 2010). Hier soll besonders auf ein Standardwerk zur Stimme und ihrer Auffälligkeiten (*Greene and Mathieson's The Voice and its Disorders*, Mathieson, 2001) zurückgegriffen werden, in dem Mathieson versucht, die normale Stimme zu definieren, damit sie im späteren Verlauf von der gestörten Stimme abgegrenzt werden kann. Die einfachste Definition sei wohl, eine normale Stimme als gewöhnlich und unauffällig zu beschreiben, wobei dies nicht eine Stimme einschließe, die besonders ‚schön‘ ist. Deshalb sei es einfacher, eine Stimme dahingehend zu betrachten, ob sie in die Grenzen einer normalen Stimme fällt. Im Folgenden listet Mathieson (2001, S. 68) Merkmale auf, die die normale Stimme aufzeigt:

- Sie ist klar, d.h. sie ist nicht rau oder ausgeprägt hauchig, sie hat keine Brüche und/oder klingt heiser. Sie ist konsistent und bricht nicht ungewollt während einer Äußerung ab.
- Sie ist innerhalb verschiedener Settings hörbar und kann auch in großem Hintergrundlärm gehört werden. Menschen mit einer normalen Stimme sollten laut rufen können und sollten eine laute Unterhaltung in einem sozialen Setting durchhalten können.
- Eine normale Stimme ist alters- und geschlechtsangemessen.
- Sie hält ihre linguistischen und paralinguistischen Rollen zur Zufriedenheit des Sprechers ein.
- Sie hat Stabilität und verändert sich nicht unerwartet in einem ihrer Parameter, weder am Anfang (Einsatz, engl. ‚onset‘) der Phonation noch während des Sprechvorgangs. Der Sprecher kann sich auf seine Stimme verlassen.
- Sie ist flexibel hinsichtlich Tonhöhe, Lautheit und Qualität.

- Sie hat Durchhaltevermögen, so dass sie normalerweise im Arbeits- und sozialen Leben ohne Verschlechterung benutzt werden kann.
- Die normale Phonation ist angenehm. Den meisten Menschen sind physiologische Empfindungen, die mit dem Sprechen assoziiert sind, unbewusst, es sei denn, sie müssen die Lautstärke sehr erhöhen oder am Limit ihres Tonumfangs phonieren.
- Akustisch generiert die Harmonics-to-Noise-Ratio-Messung (HNR) Informationen zum Verhältnis von klanglichen und nicht-klanglichen Anteilen (Geräuschanteilen) im Stimmklang. So kann ein Wert die harmonischen Anteile im akustischen Signal von den nicht-harmonischen („noise“) unterscheiden. Ein erhöhter Wert spiegelt das auditive Empfinden einer Heiserkeit wider, allerdings ist der Nutzen als objektive Einschätzung der Heiserkeit eher unbefriedigend (vgl. Mathieson, 2001, S. 75). Außerdem gibt die HNR keinen Aufschluss über die laryngealen, d.h. im Kehlkopf stattfindenden, Abläufe. Eine normale Stimme zeigt variierende Grade an Behauchtheit, was hauptsächlich ein Ergebnis eines insuffizienten (unvollständigen) Stimmlippenschlusses ist, d.h. es tritt Luft durch die Glottis, dem Raum zwischen den Stimmlippen, aus (Hirano, 1981). Wenn die Behauchtheit ein gewisses Maß übersteigt und in einer funktionell ineffizienten Phonation resultiert, kommt es zu einer klinischen Stimmstörung. Rauigkeit, die auch zwischendurch in einer ansonsten normalen Stimme auftauchen kann, ist das Ergebnis irregulärer Stimmlippenschwingung (Pulse).

Mit dem letzten Punkt wird bereits etwas sehr Wichtiges angesprochen: Die Heiserkeit ist ein wesentliches auditives Merkmal einer Stimmstörung (Dysphonie), die sich vom erfahrenen Hörer bereits auditiv in Behauchtheit und Rauigkeit unterscheiden lässt. Diese Unterschiede lassen sich allerdings instrumentell (noch) nicht so einfach wie auditiv darstellen – dies spricht für die Wichtigkeit der auditiven Beurteilung und die Unzulänglichkeit instrumentell-akustischer Messungen. Weitere Vorteile der auditiven Beurteilung sind, dass eine ganzheitliche und gestalthafte Beobachtung möglich ist und der Hörer den Fokus der Betrachtung frei festlegen kann (vgl. Anders, 2012). Dies trifft nicht nur auf klangliche Merkmale, sondern auch auf andere Merkmale des Sprechens zu, weshalb sich die aktuellen Studien in der vorliegenden Arbeit auf perzeptuell-auditive Beschreibungen konzentrieren (existierende Inventare dazu werden in Kapitel 2.2 dargestellt).

Mathieson stellt in Folge dieser Überlegungen eine Tabelle zusammen, die ein „Vocal Profile“ (Mathieson, 2001, 81ff.) beinhaltet – hier mit der Übersetzung ins Deutsche (siehe Tabelle 1). Darin werden bereits wichtige Merkmale der Stimmbeschreibung genannt, die sich in weiteren Inventaren (Kapitel 2.2) finden.

Vocal parameters	Acoustic measure	Vokale Merkmale	Akustische Messungen
Vocal note quality	Harmonics to noise ratio (HNR)	Klangqualität	HNR
Habitual pitch	Speaking fundamental frequency (SF ₀)	(Sprech-)Stimmhöhe	Grundfrequenz (F ₀)
Pitch • pitch range • pitch instability	Voice range profile Jitter (pitch perturbation)	Tonhöhe • Umfang • Instabilität	Stimmumfangsprofil Jitter (Frequenzschwankung)
Loudness • loudness instability	Intensity Shimmer (amplitude perturbation)	Lautheit • Instabilität	Intensität Shimmer (Amplitudenschwankung)
Resonance	Spectrum/spectral peaks	Resonanz	Spektrum/spektrale Spitzen
Flexibility (pitch and loudness)		Flexibilität (Tonhöhe und Lautheit)	
Stamina		Durchhaltevermögen	

Table 1: „Vocal profile parameters“ (Mathieson, 2001, S. 82). Die stimmlichen Parameter werden möglichen akustischen Messungen gegenübergestellt. Erläuterungen siehe Text. Spalte 3 und 4 sind eine Übersetzung von der Autorin.

Im Folgenden sollen diese und weitere wesentliche Merkmale der Stimmbeschreibung und deren anatomisch-physiologische und akustische Korrelate dargestellt werden. Diese Auflistung dient dem besseren Verständnis der Merkmale, die im Fragebogen der vorliegenden Arbeit verwendet wurden. Grundlage für die Zusammenstellungen bilden sowohl Mathieson (2001) als auch die Kriterien zur perceptiven Beurteilung der Stimme von Anders (2012) sowie Nawka und Wirth (2008).

- Die (mittlere) Sprechstimmhöhe hängt vom Geschlecht und Alter sowie der Art der Kommunikation und dem emotionalen Zustand des Sprechers ab. Sie kann sich situationsbedingt ändern, zum Beispiel in einer lauten Umgebung, beim lauten Vorlesen oder Sprechen am Telefon. Die Stimmhöhe und die Veränderungen der Tonhöhe werden durch die Stimmlippenlänge und Spannung, den Luftdruck unterhalb der Stimmlippen (subglottischer Luftdruck) und bestimmter Muskelaktivität im Kehlkopf bestimmt (Mathieson, 2001, S. 76). Im Zusammenhang mit der Stimmhöhe steht der Tonhöhenumfang. Um die jeweiligen Anforderungen des Sprechens und Singens zu erfüllen, ist eine entsprechende Flexibilität im Umfang erforderlich (Mathieson, 2001, S. 77). Sie kann subjektiv als die durchschnittliche Tonhöhenmitte einer Äußerung (im Sinne eines Grundtons) empfunden und beschrieben werden. Grundsätzlich gilt, dass mit einer höheren Sprechspannung eine höhere Sprechstimmhöhe einhergeht. Bei emotional geprägten Äußerungen liegt sie somit meistens höher als bei sachlicher Sprechweise (Anders, 2012).
- Unregelmäßigkeiten im Stimmklang, die als Heiserkeit wahrgenommen werden, lassen sich durch die akustischen Messwerte Jitter und Shimmer, d.h. Irregularitäten bzw. Instabilitäten in den Frequenz- oder Amplitudenperioden, darstellen. Die Messwerte, die eine gestörte Stimme diagnostizieren, werden allerdings verschieden interpretiert und variieren in den Untersuchungen, zum Beispiel abhängig von den Aufnahmebedingungen (Boersma & Weenink, 2017; Friedrich & Dejonckere, 2005; Zelcer, Henri, Tewfik & Mazer, 2002).

- Flexibilität und Durchhaltevermögen werden bei Mathieson (2001, S. 81) als nebengeordnete Kategorien verstanden. Jeder Parameter kann isoliert beschrieben werden und ein Parameter kann sich verändern, während die anderen stabil bleiben. Allerdings sind die Parameter in der Realität voneinander abhängig und die Veränderung des einen kann einen anderen beeinflussen.
- Stimmeinsätze (engl. ‚onsets‘) und Stimmabsätze (‚offsets‘) werden in der Literatur verschieden beschrieben. Mathieson (2001, S. 83) nennt den harten (‚hard glottal attack‘) und den weichen Einsatz (‚soft attack‘ / ‚breathy onset‘). Der harte Einsatz erscheint, wenn die Stimmlippen voll geschlossen sind, so dass die Ausatmung unterbrochen ist und explosionsartig freigelassen wird. Beim weichen Einsatz sind die Stimmlippen beim Einsatz der Stimmgebung nicht völlig geschlossen und die Luft passiert kurz vor dem Stimmeinsatz ohne Stimmlippenvibration die Öffnung. Anders (2012) unterscheidet zwischen physiologischen und pathologischen Einsätzen: Als physiologisch beschreibt er den festen Einsatz (der im Deutschen vor Vokalen, z.B. am Wortanfang, vorkommt), den weichen (der vor anlautenden stimmhaften Konsonanten und beim Singen vor anlautenden Vokalen gebildet wird) und den gehauchten (der nur beim Anlaut /h/ vorkommt). Pathologische Einsätze hingegen sind hart/gepresst, knarrend, behaucht und elidiert (d.h. ausgelassen; Anders, 2012, 36ff.). Steigt in emotionalen Zuständen die Spannung des Sprechers, führt dies auch zu vermehrt harten Einsätzen, sinkt sie, werden die Einsätze weicher und/oder behauchter.
- Die behauchte/verhauchte Stimme (‚breathy voice‘): Während der behauchten Phonation sind die Stimmlippen nicht völlig geschlossen (sie sind nicht adduziert), aber sie schwingen. Die hindurch strömende Luft ist hörbar, was den Eindruck der behauchten Stimme vermittelt. Die Lautheit der Stimme ist i.d.R. auch eingeschränkt.
- Die raue Stimme (‚harsh voice‘): Während der Produktion der rauhen Stimme sind die Stimmlippen stärker adduziert als in der behauchten Stimme. Die Stimmqualität wird auch als behaucht wahrgenommen, allerdings mit einer harten Qualität (‚hard edge‘, vgl. Mathieson, 2001, S. 84).
- Die gepresste Phonation (‚pressed voice‘): Hier sind die Stimmlippen fest/kräftig adduziert mit zusätzlichem erhöhten subglottischen Luftdruck (dem Druck unterhalb der Stimmlippen). Die Stimmgebung ist angestrengt und kann mit steigender Lautstärke und längerem Gebrauch unangenehm für den Sprecher werden. Die gepresste Stimme geht oft mit harten oder geknarnten Einsätzen einher. Gepresste Anteile können auch im Bereich der normalen Stimme vorkommen, dieses Stimmverhalten ist jedoch bereits inductiv für eine Dysphonie.

- Das Knarren kann als eine Extremform der Rauigkeit betrachtet werden. Es tritt auf, wenn keine Periodizität im Klang mehr erreicht wird. Klanglich kann es mit einem Türknarren assoziiert werden.
- Die Resonanz wird bei Mathieson nur hinsichtlich des klassischen Gesangs beschrieben, bezogen auf die Technik des sog. gedeckten Singens und auf den Sängerformanten (welcher ein besonderes Energiespektrum des klassischen Sängers bezeichnet, um eine Tragfähigkeit der Stimme zu erreichen; Sundberg, 1974). Als verwandtes Merkmal ist die Klangfülle zu nennen, die auch auf die Sprechstimme bezogen werden kann, d.h. sich auf das Klangspektrum bezieht.
- Die Nasalität entsteht im Ansatzrohr durch eine Senkung des Velums, d.h. durch einen Abschluss des Nasenraumes und somit der Nasenresonanz (sog. geschlossenes Näseln).

All diese Merkmale (und viele mehr) wirken zusammen und geben wichtige kommunikative Informationen wieder: „At any particular moment, an individual’s voice reflects: the structure and function of the vocal tract; the person’s neurological and psychological status; the auditory and kinaesthetic feedback that he or she is receiving; and the behaviour of the listener or audience“ (Mathieson, 2001, S. 91). Um dies möglich zu machen, müssen sich die Merkmale jederzeit der Situation anpassen können, d.h. flexibel sein. Als Folgerung aus dieser Flexibilität und da die Stimme von körperlichen Bedingungen abhängt, ist die Grenze zwischen einer normalen und einer gestörten Stimme schlecht zu definieren. Mathieson geht von einem „normal-disordered continuum“ (2001, S. 82) aus, was die Möglichkeit einschließt, dass eine im Allgemeinen normale Stimme von Zeit zu Zeit als ‚unnormal‘ wahrgenommen wird, wenn nur ein Momenteindruck von der Stimme außerhalb des Kontexts gehört wird.

2.1.2 Das ‚normal-gestört‘-Kontinuum

Mit diesem Kontinuum soll die Variabilität der Stimme allgemein und die Überschneidungen der normalen und der gestörten Stimme abgebildet werden – auch unter Berücksichtigung der Gründe für die auftretenden stimmlichen Eigenschaften und den Kontext, in dem sie auftreten. Die Überschneidungen zwischen der normalen und der gestörten Stimme treten aus verschiedenen Gründen auf, die Mathieson (2001, S. 126) auflistet:

- Physiologische Anforderungen: Wenn physiologische Anforderungen die Überhand über die Phonation gewinnen, können normale Stimmen sehr unnormale klingen, z.B. durch verschiedene Repertoires. Das bedeutet, dass ein untrainierter Sänger beispielsweise nicht die Leistungen eines Opernsängers zeigen kann und dies somit keine Anomalie darstellt. Wenn jedoch der Opernsänger seine sonst üblichen Leistungen verliert, kann es als eine Störung bezeichnet werden.

- Variabilität der stimmlichen Eigenschaften: Manche Merkmale werden bei vorübergehendem Auftreten als normal bezeichnet, wenn sie konsistent werden aber als Störung. Ein Stimmknarren ist beispielsweise am Aussageende unauffällig, ein dauerhaftes Knarren kann jedoch Schäden an den Stimmlippen hervorrufen.
- Altersbedingte Veränderungen: Hier spielen vor allem hormonelle Veränderungen eine Rolle.
- Inkonsistente stimmliche Anomalitäten: Eine gestörte Stimme ist nicht notwendigerweise in ihren Anomalien konsistent und manchmal kann die Stimme in einen normalen Zustand zurückkehren. Eine Stimme kann zum Beispiel morgens normal sein und sich bis zum Abend dysphonisch entwickeln.
- Abgrenzung der normalen und gestörten Stimme: An der Grenze von normal zu gestört kann dem Sprecher selbst seine Stimme unnormal vorkommen, während ein Hörer dieses noch nicht als auffällig wahrnimmt. Auf der anderen Seite kann ein Hörer eine Veränderung bemerken, die dem Sprecher noch nicht bewusst ist. Deshalb ist die Abgrenzung schwer vorzunehmen, obwohl eine Aphonie (völliges Versagen der Stimme) oder eine schwere Dysphonie meist offensichtlich für den Hörer sind. Die Kriterien der Einteilung hängen deshalb auch von sozialen, kulturellen und Belastungsnormen sowie objektiven Evaluationen ab.

Diese Zusammenhänge lassen sich auch auf den künstlerischen Kontext übertragen, in dem Künstler (Schauspieler, Sänger etc.) versuchen, die Ausdrucksmöglichkeiten der Stimme auszunutzen. Interessant ist in diesem Zusammenhang die Darstellung von Shewell (2009, S. 45), die ebenfalls die normale gegen die ‚abnormale‘ Stimme auf ein Kontinuum setzt (der Begriff ‚abnormal‘ wird in diesem Absatz von Shewell aus dem Englischen übernommen, er ist aber nicht ästhetisch-normativ zu verstehen). Allerdings bezieht sie sich auf die Stimmarbeit im künstlerischen Kontext und beschreibt, dass nur die normale Stimme einer „aesthetic development voice work“ unterzogen wird, während die ‚abnormale‘ Stimme einer „therapeutic voice work“ (Shewell, 2009, S. 45) unterzogen werden sollte. In ihrem Modell zur ‚voice work‘ werden bestimmten Berufen (Therapeuten, Gesangspädagogen, Wissenschaftler) bei der Arbeit mit der normalen und der ‚abnormalen‘ Stimme unterschiedliche Aufgaben zugewiesen. Es soll zeigen, für welches Stadium der Stimme welcher Beruf zuständig ist. Interessanterweise berücksichtigt sie jedoch die Beschäftigung mit der ‚abnormalen‘ Stimme im künstlerischen Kontext nicht, was fragwürdig ist. Es könnte vermutet werden, dass gerade dieses ‚Abnormale‘ das ästhetische Erleben verstärkt oder verändert. Im Folgenden soll deswegen genauer auf die sogenannten Stimmstörungen eingegangen werden, die im medizinischen Bereich als gestört und somit als behandlungswürdig definiert werden, aber in ästhetischen Kontexten möglicherweise eine andere Rolle spielen.

2.1.3 Merkmale der gestörten Stimme

Stimmen habe eine breite Ausdruckspalette mit vielfältigen klanglichen Variationen, weshalb auch eine gestörte Stimme eine Vielzahl an Veränderungen aufweisen kann: „Voice disorders range from complete absence of the voice (aphonia) to varying degrees of vocal impairment (dysphonia). Abnormalities can involve one or more of the vocal parameters: habitual pitch, pitch range, loudness, vocal note quality, resonance, flexibility and stamina. As a result, disordered voices can range from being functionally ineffective through varying degrees of inefficiency to being merely aesthetically displeasing in their least severe form.” (Mathieson, 2001, S. 121). Die hier benannten vokalen Parameter der gestörten Stimme fasst Mathieson in einer Tabelle zusammen, die in engem Zusammenhang mit den oben beschriebenen Parametern der normalen Stimme steht. Neben den auditiven Merkmalen werden auch die Mechanismen im Vokaltrakt bzw. Kehlkopf dargestellt. Die Tabelle 2 ist für die vorliegende Arbeit interessant, da sie gleichzeitig als eine Auflistung der Ausdrucksmöglichkeiten einer Stimme betrachtet werden kann und eine Bewertungsskala (z.B. ‚too high‘ und ‚too low‘) beinhaltet, die in den vorliegenden Untersuchungen eingesetzt wurde.

Acoustic parameter variations and underlying laryngeal behaviours

(The vocal tract behaviours listed as possible dysfunction can be caused by structural, functional, neurological and emotional factors.)

Parameter affected	Disordered acoustic feature	Possible vocal tract dysfunction
Habitual pitch (speaking fundamental frequency/SF₀)	<ul style="list-style-type: none"> • Too high • Too low • Unstable • Pitch breaks 	<ul style="list-style-type: none"> • Vocal folds excessively tightened • Vocal folds : <ul style="list-style-type: none"> - excessively short and thick - increased vocal fold mass • Vocal fold tension unstable • Vocal fold tension unstable
Pitch range	<ul style="list-style-type: none"> • Restricted to upper or lower frequencies • Lowered upper and lower frequencies • Raised upper and lower frequencies • Excessive pitch variation • Insufficient pitch variation • Reduced control 	<ul style="list-style-type: none"> • Insufficient vocal fold adjustment • Vocal folds excessively short and thick or increased vocal fold mass • Increased vocal fold tension • Intermittent excessive lengthening and shortening of the vocal folds • Minimal variation in length of the vocal folds • Changes in vocal fold structure or in related Systems, e.g. respiratory, neurological
Loudness	<ul style="list-style-type: none"> • Too loud • Too quiet • Unstable volume 	<ul style="list-style-type: none"> • Excessive subglottal air pressure and vocal fold tension • Insufficient subglottal air pressure and inadequate vocal fold tension and adduction • Variable vocal fold adduction and tension and/or variable subglottal air pressure
Vocal note quality	<ul style="list-style-type: none"> • Rough • Breathy • Creaky • Voiceless segments 	<ul style="list-style-type: none"> • Chaotic vocal fold vibrations or ventricular fold involvement • Turbulent air within the glottis/glottic chink • Hyperadducted vocal folds • Intermittent incomplete vocal fold adduction/glottis chink
Vocal flexibility	<ul style="list-style-type: none"> • Reduced flexibility/control <ul style="list-style-type: none"> - pitch and/or - loudness 	<ul style="list-style-type: none"> • Limited range of vocal fold lengthening and/or shortening • Limited variation of subglottal air pressure and/or vocal fold tension/adduction

Vocal stamina	<ul style="list-style-type: none"> • Inability to sustain vocal function of the type and for the time required without vocal deterioration 	<ul style="list-style-type: none"> • Impaired respiratory and/or vocal fold function • Deterioration of vocal fold mucosa status with use
----------------------	---	---

Tabelle 2: Vokale Merkmale in extremer bzw. als gestört interpretierter Ausformung und ihre anatomisch-physiologischen Korrelate (Mathieson, 2001, S. 122f.).

2.2 Evaluation von Stimmen und stimmlichen Funktionen

Die bisher dargestellten Definitionsversuche einer normalen und einer gestörten Stimme bestimmen auch die Inventare zur Beschreibung von Stimmen, je nachdem, in welchem Kontext und zu welchem Zweck Stimmen evaluiert werden sollen. Es werden nun zuerst die Erhebungsinstrumente aus der klinischen Praxis dargestellt, die von einem stärkeren Bestreben nach Standardisierung bestimmt sind. Darunter finden sich sowohl Inventare, die nur einige wenige Parameter untersuchen, aber auch solche, die um eine umfassende Darstellung der Stimme bemüht sind. Für die Forschung ist besonders die Darstellung und Beschreibung klanglicher Merkmale von Interesse, was gleichzeitig auch die meisten Schwierigkeiten beinhaltet, auf die danach eingegangen wird. Damit erweitert sich auch der Kontext der Inventare auf andere Wissenschaften wie die Phonetik und Sprechwissenschaft.

2.2.1 Klinik

Zur Diagnostik einer Stimme und zur Entscheidung über die folgende Behandlung gehören eine Reihe von Erhebungen, zu denen neben instrumentell-akustischen und perzeptuellen auch physiologische und psychische Parameter sowie die Lebensumstände des Patienten gehören. So umfassen bereits einige Inventare zur Stimmfunktion auch körperliche Auffälligkeiten, die direkte Einflüsse auf die Phonation haben (z.B. falsche Körperhaltung oder Atemtechnik und Räuspern). Für die Erhebung einer Stimme in einer Therapiesitzung, zum Beispiel, schlägt Mathieson (2001) die folgenden Aspekte vor:

Informal perceptual evaluation of phonation

(The aspects of phonation can be observed informally in general conversation with the patient.)

Vocal behaviours and parameters	Possible features observed
Fundamental vocal note quality	Rough Breathy 'Wet' Consistent quality Varies according to pitch/loudness/phonation/or vegetative behavior/emotion
Habitual pitch	Too high/low Unstable
Pitch range	Monopitch Limited in upper range Flexible intonation Extremes of pitch range

Loudness	Too quiet/loud Loudness decay, i.e. voice becomes quieter at end of utterance, or as conversation progresses
Resonance	Hyper-/hyponasal
Onset	Hard glottal attack may coexist with breathy output
Vocal habits	Vigorous throat-clearing and coughing
Extrinsic laryngeal muscles	Observation of the neck during conversation can reveal tense strap muscles
Vocal stamina/fatigue	Voice deteriorates/improves with use during interview
Aberration	Normal voice may be heard on vegetative tasks, e.g. laughing, but not during phonation Voice does not reflect laryngoscopic findings, or is much better/worse than expected
Posture	Excessive chin raising or lowering affecting laryngeal position and movement Jutting head forward affecting laryngeal movement and dimensions of supralaryngeal spaces Stooped, rounded shoulders affecting chest expansion

Tabelle 3: *Perceptive Evaluation der Stimmproduktion in einer informellen Situation (Mathieson, 2001, S. 418).*

Diese Aufstellung fasst einige der in Kapitel 2.1.1 ausführlich beschriebenen Merkmale zusammen. Die Stimme wird damit nicht nur beschrieben, sondern auch bewertet, wobei sowohl tendenziell negative Konnotationen vorkommen, wie ‚too high/low‘ oder ‚limited in upper range‘, als auch positive Konnotationen wie ‚consistent quality‘ (siehe Tabelle 3). Es wird dem Therapeuten offen gelassen, wie er die vokalen Merkmale genauer innerhalb der Kategorien (Spalte 1, Tabelle 3) bewertet. Außerdem wird deutlich, dass dieser Evaluationsbogen primär als ein Leitfaden für den Therapeuten zu verstehen ist und als Grundlage für die Therapie fungieren soll. Für die Diagnostik von Stimmstörungen eignet sich diese Form der Erhebung weniger. Für solche Situationen stehen andere Inventare zur Beurteilung der Stimme zur Verfügung, deren Ziel eine standardisierte Erhebung ist, die bestenfalls ‚harte Daten‘ produziert, wie die folgenden Inventare.

Die Auswahl des Erhebungsinstrumentes hängt davon ab, welche Aspekte der Stimme untersucht werden sollen. Besonders häufige Verwendung in der klinischen Praxis findet die GRBAS-Skala (Hirano, 1981), die auf fünf Ebenen die Heiserkeit evaluiert: G = ‚overall grade or degree‘ (Grad der stimmlichen Anomalität), R = ‚rough‘ (Rauigkeit), B = ‚breathy‘ (Behauchtheit), A = ‚asthenic‘ (Asthenie), S = ‚strained quality‘ (Spannung). Wendler und Anders (1986) sowie Nawka, Anders und Wendler (1994) haben später gezeigt, dass einige Parameter der Skala obsolet sind. Spannung scheint aufgrund der starken Streuung der Daten keine eindeutig zu bestimmende Größe zu sein und Asthenie nicht unabhängig von der Behauchtheit zu sein. Daher hat sich in deren klinischer Praxis die Beurteilung auf der Ebene von drei Qualitäten herausgestellt: R = Rauigkeit, B = Behauchtheit, H = Heiserkeit (RBH).

Um Phoniater ein Training für den Umgang mit der RBH-Skala zu ermöglichen, veröffentlichten Nawka und Anders (1996) eine Compact Disc mit Beispielen, so dass ein Maßstab zur Einordnung aufgebaut werden kann. Die Bewertung der drei Qualitäten RBH erfolgt auf einer 4-stufigen Skala, wobei 0 = nicht vorhanden, 1 = leicht oder geringgradig, 2 = mittelgradig und 3 = hochgradig vorhanden bedeutet. „Selbstverständlich bedeutet jede Einschätzung der Stimme als rau oder behaucht, daß es sich um eine heisere Stimme handelt. [...] Der Gesamtheiserkeitseindruck kann nicht besser sein als

die Bewertung einer Komponente, der Rauigkeit oder der Behauchtheit“ (Nawka & Anders, 1996, S. 19). So können die Rauigkeit und die Behauchtheit als untergeordnete Komponenten der Heiserkeit angesehen werden, was in Untersuchungen dazu genutzt werden kann, dass entweder nur die Heiserkeit, oder die Kombination von Rauigkeit und Behauchtheit erhoben werden kann. Interessanterweise stellen die Autoren heraus, dass „die Rauigkeit meist stärker ausgeprägt empfunden wird als die Behauchtheit.“ (Nawka & Anders, 1996, S. 19).

Die Einteilung nach RBH hat in der Folge zu einheitlichen sowie zu longitudinal reproduzierbaren Einschätzungen geführt. Anders, Hollien, Hurme, Sonninen und Wendler (1988) zeigten sogar, dass die Bewertung des Heiserkeitsgrades eine weitgehende Übereinstimmung zwischen den Urteilen von professionell geschulten Hörern und von Laien nach Training sowie auch zwischen Beurteilern unterschiedlicher Nationalitäten hervorbrachte. Dies zeigt, dass Training in der auditiven Bewertung von Stimmen nötig, aber auch möglich ist, denn besonders die Evaluation der Stimmqualität gilt allgemein als eine schwierige Aufgabe, die einschlägige Erfahrung braucht (Anders, 2012; Bassich & Ludlow, 1986; Mathieson, 2001; Richter, 2014; Seidner, Dippold & Fuchs, 2009), vgl. auch Kapitel 2.2.3.

Als ein weiterer mit der GRBAS-Skala verwandter Index ist der *CAPE-V – Consensus Auditory-Perceptual Evaluation of Voice* zu nennen, der sechs vokale Merkmale evaluiert: ‚Overall Severity‘, ‚Roughness‘, ‚Breathiness‘, ‚Strain‘, ‚Pitch‘ und ‚Loudness‘. Dazu gekommen sind somit Tonhöhe und Lautheit, außerdem ist die Bewertungsskala eine visuelle analoge Skala und keine Intervallskala (Kempster, Gerratt, Verdolini Abbott, Barkmeier-Kraemer & Hillman, 2009). Mit dieser Skala nähert sich die Klinik einer standardisierten, umfassenderen Stimm-Evaluation an, was sowohl die GRBAS- als auch die RBH-Skala nicht bieten können.

2.2.2 Phonetik und Sprechwissenschaft

Weitere Inventare kommen aus der Phonetik, wie das *VPA (A perceptual protocol for the analysis of vocal profiles)*, ein Analyse-Protokoll von Laver (Laver, 1980; Laver, Wirz, Mackenzie & Hiller, 1981). Dieses Profil ist das erste umfassende Inventar für die phonetische Beschreibung von Sprechercharakteristika, die basierend auf der Anatomie und der gelernten Muskeleinstellung beschrieben werden („settings“; Laver, 1994, S. 115). Neben Einstellungen im Vokaltrakt (z.B. die Zunge und Lippen betreffend) finden sich Beschreibungen des Klangs hinsichtlich der velopharyngealen Einstellungen (bezogen auf den Bereich des weichen Gaumens), die als „nasal, audible nasal escape und denasal“ (Laver, 1994, S. 154) beschrieben werden. Auch kann die Position des Kehlkopfes (Larynx) gehoben oder gesenkt sein. Klangqualitätsmerkmale wie rau und flüstern werden in die „phonation types“ „harshness, whisper(y), creak(y), falsetto“ und „modal voice“ (Laver, 1994, S. 199) eingeteilt. Die Spannung („tension“) wird supra- und laryngeal mit „tense und lax“ (Laver, 1994, S. 154) beschrieben. Prosodische Merkmale beziehen sich auf die Tonhöhe, ihren Umfang und

die Variabilität, „pitch (high mean, low mean, wide range, narrow range, und high and low variability)“ (Laver, 1994, S. 154); auch die Lautheit bezieht sich auf diese ‚settings‘. Außerdem wird noch ein möglicher Tremor (nur wenn zugegen, ‚present‘) evaluiert. Zusätzliche „comments“ beziehen sich auf „breath support, continuity (interrupted), rate (fast, slow) und rhythmicity“ (Laver, 1994, S. 154). Alles wird mit Ratingskalen erfasst, manchmal die Begriffe ‚appropriate‘ und ‚inappropriate‘ umfassend. Damit enthält das VPA wesentliche Begriffe, die auch in späteren Inventaren vorkommen. Während der Vorteil des Inventars eben diese multimodale und tiefgehende Evaluation der sprecherischen Charakteristika ist, führte dies allerdings zu Problemen in der Reliabilität (Webb et al., 2004), vgl. Kapitel 2.2.3.

Anschließend an diese Beschreibung anatomisch-physiologischer Zustände gibt es Versuche, definitive, deskriptive Terminologien zu erstellen und die Merkmale der Stimmqualität mit phonetischen Symbolen in Form der *Voice Quality Symbols (VoQS)* von Ball, Esling und Dickson (1995) darzustellen und zu systematisieren (oder auch Vieregge, Pahn & Schutte, 1996). Diese Symbole können eine normale und eine gestörte Stimme beschreiben und umfassen Zeichen, die sich drei Kategorien zuordnen lassen (und dem VPA ähnlich sind): Erstens den Luftstrom-Typen, d.h. wie der Luftstrom beim Phonieren erzeugt wird, zweitens den Phonationstypen, die Register, die Flüsterstimme, und die Knarrstimme umfassen, und drittens die supralaryngealen Settings, also alle Zustände des Ansatzrohres und der Artikulatoren, die den Klang beeinflussen, wie einen gehobenen/gesenkten Kehlkopf, eine nasalierte Stimmgebung etc.

Das Inventar, das die Grundlage für die vorliegende Arbeit bildete, ist auf der Basis verschiedener Inventare, wie den bisher beschriebenen, entstanden. Der *Katalog zur auditiven Merkmalsbeschreibung von Sprechausdrucksmustern* (Bose, 2001, S. 301f.), kurz ‚Merkmalskatalog‘, ist, wie der Titel bereits sagt, auf eine möglichst umfassende Beschreibung des Sprechausdrucks ausgerichtet und umfasst somit eine große Anzahl verschiedener Merkmale (N = 36). Diese sind in sechs Kategorien eingeordnet, die aus bestehenden Inventaren entnommen und erweitert wurden. Ziel des Katalogs ist es, den „stimmlich-artikulatorischen Ausdruck“ (Bose, 2010, S. 29) umfassend zu beschreiben. Er umfasst die folgenden Kategorien, jeweils mit ihren vielfältigen Modifikationen: Sprechtonhöhe, Lautheit, Stimmklang, Sprechgeschwindigkeit, Artikulation und weitere Sprechausdrucksmerkmale (Komplexwahrnehmungen). Beschreibungen der einzelnen Merkmale finden sich bei Bose (2001, 2003, 2010) und für die übernommenen Merkmale im Verlauf der vorliegenden Arbeit. Der Katalog ist nicht standardisiert und dafür gedacht, dass er verschiedenen Fragestellungen angepasst werden kann. Er ist in seiner Anwendung durchaus breit und reicht vom künstlerischen zum emotionalen Kontext.

Die bisher vorgestellten Inventare setzen eine Expertise zur Nutzung voraus. Es gibt jedoch ein Inventar, welches für Nicht-Experten (nicht trainierte Hörer) erstellt wurde: Die *Geneva Voice Perception Scale* (GVPS; Bänziger, Patel & Scherer, 2014), ein Fragebogen zur Beschreibung emotionaler Aussagen. Die enthaltenen Merkmale umfassen ‚loudness‘, ‚pitch‘, ‚intonation‘, ‚sharpness‘, ‚articulation‘, ‚roughness‘, ‚instability‘ und ‚speech rate‘. Außerdem wurden die akustischen Parameter der acht untersuchten Basisemotionen untersucht. Mit einer Multiplen Diskriminanzanalyse (MDA) wurde festgestellt, dass sowohl die auditiven als auch die akustischen Parameter zu einer guten Unterscheidung der Emotionen führten. Positive Emotionen können allerdings besser auditiv als akustisch diskriminiert werden, was zeigt, dass sich Valenz schlecht akustisch diskriminieren lässt. Die erhobenen Emotionen waren die Basisemotionen Ärger, Freude, Trauer, Verzweiflung und Angst – wobei jeweils die niedrige und hohe Erregung untersucht wurde (kalter und heißer Ärger etc.). Das Inventar kann als standardisiert bezeichnet werden, allerdings wurde es auf Französisch entwickelt und nur zusammen mit der englischen Übersetzung publiziert.

2.2.2.1 Beschreibungen der Klangqualitäten gesprochener Sprache

Besonders die Klangqualitäten von Stimmen beschäftigen die Wissenschaft seit langem. Deshalb soll im Folgenden ein Rückgriff auf historische Quellen gemacht werden, die bereits umfassende Beschreibungen zur Stimme beinhalten und sich bis heute erhalten haben. „Denn Stimmen sind gespannt wie Saiten, die entsprechend der jeweiligen Berührung reagieren, hoch und tief, schnell und gemessen, laut und leise. [...] Zusätzlich leiten sich von ihnen auch diese Varianten her: sanft oder rau, gepresst oder verströmend, getragen oder abgehackt, dumpf oder kreischend, mit wechselndem Ton an- und abschwellend“ (Marcus Tullius Cicero, *Brutus*; zit. nach Foehr-Janssens, 2008, S. 129). In der antiken Rhetorik finden sich einige Quellen mit ausführlichen Zusammenstellungen von Begriffen zur Beschreibung der Stimmqualität (für eine Übersicht siehe Kranich, 2016). Eine sehr umfassende stammt aus dem *Onomasticon* von Iulius Pollux mit folgenden Begriffen und alternativen Übersetzungen: Hoch (erhaben), klar (glänzend), weit (inhaltliche Weite), tief (gewichtig, nicht inhaltlich gemeint), hell (strahlend, erhellend im rhetorischen Sinne), rein (gesund, sauber, gepflegt), süß (im Sinne von Geschmack), lieblich (im Sinne von Empfindung), wohlklingend (Stimme und Sprache), geschmeidig (biegsam, anpassungsfähig), gleitend (schön gewunden, umfließend, rollend; sprachlich und sprecherisch), misstönend (schwirrend, scharf, zischend, laut misstönend), deutlich (einleuchtend, verständlich; inhaltlich), deutlich (klar, hell; sprecherisch deutlich), dunkel (evtl. inhaltliche Verdunklung), sich erhebend (dämmrig positiv, dumpf negativ zu verstehen), kraftlos (klein, schwach), misstönend (unharmonisch), missklingend (unmoduliert, unharmonisch), rau (hart, kratzend; ungeschult), heiser (schwach; krankhaft), metallisch (Verweis auf Erz und Kupfer), schrill (scharf, spitz, schneidend) (zit. nach Kranich, 2016, S. 38ff.). Diese Aufstellung ist auch aus heutigem Blickwinkel sehr umfangreich. Interessant ist, dass sich die Begriffe nicht nur auf sachliche

Beschreibungen beziehen, sondern wertende Urteile einbeziehen. Außerdem beziehen sie nicht nur die Beschreibung sprecherischer Merkmale ein, sondern (und teilweise gleichzeitig) auch inhaltliche, den Vortrag betreffende. So kann sowohl die Stimme als auch die Sprache ‚wohlklingen‘ und ‚gleitend‘ sprachlich und sprecherisch gemeint sein; ‚hell‘ beschreibt weniger die Klangfarbe, als den erhellenden Vortrag. Die Klangqualitätsmerkmale ‚rau‘ und ‚heiser‘ werden hier zwar klanglich mit ‚kratzend‘ beschrieben aber bereits mit ‚ungeschult‘ bewertet und im Sinne von ‚krankhaft‘ verwendet.

Es mangelt in der Literatur nicht an umfangreichen und kreativen Listen mit Begriffen zur Beschreibung von Stimmen und deren klanglicher Qualität. Oft werden diese Begriffe verschiedenen Sinnesmodalitäten entlehnt, wie Nessel (1960) in einer Meta-Analyse mit Texten des ‚deutschen Fachschrifttums‘ herausfindet:

- „Akustisch: Krächzend, kratzend, knarrend, rasselnd, prasselnd, schmirgelnd, fauchend, hauchig, verhaucht, scheppernd, scherbelnd, gesprungen (Topf), nasal, piepsend, pfeifend, röchelnd, brummend, blechern, gellend, kreischend, tonlos, klangarm
- Optisch (bildlich vergleichend): belegt, verschleiert, gepresst, abgeschnürt, gestopft, kloßig, gaumig, halsig, kehlig, flatternd, schwebend, wackelig, zittrig, matt, grell, flach, hohl, fädig (Faden in der Stimme)
- Gustatorisch: muffig, muffelig
- Taktile: rau, scharf, stumpf, hart, kalt, dünn, spitz, schwer, schneidend“ (Nessel, 1960, S. 5f.).

Beim Lesen der Begriffe fällt auf, dass einzelne bereits auf einen emotionalen Zustand des Sprechers oder sogar auf Persönlichkeitsmerkmale hindeuten. Die Zuschreibung gewisser Merkmale und deren Wirkung findet sich auch schon in der Antike: „Zur Beschreibung stimmlicher Merkmale geht Aristoteles von Gefühlsqualitäten aus, die sich für ihn in Polaritäten wie *warm-kalt* oder *hart-weich* manifestieren. Nach diesem Prinzip könne man die Stimme nach *hoch-tief*, *stark-schwach* oder *sanft-rau* beschreiben. Auch Kriterien entsprechend der Farbqualitäten seien für die Stimme geeignet“ (Kranich, 2016, S. 30).

Differenzierte Diskussionen von Klangbeschreibungen und deren Konfundierung und Korrelationen sind seltener. Eine Ausnahme davon bildet Stumpf, der die Phänomene der Tonhöhe und der Klangfarbe zu differenzieren versucht: „Die Höhe, die wir mit Helligkeit identisch setzen, involviert die gewaltigen Unterschiede von den tiefsten, dunkelsten bis zu den höchsten, hellsten Tönen. [...] Tiefen Tönen ist eine gewisse Breite und Verschwommenheit eigen, mit der sie uns wie ein Medium umfassen, während die hohen un stetig immer spitzer werden [...]. Wenn [Hermann von] Helmholtz die tiefen Töne ‚dumpf‘ nennt, so ist damit vielleicht ihre Dunkelheit und ihr Volumen gleichzeitig gemeint“ (Stumpf, 1926, S. 396). „Auf jeder Tonhöhe, abgesehen von den höchsten Lagen (etwa von c^3 ab), gibt es eben relativ scharfe, mittelscharfe, weiche und sehr weiche Klänge“ (Stumpf, 1926,

S. 403). Diese Differenzierung und die entsprechenden Begriffe finden sich im Laufe der Jahre wieder, zum Beispiel bei Neppert, der Schallqualitäten auf einer Skala von dunkel bis hell einordnet und als zusätzliche Beschreibungen zu dunkel noch „stumpf, dumpf, warm, o.ä.“ und für hell noch „scharf, spitz, o.ä.“ anführt (Neppert, 1999, S. 65).

Eine aktuelle Untersuchung zur Klangqualität, in der auditive Merkmale der Stimmqualität auf Basis mehrerer Hörer betrachtet werden, kommt aus der Sprechwissenschaft. Kranich (2014a, 2014b) untersuchte Beschreibungen der Klangqualität gesprochener Sprache und versuchte Begriffe in mehreren aufeinanderfolgenden Schritten zu evaluieren. Das Ziel war die „Erstellung und Evaluation eines Klangqualitätsinventars gesprochener Sprache“ (Kranich, 2014a, S. 40), welches in seiner finalen Version in Tabelle 4 dargestellt ist.

	Merkmal vorhanden	Teilweise vorhanden	Nicht vorhanden
Klangfülle			
Gespanntheit			
Helligkeit			
Knarren			

Tabelle 4: Fragebogen zur Bewertung der Klangqualität gesprochener Sprache nach Kranich (2014b).

Bei der Bewertung dieser Ergebnisse muss berücksichtigt werden, dass Kranichs Inventar auf Expertenmeinungen beruht, d.h. die Merkmale sind auf Basis einer Expertengruppe aufgestellt worden und dann Studierenden der Sprecherziehung zur weiteren Evaluation vorgelegt worden (andere Gruppen wurden nicht weiter definiert). Letztere bezeichnet er auch als ‚Nicht-Experten‘, die allerdings als semi-professionelle Hörer bezeichnet werden können, da sie bereits mit dem Gegenstand Stimme vertraut sind und explizit Begriffe zur Stimmbeschreibung gelernt haben, was bei Laien nicht der Fall ist.

Es ist auffällig, dass eine unipolare Skala verwendet wurde, was Kranich damit begründet, dass Klangqualitätsmerkmale wahrscheinlich eher skalar als binär aufzufassen seien. Damit stellt Kranich nur Merkmale ohne ihre Ausprägungen dar, z.B. nicht Klangfülle mit den Polen klangvoll-klangarm, sondern sagt, dass Klangfülle vorhanden oder nicht vorhanden ist. Diese Evaluation scheint kognitiv recht anspruchsvoll und zum Teil nicht eindeutig zu sein: Was bedeutet es, wenn Helligkeit ‚nicht vorhanden‘ ist? Außerdem wird auch in der folgenden Gruppenbefragung deutlich, dass eine Mittelkategorie in der Bewertungsskala „als ein Zeichen der Unsicherheit“ interpretiert werden kann (Kranich, 2014b, S. 29). Mit Chi-Quadrat-Tests wurden die Verteilungen über die 2-3 Antwortstufen für jede einzelne Aufnahme verglichen, die, trotz Auffälligkeiten, zu einer Annahme der bestehenden Skalen führten (Kranich, 2014b, S. 34f.). Kranichs Fazit lautet: „Für eine Beurteilung scheint demnach eine internalisierte Norm vorhanden zu sein, wonach für Klangfülle, Gespanntheit und Helligkeit eine mittlere Stufe ausschlaggebend ist. Fällt ein Klangqualitätsmerkmal nach einer Seite auf, so ist es markiert und wird damit einheitlicher bewertet. Für das Merkmal Knarren ist hingegen eine nicht

knarrende Klangqualität die Norm“ (Kranich, 2014b, S. 38). Überprüft wurden diese auditiven Merkmale mit einer Formant- und Spektral-Analyse und es zeigten sich visuell „einige reproduzierbare Merkmale [...] besonders für Klangfülle und Helligkeit“ (Kranich, 2014b, S. 38).

Methodisch interessant ist Kranichs Vorgehen zur Überprüfung der Urteilerübereinstimmung, die er in den ersten Studien mit einem eigens kreierten „Gewissheitsgrad“ (1 = hoch, 0 = niedrig; Kranich, 2014b, S. 27f.) berechnet. Dieser zeigt, dass sich die Experten und auch die sogenannten Nicht-Experten eher uneinig sind (der Gewissheitswert hat ein Maximum von 0,57). Die Ursache für die Uneinigkeiten liegt wohl am ehesten in der Wahl des Untersuchungsmaterials (die Interjektion /hm/), worauf Kranich hinweist und eine Überprüfung mit längeren Aussagen anschließt. In den folgenden Studien wird ein anderes Studiendesign verwendet, welches paarweise Vergleiche von Aussagen beinhaltet und somit erwartungsgemäß zu einer besseren Übereinstimmung führt (vgl. Gerratt, Kreiman, Antonanzas-Barroso & Berke, 1993).

Natürlich ist die Stimulus-Auswahl entscheidend für die Wahl der Merkmale – in Kranichs Studie schied deshalb die Nasalität schon früh aus und die Merkmale rau und behaucht wurden u.a. deshalb entfernt, da sie im RBH-System enthalten sind. Unter diesem Aspekt ist die Zielsetzung, dass das Erhebungsinstrument „wesentliche Aspekte dieser prosodischen Eigenschaft“ (Kranich, 2014b, S. 21) umfassen und mit möglichst wenigen, aber relativ gut bestimmbar Dimensionen auskommen sollte, nur teilweise erfüllt worden. In letzter Instanz hätten die Geräuschanteile bei dieser Zielsetzung wieder in das Inventar integriert werden müssen. Die Studie zeigt in letzter Konsequenz nicht, ob diese ausgewählten Merkmale in ihrer Dimensionalität ausreichen, um wirklich als ‚wesentliche Aspekte‘ bezeichnet zu werden.

Welche Aspekte der Stimmbeschreibung als wesentlich gelten, hängt natürlich stark davon ab, welches Ziel eine Untersuchung verfolgt. Ketzmerick (2007), zum Beispiel, führte eine *auditive und apparative Charakterisierung von Stimmen* durch und verglich 15 Stimmen auf Basis von sechs Merkmalen mithilfe von Korrelationsanalysen und visuellen Eindrücken, wobei die Merkmale Stimmlage, Sprechtempo, Stimmvolumen und Stimmklang am besten zur Charakterisierung beitrugen. In den Studien von Bänziger et al. (2014) und Hellbernd und Sammler (2016) galten die Merkmale als wesentlich, auf deren Basis zwischen bestimmten Emotionen bzw. Intentionen in der Sprache unterschieden werden konnten, was mithilfe von Diskriminanzanalysen untersucht wurde (vgl. Kapitel 2.4). In der vorliegenden Arbeit ist u.a. die charakteristische Beschreibung von Stimmen Ziel der Untersuchungen, wozu ebenfalls Diskriminanzanalysen zur Überprüfung des entwickelten Inventars herangezogen werden (Kapitel 5.2.7 und 6.2.4).

2.2.3 Reliabilität bei der Bewertung vokaler Merkmale

Die genannte Schwierigkeit in Bezug auf die Beschreibungsebene stimmlicher Merkmale führt zum Problem der Reliabilität in der Bewertung solcher Merkmale. Es ist naheliegend, dass kleinere Inventare, die sich auf drei Merkmale beziehen und sich auf eine Kategorie der Beschreibung (z.B. Geräuschhaftigkeit im Stimmklang) beziehen, erst einmal zu einer stärkeren Reliabilität führen als größere mit bis zu 20 Merkmalen aus verschiedenen Kategorien (vgl. Webb et al., 2004, die eine hohe Reliabilität für die GRBAS-Skala ermittelten, aber eine niedrige für das VPA und das *Buffalo Voice Profile*). Je nach Anwendungs- und Forschungsbereich ist eine hohe Inter- und Intrarater-Reliabilität mehr oder minder erstrebenswert und muss bei der Erstellung des Inventars berücksichtigt werden. Dazu gehört nicht nur die Wortwahl, sondern auch die ausgewählte Skala. Die Interrater-Reliabilität wird in der perceptiven stimmlichen Beurteilung als problematisch angesehen. Kreiman und Gerratt (1998) zeigten, dass das „interrater agreement [...] cannot reflect variations or patterns of agreement for specific voice samples“ (Kreiman & Gerratt, 1998, S. 1598), und berechneten einen Prozentsatz der übereinstimmenden Ratings für alle Stimmen und alle Teilnehmer innerhalb zweier Skaleneinheiten (Kreiman, Gerratt & Ito, 2007). In anderen Studien zeigten sie, dass Stimmen mit Ankerbeispielen hinsichtlich Rauigkeit zuverlässiger als ohne bewertet werden (Kreiman, Gerratt, Kempster, Erman & Berke, 1993). Hier wird deutlich, dass ohne Ankerpunkte weniger genaue internalisierte Standards zur Beurteilung herangezogen werden, die die Ratings (und somit die Reliabilität) beeinflussen, als mit Ankerbeispielen. Im systematischen Vergleich verschiedener Studien-Designs zur Beurteilung von Stimmqualität (hier Behauchtheit und genereller ‚noise‘) stellen die Autoren fest: „interrater variability is an issue of task design, not of listener unreliability“ (Kreiman et al., 2007, S. 2354). Neben einer grundlegenden Ablehnung verbreiteter Reliabilitätsmaße (Cronbachs Alpha, Cohens Kappa und Intra-Class-Correlation, ICC) sei es vom Studiendesign abhängig, welche Maße zur Überprüfung der Übereinstimmung herangezogen werden (und ob sie überhaupt sinnvoll sind).

Andere Möglichkeiten der Überprüfung zeigten sich in der dargestellten Untersuchung von Kranich (2014b) und auch in der vorliegenden Arbeit wurden, je nach Studie und Ziel der Untersuchung, unterschiedliche Methoden angewendet, welche die Reliabilität der Ergebnisse erhöhen sollten. Dabei spielt vor allem eine Rolle, ob die Interrater-Reliabilität des ganzen Inventars oder einzelne Merkmale untersucht werden sollen. So eignen sich paarweise Korrelationen der Teilnehmer-Ratings (Merrill & Larrouy-Maestri, 2017; Merrill, 2019), aber auch Verteilungen der Daten sowie Abweichungen vom mittleren Wert einer Skala, die zur Interpretation von Übereinstimmung und Einheitlichkeit der Ratings herangezogen werden können (Kapitel 5.2.4 und 6.2.3).

Aufgrund des gezeigten Trainingseffektes bei der Beurteilung von Stimmen musste auch der Einfluss der Expertise beachtet werden, da in den vorliegenden Studien Experten und Nicht-Experten untersucht wurden. Bisher zeigten vor allem die Studien eine gute Reliabilität, die sowohl

Expertenevaluationen als auch Inventare mit wenigen Items umfassten. Allerdings konnte nach einem Training in der Anwendung des VPA (Laver, 1980) und dem *Vocal Profile Analysis Scheme* (Shewell, 1998) eine höhere Reliabilität erzielt werden. Für Laien zeigte sich außerdem, dass die Nutzung der RBH-Skala gelernt werden kann (Anders et al., 1988; Nawka & Anders, 1996). Auch konnte die GVPS (Bänziger et al., 2014) mit Laien ohne vorheriges Training entwickelt werden, weswegen dieser Ansatz auch für die vorliegenden Arbeit gewählt wurde.

2.3 Singstimmen

Besonders Singstimmen im künstlerischen Kontext sind Gegenstand der Bewertung durch Experten. Es gibt anscheinend bisher keine Inventare, die Merkmale der Singstimme zur Bewertung durch Nicht-Experten beinhalten. Das liegt wohl daran, dass der Fokus auch hier bisher auf der klinisch-pädagogischen Arbeit lag. In Anknüpfung daran schreibt Richter zum analytischen Hören und der Hörschulung: „Jeder an der Beurteilung oder Betreuung von Stimmen Beteiligte muss sein Gehör intensiv schulen. Diese Schulung erfordert ein jahrelanges geduldiges Training. [...] Das Hören sollte so geschult werden, dass es analytisch die drei Ebenen des Instruments Stimme, nämlich Tonanregung (Atmung), Ton-/Klang-Produktion (Kehlkopf) sowie Ton-/Klangformung (Resonanzräume) [...] so differenziert wie möglich unterscheiden kann. [...] Die Fragen der hörenden Stimmbeurteilung sollten immer hierarchisch in folgender Reihenfolge ablaufen: Erstens, was hören wir? Und, zweitens, wie wird dieser Klang vermutlich erzeugt und modifiziert? Erst dann sollte die Frage gestellt werden: Wie sind die einzelnen Elemente zu bewerten im Sinne von stilistisch passend/nicht passend oder richtig/falsch oder unschädlich/schädlich oder auch gesund/krank. [...] In diesem Höreindruck sollten [...] Aspekte der Tonbildung (Stimmeinsatz und -absatz), der Tonqualität, des Stimmklangs (z.B. Timbre), der Tragfähigkeit, der Intonationsfähigkeit, der Registerreinheit und auch des Vibratos einfließen“ (Richter, 2014, 66f.). Während einige Merkmale an bisherige Stimmbeurteilungen erinnern, beziehen sich andere hier bereits auf Merkmale des Singens (Intonation und Vibrato). Auch geht die Bewertung über das bisherige Kontinuum zwischen normal und gestört hinaus und fragt, ob die Stimme stilistisch passend bzw. als richtig und falsch bewertet wird (wahrscheinlich beim Singen u.a. die Intonation betreffend).

Einige der bisher vorgestellten Inventare umfassen auch Merkmale, mit denen die (gestörte) Singstimme beschrieben werden kann. So wird z.B. das Inventar von Shewell (2009) für den generellen Einsatz der Stimme im künstlerischen Kontext verwendet (siehe Kapitel 2.1.2). Viele Phoniater behandeln nicht nur die gestörte Sprech-, sondern auch die Singstimme bzw. Sängerstimme (der Unterschied liegt im Training: die Singstimme ist die untrainierte Stimme, die Sängerstimme die trainierte; vgl. Seidner & Wendler, 2010). Probleme der Sängerstimme werden vorwiegend im klassischen Fach beschrieben, da die Stimme in der populären Musik bisher im klinischen Bereich eher

unterrepräsentiert ist bzw. ihr in der Phoniatrie auch gerne von vorneherein eine inhärente Störung zugewiesen wird (Kapitel 2.5.2). Phonetisch-akustische Vergleiche zwischen klassischen und populären Gesangsstilen zeigten, dass z.B. im Country oder Western die Spannung im Rachen den Sängerformanten eliminiert (eine Frequenzverstärkung, die zu einer verstärkten Tragfähigkeit führt; Mathieson, 2001, S. 566ff.; Sundberg, 1974, 1987). Generell wird gezeigt, dass die Sänger populärer Musik die Spannung im Bereich des Kehlkopfes erhöhen, um zum Beispiel die gewünschte Lautstärke zu erzeugen (im Gegensatz zum klassischen Gesang, der die Resonanzräume anders ausnutzt). Allerdings nimmt diese Spannung durch Training wieder ab (vgl. Mathieson, 2001, S. 567).

Es gibt mehrere Ansätze, die die Merkmale der Singstimme beschreiben, einerseits aus dem vokalpädagogischen Bereich, andererseits aber auch im Rahmen wissenschaftlicher Untersuchungen. So versuchten sich Ekholm, Papagiannis und Chagnon (1998) daran, das Problem der Kommunikation zwischen Vokalpädagogen und Wissenschaftlern hinsichtlich der Beschreibung von klassischen Singstimmen zu erleichtern. Erprobt wurden die Kriterien ‚resonance/ring‘, ‚color/warmth‘, ‚clarity/focus‘ und ‚appropriate vibrato‘, die für die Wissenschaft mit entsprechenden akustischen Korrelaten versehen wurden. Auch Henrich et al. (2008) haben auf der Basis eines multidisziplinären Ansatzes ein ‚listening sheet‘ zur Beschreibung der Stimmqualität im ‚Western Lyrical Singing‘ erstellt. Unterteilt ist der Evaluationsbogen in zwei Abschnitte: „Perception of vocal gesture or vocal technique“ (Henrich et al., 2008, S. 77) und „Perception of sound“ (Henrich et al., 2008, S. 78). Der erste Teil enthält Parameter zu Atmung, Vibrato, Einsätzen und Absätzen („balanced, breathy, glottal, glottal strong“), zu ‚glides‘ am Anfang und am Ende, intonatorischen Aspekten und zum Rhythmus. Außerdem enthält er Fragen zum ‚placement‘ der Stimme, ob eher vorn oder hinten platziert, offen oder gedeckt, nasal und Twang, gepresst oder ‚breathy‘ phoniert wird. Zusätzlich wird nach „close to speech“ und „close to singing“ gefragt (Henrich et al., 2008, S. 77). Der zweite Teil enthält phonetische Aspekte, die die Vokal- und Konsonantbehandlung thematisieren (nah und kontrastiert, erkennbar / nicht erkennbar, kurz und lang), außerdem suprasegmentale Aspekte die Phrasierung betreffend und ein Merkmal zur Verständlichkeit (graduell ja-nein), wobei einige dieser Merkmale eher für Muttersprachler relevant seien. Weitere Bereiche fallen unter „sound intensity and pitch“ und umfassen „pitch perception, loudness assessment“ (Effizienz, Power, Sängerformant) und „voice range possibilities“ (die dynamische Gestaltung tiefer und hoher Töne; Henrich et al., 2008, S. 78). Die Merkmale zu „sound colour“ fragen das Timbre mit den Items „high pitched-low pitched, with ring-without ring, balanced-unbalanced“ und „homogeneous-inhomogeneous“ (alle auf einer 5-stufigen Skala) sowie zusätzlich mit „dark“ und „light“ (ohne Skala) ab. Fast alle Parameter werden auf einer 5-stufigen-Skala bewertet, einige zusätzlich mit der Option N/A („not applicable“) (Henrich et al., 2008, S. 78).

Natürlich kommt es dabei zu Überschneidungen mit den Beschreibungen der Sprechstimme, was die Verwandtschaft zwischen der Sprech- und der Singstimme betont: Bewertungskriterien sind Klangqualitäten (die ebenso Behauchtheit mit einbeziehen), artikulatorische Aspekte, Stimmlage und Umfang. Auch in der Ausbildung der Stimme für einen Beruf können Vergleiche in der Behandlung von Sängern und Schauspielern hergestellt werden, da in beiden die Tragfähigkeit der Stimme eine große Rolle spielt. Eine Besonderheit beim Singen sind hingegen resonatorische Aspekte, die hier als ‚placement‘ bezeichnet werden, sowie der Sängerformant (wobei Nawka, Anders, Cebulla & Zurakowski, 1997 auch den Sprecherformanten belegen konnten). Aspekte der Rauigkeit finden sich im klassischen Gesang nicht, sind aber wichtiges Merkmal populärer Gesangsstile (Kapitel 2.3.1).

Ein Extrem der Beschreibung klassischer (Opern-)Stimmen findet sich bei Kesting (2008a, S. 2494–2499). Erläuterungen zu gesangstechnischen Begriffen erstrecken sich hier auf fast sechs Seiten, die auch historisch geprägte Begriffe (aus dem Italienischen) beinhalten, die auf das Beschreiben einer historisch-informierten, aufführungspraktischen Gestaltung des Gesangsvortrags ausgerichtet sind.

Für die populäre Musik gibt es im Gegensatz dazu einige wenige Gesangsschulen, die Anleitungen dafür geben, wie gewisse Stile gesungen und welche technischen Aspekte hierbei berücksichtigt werden sollten. Zu nennen sind die *Komplette Gesangstechnik* (engl. *Complete Vocal Technique*, CVT; Sadolin, 2009) und *Sing Anything – Mastering Vocal Styles* (Latimerlo & Popeil, 2012). Hier wird allerdings weniger auf die Beschreibung klanglicher Ergebnisse gesetzt als auf die Beschreibung anatomisch-physiologischer Einstellungen, die vorgenommen werden sollten, um ein bestimmtes klangliches Ergebnis zu erhalten. Besonders die Gesangstechnik von Sadolin erfährt in den letzten Jahren Unterstützung aus der phoniatischen Forschung, die die technischen Aspekte mit physiologischen Vorgängen in Verbindung bringt (McGlashan, Thuesen & Sadolin, 2017; Thuesen, McGlashan & Sadolin, 2017).

2.3.1 Untersuchungen zur Singstimme in der populären Musik

Federführend in der Beschreibung von Singstimmen in der populären Musik im deutschsprachigen Raum ist die Arbeit der Gruppe um Pfeleiderer zum *Vokalen Ausdruck in der populären Musik der USA, 1900-1960* (Pfeleiderer et al., 2015), die einen musikwissenschaftlichen Zugang wählt und popular-musikalischen Gesang hinsichtlich kultureller und identitätsbildender Aspekte untersucht. Hänel entwickelt darin, so beschreibt es Pfeleiderer, „Perspektiven einer deskriptiv-analytischen Terminologie der vokalen Gestaltung populärer Musik“ (Pfeleiderer et al., 2015, S. 18). Es wird zuerst deutlich gemacht, dass sich der vokale Ausdruck nicht nur auf das Singen bezieht, sondern die populäre Musik Ausdrucksmittel verwende, „die sich einer klaren Zuordnung zum Singen entziehen. In populärer Musik wird gebrüllt, geschrien, gejammert, genäsel, gesäuselt und geflüstert, manche Stimmen sind heiser, manche dünn und zittrig“ (Hänel, 2015, S. 53). Auch stellt Hänel heraus, dass der Gesang populärer

Musik lange nicht als Gesang akzeptiert wurde und immer noch Gegenstand ästhetischer Debatten sei (vgl. Hähnel, 2015, S. 53). Er analysiert verschiedene Stimmen populärer Musik und arbeitet wichtige Merkmale des vokalen Ausdrucks heraus. Als eine der wichtigsten Klangeigenschaften stellt er den Twang heraus, der mit Näseln übersetzt werden kann. Allerdings beinhaltet der Twang auch eine Klangschärfe, da der Twang mit einer Enge im Kehlkopf (einer Art Trichter; vgl. Sadolin, 2009) erzeugt wird, wodurch sich das Klangspektrum ändert und der deutsche Begriff des Näsels nicht alle Aspekte abdeckt. Als nächsten Aspekt bespricht Hähnel (2015) das laute Singen. Während Lautstärke (wie schon oben beschrieben) in der klassischen Musik über die Resonanzräume erreicht wird, wird sie beim Singen populärer Musik durch einen höheren subglottischen Luftdruck erzeugt und somit durch höhere (oft auch hörbare) Spannung. Als Beispiel der Ausdrucksformen werden das Rufen bzw. Shouting und das Belting genannt, was vor allem in höherer Lage genutzt und hörbar wird. Auch beim leisen Singen gibt es spezielle poplarmusikalische Ausdrucksmittel, zu denen das Crooning (Flüstern) sowie das Moaning, welches ein Stöhnen, Jammern, Klagen und Seufzen darstellt, gehören. „Beim Moaning wird im Legato und melismatisch, meist mit geschlossenen Lippen, auf einer absteigenden Tonfolge gesungen oder auch nur gebrummt beziehungsweise gesummt. Damit verbunden werden häufig die Begriffe Groaning (engl. groan = stöhnen) und Crying, wobei Letzteres nicht auf eine geringe Lautstärke verweist, sondern auf eine brüchige Stimme, die mitunter auch hörbare Registerwechsel beinhaltet“ (Hähnel, 2015, S. 60f.). Der Registerwechsel kann ebenfalls als ein eigenes Ausdrucksmittel genannt werden und umfasst beispielsweise das Jodeln, kann aber auch zum Ausdruck des Heulens oder Jammerns eingesetzt werden. Auch Rauigkeit und Behauchtheit spielen eine wichtige Rolle in der populären Musik, was besonders im Kontext klinischer Forschung und der oben beschriebenen Inventare zur Beschreibung dieser ‚pathologischen‘ Merkmale thematisiert wurde. Die Gesangsweisen des Growling und Screaming zeichnen sich ebenfalls durch eine sehr raue Stimmgebung aus. Diese Techniken und Merkmale stehen im Gegensatz zum klassischen Gesang, in dem die vokalen Ausdrucksmittel in Form einer „legitimen“ Gesangstechnik normiert wurden (Hähnel, Marx & Pfeleiderer, 2014, S. 3). „War bisher eine ‚reine‘ oder ‚klare‘ Stimme das Ideal, so ähnelt der Stimmklang in der populären Musik einerseits dem der Sprechstimme, andererseits wird aber auch ein großes, nicht zuletzt die Extreme suchendes Spektrum an stimmlichen Ausdrucksformen vom Flüstern über das Singen bis zum Schreien ausgelotet, wobei der regelmäßige Obertonaufbau durch zusätzliche Manipulationen gestört wird: Rauheit und Behauchtheit sind gängige Praxis. Vormalig als Stimmfehler pathologisierte Charakteristika wie Jitter, Shimmer oder Twang wurden für einige Genres Stil prägend und über diesen Weg salonfähig“ (Hähnel, 2015, S. 71).

Für diese herausgearbeiteten Ausdrucksmittel wurden in späteren Analysen Belege herangeführt. In einem methodisch ausgerichteten Artikel aus derselben Arbeitsgruppe wurden Analysetools vorgestellt, mit denen die Höreindrücke überprüft werden können (Hähnel et al., 2014, S. 19). Hier

werden neben der rauen Stimmgebung noch weitere Merkmale des Singens populärer Musik wie die klangliche Gestaltung von Vokalen, das individuelle Vibrato und das Gleiten zwischen Tonhöhen genannt. Auch wird noch einmal die Sprechnähe verdeutlicht: „So liegt eine Besonderheit des populären Stimmgebrauchs in der Annäherung an die alltägliche Sprechstimme, im Wechsel zwischen gesprochenen und gesungenen Wörtern oder Silben, oder aber in einem fließenden Übergang zwischen Sprechen und Singen. Dies zeigt sich u.a. im Gleiten zwischen den Tonhöhen, einem freien Umgang mit dem Metrum, das im vokalen Vortrag gedehnt, gestaucht oder einfach ignoriert wird, sowie in einem individuell variierenden ‚natürlichen‘ Vibrato. Bei diesen alltäglichen und sprechnahen Qualitäten des Gesangs geht es nicht nur um einen hohen Grad der Textverständlichkeit — der de facto auch bei Popstimmen nicht immer vorhanden ist. Vielmehr wird hierdurch eine individuelle, nicht selten emotional aufgeladene Gestaltungsweise möglich, die zur Grundlage eines charakteristischen Personalstils und damit der Wiedererkennbarkeit eines Sängers werden kann“ (Hähnel et al., 2014, S. 1f.).

Für die vorliegende Arbeit ist an diesen Erörterungen sowohl die dargestellte Sprechnähe des Singens von Bedeutung als auch die Aufstellung wichtiger Merkmale zur Beschreibung des Singens populärer Musik, welche für ein zu entwickelndes Inventar berücksichtigt werden müssen. Es zeigen sich bereits direkte Überschneidungen mit Sprechstimm-Inventaren, wie die Evaluation von gleitenden Tonhöhenveränderungen, individueller Artikulation und dem ‚natürlichen‘ Vibrato. Hier muss nach weiteren Übereinstimmungen gesucht werden, um sicherzustellen, dass die nötigen Beschreibungsebenen im Fragebogen enthalten sind, und zugleich leicht verständliche Begriffe für die beschriebenen Merkmale gefunden werden, die dann auch eine Beschreibung von Stimmen durch Nicht-Experten ermöglichen (vgl. Kapitel 2.6.1).

2.4 Stimme und Emotion

Inwiefern Musik in der Lage ist, Emotionen einerseits überhaupt auszudrücken, andererseits beim Hörer auszulösen und in welchem Zusammenhang diese mit der Sprache stehen, ist Gegenstand historischer Debatten und Forschungen, die bis zum heutigen Tage anhalten. Allerdings finden sich nur wenige Ansätze, die den Gesang in seinem emotionalen Ausdruck und seiner Wirkung betrachten. Deshalb werden im Folgenden einige einschlägige Studien zu Emotionen in Musik und Sprache dargestellt, bevor Verbindungen zum Gesang hergestellt werden. (Eine sprechwissenschaftliche Abhandlung zum „sprechstimmlichen Emotionsausdruck und -eindruck“ findet sich bei Bose, 2010, Kapitel 3.3).

Eine einflussreiche Studie ist die von Banse und Scherer (1996), die die vokalen Ausdrucksmuster verschiedener Emotionen der gesprochenen Sprache in ihrer akustischen und auditiven Darstellung auf den Ebenen Intensität und Valenz beschreibt. Die Emotionen umfassen u.a. ‚happiness‘, ‚anxiety‘,

‚shame‘, ‚pride‘, ‚sadness‘, ‚disgust‘, aber auch ‚interest‘ und ‚elation‘. Getestet wurden emotionale Äußerungen von Schauspielern. In anderen Studien wurden unterschiedliche Aspekte dieses Beitrags mittels verschiedener Methoden weiter vertieft. Unter anderem wurden phonetische Untersuchungen angestellt, indem physiologische Vorgänge mit einem Glottisspektrum und einem Laryngogramm abgebildet und die Steuermechanismen des Larynx für die untersuchten Basisemotionen analysiert wurden; Weiterhin wurden Besonderheiten in der Artikulation und temporale Parameter dargestellt (Klasmeyer, 1999; Klasmeyer & Sendlmeier, 2000; Sendlmeier, 2002, 2016).

Neben der bereits genannten GVPS (Bänziger et al., 2014; siehe Kapitel 2.2.2) ist eine Studie von Hellbernd und Sammler (2016) hervorzuheben, die keine Basisemotionen, sondern, basierend auf der Sprechakt-Theorie, folgende Intentionen untersuchten: ‚criticism‘, ‚doubt‘, ‚naming‘, ‚suggestion‘, ‚warning‘ und ‚wish‘. Mit einer Diskriminanzanalyse wurden diese Sprechakte auf Basis akustischer Merkmale und auditiver Beschreibungen klassifiziert. Dies gelang auf beiden Ebenen sehr gut, auch konnten die Sprechakte auf den Dimensionen ‚valence‘ und ‚arousal‘ abgebildet werden, was bedeutet, dass auch eine emotionale Komponente zwischen den Sprechakten unterschied.

Weiterhin wurden die Emotionen der gesprochenen Sprache mit den Emotionen in der Musik verglichen. Ein einflussreicher Beitrag stammt von Juslin und Laukka (2003), die zu folgendem Schluss kommen: „In particular, we predict that future research will confirm that music performers communicate emotions to listeners by exploiting an acoustic code that derives from innate brain programs for vocal expression of emotions. In this sense, at least, music may really be a form of heightened speech that transforms feelings into ‘audible landscape’” (Juslin & Laukka, 2003, S. 805). Der Gesang nimmt dabei eine Sonderstellung ein, da er andere Ausdrucksmöglichkeiten als ein Musikinstrument hat: Die Stimme kann sich Mitteln der Musik und der Sprache bzw. des Sprechens bedienen.

In diesem Sinne befassten sich Parada-Cabaleiro et al. (2017) mit den ‚musical features‘ im Gesang und untersuchten, welche elektronisch manipulierten Merkmale der Singstimme (‚rhythmic-melodic contour‘, ‚musical syntax‘ und ‚tempo‘) in der Lage sind, Emotionen auszulösen. Sie zeigen, dass das wahrgenommene Erregungslevel im A-cappella-Singen nicht von den untersuchten ‚musikalischen Parametern‘ abhängt. Die Wahrnehmung der Valenz allerdings steht in Zusammenhang mit den Parametern.

Zum emotionalen Ausdruck im Gesang beschreibt Klasmeyer (1999, 27ff.) unter Bezug auf Rapoport (1996) verschiedene stimmliche Mechanismen, mit denen emotionale Erregung im Gesang erzeugt werden kann. Dazu gehört der behauchte Stimmeinsatz, der Sängersformant, das Vibrato, die Tonhöhenbewegungen (wozu das Glissando, ein gleitender Tonwechsel, und das Sforzando, ein abrupter Anstieg der Tonhöhe von einer Note zur nächsten, gehören) und der Glottisschlag (wohl im

Sinne eines harten Einsatzes gemeint, vgl. Anders, 2012). Rapoport (1996) untersuchte Passagen aus der Oper und dem Kunstlied und typische stimmliche Mechanismen für neutrale, ruhig-traurige und ausdrucksvoll-expressive Abschnitte: Während neutrale Passagen keinen auffälligen Stimmgebrauch aufweisen, zeigen ruhige, entspannte oder traurige Abschnitte lange, stationäre Töne, zu deren Beginn leise phoniert wird (auch oft behaucht). Im mittleren Verlauf des Tons wird die Stimme kräftiger und im letzten Drittel des Tons kommt das Vibrato hinzu. Für ausdrucksvoll-expressive Passagen findet sich bereits im Beginn des Tons ein Vibrato und ein gradueller Anstieg der Tonhöhe, welcher dem Hörer den Eindruck von Bewegtheit in der Stimme vermittelt. Klasmeyer diskutiert die Funktion des Vibratos als ein Stilmittel, das anscheinend häufiger in Passagen mit höherer Erregung vorkommt, gleichzeitig aber nicht spezifisch für die Valenz zu sein scheint. Möglicherweise sei der Einsatz des Vibratos lediglich eine Strategie, das Orchester zu übertönen.

Klasmeyer untersucht auch die akustischen Korrelate des emotionalen Ausdrucks und findet Verbindungen zwischen Sprache und Gesang: Die Behauchung der Stimme mit starker spektraler Dämpfung dient als Ausdrucksmittel für Ruhe und Trauer, die Betonung höherer Teiltöne als Zeichen höherer Erregung (Freude, Ärger, Angst), die Tonhöhenänderungen (Glissando) als Ausdruck von Freude und Ärger sowie Perturbationen der Grundfrequenz (Jitter, wahrgenommen als Rauigkeit) bei Angst und Trauer. Das Vibrato des Gesangs kommt im Sprechen nicht vor und kann nicht verglichen werden (vgl. Klasmeyer, 1999, S. 223).

Es kann festgehalten werden, dass die Untersuchungen zu den Ausdrucksmitteln im Gesang und ihrer Wirkung auf den Hörer nicht besonders weit gediehen sind. Es fehlen nicht nur Untersuchungen zur Beschreibung und ästhetischen Bewertung verschiedener vokaler Merkmale im Gesang, sondern auch zur Unterscheidung vom Emotionsgehalt der Musik und den im Hörer ausgelösten Gefühlen.

2.4.1 ‚Perceived and felt emotions‘ und ästhetische Emotionen

Es ist auffällig, dass in der Emotionsforschung der Sprache primär über die Ausdrucksbeschreibung der Basisemotionen geforscht wird. Dabei konnte gezeigt werden, dass der Hörer die Emotionen wahrnimmt und korrekt beschreiben kann (und anscheinend auch körperlich nachvollzieht, wie bildgebende Verfahren zeigen, z.B. in der Musik, Koelsch, 2014). Es ist allerdings weniger davon auszugehen, dass ein Studienteilnehmer von einer gehörten ängstlichen Äußerung wirklich Angst bekommt. Gerade in der Kunst allerdings kann sich nicht nur auf Basisemotionen beschränkt werden und es ist von Interesse, was im Hörer/Zuschauer ausgelöst wird. Im populärwissenschaftlichen Bereich der ästhetischen und künstlerischen Forschung findet sich oft der Begriff der ‚emotional intelligence‘ des Hörers, von der angenommen wird, dass sie eine Voraussetzung für das Verstehen des sängerischen Ausdrucks ist (Chapman, 2017). Hier scheint es, als würde von einer 1:1-Beziehung von Ausdruck und Eindruck ausgegangen, dass also das Verstehen zum Nachfühlen und Selbstfühlen

wird. Es wird dabei zu wenig berücksichtigt, dass im Hörer Gefühle ausgelöst werden, die nicht unbedingt im direkten Verhältnis zum Ausdruck des Sängers stehen müssen – auch wenn dieser vom Hörer evtl. korrekt identifiziert werden kann. In der Musikforschung wird dieser Unterschied wesentlich stärker betrachtet, was möglicherweise auch deshalb der Fall ist, da der Emotionsausdruck in der Musik nicht so eindeutig ist wie oft in der Sprache. Um zwischen dem wahrgenommenen Ausdruck und dem ausgelösten Gefühl des Hörers zu differenzieren wird im Englischen der Begriff ‚perceived‘ oder ‚expressed emotion‘ zur Beschreibung des Ausdrucks verwendet und ‚felt emotion‘ zur Beschreibung des Gefühls, das ausgelöst wird. Diese müssen nicht in einem direkten, einseitigen Verhältnis zueinanderstehen. Bisherige Abhandlungen zum Unterschied (Evans & Schubert, 2008; Gabrielsson, 2002; Schubert, 2013) untersuchten die ‚perceived and felt emotions‘ am Beispiel von Basisemotionen und dem Dimensionsmodell (Valenz, Erregung und Dominanz), welche allerdings die für den künstlerischen Bereich wichtigen ästhetischen Emotionen unbeachtet lassen – wohl aus dem Grund, dass die ästhetischen Emotionen sich auf die ‚felt emotions‘ beziehen und die Entsprechung zu den ‚perceived emotions‘ schwer zu untersuchen ist. Speziell für die Evaluation für die ästhetischen Emotionen in der Musik ist die *Geneva Emotion Music Scale* (GEMS; Zentner, Grandjean & Scherer, 2008) als eines der wichtigsten Erhebungsinstrumente zu nennen, das im Kontext klassischer Musik evaluiert wurde. Später hinzugekommen ist die *GEneva Music-Induced Affect Checklist* (GEMIAC), die 14 sog. Gefühlsklassen beinhaltet („classes (or families) of feelings“, Coutinho & Scherer, 2017, S. 386). Ebenfalls zu nennen ist das Inventar AESTHEMOS (Schindler et al., 2017), das zur Beschreibung der ästhetischen Gefühle in verschiedenen Kontexten geeignet ist. Beide Inventare erheben das individuelle ästhetische Erleben und beinhalten eine Reihe von Zuständen, die beim Betrachten oder Hören eines ästhetischen Objektes entstehen können. Sie sind noch nicht direkt für die Stimme erprobt worden, sollten aber auch für Stimmen im Kontext künstlerischer Performances geeignet sein. Im Folgenden soll nun auf ästhetische Aspekte eingegangen werden, die die Stimme und das Singen und Sprechen betreffen.

2.5 Stimme und Ästhetik

In der Antike wurde als erste Qualität der Stimme des Redners, laut Foehr-Janssens, die Sanftheit (lat. *suavitas*) benannt: „der Stimmtone soll angenehm sein, nicht heftig oder scheltend“ (Foehr-Janssens, 2008, S. 131). Sanft und angenehm – diese Zustände bezeichnen bereits den Sprecher und den Hörer, denn wer, wenn nicht dieser, sollte die sanfte Stimme sonst als angenehm empfinden? Die nun behandelten ästhetischen Aspekte des Singens und Sprechens beziehen sich deshalb auf den Hörer und dessen Bewertung einer Stimme als ‚schön‘. In Abgrenzung zur schönen Stimme wird außerdem die gute Stimme diskutiert, in deren Beurteilung technische Aspekte hineinspielen.

2.5.1 Die schöne und die gute Stimme

„Manche Leute denken, es ist besser, hoch, laut, leise oder schön zu singen. [...] Nicht jeder hat eine Stimme zum Singen, wir können immer an einer Stimme mit einem guten Meister arbeiten, aber wir können nichts tun, wenn es nichts gibt“ (Bacilly, 1668, S. 37). Dies schreibt Bénigne de Bacilly im Jahr 1668 in seiner Monographie *Remarques curieuses sur l'art de bien chanter*, die einen starken Einfluss auf die Gesangstechnik in Frankreich hatte. Bacilly grenzt die schöne und die gute Stimme voneinander ab und beschreibt (hier zusammengestellt und übersetzt von der Autorin) die schöne Stimme („belle voix“) als angenehm für das Ohr wegen ihrer Schärfe und Weichheit (und vor allem die schöne ‚cadence‘; dt. Tempo, Rhythmus, Wohlklang u.a.) und im Gegensatz dazu die gute Stimme („bonne voix“) damit, dass sie nicht diese Sanftheit/Süße und natürlichen Wohlklang hat; allerdings wird sie durch ihre Kraft, ihre Festigkeit und durch die Bereitschaft, bewegt zu singen, nicht müde – das sei die Seele des Singens, zu der diese schönen natürlichen Stimmen normalerweise nicht fähig seien. Außerdem unterscheidet er zwischen der schönen und der hübschen Stimme („jolie voix“). Während die schöne Stimme große Harmonie und große Weite habe, sei die hübsche Stimme eine kleine Stimme (vgl. Bacilly, 1668, S. 38f.).

Während die schöne Stimme vor allem angenehm ist, wird deutlich, dass die gute Stimme etwas mit Training zu tun hat. Diese Eigenschaft wird der schönen ‚natürlichen‘ (und untrainierten) Stimme abgesprochen. Interessanterweise vertritt Bacilly (1668) den Standpunkt, die Güte der Stimme lasse sich verstandesmäßig, die ‚Schönheit‘ dagegen nur gefühlsmäßig beurteilen. Dieser Umstand wird auch Jahrhunderte später weiter bekräftigt, indem die schöne Stimme als etwas bezeichnet wird, was in fachwissenschaftlichen Abhandlungen nichts verloren hätte (Panconcelli-Calzia, 1961).

Allerdings finden sich dennoch einige wenige Versuche der Beschreibung der schönen Stimme, wie in dem folgenden Beitrag, der geschrieben wurde, als das Synchronisieren des Tonfilms als „neue Aufgabe“ beschrieben wurde (Biehle, 1955, S. 66). Der Autor formuliert auf drei Seiten „Grundsätzliches zur Stimmästhetik“ (Biehle, 1955, S. 66–69): „In der Gesangskunst steht die Forderung nach schönem Stimmklang obenan, deshalb auch das Fachwort *Bel canto*“ (Biehle, 1955, S. 66). Hier ist anzumerken, dass der Begriff des *Belcanto* nicht immer nur eine schöne Stimme beschrieben hat, sondern im ursprünglichen Wortgebrauch wohl den „richtigen“ oder „angemessenen“ Gesang beschrieb (Seedorf, 2016a, S. 75). Biehle schreibt, dass Schönheitsbegriffe zudem ungenau und abhängig vom zeitlichen Wandel und kulturellen Einflüssen seien. Der „schöne Ton“ (oder das schöne Timbre) sei zudem leider oft „mit einem Ansatzfehler“ gepaart. Er beschreibt für jede Stimmgattung diesen Ansatzfehler, der z.B. ein Knödeln beim Tenor oder ein kehliger, scharfer Ton für den Sopran ist. Damit werden klangliche Eigenschaften genannt, die vermutlich zur Zeit des Autors abgelehnt wurden. „Die genannten Ansatzfehler verhindern einen offenen, quellenden Ton und die Automatik der Kehle“ (Biehle, 1955, S. 68).

Weiterhin nennt er drei „Grundeigenschaften“, mit denen die „Stimmästhetik“ operieren müsse, „nämlich Reinheit, Ruhe, und Rundung bilden die Voraussetzungen zum echten Ton“ (Biehle, 1955, S. 67). Unter der Reinheit des Tons ist die absolute Intonationsreinheit zu verstehen. Mit der Formulierung „die Ruhe des Tons ist gegeben durch eine physikalische Eigenschaft, seine rhythmische Gliederung“ (Biehle, 1955, S. 67) wird das Vibrato beschrieben, dessen Schwingung laut Autor 5,4mal pro Sekunde erfolgen darf. Als drittes fordert er vermutlich etwas wie den Vokalausgleich: „Mit Rundung des Tones ist die Rundung des Vokals gemeint: eine nicht spitze oder flache, sondern runde Form“ (Biehle, 1955, S. 67). Außerdem wird vom Opernsänger Verständlichkeit erwartet. Die ‚ästhetischen‘ Forderungen, die den Ansatzfehler betreffen, gelten auch für die Sprechkunst, d.h. für den Schauspieler. Rein technisch müsse auch auf die Modulationsfähigkeit geachtet werden, dass z.B. bei voller Natürlichkeit auf „naturalistisches Kreischen oder Schreien“ umgeschaltet werden könne. Als das „tiefe Geheimnis“ der Sprechkunst benennt er, dass die Sprache einen „musikalischen Klang“ annimmt ohne in Singen zu verfallen (vgl. Biehle, 1955, S. 68). Der Autor beendet seine „stimmästhetischen Grundfragen“ mit dem Hinweis, dass „alle weiteren Erfordernisse“ sich hieraus „zwanglos“ ergäben (Biehle, 1955, S. 69). Es ist interessant, wie der Autor seine ästhetischen Klangvorstellungen mit einer gewissen Selbstverständlichkeit beschreibt. Dass sie heute noch einigermaßen nachvollziehbar sind, hängt damit zusammen, dass bestimmte technische Aspekte des klassischen Gesangs im Zentrum dieser Abhandlung stehen, die in Grundzügen auch heute noch ähnlich beurteilt werden, da der klassische Gesang seit langem normiert ist.

Es finden sich auch Abhandlungen zur Stimme im Allgemeinen wie eine Reihe von Publikationen von Panconcelli-Calzia aus den Jahren 1947-1961, die sich ausführlich mit der „Schönheit, der Güte und dem Wohlklang“ einer Stimme beschäftigen (Panconcelli-Calzia, 1961, S. 47f.). Er kritisiert die scharfe Trennung der schönen und der guten Stimme bei Bacilly, betont (1947) aber selbst, dass es wichtig sei, zwischen Schönheit und Güte der Stimme zu unterscheiden. Jedoch behauptet er, dass es falsch sei, in fachwissenschaftlichen Werken von Schönheit zu sprechen, da es sich um einen ästhetischen Begriff handele, wohingegen Güte ein physiologischer sei (Panconcelli-Calzia, 1947, S. 128ff.; Panconcelli-Calzia, 1961, S. 47). Panconcelli lehnt nicht nur den Begriff der Schönheit einer Stimme im Kontext fachwissenschaftlicher Werke ab, sondern auch die Beschreibung des Wohlklangs einer Stimme, der nicht anatomisch-physiologisch zu begründen sei.

Weiterhin kritisiert Panconcelli-Calzia, dass sich noch 1960 Bestrebungen finden, eine „‘ästhetisch-künstlerische‘ [!] Beurteilung einer Stimme auf Grund von ‚objektiven Kennzeichen‘ zu erzielen!“ (Panconcelli-Calzia, 1961, S. 47). Dieses Bestreben sei eine „Utopie“, da man kein Produkt der Einbildung („Figment“) objektiv begründen könne. Außerdem unterscheide sich diese Einschätzung zwischen den Kulturen. Beispielhaft nennt er einen Bericht, der eine gewisse Teiltonanzahl im Klang als objektive Kennzeichen beschreibt. Daraus schließt er, dass die „Schönheit kein einheitlich-gültiges

Kriterium für eine rationell-physiologische Stimmforschung“ sein kann, außerdem nimmt er zur Kenntnis, dass sich bisher keine Definition der schönen Stimme finden ließ – und dies auch vermieden wurde (Panconcelli-Calzia, 1956, S. 25). Er sagt, die Schönheit der Stimme sei „ausschließlich Kompetenz der Ästhetiklehre“ (Panconcelli-Calzia, 1956, S. 25). Dadurch, dass Stimmforscher unseres Kulturkreises eine normative Betrachtung der Schönheit einer Stimme einnahmen, kämen sie zu falschen Schlüssen, sprächen unberechtigterweise von der schönen oder der wunderbaren Stimme und ließen die Betrachtung der „guten“ Stimme aus. „Man hat ganz übersehen, daß eine schöne Stimme ebenso wenig gut sein kann, wie eine gute Stimme schön“ (Panconcelli-Calzia, 1956, S. 25).

Im Folgenden beschreibt er, wie sich die „Güte“ der Stimme feststellen lasse. Dies erfolge auf Basis des Gehörs, wobei nicht das periphere Hörorgan das wichtigste sei, sondern eine „akustische Empfindsamkeit, Bereitschaft und Aufmerksamkeit“ (Panconcelli-Calzia, 1956, S. 25). Diese Fähigkeit nennt er eine „Begabung“, die angeboren sei und sich entwickeln und verfeinern, aber nicht aneignen lasse. Er behauptet, Laien könnten darin erstaunlich gut sein, während Kapellmeister (und andere musikalische Menschen) dabei versagen, „weil sie die Stimme vorwiegend ästhetisch-künstlerisch zu betrachten und zu beurteilen pflegen“ (Panconcelli-Calzia, 1956, S. 25).

Panconcelli-Calzia hat mit seiner Publikation eine sogenannte „typische Fiktion“ erstellt, mit der er annimmt, „ein Kriterium gefunden zu haben, um eine Stimme als normal, als gut zu bezeichnen“ (Panconcelli-Calzia, 1956, S. 28). Er sagt, es sei über die Jahre ein bestimmtes Idealbild der guten, normalen Stimme entstanden. „Danach lässt sich eine Stimme als ‚gut‘ bezeichnen, wenn sie ausschließlich unter Inanspruchnahme der für die jeweilige Leistung nötigen Muskulatur in harmonischem Ausgleich der Atmungs-, Kehlkopf- und Ansatzrohrfunktion gebildet wird. Die ‚gute‘ Stimme hört sich frei von Nebengeräuschen, Druck, Dauer-Fehl-Überspannung an, klingt in jeder Höhe beliebig kräftig oder leise, weittragend, fließt resonanzreich, weich und anstrengungslos. Ein weiteres Merkmal der ‚guten‘ Stimme ist, daß sie, abgesehen von der ‚physiologischen‘ Müdigkeit, keine pathologischen Erscheinungen aufkommen läßt. Je nachdem sich eine Stimme diesem Ideale, diesem Typus mehr oder weniger nähert, ist ihre Güte zu beurteilen“ (Panconcelli-Calzia, 1956, S. 28). Er behauptet, dass diese typische Fiktion justifizierbar sei, was er auf Basis der gestörten Stimme, also durch Negativ-Definition, zu belegen versucht: „eine funktionell gestörte Stimme zeichnet sich durch das ganze oder teilweise Fehlen der soeben erwähnten Kennzeichen aus“ (Panconcelli-Calzia, 1956, S. 28).

Dies ist die ausführlichste Beschreibung der Merkmale einer guten Stimme, die zugleich die oben berichteten Merkmale der gestörten Stimme integriert und die gute Stimme davon abgrenzt. Die Annahmen zur ‚guten‘ Stimme sind auch denen von Mathieson (2001) zur ‚normalen‘ Stimme erstaunlich ähnlich (siehe Kapitel 2.1.1). Auch Phoniater machen sich Gedanken über ästhetische

Aspekte des Singens und in welcher Verbindung sie zur gesunden und trainierten (oft als ‚gut‘ bezeichneten) Stimme stehen. So fragen Mathelitsch und Friedrich: „Was ist eine ‚schöne‘ Stimme?“, und antworten: „Über die Bedeutung der Schönheit einer Stimme kann wohl nie ein Konsens erreicht werden. [...] [Es ist] angedeutet worden, wie kulturelle (Entwicklung der Oper) oder technische (elektrische Verstärker-, Aufnahme- und Wiedergabemöglichkeiten) Bedingungen die Anforderungen und Möglichkeiten der menschlichen Stimme beeinflussen und damit auch, was zu verschiedenen Zeiten als schöne Stimme akzeptiert wird“ (Mathelitsch & Friedrich, 1995, S. 145). Als Beispiel wird der Kastratengesang genannt, der zu einer bestimmten Zeit als das höchste Gesangsideal angesehen wurde. „Insgesamt ist die schöne und leistungsfähige Stimme das Resultat eines Zusammenwirkens einer Vielzahl angeborener und erworbener äußerer und innerer Faktoren“ (Mathelitsch & Friedrich, 1995, S. 145). Es wird im Text von Mathelitsch und Friedrich deutlich, dass sich die Beschreibung der schönen Stimme fast gänzlich auf die klassische Stimme bezieht, die Stimme der populären Musik wird nur am Rande erwähnt. Zugleich ist nicht gesetzt, dass Leistungsfähigkeit (und Beweglichkeit) und Schönheit zusammenhängen müssen. Auch weiter im Text geht es um „optimale Voraussetzungen im Stimmapparat“ und „jahrelanges konsequentes Training mit Hilfe geschulter Gesangspädagogen“ (Mathelitsch & Friedrich, 1995, S. 145). Dies weist eindeutig auf ein Gesangsideal im Bereich der klassischen Musik hin.

Mit der Gesamtheit der klanglichen Ausdrucksmöglichkeiten der Stimme und dem Gefallen jenseits der klassischen Musik beschäftigen sich nur wenige Schriften. Einzig die heisere Stimme im populär-musikalischen Kontext erfährt, besonders im klinischen Kontext, größere Aufmerksamkeit.

2.5.2 Zur Ästhetik der heiseren Stimme

„Seine Stimme besaß eine ausgeglichene Intonation, war jedoch von Natur aus etwas rau. Aber eben dieser Defekt verwandelte sich gerade für ihn in einen Vorzug. In seiner pathetischen Klage machte sich etwas Kläglich-Rührendes bemerkbar, wohl geeignet, ihm Vertrauen zu erwerben und Mitleid zu erregen“ (Marcus Tullius Cicero, *De oratore*, zit. nach Foehr-Janssens, 2008, S. 129). Hier wird bereits in der Antike die Rauigkeit als Defekt der Stimme dargestellt, der aber nicht ausschließlich negativ bewertet, sondern unter gewissen Umständen als stimmig wahrgenommen wurde. Dass die heisere Stimme eine nicht per se negative Funktion haben muss, liegt schon allein darin begründet, dass sie in der populären Musik verwendet wird (siehe Kapitel 2.3.1). Phoniater befinden sich hier in einer Zwickmühle, da die Heiserkeit sowohl als Ausdrucksmittel eingesetzt wird, aber auch als das Leitsymptom einer Stimmstörung gilt, die im schlimmsten Fall zu einem Versagen der Stimme führen kann. Bei der Diskussion von Heiserkeit als Ausdrucksmittel konstatieren Seidner und Büttner: „Der klinische Überblick macht es einem eigentlich unmöglich, an ästhetische Wertungen zu denken“ (Seidner & Büttner, 1998, S. 72). Es sei zu fordern, dass jede anhaltende, nicht spontan zurückzunehmende Heiserkeit abgeklärt werden müsse und erst nach eventueller Behandlung „für den

ästhetischen Bereich freigegeben“ werden könne (Seidner & Büttner, 1998, S. 72). „Beurteilungen von Sängerstimmen können nur dann einigermaßen vergleichbar erfolgen, wenn man ästhetische und physiologische Aspekte zu trennen versucht“ (Seidner, 2012, o.p.). Auch wird deutlich, dass Seidner als Phoniater der Heiserkeit persönlich eine eher negative Konnotation beimisst, was durch Äußerungen wie „Populärmusik, die unser Leben in fast beängstigender Weise dominiert“ (Seidner, 2012), verdeutlicht wird.

In einer Untersuchung versuchen Seidner und Büttner (1998) der Frage nachzugehen, ob die positive Bewertung einer heiseren Stimme abhängig vom beigemessenen Krankheitswert ist, und untersuchten dies an geschulten und nicht geschulten Hörern. Es ist gleich anzumerken, dass die methodische Darstellung der Untersuchung nicht den wissenschaftlichen Standards entspricht, d.h. der Bericht ermöglicht keine Replikation der Studie. Es werden u.a. keine Angaben zur Teilnehmerzahl gemacht und die rein deskriptive Auswertung und Darstellung ist unvollständig. Zusammenfassend stellt Seidner (2012) die Studie und ihre Ergebnisse wie folgt vor: „Wir ließen Klangbeispiele von 16 Sängern unterschiedlichen Genres von insgesamt 6 Hörergruppen auditiv beurteilen und nach folgenden Merkmalspaaren graduiert einschätzen: gesund oder krank, nicht heiser oder heiser, gefällt mir oder gefällt mir nicht. Auch wenn die Merkmale Heiserkeit und Krankheit von allen Probanden miteinander in Verbindung gebracht worden waren, so ergab sich doch als erstaunliches Resultat, dass die an Hörbeurteilungen gewöhnten Probandengruppen den Krankheitswert niedriger einstufen als den Heiserkeitswert. Den meisten Hörern gefielen solche Stimmen am besten, die zwar als leicht heiser, zugleich aber als gesund eingeschätzt worden waren“ (Seidner, 2012, o.p.). Dies wurde auf einer visuellen analogen Skala (5 cm) mit der positiven Tendenz nach links und der negativen nach rechts erhoben. Ausdrücklich wurde darum gebeten, nicht die Musik, sondern den Stimmklang zu beurteilen. Es wurde darauf hingewiesen, dass dies fast nicht realisierbar sei (Seidner & Büttner, 1998, S. 72).

Die deskriptive Betrachtung der Daten zeigt, dass die Teilnehmer, die weniger mit auditiver Evaluation von Stimmen zu tun haben (hier die ‚Fußballspieler‘), Krankheit und Heiserkeit weniger unterscheiden konnten, wobei die ‚Studenten allgemein‘ dagegen deskriptiv einen Unterschied erkennen lassen. Anhand der Standardabweichungen wurde die Homogenität der Beurteilungen betrachtet, die zeigt, dass sich die höchste Übereinstimmung bei der Heiserkeitsbeurteilung findet (mittlere Standardabweichung, SD = ca. 0,34), mittlere bei der Krankheitsbeurteilung (mittlere SD = ca. 0,42) und die geringste bei der ästhetischen Einschätzung (mittlere SD = ca. 0,5). Da keine Mittelwerte berichtet wurden, ist die SD entsprechend wenig aussagekräftig (Seidner & Büttner, 1998, S. 74).

Als bedeutendstes Ergebnis wird diskutiert, dass die Heiserkeit höher als der Krankheitswert eingeschätzt wird, und das sogar von klinisch geschulten Teilnehmern. Die Autoren interpretieren dies so, dass Singstimmen wohl anders eingeschätzt würden: „Das Heiserkeitsphänomen wird dann

überhört und als künstlerisches Ausdrucksmittel angenommen“ (Seidner & Büttner, 1998, S. 75). Auch stellen Seidner und Büttner ihre eigenen Eindrücke dar, indem sie sagen, dass der heisere Stimmklang zwar von vielen Menschen als schön eingeschätzt wird, „als physiologisch orientierte und klinisch erfahrener Hörer leidet man erheblich unter den Hyperfunktionen [...] und man fragt sich manchmal, wie kranke und dekompenzierte Stimmen überhaupt gefallen können“ (Seidner & Büttner, 1998, S. 75). Eine mögliche Antwort geben sie selbst, indem sie vermuten, dass es um die Vermittlung starker Emotionen gehe. Fazit ist, dass die Heiserkeit im klinischen Kontext ernst genommen werde und keine allgemeine Akzeptanz stattfinden solle.

Ob eine heisere Stimme als schön bewertet werden würde, ist auch nach dieser Untersuchung unklar, ihr wird allerdings während des Singens ein geringerer Krankheitswert zugesprochen und leicht heisere Stimmen werden gemocht. Es müsste dabei vermutlich auch zwischen dem Gefallen einer rauhen und einer behauchten Stimme unterschieden werden, da die Rauigkeit meist als stärker ausgeprägt empfunden wird als die Behauchtheit (vgl. Nawka & Anders, 1996), was evtl. auch daran liegt, dass mit Rauigkeit eine höhere körperliche Spannung einhergeht, die laut Panconcelli-Calzia nicht erstrebenswert sei. Außerdem scheinen mit der behauchten Stimme mehr positive Eigenschaften verbunden zu werden, wie Studien zur Attraktivität zeigen (allerdings mit inkonsistenten Ergebnissen; vgl. Babel, McGuire & King, 2014, 3). Dass die Wahrnehmung einer schönen Stimme kulturell und zeitlich gebunden ist, wurde bereits beschrieben, auch müssen situative Faktoren berücksichtigt werden, sodass (in Anlehnung an das Kontinuum einer normalen und gestörten Stimme) eine Stimme evtl. in einer Situation als schön wahrgenommen werden kann, in einer anderen nicht. Die Einflussfaktoren für das Bewerten einer gesungenen Performance als ‚schön‘ stellt Himonides (2009) in einem Modell zur ‚vocal beauty‘ zusammen. Dafür bezieht er nicht nur den klassischen Gesang ein, wobei er aber auf nur wenige Studien zu Gesangsstilen in der populären Musik zurückgreifen kann. Besonders werden die Gefühle betrachtet, die beim Hören von Musik ausgelöst werden (‚felt emotions‘). Die Verbindung von Emotionen in Sprache und Musik sowie die Verbindung von Gesang und Musik werden gegenübergestellt und problematisiert, vor allem hinsichtlich einer nicht praktikablen Trennung der ausgelösten Emotionen durch Musik und Gesang: “The human voice, besides being a musical instrument, is also a most powerful agent for the expression of emotions outside the musical context and, unfortunately, these cannot be isolated” (Himonides, 2009, S. 38).

Die folgenden Elemente wirken nach Himonides auf die Wahrnehmung des Schönen im Gesangsvortrag ein (“Theoretical elements that ‘construct’ the perception of beauty in a singing performance”):

- “musical structure
- effects of acoustic signal manipulations
- production of acoustic signal
- listener characteristics
- sociocultural context
- human musical development
- neuropsychobiological processing of lyrics and music” (Himonides, 2009, S. 44).

2.5.3 Präferenz und die ideale Stimme

So wie die ‚normale‘ Stimme ist auch die ideale Stimme situationsabhängig zu bewerten, d.h. sie wird in einer bestimmten Situation oder einem bestimmten Kontext als normal oder als ideal bezeichnet und kann je nach Kontext andere Parameter aufweisen. Die Präferenz (z.B. das Gefallen) und die Ablehnung können sich ebenso kontextuell verändern und sind abhängig von kulturellen und sozioökonomischen Faktoren. Zur idealen Stimme schreibt Hollien: “In all fairness, it must be obvious by now that the concept of ‘ideal voice quality’ cannot be established on any basis close to universal. There simply are no characteristics that will always separate ideal voice from average voice – and do so in the many, many situations in which it is relevant to the process” (Hollien, 2000, S. 16). Das heißt aber auch, dass sich in verschiedenen Situationen durchaus Konzepte eines Ideals entwickeln können bzw. entwickelt haben, wie zum Beispiel bei einzelnen Personen (oft berühmte Schauspieler oder Sänger) oder auch innerhalb ganzer Gattungen wie dem klassischen (Opern-)Gesang. Ob solch eine Stimme in dem Zusammenhang auch als ‚normal‘ bezeichnet wird, ist eine andere Frage. Bisher wurde die normale Stimme nur im wissenschaftlichen Kontext beschrieben, um von ihr ausgehend das Abweichende zu beschreiben. Im Alltag könnte davon ausgegangen werden, dass ‚normal‘ auch als Synonym für unauffällig und damit fast schon für langweilig verwendet wird. Eine ‚ideale‘ Stimme allerdings deutet schon auf eine Präferenz hin – d.h. eine Stimme wird präferiert, wenn sie beispielsweise ideal für eine gewisse Situation ist. Allerdings kann ein Opernsänger ideal für eine Rolle sein, er muss aber noch nicht gefallen – auch wenn dies oft impliziert wird. Auch kulturelle Abhängigkeiten verdeutlichen dies: Eine chinesische Opernsängerin in der chinesischen Oper kann ideal sein, gefällt einem europäischen Hörer aber deshalb nicht automatisch. So ist auch in experimentellen Settings bereits der Kontext vorgegeben: Werden Stimmen in der populären Musik untersucht, wird eine andere Vergleichsebene herangezogen, als wenn klassische Stimmen bewertet werden. Dies gilt auch für das Vorlesen von Texten im Vergleich zur freien Rede und vieles andere mehr.

Während Hollien (2000) eine Reihe Studien zusammenfasst, die eine ideale Stimme oder eine normale Stimme zu definieren versuchten, wurde das Gefallen in Bezug auf die Stimme nur wenig erforscht. Diese Studien haben auch in der Regel einen anderen Kontext (z.B. die Stimmwahrnehmung und -bewertung von Transgender-Menschen; Hancock, Krissingner & Owen, 2011) oder sie bewerten Stimmen im direkten Vergleich. Implizite Ankerpunkte finden sich zum Beispiel in Studien zur sozialen Akzeptanz von Dialektsprechern (zum Beispiel der Twang, Kapitel 2.3.1, im Südstaaten-Akzent der USA).

Im Kontext vorgelesener Texte untersuchten Hollien, Gelfer und Carlson (1991) die ‚listening preferences‘ von 80 Sprechern (40 weiblich) abhängig von der Lautstärke der Stimme. Es ist hervorzuheben, dass die 5-stufige Skala der Gefallens-Ratings von ‚very much liked‘ bis ‚very much disliked‘ rangierte, was über den sonst üblichen Pol des ‚gefällt nicht‘ hinausgeht. Die Ergebnisse zeigen, dass die meisten Hörer ein mittleres Intensitäts-Level sowie tiefere Sprechstimmen bevorzugen. Es fanden sich weder systematische Unterschiede im Gefallen zwischen älteren und jüngeren Sprechern noch zwischen Männern und Frauen (jedoch gab es eine Tendenz, dass Männerstimmen grundsätzlich von allen eher präferiert wurden). „Medium intensity voices in combination with any pitch level were almost always preferred regardless of age of the listener“ (Hollien et al., 1991, S. 167). Interessanterweise liegt das Mittel der Gefallens-Ratings im neutralen Bereich: Mittelwert = 3,08 (auf der 5-stufigen Skala), SD = 1,07.

Diese Ergebnisse wurden in späteren Studien teilweise bestätigt und erweitert. Alterseffekte fanden Goy, Pichora-Fuller und van Lieshout (2016), in deren Studie es um den Einfluss des Alters sowohl der Evaluierer als auch der Sprecher ging. Die Autoren unterteilen die Items in ‚speech quality‘ (= Sprechweise) und ‚voice quality‘ (= Stimmklang). Auf einer 7-stufigen Skala mit einer mittleren Kategorie „neutral“ wurden folgende Merkmale für „voice quality“ bewertet: „very pleasant/very unpleasant, very smooth/very rough, very powerful/very weak“. Für die „speech quality“ wurden bewertet: „very pleasant/very unpleasant, very natural/very unnatural, very clear/very unclear, very easy to understand/very difficult to understand, very loud/very soft“ (Goy et al., 2016, S. 1651). Interessanterweise ist für die Hörer beider Altersklassen eine angenehme Stimme nicht rau. „Pleasantness was associated with naturalness, loudness and suitability as an audiobook reader in the same way for both younger and older listeners. For younger listeners, the perceived age of the talker was associated with speech qualities (pleasantness and loudness), whereas for older listeners, perceived age was associated with voice quality (roughness)“ (Goy et al., 2016, S. 1657).

Besonders die Sprechstimmlage war wiederholt Gegenstand von Studien zur Attraktivität. Gerade im interkulturellen Bereich konnten hier bemerkenswerte Unterschiede gefunden werden, die stereotype Frauen- und Männerbilder unterstützten. Re, O'Connor, Bennett und Feinberg (2012) fanden mit

einem interessanten Studien-design („method of constant stimuli“; Gescheider, 1997) heraus, dass Frauen tiefe, aber nicht zu tiefe Männerstimmen attraktiver finden und Männer hohe Frauenstimmen tiefen Frauenstimmen vorziehen. Während der Untersuchung wurden je zwei Stimmen paarweise dargeboten und die Teilnehmer hatten die Aufgabe, zunächst die höhere Stimme, dann die attraktivere und in Aufgabe drei die mehr maskuline bzw. feminine Stimme (für Frauen bzw. Männer) herauszusuchen.

Um sich der Ästhetik der Stimme in der populären Musik zu nähern, soll auf eine der wenigen Studien zum (ästhetischen) ‚Wert‘ populärer Musik eingegangen werden, welche einen Einblick gibt, wie wenig die bisher dargestellten Kriterien der Stimmbeschreibung in der Bewertung populärer Musik eine Rolle spielen, nämlich auf die umfangreiche Dissertation von Appens (2007) und besonders auf die Aspekte zur Stimme. Er untersucht inhaltsanalytisch Rezensionen auf der Internet-Plattform amazon zu verschiedenen Alben (Kritiker- und Verkaufsbestseller) und kategorisiert die geäußerten Wertungen. Insgesamt thematisiert jede dritte Rezension interpretatorische Qualitäten. Die Kategorie der interpretatorischen Qualitäten „besteht zu 59 % aus Aussagen über den Gesang, das Rappen und die Stimme, 22 % der Wertungen beziehen sich auf die Qualität des Instrumentalspiels und des Zusammenspiels der Musiker, 19 % der Urteile widmen sich Instrumentierung und Arrangement sowie der Klanggestaltung im Tonstudio“ (von Appen, 2007, S. 105). Es fällt allerdings auf, dass es Unterschiede zwischen den Alben gibt, d.h. es scheint Unterschiede in der Priorität der Merkmale abhängig vom Album oder dem Künstler zu geben.

Von Appen zeigt, dass bei der Wertung von Stimme und Gesang „konventionelle Wertmaßstäbe der abendländischen Kunstmusik“ (von Appen, 2007, S. 107) nur eine geringe Rolle spielen. Das geht damit einher, dass „die Ansprüche der Rezipienten vergleichsweise gering“ seien. „Die ‚schöne‘ und die geschulte Stimme, der Wohlklang, der große Stimmumfang oder die selten gemeisterte Tonhöhe – dies sind Kriterien, die die amazon-Kritiker eher selten bemühen“ (von Appen, 2007, S. 107). Dahingegen spielt der emotionale Ausdruck der Stimme/des Gesangs (die in den Rezensionen nicht differenziert werden) eine größere Rolle. Emotionalität und Ausdrucksstärke werden auch eher durch einen „‘ungekünstelten‘ und ‚unpräzisen‘ Gesang“ hervorgerufen als durch einen „wohlklingenden, kunstvoll genannten“ (von Appen, 2007, S. 108f.). Der persönliche Ausdruck des Interpreten überwiegt gegenüber einer eindrucksvollen Darstellung (von Appen, 2007, S. 108f.). Am Beispiel Bob Dylans wird dies verdeutlicht, dessen Stimme als ‚kaputt‘ bezeichnet werden könne, also als weder schön noch wohlklingend. Dies hat allerdings keine negativen Auswirkungen auf die Rezensenten, „im Gegenteil: Sie ermöglichen es Dylan, eine Ausdrucksintensität zu erreichen, für die es ihrer Ansicht nach nichts Vergleichbares gibt“ (von Appen, 2007, S. 108). Diese Einschätzung von Appens zur Stimme Dylans ist klinisch geprägt, da er sie als kaputt bezeichnet. So eine Assoziation wird wohl nie einem Verehrer der Stimme Dylans einfallen: „Ich liebe Bob Dylans Stimme“, schreibt

Mittelstenscheidt (2009, S. 31) in einem Artikel in der Zeitschrift *Musik und Ästhetik* und erörtert ihr persönliches Verhältnis zur Stimme Dylans – ein Krankheitswert wird nicht erwähnt.

Zusätzlich ist noch zu erwähnen, dass bei von Appen am Beispiel von Eminem der Sprechgesang behandelt wird (hier im Rap-Stil), der nicht dazu geeignet scheint „Emotionen wie Trauer, Liebe u. dgl. sinnlich zu vermitteln“ (von Appen, 2007, S. 109), weshalb sich die Rezensenten auf den emotionalen Gehalt der Texte beziehen.

Besonders werden noch die Sängerinnen Anastacia und Norah Jones beschrieben. Anastacias Stimme wird als ‚kräftig, ausdrucksstark und charakteristisch‘ beschrieben und ihre ‚Flexibilität‘ wird betont. Norah Jones‘ Stimme wird jedoch als einzige mit dem Attribut ‚schön‘ bezeichnet, neben den Begriffen ‚gefühlvoll, samtig schön, sanft, zart, dunkel, betörend, ruhig und erotisch, beruhigend, wunderbar soulig, rau oder rauchig‘ (vgl. von Appen, 2007, S. 106). Interessant ist noch der Hinweis, dass sich zu den männlichen Stimmen kaum charakterisierende Beschreibungen fanden (von Appen, 2007, S. 107).

In einer weiteren Untersuchung zu den „Wertungskriterien der Deutschland sucht den Superstar-Jury“ (DSDS) interpretiert von Appen (2005) die Ergebnisse der vorliegenden Analyse vor dem Hintergrund sozialer Milieus und kulturindustrieller Strategien. Für die vorliegende Arbeit sind die Ergebnisse der Inhaltsanalyse interessant, für die 148 Jury-Kommentare transkribiert und inhaltsanalytisch ausgewertete wurden. Von Appen fand 214 wertende Aussagen, die er in Kriterien zusammenfasste (siehe Tabelle 5). Interessanterweise finden sich in 20 % der Aussagen „traditionelle Vorstellungen darüber, was ‚guter Gesang‘ sei“ (von Appen, 2005, S. 190), was damit die am häufigsten vertretene Kategorie ist. Darin fanden sich vor allem Aussagen zur Intonation, aber auch zum Rhythmus und zur Phrasierung. „Daneben berücksichtigte die Jury die gesangstechnische Schwierigkeit des ausgewählten Stückes“ (von Appen, 2005, S. 191).

Anteil	Kategorien
20 %	Handwerklich-technische Aspekte des Gesangs
15 %	<i>Emotionale Aspekte des Gesangs</i>
12 %	Visuelles Erscheinungsbild des Kandidaten
11 %	Visuelle Aspekte der ‚Performance‘, Tanz, Bühnenpräsenz
7 %	Vermittlung von Spaß, Lockerheit, Souveränität
6 %	<i>Qualitäten der Stimme</i>
< 6 %	Einsatz, sich Mühe geben; Originalität der Interpretation; Ausdruck von ‚Power‘, Individualität, Persönlichkeit des Kandidaten
15 %	Sonstiges

Tabelle 5: „Ergebnisse der quantitativen Inhaltsanalyse (n = 214 wertende Aussagen)“ (von Appen, 2005, S. 191).

Die zweite Kategorie der emotionalen Aspekte des Gesangs wurde von den Juroren als für sie am wichtigsten benannt (obwohl mit 15 % quantitativ an zweiter Stelle). „Dabei entsteht der Eindruck, Emotionalität sei messbar und ein Zuviel an ‚Gefühl‘ nicht möglich [...]. Kein Kandidat wurde [...] dafür kritisiert, dass er es mit der Emotionalität übertrieben habe“ (von Appen, 2005, S. 191). Besonders

betont wurde, wenn die Jury angeblich selbst tief berührt gewesen sei. Eine fehlende Ausdruckskraft des Vortrags wurde dementsprechend bemängelt. Von Appen beschreibt, dass die Gefühlsbeschreibungen auf das „Liebes-Gefühl“ beschränkt blieben, „das Assoziationen an Sanftheit, Sehnsucht, Zärtlichkeit oder Melancholie weckt [...] Bemerkenswert ist weiterhin, dass das Interesse dabei nie den tatsächlichen Gemütszuständen der Interpreten galt. Verlangt wurde nicht persönlicher Ausdruck, sondern das Darstellen oder Imitieren von Ausdruck, so wie man es von einem Schauspieler erwartet“ (von Appen, 2005, S. 192). Zum visuellen Erscheinungsbild der Kandidaten wird, so von Appen, vorwiegend erwartet, zu „den in den Medien verbreiteten männlichen und weiblichen Schönheitsidealen“ zu gehören (von Appen, 2005, S. 192). „Die Eigenschaften der Stimme beurteilte die Jury selten und dann sehr undifferenziert (,Ich finde deine Stimme grundsätzlich gut‘)“ (von Appen, 2005, S. 193).

Im Verlauf des Artikels vergleicht von Appen die DSDS-Jury-Bewertungen mit amazon-CD-Rezensionen zu a) Bestseller-CDs, die nicht mit DSDS assoziiert sind (N = 120 wertende Aussagen) und b) DSDS-CDs. Dabei fällt auf, dass bei den Bestseller-CDs emotionale Aspekte (,Gefühl‘) wesentlich seltener genannt werden (3 %), ähnlich wie Aspekte der Gesangsqualität und der Stimme (5 %). Im Vordergrund stehen hier Qualitäten der Komposition und der Songtexte (die aber selbsterklärend bei DSDS keine Rolle spielen), die mit den Kriterien „neu/alt, typisch/untypisch, überraschend/gewohnt“ (von Appen, 2005, S. 195) in 18 % der Aussagen bewertet werden. Im Vergleich zu DSDS-CD-Rezensionen (N = 116) fällt auf, dass sich die Rezensenten in Bezug auf die handwerklich-technische Seite des Gesangs einig sind (,möglicherweise beeinflusst durch die TV-Jury“, von Appen, 2005, S. 197), bezüglich des „sentimentalen ,Gefühls‘ nicht ganz so deutlich mit der Jury übereinstimmen“ (von Appen, 2005, S. 197; 15 % Jury vs. 9 % CD-Rezensionen). Von Appen diskutiert den Umstand, dass nicht deutlich wird, ob die Kriterien von der Jury gemacht wurden oder die Rezensenten diese kreierten.

Es wurde in diesem Kapitel deutlich, dass ästhetische Urteile zur Stimme nur wenig erforscht sind. Besonders nicht speziell geschulte Hörer und ihre ästhetischen Urteile in Hinsicht auf die Stimme sind selten untersucht worden. Das zeigt, dass diese Hörer gerade nicht auf traditionelle oder sachliche Beschreibungen der Stimme zurückgreifen, sondern besonders häufig die ausgelösten Gefühle einer Stimme beschreiben.

Etablierte Wissenschaftler äußerten sich zwar zu Konzepten der idealen, schönen und guten Stimme, welche erwartungsgemäß sehr breit und von vielen Faktoren abhängig sind (wie am Modell von Himonides (2009) gesehen wurde), die empirische Forschung ist zu diesem Thema jedoch spärlich. Nur wenige Merkmale der Stimme wurden hinsichtlich ihrer Präferenz untersucht: Es finden sich Untersuchungen zur präferierten Stimmlage und zur Heiserkeit als spezifischem stimmlichen Merkmal. Dabei lassen sich sowohl die Präferenz als ein Vergleichs- bzw. Vorzugsurteil zu einem Objekt (,ich mag

das hier lieber als jenes') als auch das Gefallen als eine generelle positive Einstellung gegenüber einem Objekt in experimentellen Settings untersuchen, sowohl bezogen auf einzelne Merkmale als auch auf den gesamten Klang einer Stimme. Möchte man mehrere Stimmen oder Merkmale miteinander vergleichen und untersuchen, welche Stimme oder Merkmalsausprägung von den dargebotenen am meisten gefällt, sollte man die Präferenz als ein situatives, spontanes Urteil abfragen. Bei Bewertungen von Stilen oder dem gesamten Werk eines Interpreten fragt man jedoch nicht nur nach einem spontanen Urteil, sondern nach der Einstellung zu diesem Objekt, die von vielen musikimmanenten und -exmanenten Faktoren wie Hörerfahrungen, Assoziationen, Vorurteile usw. beeinflusst wird. In den vorliegenden Studien wird deswegen größtenteils nach Präferenzurteilen gegenüber Stimmmerkmalen gefragt.

2.6 Zusammenfassung und Ausblick auf die aktuelle Studienreihe

Es wurde der Forschungsstand aus verschiedenen Wissenschaftsbereichen, genauer gesagt aus Phoniatrie, Sprechwissenschaft, Phonetik, Musikwissenschaft und Emotionsforschung, zusammengestellt, die mittels verschiedener Methoden wie akustischer und auditiver Analyse, mit Experimenten und mithilfe der Inhaltsanalyse den Gegenstand Stimme untersuchten. Bestehende Inventare zur Kategorisierung und Klassifikation verschiedener Stimmmerkmale sowohl für das Sprechen als auch das Singen wurden vorgestellt, mittels derer vokale Profile für Forschung, Diagnostik, Gesangs- und Sprechpädagogik o.Ä. erstellt werden können. Außerdem wurden theoretische und empirische Ansätze zur Definition und den Merkmalen der normalen Stimme im Gegensatz zur gestörten Stimme sowie der idealen, schönen und guten Stimme dargestellt.

Es wurde deutlich, dass es zwar eine Reihe von Inventaren zur Beschreibung stimmlicher Merkmale gibt, ästhetische Urteile aber wenig oder gar nicht mit ihnen in Beziehung gesetzt und untersucht wurden. Auch sind die Inventare auf bestimmte Anwendungsbereiche und Gegenstände beschränkt. Dies waren bisher überwiegend die Sprechstimme sowie die klassische Gesangsstimme. Die Vielzahl an möglichen Sprech- und Gesangssituationen mit jeweils eigenen Bedingungen und Stimmmerkmalen wurde bisher kaum in der Forschung berücksichtigt. Je nach untersuchtem Gegenstand und vorhandenem Kontext muss auch angenommen werden, dass andere, dem jeweiligen Stil angemessene Standards zur Bewertung angelegt werden. Diese beeinflussen die vom Sänger verwendete Technik sowie die intendierte Wirkung und somit, auf Rezipientenseite, die Bewertung der Stimmmerkmale und das Gefallens-Urteil.

Der Fokus der bisherigen Untersuchungen der Stimme lag vor allem auf Experteneinschätzungen im Sinne von Auffälligkeiten der Stimme oder pathologischen Eigenheiten. Phoniatische Arbeiten versuchen primär, gesunde und gestörte Stimmen mit konkreten Merkmalen zu beschreiben und diagnostizieren zu können. Wie jedoch Laienhörer, also Menschen, die nicht spezifisch stimmlich

geschult sind, Stimmen wahrnehmen und welche Aspekte an Stimmen für sie wichtig sind, wurde bislang kaum untersucht. Spielen für Fans bestimmter Musik spezifische Stimmmerkmale eine Rolle, oder ist es ihnen für eine positive Bewertung eines Sängers oder einer Sängerin wichtiger, welche emotionale Wirkung die Stimme auf sie hat? Welche Begriffe verwenden sie, um eine Stimme (vielleicht im Gegensatz zu einer anderen) zu beschreiben und können sie auch als ungeschulte Hörer Stimmauffälligkeiten wahrnehmen und benennen?

Es lässt sich daher feststellen, dass ein Inventar vokaler Merkmale fehlt, das von verschiedenen Hörergruppen für die Erstellung von Stimmprofilen genutzt werden kann. Diese Profile und die entsprechenden vokalen Merkmale können dann für verschiedene Forschungsfragen mit anderen Variablen in Verbindung gesetzt werden, zum Beispiel mit ästhetischen Urteilen. Außerdem ist auffällig, dass sich die Musikforschung bisher auf die Untersuchung des Gefallens beschränkt und die Ablehnung kaum berücksichtigt (Bryson, 1996, S. 894f.; Greasley et al., 2013, S. 406). Gerade darin wird jedoch Potential gesehen, Hörer zu differenzierten Beschreibungen von Musik zu veranlassen, wie in aktuellen Untersuchungen gezeigt wurde (Ackermann, 2018) – dies sollte auch auf die Stimme zutreffen. Anhand der vorliegenden Studien ist außerdem anzunehmen, dass Nicht-Experten andere Kriterien zur Beurteilung von Stimmen anlegen bzw. für wichtig erachten.

2.6.1 Auswahl der Merkmale, Skalen und Musikstile

Auffällig ist, dass das Merkmal der Rauigkeit von antiken Quellen bis heute in vielen Inventaren zur Evaluation von Stimmen sehr häufig Verwendung findet – Ausnahme ist die Bewertung des klassischen Gesangs. Im populären Gesang hingegen ist es ein wesentliches Ausdrucksmittel. Ebenso wird die Behauchtheit als ein Kriterium genannt, wenngleich die Rauigkeit als das prominentere Merkmal von beiden beschrieben wird. Beide Begriffe fallen unter den Oberbegriff der Heiserkeit, welche allerdings keine neutrale Konnotation mehr hat, sondern einen Krankheitswert impliziert. Da in der vorliegenden Arbeit möglichst konnotationsarme Begriffe zur Beschreibungen von Stimmen verwendet (und entwickelt) werden sollen, wurden die Begriffe ‚rau‘ und ‚behaucht‘ verwendet. Diese vor allem im populären Gesang häufig vorkommenden Klangqualitäten sollen vor allem in ihrer ästhetischen Wirkung untersucht werden, da ihnen, trotz der von Phoniatern regelmäßig angesprochenen pathologischen Komponente, offenbar eine ästhetische Funktion zukommt.

Obwohl also der Krankheitswert einer Stimme hier keine zentrale Rolle spielen soll, müssen trotzdem die Inventare aus der Forschung und Diagnostik in der vorliegenden Arbeit berücksichtigt werden. Vor allem klinische Inventare beschreiben prominente und auffällige Merkmale der Stimme – ein Inventar zur Beschreibung von Stimmstörungen kann, ungeachtet des Krankheitswertes, eine mögliche Ausdruckspalette einer Stimme beschreiben: So finden sich auch im Alltag und in verschiedenen Gesangsstilen (auch der klassischen Musik) ein nasaler oder gepresster Stimmklang sowie harte und

weiche Stimmeinsätze u.v.m. Auch ist zu beachten, dass diese Merkmale nur von einigen Hörern als unangenehm oder sogar pathologisch aufgefasst werden, während sie für andere ansprechend sein können.

Anspruch der vorliegenden Arbeit war es, von einem möglichst breiten und detaillierten Inventar zur Beschreibung und Bewertung vokaler Merkmale auszugehen. Daher wurde auf den sehr umfangreichen Katalog von Bose zur Beschreibung des „stimmlich-artikulatorischen Ausdrucks“ (Bose, 2010, S. 29) zurückgegriffen, der bereits Merkmale aus verschiedenen Inventaren und Fachbereichen vereint und um eine konnotationsarme Beschreibungsebene bemüht ist. Deshalb werden nicht nur Klangqualitäten (z.B. rau oder hell) untersucht, sondern auch artikulatorische Merkmale, Form und Varianz der Tonhöhenveränderungen sowie „Komplexwahrnehmungen des Sprechausdrucks“ (Bose, 2001, S. 280) wie die Phonationsform (Singen, Sprechen, Rufen etc.) und die Spannung oder der allgemeine Ausdruck.

Da der Katalog auf die Sprechstimme ausgerichtet ist, mussten für den musikwissenschaftlichen Zusammenhang Aspekte des Singens integriert werden. Die Überschneidungen in den Merkmalsbeschreibungen sind groß, da dieselben Organe an der Produktion des Singens und Sprechens beteiligt sind, allerdings finden sich im Gesang Fachbegriffe, die mit den Begriffen des Katalogs abgeglichen werden mussten. Das betrifft beispielsweise den Twang, der als ein Ausdrucksmittel des populären Gesangs gilt. Ausgehend von Beschreibungen zur Physiologie in der Stimmforschung ist der Twang im Katalog von Bose mit der physiologisch-anatomischen Beschreibung der faukalen Enge repräsentiert (lat. fauces = der Rachen). Die Facette des ‚nasal twang‘ findet sich in einer Kombination der Enge und der Nasalität wieder, welche ebenfalls im Katalog enthalten ist. Für die Stimmexperten der Sprechwissenschaft in der ersten Studie zur Fragebogenentwicklung (Kapitel 4.1.5) konnte somit die faukale Enge als Merkmal für dieses Ausdrucksmittel verwendet werden. Für die Nicht-Experten wurde als Ausgangspunkt lediglich der Begriff ‚eng‘ verwendet (Kapitel 4.1.4). Ein anderes wichtiges Merkmal des Singens und ein angeblich wichtiges Unterscheidungsmerkmal von Singen und Sprechen, die gleitenden Tonhöhenveränderungen (in der Musik/dem Gesang mit dem Begriff ‚Glissando‘ bezeichnet), ist mit den Merkmalsausprägungen ‚sprunghaft-gleitend‘ vertreten. Das Vibrato hingegen kommt im Sprechen nicht vor. Damit wird eine regelmäßige Veränderung der Tonhöhe in regelmäßigen Abständen bezeichnet. Im Sprechen kann allerdings ein Zittern (unregelmäßig) in der Stimme auftreten (durch einen Tremor oder in gewissen emotionalen Momenten wie Trauer oder Furcht). Dieses Merkmal wird in der GVPS (Bänziger et al., 2014) als eine Instabilität beschrieben. Das Merkmal ‚Instabilität‘ wurde für den Fragebogen für die Nicht-Experten übernommen, da auch Bänziger et al. (2014) ihren Fragebogen mit Laien entwickelt haben – in der vorliegenden Arbeit allerdings wurden die Ausprägungen Vibrato und Zittern verwendet.

Damit das Ziel der Entwicklung eines Inventars zur Beschreibung von Stimmen verfolgt werden konnte, wurde eine beschreibende Bewertungsskala genutzt („Welche Ausprägung hat das Merkmal?“). Bestimmte ästhetische Werturteile zu Stimmen wurden gesondert abgefragt und hatten in den vorliegenden Studien jeweils einen anderen Schwerpunkt: von der Ablehnung (Kapitel 4) zum breiteren Gefallen und Nicht-Gefallen (Kapitel 5) mit zusätzlich einer Bewertung ästhetischer Emotionen (Kapitel 6). Die ästhetischen Emotionen wurden hier, wie in Kapitel 2.4.1 erläutert, als ‚felt emotions‘ abgefragt. Die Teilnehmer der Studien hatten angemerkt, dass sie neben dem Gefallen auch die Wirkung der Stimmen angeben möchten. Dazu wurde in der letzten Studie auf Basis der beschriebenen Inventare GEMS und AESTHEMOS, aber auch im Abgleich mit in den vorherigen Studien genannten Emotionen, eine Abfrage einiger ästhetischer Emotionen in den Fragebogen aufgenommen. Als Ergänzung etablierter Erkenntnisse aus der Präferenzforschung wurden auch weitere Aspekte wie der Einfluss von Vorinformation, Hörerfahrung und musikalische Vorbildung auf die Bewertung von Gesang und Musik untersucht.

Zur Entwicklung des Inventars zur Stimmbeschreibung und der Untersuchung der Zusammenhänge von Stimmerkmalen mit ästhetischen Werturteilen wurden nicht nur quantitative Ansätze berücksichtigt. Es wurde ein breiter, offener Zugang gewählt, der verschiedene methodische Ansätze einschließt. Da die bisherige Forschung keine verlässlichen Daten zu den von Laien verwendeten Begrifflichkeiten bei der Stimmbeschreibung bot, wurde zunächst mit einer Interview-Studie (Kapitel 4) untersucht, wie Nicht-Experten überhaupt über Stimmen sprechen, wenn sie gebeten werden, ihr ästhetisches Urteil zu einer Stimme anhand des Klangbilds zu begründen. Dazu wurde mithilfe von Interviews den Einstellungen und Beschreibungen der Nicht-Experten spezifisch zu abgelehnten Stimmen nachgegangen. Die Aussagen der Teilnehmer wurden nach dem Interview mit einem anschließenden Fragebogen trianguliert, um zum einen den Nicht-Experten weitere Ausdrucksmöglichkeiten zur spezifischeren Beschreibung der Stimmen an die Hand zu geben, zum anderen aber auch, um die Anwendbarkeit bestimmter Merkmalsformulierungen zu überprüfen. Im Anschluss an diese Studie wurde der Fragebogen nach den Rückmeldungen der Teilnehmer angepasst sowie die Ablehnung als ästhetisches Urteil um das Gefallen als positive Dimension erweitert und an einer größeren Gruppe Teilnehmer überprüft (Kapitel 5). Auch in dieser Studie wurden, wie bereits im Interview, quantitative und qualitative Elemente verbunden, um mit offenen Fragen möglichen Schwierigkeiten im Umgang mit dem Fragebogen nachzugehen. Der wiederum veränderte Fragebogen wurde letztendlich mit einer großen Anzahl Teilnehmer verschiedener Professionen (Experten der Sing- und Sprechstimme) und Nicht-Experten in einer Online-Studie (Kapitel 6) getestet, um die Merkmale in ihrer Anwendbarkeit zu überprüfen und allgemeinere Aussagen zum Gefallen vokaler Merkmale treffen zu können. Auch wurden erstmals die ästhetischen Emotionen integriert. Durch das

gewählte Studiendesign konnten alle ästhetischen Urteile mit den wahrgenommenen Merkmalen in Verbindung gesetzt werden.

Zum Schluss ist anzumerken, dass akustische Analysen in der vorliegenden Arbeit keine Beachtung fanden, da zum einen grundsätzlich die Zuordnung von auditiven zu akustischen Ereignissen nicht unproblematisch ist und zum anderen in den vorliegenden Studien auch nicht möglich war, da ausschließlich Stimmen im Kontext von Musik untersucht wurden, d.h. Instrumente und Stimme waren nicht getrennt (keine A-cappella-Aufnahmen), was eine akustische Analyse klanglicher Parameter unmöglich macht. Die perceptuell-auditive Einschätzung verschiedener Hörer war Basis der Befragung, da eben diese Eindrücke Einfluss auf das Inventar zur Beschreibung von Singstimmen haben und die ästhetischen Urteile auf dieser Wahrnehmung basieren.

3 Prolog: Zuordnung von Stimmen und ihren Beschreibungen

Wenn wir über Stimmen berichten, sind wir abhängig von dem uns zur Verfügung stehenden Vokabular. Während der Laie hier schnell an seine Grenzen gerät, stehen im künstlerischen Bereich besonders dem Musikrezensenten (Musikwissenschaftler, Journalist etc.) sowohl präzise Begriffe, als auch ein großes Repertoire an Metaphern zur Verfügung, mit denen klangliche Besonderheiten beschrieben werden können. Mit der folgenden Untersuchung wurde der grundlegenden Frage nachgegangen, ob Beschreibungen von Stimmen den auditiven Eindruck einer Stimme überhaupt charakteristisch abbilden können. Ausgangspunkt waren Beschreibungen zu verschiedenen Sängern aus Musikrezensionen und der musikwissenschaftlichen Forschung. Diese wurden zusammen mit entsprechenden Einspielungen der beschriebenen Sänger dargeboten und sollten zugeordnet werden. Außerdem wurden die Teilnehmer gebeten, zwei der Sänger mit eigenen Worten charakteristisch zu beschreiben. Damit sollte untersucht werden, inwiefern eine Einheitlichkeit in den Beschreibungen oder eine größere Varianz an Begriffen auftritt, wie sehr Kategorien zur Beschreibung erkennbar werden und ob eine Differenzierung zwischen den Stimmen anhand der verwendeten Begriffe möglich ist.

Über einen Zuordnungstest sollte untersucht werden, inwiefern die drei dargebotenen Sänger oder Sängerinnen den drei vorliegenden Beschreibungen korrekt zugeordnet werden können. Außerdem sollten im Text die für die Zuordnung relevanten Eigenschaften/Beschreibungen des Sängers markiert werden. Somit konnten neben der Genauigkeit der Zuordnung auch charakteristische Merkmale der Stimmen untersucht werden. Damit wird auch eine Grundannahme von Musikkritikern überprüft, die den Anspruch an ihre Rezensionen haben, die Besonderheiten einer Stimme so zu beschreiben (und beschreiben zu können), dass im besten Fall „der Leser die Stimme im Kopf klingen hört“ (J. Kesting, persönl. Mitteilung, 11.10.2016). Bei näherer Betrachtung der Musikkritiken von Kesting (2008a) fallen neben technischen, musik-stilistischen Begriffen auch sehr weite Assoziationen einer Stimme auf, die zum Beispiel das besonders beeindruckenden Timbre der Stimme Plácido Domingos mit den Worten beschreibt: „Ihre Farbe war die eines dunkelroten Vino de Rioja“ (Kesting, 2008b). Es soll nun mit einem qualitativen Ansatz überprüft werden, ob diese Form der Stimmbeschreibung so charakteristisch ist, dass eine Zuordnung von drei Sängern desselben Musikstils zu ihren Beschreibungen möglich ist. Dazu wurden drei einflussreiche klassische Tenöre und drei Gospel-Sängerinnen ausgewählt.

3.1 Methoden

Die Stimmen und Stücke wurden nach der Verfügbarkeit qualitativ hochwertiger und repräsentativer Beschreibungen ausgewählt. Für die klassischen Stimmen wurden Beschreibungen Kestings aus dessen mehrbändiger Monographie *Die großen Sänger* (Kesting, 2008a) gewählt: Plácido Domingo (geb. 1941;

Kesting, 2008b) und Luciano Pavarotti (1935-2007; Kesting, 2008c). Die andere Quelle, aus der auch die Beschreibungen für die Gospel-Sängerinnen gewählt wurden, ist das *Lexikon der Gesangsstimme* (Mecke, Pfeleiderer, Richter & Seedorf, 2016), woraus die Beschreibung zu Jussi Björling (1911-1960) von Brandenburg gewählt wurde, der u.a. auch Kesting zitiert (Brandenburg, 2016). Es wurden eher bekannte Sänger ausgesucht, damit die Teilnehmer eine Chance hatten, die Stimmen zu erkennen und zu benennen. Außerdem sollte von den drei Tenören jeweils eine Aufnahme des gleichen Werkes in vergleichbarer Qualität vorliegen. Die Wahl fiel auf die Arie *Vesti La Giubba* aus der Oper *Pagliacci* (1892) von Ruggero Leoncavallo. Die Auszüge aus den Texten zu den einzelnen Sängern wurden so gewählt, dass eine Identifikation aus anderen Gründen als stimmlichen Beschreibungen nicht möglich war, d.h. es wurden nur klangliche und gesangstechnische Merkmale beschrieben. Dies galt auch für die Gospel-Sängerinnen. Gewählt wurden drei einflussreiche afroamerikanische Sängerinnen: Aretha Franklin (1942-2018), Mahalia Jackson (1911-1972) und Bessie Smith (1894-1937). Pfeleiderer schrieb die Artikel zu Bessie Smith (Pfeleiderer, 2016b, Auszüge von S. 559f.) und Aretha Franklin (Pfeleiderer, 2016a, Auszüge von S. 234); Hähnel, ein Mitarbeiter von Pfeleiderer, schrieb den Artikel zu Mahalia Jackson (Hähnel, 2016, Auszug von S. 302). Hier konnte nicht für alle Sängerinnen das gleiche Stück gewählt werden, da es den Beschreibungen nicht entsprochen hätte, die sich teilweise auf bestimmte, charakteristische Stücke der Sängerinnen beziehen. Die Stücke wurden den Texten entsprechend ausgewählt: Aretha Franklin: *Precious Lord, Take My Hand*, vom Album *Amazing Grace*, Atlantic 1972; Mahalia Jackson: *There's Not a Friend Like Jesus*, 1947, vom Album *The Complete Mahalia Jackson*, Integrale 1998; Bessie Smith: *I Need a Little Sugar in My Bowl*, 1931, vom Album *The Complete Recordings, Vol. 1-5*, Columbia 1991-1996.

Es wurden 22 Schulmusik-Studierende der Universität Kassel befragt. Musikstudierende zeichnen sich durch ein entsprechend geschultes Gehör aus und haben ein Vokabular zur musikalischen Klangbeschreibung erlernt, welches das Verstehen gewisser Fachtermini in den vorliegenden Texten gewährleistet. Die sängerische Vorbildung der Teilnehmer reicht vom Gesang als Nebenfach zum Gesang als Hauptfach, mit unterschiedlichen Interessen hinsichtlich populärer und klassischer Musik. Die Studie wurde in Form einer Gruppentestung durchgeführt. Die Einspielungen wurden über Lautsprecher präsentiert und die Bewertung erfolgte auf Blanko-Papier und zwei Fragebögen.

Die Teilnehmer hatten zwei Aufgaben zu bearbeiten. Die erste Aufgabe war die freie Beschreibung der Stimme Jussi Björlings und anschließend der Stimme Bessie Smith' mit der Instruktion: „Bitte schreiben Sie in Stichworten alles auf, was Ihnen zu der jeweiligen Stimme einfällt. Wählen Sie Ihre Beschreibungen so, dass die Stimme möglichst komplett und charakteristisch beschrieben wird.“

Die zweite Aufgabe bestand darin, Stimmen der passenden (Kritiker-)Beschreibung zuzuordnen. Die Teilnehmer bekamen Auszüge aus den Texten zum jeweiligen Sänger, jeweils drei auf 1-2 Seiten. Die

Beschreibungen wurden zu Beginn von der Studienleiterin vorgelesen (die Teilnehmer lasen still mit). Begonnen wurde mit der Einordnung der klassischen Stimmen. Jede Aufnahme wurde zunächst zweimal hintereinander abgespielt und danach alle drei noch einmal nacheinander. Die Teilnehmer sollten zum einen vermuten, um welchen Sänger es sich handeln könnte (Angabe des Namens) sowie welche Stimme zu welcher Beschreibung passt (die Stimmen wurden durchnummeriert). Anschließend sollten sie die Begriffe oder Phrasen unterstreichen, die ihre Zuordnung am stärksten beeinflusst haben. Nach der Zuordnung der klassischen Stimmen wurde das gleiche Verfahren für die drei Gospel-Stimmen angewendet. Die Reihenfolge der Sänger war: Björling (Beschreibung C), Domingo (Beschreibung B), Pavarotti (Beschreibung A); Smith (Beschreibung F), Franklin (Beschreibung D), Jackson (Beschreibung E).

Die Auswertung und Interpretation der Daten folgte folgenden Leitfragen:

- Welche Begriffe verwenden die Teilnehmer zur Beschreibung der Stimmen? Welche Kategorien werden dabei häufiger besetzt? Gibt es Unterschiede in der Beschreibung einer klassischen und einer nicht-klassischen Stimme?
- Gibt es Überschneidungen zwischen der Beschreibung der Teilnehmer und der Autoren?

3.2 Auswertung und Ergebnisse

Da nicht alle Teilnehmer alle Aufgaben vollständig erfüllt haben, wurde Nr. 21 komplett aus der Analyse ausgeschlossen sowie Teile folgender Datensätze: 16 A, C, D, 22 C, 23 A-C. Die Buchstaben verweisen auf die Sänger. Bei den ausgeschlossenen Teilen der Datensätze waren einzelne Aufgaben nicht korrekt ausgeführt worden. Am Ende flossen von 21 Personen die Beschreibungen/Daten ein.

Die Beschreibungen wurden in Tabellen übertragen und kodiert. Da zum Teil Begriffe und Beschreibungen zu mehreren der Kodierungen gehören konnten, wurden in solchen Fällen doppelte Kodierungen verwendet. Die Beschreibungen der Teilnehmer berührten Aspekte aus den folgenden sechs Kategorien:

1. Stimmlage und Stimmgattung
2. Klangliche Beschreibung (inkl. Resonanz und Dynamik)
3. Technische Angaben (inkl. Spannung, Stimmstörungen)
4. Stilistische Beschreibung (inkl. Vibrato)
5. Artikulation und Text/Sprache
6. Ausdruck

Kesting selbst nennt vier Kategorien, die er als Skizze für seine Beschreibungen verwendet (J. Kesting, persönl. Mitteilung, 11.10.2016):

1. Stimmmaterial (Umfang, Klang)
2. Technik (Intonation, Stimmführung)
3. Artikulation (Musikalität, Stil)
4. Gestaltung (Charakter, Drama)

Die gefundenen Kategorien aus den Beschreibungen der Studienteilnehmer unterscheiden sich interessanterweise nur wenig von denjenigen, die Kesting als wichtig für eine Stimmbeschreibung ansieht. Allerdings fehlt den Teilnehmern (im Vergleich zu Kesting und anderen Experten) neben der detailreichen Beschreibung der Stimme auch die Kenntnis des Stils, wie die Arie *Vesti La Giubba* stilistisch angemessen interpretiert wird, d.h. sie beurteilen hier weniger eine historisch-informierte Aufführungspraxis, als dass sie ihren Eindruck beschreiben.

3.2.1 Offene Beschreibungen zum klassischen Sänger

Für die Stimme Björblings gab es 75 Beschreibungen, die sich wie folgt in die Kategorien einteilen ließen (hier wörtlich wiedergegeben, die Anzahl der Nennungen findet sich in Klammern, Einmalnennungen werden nicht extra bezeichnet):

- *Stimmlage und Stimmgattung*
 - Mann (8x), Tenor (7x), Bariton (5x), Kombination Bass-Bariton (2x), Bass (1x), hohe Lage, mitteltief
- *Klangliche Beschreibung (inkl. Resonanz und Dynamik)*
 - Klang: dunkel/angedunkelt (3x), eng/gequetscht in hoher Lage (2x), Frosch im Hals, helle Tiefen, kernig, klar, Klangfarbe Tendenz zu quäkig, warmer Klang, knödelt, knarzend, manchmal belegt
 - Resonanz: kräftig/kraftvoll (7x), volle Resonanz/Maskenresonanz (2x), scharf, schreiend, (stimm)stark, Bruststimme
 - Dynamik: laut/forte (8x), kaum Steigerung/konstant laut (2x)
- *Technische Angaben (inkl. Spannung, Stimmstörungen)*
 - Spannung: angestrengt/anstrengend (6x), in der Höhe/Mittellage etwas gepresst (2x), zu viel Druck in hoher Lage, ungesund
 - Technik: klassische/Opernausbildung (4x), hinten gesungen, gute Höhe, Mund weit auf, sauber, gute Stütze, großer Tonumfang, direkte Ansprache der Töne

- *Stilistische Beschreibung (inkl. Vibrato)*
 - Stil: Töne werden lang gehalten/viele lange Töne (4x), Glissando, Legato, Sprechgesang
 - Vibrato (17x), Zusätze: angenehm (2x), sehr stark (2x), sehr stark betont, auffällig, natürlich, extrem, forciert, wenig
- *Artikulation und Text/Sprache*
 - deutliche Aussprache (5x), gerolltes r (4x), italienisch (3x), Vokale: ä sticht hervor, Schwa-Laut stark, Vokale ä und a eher dunkel, offen, Vokale o, i, u, ae sehr hell gefärbt, Konsonanten: stimmhaftes s, sch und k etwas scharf (2x)
- *Ausdruck*
 - Gelächter/Lachen/lachend (5x), Dramatik/dramatisch (4x), emotional (3x), ausdrucksstark (2x), pathetisch/Pathos (2x), aufgeregt, bedrohlich, Charakter böse/gehässig, gewaltig, leidend/traurig, leidenschaftlich/melancholisch, Mafia, pompös, Schmalz, sehnsüchtig, stellenweise leicht hysterisch, vielfältig: wütend, jammernd, dramatisch, vorantreibend

Die Zusammenfassung mehrfach genannter Begriffe wurde hier sehr eng vorgenommen (siehe z.B. Gelächter/Lachen/lachend), wodurch die Anzahl gleicher Nennungen einer Eigenschaft relativ klein gehalten wurde, um so viel Information wie möglich zu erhalten. Am häufigsten fand sich die Nennung des Vibratos (17-mal von 21 möglichen Nennungen), alle anderen Eigenschaften zeigten wenig genaue Übereinstimmungen (< 9). Insgesamt beschreibt eine Reihe von Begriffen den Eindruck des kraftvollen, lauten, resonanzreichen und angestregten Singens, was zusammengefasst einen durchaus ähnlichen Eindruck wiedergibt, wenn es auch begrifflich unterschiedlich gefasst wurde. Im Vergleich zu der Beschreibung von Brandenburg (2016) finden sich eher negativ konnotierte Begriffe von vergleichbaren Eigenschaften. Die folgende Gegenüberstellung macht dies deutlich (die von Brandenburg verwendeten Begriffe sind kursiv gedruckt, dahinter finden sich vergleichbare Begriffe der Teilnehmer):

- *große Schallkraft, Strahlkraft, Durchschlagskraft, präziser Stimmsitz* vs. Lautheit, Kraft, volle Resonanz
- *wundervoll leicht ansprechende Höhe, ansatzloses Attackieren der hohen Töne, ausgezeichnete Atemkontrolle* vs. gute Höhe/Stütze, direkte Ansprache der Töne, eng/gequetscht/zu viel Druck/gespresst in hoher Lage, angestrengt, scharf, schreiend, ungesund,
- *silbriger Klang, brillante Höhe* vs. scharf, schreiend, quäkig, dunkel/abgedunkelt, knödelt, helle Tiefen, klar, kernig, warm
- *nahezu perfekte Gesangstechnik* vs. klassische (Opern-)Ausbildung
- *schnörkelloser Gesangsstil* vs. Glissando, Legato, Sprechgesang, Vibrato wird meist positiv beschrieben (angenehm, natürlich), aber auch negativ mit sehr stark/betont, extrem, forciert

Aus diesen (z.T. groben) Zuordnungen wird die gegensätzliche Konnotation der Begriffe deutlich: Während Brandenburg die Stimme Björblings äußerst positiv beschreibt, nutzen die Teilnehmer teilweise negative Beschreibungen. Die Artikulation wird im Text Brandenburgs nicht spezifischer thematisiert. Die Teilnehmer benennen die italienische Sprache, eine deutliche Aussprache und das gerollte (Zungenspitzen-)/r/. Außerdem beschreiben sie mehrfach den Ausdruck des Sängers bezogen auf diese spezifische Arie, was natürlich in dieser Spezifität nicht von Brandenburg erwähnt wird, da er die Arie nicht behandelt.

Grundsätzlich lässt sich sagen, dass die Teilnehmer neben neutralen Beschreibungen und wenigen positiven Beschreibungen (vor allem in Ausdrucksbeschreibungen) eher negativ auffällige Merkmale benennen, die sich vor allem auf klangliche Merkmale beziehen. Die genauen Gründe für diese negativ konnotierten Beschreibungen bleiben in dieser Untersuchung offen. Durch eine nachträgliche Abfrage wurde kontrolliert, dass den Teilnehmern weder der Sänger noch die gesungene Arie bekannt waren. Möglicherweise spiegelt sich in dieser Unkenntnis des Stücks und des Sängers ein gewisses Desinteresse oder sogar eine Ablehnung des Musikstils, die zusätzlich zu den negativen Einschätzungen beiträgt. Da das Gefallen nicht extra abgefragt wurde, kann hier nur spekuliert werden. Daher stellt sich die Frage, ob die spezifische Gesangsdarbietung von Björbling nicht gefällt oder ob generell eher Merkmale beschrieben werden (können), die negativ auffallen. Zwar kann diese Frage mit der vorliegenden Befragung nicht beantwortet werden, es kann aber vergleichend der Gesang von Bessie Smith angesehen werden, der der Populärmusik zugeordnet werden kann und somit einen anderen Gesangsstil repräsentiert.

3.2.2 Offene Beschreibungen zum Gospel-Sänger

Für die Stimme Smiths gab es 69 Beschreibungen, die sich wie folgt auf die Kategorien verteilten:

- *Stimm Lage und Stimmgattung*
 - Frau (9x), Alt (5x), tief (5x), Kombination Alt/Mezzo (2x), weiblich (2x), eher männlich
- *Klangliche Beschreibung (inkl. Resonanz und Dynamik)*
 - Klang: kehlig (2x), nasal (2x), warm, weich, Overdrive
 - Geräuschhaftigkeit: kratzen/kratzig (4x), dreckig (2x), rasping (2x), anzerren, ‚arr‘, geräuschvoll, knurren bei i, räuspernd, abrotzen
 - Resonanz: stark (2x), volle Stimme/röhrig (2x), ausfüllend, kräftig
 - Dynamik: mp/mf
 - Register: Bruststimme (5x), Belting; Zusätze: viel, primär, Bruststimmlastig, ohne Kopfstimme
- *Technische Angaben (inkl. Spannung, Stimmstörungen)*
 - Technik: Töne: geringer Ambitus (2x); keine klassische Ausbildung; hinten geschlossen, trotzdem locker; strahlt Gesundheit aus

- *Stilistische Beschreibung (inkl. Vibrato)*
 - Stil: Blues (4x), moderner/frühes 20. Jh. (3x), Chanson (2x), Jazz, Swing, Blue Notes, leichte Glissandi; Anmerkungen zum Gesprochenen (8x): Parlando, singt erzählend, sprechend, Sprechgesang (3x), Sprechstimme eingebaut/kurz (2x)
 - Glissando (6x) in Form folgender Beschreibungen: schleift Töne an, Anschleifen der Töne, Slides
 - Vibrato (11x), Zusätze: bei gehaltenen Tönen, leichtes Vibrato (3x), schnelles Vibrato (2x)
- *Artikulation und Text/Sprache*
 - weiche Artikulation, enge Aussprache, Englisch (3x), Konsonanten: stimmhafte Konsonanten werden ausgekostet/zieht Laute/singt auf Klinger ‚lovinnn‘ (3x), Verständlichkeit nur teilweise gegeben (2x), Vokale: viele o/ö Laute
- *Ausdruck*
 - Abwechslungsreich/facettenreich (2x), ausdrucksstark (2x), emotional (2x), gemütliche Stimme/Tempo moderato (2x), selbstbewusst (2x), variiert nicht stark in der Tonhöhe (2x), authentisch (erzeugt Bilder beim Zuhören), besonders, durchlässig, lässig, modern, Raunen, verführerisch
- *anderes*
 - Ethnisch: farbige Stimme, schwarz, Südstaaten

Grundsätzlich scheinen die Teilnehmer die Darbietung von Bessie Smith besser zu bewerten als die von Björling, denn es finden sich keine eindeutig negativ konnotierten Begriffe. Neben den neutralen Beschreibungen einer Alt-Stimme ohne klassische Ausbildung und einer stilistischen Einordnung in den Blues wird die positive Beschreibung eher im Ausdruck deutlich: abwechslungsreich, authentisch, selbstbewusst, lässig, gemütlich, emotional. Begriffe zur Beschreibung des Ausdrucks werden häufiger angegeben als andere Kategorien.

Besonders interessant ist, dass es eine Reihe von Anmerkungen zu Sprechanteilen im Singen gibt. Eine Anmerkung zum Sprechgesang fand sich bei Björling nur einmal. Während klangliche Merkmale eher weniger genannt werden (nur kehlig, nasal, warm, weich), wird eine Geräuschhaftigkeit im Klang öfter benannt, d.h. das Merkmal der Rauigkeit wird mit verwandten Begriffen umschrieben. Wie bei Björlings fallen die meisten Angaben auf das Vibrato. Häufiger als bei Björling werden bei Smith der Stil, das Glissando und die Geräuschanteile genannt, bei Björling hingegen werden insgesamt mehrere und andere klangliche Merkmale benannt.

Der Vergleich mit der Beschreibung von Pfeleiderer (2016b) lässt einige Übereinstimmungen mit dem Höreindruck der Teilnehmer erkennen:

- *kräftige Stimme, Blues-Shouterin* vs. kräftig, ausfüllend, stark, volle Stimme, röhrig, Belting, Bruststimme,

- *Tonumfang (größer für die Stimme, geringer im Lied, hier nur von einer Oktave)* vs. Töne geringer Ambitus, variiert nicht stark in Tonhöhe
- *Gleiten/Glissandi (hier differenzierter/detaillierter beschrieben)* vs. schleift Töne an etc.
- *schnelles Vibrato* vs. Vibrato schnell, leicht

Im nicht abgedruckten Teil des Texts von Pfeleiderer stehen noch weitere Angaben, die sich mit denen der Teilnehmer decken, z.B. zum „langsamen Tempo“, der Artikulation und der Bezeichnung Smith's als „Empress of the Blues“ sowie zu ihrem „emotional intensiven Vortrag“ (Pfeleiderer, 2016b, S. 561). Die Konsonantendehnungen werden allerdings nicht benannt, nur eine „bisweilen auch recht eigenwillig[e]“ Artikulation (Pfeleiderer, 2016b, S. 561). Auch kann der Text als eher technische Beschreibung der Stimme bewertet werden, der wenige Angaben zum Ausdruck bzw. der Rezeption der Gesangsvorträge macht.

3.2.3 Zuordnungstest

Für die Auswertung des Zuordnungstests wurden die Fragebögen in Tabellen übertragen und die unterstrichenen Aussagen kodiert und quantitativ ausgezählt. Die Anzahl an Markierungen (d.h. unterstrichene Textstellen) kann in den Textauszügen im Anhang 8.1.1 nachgesehen werden.

Die einzige Vermutung zum Namen eines Sängers war „Pavarotti“, wenn auch mit Fragezeichen versehen und fälschlicherweise zur Beschreibung Björlings angeführt. Kein anderer Teilnehmer hat Vermutungen zu den Namen der Tenöre oder der Sängerinnen angestellt. Eine Abfrage nach Abschluss der Studie hat ergeben, dass Björling gänzlich unbekannt war und Bessie Smith und Mahalia Jackson nur teilweise bekannt waren.

Die Zuordnungen waren zwischen den klassischen und den Gospel-Sängerinnen unterschiedlich erfolgreich. Die drei Tenöre konnten den Beschreibungen nicht zugeordnet werden: Luciano Pavarottis Beschreibung wurde nur in sieben (von 20) Fällen korrekt zugeordnet, in acht Fällen wurde seine Beschreibung mit Domingos und in fünf Fällen mit Björlings verwechselt. Placido Domingos Beschreibung wurde in nur sechs (von 21; die unterschiedlichen Summen ergeben sich durch die nicht vollständig ausgefüllten Fragebögen) Fällen ihm selbst und weiteren sechs Pavarotti zugeschrieben, allerdings in neun Fällen Björling. Björlings Beschreibung wurde in sieben (von 19) Fällen korrekt zugeordnet, in weiteren sieben Pavarottis und in fünf Fällen Domingos. Das bedeutet, dass keine direkte Verbindung der Beschreibungen und der Hörbeispiele hergestellt werden konnte (die Details der Auswertung können im Anhang 8.1.2 eingesehen werden).

Aus der näheren Betrachtung der Markierungen können weitere Erkenntnisse zur Genauigkeit der Zuordnungen von Text und Sänger gewonnen werden. Die Teilnehmer sollten die charakteristischen Stellen der Rezensionen markieren, die den Ausschlag für ihre Zuordnung gegeben haben, also von ihnen als charakteristisch für die jeweilige Stimme gesehen wurden. Deshalb wurde die Anzahl an

Markierungen ausgezählt und ausgewertet, welchen Sängern welche Merkmale zugeordnet wurden. So konnte zum Beispiel ein Merkmal im Beschreibungstext zu Pavarotti auch als charakteristisch für Domingo und/oder Björling markiert werden, je nachdem welchem Sänger der Text zugeordnet wurde. Die Codes in Klammern beziehen sich auf die Tabellen im Anhang 8.1.2.

Pavarotti und Domingo wurden auf der Basis fast aller Merkmale vertauscht; nur die Anspannung (A13) wurde häufiger korrekterweise als Indiz für Pavarotti markiert. Wurde allerdings eine Beschreibung Domingos oder Pavarottis dem Sänger Björling zugeordnet, wurden als charakteristische Merkmale ebenfalls die Anspannung (A13)/Anstrengung (B12+B13) markiert, aber auch der ‚Macho-Klang‘ (B8) sowie die kehligen Einfärbungen (B9). Korrekte Zuordnungen von Björling fanden eher auf der Basis des silbrigen Klangs (C9) und einer brillanten Höhe (C10) statt, allerdings auch nicht eindeutig.

Im Gegensatz zu den Tenören wurden die drei Gospelsängerinnen mehrheitlich korrekt zugeordnet: Aretha Franklin wurde 15-mal (von 20) korrekt zugeordnet und nur zweimal mit Bessie Smith und dreimal mit Mahalia Jackson verwechselt. Jackson wurde elfmal (von 21) korrekt zugeordnet, allerdings siebenmal mit Smith und dreimal mit Franklin verwechselt. Gleichermaßen wurde Smith zwölfmal (von 21) korrekt identifiziert und siebenmal mit Jackson verwechselt, nur zweimal mit Franklin. Franklin war somit am besten zu ihrer Beschreibung zuzuordnenden, während Jackson und Smith mehrfach verwechselt wurden.

Nach den Markierungen in den Beschreibungen wurde Franklin wegen des Belts (die Bruststimme in höheren Lagen) als besonders charakteristische Eigenschaft korrekt zugeordnet, d.h. anhand der weit ausgedehnten Bruststimme bis zum g^2 , dem nasalen Stimmklang, dem sukzessiven Ansteigen von Tonhöhe und Lautstärke, den Dehnungen von Vokalen und Konsonanten sowie den Glissandi und Melismen. Smith teilt sich kaum Merkmale mit ihr.

Im Falle von Jackson wurden nicht so häufig die gleichen Aussagen unterstrichen, auf deren Grundlage sie zugeordnet wurde. Elfmal wurde sie aufgrund der Beschreibung einer tiefen Stimme zugeordnet, die aber auch viermal mit Smith assoziiert wurde. Die Rauheit wurde ihr achtmal (viermal Smith) zugeordnet, zusätzlich ein Anschleifen der Töne von unten. Das regelmäßige Vibrato wurde ihr fünfmal, Smith hingegen siebenmal zugeordnet. Nur ihr wurde zweimal das Merkmal der ‚Blues-Sängerin‘ zugeordnet, den anderen nie.

Smith wurde vor allem wegen der Beschreibung des schnellen Vibratos korrekt zugeordnet (zwölfmal, plus die Vibrati an Phrasenenden viermal, im Vergleich zu fünfmal schnelles Vibrato für Jackson), aber auch wegen der auf- und abwärtsgleitenden Glissandi (achtmal bzw. sechsmal und nur ein- bis zweimal bei den anderen) und der kräftigen Stimme (viermal), allerdings fast gleichauf mit Jackson (dreimal).

Da die Beschreibungen Doppelungen enthielten, ist vor allem die Kombination der Merkmale von Bedeutung, d.h. sowohl für Franklin als auch Jackson werden Beschreibungen mit Glissandi

unterstrichen; auch werden immer zwei Sängerinnen mit Gospel-, Blues- oder Gospelsängerinnen assoziiert. Weshalb Jackson und Smith öfter miteinander verwechselt werden als mit Franklin kann nur vermutet werden: Es könnte daran liegen, dass sie in einer tieferen Lage singen und somit der Belt nur bei Franklin überhaupt deutlich hervortritt.

3.3 Diskussion

Die Aufgabe, drei Stimmen den entsprechenden Beschreibungen zuzuordnen, zeigte das eindeutige Ergebnis, dass dies den Teilnehmern nicht möglich ist: Die Tenöre konnten gar nicht und die Gospel-Sängerinnen nur eingeschränkt den entsprechenden Beschreibungen zugeordnet werden. Während diese direkte Zuordnung eines Sängers zu einem vollständigen Text nicht erfolgreich war, gab es einige Merkmale, die eine Zuordnung erleichterten, weil sie charakteristisch genug waren, um auf nur eine Stimme zuzutreffen. Die Ergebnisse werden im Folgenden auch unter Berücksichtigung methodischer Restriktionen diskutiert.

Beim Betrachten der Verteilung an unterstrichenen Aussagen zeigt sich, dass eigentlich alle Aussagen des Textes mindestens einmal unterstrichen wurden und sich oft nur wenige Übereinstimmungen in derselben Textstelle fanden. Dies kann verschiedene Ursachen haben und an der Relation der Anzahl an Teilnehmern ($N = 21$) zur Länge der Texte liegen, d.h. die Möglichkeiten zum Unterstreichen waren unverhältnismäßig groß, sodass nur wenige Markierungen übereinstimmen können, erst recht dann, wenn sowieso Uneinigkeit herrscht. Die Beschreibungen, vor allem die der klassischen Sänger, sind zum Teil sehr allgemein und die Auswahlmöglichkeiten somit sehr unspezifisch, weswegen die Texte jeweils auf mehrere Stimmen zutreffen können. Obwohl davon ausgegangen werden kann, dass die Autoren der Texte eine große Kenntnis von den Sängern haben und diese auch wortgewandt darstellen können, waren alle Beschreibungen nicht eindeutig genug, um sie sicher mit einer Stimme zu verknüpfen. Inwiefern Kenner (oder Fans) die Stimmen eines Domingo oder Pavarotti unterscheiden können, kann nur vermutet werden, womöglich, weil sie anhand von Aufnahmen gelernt haben, diese beiden voneinander zu unterscheiden. Möglicherweise unterstützt eine Beschreibung schon vorhandenes Expertenwissen zu der Stimme eines berühmten Opernsängers oder fasst sie für Fans in nachvollziehbare Worte, wenn diese wissen, auf wen sich die Rezension bezieht. Für die Hörer in der vorliegenden Studie allerdings weist offenbar die Darbietung durch den legitimierte klassischen Gesang keine ausreichenden Spezifika auf, auf deren Basis eine Zuordnung ausgeführt werden kann. Eine geringe Kenntnis bzw. mangelndes Interesse der Hörer kann ein Grund für die schlechte Zuordnung gewesen sein, allerdings besteht auch die Möglichkeit, dass die Beschreibungen bei den Sängerinnen besser zu den Stimmen passten, da sie zusätzlich auf die Songs zugeschnitten waren (bis auf Smith). Die Beschreibungen für die Sängerinnen zeichnen sich durch eher technische und weniger ausschmückende Beschreibungen als bei den Tenören aus, auch waren sie kürzer, teilweise prägnanter

und dadurch aussagekräftiger. Allerdings wurde lediglich Aretha Franklin mit hoher Zahl korrekt zugeordnet, während Jackson und Smith ebenfalls öfter verwechselt wurden. Es gab besonders bei den Sängerinnen Überschneidungen bei den Glissandi und der tiefe Stimme.

Auch wenn es nicht sonderlich überrascht, dass es einfacher ist, stärker voneinander verschiedene Stimmen charakteristisch zu beschreiben, und dass sie somit auch besser unterschieden und zugeordnet werden können (wie Franklin im Vergleich zu Jackson oder Smith), so ist doch bezeichnend, wie gering die richtige Zuordnungsrate selbst bei musikalisch gebildeten Teilnehmern ist. Möglicherweise erschwert der standardisierte, legitimierte Gesangsstil der klassischen Tenöre (vgl. Hähnel et al., 2014), die sich teilweise bemühen einen historisch-informierten Aufführungsstil umzusetzen, die korrekte Zuordnung zusätzlich. Im Gesang populärer Stile besteht in dieser Hinsicht immerhin etwas mehr Spielraum – der individuelle Stil ist hier prägender und somit auch charakteristischer.

An den offenen Beschreibungen der Stimmen konnten allerdings gewisse Überschneidung mit den Vorlagentexten gesehen werden. Während sich bei Bessie Smith sogar ähnliche Begriffe zwischen Hörern und dem Autor finden, werden bei Björling zumindest ähnliche Konzepte erkennbar, die der Autor unter Verwendung eindeutiger Fachbegriffe beschreibt, während die Hörer wesentlich einfachere Begriffe verwenden. Wesentliche Begriffsunterschiede zwischen beiden Sängern widerspiegeln bisher dargestellte Unterschiede klassischer und populärer Gesangsstile, wie die Geräusche, Glissandi, die individuelle Artikulation und die Sprechnähe (siehe Kapitel 2.3.1). Auch finden sich einige bereits in anderen Kontexten genannte Begriffe zur Stimmbeschreibung (siehe Kapitel 2), wie eng, quäkig, scharf und gepresst (zu viel Druck). Die Hörer beschränken sich bei ihrer Beschreibung der Stimmen allerdings nicht auf klangliche Eindrücke, sondern beschreiben auch den allgemeinen Ausdruck des Sängers und seine Wirkung auf den Hörer.

Es soll nun nicht besprochen werden, welche Aufgaben und Funktionen Kritiken (im Journalismus) haben, aber nach diesen Ergebnissen kann das Selbstverständnis des Kritikers in Frage gestellt werden, der davon ausgeht, eine Stimme mit Worten hörbar zu machen. Kritiken funktionieren am besten bei einem Hörer, dem der Sänger bekannt ist und der bereits eine klangliche Vorstellung vom Sänger hat. So ist davon auszugehen, dass die Beschreibung der Stimme den bereits vorhandenen Höreindruck untermalt. Allerdings kann auf Basis der vorliegenden Ergebnisse davon ausgegangen werden, dass die Beschreibungen austauschbar und die Texte eher als ein Stück Prosa zu verstehen sind, dessen Worte an sich einen Wohlklang bilden. Als charakteristische, zuordnungsfähige Beschreibung einer Stimme funktionieren sie nicht. Auch kann weiterführend die Frage gestellt werden, ob Hörer doch nicht so gut im Unterscheiden von Stimmen sind, wie sie selbst oder Experten es glauben zu sein. Inwiefern ein

geschulter, interessierter Hörer in einem kontrollierten Experiment einen Domingo wirklich zuverlässig von einem Pavarotti unterscheiden kann, kann also ebenfalls in Frage gestellt werden.

4 Beschreibung und Ablehnung vokaler Merkmale von Singstimmen – Interview-Studie

Die vorliegende Studienreihe verfolgt zum einen das Ziel, ein Inventar zu erstellen, mit dem Nicht-Experten Singstimmen charakteristisch beschreiben können. Zum anderen sollen die erhobenen vokalen Merkmale von Singstimmen mit ästhetischen Urteilen und ausgelösten Gefühlen in Verbindung gesetzt werden. Dazu wurde im ersten Schritt ein breiter Ansatz gewählt, der qualitative und quantitative Methoden verbindet. Die Teilnehmer äußerten Stimmbeschreibungen zuerst in Interview-Sessions im Kontext abgelehnter Stimmen. Im Anschluss wurde ihnen eine Liste mit Experten-Begriffen zur Beschreibung des vokalen Ausdrucks vorgelegt, mit denen überprüft werden sollte, welche Begriffe die Teilnehmer verstehen und korrekt nutzen können. Dazu diente ein direkter Vergleich der Teilnehmer-Ratings mit Experten-Ratings zu jeder untersuchten Stimme.

Die Beschreibung der abgelehnten Stimmen wurde nicht nur auf stimmliche Merkmale beschränkt, sondern auf Basis eines Leitfadenterviews auch die Gründe für die Ablehnung erfragt. Somit liegt mit der vorliegenden Studie eine erste, umfassende Arbeit zur Ablehnung von Singstimmen vor, die sich nicht nur auf vokale Merkmale bezieht, sondern auch auf persönliche und emotionale Gründe. Für eine genaue Einführung in die Studienreihe sei der Leser auf Kapitel 2.6 verwiesen.

4.1 Methoden

4.1.1 Auswahl der Forschungsmethode

Um sich dem Gegenstand Stimme zu nähern, wurde eine Methoden-Triangulation eingesetzt, eine methodische Technik, die zu einer breiteren und tieferen Erfassung des Untersuchungsgegenstands führt (Denzin, 1994; Flick, 2007) und in den vorliegenden Untersuchungen eine Kombination qualitativer und quantitativer Forschungsansätze beinhaltet. Zu Beginn des Forschungsprojekts wurden die Forschungsfragen formuliert (vgl. Flick, 2007, S. 132), sodass daraus ein Konzept für die Durchführung der Untersuchung entwickelt werden konnte:

- Wie beschreiben Nicht-Experten vokale Merkmale?
- Welche Merkmale der Stimme werden abgelehnt?
- Wie wird die Ablehnung der Stimme begründet?

Nach der Formulierung der Gesamtfragestellungen wurden spezifische Forschungsfragen formuliert und unter Berücksichtigung der Untersuchungsgruppe ein Forschungsdesign entwickelt (vgl. Flick, 2007, S. 134).

- Inwiefern ist die Ablehnung einer Stimme assoziiert mit der Ablehnung des Musikstils, Texts und anderer möglicher Einflussfaktoren wie dem Eindruck von der Persönlichkeit des Sängers, Assoziationen des Liedes mit spezifischen Situationen etc.?

- Wie unterscheiden sich die beschriebenen Merkmale zwischen Experten und Nicht-Experten?

Die Studie zielte darauf ab, eine offene und unbeeinflusste Merkmalsbeschreibung von abgelehnten Stimmen sowie die Gründe dafür und Reaktionen auf die Stimme zu erfassen. Als Basisdesign kann im Sinne Flicks von einer Vergleichsstudie gesprochen werden: „Bei vergleichenden Studien wird der Fall nicht in seiner Komplexität und Ganzheit betrachtet, sondern eine Vielzahl von Fällen in Hinblick auf bestimmte Ausschnitte. [...] Konsequenterweise stellt dieses Design einen Weg zur fokussierten und damit auch begrenzten komparativen Analyse von Themen oder Erfahrungsweisen dar.“ (Flick, 2007, S. 179). Beim Sampling der Teilnehmer wurde darauf geachtet, dass sie die abgelehnten Stimmen selbst aussuchten, damit dieser Vergleich möglich gemacht werden konnte. Zur Datenerhebung eignete sich hier besonders das Interview, in das auch gezielte Fragen eingebunden wurden. Ausgewertet werden sollte mit computergestützten, kodierenden Verfahren wie der Inhaltsanalyse (Kapitel 4.1.6). Der Forschungsprozess ist linear (vgl. Flick, 2007, S. 186), d.h. die Forschungsschritte waren vorab festgelegt und klar aufgeteilt und erlaubten eine Kombination qualitativer und quantitativer Forschungsansätze. Auch das Sampling wurde vorab festgelegt, da gezielt Nicht-Experten im Umgang mit der vokalen Ausdrucksbeschreibung untersucht werden sollen. Als Form des Interviews eignet sich das fokussierte Interview (Flick, 2007, S. 195; Misoch, 2015, S. 83ff.), welches auf Merton und Kendall (1979) zurückgeführt wird. Wichtig ist, dass sich das Interview auf eine konkrete Situation oder einen bestimmten Stimulus bezieht, „wobei das darauffolgende Interview die Reaktionen des Individuums auf diesen Stimulus untersucht“ (Misoch, 2015, S. 83). Dem wird ein Leitfaden zugrunde gelegt, welcher vier Kriterien der Interviewdurchführung genügen müsse: Nichtbeeinflussung der Interviewpartner, Spezifität der Sichtweise und Situationsdefinition aus deren Sicht, Erfassung eines breiten Spektrums der Bedeutungen des Stimulus sowie Tiefgründigkeit und personaler Bezugsrahmen aufseiten des Interviewten (Flick, 2007, S. 195; Merton & Kendall, 1979, S. 178; Misoch, 2015, S. 84f.).

- Die *Nichtbeeinflussung* soll durch verschiedene Frageformen realisiert werden wie unstrukturierte Fragen („Was fiel Ihnen an diesem Film besonders auf?“) und halbstrukturierte Fragen, die z.B. auf einen konkreten Gegenstand hinweisen („Was empfanden Sie in dem Moment als...“) (vgl. Flick, 2007, S. 195f.).
- Das Kriterium der *Spezifität* soll durch eine ‚explizite Bezugnahme auf die Stimulusituation‘ realisiert werden („Was war es genau, das Ihnen diesen Eindruck vermittelte?“). Als allgemeine Regel formulieren Merton und Kendall (1979, S. 191), dass die spezifizierenden Fragen so explizit sein sollen, dass der Befragte nicht umhinkann, seine Antworten unmittelbar auf bestimmte Aspekte der Stimulusituation zu beziehen (vgl. Flick, 2007, S. 197).

- „Die *Erfassung eines breiten Spektrums* zielt darauf ab, dass alle für die Fragestellung relevanten Aspekte und Themen im Lauf des Interviews angesprochen werden“ (Flick, 2007, S. 197).
- *Tiefgründigkeit und personaler Bezugsrahmen*: Hier sei es Ziel, „ein Höchstmaß an selbstenthüllenden Kommentaren des Informanten“ (Flick, 2007, S. 198) darüber zu erhalten, wie er das Stimulusmaterial erlebt hat (vgl. auch Merton & Kendall, 1979).

Das Ziel dieser Methode ist es, subjektive Sichtweisen zu untersuchen, um z.B. Hypothesen für spätere quantitative Untersuchungen zu generieren, aber auch um experimentelle Ergebnisse vertiefend zu interpretieren. Im Folgenden werden das Sampling der Teilnehmer, der Leitfaden und der Ablauf der Interviews, der Fragebogen sowie die Auswertung dargestellt.

4.1.2 Teilnehmer

Die Teilnehmer wurden über die Datenbank des Max-Planck-Instituts für empirische Ästhetik Frankfurt/Main (MPIEA) rekrutiert. Der Text der Ankündigung war in Auszügen wie folgt:

Gibt es eine Stimme eines Sängers oder einer Sängerin, die Sie absolut nicht mögen? Dann laden wir Sie herzlich ein, an einer Interviewstudie am Max-Planck-Institut teilzunehmen, mit der wir versuchen, die Wirkungen von Stimmen zu ergründen. Das Interview dauert etwa eine halbe Stunde und findet in den Räumen des Instituts statt. Für die Teilnahme wird eine Aufwandsentschädigung in Höhe von 10 € ausgezahlt.

Haben Sie Interesse? Dann bitten wir Sie nun, uns zur Vorbereitung Ihres Besuchs eine **Aufnahme eines Stückes oder Songs zu schicken, in der Sie die Stimme eines professionellen Sängers oder einer Sängerin nicht mögen**. Die gewählte Aufnahme können Sie uns per URL (YouTube oder ähnliche Portale) oder als MP3- oder wav-Datei als Mailanhang an die Projektadresse senden. Sollten Sie zwei Einspielungen verschiedener Sänger haben, können Sie auch gerne zwei schicken. Bitte beachten Sie dabei:

Der Fokus liegt dabei auf stimmlichen Charakteristika einer professionellen Stimme (d.h. wie die Stimme klingt etc.) und nicht der Performance (falsche Töne, Versingen etc.).

Bitte wählen Sie jedoch **keine Aufnahmen**, die **Growlen** oder **Screamen** (Gesangstechniken im Metal) oder **elektronisch bearbeitete** Stimmen (z.B. Verzerrer) beinhalten.

Außerdem soll die Stimme, um die es geht, auf großen Teilen der Aufnahme klar zu erkennen sein, das heißt auch gut von etwaigen anderen Stimmen zu differenzieren sein (wenig Background-Sänger, kein Chor, keine Duetten o.ä.).

Nachdem Ihre Aufnahme bei uns eingegangen ist, senden wir Ihnen per Mail Terminvorschläge zu. In dem Interview werden wir Sie dann zu der von Ihnen eingesandten Aufnahme befragen.

Die Teilnehmer meldeten sich zuerst per Email mit einem Stimmbispiel, das den vorgegebenen Kriterien entsprechen musste und von der Studienleiterin angehört und mit den Vorgaben abgeglichen wurde. Folgende weitere Kriterien führten zum Ausschluss: zu starke Bearbeitung der Stimme („Auto tune“, Verzerrer), Stimme war schlecht zu hören (zu viel Hintergrundmusik oder zu dominante Backgroundsänger) oder sang nicht allein (Backgroundsänger, Duopartner). In einem Fall wurde eine Sängerin ausgeschlossen, die nur regional bekannt war. Die Teilnehmer hatten die Gelegenheit

mehrere Beispiele einzureichen oder im Falle einer Ablehnung der ersten Stimme eine zweite einzureichen. In den meisten Fällen entsprach die Auswahl bereits beim ersten Einreichen den Kriterien. Im Falle von der Aufnahme *An Angel* von der Kelly Family wurde nur der Gesangspart des Angelo (sein Bruder Paddy singt im Wechsel mit ihm) verwendet, das heißt, sie wurde vor dem Interview zusammengeschnitten.

20 Teilnehmer wurden rekrutiert, von denen drei jeweils zwei Stücke verschiedener Sänger mitbrachten, d.h. insgesamt floss die Beschreibung von 23 Stimmen in die Studie ein (siehe Tabelle 6). Die Teilnehmer (15 weiblich) waren im Mittel 38,55 Jahre alt, Standardabweichung 16,39. Unter den Teilnehmern waren 13 Studierende, drei Angestellte, drei Rentner, ein Berufsschüler; die Bildungsabschlüsse waren folgendermaßen verteilt: Zwei Teilnehmer mit Promotion, elf mit Universitätsabschluss, vier mit Abitur, zwei mit Fachhochschulreife, einer mit Berufsausbildung. Zwei haben/hatten beruflich mit Musik zu tun (Musikwissenschaftler), sechs spiel(t)en ein Instrument. Sieben Teilnehmer singen nicht, zehn singen nur für sich, drei unter Anleitung oder auch öffentlich.

Die mitgebrachten Songs konnten alle verschiedenen Stilen der populären Musik zugeordnet werden, bei zwei Sängerinnen liegt die Vermutung einer klassischen Ausbildung nahe (Sarah Brightman, Anneliese Rothenberger). Unter den SängerInnen waren neun weibliche, 13 männliche und ein Knabe (Angelo Kelly), vertreten mit 13 deutschen Titeln.

DV	Sänger/Künstlername (Name)	Stück/Song	Geschlecht
01	Kelly Family (Angelo Kelly)*	An Angel	m (Knabe)
02	Lena Meyer-Landrut	Satellite	w
03	Xavier Naidoo	Abschied nehmen	m
04	Gitte Haenning	Ich habe die Liebe verspielt	w
05	Helge Schneider	Katzenklo	m
06	Canned Heat (Adam Wilson)	Going up the country	m
07_1	Samuel Harfst	Privileg	m
07_2	Casper (Benjamin Griffey)	Hinterland	m
08	Anneliese Rothenberger	In mir klingt ein Lied	w
09	Annett Louisan	Das Spiel	w
10_1	Tim Bendzko	Nur noch kurz die Welt retten	m
10_2	Udo Lindenberg	Durch die schweren Zeiten	m
11	Maroon 5 (Adam Levine)	Misery	m
12	Bob Dylan	Don't think twice it's all right	m
13	Michelle	Paris	w
14	Sarah Brightman	Think of me	w
15	Anastacia	I'm outta love	w
16	Mariah Carey	My all	w
17	Herbert Grönemeyer	Männer	m
18_1	Max Herre	Blick nach vorn	m
18_2	Nena	Nur geträumt	w
19	Passenger (Mike Rosenberg)	Let her go	m
20	The Tallest Man on Earth (Kristian Matsson)	Love is all	m

Tabelle 6: Liste der mitgebrachten Stücke. * Auszug/Zuschnitt Angelo Kelly.

Die Interviews fanden in einem Raum am MPIEA statt, der extra für die Durchführung von Interviews eingerichtet ist. Die Teilnehmer und der Interviewer saßen bequem in einer Sofaecke mit Tisch. Die Aufnahmen wurden zu Beginn des Interviews gemeinsam über die Lautsprecher im Raum angehört.

Ein Interviewer interviewte jeweils eine Person. Zehn Interviews wurden von einem jungen Mann und zehn von einer jungen Frau durchgeführt, um einem Interviewer-Bias entgegenzuwirken. Ein Interview dauerte ca. 20 Minuten, im Fall der drei Teilnehmer, die zwei Stücke mitgebracht hatten, entsprechend doppelt so lange.

4.1.3 Ablauf des Interviews und Interviewleitfaden

Nach einer Begrüßung und dem Bedanken für die Teilnahme wurde über den Ablauf der Sitzung informiert (Formalien, Interview mit Anhören der Aufnahme, abschließender Fragebogen). Zuerst wurden das Studienteilnehmerinformationsblatt, die Einwilligungserklärung für Interview und Datenverwendung und die Einwilligungserklärung zur Aufzeichnung des Interviews besprochen und ausgefüllt. Auch wurde darauf hingewiesen, dass keine der Fragen beantwortet werden muss, wenn der Teilnehmer dies nicht will, und dass das Interview jederzeit ohne Angabe von Gründen abgebrochen werden kann, ohne dass dem Teilnehmer dadurch Nachteile entstehen. Sie füllten dann einen Fragebogen zu ihren demographischen Daten aus (Alter, Geschlecht, höchster Bildungsabschluss und aktuelle berufliche Situation). Weiterhin wurde gefragt, ob der Teilnehmer beruflich mit Musik zu tun hat (ja/nein), ob er ein Instrument spielt (ja/nein) und ob er ‚regelmäßig‘ singt (ja, nur für mich; ja, unter Anleitung/Zuhörern (z.B. Chor, Band, Solo, im Unterricht); nein). Die in dieser und den folgenden beiden Studien verwendeten Methoden folgen den Richtlinien des Ethik-Rahmenantrages des Max-Planck-Instituts für Empirische Ästhetik, Frankfurt/Main.

Außerdem wurde die Form der Befragung besprochen: „Bitte lassen Sie sich nicht davon irritieren, wenn Sie das Gefühl haben, dass ich eine Frage mehrfach nur minimal anders formuliert stelle, und antworten Sie einfach noch einmal darauf. Möglicherweise treffen einige der Fragen auf die von Ihnen ausgewählte Aufnahme nicht ganz zu. Sollte das der Fall sein, sagen Sie das einfach.“ Vor dem Beginn des Interviews wurden etwaige noch offene Fragen geklärt. Erst danach wurde vor den Augen des Teilnehmers das Aufnahmegerät aktiviert und mit dem Interview begonnen.

Die Interviews basierten auf einem Leitfaden und entsprachen der Form eines Fokusinterviews (siehe Kapitel 4.1.1). Danach wurden direkt im Anschluss zwei Seiten Fragebogen (ebenfalls unter Aufzeichnung) vorgelegt und bearbeitet und eine Nachbesprechung schloss das Interview ab: „Haben Sie alle Fragen und Begriffe verstanden? Wie erging es Ihnen mit dem Fragebogen? Wie fanden Sie das Interview?“

Der Verlauf des Interviews bestand dementsprechend aus zwei Teilen. Der erste Teil fragte nach den Beschreibungen des Teilnehmers (qualitativ), der zweite Teil legte einen Fragebogen mit auszuwählenden Merkmalen zur Stimme vor (quantitativ). Der qualitative Teil des Interviews bestand aus 15 Fragen (der Leitfaden ist im Anhang 8.2 einzusehen), die zunächst auf eine Beschreibung der abgelehnten stimmlichen Merkmale abzielten, um anschließend auf Reaktionen und Erinnerungen einzugehen. Danach folgten einige Fragen zu anderen möglichen Gründen für die Ablehnung, die nichts mit dem Klang der Stimme oder ihren Merkmalen zu tun hatten. Stimmliche Merkmale wurden abgefragt mit ‚Was mögen Sie an der Stimme nicht?‘ und mit genaueren Nachfragen weiter eruiert, was an diesem Merkmal genau nicht gemocht wird, wobei jedes Merkmal gesondert abgefragt wurde. Außerdem wurde gefragt, wie die Stimme generell klingt, um noch eine weitere Beschreibungsebene zu eröffnen, d.h. zum einen noch einmal die Gelegenheit zur Beschreibung der abgelehnten Merkmale zu bekommen, aber auch um die Stimme generell zu beschreiben, neutral und ungeachtet der Ablehnung. Zugleich konnte damit untersucht werden, ob es den Teilnehmern möglich war, die Stimmen aus der Distanz zu beschreiben und sich nicht nur auf die abgelehnten Eigenschaften zu beziehen. Auch sollten die genannten Merkmale in eine Rangfolge gebracht werden, wie es auch im Fragebogen der Fall sein würde.

Fragen im zweiten Block zielten darauf ab, ob körperliche Reaktionen auftreten und sich die Ablehnung körperlich äußert bzw. festmachen lässt. Von Interesse war hier besonders die Nennung spezifischer Körperteile, die evtl. den Eindruck des Gehörten im Sinne eines sogenannten ‚funktionellen Nachvollzugs‘ (vgl. Nespital, 2013) festmachen lassen, im Sinne von: ‚mir schnürt das die Kehle zu‘, ‚da krieg ich einen Kloß im Hals‘ etc. Fragen zu Assoziationen sollten Verbindungen herstellen zu z.B. Umgebungsgeräuschen oder Klängen, die evtl. den beschriebenen Eindruck unterstützen. An die Assoziationen anknüpfend sollten mögliche Erinnerungen erfragt werden, die mit dem Stück verbunden werden. Wie die Fragen des letzten Blockes zielte auch dies darauf ab, herauszufinden, ob möglicherweise andere außer-stimmliche Gründe zur Ablehnung führten oder beitrugen, z.B. persönliche Verbindungen und Erfahrungen mit dem Sänger (‚ich habe den mal getroffen und der war mir unsympathisch‘ etc.) oder private/familiäre Hintergründe (‚das war der Lieblingssänger meiner Ex-Freundin‘ etc.). Außerdem wurde der Einfluss sowohl des Musikstils als auch des Texts auf die Ablehnung untersucht und von der Ablehnung stimmlicher Merkmale abgegrenzt. Die Frage, ob der Sänger evtl. nur in diesem Stück nicht gefällt, sollte diesen Ansatz unterstützen.

4.1.4 Fragebogen

Der Fragebogen im Anschluss bestand ebenso aus zwei Teilen, zum einen aus einer angepassten/gekürzten Version des Merkmalskatalogs nach Bose (2001) (vgl. Kapitel 2.6.1) und zum anderen aus einigen zusätzlichen Merkmalen, die eher generelle Beschreibungen von Stimmen beinhalten. Diese waren auf die Ablehnung ausgelegt, d.h. sie umfassten Beschreibungen zu einem

eventuell wahrgenommenen Krankheitswert stimmlicher Merkmale oder menschliche, persönliche Qualitäten wie der Frage nach Authentizität. Die Merkmale wurden aufgrund von Ergebnissen anderer Studien zusammengestellt (z.B. Ackermann, 2018; von Appen, 2007), in denen die Begriffe verwendet wurden, um den Eindruck von Stimmen zu beschreiben. Die Merkmale waren eventuell für den Teilnehmer leichter anzuwenden und konnten im Anschluss in der Analyse mit dem ersten Teil des Fragebogens in Verbindung gesetzt werden.

Wichtig war, dass nur die Merkmale angekreuzt wurden, die für die Ablehnung relevant waren, und nicht allgemein die Stimme anhand des Fragebogens bewertet werden sollte. Die Teilnehmer sollten erst für sich den Fragebogen ausfüllen und im Anschluss Rückfragen zu den Items stellen, die sie nicht verstanden hatten (durch das Fragezeichen gekennzeichnet). Der Interviewer hatte eine Liste mit Zusatzstichworten vorliegen und konnte diese bei Bedarf als Erklärungen angeben. Nicht zu allen Begriffen gab es Hilfestellungen und es wurde kontrolliert, dass keine weiteren Erklärungen gegeben wurden (siehe Anhang Tabelle 38).

Die Anweisung zum Fragebogen war wie folgt: „Beschreiben Sie anhand der Merkmale in der Liste, was Ihnen an der Stimme nicht gefällt. Beeinflusst ein Merkmal das Gefallen *nicht* bzw. spielt ein Merkmal keine Rolle, kreuzen Sie die Spalte **nicht wichtig (n.w.)** an. Können Sie mit einem Merkmal nichts anfangen bzw. verstehen Sie nicht, was damit gemeint ist, kreuzen Sie die Spalte mit dem Fragezeichen (**?**) an. Sie können diese Merkmale mit dem Interviewer besprechen.“

			n.w.	?	R	
Die Stimme ist zu...	<input type="checkbox"/>	tief	<input type="checkbox"/>	hoch	<input type="checkbox"/>	<input type="checkbox"/>
	<input type="checkbox"/>	dunkel	<input type="checkbox"/>	hell	<input type="checkbox"/>	<input type="checkbox"/>
	<input type="checkbox"/>	laut	<input type="checkbox"/>	leise	<input type="checkbox"/>	<input type="checkbox"/>
	<input type="checkbox"/>	gespannt	<input type="checkbox"/>	ungespannt	<input type="checkbox"/>	<input type="checkbox"/>
	<input type="checkbox"/>	klangvoll	<input type="checkbox"/>	klangarm	<input type="checkbox"/>	<input type="checkbox"/>
	<input type="checkbox"/>	weit	<input type="checkbox"/>	eng	<input type="checkbox"/>	<input type="checkbox"/>
	<input type="checkbox"/>	weich	<input type="checkbox"/>	hart	<input type="checkbox"/>	<input type="checkbox"/>
	<input type="checkbox"/>	dumpf	<input type="checkbox"/>	schrill	<input type="checkbox"/>	<input type="checkbox"/>
	<input type="checkbox"/>	behaucht			<input type="checkbox"/>	<input type="checkbox"/>
	<input type="checkbox"/>	rau			<input type="checkbox"/>	<input type="checkbox"/>
	<input type="checkbox"/>	gekarrt			<input type="checkbox"/>	<input type="checkbox"/>
	<input type="checkbox"/>	nasal (Näseln)			<input type="checkbox"/>	<input type="checkbox"/>
	Die Laute sind zu...	<input type="checkbox"/>	gedehnt	<input type="checkbox"/>	verkürzt	<input type="checkbox"/>
Die Tonveränderungen sind zu...	<input type="checkbox"/>	sprunghaft	<input type="checkbox"/>	gleitend	<input type="checkbox"/>	<input type="checkbox"/>
Das Singen erinnert mehr an...	<input type="checkbox"/>	Sprechen	<input type="checkbox"/>	Rufen	<input type="checkbox"/>	<input type="checkbox"/>
Die Stimme hat ein unpassendes...	<input type="checkbox"/>	Vibrato	<input type="checkbox"/>	Zittern	<input type="checkbox"/>	<input type="checkbox"/>
Die Stimmeinsätze sind zu...	<input type="checkbox"/>	weich	<input type="checkbox"/>	hart	<input type="checkbox"/>	<input type="checkbox"/>
Die Artikulation ist zu...	<input type="checkbox"/>	präzise	<input type="checkbox"/>	unpräzise	<input type="checkbox"/>	<input type="checkbox"/>
Die Stimme moduliert im Klang zu...	<input type="checkbox"/>	wenig	<input type="checkbox"/>	viel	<input type="checkbox"/>	<input type="checkbox"/>
Die Stimme moduliert die Töne zu...	<input type="checkbox"/>	wenig	<input type="checkbox"/>	viel	<input type="checkbox"/>	<input type="checkbox"/>

Tabelle 7: Fragebogen zur Beschreibung der abgelehnten Merkmale der Stimme mit zusätzlichen Angaben zum Verständnis und der Bedeutung (n.w. = nicht wichtig, ? = Bedeutung unklar, R = Ranking).

Nach möglichen Rückfragen wurden die Merkmale einem Ranking unterzogen: „Wie wichtig ist ein einzelnes Merkmal für die Beschreibung Ihres Missfallens? Schätzen Sie die wichtigen Merkmale auf

einer Skala von 1 (sehr wichtig), 2 (wichtig) bis 3 (nicht so wichtig) ein und tragen das Ergebnis in die letzte Spalte („Ranking“) ein.“

Nach dem Abschluss des Rankings wurde der zweite Teil des Fragebogens ausgefüllt (abgedruckt auf der Rückseite des ersten, damit nicht verglichen werden konnte): „Und zum Abschluss: Kreuzen Sie an, was auf Ihre ausgewählte Stimme zutrifft.“

- Die Stimme klingt...**
- angestrengt.
 - nicht authentisch.
 - alt.
 - langweilig.
 - krank.
 - schmalzig.
 - pathetisch.
 - kaputt.
 - dramatisch.

Tabelle 8: Teil 2 des Fragebogens.

4.1.5 Expertenbefragung

Um die Einschätzungen der Teilnehmer in Hinsicht auf die Eigenschaften der abgelehnten Stimmen vergleichen zu können, wurden drei promovierte Sprechwissenschaftler (Seminar für Sprechwissenschaft und Phonetik, Universität Halle) für die Experteneinschätzung der 23 Stimmen befragt. Den drei Experten war der originale Merkmalskatalog von Bose (2001) gut bekannt und sie waren überdies fachlich auch im Beurteilen von Singstimmen erfahren. Damit konnte gewährleistet werden, dass sie die Begriffe ähnlich verstehen. Den Experten wurde der gleiche Fragebogen (Teil 1) wie den Teilnehmern vorgelegt, anhand dessen sie alle Stimmen beurteilen sollten. Hier war wichtig, dass die Experten nicht bewerten (nach Gefallen oder Ablehnung), sondern die Stimme neutral/objektiv beschreiben. Bei allen Merkmalen sollte eine tendenzielle Ausprägung angegeben werden. Für den Fall, dass eine Einschätzung aus bestimmten Gründen nicht möglich war, konnte dies ebenfalls angegeben (in einer Spalte „NN“) und begründet werden (es gab eine Spalte für Kurzkommentare und am Ende des Fragebogens Raum für ausführlichere Kommentare). Die Anweisungen an die Experten umfasste zusätzliche Angaben zu den Items, die abweichend vom Original waren und können im Anhang 8.2.1 eingesehen werden.

Die Experten tätigten die Beschreibung unabhängig voneinander. Der Fragebogen und die Audiodateien wurden per Email verschickt und sie meldeten zurück, dass die Evaluation zwischen zwei und drei Stunden gedauert habe.

4.1.6 Analyse

Die Interviews wurden folgend der Qualitativen Inhaltsanalyse ausgewertet (Mayring, 2010, 2012). „Das Ziel der qualitativen Inhaltsanalyse ist die systematische Bearbeitung von Kommunikationsmaterial“ (Mayring, 2012, S. 468). Mayring formuliert für die Auswertungsverfahren bestimmende Grundsätze:

- „Das zu analysierende Material wird in seinem Kommunikationszusammenhang eingebettet verstanden [...]
- Die besondere Systematik der Inhaltsanalyse besteht in der Regelgeleitetheit (nach vorher formulierten Ablaufmodellen vorgehend), in der Theoriegeleitetheit (theoretisch abgesicherten Fragestellungen und Codierregeln folgend) und im schrittweisen, den Text in einzelne Analyseeinheiten zergliedernden, an Kategorien(-Systemen) orientierten Vorgehen [...].
- Die qualitative Inhaltsanalyse hat dabei den Anspruch, sich auch an Gütekriterien wie der Interkoderreliabilität zu messen. [...] Das Ziel bleibt aber, nachweislich zu ähnlichen Ergebnissen zu kommen.
- Dabei will sich die qualitative Inhaltsanalyse quantitativen Analyseschritten nicht verschließen, sondern versucht sie in den Analyseablauf begründet einzubeziehen“ (Mayring, 2012, S. 471).

An folgendem Ablaufmodell nach Mayring (Abbildung 1) wurde sich in der vorliegenden Studie orientiert. Durch die konkret formulierte Fragestellung und das Fokusinterview konnten bereits allgemeine Kategorien definiert werden. Unterkategorien und weitere Kategorien wurden schrittweise induktiv entwickelt, das heißt aus dem Material heraus (vgl. Mayring, 2010, S. 84, 2012, S. 472).

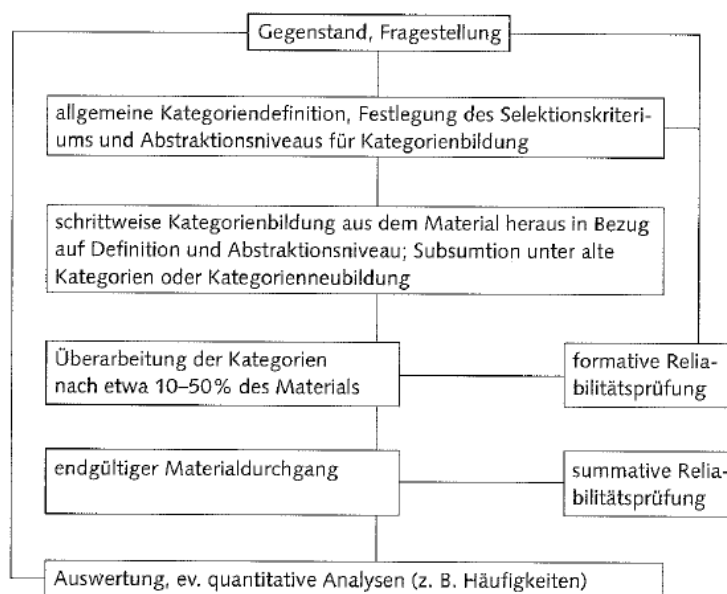


Abbildung 1: Ablaufmodell qualitativ-inhaltsanalytischer Verfahren am Beispiel induktiver Kategorienbildung (aus: Mayring, 2012, S. 472).

Die Interview-Studie hatte ein klares Ziel und verfolgte einen hypothesenbasierten Ansatz. Die Fokusinterviews wurden aufgezeichnet und transkribiert. Es handelte sich um eine weite Transkription, bei der Lacher, Hesitationen, evtl. Demonstrationen und Unverständliches notiert und angemerkt wurden. Verkürzte Silben wurden ebenso übernommen (eine = ´ne usw.).

Die Auswertung der Interviewdaten erfolgte mit der Computer-Software *Atlas.ti* (Version 7, Scientific Software Development, Berlin). An der Transkription und Kodierung waren drei Personen beteiligt, von denen die Versuchsleiterin die Leitung übernahm. Die drei Personen kodierten zu Beginn gemeinsam zwei Interviews, wobei die Kodierung besprochen und vereinheitlicht wurde. Anschließend arbeiteten alle Kodierer einzeln und setzten sich in regelmäßigen Abständen zusammen, um über Schwierigkeiten, Unklarheiten und den Kodierprozess zu sprechen und diese zu klären bzw. den Umgang damit zu vereinheitlichen. Die erste Kodierung folgte aufgrund der beschriebenen Forschungsfragen bereits einer grundlegenden Struktur:

- Abgelehnte Eigenschaften (Frage 3)
- Allgemeine Beschreibung der Stimme (Frage 6)
- Gründe für die Ablehnung (Fragen 4+5)
- Körperliche Reaktionen (Fragen 7+8)
- Abfragen zum Sänger (Fragen 2+11+13)
- Musikstil (Frage 12)
- Assoziationen mit dem Sänger/dem Stimmklang (Frage 9)
- Erinnerungen (Frage 10)
- Nachfragen/Verständnisfragen zum Fragebogen
- Feedback zum Fragebogen

Die Codes wurden permanent beim Kodieren ergänzt und das Codesystem regelmäßig angepasst, indem Codes zusammengefasst und überarbeitet wurden.

Die Daten der Fragebögen wurden in ein Tabellenkalkulationsprogramm übertragen. Die Daten wurden deskriptiv und nicht inferenzstatistisch ausgewertet (dieser Schritt der Überprüfung wurde in den Folgestudien unternommen). Die Auswertung der Fragebögen wurde mit den Interviews und den Experteneinschätzungen trianguliert und hinsichtlich folgender Fragestellungen untersucht:

- Welche Merkmale werden im Interview abgelehnt? Welche Gründe und Reaktionen für/auf die Ablehnung finden sich?
- Welche Begriffe wurden zur Beschreibung der Stimme bzw. der Ablehnung verwendet? In welche Kategorien lassen sie sich einordnen?

- Wie können die Begriffe (aus Expertensicht) verstanden werden? Waren die Teilnehmer in der Lage, ihren Eindruck umfassend zu beschreiben?
- Vergleich der Beschreibungen im Interview mit den Merkmalen im Fragebogen: Lassen sich die Beschreibungen durch die ausgewählten Begriffe erklären? Stimmen sie überein? Ergänzen sie sich?
- Vergleich der ausgewählten Merkmale des Teilnehmers mit denen der Experten: Stimmen die Merkmalsausprägungen überein? Und wenn nicht, wie wurden sie verstanden?
- Welches waren Merkmale der Ablehnung im Fragebogen? In welchem Verhältnis stehen sie zu den genannten im Interview?

4.2 Ergebnisse

Zuerst werden die Ergebnisse der Interviews dargestellt, beginnend mit den begleitenden Fragen, bevor die Merkmale der Sänger betrachtet und in den Zusammenhang zum Fragebogen gebracht werden.

4.2.1 Einfluss von Hintergrundwissen

4.2.1.1 *Sänger*

4.2.1.1.1 Stimme in anderen Kontexten

87 % der Teilnehmer (N = 23 bewertete Sänger) würden das Stück wahrscheinlich eher mögen, wenn es von einem/r anderen Sänger/in gesungen würde. Nur zwei Teilnehmer verneinten die Frage zur Gänze, die verbleibenden konnten es sich zwar teilweise nicht mit einem anderen Sänger vorstellen, gaben aber auch vereinzelt weitere nötige Veränderungen am Lied an sich an, damit es gemocht wird (den Text, das Arrangement; DV05, 07_2, 16). 82,6 % mögen diese Stimme auch in anderen Stücken nicht, bzw. können es sich nicht vorstellen, dass ihnen die Stimme gefallen würde. Die anderen Teilnehmer relativierten vorsichtig ihre Stimmablehnung (die Stimme sei in anderen Stücken etwas weniger unangenehm; DV02, 04, 07_2). Nur ein Teilnehmer gab an, die Stimme in anderen Stücken positiv zu bewerten: „Bob Dylan könnte eigentlich gut singen, also es gibt Stücke, da kann er gut singen“ (DV12).

4.2.1.1.2 Hintergrundwissen, Erinnerungen und Assoziationen

Die Kodierung von ‚Hintergrundwissen‘ ist nicht explizit mit einer der Leitfadenfragen verknüpft. Dennoch wurde kodiert, wenn ein Teilnehmer Hintergrundwissen über den Sänger hatte, welches meistens bei den Fragen 2) „Wer ist der Sänger?“ oder 13) „Was denken Sie über den Sänger? / Wie ist er?“ auftrat. Hier werden sowohl Hintergrundinformationen (aus externen Quellen) als auch Assoziationen (reine Vermutungen) benannt.

Nur bei 26 % der Teilnehmer zeigt sich in den Antworten (z.T. nicht überprüfbares) Hintergrundwissen über den Sänger, d.h. die Teilnehmer führen Belege an, indem sie sich auf Informationen aus den

Medien stützen, z.B. generell aus dem Internet (DV06, 14), ein gesehenes Interview (DV05, 18_2) oder Abbildungen (DV08, 11). Diese Hintergrundinformationen werden eher neutral gehalten, d.h. sie beziehen sich auf präsentierte Fakten, wie einen gewissen Erfolg (DV05, 08, 12, 06). Öfter als auf Fakten gestützte Meinungen werden Vermutungen über den Sänger angestellt (wozu die Teilnehmer auch ermutigt wurden) und darüber Rückschlüsse auf die Persönlichkeit gezogen (intelligent, selbstbewusst, sympathisch, bestimmten Gender-Vorurteilen entsprechend; DV05, 11, 14, 17, 08).

Einige Assoziationen gehen relativ weit (z.B. Assoziationen zu Herkunft, Alter und Äußerlichem) und sind nicht von weiterem Interesse, wenn sie nicht mit der Ablehnung in Verbindung stehen oder mit der Stimme in Verbindung gebracht werden, z.B. dass eine Sängerin evtl. Französin ist (DV09). Zum Teil finden sich jedoch auch andere, in Zusammenhang mit der negativen Bewertung stehende Vorurteile über die Sänger oder Unterstellungen zu ihren Intentionen: Zum Beispiel wird dem Künstler unterstellt, dass er nur in Abhängigkeit vom Publikumsgeschmack handele (DV08) bzw. „Es klingt für mich nicht, als wenn er Lust hätte zu singen, sondern als wenn er einfach nur da ist, um was zu machen, weil das sein Job ist und schnell, schnell und nicht, weil er irgendwie Emotionen da rein bringen will, was finde ich zumindest mit Musik zu tun hat, wenn man singt“ (DV10 über Lindenberg), oder es wird bedauert, dass der Sänger „das Potenzial der menschlichen Stimme nicht ausgenutzt hat“ (DV12 über Dylan). In diesen Fällen sind die Assoziationen erweiterte Ausführungen sowie Begründungen des Ablehnungsurteils.

In Verbindung mit dem Stimmeindruck stehen zum Beispiel folgende Aussagen: „Tja, sie ist eigentlich alt, also im mittleren Alter, aber klingt wie ein kleines Mädchen, also das finde ich irgendwie negativ.“ (DV13 über Michelle) oder „Das ist ein Weichei (lacht) [...] ich denke mir einfach, diese Stimme ist für mich schwächlich [...] das ist weinerlich“ (DV18_1 über Herre), „Ich finde, er hört sich unterdrückt an. Schon leidenschaftlich, aber irgendwie unterdrückt“ (DV01 über Kelly), womit keine körperliche Unterdrückung beschrieben wird, sondern eine nicht freie Stimmgebung, „Also jetzt ohne ihn recherchiert zu haben oder ihn zu kennen würde ich sagen, dass das ein alter Mann ist, der versucht cool zu sein, indem er seine Stimme noch tiefer macht und über irgendwelche Sachen singt, die sein Leben beeinflussen, aber so wie er seine Stimme verändert, klingt das für mich so, als hätte er das nie erlebt, sondern singt einfach nur irgendwas, was er denkt, was gut ankommt“ (DV10 über Lindenberg). Letztere Aussage bezieht bereits Faktoren der Glaubwürdigkeit einer Sängerpersönlichkeit ein und begründet über die kritisierte mangelhafte Authentizität zusätzlich die Ablehnung, wie auch in den folgenden Beispielen: „Ich stelle ihn mir immer so als kleinen Typ vor, der irgendwie versucht, durch seine Stimme sich als was Besonderes darzustellen. Weil die Stimme ja schon an sich besonders klingt und schnell wiedererkannt wird, aber sie gefällt mir einfach nicht“ (DV20 über The Tallest Man on Earth); „Irgendwie oberflächlich, sehr darauf bedacht, was andere Leute von ihr denken und langweilig. [...] Also sie ist mir auch nicht so sympathisch, deswegen“ (DV02 über Lena). Damit folgen die

Teilnehmer hier einer Authentizitätsästhetik, in der eine (wahrgenommene) Ehrlichkeit sowohl in der Musik als auch im Auftritt der Künstler ein wichtiges musikexmanentes Qualitätsmerkmal ‚guter‘ Musik ist (Berli, 2014, S. 191ff.; von Appen, 2007, S. 115ff.).

Es gibt aber auch Zuschreibungen positiver Eigenschaften, wie „sie ist ja auch anscheinend eine gute Sängerin“ (DV16 über Mariah) oder „Ich glaube, dass der nicht blöd ist. Ich glaube, dass der intelligent ist und ich glaube, dass der einfach eine völlig andere Art von Humor hat als ich“ (DV05 über Helge), d.h. trotz Anerkennung von positiven Merkmalen (Humor und Gesangstechnik) wird festgestellt, dass es einfach nicht dem eigenen Geschmack entspricht. In ähnlicher Weise wird sogar eine gewisse Faszination für die Stimme geäußert, wobei dennoch das Zuhören als ‚nicht schön‘ bezeichnet wird: „Aber auf der anderen Seite [...] ist es auch interessant. [...] Also es ist interessant, aber man findet es nicht schön zum Anhören längere Zeit dann.“ (DV07), „Ich kann mir Schöneres vorstellen zu hören“ (DV11) oder „es ist halt einfach, ja, nicht so schön (lacht)“ (DV15). Hier werden ein interessanter Stimmklang und ein nicht schöner Stimmklang gegenübergestellt.

Nur ein Teilnehmer (DV03 über Naidoo) erzählt davon, den Sänger in einer einmaligen Begegnung persönlich kennengelernt zu haben. Da er negative Erfahrungen berichtet, ist hier nicht auszuschließen, dass die Ablehnung durch das beschriebene, persönliche Erlebnis ausgelöst wurde.

34 % der Teilnehmer können von konkreten Erinnerungen im Zusammenhang mit dem Stück berichten. Neben vielen neutralen Erinnerungen, die auch schon teilweise in den Punkt Hintergrundwissen eingeflossen sind, wie das Gewinnen des Eurovision Song Contest (DV02) oder dass ein Song ein Hit war (DV05), gab es einige wenige persönliche Verbindungen. Diese waren nicht per se negativ, sondern auch neutral beschreibend gemeint, wie mit Rückfragen geklärt wurde. Andere Erinnerungen standen mit familiären Erinnerungen in Verbindung: „Nur an frühe Kindheit, mit meiner Mutter. Die hat das mal gehört gehabt. Und da hat es mich auch schon genervt“ (DV01). Zum Teil werden die Stücke auch mit der Erinnerung an eine bestimmte Zeit/einen Zeitgeist/etwas Zeittypisches verbunden, z.B. mit dem Fernsehen bzw. Musiksendungen (Peter Alexander, André Rieu, viele Streicher im Arrangement; DV08). Der Fernsehkonsum wird dabei als negativ dargestellt, d.h. es kann vermutet werden, dass die Ablehnung für das Lied aufgrund dieser assoziativen Verknüpfung mit der Zeit im Elternhaus zusammenhängt (DV08). Einmal wird die besondere Auffälligkeit der Ablehnung in einer eigentlich positiven Situation beschrieben: „als das Album rauskam, kam im Fernsehen da ja irgendwie so viel Werbung dazu und da war ich zu Hause mit meiner Familie und wir saßen davor und dann kam das halt, so ein kurzer Ausschnitt von dem Lied [...] das war ein schöner Moment irgendwie, aber die Musik hat mich in dem Moment halt irgendwie gestört“ (DV09).

Ein weiterer Teilnehmer verdeutlicht die Ablehnung der Stimme im Gegensatz zu einer positiven Bewertung des Stückes, welches mit anderen Sängerinnen sehr gemocht wird und auch selbst

einstudiert wurde und zu bestimmten Anlässen gehört/gesungen wurde, die zwar auch mal traurige Anlässe waren, aber trotzdem in positiver (z.B. in Trost spendender) Weise genutzt wurden (DV14).

4.2.1.2 Musikstil

Die doppelte Abfrage (innerhalb von Frage 12) nach dem Mögen des Musikstils fiel vergleichbar aus: 60,9 % der Teilnehmer mögen diese Art von Musik bzw. 73,9 % mögen den Musikstil. Verneinungen werden in manchen Fällen nicht weiter begründet, manchmal wird nur hinzugefügt, dass dieses Stück einfach nicht zu der Art von Musik gehört, die vom jeweiligen Teilnehmer gern gehört wird. Die Einstellung zu dem Musikstil an sich kann und soll hier nicht weiter untersucht werden.

4.2.1.3 Text

Der Textinhalt war den meisten Teilnehmern bekannt und konnte von 69,9 % wiedergegeben werden. Die anderen Teilnehmer begründeten ihre Unkenntnis meist damit, dass sie dem Text bisher keine Aufmerksamkeit geschenkt hätten. Auf die Frage (14) „Was denken Sie über den Text des Stückes?“ konnte herausgearbeitet werden, dass 56,5 % den Text mögen, d.h. sie äußern sich darüber positiv. Die anderen deuten Kritik an, wie „das ist so ein Text, der halt relativ schnell irgendwie genau für diesen Zweck halt runtergeschrieben wurde, von irgendeinem Songwriter [...] Der Text verkauft sich halt ganz gut“ (DV02). Andere beschreiben den Text als „nicht sehr geistreich“ (DV03) oder „banal“ (DV03, 05), „kitschig, typisch Schlager“ (DV13) oder „inhaltlich nicht so faszinierend oder toll, dass er mir gefallen würde.“ (DV10). Polemisch formulierte Kritik oder deutlich abwertende Sprache findet sich nur einmal, wie „Teenager-Scheiße [...] Das ist für mich [...] ein nicht ernst zu nehmender Text, der eine bestimmte Zielgruppe anspricht, mich nicht“ (DV18_2). Es muss jedoch angemerkt werden, dass keine extremen Texte vorlagen, d.h. keine radikalen politischen Inhalte o.ä.

4.2.2 Gründe für die Ablehnung und Reaktionen auf die Stimmen

Als Begründungen für die Ablehnungen wurden alle Argumente kodiert, die die Teilnehmer anführten, um ihre Ablehnung der jeweiligen spezifischen Stimme zu erklären, also u.a. die Antwort auf die Frage „Warum mögen Sie diese Stimme nicht?“. Dabei stehen externe Zuschreibungen ebenso im Fokus wie konkrete klangliche, musikalische oder textliche Aspekte, die kritisiert werden oder stören und somit zu dem negativen Urteil über die Stimme führen.

Von der Begründung der Ablehnung ließen sich konkrete Reaktionen auf die Konfrontation mit der abgelehnten Stimme unterscheiden. Während die Gründe für die Ablehnung als Teil der Urteilsbildung eine Mischung aus Meinungen, Verallgemeinerungen, Vorurteilen, zurückliegenden Erfahrungen und weiterem darstellen, sind Reaktionen verhaltens- und körpernah und folgen zeitlich unmittelbar auf die Konfrontation mit der abgelehnten Musik. Sie beinhalten, wie sich der Teilnehmer verhält, wenn er mit der Musik konfrontiert ist, was sie in ihm auslöst oder welche Verhaltensimpulse er hat, wenn er sie hört. Reaktionen können zu Begründungen für die Ablehnung werden, wenn es um Vermeidung

eben dieser (unangenehmen) Reaktionen geht: ‚Diese Musik löst etwas Unangenehmes in mir aus, daher mag ich sie nicht und will sie nicht hören‘. Emotionale Gründe können sich sowohl auf den wahrgenommenen Ausdruck (engl. ‚perceived emotions‘) als auch auf das ausgelöste Gefühl (engl. ‚felt emotions‘, vgl. Gabrielsson, 2002) beziehen, Reaktionen beziehen sich nur auf das Gefühl.

In der Kodierung wurde darum zwischen Begründungen für die Ablehnung (warum der Teilnehmer die Stimme nicht mag) und Reaktionen auf die Konfrontation mit der abgelehnten Stimme (was passiert, wenn er sie hört) unterschieden. Die Teilnehmer selbst unterschieden diese beiden Aspekte nicht oder nur selten. Die von den Teilnehmern genannten Gründe lassen sich in die Oberkategorien ‚emotionale Gründe‘ und ‚sachliche bzw. klangliche Gründe‘ einteilen. Entgegen den Ergebnissen von Greasley et al. (2013) gaben die Interviewteilnehmer keine Begründungen über die Funktionen der Musik oder das Identifikationspotential der Musik ab, was jedoch vermutlich an dem Fokus der Interviews auf Stücke liegt, die dezidiert aus klanglichen Gründen abgelehnt werden. Die beschriebenen Reaktionen waren emotional oder körperlich, was der Argumentation aufgrund von affektiver Wirkung bei Greasley et al. (2013) entspricht. Ähnlich wie bei Berli (2014) verwenden die Teilnehmer also sowohl musik-immanente als auch musikexmanente Gründe, auf die im Folgenden näher eingegangen werden soll.

4.2.2.1 *Emotionale Gründe und Reaktionen*

Die beiden Beschreibungen „unangenehm/nicht angenehm“ (DV13, 17, 18_1, 11, 20) und löst „Unbehagen“ aus (DV01), und, im weiteren Sinne, dass die Stimme nicht ‚harmonisch‘ klinge, werden am häufigsten genannt, woraufhin sich die Teilnehmer „gestört“ fühlen (DV16, 01).

Oft wird als Grund für die Ablehnung ein „unnatürlicher“ (DV06, 02, 20) bzw. wenig „authentischer“ Eindruck mit zusätzlichen Beschreibungen benannt (DV03, 10, 12, 14, 02, 08, 09, 13). Auch wenn dabei kein konkretes emotionales Erleben benannt wird, wurde diese Art von Eindrücken unter emotionalen Gründen kodiert, da es sich hierbei eher um ‚gefühlte Eindrücke‘ statt um rational überdachte Begründungen handelt. Die (wahrgenommene) Ehrlichkeit und Aufrichtigkeit des Vortrags und vor allem des Sängers spielt offenbar bei vielen Teilnehmern eine wichtige Rolle bei der Bewertung eines Stücks und einer Stimme.

Oft wird in verschiedenen Formen benannt, dass der Vortrag keine Emotionen auslöse (DV03, 18_1, 18_2, 20, 10) und im weiteren Sinne, dass er langweilig sei (DV05, 11, 18_2). Außerdem, dass es keinen Spaß mache, zuzuhören (DV16), dass es etwas schlechte Laune mache (DV17, 19) und dass er keine positiven Emotionen auslöse (DV20) und nicht „einschmeichelnd“ sei (DV19).

Als emotionale Reaktionen werden benannt, dass das Zuhören „nervig“ sei bzw. „nervt“ (DV01, 02, 10, 16, 17, 20) und dass es „stressig“ sei bzw. Stress auslöst (DV01, 17), „irgendwie aufgebracht“ (DV10) mache und „anstrengend“ (DV07) sei. Einmal ist ein Teilnehmer von der Stimme „abgestoßen“ (DV17), einmal löst sie Ekel aus (DV15). Weitere Reaktionen sind Rückzug bzw. Vermeidung („innerlich ziehe

ich mich zurück“, DV19), Wut oder Aggression („daraus resultieren eher Wut und Unbehagen, Stress“, DV01; „Meine Ungeduld ist eigentlich eine Form von Aggression.“ DV03; Aggression nennen auch DV04, 05, 06, 15; wütend werden/Wut DV01, 16), Widerwillen (DV03, 06, 14), Abwehr (DV18_1, 18_2, 19) oder eine negative emotionale Reaktion auf den Sänger („Abneigung gegen den Menschen, den ich überhaupt nicht kenne“, DV19; „Abneigung“ auch DV02). Es wird einmal bemerkt, dass Musik gehört werden kann, um „eine bestimmte Stimmung herbeizuführen“, woraufhin gesagt wird, dass die Stimme „mich runterzieht [...] es verändert meine Stimmung [...] wenn der [...] jammert, dann ist mir auch zum Jammern“ (DV18_1).

Bei Berli (2014) werden diese Arten von stimmungs- oder emotionsbezogenen Begründungen des getroffenen musikalischen Werturteils unter Legitimierungen in Bezug auf habitualisierte Hörstrategien gefasst (Berli, 2014, S. 195ff.). Auch wenn er die Werturteile primär in ihrer positiven Ausprägung untersucht hat, dienen die emotionale Qualität und die Fähigkeit zur Stimmungserzeugung auch bei der Ablehnung der hier ausgewählten Stücke zur Begründungsstrategie vieler Teilnehmer. Berli thematisiert auch die „Aufrichtigkeit der Darstellung“ sowie Authentizität als musikexmanente Qualitätskriterien als eine weitere Legitimierungsebene (Berli, 2014, S. 191ff.).

Auch von Appen (2007) betont bei seiner Untersuchung von Onlinekommentaren zu Musikalben im Internet, dass das Authentizitätsargument bei der Bewertung musikalischer Werke eine wichtige Rolle spielt. Besonders im Fall von negativen Rezensionen wird diese Argumentationsstrategie zur sozialen Distinktion und Abwertung der Gegenseite verwendet (von Appen, 2007, S. 117ff.).

4.2.2.2 Sachliche (klangliche) Gründe

Bei den sachlichen oder musikimmanenten Begründungen wird am häufigsten benannt, dass bestimmte Aspekte der Musik klanglich nicht zusammenpassen, oft das Lied und die Stimme, z.B. die Stimmung des Stückes und die Stimme (DV14, 01). Teilweise wird eine Disharmonie, auch im Vergleich zu einem „melodischen“ Stück, beschrieben (DV04, 14, 01, 10). Zum Teil wird eine konkrete klangliche Eigenschaft beschrieben, hier ein Kratzen: „Also es ist irgendwie ein Störgeräusch, also wie wenn man ein schönes Bild hat, mit einer Blume und einer Landschaft, und dann kommt so Fernsehrieseln rein in dieses Bild“ (DV01).

Manche Teilnehmer nennen technische Mängel, die die Intonationssicherheit (DV08, 10) und das Klangbild (DV14, 18_2) des Sängers betreffen, fehlende Variabilität (DV19) und Kraft (DV18_2) und dass es angestrengt klingt (DV09, 17). Auch klänge es als wenn man mit den Fingernägeln über eine Tafel kratze (DV01).

Dabei beschreiben die Teilnehmer mit einer Ausnahme primär das Fehlen bestimmter musikalischer oder klanglicher Qualitäten (siehe auch von Appen, 2007, S. 98). Wie durch die Fragestrategie schon gefordert, fokussieren sich die Teilnehmer in ihren Antworten primär auf interpretatorische Mängel

sowie auf den Zusammenklang oder Gesamteindruck der Stimme mit der Musik. Ähnlich wie in den Ergebnissen von Appens fehlen kompositionsrelevante Kriterien in den Begründungen der Teilnehmer völlig.

4.2.2.3 Reaktionen auf die Stimme

4.2.2.3.1 Handlung und Verhalten

65,2 % der Teilnehmer beschreiben das Bedürfnis, sich zu entfernen oder zu distanzieren, indem sie z.B. das Radio/die Anlage ausschalten bzw. den Sender wechseln, versuchen es „auszublenden“ (DV14), verspüren einen „Fluchreflex“ (DV03). Einmal wird beschrieben, dass es ein bewusstes Bemühen erfordert, sitzenzubleiben und zuzuhören (DV08) oder nur mittels anderer Handlungen möglich ist (Augen zukneifen, DV15; anderer Beschäftigung nachgehen, DV10).

4.2.2.3.2 Körperliche Reaktionen und Gründe

Der unter ‚emotionale Gründe‘ bereits beschriebene Eindruck des Unangenehmen wird von zwei Teilnehmern mit „Unwohlsein“ (DV10, 18_1) beschrieben. „Anstrengend“ wird genannt (DV03, 07, 18_1), allerdings ist hier nicht immer eindeutig, ob es emotional oder körperlich anstrengend ist. Eindeutig körperlich sind zusätzliche Beschreibungen, dass man sich beim Hören „verspanne“ (DV16, 14, 18), „verkrampfe“ (DV03, 14, 18_1) oder „angespannt“ sei (DV05, 07, 14, 15, 18_1, 18_2). Außerdem wird genannt, dass die Stimme ungeduldig (DV03, 08, 20) und unruhig mache und die Teilnehmer sich deshalb nicht konzentrieren könnten (DV01) oder nervös würden (DV10, 18_1).

Körperliche Reaktion im Sinne einer tatsächlichen physischen Reaktion werden beobachtet in Form eines Zuckens (DV14), dass man die Arme verschränkt (DV02), die Hände zu Fäusten ballt und die Augen zusammenkneift (DV15) und dass sich die Atmung verschlechtert („dieser Druck, der bildet sich vor allen Dingen im Brustkorb aus“, DV18_1). Dreimal wird berichtet, dass es in den Ohren weh tue (DV15, 16, 18_1), einmal, dass sich die Nackenhaare aufstellen (DV01).

Die körperlichen Reaktionen sind selten einem bestimmten Organ oder Körperteil zugeordnet. Im Sinne eines funktionellen Nachvollzugs einer Stimme wird in der Sprechwissenschaft davon ausgegangen, dass (vor allem gestörte) Stimmen einen (negativen) Einfluss auf die Stimme des Hörers haben. Er wird mit dem intuitiven Nachvollziehen von Bewegungen im Sinne von Empathie und Nachahmung definiert. Dies wurde auch am Beispiel von Singstimmen in ihrer Wirkung auf die Sprechstimme des Hörers untersucht (Nespital, 2013). Interessanterweise werden in der vorliegenden Studie keine Aussagen bezogen auf stimmliche Reaktionen genannt, d.h. Reaktionen, die Organe betreffen, die direkt an der Stimmproduktion beteiligt sind, wie der Rachen oder die Kehle. Lediglich die Atmung/der Brustkorb wird einmal genannt, d.h. dass direkt die auffällige Atmung und der Atemdruck des Sängers körperlich nachvollzogen werden. In der aktuellen Studie hätte vermutet werden können, dass ein auffälliger Stimmgebrauch, der bei einigen Sängern deutlich hörbar ist, im Sinne einer Nachahmung

empfundene Wirkung beiträgt. Entweder wird der funktionelle Nachvollzug nicht explizit wahrgenommen oder er hat keinen solch starken Einfluss auf die Reaktion des Hörers wie vermutet wird.

4.2.3 Abgelehnte Eigenschaften und Stimmbeschreibungen

Tabelle 9 zeigt die benannten abgelehnten Eigenschaften, sortiert nach den Kategorien des Merkmalskatalogs (Bose, 2001; Tabelle 37 im Anhang 8.2 zeigt die Merkmale sortiert nach Teilnehmer). Die Merkmale lassen sich weitestgehend gut in die Kategorien einteilen. Merkmale, die sich nicht eindeutig zuordnen ließen und die auch in ihrer Bedeutung nicht klar wurden, sind mit Fragezeichen versehen. Diese Auflistung beinhaltet noch nicht die Information, ob die Begriffe auch so verwendet wurden, wie es Experten tun würden.

Tonhöhe	Lautheit	Stimmklang	Geräusch	Artikulation/ Lautdauer
<ul style="list-style-type: none"> • wechselnde Tonlage (DV10, 16, evtl. 20) • falsche Tonlage (DV04, 05) • unsauber (i.S. Intonation DV06, i.S. von Form, Tonverbindungen DV08) • hoch (DV06, 17) • mangelnde Tiefe (DV19) • monoton (DV02, 05) • singt Töne von unten an (DV08) 	<ul style="list-style-type: none"> • (schreiend) 	<ul style="list-style-type: none"> • nasal, verschnupft (DV02, 07, 09, 14) • knödeln (DV08) • gepresst (DV10, 17, 18_1, 20) • wenig weich (DV19) • dünn (DV17) • spitz (DV06) • quakend, quäkig (DV11, 15, 18_1, 20) • quietschend (DV13) • piepsig (DV13) • blechern (DV19; ~ „schrebbelig“, DV04) • substanzlos (?) (DV18_2) 	<ul style="list-style-type: none"> • Geräusch vor Tonanfang (DV16) • Kratzen (DV01) • rau (DV07, 20) • scharrend (DV14) • nicht klar (DV08, 11) • unsauber (i.S. Geräusch) (DV01) 	<ul style="list-style-type: none"> • Akzent (DV02) • Verständlichkeit (DV02) • gezogen (?) (DV01) • singt Vokale nicht aus (?) (DV14) • kurze Tondauer (?) (DV03)
Phonationsform	Rhythmus	Akzentuierung	Spannung	Assoziationen / Ausdruck
<ul style="list-style-type: none"> • gesprochen (DV12) • schreiend (DV16) 	<ul style="list-style-type: none"> • abgehackt (DV10) 	<ul style="list-style-type: none"> • Betonung (?) (DV07, 08) 	<ul style="list-style-type: none"> • kraftlos, schwach (DV18_1) • (auch gepresst, siehe Stimmklang) 	<ul style="list-style-type: none"> • unmännlich (DV11, 18_1, 19) • kindlich (DV01, 02, 18_1, 19) • arrogant (DV14) • unsympathisch (DV14) • desinteressiert („flapsig DV10; „hingerötzt“ DV12) • langweilig (DV02) • weinerlich/wehleidig, jammern (DV03, 18_1) • furchterregend (DV05) • übertrieben weiblich (DV13)

Tabelle 9: Abgelehnte Eigenschaften der Stimmen im Interview, sortiert in Kategorien.

Die meisten Beschreibungen beziehen sich auf den Stimmklang (N = 29). Manche Begriffe lassen sich zwar auch in andere Kategorien einteilen, beispielsweise Begriffe wie ‚gepresst‘ und ‚quakend‘, die sich auch auf die Spannung beziehen könnten, werden aber vom Hörer eher auf Basis des klanglichen Eindrucks beschrieben (und nicht auf Basis eines funktionellen Nachvollzugs, siehe Kapitel 4.2.2.3.2). Die Beschreibungen des Stimmklangs (N = 21) wurden unterteilt, so dass die Geräuschanteile (N = 8)

in einer extra Spalte aufgelistet wurden. Die genannten Begriffe zu den Geräuschen beziehen sich fast alle auf eine wahrgenommene Rauigkeit im Stimmklang. Auch Aussagen wie ‚gepresst‘ können eine Geräuschhaftigkeit beinhalten. Begriffe wie ‚quakend‘, ‚piepsig‘ und ‚quietschend‘ beziehen sich oft auf die gleiche Klangqualität, weswegen diese Beschreibung auch recht häufig zu finden ist (N = 6). Beschreibungen zur Tonhöhe/Tonlage finden sich zehn (ob ‚falsch‘, zu hoch oder zu tief oder wechselnd oder monoton) und insgesamt 13 Aussagen zum Gesamteindruck des Ausdrucks, wobei darin sieben Beschreibungen eines kindlichen Stimmklangs (in Form von nicht männlich oder übertrieben weiblich und kindlich) enthalten sind. Unklar sind die Einordnungen in der Kategorie Artikulation und Lautdauer. Während nur von einem Teilnehmer ein Akzent und eine Unverständlichkeit genannt wurden, bezogen sich andere Aussagen eher diffus auf Tonhaltedauern. Auch der Begriff ‚Betonung‘ tritt zweimal auf, ohne dass er eindeutig einer Akzentuierung zugeordnet werden kann. Um die Verwendung (und Bedeutung aus Sicht der Teilnehmer) erklären und verstehen zu können, ist eine Betrachtung der genannten Begriffe im Vergleich zu den weiteren Erklärungen innerhalb des Interviews, den angekreuzten Merkmalen im Fragebogen und dem Vergleich zu den Experten nötig (Kapitel 4.2.4).

Zunächst lässt sich schlussfolgern, dass die Teilnehmer besonders oft Merkmale aus den Kategorien Tonhöhe/Stimmlage, Stimmklang (inkl. Geräuschanteile) und Beschreibungen des Gesamteindrucks nennen. Merkmale der Artikulation (Lautdauer und Betonung) scheinen zwar von Bedeutung zu sein, es ist jedoch noch nicht klar, was genau damit beschrieben wird bzw. was die Teilnehmer mit den verwendeten Begriffen genau meinen. Textverständlichkeit an sich scheint eine untergeordnete Rolle zu spielen.

4.2.3.1 Assoziationen zum Stimmklang

Eine Reihe von Assoziationen zur Stimme/zum Gesamteindruck werden genannt (N = 16) sowie einige direkt zum Stimmklang: Die ‚quäkige‘ Stimme, z.B., erinnere „an die Chipmunks. Bisschen auch an AC/DC“ (DV11), „bei dem Quaken könnte man auch sagen, dass es irgendwo fast was Kindliches, Quengelndes hat“ (DV18_1) es klänge „als würde ich irgendwie Kindermusik [...] hören“ (DV09), erinnere „(lachend) an Donald Duck [...] und Frösche“ (DV20; auch DV15) oder in Verbindung mit dem Geschlecht, „an eine naive Frau“ (DV13). Vergleichbar damit ist der Klangeindruck des Schreiens: „diese Schreie stören mich, als wie wenn ein Kind schreien würde, oder ein Tier, oder (lacht) ähm, Bremsen quietschen“ (DV16), „ich [habe] diese Assoziation von so Vorsängerinnen in so Baptisten-Kirchen, die dann laut den Herrn preisen und dabei halt schreien, sich so in Ektase schreien“ (DV16).

In Bezug zu Geräuschanteilen im Klang finden sich folgende Assoziationen: „Ich assoziiere damit irgendwie ein gequältes Gesicht“ (DV01; in Bezug zum Kratzigen), „Diese ganz üblen I’s erinnern mich

an so `ne knarrende Tür“ (DV14), „Jemand der heiser ist oder sowas, da hat er eine ähnliche Stimme dann“ (DV07) und „Wie eine Altherren-Stimme, die er versucht noch tiefer zu machen“ (DV10),

Assoziationen mit dem Sprechgestus finden sich einmal: „irgendwie mit so Moderatoren, die ganz schnell und aufgeregt sprechen, grad wenn es so, Fußball-Kommentatoren, [...] wo sich dann die Stimme überschlägt“ (DV17).

4.2.3.2 Allgemeine Beschreibung

Die Frage zur allgemeinen Beschreibung der Stimme hatte sowohl eine mögliche Entdeckung weiterer Merkmale der Ablehnung als auch eine Erweiterung des Begriffskataloges zum Ziel, um mögliche positive oder neutrale Beschreibungen jenseits der Ablehnung aufzudecken. Interessanterweise war fast kein Teilnehmer in der Lage, über die abgelehnte Stimme aus der Distanz zu urteilen, d.h. sie mit weiteren, evtl. eher neutralen Begriffen zu beschreiben. Es kamen aber an vielen Stellen noch weitere Begriffe bzw. Beschreibungen der Ablehnung hinzu. Hier bezogen sich die Mehrzahl der Nennungen auf einen jugendlich/kindlichen Ausdruck (DV01, 10, 19), und einmal auf einen maskulinen Ausdruck einer Frauenstimme (DV15). Eine mögliche Gesangsausbildung wurde benannt, wie „klassisch“ (DV03, 14). Neutrale Beschreibungen wie die ‚Tenorstimme‘ waren sehr selten (DV03). Weitere Beschreibungen thematisierten, dass die Stimme kräftig ist (DV15), inbrünstig (DV01), hell (DV03, 19, mit positiver und negativer Konnotation), klar (DV03), weich (DV09, 10) und „versucht zu variieren“ (DV05). Die Konnotationen wurden meist deutlich. Die meisten Teilnehmer drifteten nach einem anfänglichen Versuch der neutralen Beschreibung wieder in die bereits genannten abgelehnten Merkmale ab.

4.2.3.3 Abgelehnte Eigenschaften im Fragebogen

Die Anzahl der überwiegend abgelehnten Eigenschaften im Fragebogen wird im Folgenden im Verhältnis beider Merkmalsausprägungen dargestellt (z.B. klangarm (N = 13) und klangvoll (N = 1) stehen im Verhältnis 13:1):

- Klangfülle: klangarm-klangvoll (13:1)
- Modulationen im Klang: wenig-viel (12:5),
- Artikulationspräzision: unpräzise-präzise (11:2),
- Faukale Distanz: eng-weit (10:0),
- Resonanz: schrill-dumpf (10:1),
- Spannung: gespannt-ungespannt (10:6),
- Modulationen der Töne: wenig-viel (9:6),
- Phonationsform: sprechen-rufen (9:3),
- Lautdauer: gedehnt-verkürzt (7:3),
- Stimmlage: hoch-tief (7:5),

- Klangfarbe: hell-dunkel (5:3),
- Stimmklang: hart-weich (5:3),
- Lautheit: laut-leise (5:3),
- Tonhöhenveränderungen: gleitend-sprunghaft (4:3),
- Vibrato: Zittern-Vibrato (4:2).

Die unipolaren Merkmale wurde wie folgt angekreuzt: nasal (N = 13), rau (N = 7), geknarrt und behaucht (N = 5). Es kann geschlussfolgert werden, dass, wenn in den Paaren negativ konnotierte Wörter vorkommen, diese eher angekreuzt werden, z.B. klangarm, unpräzise, eng, hart, laut. Allerdings weisen die Stimmen diese Eigenschaften laut Experteneinschätzung auch öfters auf (57-86 % korrekt, d.h. Übereinstimmung mit den Experten; siehe Kapitel 4.2.4), somit handelt es sich nicht per se um eine Tendenz zum Ankreuzen negativ konnotierter Wörter.

Das Ranking der Merkmale, also welches am stärksten ausschlaggebend für die Ablehnung ist, zeigt, dass nasal (N = 11 von insgesamt N = 13) am häufigsten auf Rang 1 gewertet wurde, gefolgt von Klangfarbe, Modulationen im Klang (N = 9), Modulationen in den Tönen, Stimmklage, Klangfarbe, Klangfülle und Resonanz (N = 8). Auf Rang 2 wird am häufigsten der Stimmklang (N = 9) angekreuzt, wobei die Einsätze (N = 5) dazugezählt werden können, dann erneut die Modulationen im Klang (N = 7), weiterhin Spannung, faukale Weite, Tonhöhenveränderungen und Artikulation (N = 6). Auf Rang 3 wählten die Teilnehmer vor allem die Phonationsform (N = 9), die Lautheit (N = 8) und die Spannung (N = 6), die auch bereits sechsmal auf Rang 2 erschien.

Zusammenfassend lässt sich sagen, dass die Nasalität als ein wichtiges Merkmal der Ablehnung angegeben wird, ebenso wie die Modulationen in Klang und Tönen und einige andere klangliche Merkmale. Es ist zu beachten, dass nicht alle Merkmale einem Ranking unterzogen wurden. Besonders die Merkmale, zu denen es Rückfragen gab, wurden oft nicht bewertet (aus Unklarheit), wie zum Beispiel ‚knarren‘ und ‚behaucht‘. Andere Merkmalen waren nur gering vertreten (N < 5) und waren gleichmäßig über die Ränge verteilt.

Diese quantitative Auflistung angekreuzter Merkmale kann jedoch alleinstehend erst einmal nicht interpretiert werden, da noch nicht deutlich wird, ob die angekreuzten Begriffe im wirklichen Sinne (d.h. im Expertensinne) verstanden wurden und ob sie wirklich den Eindruck des Hörers wiedergeben, den er vorher im Interview beschrieben hat.

4.2.4 Auswertung der Merkmale

4.2.4.1 Vergleich der Teilnehmer- und Experten-Einschätzung

Die Ratings eines Teilnehmers wurden mit den Ratings der drei Experten (siehe Anhang 8.2.1) deskriptiv verglichen. Stimmte die Einschätzung eines Teilnehmers mit zwei der Experten überein, galt das Merkmal als korrekt bewertet. Es konnten natürlich nur die Merkmale verglichen werden, die ein

Teilnehmer auch angekreuzt hatte (vgl. Kapitel 4.1.4). Die folgende Abbildung 2 zeigt die Häufigkeiten der korrekt/inkorrekt angekreuzten Merkmale im Fragebogen.

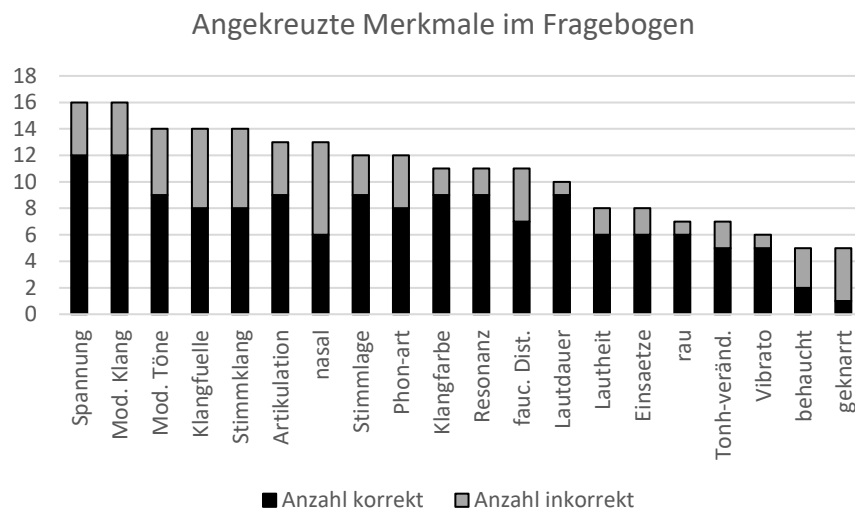


Abbildung 2: Balkendiagramm der Häufigkeiten (von N = 23) der angekreuzten Merkmale im Fragebogen. Anzahl (y-Achse) der korrekt angekreuzten Merkmale (x-Achse) in schwarz, inkorrekte in grau.

Zwei Merkmale, geknarrt und behaucht, werden sowohl selten genannt als auch kaum korrekt verwendet. Nasal wird zwar häufiger verwendet, aber nur in < 50 % der Fälle korrekt angewendet. Die anderen Merkmale werden, wenn sie verwendet werden, weitestgehend im Vergleich mit den Experten korrekt verwendet. Klangfülle und Stimmklang werden mit knapp über 50 % korrekt verwendet, Modulationen der Töne und faukale Weite mit etwas über 60 % (eine Übersicht der prozentualen Angaben findet sich im Anhang Tabelle 39). Einige Merkmale werden nur selten genannt, weshalb die korrekte Verwendung mit Vorsicht zu betrachten ist.

Eine Übereinstimmung zwischen Experten und Teilnehmern bedeutet jedoch nicht automatisch, dass die von den Teilnehmern angekreuzten Begriffe im Expertensinne verstanden wurden. Es ist möglich, dass diese Übereinstimmung zufällig zustande kommt, da die Teilnehmer andere Konzepte und Bedeutungen mit den angekreuzten Begriffen verbunden haben. Deswegen ist es wichtig, die Fragebogenantworten der Teilnehmer mit den Beschreibungen aus den Interviews zu vergleichen, um Anhaltspunkte für eine korrekte oder inkorrekte Verwendung der Merkmale zu finden.

Insgesamt wurden nicht viele Rückfragen zu den Merkmalen gestellt. Am häufigsten wurde nach der Bedeutung der faukalen Distanz gefragt (N = 8), d.h. eng und weit (der Begriff ‚faukale Distanz‘ kam nicht im Fragebogen vor), sowie nach den Einsätzen (N = 5), die zur Kontrolle zum Stimmklang abgefragt wurde, welcher keine Rückfragen bekam. Weitere Rückfragen gab es zu den Modulationen in den Tönen (N = 5) und im Klang (N = 4), zur Spannung (N = 4), zum Vibrato (N = 3) und zu behaucht (N = 3), geknarrt (N = 2), nasal, Lautheit und Tonhöhenveränderung (jeweils N = 1).

4.2.4.2 *Beziehungen zwischen Interviews und Fragebogen*

Zuerst wurde jedes Interview sehr ausführlich einzeln ausgewertet und versucht, auf der Basis der verschiedenen Aussagen und Ratings eine detaillierte Darstellung der Beschreibungsebenen der Teilnehmer zu finden. Auch sollte individuell nachvollzogen werden, welcher Begriff wie gemeint ist und wie welche Eigenschaften der Stimme beschrieben wurden. Konkret wurden zuerst die Aussagen eines Teilnehmers im Interview zu den abgelehnten Merkmalen und der allgemeinen Beschreibung der Stimme zusammengefasst und dann a) die Merkmale im Fragebogen mit denen im Interview in Verbindung gesetzt, d.h. die erhobenen Merkmale im Fragebogen wurden als eine zusätzliche Erklärungsebene für die Aussagen im Interview herangezogen und b) diese Gesamteinschätzung mit den Experteneinschätzungen verglichen. Für die Gewichtung abgelehnter Merkmale wurde das Ranking der abgelehnten Merkmale aus dem Interview und dem Fragebogen analysiert. Auf dieser Basis konnten danach die Merkmale separat untersucht werden, so dass nachvollzogen werden konnte, wie ein Merkmal von den Teilnehmern verstanden wurde und ob es Übereinstimmungen zwischen den Teilnehmern gab. Hierbei wurden auch die Merkmale des zweiten Fragebogens berücksichtigt und zur zusätzlichen Abklärung verwendet.

Im Folgenden wird jedes Merkmal detailliert diskutiert und die Änderungen für den überarbeiteten Fragebogen beschrieben. Es ist noch einmal anzumerken, dass die Teilnehmer nur die Merkmale angekreuzt haben, die für ihre Ablehnung der Stimme relevant waren, während die Experten eine Beschreibung anhand sämtlicher Merkmale vornahmen. Jeder Abschnitt enthält statistische Angaben zu den Häufigkeiten und den korrekten Verwendungen sowie Unklarheiten, die Bemerkungen / Belegstellen aus den Interviews zur Klärung der Bedeutung bzw. Interpretation des Merkmals sowie eine Schlussfolgerung für den neuen Fragebogen. Die unter ‚Statistik‘ genannten Angaben sind die in Prozent angegebenen Häufigkeiten aus Abbildung 2.

4.2.4.3 *Merkmal Modulationen in Klang und Tönen (wenig-viel)*

- Statistik: Mod. Klang 74 % Nennungen, davon 75 % korrekt, 17 % unklar, ‚wenig‘ 12:5 Nennungen
- Statistik: Mod. Töne 65 % Nennungen, davon 64 % korrekt, 22 % unklar, ‚wenig‘ 9:6

Die Merkmale ‚Modulationen in Klang‘ und ‚Modulationen in Tönen‘ werden gemeinsam betrachtet. In keinem Fall wurden die Modulationen konträr zueinander bewertet, d.h. eine Dissoziation kam nicht vor, obwohl sie möglich gewesen wäre. Die Beschreibungen im Interview wurden nicht eindeutig den Modulationen in Klang und Tönen zugeordnet (z.B. monotone Stimmlage ist eher ein Merkmal der Töne, nicht des Klangs; DV05).

‚Geringe‘ Modulationen (ohne Unterscheidung in Klang und Töne) wurde eindeutig mit monoton (DV02, 05), langweilig (DV02, 18_1), ohne verschiedene Tonhöhen, Stimmvolumen nicht genutzt (DV12), kein Klangbild (DV18_2), flach (DV04, 19), singt in einer Tonlage (DV04), keine Variabilität

(DV19) assoziiert. ‚Häufige‘ Modulationen wurden mit ständigem Hoch und Runter (Hin und Her) und wechselhaft (weich vs. schreien) assoziiert (DV16). Insgesamt waren zehn Nennungen nicht eindeutig, d.h. es gab keine Entsprechungen im Interview oder die Einschätzung war nicht korrekt. In Einzelnennungen gab es auch konträre Assoziationen, d.h. der Klang wurde im Interview benannt, die Töne aber im Fragebogen angekreuzt oder es wurde ‚wenig‘ angekreuzt, aber im Interview ein häufiger Wechsel von Klängen beschrieben. Zusammengenommen wurde das Merkmal ‚wenig‘ doppelt so oft gewählt wie ‚viel‘ (21:11). Die korrekte Verwendung von klanglichen Modulationen war etwas höher als die von Tonhöhen-Modulationen (75 % und 64 %). In manchen Fällen könnte die Modulation als eine Beschreibung des Ausdrucks verwendet worden sein, z.B. zur Beschreibung von weinerlich oder wehleidig (DV03, 18_1). Die Beschreibung der Singweise des Interpreten als langweilig (DV02, 05) oder lustlos (DV07_2, 12) oder die Feststellung, dass er emotional nichts transportiere, sind auch eher Beschreibungen eines Ausdrucks.

Es kann geschlussfolgert werden, dass beide Merkmale nicht unterschieden zu werden scheinen, da die Ratings meist übereinstimmen. Sie werden wahrscheinlich verwendet, um den Eindruck einer interessanten oder uninteressanten stimmlichen Leistung zu beurteilen. Auch wird von den Teilnehmern häufig ‚wenig‘ als Modulation angegeben, was dem uninteressanten Eindruck entsprechen könnte und der Ablehnung eines negativ konnotierten Ausdrucks entspricht.

Das Merkmal wird für den neuen Fragebogen geändert, so dass nach dem Ausdruck der Stimme gefragt wird, der mit den Polen ‚wechselhaft-gleichförmig‘ beschrieben werden kann. Damit wird wieder etwas an das Original von Bose angeknüpft, wo der „Sprechausdruck insgesamt“ mit „geschlossen-wechselhaft“ und die „Stimmklangmodifikationen“ mit „wechselnd-gleichförmig“ beschrieben werden (Bose, 2001, S. 301f.).

4.2.4.4 Merkmal Spannung (gespannt-ungespannt)

- Statistik: 70 % Nennungen, 75 % korrekt, 17 % unklar, ‚gespannt‘ 10:6.

Gespannt wurde mit angestrengt (DV07_1, 09), quäkig (DV11, 18_1, 20), schreiend (DV16) und gepresst (DV16, 17, 18_1, 20) zusammen genannt; ungespannt zusammen mit schwach, wenig kraftvoll (DV06, 18, nicht ganz eindeutig), lustlos (DV12) und wehleidig (DV18_1). Das Merkmal gepresst, welches nicht im Fragebogen vorkam, wurde vier Mal im Interview genannt, von den Experten zweimal, immer im Zusammenhang mit ‚gespannt‘. ‚Gespannt‘ wird mit ‚gepresst‘ gleichgesetzt, aber auch ‚ungespannt‘ wurde gewählt, wenn auch seltener (10:6). Das Merkmal Spannung kann mit den Polen ‚gespannt-ungespannt‘ erhalten bleiben.

Als neues und ergänzendes Merkmal wird der Begriff ‚gepresst‘ aufgenommen. Da der Versuch unternommen werden sollte, bipolare Items zu bilden, wurden die Pole ‚locker-gepresst‘ gebildet (vgl. Bojack-Weber, 2012, S. 56).

4.2.4.5 Merkmal faukale Distanz (eng-weit)

- Statistik: 43 % Nennungen, 64 % korrekt, 35 % unklar, ‚eng‘ 10:0

Faukale Distanz ist ein interessantes Merkmal, da die Begriffe ‚eng-weit‘ kaum Raum für stimmlich-klangliche Assoziationen geben, die für Nicht-Experten wichtig sind. Der Pol ‚eng‘ scheint greifbarer, der Pol ‚weit‘ wird nicht verwendet. Die Rückfragen und die gegebenen weiteren Begriffe führen in keinem Fall zu einem besseren Verständnis. Das Merkmal wurde in 43 % der Fälle verwendet und von diesen in immerhin 64 % der Fälle korrekt. Auch beschreibt es eine wichtige Erscheinung bei Sängern, den Twang (Kapitel 2.3.1), der auch in Kombination mit einer Nasalität in Form des ‚nasal twang‘ auftreten kann (auch das Knödeln sei an dieser Stelle erwähnt). Damit ist es ein wichtiges klangliches Merkmal, das eine Beschreibungsebene im Fragebogen braucht.

Eng (DV02, 06, 07_1, 08, 11, 14, 16, 17, 18_1, 20) wird in durchaus passenden Kombinationen mit quäkig/quakend verwendet (DV11, 18_1, 20), mit gepresst (DV17, 18_1, 20), mit hoch (spitz) (DV06, 17), mit weinerlich/jammern (DV03), es wird aber auch uneindeutig verwendet, evtl. einmal um Schreien darzustellen (DV16) oder repräsentiert evtl. eine auffällige Klangbildung (z.B. bei Brightman den Lippenbreitzug bei /i/; DV14). Eng und quäkig (auch möglich in der Kombination mit gepresst und hoch) sind sich sehr nahe (zum Beispiel im Twang). Das Merkmal ‚weit‘ kann aufgrund der Datenlage an dieser Stelle nicht weiter untersucht werden. Es stellt sich die Frage, ob ‚eng‘ durch ‚quäkig‘ ersetzt werden kann. Dazu muss aber auch das Merkmal ‚nasal‘ vergleichend untersucht werden.

4.2.4.6 Merkmal Nasal

- Statistik: 57 % Nennungen, 46 % korrekt, 4 % unklar

In neun von 13 Fällen wurde im Fragebogen von den Teilnehmern ‚eng‘ und ‚nasal‘ zusammen angekreuzt. Viermal wurden Begriffe verwendet, die sich auf die Nase an sich beziehen (nasal, verschnupft, zugehaltene Nase, genäsel; DV01, 07_1, 09, 14), allerdings stimmen die Experten dieser Einschätzung oft nicht zu. Der Klangeindruck von ‚quäkig, quaken‘ (DV11, 15, 18_1, 20) wurde viermal zusammen mit ‚nasal‘ benannt, einmal wird auch „knödelt“ (DV08) genannt. Die Vermischung von einer faukalen Enge und einer wahrgenommenen Nasalität sowie Klangbeschreibungen wie ‚quäkig‘ lassen vermuten, dass die beiden Begriffe für die Teilnehmer zusammenhängen oder nicht differenziert werden können. Weitere interessante Belegstellen finden sich folgende: DV18 sagt in der Nachbefragung, das Nasale mache das Quaken aus. DV11 versucht ‚quäkig-kratzend‘ gegen ‚rauchig-kratzig‘ abzugrenzen. DV15 beschreibt das Quaken mit „es ist, als wenn der Ton von hier, hier oben im Hals kommt“.

In den Fällen, in denen ‚nasal‘ vom Teilnehmer angekreuzt wurde, liegt meistens eine von Experten eingeschätzte faukale Enge vor. Der Twang kommt bei fünf Sängern vor, deren Klang von den Teilnehmern als ‚nasal‘ beschrieben wurde. Es kann sein, dass die Nasalität und die faukale Enge nicht

auditiv unterschieden werden. Daher ist es sinnvoll, die Begriffe ‚eng‘ und ‚nasal‘ bei der Überarbeitung des Fragebogens zusammenzufassen und den Begriff ‚quäkig‘ zu übernehmen, da er in den Interviews oft genannt wird. Es ist noch zu erwähnen, dass Nasalität auch mit faukaler Weite zusammengehen kann, wie es z.B. bei Sarah Brightman (und vielen klassischen Sängern) der Fall ist.

Als neues Merkmal wird eine Kombination aus faukaler Distanz und Nasalität geschaffen: ‚Die Stimme klingt eher dumpf/weit-quäkig/nasal.‘ Das Merkmal ‚dumpf‘ wird im Zuge der Resonanz noch diskutiert.

4.2.4.7 Merkmale Resonanz (schrill-dumpf)

- Statistik: 47 % Nennungen, 81 % korrekt, 0 % unklar, ‚schrill‘ 10:1

‚Schrill‘ beschreibt eine Frequenzverstärkung, die positiv betrachtet zu einer Brillanz in der Stimme führt und Tragfähigkeit fördert oder einen metallischen Klang erzeugt, d.h. in der negativen Konnotation kann ‚schrill‘ auch mit „blechern“ (DV19) o.ä. gleichgesetzt werden. ‚Schrill‘ wird sehr häufig verwendet (10mal), das Gegenteil ‚dumpf‘ hingegen nur einmal (und, laut Expertenmeinung, inkorrekt). Im Fragebogen wird ‚schrill‘ zusammen mit ‚hart‘ (3 x) und ‚klangarm‘ (2 x) verwendet, in den Interviews in Verbindung mit resonatorischen Eigenschaften wie „klingt wie aus dem Eimer“ und „schreibelig“ (DV04), quakt (DV11, 15) oder kraftlos/dünn (DV18_2, 17). Letztere sind nicht eindeutig und können auch dem Merkmal Klangfülle zugeordnet werden. Als ‚dumpf‘ wurden nur vier Sänger von den Experten eingeschätzt: Mariah Carey (wechselt), Udo Lindenberg, Alan Wilson (Canned Heat) und Helge Schneider. Die Merkmalsausprägung ‚dumpf‘ war somit in der Auswahl der Lieder nur selten vertreten. Die Ausprägung ‚schrill‘ konnte nicht eindeutig einer klanglichen Eigenschaft zugeordnet werden und scheint mehr wegen ihrer negativen Konnotation zur Beschreibung gewählt worden zu sein. Die Merkmale werden, wie oben beschrieben, neu kombiniert zu ‚dumpf/weit-quäkig/nasal‘.

4.2.4.8 Merkmale Klangfülle (klangarm-klangvoll)

- Statistik: 61 % Nennungen, 57 % korrekt, 0 % unklar, ‚klangarm‘ 13:1

Klangarm scheint ein wichtiges Merkmal zu sein, es wird oft auf Rang 1 der Ablehnung gesetzt. Klangvoll kann nicht bewertet werden, da es nur eine Eintragung gab. Es stellt sich auch die Frage, ob das Merkmal der Resonanz nicht auch durch die Klangfülle abgedeckt werden kann, womit ein wichtiges Merkmal erhalten bleiben könnte. Die Alternative ‚resonanzreich-resonanzarm‘ wurde verworfen, da das Fremdwort ‚Resonanz‘ vermieden werden sollte. Es bleibt im neuen Fragebogen erhalten, um es weiter zu erproben.

4.2.4.9 Merkmal Artikulation (unpräzise-präzise)

- Statistik: 57 % Nennungen, 69 % korrekt, 0 % unklar, ‚unpräzise‘ 11:2

‚Unpräzise‘ wird zur Beschreibung von vermutetem Desinteresse des Sängers (DV07_2, 10_2, 12) bzw. einer eingeschränkten Verständlichkeit aufgrund von unterspannter Artikulation (DV02, 06, 12, 18_1) verwendet. Zusätzlich wird ‚unpräzise‘ zur Beschreibung eines individuell ausgeprägten Artikulationsmusters verwendet, das aber eigentlich präzise ist (DV02, 14), z.B. bei Lena Meyer-Landrut, die mit aufgesetztem britischen Akzent singt, oder bei Sarah Brightman, deren Klangformung der Vokale /u/ und /i/ nicht der Erwartung des Teilnehmers entsprechen – d.h. der Eindruck einer unpräzisen Artikulation ist nicht zu verallgemeinern (die Experten stimmen nicht mit der Einschätzung überein) und es wird eine Wertung auf Basis individuell subjektiver Hörgewohnheiten getroffen. Der Gegenpol ‚präzise‘ (DV08, 17) tritt lediglich zweimal mit unklarer Verwendung auf. Die Experten konnten beide Male keine Entsprechung in der Stimme finden. Das Merkmal Artikulation könnte mit spezifischeren Polen möglicherweise besser beschrieben werden, indem Auffälligkeiten anstelle der Pole ‚präzise-unpräzise‘ als Charakteristikum verwendet werden.

Für den neuen Fragebogen soll das Merkmal artikulatorische Besonderheiten untersuchen: ‚Die Artikulation ist eher markant-schlicht.‘

4.2.4.10 Weitere Merkmale

Das Merkmal Stimmklang, ‚hart-weich‘ (61 % Nennungen, 57 % korrekt, 0 % unklar), wird beibehalten. Das Merkmal zu den Einsätzen wird herausgenommen, da es redundant ist und von den Teilnehmern auch ohne diese Spezifikation beurteilt werden konnte.

Das Merkmal Phonationsform (Sprechen-Rufen) kann beibehalten werden.

Stimmfarbe (tief-hoch) und Klangfarbe (dunkel-hell) werden beibehalten, da sie Merkmale sind, die bereits im Interview von den Teilnehmern angesprochen wurden, also ins aktive Beschreibungsvokabular der Hörer gehören. Beide werden oft auf Platz 1 der Rangliste abgelehnter Merkmale gesetzt. Außerdem stellt vor allem die Stimmfarbe ein intuitiv zu bewertendes Merkmal dar. Klangfarbe kann durchaus noch mit Stimmfarbe verwechselt werden bzw. ist nicht gänzlich von der Stimmfarbe zu unterscheiden, zeigt aber noch eine weitere Facette auf.

Die Ausprägungen des Merkmals Tonhöhenveränderungen, ‚sprunghaft-gleitend‘ (30 % Nennungen, 71 % korrekt, 4 % unklar), werden in den Interviews und im Fragebogen benutzt: DV08 beschreibt „singt die Töne von unten an“ und nennt die gleitenden Tonhöhenveränderungen; DV12 beschreibt das Singen von Bob Dylan als „hingerotzt, lustlos“ und „gleitend“; bei Mariah Carey fällt DV16 ein „ständiges Hoch und Runter“ auf; auch die Beschreibungen „wellenförmig“ und „gleitend“ kommen vor (DV01 inkorrekt, DV08, 12, 16), ebenso ‚sprunghaft‘ (DV03, 10_1, 17). Das Merkmal bleibt erhalten.

Beschreibungen der Lautdauer, ‚gedehnt-verkürzt‘ (43 % Nennungen, 90 % korrekt, 0 % unklar), werden in den Interviews eher diffus verwendet, wie „gezogen“ (DV01), „singt Vokale nicht aus“ (DV14) und „kurze Tonhaldedauer“ (DV03) und auch die Verbindungen im Fragebogen werden nicht deutlich: Dort geht ‚gedehnt‘ (7:3) mit ‚unpräzise‘ einher (DV02, 06, 07_1, 16), aber auch einmal mit ‚präzise‘ (DV08); zweimal geht ‚verkürzt‘ mit ‚unpräzise‘ einher (DV03, 10_1) und einmal mit ‚präzise‘ (DV17). Da beim Singen die Laute (also vor allem die Vokale) typischerweise gedehnt werden, wird nicht klar, warum die Ablehnung hinsichtlich der Lautdauer auftritt, da es in den Interviews kaum Hinweise darauf gibt. Möglicherweise wurde vom Teilnehmer Lautdauer mit Artikulation verwechselt, allerdings nicht mit konsistenten Verbindungen (präzise/verkürzt etc.). Da das Merkmal offenbar nicht verstanden wird oder keine Rolle spielt, wird es in der nächsten Befragung nicht mehr verwendet.

Vibrato (N = 2) und Zittern (N = 4) scheinen in der Beschreibung der gewählten Stimmen kaum oder keine Rolle zu spielen, sie werden im Interview nicht genannt und stehen nicht in Verbindung zu den Nennungen im Fragebogen. Einige wenige Nachfragen lassen vermuten, dass die Teilnehmer den Begriff des Vibrato nicht kennen, es scheint also abhängig von der Expertise und auch dem Interesse zu sein, denn vermutlich ist es im Kontext der klassischen Musik deutlich relevanter. Nicht-Sänger, zum Beispiel, sind sich in Ratings zum Vibrato sehr uneinig und ihre Einschätzung einer Performance ist nicht vom Vibrato abhängig – ganz im Gegensatz zu Sängern (Reddy & Subramanian, 2015). Das Merkmal wird nicht weiter berücksichtigt.

Lautheit (laut-leise) wird im Interview nie und im Fragebogen nur selten genannt (35 % Nennungen, 75 % korrekt, 4 % unklar), wird aber, wenn es bemerkt wird, an dritte Stelle der Rankings gesetzt, was ihm eine gewisse Bedeutung zuspricht. Die Lautheit von Aufnahmen ist nicht unproblematisch zu bestimmen, allerdings geht es hier um eine wahrgenommene Intensität, die evtl. auch anders beschrieben werden kann, z.B. als zurückhaltend oder aufdringlich. Laut (DV01, 04, 07_1, 13, 15) wird in Verbindung mit rufend genannt (DV04, 13) sowie einmal mit quakend (i.S. des Twang; DV15), zweimal kann keine eindeutige Zuordnung erfolgen. Leise (DV03, 07_1, 18_1) kommt in Verbindung mit weinerlich vor (DV03) und jammernd/wehleidig (DV18_1), einmal wird es unklar verwendet und entgegen der Experten-Einschätzung (DV07_1). Somit stellt sich die Frage, ob mit laut und leise eher Persönlichkeitsmerkmale beschrieben werden, wie zurückgenommen und aufdringlich. Das Merkmal bleibt im neuen Fragebogen erhalten und wird weiter evaluiert.

4.2.4.11 Geräuschanteile

- Rau, 30 % Nennungen, 86 % korrekt, 0 % unklar

„Rau“ wurde im Fragebogen fast immer korrekt verwendet und es gab keine Rückfragen. „Behaucht“ wird nur selten und dann nicht korrekt verwendet, „geknarrt“ zeigt ebenso ein schlechtes Ergebnis. „Rau“ wird zweimal korrekt im Interview genannt (DV07_2, 20), „behaucht“ wird nicht genannt (auch nicht Hauch, Luft etc.). Die Rauigkeit wird mit einem Geräusch im Klang beschrieben, synonym verwendet zu Kratzen (DV01), kein klarer Klang (DV07_1), trockener Hals (DV07_2), gepresst (DV10, 17, 20), aber auch als „Alt-Herren-Stimme“ (DV10) bezeichnet. Der Eindruck von Rauigkeit wird in nur wenigen Fällen mit einem gepressten Klang verwechselt, ebenso einer tiefen, dunklen Klangfarbe.

Es wurde überlegt, ob die beiden Begriffe „rau“ und „behaucht“ wie in der Literatur zu dem Komplex Heiserkeit zusammengefasst werden sollten (Nawka et al., 1994). Diese Idee wurde verworfen, da bewusst kein Krankheitsbezug bei der Stimmbeschreibung hergestellt werden sollte. Das Merkmal „rau“ bleibt erhalten und wird um einen Gegenpol ergänzt: „rau/kratzig-glatt“. Auf den Gegenpol „klar“ sollte verzichtet werden, da dies im Prinzip gleichzusetzen ist mit „Merkmal nicht vorhanden“ und das Item somit wieder versteckt unipolar wäre.

Grundsätzlich konnte bei der Analyse der Merkmale festgestellt werden, dass sich die Bedeutung von genannten Begriffen im Interview rückwirkend durch die Merkmale im Fragebogen erklärten. Es gab einige wenige Ausnahmen, die sich vorwiegend auf ausgefallene Wörter oder Wortkreationen ohne eindeutige Definition wie z.B. „schrebbelig“ bezogen. Das heißt, der vorgelegte Fragebogen eignet sich zur Wiedergabe des Eindrucks der Stimme, was die Überschneidungen zwischen den Begriffen und Beschreibungen zeigen. Es hat sich auch gezeigt, dass er die Ausdrucksmöglichkeiten der Teilnehmer erweitert, die oft nur zwei Begriffe zur Stimmbeschreibung frei nennen konnten. Am Vergleich mit den Expertenratings ließ sich feststellen, dass die Merkmale oft korrekt benannt wurden. Allerdings kann dies aussagekräftiger anhand einer quantitativen Auswertung von Teilnehmern und Experten-Ratings erfolgen, wie es für die weiteren Untersuchungen geplant ist.

4.2.5 Schlussfolgerungen für die Fragebogenentwicklung

Nachdem die Merkmale des ersten Fragebogens mittels der Interviews und Expertenbefragung hinsichtlich ihrer Bedeutung, Definition, Verständlichkeit und Wichtigkeit für die Nicht-Experten-Teilnehmer untersucht wurden, wurde ein überarbeiteter Folge-Fragebogen erstellt, der in einer Gruppentestung weiter evaluiert wurde. Im Folgenden werden die Überlegungen dieser Konstruktion dargestellt und diskutiert.

4.2.5.1 Feedback zum Fragebogen

Die meisten Teilnehmer empfanden den vorliegenden Fragebogen als gut lösbar: „sehr übersichtlich“, „gut strukturiert“ (DV01, 11, 19), „ist in Ordnung“ (DV04, 09, 14, 15, 17), „einfach bzw. nach

Nachfragen klar“ (DV10, 12, 16, 17, 20); DV17 sagt, dass Begriffe dabei waren, die ihm selbst nicht eingefallen wären. Probleme wurden nur hinsichtlich der Vielschichtigkeit des Fragebogens benannt, d.h. dass es nicht nur die Evaluation der Merkmale gab, sondern zusätzlich noch die Einschätzung, was wichtig ist, welche Merkmale unklar sind und wie das Ranking aussieht (DV03, 07).

Da kein Teilnehmer Probleme mit dem Fragebogen hatte, kann die Struktur, d.h. die Bipolarität und die Abfolge der Merkmale, beibehalten werden. Die kritisierte Vielschichtigkeit wird in den Folgestudien keine Rolle mehr spielen.

4.2.5.2 *Bipolarität*

Einige Merkmale lassen sich gut bipolar darstellen (z.B. hoch-tief), andere jedoch weniger (rau, behaucht). Selbst wenn auf die Ausprägung ‚wenig-viel‘ ausgewichen wird, bedeutet dies aber immer noch keine echte Bipolarität. Bänziger et al. (2014) haben dies in dieser Form umgesetzt, also eine vermeintliche Bipolarität hergestellt – und sie damit auch nur vermeintlich für die statistischen Analysen einsatzfähig gemacht. Es ist methodisch problematisch, zwischen bipolaren und unipolaren Items zu wechseln, da zum einen Teilnehmer dazu neigen, unipolare Items zu überspringen (vgl. Merrill, 2017) und zum anderen die statistische Analyse erschwert wird, da diese Merkmale in der statistischen Auswertung anders behandelt werden müssen und somit nur auf Umwegen mit den bipolaren Items verglichen werden können (vgl. Merrill & Larrouy-Maestri, 2017). Für den neuen Fragebogen sollte deshalb getestet werden, ob eine Bipolarität für alle Merkmale hergestellt werden kann. Es sollte versucht werden, alle Items des Fragebogens bipolar anzulegen, sodass sich der Studienteilnehmer zwischen den Merkmalsausprägungen entscheiden muss. Dies kann auch von Vorteil sein, da, wenn sich Teilnehmer (besonders Nicht-Experten) der Bedeutung eines Pols unsicher sind, er möglicherweise über den anderen Pol erklärt werden kann.

Als überarbeitete Merkmalspolaritäten wurden daher ‚markant-schlicht‘ für die Artikulation, ‚rau/kratzig-glatt‘ für die Geräuschanteile und ‚gepresst-locker‘ für die Spannung ausgewählt. Dabei wird in Kauf genommen, dass ‚schlicht‘ sowohl positiv im Sinne von unpräzise als auch negativ im Sinne von nichtssagend interpretiert werden kann und somit beide Möglichkeiten der Wertung enthält. Auch ‚markant‘ als Gegenpol kann sowohl positive wie negative Auffälligkeiten beschreiben.

‚Glatt‘ als Pol zu ‚rau/kratzig‘ kann mit einer glatten Persönlichkeit verbunden werden, kann aber auch lediglich als unauffällig verstanden werden, also als etwas, das keine besonderen Merkmale im Stimmklang zeigt. Das problematischste neu kombinierte Merkmal ist der Gegensatz ‚dumpf/weit-quäkig/nasal‘. Während die Kombination von ‚quäkig‘ und ‚nasal‘ nachvollziehbar ist und davon ausgeht, dass die Hörer diese beiden nicht explizit unterscheiden können, ist die Kombination ‚dumpf/weit‘ sprechwissenschaftlich angreifbar. Der Pol ist durch eine Kombination mehrerer Merkmale kreiert worden: ‚quäkig‘ verbindet die Eigenschaften ‚schrill‘ und ‚eng‘, zu denen die

eigentlichen Gegensätze ‚dumpf‘ und ‚weit‘ sind. Das heißt, hier wird eine indirekte Konstellation aufgebaut, die sich möglicherweise für die Teilnehmer nicht auf den ersten Blick erschließt. Dennoch erschien es als die sinnvollste Lösung für die Verständnisschwierigkeiten, die sich bei diesen Merkmalen in der Interview-Studie gezeigt hatten, und soll im Folgenden erprobt werden.

4.2.6 Schlussbemerkungen

In den Interviews zeigt sich, dass besonders Merkmale aus den Kategorien Tonhöhe/Stimmhöhe, Stimmklang und Beschreibungen des Gesamteindrucks genannt werden. Dies bestätigte sich in den Fragebögen, wo Modulationen in Klang und Tönen sowie Klangfülle/Resonanz sowohl oft genannt als auch als am wichtigsten für die Ablehnung bewertet werden. Die Stimmhöhe an sich wird im Fragebogen allerdings seltener genannt und auf Rang 2 gesetzt. Merkmale der Artikulation (Lautdauer und Betonung) scheinen im Interview von Bedeutung zu sein, es ist jedoch noch unklar, inwiefern sie das sind, bzw. was eigentlich beschrieben wird. Textverständlichkeit an sich scheint eine untergeordnete Rolle zu spielen im Vergleich zu Eigenheiten in der Artikulation.

Das Merkmal ‚nasal‘ führte in dieser Auswertung zu einem interessanten Ergebnis. Es wird am häufigsten im Fragebogen genannt und am häufigsten auf Platz 1 für die Ablehnung gesetzt. Im Interview wird dieser Begriff zwar nicht verwendet, es gibt aber Verweise auf ‚verschnupft‘, die allerdings oft etwas Anderes beschreiben sollen. Äquivalent dazu wird auch das Merkmal ‚nasal‘ im Fragebogen für einen anderen Klangeindruck verwendet als intendiert. Die Teilnehmer benutzen einen Begriff, den sie im Interview nicht verwenden und im Fragebogen anscheinend mit einem falschen Konzept verbinden – bewerten ihn aber gleichzeitig als den wichtigsten klanglichen Grund für ihre Ablehnung. Das Verständnis der Teilnehmer vom Begriff ‚nasal‘ kann leider mit der vorliegenden Studie nicht gänzlich geklärt werden, allerdings wird deutlich, dass wohl vor allem eine Enge im Rachen/Kehlkopf gemeint ist, durch die ein ‚quäkiger‘ Stimmklang entsteht, der, im Gegensatz zum Begriff ‚nasal‘, mehrfach in den Interviews genannt wird.

Mit 23 Sängern wurde zwar eine recht umfangreiche Anzahl an Stimmen beurteilt, allerdings nur von 20 Personen. Deshalb sollen die herausgearbeiteten Merkmale anhand einer größeren Hörergruppe überprüft werden. Da die Evaluation der Stimmen in dieser Studie auf der Ablehnung basierte, soll im nächsten Schritt auch die ästhetische Bewertung auf das Gefallen ausgeweitet werden.

5 Präferenzen vokaler Merkmale und Fragebogenentwicklung – Gruppen-Studie

Die Ergebnisse der vorherigen Studie zeigten, welche vokalen Merkmale ungeschulte Hörer zur Beschreibung der individuellen Ablehnung von Sängern heranziehen und nutzen können. Die individuellen Beschreibungen wurden durch einen umfangreichen Fragebogen ergänzt, womit den Teilnehmern Experten-Begriffe zur Beschreibung vokaler Merkmale zur Verfügung standen. Die gesammelten Einschätzungen der Teilnehmer wurden weiterhin mit denen von Stimmexperten verglichen, um die korrekte Verwendung der Begriffe einschätzen zu können. Letztendlich konnte eine erste Version des Inventars zur Beschreibung vokaler Merkmale von Singstimmen entsprechend ausgearbeitet werden. Dieses soll in der folgenden Studie einer größeren Teilnehmer-Gruppe zur Beschreibung von Sängern vorgelegt werden, damit überprüft werden kann, ob die Begriffe verständlich sind und welche für eine charakteristische Beschreibung von Singstimmen wichtig sind. Die Möglichkeit zur ästhetischen Bewertung wurde um die Gefallens-Dimension erweitert, die Ablehnung stand hier nicht mehr im Zentrum. Der bisher stark qualitative Methodenansatz verlagert sich in der folgenden Studie auf einen quantitativen, wobei neben den Ratings weiterhin die Kommentare der Teilnehmer in die Auswertung einfließen.

5.1 Methoden

5.1.1 Auswahl der Musiktitel

Es wurden 16 von den 20 Titeln der Interview-Studie ausgewählt (siehe Tabelle 10). Dies lag nahe, da zum einen die individuelle Ablehnung der Teilnehmer der ersten Studie mit dem Gruppenurteil verglichen werden sollte, und zum anderen, weil die mitgebrachten Stimmen auffällige Merkmale aufzeigten und zusätzlich die Auswahl eine gute Bandbreite stimmlicher Merkmale zur Beurteilung darstellten. Dabei wurde darauf geachtet, dass ein ausgeglichenes Verhältnis männlicher und weiblicher SängerInnen vorlag. Aussortiert wurden vier Sänger, die aufgrund der Musik schlechter hörbar waren als andere, auffälliger bearbeitet waren oder nur ausschnittsweise verwendet werden konnten (Angelo Kelly sang im Duett).

ID	SängerIn	Titel	Album (Jahr)	Label	Ausschnitt
1	Anastacia	I'm Outta Love	Not That Kind (2000)	Epic Records/Daylight Records	00:34-01:09
2	Annett Louisan	Das Spiel	Bohème (2004)	105music	00:08-01:28
3	Bob Dylan	Don't Think Twice, It's All Right	The Freewheelin' Bob Dylan (1963)	Columbia Records	00:04-00:43
4	Casper	Hinterland	Hinterland (2013)	Four Music	00:31-01:35
5	Gitte Haenning	Ich hab' die Liebe verspielt in Monte Carlo	Ich bin kein Kind von Traurigkeit (1975)	Universal Music	00:00-01:16
6	Helge Schneider	Katzeklo	Es gibt Reis, Baby (1993)	Roof Music/Electrola	00:28-01:00

7	Herbert Grönemeyer	Männer	4630 Bochum (1984)	EMI	00:04-01:17
8	Lena	Satellite	My Cassette Player (2010)	We Love Music	00:08-00:50
9	Mariah Carey	My All	Butterfly (1997)	Columbia Records	00:19-01:23
10	Max Herre	Blick nach vorn	Ein geschenkter Tag (2009)	Nesola	00:26-01:33
11	Michelle	Paris	Die Ultimative Best Of (1024)	Electrola	00:13-01:20
12	Passenger	Let Her Go	All The Little Lights (2012)	Netzwerk/Black Crow Records	00:24-01:16
13	Anneliese Rothenberger	In mir klingt ein Lied	In mir klingt ein Lied (2008)	Documents	00:16-02:15
14	Udo Lindenberg	Durch die schweren Zeiten	Stärker als die Zeit (2016)	Warner Music Germany	00:11-01:37
15	Xavier Naidoo	Abschied nehmen	Zwischenspiel – Alles für den Herrn (2002)	Naidoo Records	00:00-01:22
16	The Tallest Man on Earth	Love Is All	The Wild Hunt (2010)	Dead Oceans	00:23-01:27

Tabelle 10: Übersicht der ausgewählten Stimmen und Titel.

5.1.2 Fragebogen

Der Fragebogen (Tabelle 11) wurde aus der Interview-Studie entwickelt (Kapitel 4.2.5). Zusätzlich zur Beschreibungsebene der Merkmale wurde das Gefallen eines jeden Merkmals abgefragt, wobei die Beschreibung auf einer 4-stufigen Skala erfolgte, damit die Teilnehmer sich zwischen den Merkmalsausprägungen entscheiden mussten, und das Gefallen auf einer 5-stufigen Skala, damit ein Mittelpunkt zum Angeben einer neutralen Einstellung möglich war. Im Anschluss an den Merkmalsfragebogen gab es weitere Fragen zum allgemeinen Gefallen von Stimme, Lied, Text und Musikstil (jeweils auf einer 5-stufigen Skala), der Bekanntheit von Lied und Sänger/Sängerin (bekannt: ja/nein), wobei auch der Name des Sängers/der Sängerin abgefragt wurde (ja, er/sie heißt...).

Weitere Fragen bezogen sich methodisch auf den Fragebogen an sich: „Wie gut konnten Sie den Eindruck von der Stimme mit den gegebenen Merkmalen wiedergeben?“ (5-stufige Skala, sehr gut-gar nicht); „Fehlte Ihnen etwas?“ (nein, ja, und zwar...). Diese grundsätzliche Einschätzung zur Funktionalität des Fragebogens war wichtig, um den Fragebogen weiter zu entwickeln.

In einer Abschlussbefragung wurden die Teilnehmer gebeten, jedes abgefragte Merkmal hinsichtlich der Schwierigkeit zu bewerten und dieses Urteil zu begründen: „Wie empfanden Sie die Beschreibung der Stimmen anhand der einzelnen Merkmale? Bitte geben Sie die Schwierigkeit für jedes Merkmal an.“ (4-stufige Skala von 1-4) und „Warum fiel Ihnen die Beschreibung schwer?“ (Anhang 8.3). Besonders diese Abschlussbefragung war für die Weiterentwicklung des Fragebogens von Bedeutung, denn so konnten subjektive Eindrücke mit den eigentlichen Ergebnissen verglichen werden.

Beschreiben						Gefallen				
						Die angekreuzte Eigenschaft der Stimme...				
						gefällt mir		gefällt mir nicht		
Die Stimme ist eher...										
tief	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	hoch	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
laut	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	leise	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
klangvoll	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	klangarm	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
dunkel	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	hell	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
weich	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	hart	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
gespannt	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	ungespannt	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Die Stimme klingt eher...										
rau/kratzig	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	glatt	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
dumpf/weit	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	quäkig/nasal	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
locker	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	gepresst	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Die Veränderungen von Ton zu Ton sind eher...										
sprunghaft	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	gleitend	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Die Artikulation ist eher...										
markant	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	schlicht	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Das Singen erinnert an...										
Sprechen	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Rufen	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Der Ausdruck der Stimme ist eher...										
wechselhaft	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	gleichförmig	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Tabelle 11: Fragebogen der Gruppen-Studie.

5.1.3 Teilnehmer und Durchführung

Insgesamt nahmen 48 Teilnehmer, davon 29 weiblich, Alter Mittelwert (MW) = 40,06 Jahre (SD = 18,24), an der Befragung teil. Davon gaben 25 einen Universitätsabschluss, 20 das Abitur, einer einen Realschulabschluss und einer eine Promotion als höchsten Bildungsabschluss an. Beruflich waren 27 der Befragten Studenten, drei Angestellte, zwei Beamte, elf Rentner/Pensionäre, fünf Selbstständige. Fünf Teilnehmer haben beruflich mit Musik zu tun, 22 spielen ein Instrument, 17 singen ‚nur für sich‘ und vier unter Anleitung/mit Zuhörern.

Die Teilnehmer wurden aus der Teilnehmerdatenbank des Max-Planck-Instituts für empirische Ästhetik (MPIEA) rekrutiert und konnten sich online für einen Termin zur Gruppentestung anmelden. Dadurch variierte die Gruppengröße pro Testung. Das Kapazitätsmaximum lag bei zwölf Teilnehmern. Die Testung fand im Gruppentestungsraum des MPIEA statt, in dem jeder Teilnehmer an einem Rechner/Bildschirm sitzt und durch Trennwände von den Nachbarn getrennt ist.

Zuerst wurden die Teilnehmer über den Ablauf der Studie informiert, indem ihnen die Teilnehmerinformation ausgehändigt und erklärt wurde. Wenn sie in die Befragung einwilligten und alles verstanden hatten, gaben sie schriftlich ihr Einverständnis. Sie füllten dann einen Fragebogen zu ihren demographischen Daten aus (vgl. Kapitel 4.1.2).

Jeder Teilnehmer bewertete sechs von insgesamt 16 Stimmen, wobei es einen Sänger gab, den alle Teilnehmer bewerteten: *The Tallest Man on Earth* (Nr. 16). Anhand dieses Sängers sollten grundsätzliche Vergleiche hinsichtlich des Antwortverhaltens und einzelner Aspekte (wie z.B. den

Einfluss der Position des bewerteten Stücks im Testverlauf auf das Antwortverhalten) angestellt werden.

Die Musikbeispiele wurden über Kopfhörer (Beyerdynamic DT 770 Pro) gehört, wobei die Lautstärke individuell eingestellt werden konnte. Die Musiktitel wurden randomisiert nach vorher erstellten Listen abgespielt, damit a) die Stücke über die Studie hinweg möglichst gleichmäßig über die Teilnehmer verteilt bewertet wurden und b) dass die Darbietungsreihenfolge variierte.

Bei der Bearbeitung konnte jeder Teilnehmer sein Tempo selbst bestimmen. Nachdem ein Teilnehmer sechs Stimmen bewertet hatte, wurde die Abschlussbefragung durchgeführt. Die Testung dauerte ca. 1-1,5 Stunden und jeder Teilnehmer erhielt eine Aufwandsentschädigung von 15 €.

Während die randomisierte Darbietung der Musiktitel über den PC lief, wurden die Fragebögen auf Papier ausgefüllt. Der Vorteil an dieser Vorgehensweise war, dass alle Merkmale des Fragebogens auf einer Seite übersichtlich dargestellt und eingesehen werden konnten. Es war für diesen Schritt der Befragung wichtig, dass es die Möglichkeit gab, die Merkmale in Relation zueinander zu sehen. So konnten den Teilnehmern mögliche Überschneidungen und Redundanzen zwischen den Merkmalen sowie Unterschiede auffallen, die die Bedeutung eines Merkmals anhand eines anderen erklären/verdeutlichen konnten.

Die folgende Darstellung zeigt die Instruktionen:

Anleitung

Sie werden nun nacheinander sechs Singstimmen beschreiben und bewerten. Die Aufnahmen sind zwischen 30 Sekunden und 2 Minuten lang. Nehmen Sie sich alle Zeit, die Sie brauchen, um sich einen Eindruck von der Stimme zu verschaffen und die Merkmale zu beurteilen.

Die Merkmale umfassen eine Reihe von Eigenschaften, mit denen der Eindruck einer Stimme beschrieben werden kann. Die Antworten basieren auf *Ihrer* Wahrnehmung und Einschätzung der Stimme. Vielleicht wissen Sie bei dem einen oder anderen Merkmal nicht ganz genau, was damit gemeint ist. Versuchen Sie, sich anhand der Aufnahmen eine Vorstellung von den Merkmalen zu machen. Anmerkungen dazu können Sie uns in der Abschlussbefragung mitteilen.

Bei der Bewertung gehen Sie folgendermaßen vor: Zuerst *beschreiben* Sie die Stimme anhand eines Merkmals, bevor Sie dieses anschließend *bewerten*. Auf der linken Seitenhälfte unter „Beschreiben“ sehen Sie die Merkmale. Erscheint Ihnen die gehörte Stimmlage des Sängers zum Beispiel als eher „tief“, kreuzen Sie das Merkmal entsprechend an. Danach bewerten Sie, wie Ihnen diese eher tiefe Stimmlage *gefällt*. Während Sie sich bei der Beschreibung der Merkmale entscheiden müssen, d.h. ob die Stimmlage eher hoch oder tief ist, haben Sie beim Gefallen die Möglichkeit ein „weder noch“ in der Mitte anzugeben.

Beschreiben	Gefallen				
Die Stimme ist eher...	Die angekreuzte Eigenschaft der Stimme...				
tief <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> hoch	gefällt mir				gefällt mir nicht
	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Sie können die Merkmale in einer beliebigen Reihenfolge bearbeiten. **Bitte stellen Sie sicher, dass sie alle Merkmale bewertet haben, bevor Sie die nächste Seite des Fragebogens bearbeiten.**

Die Daten wurden in Tabellen übertragen und deskriptiv-statistisch in Microsoft Office Excel 2016 ausgewertet, die inferenzstatistischen Analysen wurden mit der Software SPSS 22 durchgeführt.

5.2 Ergebnisse

Im Folgenden wird zuerst dargestellt, welche Merkmale mit dem Gefallen assoziiert sind und inwiefern sich die Gefallens-Urteile in der Gruppen-Studie von den ablehnenden Urteilen der Interviewteilnehmer unterscheiden. Im Anschluss werden die Merkmale hinsichtlich der Fragebogenentwicklung analysiert.

5.2.1 Gefallen

Das Gefallen wurde auf zwei Ebenen evaluiert: Zum einen wurde zum Ende des Fragebogens nach dem Gefallen des Sängers, des Liedes, des Textes und des Musikstils gefragt, zum anderen wurde das Gefallen eines jeden Merkmales gesondert erfragt. Die Teilnehmer waren dazu angehalten, zuerst das Merkmal in seiner Ausprägung einzuschätzen (z.B. hoch oder tief) und dann das Gefallen eben dieser Ausprägung anzugeben (wurde z.B. ‚hoch‘ angegeben, wurde das Gefallen von ‚hoch‘ bewertet).

Das Ziel der Auswertung der Gefallens-Ratings lag darin, das Gefallen spezifischer Merkmale zu untersuchen. Das Gesamtgefallen kann mit einer Regressionsanalyse immer nur indirekt untersucht werden, d.h. welche Merkmale das Gefallen vorhersagen. Mit einer Einzelabfrage konnte überprüft werden, ob es Unterschiede zwischen dem Gesamtgefallen und dem Merkmalsgefallen gibt. Außerdem konnten daraus Rückschlüsse gezogen werden, ob evtl. eine Gesamteinschätzung des Gefallens ausreicht, oder Einzelabfragen detailliertere Aussagen zulassen.

Zuerst wurde analysiert, ob sich das Gefallen eines Sängers und der Merkmale je nach Position im Ablauf der Studie verändern. Einfaktorielle Varianzanalysen wurden berechnet, um die Zusammenhänge von Position und Gefallens-Ratings zu untersuchen. Diese Berechnungen wurden für den Song Nr. 16 durchgeführt, den alle 48 Teilnehmer an verschiedenen Positionen bewertet haben. Das Gefallen pro Merkmal und für den Sänger, das Lied etc. änderte sich nicht innerhalb einer Testung, d.h. die Position, an der die Stimme in der Testung erklang (Position 1 bis 5) ändert nichts an den Gefallens-Ratings, weder pro Merkmal, noch allgemein für den Sänger, das Lied, den Text oder den Stil ($p > 0,1$).

Außerdem sollte geschaut werden, ob die Einschätzung des Gesamtgefallens in Zusammenhang mit den Merkmalsgefallens-Ratings stehen, weiterhin, ob bestimmte Merkmale öfters gemeinsam gefallen bzw. nicht gefallen. Untersucht werden die Beziehungen 1) der Gesamtgefallens-Ratings und der Merkmalsgefallens-Ratings, 2) zwischen den Merkmalen und dem Gefallen eines Sängers und 3) die Beziehung der Merkmalsratings und der Merkmals-Gefallens-Ratings.

5.2.1.1 Relationen zwischen Merkmalen und Gefallen

Alle Merkmale gefallen entweder im Mittel ($MW < 3$) oder werden als neutral ($MW = 3$) eingestuft (Tabelle 12). Minimum und Maximum liegen bei 1-5, d.h. alle Werte sind vorhanden, und auch die Standardabweichung zeigt, dass es durchaus abgelehnte Merkmale bei einzelnen Personen gab. ‚Quäkig/nasal‘ und ‚gepresst‘ wurden eher abgelehnt als z.B. ‚laut‘. Der Gesamtmittelwert allerdings liegt bei 2,745, d.h. neutral mit leichter Tendenz zum Gefallen. Auch im allgemeinen Gefallen liegen die Werte bei neutral: Sänger ($MW = 3,12$, $SD = 1,32$), Lied ($MW = 3,16$, $SD = 1,43$), Text ($MW = 3,08$, $SD = 1,35$), Stil ($MW = 3,33$, $SD = 1,31$). Ein Intraklassen-Korrelationskoeffizient (absolute Übereinstimmung, Modell: zweifach gemischt) wurde für den Song Nr. 16 für alle 48 Teilnehmer und die vier Gefallens-Ratings berechnet. Cronbachs Alpha = 0,860 (Durchschnittliche Maße 0,855, Konfidenzintervall 0,775-0,912, $p < 0,001$) zeigt eine hohe Übereinstimmung zwischen den Teilnehmern.

Gefallen_Merkmal	Mittelwert	Standardabweichung
G_laut-leise	2,516	1,0704
G_markant-schlicht	2,572	1,1607
G_sprunghaft-gleitend	2,591	1,0863
G_tief-hoch	2,605	1,1863
G_weich-hart	2,612	1,1624
G_rau/kratzig-glatt	2,652	1,1377
G_dunkel-hell	2,669	1,1607
G_klangvoll-klangarm	2,720	1,3048
G_sprechen-rufen	2,805	1,1358
G_wechselhaft-gleichförmig	2,833	1,2244
G_gespannt-ungespannt	2,930	1,2140
G_locker-gepresst	3,058	1,2719
G_dumpf/weit-quäkig/nasal	3,117	1,1189

Tabelle 12: Mittelwert und Standardabweichung der Gefallens-Ratings pro Merkmal. Skala 1-5 von gefällt bis gefällt nicht.

Um die Entsprechungen des Gefallens pro Merkmal und dem Gesamtgefallen des Sängers zu überprüfen, wurden Spearman-Korrelationen ($N = 16$, alle Stücke) berechnet, die einen moderaten Zusammenhang zeigen (Median $r = 0,589$, $p < 0,05$), d.h. wird ein Sänger gemocht, zeigen dies die Einzelratings der Merkmale an. Lediglich die Merkmale Klangfülle ($r = 0,489$, $p = 0,055$) und Ausdruck ($r = 0,470$, $p = 0,066$) sind knapp über der $p < 0,05$ -Schwelle.

Es wurde als nächstes geschaut, welche Merkmale das Gefallen eines Sängers vorhersagen (Erläuterungen zum Verfahren finden sich in Merrill, 2017). Eine lineare Regression wurde über alle Interpretationen gerechnet (Methode: Einschluss) und zeigte ein signifikantes Modell

($F(13, 244) = 5,738, p < 0,001$), welches 23,4 % der Datenvarianz erklärte. Signifikante Prädiktoren sind die folgenden Merkmale:

- klangvoll-klangarm (Beta = 0,409, Standardfehler des Mittelwertes (SEM) = 0,091, $p < 0,001$)
- rau/kratzig-glatt (Beta = 0,195, SEM = 0,086, $p = 0,024$)
- locker-gepresst (Beta = 0,313, SEM = 0,099, $p = 0,002$)
- Sprechen-Rufen (Beta = 0,168, SEM = 0,084, $p = 0,047$)

Um die Merkmalsausprägungen zu ermitteln, wurden Spearman Korrelationen ($N > 255$) zwischen den vier signifikanten Items und den Gefallens-Ratings der Sänger durchgeführt: klangvoll-klangarm ($r = 0,343, p < 0,001$), locker-gepresst ($r = 0,297, p < 0,001$), sprechen-rufen ($r = 0,148, p = 0,017$) und rau-glatt ($r = 0,005, p = 0,941$). Somit sagen ‚klangarm, gepresst, rufen‘ (nur mit geringem Zusammenhang) und sowohl ‚glatt‘ als auch ‚rau‘ ein Nicht-Gefallen des Sängers voraus.

Es zeigte sich, dass bei einem Mittelwert pro Interpretation viel Information verloren geht, da nur wenige Merkmale im Mittel ein Nicht-Gefallen aufweisen (die einzigen Merkmale mit recht vielen Missfallens-Ratings sind ‚quäkig‘, ‚gepresst‘ und ‚gespannt‘). Um nun die Zusammenhänge zwischen den Merkmalen und dem Gefallen der einzelnen Merkmale zu untersuchen, wurden Kreuztabellen und Chi-Quadrat-Tests für jedes Merkmal mit Gefallens-Rating über alle Sänger und Merkmale berechnet. Die Werte wurden neu kodiert, damit zum einen die Mitte der Gefallens-Ratings ausgefiltert wurde, aber auch die Merkmalsausprägungen auf einen Wert reduziert wurden. Dadurch entstanden für jedes Merkmal und dessen Gefallens-Ratings nominale Variablen mit den Ausprägungen Gefallen und Nicht-Gefallen und jede Merkmalsausprägung (z.B. hoch-tief, siehe Tabelle 13). Kreuztabellen und Chi-Quadrat-Tests wurden berechnet mit einer Monte-Carlo-Simulation und Berechnung des Phi-Koeffizienten. Außerdem wurde eine z-Transformation der Daten mit Bonferroni-Korrektur vorgenommen. Die Ergebnisse der Chi-Quadrat-Tests und die Zusammenfassungen der Kreuztabellen sowie die Interpretation der Daten wurde in Tabelle 14 zusammengestellt.

			Gefallen		Gesamtsumme
			gefällt	gefällt nicht	
Merkmal	tief	Anzahl	63	13	76
		Erwartete Anzahl	52,2	23,8	76,0
	hoch	Anzahl	75	50	125
		Erwartete Anzahl	85,8	39,2	125,0
Gesamtsumme	Anzahl	138	63	201	
	Erwartete Anzahl	138,0	63,0	201,0	

Tabelle 13: Beispiel einer Kreuztabelle der beobachteten Anzahl und der erwarteten Anzahl an Ratings für das Merkmal tief-hoch mit den entsprechenden Gefallens-Ratings.

Merkmal	$\chi^2(1)$	$p \leq$	Phi	$p \leq$	Bewertung der Kreuztabelle (beobachtete Anzahl vs. erwartete Anzahl)
hoch-tief	11,513	0,001	0,239	0,001	mehr als erwartet für tief/Gefallen (beobachtete Anzahl 63 vs. erwartete Anzahl 52,5) und hoch/Missfallen (50 vs. 39,2), d.h. hoch gefällt weniger (vgl. Tabelle 13)
leise-laut		n.s.			[Generelle Anmerkung: Lautheit scheint für die Urteilsbildung nicht relevant zu sein, Laien können auch eher schlecht unterscheiden; siehe Kapitel 5.2.6.6]
klangvoll-klangarm	101,147	0,001	0,689	0,001	mehr als erwartet für klangvoll/Gefallen (110 vs. 74,5) und klangarm/Missfallen (71 vs. 35,5), d.h. klangarm gefällt weniger; kann auch aufgrund der positiven vs. negativen Konnotation der Begriffe kommen, da die Bedeutung der Begriffe teilweise unklar waren (siehe Kapitel 5.2.6.6)
dunkel-hell	21,24	0,001	0,337	0,001	mehr als erwartet für dunkel/Gefallen (55 vs. 40,8) und hell/Missfallen (58 vs. 43,8), d.h. hell gefällt weniger
weich-hart	20,26	0,001	0,321	0,001	mehr als erwartet für weich/Gefallen (96 vs. 81,6) und hart/Missfallen (39 vs. 24,6), d.h. hart gefällt weniger; Unterschiede in der Anzahl nicht so groß; auch hier spielen möglicherweise positive und negative Konnotationen der Begriffe eine Rolle
gespannt-ungespannt	18,38	0,001	0,309	0,001	mehr als erwartet für ungespannt/Gefallen (53 vs. 38,6) und gespannt/Missfallen (71 vs. 56,6), d.h. gespannt gefällt weniger
rau/kratzig-glatt	4,505	0,034	0,150	0,034	alle relativ ausgeglichen, mehr als erwartet für rau+kratzig/Gefallen (80 vs. 72,9) und glatt/Missfallen (38 vs. 30,9), d.h. glatt gefällt weniger und rau kann auch gefallen
dumpf/weit-quäkig/nasal	54,47	0,001	0,533	0,001	mehr als erwartet für dumpf+weit/Gefallen (51 vs. 27,3) und quäkig+nasal/Missfallen (96 vs. 72,3), d.h. quäkig/nasal gefällt weniger
locker-gepresst	97,75	0,001	0,703	0,001	mehr als erwartet für locker/Gefallen (67 vs. 33,6) und gepresst/Missfallen (101 vs. 67,6), d.h. gepresst gefällt weniger
sprunghaft-gleitend	9,91	0,001	-0,237	0,002	relativ ausgeglichen; mehr als erwartet für gleitend/Gefallen (95 vs. 86,2) und sprunghaft/Missfallen (25 vs. 16,2), d.h. sprunghaft gefällt weniger
markant-schlicht	38,37	0,001	0,448	0,001	mehr als erwartet für markant/Gefallen (100 vs. 80,8) und schlicht/Missfallen (42 vs. 22,8), d.h. schlicht gefällt weniger; man berücksichtige die geringe Anzahl für schlicht-Ratings und die berichteten Probleme mit den Begriffen (siehe Kapitel 5.2.6.4)
sprechen-rufen	13,69	0,001	0,277	0,001	mehr als erwartet für sprechen/Gefallen (79 vs. 67,3) und rufen/Missfallen (38 vs. 26,3), d.h. rufen gefällt weniger; Begriffsprobleme (siehe Kapitel 5.2.6.3)
wechselhaft-gleichförmig	28,79	0,001	0,383	0,001	mehr als erwartet für wechselhaft/Gefallen (63 vs. 44,7) und gleichförmig/Missfallen (69 vs. 50,7); d.h. gleichförmig gefällt weniger

Tabelle 14: Ergebnisse der Chi-Quadrat-Tests für die Merkmale und die Gefallens-Ratings der Merkmale mit entsprechender Interpretation der Kreuztabellen für jedes Merkmal.

Folgend den Phi-Koeffizienten gleichen die Ergebnisse den Regressionsmodellen, d.h. die folgenden Merkmalsausprägungen gefallen nicht:

- gepresst ($r = 0,703$)
- klangarm ($r = 0,689$),
- quäkig/nasal ($r = 0,533$),
- schlichte Artikulation ($r = 0,448$),

- monotoner Ausdruck, heller, harter, gespannter Klang, rufender Ausdruck, hohe Stimme, sprunghafte Tonhöhenveränderungen ($r < 0,4$)
- glatt ($r = 0,150$)

Wie die Rauigkeit kann auch die Phonationsform sowohl gefallen als auch nicht gefallen, was sich im Regressionsmodell am nur schwachen Zusammenhang zeigt. Die Merkmalsausprägung ‚gepresst‘ findet sich bereits in der Interview-Studie und auch die Rauigkeit war nicht immer abgelehntes Merkmal (vgl. Kapitel 4.2.3). Diese Ergebnisse scheinen recht stabil, müssen aber anhand einer veränderten Sängerauswahl weiter überprüft werden. Diese Ergebnisse zeigen auch, dass die Gesamtabfrage nach dem Gefallen des Sängers ausreichend und eine Einzelabfrage pro Merkmal nicht nötig ist. Allerdings muss der Einfluss des Liedes/der Musik auf die Bewertung des Gefallens berücksichtigt werden (vgl. Kapitel 6.2.6).

5.2.1.2 Gefallen und Missfallen einzelner Sänger

Folgende Aufstellung soll noch einen Blick darauf werfen, inwiefern sich die Sänger in den Gesamtgefällens-Ratings unterscheiden. Die Übersichtstabelle (Tabelle 15) aller Ratings und aller Sänger zeigt bereits, dass sich die Gefällens-Ratings nur vereinzelt zwischen Sänger, Stil, Lied und Text unterscheiden und es nur wenige Sänger gibt, die auf allen Ebenen nicht gemocht werden.

ID	SängerIn	Sänger_Gefallen		Lied_Gefallen		Text_Gefallen		Stil_Gefallen	
		MW	SD	MW	SD	MW	SD	MW	SD
1	Anastacia	3,125	1,0247	2,938	1,3401	3,188	1,0468	3,250	1,2383
2	Annett Louisan	3,333	1,4960	3,533	1,5523	2,733	1,2799	3,467	1,4573
3	Bob Dylan	2,700	0,9234	2,900	1,2096	2,900	1,0712	3,050	1,3945
4	Casper	3,333	1,5430	3,533	1,5523	3,067	1,1629	3,200	1,4243
5	Gitte Haening	3,667	1,3452	3,800	1,3202	3,600	1,4041	4,067	1,0998
6	Helge Schneider	3,188	1,4705	3,688	1,2500	3,500	1,4606	4,000	1,2111
7	Herbert Grönemeyer	2,941	1,2976	2,471	1,3747	2,294	1,2127	3,235	1,2005
8	Lena Meyer-Landrut	2,944	1,4337	3,389	1,4200	3,667	1,4552	3,389	1,3779
9	Mariah Carey	3,118	1,1114	3,059	1,3449	3,000	1,1180	3,176	1,1851
10	Max Herre	2,714	1,3260	2,867	1,5055	2,800	1,5213	3,133	1,2459
11	Michelle	4,059	1,1440	4,176	1,1311	4,176	1,0744	4,412	0,8703
12	Passenger	2,375	1,3102	2,250	1,5275	2,313	1,3022	2,688	1,3022
13	Anneliese Rothenberger	3,438	1,3150	3,875	1,0878	3,750	1,0646	3,688	1,2500
14	Udo Lindenberg	2,667	1,2344	2,733	1,2228	2,867	1,4573	2,933	1,2228
15	Xavier Naidoo	3,188	1,4245	2,813	1,3769	2,688	1,5370	2,938	1,2366
16	The Tallest Man on Earth	3,222	1,3086	2,667	1,4552	2,833	1,1504	2,889	1,2783

Tabelle 15: Mittelwerte (MW) mit Standardabweichung (SD) des Gefallens von Sänger, Lied, Text und Stil pro Stück. Skala 1-5 von gefällt bis gefällt nicht.

Auch die Spearman-Korrelationen verdeutlichen, dass die vier Gesamtgefällens-Einschätzungen in einem starken Zusammenhang stehen ($r_{\min} = 0,644$, $r_{\max} = 0,798$; Tabelle 16).

		Lied_ Gefallen	Sänger_ Gefallen	Text_ Gefallen	Stil_ Gefallen
Lied_Gefallen	Korrelationskoeffizient	1,000	,798**	,772**	,767**
	N	257	256	257	257
Sänger_Gefallen	Korrelationskoeffizient	,798**	1,000	,656**	,651**
	N	256	256	256	256
Text_Gefallen	Korrelationskoeffizient	,772**	,656**	1,000	,644**
	N	257	256	257	257
Stil_Gefallen	Korrelationskoeffizient	,767**	,651**	,644**	1,000
	N	257	256	257	257

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

Tabelle 16: Korrelationen nach Spearman zwischen den Gesamtgefallens-Ratings.

Eine einfaktorielle ANOVA (alle Stücke, Variable Sänger_Gefallen) bestätigt, dass sich die Stücke hinsichtlich des Gefallens des Sängers voneinander unterscheiden ($F(15) = 1,756$, $p = 0,042$). Tukey-HSD-Post-hoc-Tests zeigen, dass sich die Stücke 11 und 12 unterscheiden, d.h. Michelle (MW = 4,059) von Passenger (MW = 2,375) (Mittlere Differenz = 1,684, SD = 0,453, $p = 0,022$). Alle anderen Vergleiche bestehen die Korrektur nicht. Die nicht-korrigierten Daten zeigen jedoch eine Tendenz, dass Michelle sich noch von zwei weiteren Interpreten unterscheidet (Dylan und Lindenberg) und Passenger noch von fünf anderen. Somit werden die mit den höchsten und die mit den niedrigsten Gefallens-Ratings unterschieden. Da die vier Gesamtgefallens-Ratings hoch korrelieren, ist davon auszugehen, dass die Sänger-Ratings von dem Gefallen des Musikstils und des Liedes stark beeinflusst sind.

Mit t -Tests für unabhängige Stichproben wurden die Merkmale verglichen, in welchen sich Michelle und Passenger unterscheiden: Michelle ruft eher laut, hart, gespannt, gepresst, mit sprunghaften Tonhöhenveränderungen. Passenger hat eher einen sprechenden Gestus, leise, weich, ungespannt, locker, mit gleitenden Tonhöhenveränderungen. Ansonsten werden sie beide mit hoher Stimmlage sowie mit hellem, quäkigem Stimmklang beschrieben. Diese Berechnung erfolgte auf den Merkmals- und nicht den Gefallens-Ratings. Die Trennung der Sänger auf Basis der Merkmale wird in Kapitel 5.2.7 mit einer Diskriminanzanalyse weiter untersucht.

In der aktuellen Studie wurden die vier Fragen nach den Gesamtgefallen direkt nacheinander abgefragt, was die teilweise hohen Korrelationen im Gefallen erklären könnte. Für die Folgestudie ist deshalb wichtig, dass die Fragen nach dem Gesamtgefallen getrennt voneinander erfolgen müssen, um die hier gesehenen Einflüsse aufeinander zu kontrollieren.

5.2.1.3 Bekanntheit des Sängers

Bereits in der Interview-Studie wurden Fragen zur Bekanntheit des Sängers genutzt, um die Ablehnung stimmlicher Merkmale von einer möglichen Ablehnung der Person des Sängers abzugrenzen. Da die Ablehnung in der Gruppen-Studie nicht mehr im Fokus stand, wurde die Frage nach der Bekanntheit auf ja/nein reduziert, mit der zusätzlichen Frage nach dem Namen des Sängers. 10-14 korrekte Nennungen gab es für Herbert Grönemeyer, Udo Lindenberg, Lena Meyer-Landrut, Xavier Naidoo und Helge Schneider. Es gab noch 4-6 korrekte Nennungen für Bob Dylan, Anastacia, Annett Lousian, < 4

für Gitte, Casper, Passenger, Mariah Carey und The Tallest Man on Earth. Max Herre und Michelle sowie Anneliese Rothenberger wurden von keinem Teilnehmer erkannt. Michelle wurde dreimal mit Helene Fischer verwechselt.

Es ist interessant zu bemerken, dass nicht alle Teilnehmer die Namen wussten, aber teilweise dennoch die Sänger korrekt beschreiben konnten, zum Beispiel, dass jemand Juror in der Fernsehsendung ‚Deutschland sucht den Superstar‘ war. Das heißt, eine Bekanntheit ist nicht nur vom Namen abhängig. Im Fall von Verwechslungen mit anderen, stilgleichen Sängern (wie Helene Fischer und Michelle) wurde jedoch nur vermeintlich der Sänger erkannt, was darauf hinweist, dass der Teilnehmer nicht wirklich mit dem Sänger vertraut ist. Bekanntheit setzt sich also aus mehreren Facetten zusammen, die für eine höhere Differenzierung abgefragt werden müssten. Diese genauere Ausdifferenzierung würde jedoch deutlich mehr als nur eine Frage erfordern, weswegen in der Folgestudie nur abgefragt wurde, ob der Sänger bekannt ist, ohne weitere Rückfragen zu stellen.

5.2.2 Abhängigkeit der Bewertungen von der Position der Stimme im Testverlauf

Bevor die Merkmalsratings analysiert werden können, musste wie bei den Gefallens-Ratings getestet werden, ob die Bewertungen der Stimmen sich gegenseitig beeinflussten. Es wird immer wieder diskutiert, dass Stimmen auf Basis von Ankerbeispielen (Gerratt et al., 1993) beurteilt werden, d.h. dass eine Stimme im Vergleich zu einer anderen bewertet wird. In der aktuellen Studie wurde eine einfaktorielle Varianzanalyse durchgeführt, um zu testen, ob sich die Ratings der einzelnen Merkmale unterscheiden, je nachdem, an welcher Position (1-5) im Verlauf einer Testung die Stimme Nr. 16 beurteilt wurde. Es finden sich keine Unterschiede, die Verteilungen aller Merkmale sind über die Positionen hinweg identisch ($p > 0,265$). Das bedeutet, dass die Merkmalsratings unabhängig davon durchgeführt wurden, welche Stimme davor gehört worden war.

5.2.3 Zusammengehörigkeit von Merkmalen

Im ersten Schritt zur Entwicklung des Fragebogens wurde eine Hauptkomponentenanalyse (PCA) durchgeführt, um Merkmalskombinationen zu ermitteln. Der Screeplot (Abbildung 3) zeigt drei Faktoren, die einen Eigenwert > 1 aufweisen. Die erste Komponente erklärt 24,65 % der Datenvarianz, die zweite erklärt 17,92 % der Varianz, die dritte erklärt 11,96 % der Varianz.

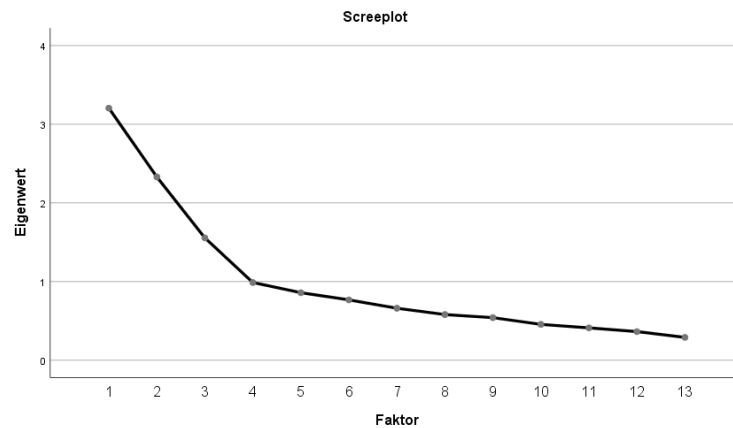


Abbildung 3: Screeplot zur Hauptkomponentenanalyse.

Merkmal	Merkmalsausprägungen	Komponente		
		1	2	3
Stimmhöhe	tief-hoch	,017	,835	-,081
Lautheit	laut-leise	,421	-,107	,530
Klangfülle	klangvoll-klangarm	-,508	-,048	,721
Klangfarbe	dunkel-hell	,055	,874	,084
Stimmklang	weich-hart	-,752	-,234	-,081
Spannung	gespannt-ungespannt	,656	,008	,115
Rauigkeit	rau/kratzig-glatt	,488	,563	-,183
Resonanz	dumpf/weit-quäkig/nasal	-,342	,436	,151
gepresst	locker-gepresst	-,764	-,005	,082
Tonhöhen- veränderungen	sprunghaft-gleitend	,589	-,153	,172
Artikulation	markant-schlicht	,080	,151	,673
Phonationsform	Sprechen-Rufen	-,437	,148	-,432
Ausdruck	wechselhaft-gleichförmig	,144	-,037	,636

Tabelle 17: Ergebnis der Hauptkomponentenanalyse. Rotationsmethode Oblimin mit Kaiser-Normalisierung.

Folgende Komponenten lassen sich aus Tabelle 17 ablesen, wobei die Phonationsform (Sprechen-Rufen) auf zwei Komponenten gleichermaßen lädt:

- Komponente 1: Stimmklang (weich-hart), Spannung (gespannt-ungespannt), gepresst, Tonhöhenveränderungen, Phonationsform (siehe Komponente 3)
 - Komponente 1 ‚Spannung‘ umfasst eher spannungsbezogene Merkmale, wie hart, gespannt, gepresst bzw. weich, ungespannt, locker sowie zwei andere Kategorien, die Tonhöhenveränderungen und die Phonationsform.
- Komponente 2: Stimmhöhe, Klangfarbe, Rauigkeit, Resonanz
 - Komponente 2 ‚Klang‘ umfasst klangliche Merkmale, wie tief, dunkel, rau und dumpf bzw. hoch, hell, glatt und quäkig.
- Komponente 3: Lautheit, Klangfülle, Artikulation, Ausdruck, Phonationsform (siehe Komponente 1)
 - Komponente 3 ‚Ausdruck‘ umfasst ausdrucksbezogene Merkmale, die die Artikulation einschließt.

Um diese Interpretationen der Komponenten weiter zu untersuchen und die Aussagen der Teilnehmer (siehe Kapitel 5.2.6) zu den Verwandtschaften von Merkmalen zu überprüfen, wurden vereinzelt Spearman-Korrelationen zwischen Merkmalen berechnet. Diese bestätigen, dass die Merkmale ‚tief-hoch‘ und ‚dunkel-hell‘ einen Zusammenhang zeigen ($N = 258$; $r = 0,629$, $p < 0,001$) sowie ‚gespannt-ungespannt‘ und ‚locker-gepresst‘ ($r = -0,547$, $p < 0,001$). Ein weiterer moderater Zusammenhang besteht zwischen den Merkmalen ‚weich-hart‘ und ‚rau/kratzig-glatt‘ ($r = -0,400$, $p < 0,001$), was den Vorschlag eines Teilnehmers untermauert, der Ausprägung ‚rau/kratzig‘ ein ‚weich‘ gegenüberzustellen. Der negative Korrelationskoeffizient zeigt, dass sich ‚hart‘ und ‚rau/kratzig‘ ähnlich sind.

5.2.4 Einigkeit und Übereinstimmung mit den Expertenratings

Als nächstes sollte geschaut werden, wie einig sich die Teilnehmer bei der Bewertung der Merkmale waren und ob ihre Ratings mit denen der Experten übereinstimmen. Für alle 206 Merkmals-Ratings (jedes Merkmal jedes Sängers) wurde folgendermaßen vorgegangen: Zuerst wurde untersucht, ob sich die Teilnehmer in ihren Ratings ähnelten (im Folgenden wird dies mit ‚Einigkeit‘ bezeichnet). Dazu wurden eine Reihe Einstichproben-*t*-Tests mit einem Testwert von 2,5, d.h. der Mitte der Rating-Skala, für jedes Merkmals-Rating durchgeführt. Ein signifikanter Test zeigte die mehrheitliche Abweichung vom Mittelwert in der Gruppe an, was die Einigkeit in einem Merkmals-Rating anzeigte. Dann wurden alle vom Mittelwert der Skala signifikant verschiedenen Merkmale mit den Experteneinschätzungen aus der Interview-Studie (Kapitel 4.1.5) verglichen. Dazu wurde deskriptiv analysiert, ob die kategoriale Einschätzung (z.B. hoch oder tief) mit denen der Experten übereinstimmt (im Folgenden der Einfachheit halber mit korrekt und inkorrekt beschrieben). Es ist zu beachten, dass für diese Studie keine gesonderten Expertenratings eingeholt wurden, da die bewerteten Merkmale weitestgehend mit der Interview-Studie übereinstimmten und hier nur Tendenzen einer Richtigkeit berücksichtigt werden sollten. Deshalb musste bei den Vergleichen berücksichtigt werden, dass die Expertenurteile in kategorialer Form vorlagen und sich außerdem zum Teil die Merkmalskombinationen in dem überarbeiteten Fragebogen von den Formulierungen der ersten Version unterschieden.

50,48 % der Einstichproben-*t*-Tests wurde signifikant, was bedeutet, dass sich bei der Hälfte aller Merkmals-Ratings die Mehrheit der Teilnehmer in einer Merkmalsausprägung einig war. Davon wichen sie nur in 9,12 % der Fälle von den Expertenratings ab. Um detailliertere Aussagen zu einzelnen Merkmalen treffen zu können, wurden die *t*-Tests nun pro Merkmal betrachtet, die Ergebnisse sind in Tabelle 18 zusammengefasst. Lediglich die Stimmlage zeigte eine relativ hohe Einigkeit mit korrekter Einschätzung bei zehn von 16 Sängern, gefolgt von Rauigkeit, Klangfarbe und Stimmklang bei jeweils acht Sängern. Die Merkmale Resonanz, Tonhöhenveränderungen, Phonationsform und Ausdruck lagen mit sieben Sängern im Mittelfeld, dicht gefolgt von Lautheit, gepresst, Spannung mit sechs korrekt bewerteten Sängern. Die Merkmale Klangfülle und Artikulation zeigten nur wenige signifikante *t*-Tests.

Merkmal	Merkmalsausprägung	Anzahl sign. <i>t</i> -Tests (von 16)	Anzahl einig+korrekt	Anzahl einig+inkorrekt	MW_SD
Stimm lage	tief-hoch	10	10	0	0,68
Rauigkeit	rau/kratzig-glatt	10	8	2	0,77
Klangfarbe	dunkel-hell	9	8	1	0,80
Stimmklang	weich-hart	9	8	1	0,76
Tonhöhenveränderungen	sprunghaft-gleitend	10	7	3	0,83
Phonationsform	Sprechen-Rufen	8	7	1	0,86
Ausdruck	wechselhaft-gleichförmig	8	7	1	0,93
Resonanz	dumpf/weit-quäkig/nasal	7	7	0	0,80
Lautheit	laut-leise	9	6	3	0,67
Spannung	gespannt-ungespannt	8	6	2	0,90
gepresst	locker-gepresst	7	6	1	0,85
Artikulation	markant-schlicht	6	3	3	1,01
Klangfülle	klangvoll-klangarm	4	3	1	0,95

Tabelle 18: Ergebnisse der Merkmalsratings hinsichtlich Einigkeit der Teilnehmer und korrekter Gruppenentscheidung im Vergleich zu den Experten. Pro Merkmal flossen die Merkmalsratings von 16 Sängern ein. MW_SD = Standardabweichung vom Mittelwert.

Grundsätzlich ist zu beachten, dass es Merkmale gibt, die bei dem einen Sänger stärker oder eindeutiger vorhanden sind als bei einem anderen. Durch die fehlende Mittelkategorie waren die Teilnehmer gezwungen, sich zwischen den beiden Polen zu entscheiden, selbst wenn die Stimme zwischen den beiden Polen wechselt oder in keiner der beiden Ausprägungen auffällig ist. Das heißt, dass eine geringe Einigkeit der Teilnehmer nicht nur auf ihre (ggf. fehlende) Kompetenz zurückgeführt werden kann (vgl. Kreiman et al., 2007). Eine geringe Anzahl an signifikanten *t*-Tests kann auch bedeuten, dass der Sänger/die Aufnahme dieses Merkmal nicht erkennen und somit bewerten lassen. Diesem Umstand wird mit der anschließenden Online-Studie (Kapitel 6) weiter nachgegangen.

5.2.5 Berichtete Schwierigkeiten bei der Bewertung

In der Abschlussbefragung wurden die Teilnehmer gebeten, auf einer 4-stufigen Skala von einfach bis schwierig anzugeben, wie schwierig die Bewertung der Merkmale für sie war. Danach gab es Raum für eine offene Antwort auf die Frage: „Warum fiel Ihnen die Bewertung schwer?“ Manche der Merkmalsausprägungen, vor allem die, die Begriffskombinationen enthielten (wie rau/kratzig) wurden einzeln abgefragt, um konkrete Rückmeldungen dazu zu bekommen.

Merkmale	MW	SD	Kommentare (von N = 48)	Kommentare in Prozent
dumpf/weit	2,7	1	21	43,75
gespannt-ungespannt	2,35	1,09	23	47,92
sprechen-rufen	2,04	1,04	16	33,34
markant-schlicht	2,02	0,92	13	27,08
sprunghaft-gleitend	1,91	0,79	6	12,5
glatt	1,9	0,9	9	18,75
wechselhaft-gleichförmig	1,86	0,75	6	12,5
dunkel-hell	1,85	0,85	14	29,17
laut-leise	1,77	0,77	13	27,08
klangvoll-klangarm	1,73	0,76	8	16,67
weich-hart	1,66	0,78	6	12,5
locker-gepresst	1,65	0,78	11	22,92
tief-hoch	1,6	0,84	14	29,17
quäkig/nasal	1,46	0,76	6	12,5
rau/kratzig	1,4	0,57	2	4,17

Tabelle 19: Deskriptive Statistik der Einschätzung der Schwierigkeit und Anzahl der Kommentare.

Fast die Hälfte der Teilnehmer äußerte sich zum Merkmalspaar ‚gespannt-ungespannt‘ (47,9 %) sowie ‚dumpf/weit‘ (43,75 %). Etwa ein Drittel kommentierte das Merkmal ‚Sprechen-Rufen‘ (33,3 %), ‚tief-hoch‘ (29,2 %, wobei hier in fast allen Fällen der Bezug zu ‚dunkel-hell‘ gemacht wurde, d.h. das Merkmal an sich war anscheinend unproblematisch), ‚dunkel-hell‘ (29,2 %) sowie auf Artikulation und Lautheit (jeweils 27,1 %; vgl. Tabelle 19).

Die Einschätzung der Schwierigkeit über alle Merkmale liegt im Mittel zwischen 1,6-2,7, d.h. sie wird als nicht besonders hoch eingeschätzt. Vier Merkmale werden im Mittel > 2 eingeschätzt: dumpf/weit, Spannung, Phonationsform und Artikulation. Die Merkmale ‚glatt‘, Tonhöhenveränderungen und Klangfarbe liegen knapp darunter. Grundsätzlich zeigt diese quantitative Übersicht, dass die Teilnehmer keine großen Schwierigkeiten mit dem Fragebogen berichteten. Die Kommentare der Teilnehmer wurden im Folgenden direkt in die Überarbeitung des Fragebogens einbezogen (Kapitel 5.2.6), eine Übersicht findet sich im Anhang 8.3.2.

5.2.5.1 Selbsteinschätzung zur Wiedergabe des Eindrucks von der Stimme

Die quantitative Abfrage danach, wie gut der Eindruck der Stimme mit den vorgegebenen Merkmalen wiedergegeben werden konnte, lieferte ein äußerst positives Ergebnis: Im Mittel konnten die Teilnehmer ihren Eindruck sehr gut wiedergeben (MW = 1,96, SD = 0,87; auf einer 5-stufigen Skala).

Auf die Rückfrage, was zur Wiedergabe des Eindrucks fehlte, bezogen sich fast die Hälfte der Kommentare (N = 21) noch einmal auf einzelne Merkmale, die im Fragebogen verwendet wurden. Dass zusätzliche Merkmale gefordert wurden, war selten: einmal wurde nach dem Tonhöhenumfang gefragt, zweimal wurden Vibrato und Tremolo vermisst und ebenfalls zweimal ein Merkmal zur Intonation sowie zur Phrasierung gewünscht.

Weitere Kommentare (N = 9) beziehen sich auf externe Bewertungskriterien wie Instrumentierung (das Zusammenspiel von Stimme und Instrumenten) oder der Musik-/Gesangsstil bzw. die Melodie an sich, manchmal wird auch eine Beurteilung der Technik des Sängers angedeutet.

Andere Anmerkungen (N = 10) lassen erkennen, dass Teilnehmer gerne mehr als nur den klanglichen Eindruck der Stimme beschreiben würden. Man möchte einerseits ausdrücken können, wie die Stimme auf einen wirkt („Merkmale, die beschreiben, was die Stimme vermittelt, nicht nur wie sie ist“, „so etwas wie Gestimmtheit“, z.B. „es klang sehr positiv“), andererseits wollen die Teilnehmer auch negative Empfindungen wiedergeben, also das Empfinden von Befremdlichkeit, Gleichgültigkeit, Abneigung oder Unwohlsein beim Hören. Auch werden Begriffspaare zur zusätzlichen Bewertung vorgeschlagen, wie „angenehm-unangenehm“, „beruhigend-nervig“, „monoton-anregend“ sowie „warm emotional-kalt emotional“. Das Bedürfnis, zu beschreiben, was die Stimme in einem auslöst und mit einem macht, also den persönlichen Eindruck wiederzugeben, scheint groß zu sein. Diese Anmerkungen sind besonders interessant, da sie Ähnlichkeiten mit den Ergebnissen der Interview-Studie aufzeigen und in der Folgestudie berücksichtigt werden sollten – es scheint für manche Teilnehmer unbefriedigend zu sein, Stimmen lediglich auf einer sachlich-klanglichen Ebene zu beschreiben, ohne ihr emotionales Erleben schildern zu können. Für Nicht-Experten-Hörer stellt offenbar ihre emotionale Reaktion auf eine Sängerstimme ein wichtiges Kriterium für deren Bewertung dar.

5.2.6 Schlussfolgerungen und Items für den neuen Fragebogen

An erster Stelle wurden bei der Entscheidung für die neue Zusammensetzung des Fragebogens die Kommentare der Teilnehmer berücksichtigt, dann die Auswertungen hinsichtlich der Einigkeit und Korrektheit und der Hauptkomponentenanalyse (PCA) bzw. der Korrelationen. Die Überschriften zeigen bereits das neue Item an.

5.2.6.1 *Spannung: gepresst-drucklos*

Mehrere Kommentare thematisieren die Überschneidung der Merkmale ‚gespannt-ungespannt‘ und ‚gepresst-locker‘ sowie spezifische Probleme mit dem Merkmal der Spannung. Die Teilnehmer sind sich in der Beurteilung dieser Merkmale auch uneinig. Da sich die berichteten Ähnlichkeiten der beiden Merkmale auch in einer Korrelation und der PCA (laden auf eine Komponente) zeigen, können die Merkmale kombiniert werden. ‚Gespannt‘ zeigt eindeutige Überschneidungen mit ‚gepresst‘, d.h. der Begriff ‚gepresst‘ kann somit auch im Sinne einer Steigerung von ‚gespannt‘ verwendet werden. Wegen des problematischen Begriffs ‚ungespannt‘, musste ein neuer Gegenpol zu ‚gepresst‘ gefunden werden. Dieser sollte einen klaren Spannungsbezug haben, z.B. ‚unterspannt, ungestützt oder entspannt‘, wobei diese Begriffe negativ oder positiv konnotiert sind oder eher dem Fachvokabular entsprechen. Als Pol soll deshalb in der Folgestudie der Begriff ‚drucklos‘ genutzt werden.

5.2.6.2 *Rauigkeit: rau/kratzig-weich*

Die Kommentare zeigen, dass der Gegenpol ‚glatt‘ zu ‚rau/kratzig‘ geändert werden muss. Während ‚rau‘ zur Beschreibung der Stimmen genutzt werden konnte, war ‚glatt‘ problematisch. Als neuer Pol wurde von einem Teilnehmer ‚weich‘ vorgeschlagen. Das Merkmal ‚weich-hart‘ zeigte immerhin eine Einigkeit der Teilnehmer bei der Hälfte der Sänger und korreliert moderat mit dem Merkmal ‚rau/kratzig-glatt‘. Deshalb soll eine Kombination der Merkmale in der Folgestudie erprobt werden.

5.2.6.3 *Phonationsform: gesprochen-gesungen*

Die Phonationsform wird immerhin bei sieben Sängern (von 16) einheitlich korrekt zugeordnet. Ausgehend von den Kommentaren scheint eine Kategorie zu fehlen, die das Singen abbildet. Möglichkeiten sind eine kategoriale Variable oder eine mittlere Kategorie ‚Singen‘. ‚Sprechen‘ kann als Pol bleiben, dem der Gesang gegenübergestellt wird, ‚Rufen‘ müsste herausgenommen werden. Dies ist interessant, wenn an die Gospel-Shouter (siehe Kapitel 3) und den Metal-Gesang gedacht wird, die durchaus die Qualität des Rufens beinhalten. Auch das Belting/der Belt ist eine Form des Rufens. Das bedeutet, dass dieses Merkmal in Abhängigkeit von der Forschungsfrage bestimmt werden sollte, also inwiefern Schreien oder Rufen wichtige zu erfragende Merkmale sind. Wenn der Begriff der ‚Phonationsform‘ vermieden werden soll, bietet sich folgende Formulierung an: ‚Die Vortragsform ist eher gesungen-gesprochen.‘ Dies bildet auch die Komponentenzugehörigkeit ab, denn Phonationsform lädt auf denselben Faktor wie Ausdruck.

5.2.6.4 *Artikulation: präzise-unpräzise und eigentümlich-gewöhnlich*

Ausgehend von den Kommentaren scheint bei vielen Teilnehmern nicht deutlich geworden zu sein, dass mit dem Merkmal ‚markant-schlicht‘ Besonderheiten der Artikulation jenseits der reinen Textverständlichkeit gemeint waren. Verständlichkeit an sich war bislang nicht im Fokus der Befragung. Stattdessen sollten individuelle Auffälligkeiten in der Artikulation entdeckt werden, die stören oder gefallen. Dies kann Verständlichkeit als einen Teilaspekt beinhalten, ohne dass darauf eindeutig eingegangen wird. Anscheinend wurde der Eindruck von Verständlichkeit darüber nicht aufgefangen und muss gesondert abgefragt werden, d.h. neben ‚präzise-unpräzise‘ wird nun ‚eigentümlich-gewöhnlich‘ zusätzlich mit aufgenommen, um den Eindruck des Auffälligen, Besonderen, Interessanten etc. in der Artikulation wiedergeben zu können.

5.2.6.5 *Resonanz und Klangfarbe: quäkig/nasal-dunkel/dumpf*

Von den Begriffen der Merkmalskombination ‚quäkig/nasal-dumpf/weit‘ wird ‚weit‘ am häufigsten kritisiert, während kaum Probleme mit ‚dumpf‘ und ‚quäkig/nasal‘ berichtet werden. Bei fast der Hälfte der Sänger sind sich die Teilnehmer einig und bewerten korrekt, inkorrekt bewerten sie im Mittel nie. Wird das Merkmal als Durchdringungsfähigkeit betrachtet, ist ‚dumpf‘ ein passender Gegenpol und

kann beibehalten werden. Da Klangfarbe und ‚dumpf-quäkig‘ auf denselben Faktor laden, kann ‚dunkel‘ als Pol ergänzt werden, dem ‚quäkig‘ gegenüberstehen kann.

Nach Aussagen der Teilnehmer scheint das Merkmal Klangfarbe redundant zu sein, d.h. es kann in einem anderen Merkmal aufgehen. Dieser Eindruck wird durch die PCA unterstützt, wo Klangfarbe und Stimmklang auf einen Faktor laden. Auch eine Spearman-Korrelation, um die Aussagen der Teilnehmer zu den Verwandtschaften von Merkmalen zu überprüfen, bestätigte, dass die Merkmale ‚tief-hoch‘ und ‚dunkel-hell‘ einen Zusammenhang zeigen. Deshalb scheint es nicht eigenständig zu sein und kann mit der Resonanz kombiniert werden.

5.2.6.6 *Klangfülle und Lautheit*

Die Merkmale ‚klangvoll-klangarm‘ und ‚laut-leise‘ scheinen aus denselben Gründen problematisch zu sein: Sie brauchen Bezugspunkte zur Beurteilung und sind in den Aufnahmen (vor allem populärer Musik) aufgrund der Aufnahmetechnik nur schwer zu beurteilen, wie aus den Kommentare ersichtlich wird. Auch zeigen sich nur bei wenigen Sängern Einigkeit und Korrektheit. Beide werden herausgenommen.

5.2.6.7 *Stimmlage: tief-hoch*

Stimmlage ist das am besten eingeschätzte Merkmal mit zehn korrekt (keine inkorrekt) eingeschätzten Stimmen. Stimmlage lädt auf eine Komponente zusammen mit Klangeigenschaften, was zeigt, dass die Stimme eher dahingehend eingeschätzt wird, wie sie klingt bzw. ist, als dass die Lage bewertet wird, in der gesungen wird. Für den neuen Fragebogen kann das Merkmal daher zusammen mit den Klangmerkmalen abgefragt werden und mit ‚Die Stimme klingt...‘ eingeführt werden.

5.2.6.8 *Ausdruck: wechselhaft-gleichförmig*

Zum einen sollte im neuen Fragebogen spezifiziert werden, dass es um den stimmlichen Ausdruck geht, da nicht ausgeschlossen werden kann, dass Teilnehmer zum Teil den emotionalen Ausdruck (mit-)bewertet haben. Zudem sollte in den allgemeinen Anweisungen mitgeteilt werden, dass mit den vorliegenden Merkmalen der Gesamteindruck der Stimme wiedergegeben werden soll. Das bedeutet, dass eine Art interner Durchschnitt gebildet wird (z.B. darüber, wie oft im Verlauf des Stückes der Ausdruck wechselt), der ein Merkmal in seinem überwiegenden Eindruck widerspiegelt.

5.2.7 Beschreibungsmuster und Diskriminanz der Sänger

Es ist u.a. ein erklärtes Ziel des zu entwickelnden Inventars, dass die ausgewählten Merkmale zu einer möglichst charakteristischen Beschreibung von Stimmen führen. Dies wurde im Folgenden mit den beschriebenen Merkmalen untersucht; sowohl auf Basis der erhobenen Merkmale als auch der neu entwickelten, sodass dieses Vorgehen für die geplante Folgestudie erprobt werden konnte.

In der vorliegenden Studie wurden 16 Sängerinnen und Sänger anhand von 13 Merkmalen in ihrem stimmlich-artikulatorischen Ausdruck bewertet. Im Folgenden soll betrachtet werden, wie gut sich die Stimmen anhand der Einschätzungen klassifizieren (trennen) lassen und welche Merkmale besonders stark zu einer Unterscheidung beitragen. Obwohl 16 Stimmen eine recht große Anzahl ist, wurden zunächst alle berücksichtigt. Bei der Vorgehensweise wurde sich an Studien orientiert, die die multiple Diskriminanzanalyse (MDA) zur Klassifikation von Emotionen (Bänziger et al., 2014) und Intentionen (Hellbernd & Sammler, 2016) in der gesprochenen Sprache untersucht haben.

Mit der MDA wird getestet, in welchem Ausmaß die Merkmalsratings zu einer Trennung der Stimmen beitragen. Das Ergebnis zeigt, dass 36,1 % der kreuzvalidierten gruppierten Fälle korrekt klassifiziert wurden (54,1 % der ursprünglichen Fälle). Am stärksten tragen dazu die Merkmale Phonationsform ($r = 0,616$), Lautheit ($r = -0,587$) und Stimmklang ($r = 0,511$) in der ersten Funktion (erklärt 35,6 % der Varianz) sowie Rauigkeit ($r = 0,774$), Stimmlage ($r = 0,535$) und Klangfarbe ($r = 0,456$) in der zweiten Funktion (erklärt 30,9 % der Varianz) bei. In weiteren Funktionen (deren Eigenwert < 1 ist, aber folgend Wilks' Lambda signifikant zur Diskrimination beitragen) treten noch ‚dumpf/weit-quäkig/nasal‘ (Resonanz; $r = 0,751$) und die Tonhöhenveränderungen ($r = 0,575$) auf.

Die Klassifizierung mit Kreuzvalidierung von 36 % ist niedrig und das Ergebnis aufgrund methodischer Restriktionen nur bedingt aufschlussreich, da sehr viele Stimmen mit recht vielen Merkmalen und nur relativ wenigen Teilnehmern beschrieben wurden. Das Ergebnis der MDA stimmt allerdings mit Überlegungen in Folge der PCA überein (die Methoden sind sich auch ähnlich). Zum Beispiel haben manche Merkmale (wie die Spannung) keinen Einfluss auf die Klassifikation und einige Merkmale der ersten und zweiten Funktion zeigen nur mittlere Korrelationskoeffizienten, d.h. einige davon könnten in anderen Merkmalen aufgehen (siehe Kapitel 6.2.4). Es ist auch noch darauf hinzuweisen, dass manche Merkmale als schwierig kritisiert wurden (wie ‚glatt‘, Lautheit, Klangfülle), die hier dennoch mit eingeflossen sind.

Die Klassifikation der Sänger (Tabelle 20) zeigt, welche Sänger besser und schlechter klassifiziert werden konnten: Am schlechtesten konnte Bob Dylan (ID 3) klassifiziert werden, am besten Casper (ID 4). Dieses Ergebnis lässt sich auch dahingehend interpretieren, dass manche Sänger besser zu beschreiben sind als andere und evtl. auffälligere Merkmale zeigen, auf deren Basis eine Diskrimination vorgenommen werden kann.

ID	SängerIn	% korrekt klassifiziert
4	Casper	85,7
13	Rothenberger	75
7	Herbert Grönemeyer	70,6
1	Anastacia	68,8
5	Gitte Haenning	68,8
2	Annett Louisan	64,3
10	Max Herre	64,3
16	The Tallest Man on Earth	61,1
14	Udo Lindenberg	60
6	Helge Schneider	56,3
9	Mariah Carey	47,1
8	Lena Meyer-Landrut	44,4
11	Michelle	35,3
12	Passenger	31,3
15	Xavier Naidoo	31,3
3	Bob Dylan	6,7

Tabelle 20: Ergebnis der Klassifikation der Sänger (ursprünglich gruppierte Fälle).

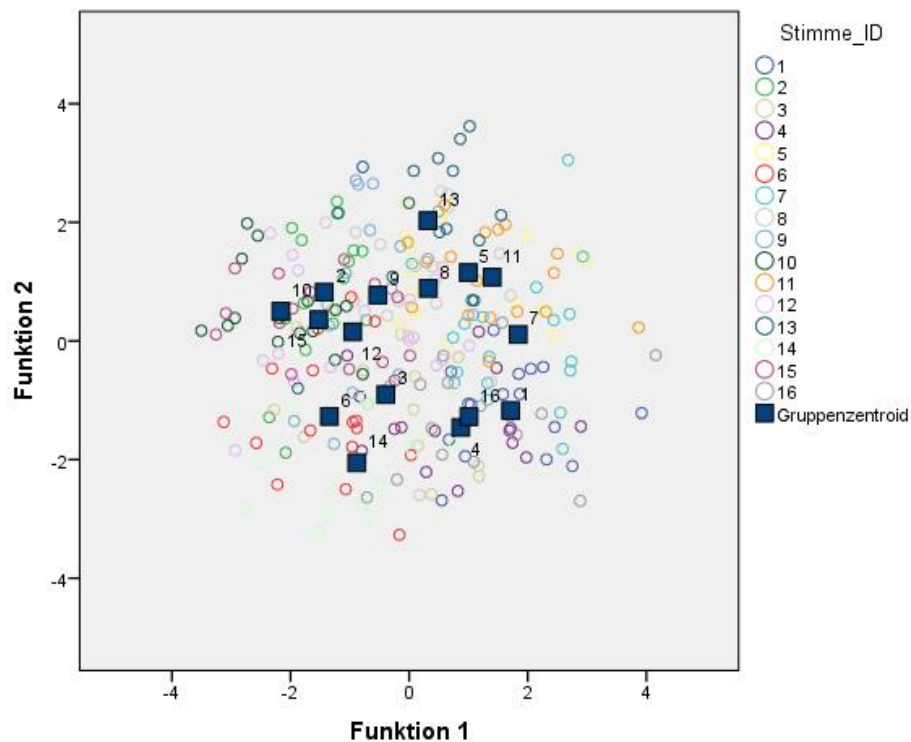


Abbildung 4: Streudiagramm der 1. und 2. Funktion der MDA über alle Stücke und alle Merkmale mit Gruppenzentroiden. Stimme_ID siehe Tabelle 20.

5.2.7.1 Explorative Berechnung einer 6-Stimmgruppen- und 8-Merkmals-Lösung zur Diskriminanz der Sänger

Die folgende Berechnung sollte Tendenzen aufzeigen, inwiefern die in Kapitel 5.2.6 gemachte Auswahl an Merkmalen für den neuen Fragebogen mit weniger Stimmen zu einer Differenzierung der Stimmen führen kann. Dies sollte zum einen die neue Merkmalsauswahl bestätigen und auch die Anzahl an Sängern aufzeigen, die zu einer besseren Diskriminanz führen. Dazu wurden zuerst mit einer Clusteranalyse Stimmen zusammengeführt. Es wurden hierarchische Cluster gebildet folgend der Ward-Methode, basierend auf der quadrierten euklidischen Distanz. Es wurde sich für eine 6-Cluster-Lösung entschieden. Die Stimmen wurden neu gruppiert: so fielen z.B. die Stimmen Nr. 5, 8, 10, 11 zusammen unter die Nr. 5.

Danach wurde eine weitere MDA unter Ausschluss der problematischen bzw. redundanten Merkmale durchgeführt. Dadurch werden nun fünf Funktionen laut Wilks' Lambda signifikant. Auf die erste Funktion laden Rauigkeit ($r = 0,768$) und Stimmklang ($r = -0,591$), auf die zweite Phonationsform ($r = 0,777$), Ausdruck ($r = -0,580$) und ‚gepresst-locker‘ ($r = 0,366$), auf die dritte Tonhöhenveränderungen ($r = -0,527$), auf die vierte Resonanz ($r = -0,589$) und auf die fünfte Funktion Stimmlage ($r = 0,721$). Es ist zu erwarten, dass sich die Funktionen verändern, je nach Stimmen und Merkmalen, die in die Analyse einfließen.

Sänger-Gruppe	% korrekt (kreuzvalidiert)
1	73,5
2	61,7
3	52,5
4	80
5	37,9
6	63,6

Tabelle 21: Ergebnis der Klassifikation der Sängergruppen.

56,4 % der kreuzvalidierten Fälle wurden ordnungsgemäß klassifiziert (61,8 % der ursprünglich gruppierten Fälle). Die Klassifikation funktionierte damit recht gut, nur die Sänger-Gruppe 5 fällt niedriger aus. Diese explorative Analyse zeigt, dass mit weniger Stimmen und Merkmalen und mit mehr Ratings die Diskrimination erwartungsgemäß besser funktioniert. Auch würden andere Stimmen, die größere Unterschiede zeigen, zu einer besseren Diskriminanz führen. Das Ergebnis der MDA zeigt eine teilweise erwartete, teilweise neue Gruppierung der Merkmale: Interessant ist nun, dass die Klangqualität ‚rau/kratzig-glatt‘ einen großen Einfluss auf die Diskriminanz hat.

Mit diesem letzten Schritt waren die Voraussetzungen zur Überprüfung des neuen Inventars geschaffen. Die vokalen Merkmale wurden auf Basis der Ratings und der Kommentare der Teilnehmer neu für die Folgestudie zusammengestellt. Der Fragebogen umfasst nun neun Merkmale, die bei einer kleineren Gruppe Sänger zu einer guten Diskriminanz führen und kann somit in eine abschließende Evaluation des Inventars zur Beschreibung von Singstimmen überführt werden.

6 Fragebogenentwicklung und ästhetische Emotionen – Online-Studie

Das neue Inventar wird nun anhand einer wesentlich größeren Anzahl Teilnehmer überprüft. Dabei sollten die eruierten neun Merkmale anhand von sechs Stimmen bewertet werden. Das Gesamtgefallen des Sängers sowie des Liedes wurde weiterhin abgefragt, allerdings nun getrennt voneinander. Zusätzlich sollten neben dem Gefallen emotionale Aspekte abgefragt werden, die sowohl in der Interview-Studie (als emotionale Gründe, Kapitel 4.2.2.1) eine Rolle spielten als auch der Gruppen-Studie von den Teilnehmern gewünscht wurden. Es gilt auch nachzuvollziehen, ob die Urteilsbildung über eine Stimme eher auf der Ebene vokaler Merkmale oder emotionaler Wirkungen erfolgt. Die Erstellung des neuen Fragebogens unter Berücksichtigung dieser Bereiche wird im Folgenden dargestellt.

6.1 Methoden

6.1.1 Auswahl der Stimmen und Items

Da die Quellenlage zur Wirkung von Stimmen hinsichtlich der gefühlten ästhetischen Emotionen gering ist, wurde für die aktuelle Befragung auf eine Auswahl an ästhetischen Emotionen aus dem Bereich der Musik und allgemein der Bewertung von Kunst zurückgegriffen (Schindler et al., 2017; Zentner et al., 2008). Die Auswahl entsprach teilweise dem, was die Teilnehmer angemerkt hatten, so die Items ‚berührt mich‘ und ‚fasziniert mich‘. Ansonsten wurde versucht, die Dimensionen der Emotionen abzubilden: Valenz, in Form von traurig und fröhlich, und Erregung, in Form von ‚energetisiert mich‘, ‚macht mich gereizt‘ und ‚entspannt mich‘ (eine Emotion, die auch im vokalen Merkmal der Spannung ‚gepresst-drucklos‘ reflektiert ist). Es ist zu klären, ob diese geläufigen ästhetischen Emotionen durch die Stimme entstehen oder durch die Musik bzw. den Song an sich. Diese Differenzierung ist jedoch nur sehr beschränkt möglich, wenn Stimme und Instrumente zusammen rezipiert werden. Da diese Studie aber auf das alltägliche Hörerleben abzielte, wurden nur Stimmen im Kontext von Instrumenten untersucht, keine A-cappella-Stimmen. Eine Differenzierung zwischen Stimme und Song kann also nur auf der Ebene des Gefallens abgefragt und darauf bezogen werden. Die Abfrage nach dem Gefallen des Sängers und des Songs wurde nun getrennt voneinander vorgenommen: Nach dem Sänger wurde gleich zu Beginn, nach dem ersten Hören, gefragt, nach dem Song erst nach dem Bewerten der Emotionen und der Merkmale.

Die Bewertung der Merkmale und Emotionen geschah mit der Anweisung, auf die Stimme zu achten und so gut als möglich die Musik außer Acht zu lassen. Es ist davon auszugehen, dass dies besonders bei den Emotionen schwierig ist, allerdings können so auf Ebene des Gefallens Rückschlüsse gezogen werden, ob die Items eher das Gefallen der Stimme oder das des Songs vorhersagen. Auch können zum

ersten Mal Beziehungen zwischen den vokalen Merkmalen und den ästhetischen Emotionen hergestellt werden.

Die Auswahl der Sänger erfolgte unter folgenden Gesichtspunkten: Die Anzahl sollte sowohl die geplante Diskriminanzanalyse ermöglichen als auch zugleich die Bearbeitungszeit der Umfrage für die Teilnehmer nicht übermäßig in die Länge ziehen. Die Gruppen-Studie hatte explorativ gezeigt, dass sechs Stimmen für die Analyse ausreichen könnten, was auch den Teilnehmern zuzumuten war. Auch wurde die Auswahl an Sängern geändert, sodass die bisher gefundenen Merkmale zum Gefallen zwischen den Studien verglichen werden konnten. Es wurde sich für sechs populäre Stimmen und Songs entschieden, bei denen aufgrund ihres Alters und ihrer Positionen in den Charts davon auszugehen ist, dass sie von den Teilnehmern gekannt werden. Der Grund dafür ist, dass schlecht zu definieren ist, was mit einer Abfrage nach der Bekanntheit eigentlich gemeint ist. Der Einfluss der Bekanntheit auf die Bewertung sollte damit primär kontrolliert und nicht spezifischer untersucht werden. Die Bekanntheit wurde, wie schon vorher erprobt (Kapitel 5.2.1.3), deshalb nur mit ‚ja‘ und ‚nein‘ erfragt. Es kann davon ausgegangen werden, dass den Teilnehmern die Songs insofern bekannt sind, als dass sie sie schon einmal gehört haben. Ob sie jedoch den Titel des Songs und den Sänger benennen könnten, ist für diese Umfrage nicht wichtig.

Es wurden drei weibliche und drei männliche Sänger ausgewählt (ungeachtet davon, ob diese auch stimmlich als männlich oder weiblich wahrgenommen werden; Tabelle 22), die in der entsprechenden Aufnahme gut zu hören sind und vom Höreindruck her wenige oder keine nachträglichen akustischen Bearbeitungen erfahren haben (ein wenig zusätzlicher Hall kann bei Houston gehört werden). Jeder Ausschnitt wurde so gewählt, dass er ca. 30 Sekunden lang ist und einen charakteristischen Teil des Songs darstellt und in dem die Stimme sich in Klangbild und Technik nicht zu sehr ändert (was oft zwischen Strophe und Refrain der Fall ist; Seedorf, Pfeleiderer, Richter & Mecke, 2016, S. 577). Die Doppelung Dylans mit den vorherigen Studien hatte keine weiteren Gründe.

Sänger	Titel	Album	Erschienen	Ausschnitt
Bob Dylan	Don't Think Twice it's all Right	The Freewheelin' Bob Dylan	1963	00:07-00:43
Elvis Presley	Love Me Tender	Love Letters from Elvis	1971	00:02-00:41
James Brown	I Got You (I Feel Good)	Out of Sight	1964	01:26-02:03
Tina Turner	What's Love Got to Do with it	Private Dancer	1984	00:45-01:12
Wanda Jackson	Let's Have a Party	Wanda Jackson	1960	00:04-00:33
Whitney Houston	One Moment in Time	One Moment in Time: 1988 Summer Olympics Album	1988	00:59-01:30

Tabelle 22: Ausgewählte Sänger für die Online-Studie.

6.1.2 Ablauf

Die Studie wurde über soziale Netzwerke wie Facebook beworben und mit dem Titel *Studie zur Stimmästhetik* angekündigt (die Ankündigung findet sich im Anhang 8.4), mit der die Wirkung und das

Gefallen von Singstimmen der populären Musik untersucht werden sollte. Es wurde darauf hingewiesen, dass kein Vorwissen zur Teilnahme nötig ist. Die Durchführung dauerte ca. 20 Minuten.

Zuerst wurden demographische Daten erfasst. Neben Geschlecht und Alter wurde der sozioökonomische Status (Ausbildung, Beruf) sowie die Muttersprache (deutsch ja/nein) erfasst. Die musikalische Vorbildung und die Hörgewohnheiten wurden mithilfe des *MUSE Questionnaire* (Chin & Rickard, 2012) erfasst, der von der Autorin ins Deutsche übersetzt und adaptiert wurde. Außerdem wurde die Vorerfahrung im Umgang mit der Stimme erfragt. Dazu wurde eine offene Frage mit Beispielen gestellt: „Haben Sie privat oder beruflich mit Stimmen zu tun? Z.B. Tätigkeit als SängerIn, SchauspielerIn, SprecherzieherIn, StimmtherapeutIn, GesangspädagogIn; Studium / Ausbildung / Abschluss Sprechwissenschaft, Sprecherziehung, Logopädie etc.; Wie lange schon? In welchem Stil? In welchem Fach? Haben/hatten Sie Gesangsunterricht?“.

Danach wurde allgemein nach dem Gefallen von Musikstilen gefragt, allerdings mit Bezug zum Gesang: „Wie gefällt Ihnen der Gesang in den folgenden Musikstilen?“ Zur Auswahl standen „Klassik, Jazz, Blues/Soul, Pop, Country, Rock, Metal, Rap/Hip Hop, Musical, Folk/Singer-Songwriter (Liedermacher) und Weltmusik“, die auf einer 5-stufigen Skala bewertet wurden, von „gefällt mir“ bis „gefällt mir nicht“ (wie alle folgenden Fragen zum Gefallen). Im Anschluss wurde nach konkreten SängerInnen gefragt, die gemocht und nicht gemocht wurden. Es konnten bis zu drei SängerInnen genannt werden. Diese Abfrage sollte ein Abbild aktuell eher gemochter und weniger gemochter Sänger geben.

Danach begann die Bewertung der sechs ausgewählten Sänger, die in randomisierter Reihenfolge präsentiert wurden. Zuerst wurde der Ausschnitt komplett angehört, dann erschien die Frage nach dem Gefallen der Stimme auf einer neuen Browser-Seite. Auf einer weiteren Seite wurde nach den Emotionen gefragt: „Welche Gefühle löst die Stimme in Ihnen aus?“ (Tabelle 23). Dazu konnte die Aufnahme wiederholt angehört werden.

	(1) überhaupt nicht	(2) eher nicht	(3) mittelmäßig	(4) ziemlich	(5) sehr stark
berührt mich	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
fasziniert mich	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
entspannt mich	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
energetisiert mich	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
macht mich fröhlich	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
macht mich traurig	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
macht mich gereizt	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Tabelle 23: Bewertung der ästhetischen Emotionen. Die Emotionen wurden für jeden Teilnehmer randomisiert.

Nach der Bewertung der Emotionen (alle Items mussten beantwortet werden) wurden die vokalen Merkmale beschrieben: „Welche Eigenschaften hat die Stimme? Nun geht es um eine sachliche Beschreibung der Stimme. Bitte beschreiben Sie den Gesamteindruck der Stimme anhand der vorgegebenen Merkmale. Es ist hilfreich, die Aufnahme während der Beschreibung anzuhören.“ Diese

wurden auf einer Seite dargestellt, damit sie immer im Vergleich zueinander bewertet werden konnten (Tabelle 24).

Die Stimme klingt eher...							
tief	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	hoch
rau/kratzig	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	weich
gepresst	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	drucklos
quäkig/nasal	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	dunkel/dumpf
Die Veränderungen von Ton zu Ton sind eher...							
sprunghaft	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	gleitend
Die Artikulation ist eher...							
präzise	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	unpräzise
eigentümlich	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	gewöhnlich
Die Vortragsform ist eher...							
gesungen	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	gesprochen
Der stimmliche Ausdruck ist eher...							
wechselhaft	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	gleichförmig

Tabelle 24: Beschreibung der vokalen Merkmale auf einer 6-stufigen Skala.

Auf einer weiteren Seite wurde das Gefallen des Songs bewertet (auf dieselbe Weise wie das Gefallen der Stimme). Danach wurde die Bekanntheit erfragt: „Kennen Sie den Song? ja/nein“ und „Kennen Sie den Sänger? ja/nein“.

Dieser Vorgang wiederholte sich, bis alle sechs Stimmen bewertet wurden. Abschließend gab es die Möglichkeit für einen Kommentar. Im Anschluss konnte an einer Verlosung teilgenommen werden, in der jeder zehnte Teilnehmer einen amazon-Gutschein in Höhe von 10 Euro gewann.

Die Daten wurden als .csv-Datei exportiert und mit Microsoft Excel 2016 und SPSS Version 23 ausgewertet.

6.1.3 Teilnehmer

216 Teilnehmer (weiblich = 138, männlich = 77; keine anderen Angaben) flossen in die Auswertung ein. Die Teilnehmer waren im Mittel 31,29 Jahre alt, Standardabweichung (SD) = 12,22, Umfang 18-68 Jahre.

Die demographischen Daten verteilten sich wie folgt: Die meisten Teilnehmer hatten als höchsten Bildungsabschluss das Abitur (N = 87), weitere einen Bachelor (N = 41), Master o.ä. (N = 58) oder eine Promotion (N = 9). Einen niedrigeren Abschluss als Abitur hatten neun Teilnehmer, einen anderen als aufgelistet 14 Teilnehmer. Die aktuelle berufliche Situation gestaltete sich folgendermaßen: Der Großteil der Teilnehmer war Azubi / SchülerIn / StudentIn (N = 113), andere waren angestellt (N = 67) oder selbstständig (N = 33), nur wenige waren arbeitslos (N = 4), RenterIn/PensionärIn (N = 3) oder sonstiges (N = 14).

Neben der musikalischen Vorbildung sollte die Vorerfahrung im Bereich Stimme erfragt werden, um die Gruppe in solche mit und ohne Stimmerfahrung einzuteilen. Auf die Frage „Haben Sie privat oder beruflich mit Stimmen zu tun?“ sollte frei geantwortet werden und es ließen sich drei Gruppen ableiten:

- 53,24 % (N = 115) ohne Stimmerfahrung
- 19,98 % (N = 41) mit Stimmerfahrung ‚Sprechen‘ (z.B. Sprechwissenschaftler, Logopäden, Schauspieler)
- 27,78 % (N = 60) mit Stimmerfahrung ‚Singen‘ (z.B. professionelle Sänger, Gesangsunterricht im Studium)

Zu den einzelnen Gruppen zählen sowohl die mit Hochschulabschlüssen als auch die, die sich im Studium befinden. Unter die Gruppe der Singstimm-Erfahrenen fielen auch Teilnehmer, die seit vielen Jahren im Chor singen und Gesangsunterricht hatten (i.d.R. Chorerfahrung seit mehr als acht Jahren plus Einzelunterricht für zwei Jahre) und im Singen populärer Gesangsstile erfahren waren.

Nur elf Teilnehmer waren keine deutschen Muttersprachler, weshalb sie wie die anderen in die Auswertung einbezogen wurden und keine Gruppenvergleiche zwischen den Mutter- und Nicht-Muttersprachlern angestellt wurden.

6.2 Ergebnisse

6.2.1 Deskriptive Statistik der Merkmale

Die Liniendiagramme (Abbildung 5) zeigen die Häufigkeiten der Ratings pro Merkmal und Sänger auf der 6-stufigen Ratingskala. Jede Linie stellt die Häufigkeiten der Ratings für einen Sänger dar. Viele Linienverläufe zeigen Anstiege auf der einen oder anderen Seite der Ratingskala an, öfters mit einem ‚Knick‘ im Verlauf bei den extremen Ausprägungen eines Merkmals (Werte 1 und 6). Manche Merkmale zeigen glockenartige Verteilungen, wie zum Beispiel das Merkmal Stimmlage, andere klare Anstiege auf einer Seite, wie zum Beispiel das Merkmal Ausdruck. An diesen Verteilungen der Daten wird ersichtlich, dass die Daten nicht normalverteilt sind, was auch nicht zu erwarten war. Deshalb basieren die folgenden Analysen auch nicht auf dem arithmetischen Mittel, sondern Rängen. Der Mittelwert wurde höchstens zu illustrativen Zwecken verwendet. Linearen Vergleichen wurden die Rohdaten zugrunde gelegt (ersichtlich an einem hohen N), es wurde nicht über die Sänger gemittelt. Das Verfahren ist identisch mit Merrill & Larrouy-Maestri, 2017. Weitere deskriptive Statistiken zu allen bewerteten Sängern und Merkmalen finden sich im Anhang 8.4.

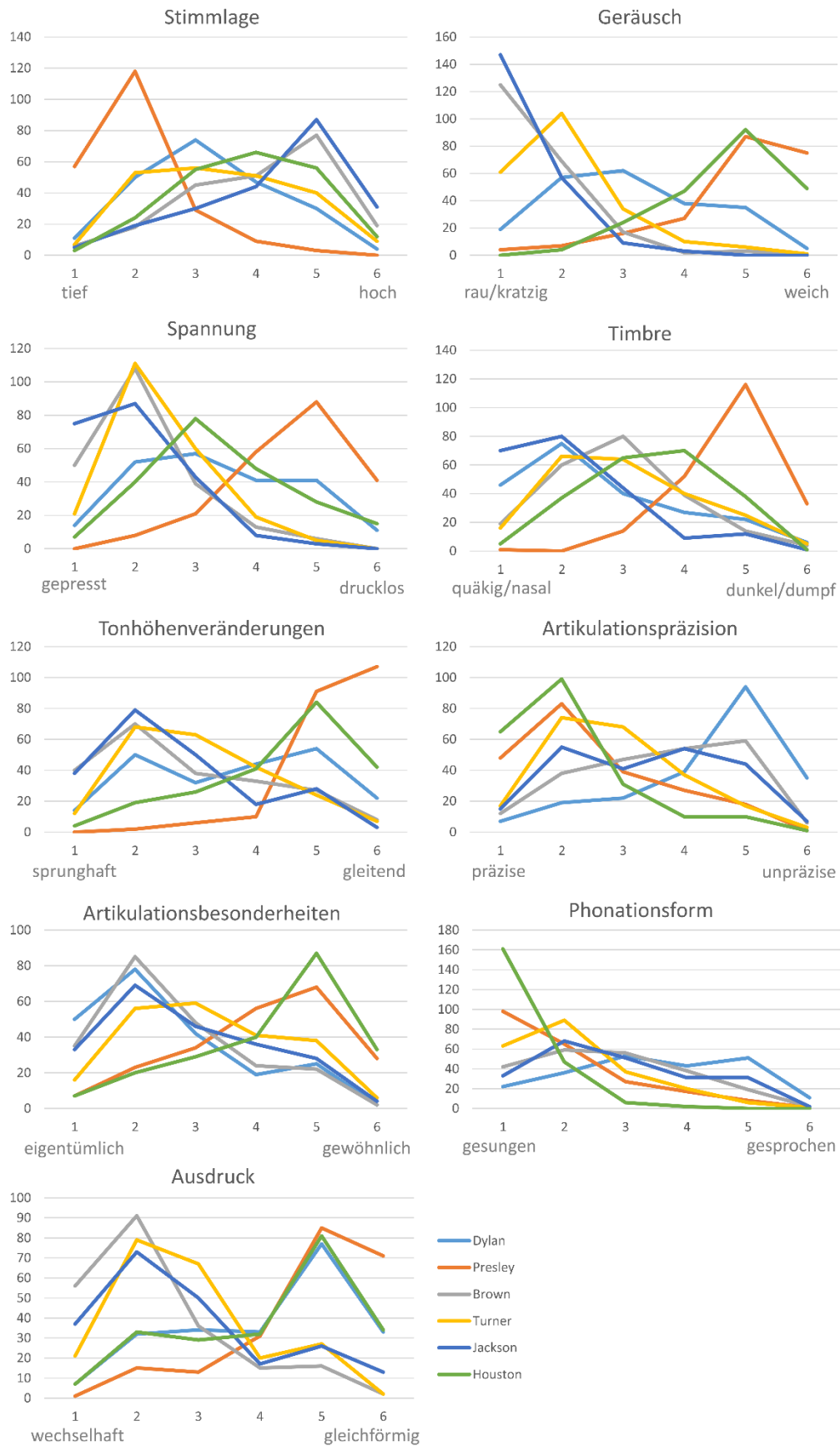


Abbildung 5: Liniendiagramme aller Merkmale, geplottet für alle Stimmen. Auf der x-Achse sind die Skalenpunkte mit den Merkmalsausprägungen abgetragen, auf der y-Achse die Häufigkeiten.

6.2.2 Expertise

Um den Einfluss der Expertise auf die Bewertung der vokalen Merkmale zu untersuchen, wurde basierend auf der Gruppeneinteilung (ohne Stimmerfahrung, mit Sprechstimmerfahrung, mit Singstimmerfahrung) eine einfaktorielle ANOVA gerechnet. Der Test wird für Artikulationsbesonderheiten signifikant ($F(2) = 4,905$, $p = 0,008$) sowie für Artikulationspräzision ($F(2) = 3,070$, $p = 0,047$) und Phonationsform ($F(2) = 3,286$, $p = 0,038$). Scheffé-Post-hoc-Tests (der die ungleichen Stichprobenumfänge der Gruppen berücksichtigt) zeigen, dass sich lediglich die Nicht-Experten signifikant von den Gesangserfahrenen (Mittlere Differenz = $-0,258$, $SD = 0,095$, 95% Konfidenzintervall = $-0,49 - -0,02$, $p = 0,026$) unterscheiden. Die beiden Expertengruppen unterscheiden sich nicht. Das Merkmal der Artikulation kann hier nicht weiter untersucht werden, da es sich bei der Beschreibung der Stimmen um deutsche Muttersprachler handelte, die englische Liedtexte beurteilten. Es kann vermutet werden, dass besonders die gesangserfahrenen Experten durch eigene Erfahrungen im Singen englischsprachiger Texte eine andere Vorstellung davon haben, was eine ausgefallene Artikulation ist. Diesem Unterschied muss in Zukunft mit deutschen Musiktiteln oder mit englischen Muttersprachlern und einem ins Englische übersetzten Inventar nachgegangen werden.

Die drei Gruppen unterscheiden sich nicht hinsichtlich der Emotions-Ratings ($p_{\min} = 0,095$, $p_{\max} = 0,750$).

An dieser Stelle ist herauszustellen, dass es im aktuellen Fragebogen nicht mehr nur um mögliche Schwierigkeiten der Nicht-Experten mit den Merkmalen geht, sondern nun auch um die der Experten: So teilte ein Experte im abschließenden Kommentar (siehe Kapitel 6.2.7) mit, dass ihm die Pole zum Timbre (quäkig/nasal-dunkel/dumpf) nicht zwei- sondern vierdimensional erscheinen. Dies erklärt sich damit, dass Experten Stimmen detaillierter beschreiben und die vier Dimensionen dieses Merkmals gesondert voneinander hätten beurteilen können, wo untrainierte Hörer eher keine Unterschiede hören. Dies war das einzige berichtete Problem in dieser Hinsicht.

6.2.3 Einheitlichkeit der Beurteilungen

Es ist ein wesentlicher Schritt zur Entwicklung des Inventars zu überprüfen, wie einheitlich die einzelnen Merkmale beurteilt werden konnten. Dies ist nicht nur davon abhängig, wie gut die Teilnehmer einen Begriff verstehen und das Merkmal hörend evaluieren können (Kreiman et al., 2007), sondern auch von der Ausprägung des Merkmals bei einem bestimmten Sänger. Die Ratings auf der vorliegenden bipolaren Bewertungsskala wurden dazu in zweierlei Hinsicht ausgewertet: Zum einen, wie sich die beobachtete Verteilung der Merkmale über die 6-stufige Skala von einer hypothetischen Verteilung unterscheidet, und zum anderen, wie stark der Median der Ratings von der Mitte der Skala abweicht. Eine Verteilung, die eine Uneinheitlichkeit der Teilnehmer abbildet, sollte also höhere Werte für die mittleren Skalenpunkte 3 und 4 erwarten, und geringere für die extremeren Skalenpunkte 2

und 5 bzw. 1 und 6. Die erwarteten/hypothetischen Häufigkeiten wurden somit einer Normalverteilung angepasst, da nicht anzunehmen ist, dass alle Kategorien mit derselben Wahrscheinlichkeit auftreten (vgl. Abbildung 6). Weicht eine Verteilung signifikant davon ab, zeigt das, dass die Ratings anders verteilt sind. Die Berechnung der Abweichung von der Mitte der Skala dient dazu, zu zeigen, ob sich die Mehrheit der Teilnehmer für eine Merkmalsausprägung entschieden hat. Dies deckt auch Verteilungen auf, die zwar signifikant verschieden sind, aber trotzdem eine Uneinheitlichkeit bei den Teilnehmern darstellen, wie zum Beispiel U-förmige Verteilungen. Beobachtete und hypothetische Häufigkeiten der Ratings wurden mit Chi-Quadrat-Tests separat für jede Stimme und jedes Merkmal verglichen ($N = 216$). Der kleinste erwartete Wert entspricht somit $N = 18$ (auf den Skalenpunkten 1 und 6). Alle folgenden Berechnungen zur Verteilung können mithilfe der Abbildung 5 nachvollzogen werden.

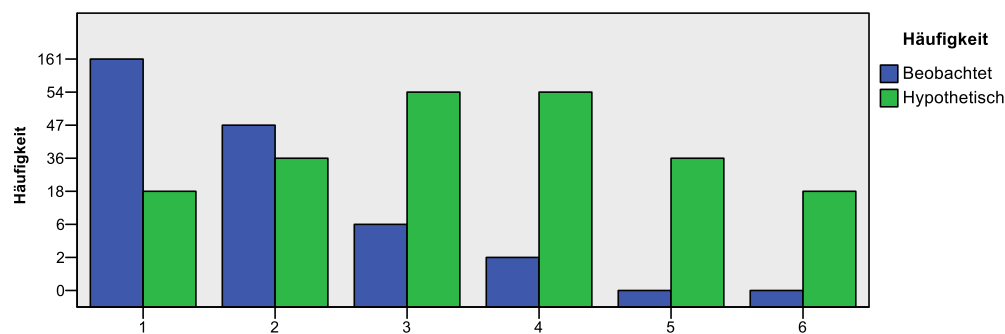


Abbildung 6: Balkendiagramm der Anzahl der Ratings (Häufigkeiten, y-Achse) für das Merkmal Phonationsform bei Houston (blau) auf der Ratingskala von 1 (gesungen) bis 6 (gesprochen; x-Achse) im Vergleich zur hypothetischen Verteilung (grün).

Zur Berechnung der Effektgröße wurde Cramers V verwendet, da es den Freiheitsgrad berücksichtigt.

Es wurde nach folgender Formel berechnet: $V = \sqrt{\frac{\chi^2}{n \cdot df}}$, wobei n die Anzahl der Beobachtungen darstellt. Es ist zu beachten, dass die Effektstärke auch Werte > 1 annehmen kann (anders als bei einem Korrelationskoeffizienten). Bei dem vorliegenden Freiheitsgrad $df = 5$ (basierend auf der 6-stufigen Skala) repräsentiert ein Cramers $V = 0,04$ einen kleinen Effekt, Cramers $V = 0,13$ einen mittleren Effekt und Cramers $V = 0,22$ einen großen Effekt (Cohen, 1988, S. 222). In Tabelle 25 sind die Ergebnisse der Chi-Quadrat-Tests und der Effektgrößen dargestellt. Von den 54 berechneten Chi-Quadrat-Tests zeigten 40 einen großen Effekt, 13 einen mittleren und einer einen kleinen Effekt.

Merkmal	χ^2 (5)	$p \leq$	Cramers V	Merkmal	χ^2 (5)	$p \leq$	Cramers V
Dylan				Turner			
Stimmelage	28,37	0,001	0,162	Stimmelage	19,935	0,001	0,136
Geräusch	27,648	0,001	0,160	Geräusch	315,481	0,001	0,540
Spannung	14,713	0,012	0,117	Spannung	224,796	0,001	0,456
Timbre	116,38	0,001	0,328	Timbre	43,454	0,001	0,201
Tonh.-ver.	27,037	0,001	0,158	Tonh.-ver.	45,333	0,001	0,205
Art.-präz.	147,38	0,001	0,369	Art.-präz.	71,676	0,001	0,258
Art.-bes.	148,824	0,001	0,371	Art.-bes.	23,037	0,001	0,146
Phon.-form	12,12	0,033	0,106	Phon.-form	258,343	0,001	0,489
Ausdruck	81,935	0,001	0,275	Ausdruck	92,87	0,001	0,293
Presley				Jackson			
Stimmelage	368,602	0,001	0,584	Stimmelage	111,574	0,001	0,321
Geräusch	327,241	0,001	0,550	Geräusch	1076,417	0,001	0,998
Spannung	164,741	0,001	0,391	Spannung	342,426	0,001	0,563
Timbre	272,037	0,001	0,502	Timbre	275,407	0,001	0,505
Tonh.-ver.	652,713	0,001	0,777	Tonh.-ver.	112,157	0,001	0,322
Art.-präz.	154,083	0,001	0,378	Art.-präz.	22,157	0,001	0,143
Art.-bes.	52,989	0,001	0,222	Art.-bes.	62,602	0,001	0,241
Phon.-form	455,602	0,001	0,650	Phon.-form	65,824	0,001	0,247
Ausdruck	291,981	0,001	0,520	Ausdruck	87,898	0,001	0,285
Brown				Houston			
Stimmelage	65,417	0,001	0,246	Stimmelage	32,296	0,001	0,173
Geräusch	789,981	0,001	0,855	Geräusch	204,519	0,001	0,435
Spannung	279,185	0,001	0,508	Spannung	20,778	0,001	0,139
Timbre	57,074	0,001	0,230	Timbre	32,565	0,001	0,174
Tonh.-ver.	79,713	0,001	0,272	Tonh.-ver.	132,565	0,001	0,350
Art.-präz.	25,713	0,001	0,154	Art.-präz.	313,454	0,001	0,539
Art.-bes.	119,75	0,001	0,333	Art.-bes.	113,787	0,001	0,325
Phon.-form	73,759	0,001	0,261	Phon.-form	1286,157	0,001	1,091
Ausdruck	223,75	0,001	0,455	Ausdruck	97,981	0,001	0,301

Tabelle 25: Ergebnisse der Chi-Quadrat-Tests pro Sänger und Merkmal. (Tonh.-ver. = Tonhöhenveränderungen; Art.-präz. = Artikulationspräzision, Art.-bes. = Artikulationsbesonderheiten.)

Neben den Verteilungen wurde die Einheitlichkeit der Beurteilungen über Einstichproben- t -Tests festgestellt. Es ist davon auszugehen, dass sich der Mittelwert der Gruppe vom mittleren Wert der Skala unterscheidet, wenn sich die Teilnehmer in der Ausprägung eines Merkmals einig sind bzw. ein Merkmal bei einem Sänger gut zu erkennen war. Mit t -Tests wurde überprüft, welche Merkmale sich signifikant vom Mittelwert der Skala (MW = 3,5) unterscheiden und sich somit mehr Teilnehmer für eine Merkmalsausprägung entschieden haben. Die Ergebnisse finden sich in Tabelle 26.

Merkmal/Sänger	Dylan	Turner	Houston	Brown	Presley	Jackson	Merkmale: Anzahl
Art.-bes.	0,371	0,146	0,325	0,333	0,222	0,241	5/1/0
Art.-präz.	0,369	0,258	0,539	0,154°	0,378	0,143°	4/2/0
Ausdruck	0,275	0,293	0,301	0,455	0,520	0,285	6/0/0
Stimmlage	0,162	0,136°	0,173	0,246	0,584	0,321	3/3/0
Geräusch	0,160	0,540	0,435	0,855	0,550	0,998	5/1/0
Phon.-form	0,106°	0,489	1,091	0,261	0,650	0,247	5/0/1
Spannung	0,117°	0,456	0,139°	0,508	0,391	0,563	4/1/1
Timbre	0,328	0,201	0,174°	0,230	0,502	0,505	4/2/0
Tonh.-veränd.	0,158°	0,205	0,350	0,272	0,777	0,322	4/2/0
Sänger: Anzahl	4/3/2	5/4/0	6/3/0	8/1/0	9/0/0	8/1/0	

Tabelle 26: Cramers V folgend der Chi-Quadrat-Tests für Sänger und Merkmale. Die Farbe zeigt die Anzahl großer/mittlerer/kleiner Effektgrößen. Die Markierungen ° zeigen einen nicht signifikanten Einstichproben t-Test.

Diese Ergebnisse werden nun pro Merkmal betrachtet:

- Ausdruck: Das Merkmal Ausdruck zeigt für alle Sänger große Effekte, d.h. die Ratings weichen von der hypothetischen Verteilung ab. Die Bewertungen weichen auch im Median signifikant von der Mitte der Ratingskala ab.
- Geräusch: Die Verteilungen von ‚rau/kratzig-weich‘ zeigen bei fünf Sängern große Effekte. Bei Dylan zeigt sich ein mittlerer Effekt. Als ‚weich‘ werden Presley und Houston bewertet, die anderen als eher ‚rau‘.
- Phonationsform: Die Verteilungen von ‚gesungen-gesprochen‘ zeigen nur bei Dylan einen kleinen Effekt, alle anderen zeigen große Effekte. Das heißt, alle Sänger bis auf Dylan singen, bei Dylan sind sich die Teilnehmer uneinig, was sich auch im t-Test zeigt.
- Spannung: Für das Merkmal ‚gepresst-drucklos‘ wurden bei vier Sängern große Effekte gefunden, bei Houston ein mittlerer und bei Dylan ein kleiner. Während Jackson, Brown und Turner als ‚gepresst‘ eingestuft wurden, wurde Presley als ‚drucklos‘ bewertet. Bei Dylan und Houston wurde folgend dem t-Test keine Einigkeit erzielt. Interessant ist, dass sich bei Houston viele Ratings auf den Skalenpunkten 3 und 4 finden, was widerspiegelt, dass sie weder drucklos noch gepresst singt, was als eine angemessene Spannung interpretiert werden könnte.
- Artikulationspräzision: Die Artikulationspräzision zeigte vier große und zwei mittlere Effekte, wonach Houston, Presley und Turner eine präzise Aussprache haben und Dylan eine unpräzise. Bei Brown und Jackson wurde keine Einigkeit erzielt.
- Tonhöhenveränderungen: Die Verteilungen für die Tonhöhenveränderungen zeigten große Effekte im Falle von vier Stimmen und mittlere bei zwei Stimmen, wobei bei Dylan keine Einigkeit bei den Teilnehmern vorlag. Tonhöhenveränderungen wurden bei Presley und Houston als ‚gleitend‘ beschrieben, bei Jackson, Brown und Turner als ‚sprunghaft‘.
- Artikulationsbesonderheiten: Das Artikulationsmerkmal ‚eigentümlich-gewöhnlich‘ zeigte bei fünf Sängern große und bei einem Sänger mittlere Effekte. Dylan, Brown, Jackson und Turner

wird eher eine eigentümliche Aussprache zugeschrieben, Presley und Houston eine gewöhnliche.

- **Timbre:** Vier Sänger zeigten einen großen Effekt für das Merkmal ‚quäkig/nasaldunkel/dumpf‘, zwei einen mittleren, wobei sich die Teilnehmer bei Houston uneinig waren (siehe *t*-Test). Jackson, Dylan, Brown und Turner wurden mit ‚quäkig/nasal‘ beschrieben und Presley mit ‚dunkel/dumpf‘.
- **Stimmhöhe:** Für das Merkmal ‚tief-hoch‘ zeigten sich drei große und drei mittlere Effekte. Dylan und Houston ließen sich nicht gut bestimmen, obwohl sich die Teilnehmer signifikant in eine Richtung des Merkmals entschieden (beide tendenziell ‚tief‘). Bei Turner allerdings waren sich die Teilnehmer uneinig (siehe *t*-Test). Presleys Stimmhöhe wurde als ‚tief‘, Browns und Jacksons als ‚hoch‘ eingeschätzt. Die Lage ist wegen der relativ großen Tonumfänge schwierig zu beurteilen, die beim Singen bzw. in den Stücken verwendet werden. Diese Entscheidung wurde wahrscheinlich auch durch das Register beeinflusst, in dem gesungen wird, was die Wahrnehmung einer hohen oder tiefen Lage beeinflusst.

Zusammengenommen zeigten sich bei der Bewertung der Merkmale keine systematischen Auffälligkeiten. Manche Merkmale konnten bei manchen Sängern gut, bei anderen weniger gut beschrieben werden. Das bedeutet, dass die Bewertung weniger vom Verständnis der Merkmale als von der Merkmalsausprägung eines Sängers abhängt. Dylan, zum Beispiel, mischt mehr singende und gesprochene Anteile als Houston, was sich in den Ratings niederschlägt.

6.2.4 Klassifikation der Sänger

6.2.4.1 *Sänger und Merkmale*

Mit einer Diskriminanzanalyse wurden die sechs Sänger auf Basis der bewerteten Merkmale klassifiziert. Die Ratings der Teilnehmer pro Sänger und Merkmal wurden den Berechnungen zugrunde gelegt, was heißt, dass die Analysen nicht auf Mittelwerten basieren, sondern auf den Rohdaten ($N = 1296$). Es zeigten sich, welche Merkmale in welcher Kombination am besten zur Unterscheidung der Sänger beitragen (Funktionen). Eine multiple Diskriminanzanalyse (MDA) wurde über alle Stimmen zusammen durchgeführt. Chi-Quadrat-Tests mit Wilks' Lambda zeigten fünf Funktionen, die einen statistisch signifikanten Einfluss auf die Diskrimination haben (Funktionen 1 bis 4, $p < 0,001$, Funktion 5, $p < 0,01$). Nur die erste Funktion hat einen Eigenwert > 1 (2,994).

- Die erste Funktion erklärt 75,3 % der Varianz und darauf laden die Merkmale Geräusch und Spannung.
- Die zweite Funktion erklärt 17,1 % der Varianz und darauf laden Artikulationspräzision, Phonationsform und Artikulationsbesonderheiten.
- Die dritte Funktion erklärt 5,9 % und darauf laden Timbre und Stimmlage.
- Die vierte Funktion erklärt 1,3 % und darauf lädt das Merkmal Ausdruck.
- Die fünfte Funktion erklärt 0,4 % und darauf lädt das Merkmal Tonhöhenveränderungen.

62,7 % der kreuzvalidierten gruppierten Fälle wurden korrekt klassifiziert (64,0 % ohne Kreuzvalidierung). Dylan wurde in 67,1 % der Fälle (kreuzvalidiert) korrekt klassifiziert, Presley in 81,5 %, Brown in 47,4 %, Turner in 54,2 %, Jackson in 52,8 % und Houston in 73,1 %. Brown wurde in 24,1 % der Fälle als Jackson klassifiziert, Jackson in 25,5 % der Fälle als Brown. Turner wurde auch in 18,1 % der Fälle mit Brown und Jackson vertauscht.

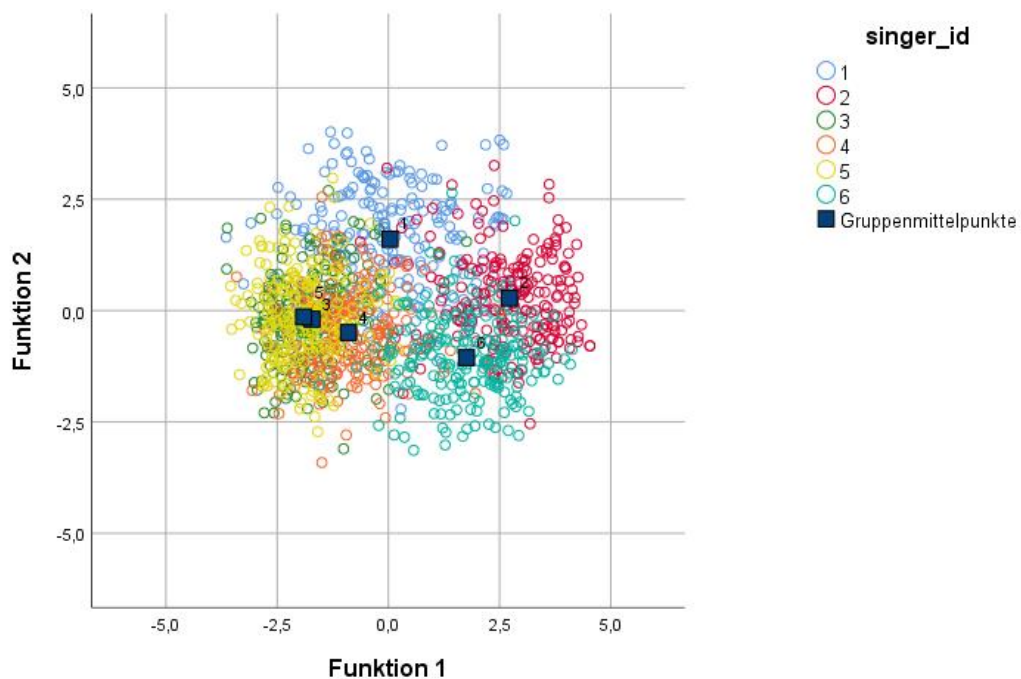


Abbildung 7: Streudiagramm zum Ergebnis der MDA mit Sängern (*singer_id*: 1 = Dylan, 2 = Presley, 3 = Brown, 4 = Turner, 5 = Jackson, 6 = Houston) und Merkmalen. Auf der x-Achse wird nach den Merkmalen der 1. Funktion der MDA unterschieden (Geräusch und Spannung), auf der y-Achse nach denen der 2. Funktion (Artikulation und Phonationsform).

Auf Basis der ersten Funktion lassen sich die Stimmen sehr gut in den Merkmalen Geräusch (rau/kratzig-weich) und Spannung (gepresst-drucklos) unterscheiden. Das Streudiagramm (Abbildung 7) zeigt, dass die Pole von Brown und Jackson in Richtung ‚rau‘ / ‚gepresst‘ und von Presley in Richtung ‚weich‘ / ‚drucklos‘ besetzt sind. Folgend Funktion 2 lassen sich die Stimmen hinsichtlich der Artikulation und der Phonationsform trennen, wo die Pole durch Dylan und Houston bestimmt werden. Allerdings ist hier die Distanz nicht so groß wie in Funktion 1 (auch reflektiert in der geringeren Varianzaufklärung).

Um die Unterschiede zwischen den Sängern auf Basis der Merkmale weiter zu untersuchen, wurde ein allgemeines lineares Modell mit Messwiederholung (N = 216) gerechnet. Die Tests der Innersubjekt-effekte zeigten signifikante Haupteffekte für Sänger und Merkmale sowie eine signifikante Interaktion für Sänger × Merkmale, d.h. die Sänger unterscheiden sich auf Basis der Merkmale.

- ME Sänger $F(1) = 338,964, p < 0,001$
- ME Merkmale $F(1) = 79,851, p < 0,001$
- IA Sänger × Merkmale $F(1) = 152,722, p < 0,001$

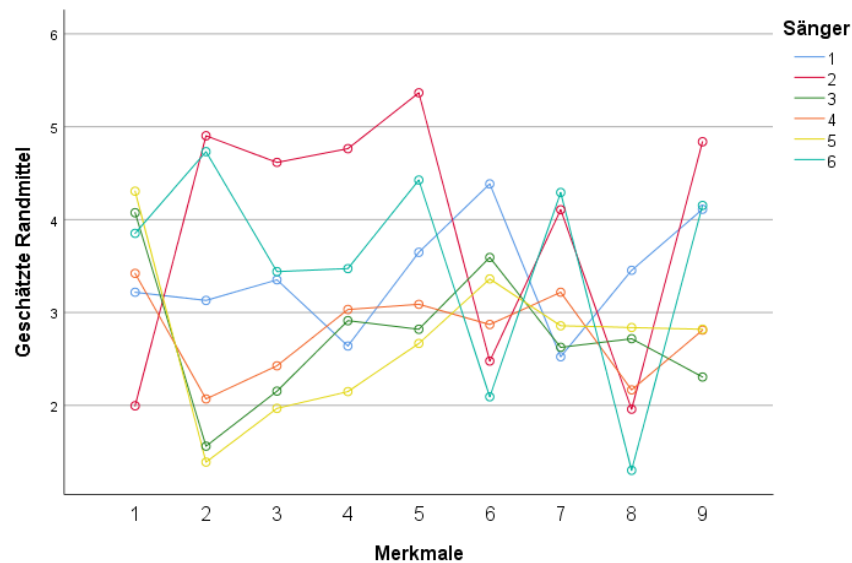


Abbildung 8: Liniendiagramm der geschätzten Randmittel der sechs Sänger (y-Achse) für jedes Merkmal (x-Achse). Sänger: 1 = Dylan, 2 = Presley, 3 = Brown, 4 = Turner, 5 = Jackson, 6 = Houston. Merkmale: 1 = Stimmlage, 2 = Geräusch, 3 = Spannung, 4 = Timbre, 5 = Tonhöhenveränderungen, 6 = Artikulationspräzision, 7 = Artikulationsbesonderheiten, 8 = Phonationsform, 9 = Ausdruck.

Das Liniendiagramm (Abbildung 8) zeigt die Unterschiede zwischen einzelnen Sängern. Übereinstimmungen mit der MDA zeigen sich insofern, als dass sich Brown, Jackson und Turner (grün, gelb, orange) untereinander recht ähnlich sind, genauso wie Presley und Houston (rot, türkis), Dylan (blau) weist wenig Ähnlichkeiten mit anderen Sängern auf. Breite Streuungen finden sich in den Merkmalen Geräusch (2), Spannung (3) und Timbre (4) sowie Tonhöhenveränderungen (5) und Ausdruck (9).

Anhand der Funktionen der MDA lassen sich nicht nur Ähnlichkeiten zwischen den Sängern, sondern auch Gemeinsamkeiten zwischen den Merkmalen erkennen (vergleichbar einer Hauptkomponentenanalyse). Dass sich Geräusch und Spannung (Funktion 1) ähnlich sind, zeigt sich z.B. an Jackson und Brown, die sowohl mit einer rauhen/kratzigen Stimme als auch mit hohem Druck (gepresster Stimmklang) singen. Die Ähnlichkeit in den Artikulationsmerkmalen (Artikulationsbesonderheiten und -präzision; Funktion 2) ist konzeptuell nicht verwunderlich und gruppiert sich zusammen mit der

Phonationsform, was andeuten kann, dass Singen und Sprechen unter Berücksichtigung artikulatorischer Merkmale bewertet wurden.

Timbre (quäkig/nasal-dunkel/dumpf) und Stimmlage (tief-hoch) laden zusammen auf die dritte Funktion, was zeigt, dass eine hohe Stimmlage bei den ausgewählten Stimmen mit einem quäkig/nasalen Stimmklang einhergeht (z.B. Jackson; Merkmale 1 und 4, Abbildung 8) und eine tiefe Stimmlage mit einem dunkel/dumpfen Stimmklang (z.B. Presley). Eine Mischung findet sich bei Houston, die in höheren Tönen eher zum quäkigen und in tieferen zum dunklen Klang neigt, was sich in vermehrt mittleren Ratings zeigt. Auf die vierte Funktion lädt nur das Merkmal Ausdruck, auf die fünfte Funktion nur die Tonhöhenveränderungen, wobei beide nicht mehr stark der Diskriminanz der Stimmen dienen. Die Zusammenhänge innerhalb der Merkmale werden im Folgenden weiter mit Korrelationen nach Spearman betrachtet, die die Beziehungen zwischen den vokalen Merkmalen ohne die Trennung nach Sänger aufzeigen. Es wurden alle Korrelationen signifikant, weshalb nur die $r > 0,4$ dargestellt werden:

- Ein rau/kratziger Stimmklang korreliert mit gepresst, bzw. ein weicher mit drucklos ($r = 0,641$) und einem wechselhaften Ausdruck ($r = 0,455$).
- Timbre korreliert mit Lage ($r = -0,497$), Geräusch ($r = 0,407$), Spannung ($r = 0,445$), d.h. dunkel korreliert mit tief, weich, drucklos, und quäkig mit hoch, rau/kratzig, gepresst.
- Sprunghafte Tonhöhenveränderungen korrelieren mit rau/kratzig ($r = 0,520$), gepresst ($r = 0,492$) und einem wechselhaften Ausdruck ($r = 0,476$).

Die Korrelationen unterstreichen die Ergebnisse der MDA, vor allem hinsichtlich ‚rau/kratzig-weich‘ und ‚gepresst-drucklos‘, die den stärksten Zusammenhang zeigen – allerdings stehen sie nur in einem moderaten Zusammenhang, was sich auch bei den anderen Merkmalskorrelationen zeigt. Das bedeutet, dass sich die Merkmale ergänzen und zu einem umfassenden vokalen Profil eines Sängers beitragen.

6.2.4.2 Sänger und Emotionen

Eine MDA wurde, wie für die Sänger und Merkmale, auch für die Sänger und Emotionen durchgeführt. Chi-Quadrat-Tests mit Wilks' Lambda zeigten drei Funktionen, die einen statistisch signifikanten Einfluss auf die Diskriminanz haben (Funktionen 1 und 2, $p < 0,001$, Funktion 3, $p < 0,01$).

- Die erste Funktion erklärt 94,4 % der Varianz, worauf keine Merkmale laden.
- Die zweite Funktion erklärt 3,9 % der Varianz und darauf laden die Merkmale berührt, energetisiert und fasziniert.
- Die dritte Funktion erklärt 1,4 % und darauf lädt das Merkmal entspannt.

41,7 % der kreuzvalidierten Fälle wurden korrekt klassifiziert (43,8 % der ursprünglich gruppierten Fälle). Dylan wurde in 30,1 % der (kreuzvalidierten) Fälle korrekt klassifiziert (in 21,3 % der Fälle

verwechselt mit Presley und in 19,4 % mit Houston), Presley in 60,6 %, Brown in 40,3 % (in 30,6 % der Fälle verwechselt mit Jackson), Turner in 34,3 % (oft verwechselt mit Brown, 19,9 %, und Houston, 22,2 %), Jackson in 53,2 % (31,0 % mit Brown) und Houston in 31,5 % (24,5 % verwechselt mit Turner).

Auf Basis der Emotionen lassen sich die Sänger wesentlich schlechter diskriminieren als auf Basis der Merkmale, obwohl Ähnlichkeiten zwischen beiden MDAs gesehen werden können, da auch hier die Sänger Presley (ID 2) und Jackson (ID 5) die Pole der Funktion 1 bilden (Abbildung 9). Da keine Merkmale auf die erste Funktion laden, die 94,4 % der Datenvarianz erklärt, können die Sänger kaum mithilfe der Emotionen voneinander getrennt werden. Nur mit 3,9 % erklärt die zweite Funktion die Klassifikation, wobei ‚berührt‘, ‚energetisiert‘ und ‚fasziniert‘ wichtig sind.

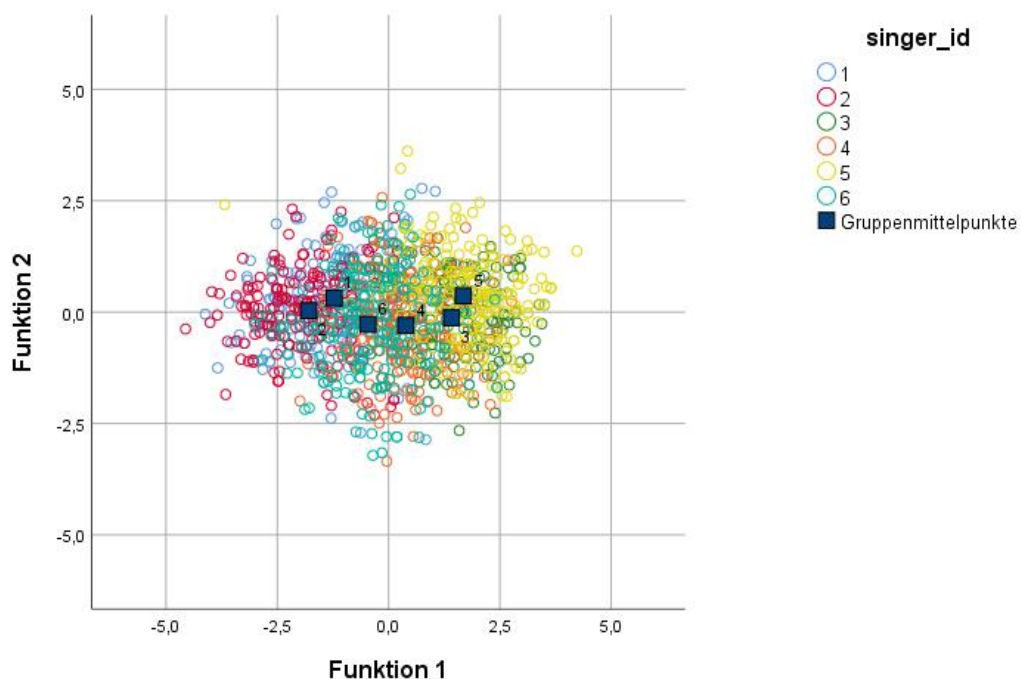


Abbildung 9: Streudiagramm der MDA für die Sänger, klassifiziert nach den Emotionsratings. Auf die 1. Funktion laden keine Emotionen, auf die 2. die Merkmale berührt, energetisiert und fasziniert. Singer_id: 1 = Dylan, 2 = Presley, 3 = Brown, 4 = Turner, 5 = Jackson, 6 = Houston.

Um zu sehen, welche Sänger sich in welchen Emotionsratings unterscheiden, wurde ein allgemeines lineares Modell mit Messwiederholung (N = 216) für Sänger und Emotionen berechnet. Die Tests der Innersubjekteffekte zeigten signifikante Haupteffekte für Sänger und Emotion sowie für die Interaktion Sänger × Emotion.

Selbst mit einem Freiheitsgrad von 1 sind alle Effekte signifikant:

- ME Sänger $F(1) = 7,054, p = 0,009$
- ME Emotionen $F(1) = 198,156, p < 0,001$
- IA Sänger \times Emotionen $F(1) = 99,032, p < 0,001$

Wichtig ist das Ergebnis der signifikanten Interaktion, die zeigt, dass sich die Sänger hinsichtlich der ausgelösten Emotionen unterscheiden, allerdings nur bestimmte Sänger in Bezug auf bestimmte Emotionen. Die MDA zeigte bereits, dass sich keine Emotionen finden, auf deren Basis alle Sänger diskriminiert werden können.

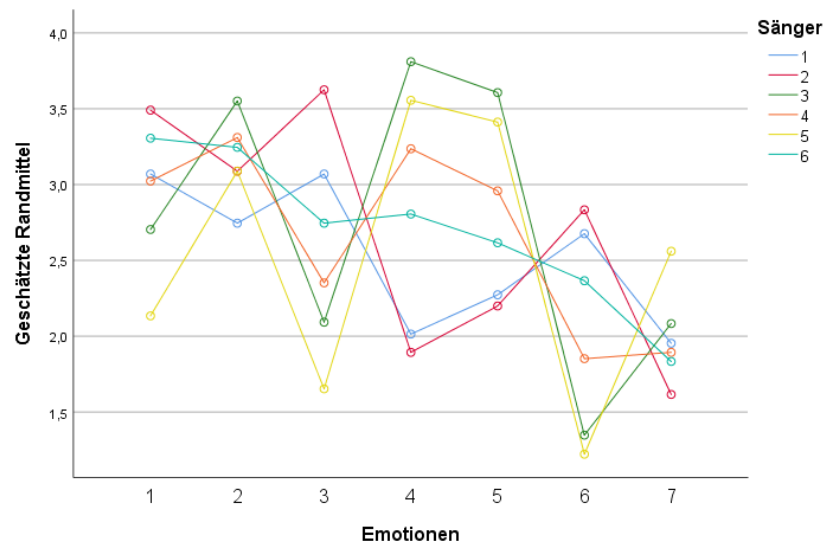


Abbildung 10: Liniendiagramm der geschätzten Randmittel der sechs Sänger (y-Achse) für jede Emotion (x-Achse). Sänger: 1 = Dylan, 2 = Presley, 3 = Brown, 4 = Turner, 5 = Jackson, 6 = Houston. Emotionen: 1 = berührt, 2 = fasziniert, 3 = entspannt, 4 = energetisiert, 5 = fröhlich, 6 = traurig, 7 = gereizt.

Abbildung 10 spiegelt die Ergebnisse der MDA insofern wider, als dass Brown, Jackson und Turner (grün, gelb, orange) Presley und Dylan (rot, blau) gegenüberstehen, während Houston (türkis) in der Mitte verläuft. Besonders weite Streuungen, bzw. Unterschiede zwischen den Sängern, finden sich für die Merkmale ‚entspannt‘ (3), ‚energetisiert‘ (4) und ‚fröhlich‘ (5).

Die detaillierten Unterschiede zwischen den Sängern bezogen auf einzelne Merkmale und Emotionen sind für die vorliegende Fragestellung nicht von Interesse, weshalb keine paarweisen Vergleiche berechnet wurden.

6.2.5 Zusammenhänge von Merkmalen und Emotionen

Um die Beziehungen zwischen den Merkmalen und den Emotionen zu untersuchen, wurden Korrelationen nach Spearman durchgeführt. Dies sollte Aufschluss darüber geben, ob bestimmte stimmliche Eigenschaften mit bestimmten Emotionen zusammenhängen. Hierbei ist zu beachten, dass die ausgelösten Emotionen nicht nur aufgrund der Stimme zustande kommen, sondern auch in

Zusammenhang mit dem Song stehen. Dieser Zusammenhang wird in Kapitel 6.2.6.2 in Verbindung mit dem Gefallen näher untersucht. Da alle Korrelationen signifikant wurden, werden hier nur die $r > 0,4$ dargestellt:

- Entspannung korreliert mit einem weichen ($r = 0,450$), drucklosen ($r = 0,510$) sowie dunkel/dumpfem Stimmklang ($r = 0,408$) und gleitenden Tonhöhenveränderungen ($r = 0,404$).
- Energetisierend korreliert mit einem rau/kratzigen Stimmklang ($r = -0,400$).
- Traurig korreliert mit einem weichen Stimmklang ($r = 0,437$).

Interessant ist an dieser Stelle anzumerken, dass das Emotions-Merkmal der Entspannung mit dem vokalen Merkmal der Spannung (gepresst-drucklos) in Zusammenhang steht, es scheint, als wirke ein druckloser Stimmklang entspannend.

6.2.6 Gefallen

6.2.6.1 Zusammenhänge von Merkmalen mit Gefallen

Lineare Regressionsmodelle (Methode: Einschluss) wurden über alle Sänger und Merkmale ($N = 1295$) sowie für jeden Sänger einzeln ($N = 216$) mit dem Gefallen der Stimme als unabhängige Variable und den neun Merkmalen als Prädiktorvariablen gerechnet (vgl. Merrill & Larrouy-Maestri, 2017 und Kapitel 5.2.1.1).

Für das Gefallen der Stimme für alle Sänger fand sich ein signifikantes Modell ($F(9, 1286) = 29,215$, $p < 0,001$), welches 17 % der Varianz in den Daten erklärt. Die folgenden Merkmale waren signifikante Prädiktoren des Gefallens der Stimme:

- Timbre (Beta = $-0,188$, SEM = $0,029$, $p < 0,001$),
- Tonhöhenveränderungen (Beta = $-0,081$, SEM = $0,025$, $p = 0,001$),
- Artikulationspräzision (Beta = $0,117$, SEM = $0,024$, $p < 0,001$),
- Artikulationsbesonderheiten (Beta = $0,097$, SEM = $0,024$, $p < 0,001$),
- Phonationsform (Beta = $0,104$, SEM = $0,025$, $p < 0,001$),
- Ausdruck (Beta = $0,137$, SEM = $0,023$, $p < 0,001$).

Es ist auffällig, dass nur ein Klangmerkmal signifikant wird (Timbre). Korrelationen nach Spearman sollten die Richtungen der Merkmalsausprägung, besonders die der niedrigen Beta-Werte, erklären, d.h. welche Merkmalsausprägung mit dem Gefallen assoziiert ist: Timbre ($r = -0,297$), Tonhöhenveränderungen ($r = -0,196$), Artikulationspräzision ($r = 0,216$), Artikulationsbesonderheiten (n.s., $r = -0,022$), Phonationsform ($r = 0,215$), Ausdruck (n.s., $r = 0,026$). Daraus lässt sich schließen, dass eine Stimme dann gefällt, wenn sie folgende Merkmalsausprägungen hat: dunkel/dumpfer Klang, gleitende Tonhöhenveränderungen, präzise Artikulation und eher eine singende Phonationsform. Außerdem können sowohl eine eigentümliche als auch eine gewöhnliche Artikulation gefallen sowie

ein wechselhafter als auch ein gleichförmiger Ausdruck. Die Relationen werden in Abbildung 11 dargestellt, in der auch die schwachen Zusammenhänge und die geringe Erklärung der Datenvarianz ersichtlich werden.

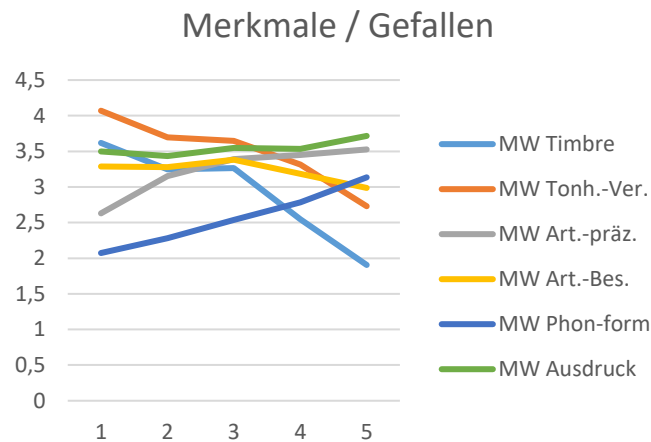


Abbildung 11: Liniendiagramm der Mittelwerte für die signifikanten Merkmale (y-Achse) der Regressionsanalyse und Gefallen (x-Achse). Gefällt = 1, gefällt nicht = 5; MW = Mittelwert.

Die Auswertung der einzelnen Sänger (siehe Tabelle 27) entsprach der allgemeinen Auswertung, d.h. es fanden sich kaum weitere Merkmale, die bei einzelnen Sängern gefallen. Nur das Merkmal Spannung trat zweimal auf, bei Turner und bei Houston, wobei in beiden Fällen ein gepresster Stimmklang nicht gefällt. Allerdings sagen nicht alle Merkmale auch bei allen Sängern das Gefallen voraus, bei Presley sagte auch gar kein Merkmal das Gefallen voraus. Timbre tritt zum Beispiel bei drei Sängern auf, andere bei ein bis zwei Sängern, womit sich gewisse Übereinstimmungen zeigen.

Sänger	$F(9, 206)$	$p \leq$	R^2	Merkmal	Beta	SEM	$p \leq$
Dylan	4,983	0,001	0,179	Timbre	-0,245	0,071	0,001
				Art.-präz.	0,189	0,065	0,004
Presley	1,836	0,064	0,074	n.s.			
Brown	2,300	0,018	0,091	Timbre	-0,161	0,082	0,050
				Art.-präz.	0,175	0,062	0,006
Turner	4,058	0,001	0,151	Spannung	-0,204	0,091	0,026
				Ausdruck	0,123	0,060	0,040
Jackson	5,586	0,001	0,196	Timbre	-0,252	0,082	0,002
				Tonh.-veränd.	-0,281	0,066	0,001
				Ausdruck	0,181	0,060	0,003
Houston	5,368	0,001	0,190	Spannung	-0,134	0,064	0,039
				Art.-bes.	0,188	0,055	0,001
				Phon.-form	0,308	0,130	0,018

Tabelle 27: Ergebnisse der linearen Regressionsmodelle von Gefallen und Merkmale für jeden Sänger.

6.2.6.2 Zusammenhänge von Emotionen und Gefallen Stimme und Song

Auf dieselbe Weise wie bei den Merkmalen sollte mithilfe von linearen Regressionsanalysen die Relationen vom Gefallen der Stimme und Gefallen des Songs mit den sieben Emotionen untersucht werden. Sowohl die Regressionen über alle Sänger als auch für jeden Sänger einzeln ergaben signifikante Modelle, die auch recht hohe Werte für die Varianzaufklärung zeigten (Tabelle 28).

Sänger	<i>F</i> (7, 208)	<i>p</i> <	<i>R</i> ²
Gefallen Stimme			
Dylan	47,904	0,001	0,617
Presley	30,096	0,001	0,503
Brown	38,576	0,001	0,565
Turner	43,121	0,001	0,592
Jackson	35,491	0,001	0,544
Houston	57,528	0,001	0,659
Gefallen Song			
Dylan	48,313	0,001	0,619
Presley	45,739	0,001	0,606
Brown	36,726	0,001	0,533
Turner	37,309	0,001	0,557
Jackson	34,419	0,001	0,537
Houston	58,351	0,001	0,663

Tabelle 28: Ergebnisse der Regressionsanalysen von Sänger und Emotionen für Gefallen Stimme und Gefallen Song.

Gefallen Stimme:

- Signifikantes Modell $F(7, 1288) = 260,824$, $p < 0,001$; 58,6 % erklärte Varianz der Gefallens-Ratings. Alle Emotionen bis auf Traurigkeit sagen das Gefallen voraus (siehe Tabelle 29, Spalte ‚Alle Sänger‘).
- Korrelationen nach Spearman mit Gefallen Stimme werden alle signifikant: berührt ($r = -0,627$), fasziniert ($r = -0,595$), entspannt ($r = -0,449$), energetisiert ($r = -0,316$), fröhlich ($r = -0,416$), traurig ($r = -0,143$), gereizt ($r = 0,532$).
- Somit gefällt eine Stimme, wenn sie berührt, fasziniert, energetisiert, nicht gereizt macht, entspannt und fröhlich stimmt. Das Merkmal traurig ist kein signifikanter Regressor, d.h. dass Traurigkeit, die aus einer der vorliegenden Stimmen resultiert, nicht das Gefallen vorhersagt.

Gefallen Song:

- Signifikantes Modell $F(7, 1288) = 245,469$, $p < 0,001$; 57,2 % erklärte Varianz. Alle Emotionen werden signifikante Prädiktoren des Gefallens (siehe Tabelle 29).
- Korrelationen nach Spearman mit Gefallen Song: berührt ($r = -0,542$), fasziniert ($r = -0,632$), entspannt ($r = -0,320$), energetisiert ($r = -0,448$), fröhlich ($r = -0,529$), traurig ($r = 0,017$, n.s.), gereizt ($r = 0,461$).
- Somit gefällt ein Song, wenn er berührt, fasziniert, energetisiert, entspannt und nicht gereizt macht. Traurigkeit korreliert nicht signifikant.

Merkmal	Stimme			Song		
	Beta	SEM	$p \leq$	Beta	SEM	$p \leq$
Alle Sänger						
berührt	-0,263	0,027	0,001	-0,211	0,029	0,001
fasziniert	-0,234	0,026	0,001	-0,296	0,028	0,001
entspannt	-0,131	0,024	0,001	-0,096	0,026	0,001
energetisiert	-0,056	0,024	0,022	-0,120	0,026	0,001
fröhlich	-0,128	0,027	0,001	-0,189	0,029	0,001
traurig	-0,004	0,025	0,875	0,083	0,027	0,002
gereizt	0,301	0,022	0,001	-0,255	0,024	0,001
Dylan						
berührt	-0,254	0,085	0,003	-0,411	0,088	0,001
fasziniert	-0,208	0,072	0,004	-0,261	0,075	0,001
entspannt	-0,142	0,068	0,040			n.s.
gereizt	0,354	0,063	0,001	0,267	0,065	0,001
Presley						
berührt	-0,318	0,071	0,001	-0,440	0,072	0,001
fasziniert	-0,185	0,060	0,002	-0,289	0,062	0,001
entspannt	-0,165	0,062	0,008	-0,144	0,063	0,024
traurig			n.s.	0,176	0,050	0,001
gereizt	0,181	0,054	0,001	0,232	0,055	0,001
Brown						
fasziniert	-0,201	0,064	0,002	-0,180	0,063	0,005
fröhlich	-0,469	0,062	0,001	-0,395	0,076	0,001
gereizt	0,209	0,056	0,001	0,254	0,055	0,001
Turner						
berührt	-0,245	0,059	0,001	-0,230	0,071	0,001
fasziniert	-0,222	0,060	0,001	-0,254	0,073	0,001
energetisiert			n.s.	-0,171	0,068	0,012
fröhlich	-0,154	0,062	0,013	-0,158	0,075	0,037
gereizt	0,289	0,053	0,001	0,273	0,064	0,001
Jackson						
berührt	-0,304	0,072	0,001			n.s.
fasziniert	-0,221	0,066	0,001	-0,323	0,065	0,001
energetisiert			n.s.	-0,150	0,070	0,034
fröhlich			n.s.	-0,189	0,080	0,019
gereizt	0,339	0,058	0,001	0,236	0,057	0,001
Houston						
berührt			n.s.	-0,225	0,078	0,004
fasziniert	-0,346	0,064	0,001	-0,310	0,076	0,001
energetisiert			n.s.	-0,127	0,060	0,035
fröhlich	-0,120	0,058	0,039	-0,264	0,069	0,001
gereizt	0,294	0,050	0,001	0,310	0,059	0,001

Tabelle 29: Prädiktoren der signifikanten Regressionsmodelle für Gefallen Stimme/Song.

Die Ratings des Gefallens einer Stimme sind in Abbildung 12 gegen die Emotionen geplottet. So wird deutlich, wie ähnlich die Emotionen zwischen den Sängern bewertet werden (ein Ergebnis, das bereits in der MDA ersichtlich wurde). Auch zeigt sich der schwache Zusammenhang der Ratings der Emotion Traurigkeit mit dem Gefallen.

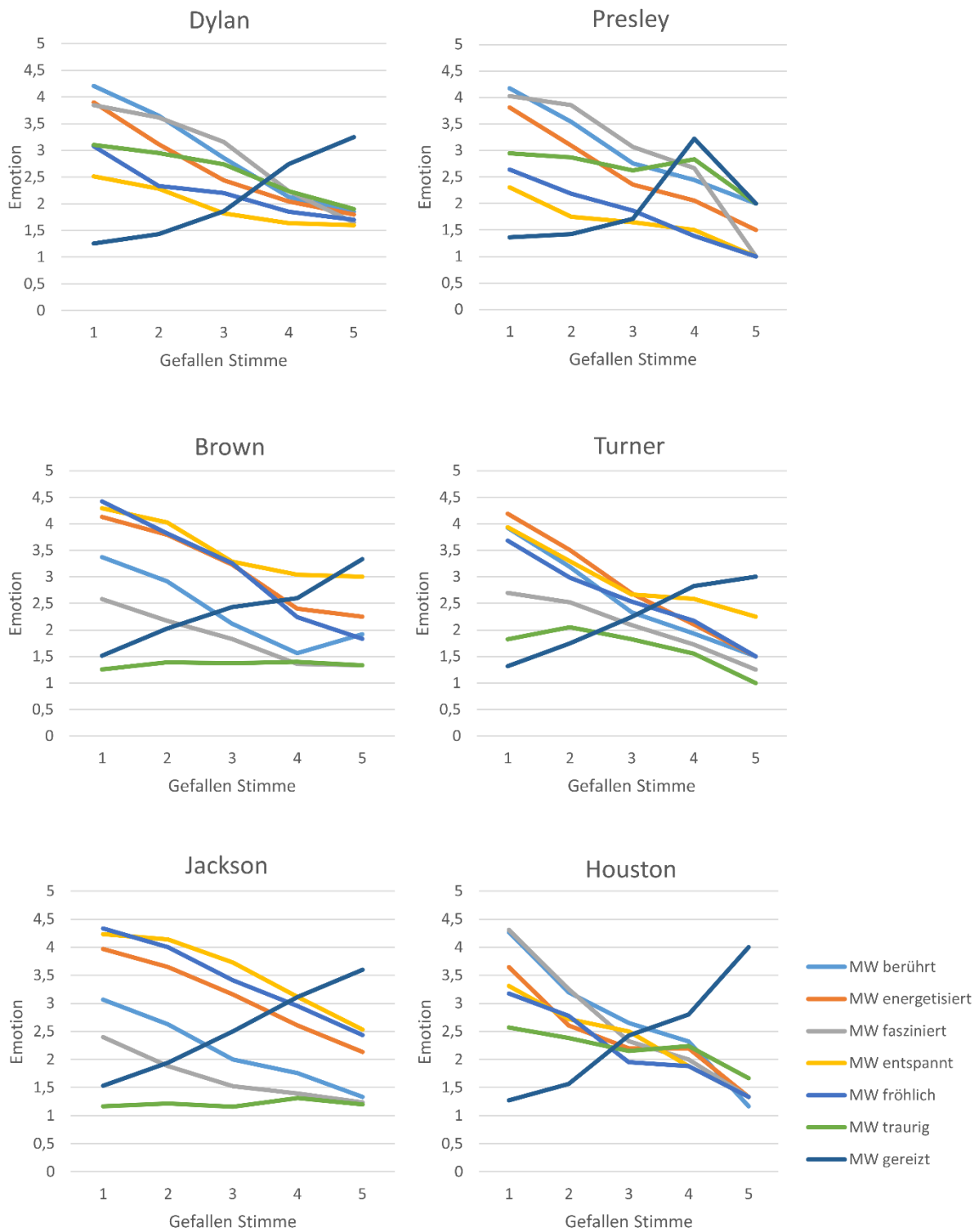


Abbildung 12: Liniendiagramm. Gefallen der Stimme jedes Sängers geplottet gegen die Mittelwerte der Emotionsratings. Gefällt = 1, gefällt nicht = 5. MW = Mittelwert.

In einem Vergleich der Emotionsratings in Verbindung zur Stimme und dem Song (Tabelle 30) wird deutlich, dass sich keine Emotion nur auf die Stimme bezieht. Allerdings scheint die Emotion ‚energetisiert‘ spezifisch für Songs zu sein (es wird nicht von einer Stimme evoziert). Für alle Sänger/Songs finden sich die Emotionen ‚fasziniert‘ und ‚nicht gereizt‘, bei fast allen das Merkmal ‚berührt‘. Auch stimmen vier von sechs Songs bzw. drei von sechs Stimmen fröhlich.

	berührt	fasziniert	energetisiert	entspannt	macht fröhlich	macht nicht traurig	macht nicht gereizt
Dylan	+ #	+ #		#			+ #
Presley	+ #	+ #		+		#	+ #
Brown		+ #			+ #		+ #
Turner	+ #	+ #	#		+ #		+ #
Jackson	+	+ #	#		#		+ #
Houston	#	+ #	#		+ #		+ #

Tabelle 30: Auflistung der Emotionen, die für jeden Sänger signifikante Prädiktoren für Gefallen Stimme (+) und Gefallen Song (#) sind.

Eine Korrelation folgend Spearman spiegelt das Ergebnis wider, dass Gefallen Stimme und Gefallen Song korrelieren ($r = 0,711$), was auch in der folgenden Abbildung 13 ersichtlich wird. Obwohl sich kaum Unterschiede in den ausgelösten Emotionen von Stimme und Song fanden, zeigt die Abbildung, dass (bis auf Dylan) Unterschiede zwischen Gefallen Stimme und Song innerhalb der Sänger bestehen. Das bedeutet, dass die Teilnehmer zwischen den Gefallens-Ratings unterschieden haben, was bei der Gruppen-Studie nicht der Fall war, d.h. die Trennung der Fragen an Anfang und Ende war sinnvoll. Die Unterschiede hielten sich allerdings in Grenzen, kein Sänger wurde im Mittel nicht gemocht.

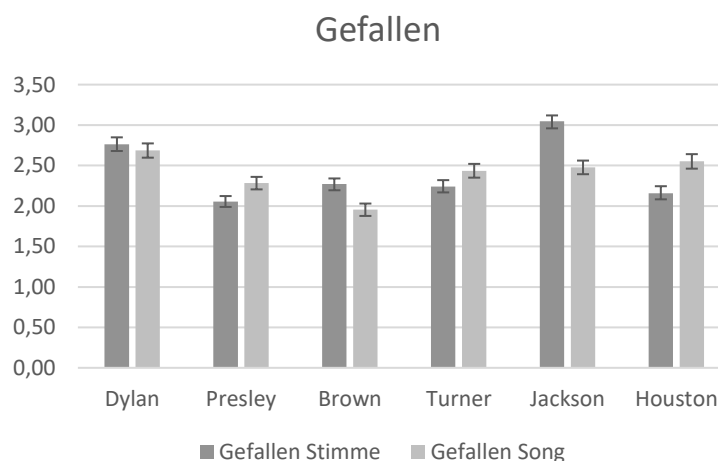


Abbildung 13: Balkendiagramm der Mittelwerte mit Standardfehler des Gefallens Stimme und Song für jeden Sänger, bewertet auf einer 5-stufigen Skala.

Zusammenfassend kann gesagt werden, dass sich sowohl aus der Interview- als auch aus der Gruppenstudie die emotionale Wirkung einer Stimme als eine wichtige Beschreibungsebene für die Teilnehmer herauskristallisierte und deshalb in die Bewertung integriert wurde. Es sollte auch im Vergleich zu den Merkmalsratings nachvollzogen werden, welche Relevanz die Emotionen für die Urteilsbildung über eine konkrete Stimme haben. Die Ergebnisse zeigen, dass die ausgewählten Emotionen keine gute Differenzierung der Sänger ermöglichten (die Diskriminanz auf Basis der Merkmale war wesentlich besser). Die Bewertungen der durch die Stimme ausgelösten Emotionen sind außerdem wesentlich durch die Emotionen konfundiert, die durch die Musik ausgelöst werden, so dass es nicht möglich war, diese beiden statistisch voneinander zu unterscheiden. Korrelationen der Emotionen mit den Merkmalen deuten eher auf einen Zusammenhang mit dem Song als mit dem Sänger hin, z.B. stammt die empfundene Energetisierung wohl eher vom Song Browns und Jacksons als dem damit assoziierten rau/kratzigen Stimmklang. Allerdings können andere Zusammenhänge durchaus nachvollzogen werden, z.B., dass ein druckloser Stimmklang entspannt. Nichtsdestotrotz kristallisierten sich keine Emotionen heraus, die spezifisch durch die Stimme evoziert werden. Die von den Teilnehmern als wichtiger Faktor der Urteilsbildung dargestellten Emotionen drücken eher einen allgemeinen Eindruck der Musik als einen spezifischen Eindruck von Stimmen aus. Weitere Forschungen dazu, inwiefern es ästhetische Emotionen gibt, deren Auftreten durch spezifische Stimmmerkmale verstärkt oder abgemildert werden, sind wünschenswert.

6.2.7 Andere Einflüsse

Zuerst sollen an dieser Stelle die Rückmeldungen der Teilnehmer betrachtet werden bzw. was sie an zusätzlichen Einflüssen auf ihre Ratings wahrgenommen haben. Nur 46 Personen nutzten die Kommentarfunktion am Ende des Fragebogens für Anmerkungen. Von diesen Teilnehmern merkten 14 an, dass ihnen die Auswahl der zur Verfügung stehenden Musikstile bei der Abfrage ihrer Gesangstilvorlieben nicht ausgereicht habe. Lediglich elf Teilnehmer berichteten, dass ihnen die Trennung von Stimme und Song schwerfiel. Zu den Merkmalen gab es sechs Anmerkungen (zweimal generelle Äußerungen wie „die Kriterien der objektiven Bewertung verstehe ich z.T. nicht“, zweimal zum Ausdruck und jeweils einmal zu ‚gepresst‘, ‚quäkig/nasal-dunkel/dumpf‘ sowie der Artikulation). Zwölf Teilnehmer gaben außerdem positives Feedback zum Fragebogen insgesamt und schrieben, dass sie die Studie interessant fanden. Die hier genannte gegenseitige Beeinflussung von Stimme und Song (bei den Emotionsratings) wurde bereits behandelt.

6.2.7.1 Einfluss musikalischer Vorbildung und Hörgewohnheiten

Die Indizes zur musikalischen Aktivität (MUSE; Chin & Rickard, 2012) wurden wie folgt ausgewertet: Der ‚Index of Music Listening (IML)‘ wird aus dem Produkt der Frequenz des Musikhörens pro Woche und pro Tag berechnet. Der ‚Index of Music Playing (IMP)‘ wird aus dem Produkt der Anzahl der Jahre, die ein Instrument (auch Gesang) gespielt wurde, den Übungsstunden pro Tag sowie dem Quotienten

dessen und der Regelmäßigkeit des Spiels berechnet. Der ‚Index of Music Training (IMT)‘ wird aus der Summe der Angaben zum höchsten Stand musikalischer Ausbildung an Schulen, anderer, außerschulischer Ausbildung und Studienabschlüssen berechnet. Die deskriptive Statistik der Daten (N = 216) zeigt bereits einige interessante Umstände auf, zum Beispiel, dass der Umfang der IMIP-Daten sehr groß ist, was bedeutet, dass Personen mit viel musikpraktischer Erfahrung unter den Teilnehmern waren (max. IMIP = 308, d.h. langjährige Berufsmusiker), allerdings zeigt der Modus von 0, dass die meisten Teilnehmer keine bis wenig musikpraktische Erfahrung hatten:

- IML: Mittelwert = 9,75, SD = 5,763, Min = 1, Max = 35, Modus = 10
- IMIP: Mittelwert = 33,769, SD = 54,832, Min = 0, Max = 308, Modus = 0
- IMT: Mittelwert = 3,708, SD = 1,597, Min = 1, Max = 8, Modus = 3

Berechnungen des Zusammenhangs zwischen musikalischer Aktivität und der Merkmalsbeurteilungen waren dementsprechend problematisch. Die Zusammenhänge zwischen den Indizes und den Bewertungen der Merkmale wurden mit linearen Regressionen für die drei Indizes und die Merkmale (N = 1296) analysiert, allerdings sind die Ergebnisse wenig aussagekräftig, da nur das Modell folgend dem IMIP signifikant wird und nur 2 % der Datenvarianz vorhersagt: $F(9, 1286) = 2,920$; $p = 0,002$. Die beiden Merkmale Stimmlage (Beta = 3,929, SEM = 1,320, $p = 0,003$) und Artikulationspräzision (Beta = -0,414, SEM = 1,188, $p = 0,042$) werden signifikant.

6.2.7.2 Einfluss Bekanntheit

Das Verhältnis von Bekanntheit Song sowie Bekanntheit Stimme wurde mit einer Kreuztabelle (Tabelle 31) und einem Chi-Quadrat-Test ermittelt (N = 1296), der signifikant wurde, d.h. die Verteilungen beider Variablen sind voneinander verschieden ($\chi^2(1) = 307,393$, $p < 0,001$; Phi-Koeffizient/Cramers V = 0,497). Die Kreuztabelle zeigt, dass wenn der Song bekannt ist, auch die Stimme bekannt ist (höhere Anzahl als erwartet). Diese Kombination tritt in 56,3 % der Fälle auf (in 95,7 % innerhalb Stimme und 71,1 % innerhalb Song). Der Fall, dass beide unbekannt sind, tritt ebenso häufiger als erwartet auf, allerdings nur in 18,4 % der Fälle. Etwas häufiger kommt es vor, dass der Song bekannt ist, der Sänger jedoch nicht (22,8 %). Es kam nur selten vor, dass der Sänger bekannt war, der Song aber nicht (nur 33mal, d.h. 2,5 % der Gesamtzahl).

		Bekanntheit Song		Gesamt	
		ja	nein		
Bekanntheit Stimme	ja	Anzahl	729	33	762
		Erwartete Anzahl	602,7	159,3	762,0
	nein	Anzahl	296	238	534
		Erwartete Anzahl	422,3	111,7	534,0
Gesamt		Anzahl	1025	271	1296

Tabelle 31: Kreuztabelle von Bekanntheit Stimme und Bekanntheit Song.

Die meisten Teilnehmer kannten dementsprechend sowohl die Stimme als auch den Song. Im Vergleich dazu waren beide nur in wenigen Fällen unbekannt. Es kann also, wie erwartet, von einer überwiegenden Bekanntheit der Stücke/Sänger ausgegangen werden.

6.2.7.3 Geschlechterunterschiede

Von den 139 weiblichen Teilnehmern fallen 70 in die Gruppe der Nicht-Experten, 34 in die Gruppe der Sprechstimm-Experten und 35 in die Gruppe der Singstimm-Experten. Von den 77 männlichen Teilnehmern fallen 45 in die Gruppe der Nicht-Experten, lediglich 7 in die Gruppe der Sprechstimm-Experten und 25 in die Gruppe der Singstimm-Experten (Tabelle 32). Das bedeutet, dass die Gruppe der Sprechstimm-Experten den stärksten Unterschied in der Geschlechterverteilung der Gruppenmitglieder aufweist, während die Gruppe der Singstimm-Experten relativ ausgeglichen ist.

		Experten			Gesamt
		nein	Sprechen	Singen	
Geschlecht	männlich	45	7	25	77
	weiblich	70	34	35	139
Gesamt		115	41	60	216

Tabelle 32: Kreuztabelle mit der Geschlechterverteilung und den Expertengruppen (Anzahl pro Gruppe).

Bei Bewertungen von Stimmen sind Geschlechterunterschiede besonders dann wichtig, wenn zum Beispiel die Attraktivität von Stimmen/Sprechern untersucht wird. Um eventuelle Geschlechtereinflüsse auszuschließen, werden oft nur weibliche oder nur männliche Sprecher für Stimuli verwendet (vgl. Grawunder, Bose, Hertha, Trauselt & Anders, 2006). In der aktuellen Studie wurde deshalb auf ein ausgeglichenes Geschlechterverhältnis in der Auswahl der Stimmen geachtet (drei Sängerinnen und drei Sänger). Es ist deshalb nur von Interesse zu überprüfen, ob der große Anteil an Frauen in der Sprechstimm-Experten-Gruppe einen Einfluss auf die Bewertung hat. Zuerst wurde mit einer binärlogistischen Regression versucht, die Merkmalsratings von Männern und Frauen vorherzusagen, was ein relativ schwaches Modell mit nur 2,2 % (Nagelkerkes R^2) aufgeklärter Varianz zeigt: In der Beschreibung der Merkmale finden sich Unterschiede zwischen den Geschlechtern ($\chi^2(9) = 21,066$, $p = 0,012$). Merkmale, die das Geschlecht vorhersagen, sind Stimmlage (Beta = -0,137, SEM = 0,051, $p = 0,007$) und Artikulationsbesonderheiten (Beta = 0,148, SEM = 0,046, $p = 0,001$). Die Unterschiede im Mittelwert der beiden Merkmale sind allerdings eher gering: Für Stimmlage liegen Frauen im Mittel (MW = 3,397, SD = 1,398) etwas niedriger als Männer (MW = 3,623, SD = 1,326), für Artikulationsbesonderheiten liegen Männer (MW = 3,106, SD = 1,392) etwas niedriger als Frauen (MW = 3,361, SD = 1,5). Es findet sich zudem kein signifikantes Regressionsmodell für Geschlecht und die Emotionsitems ($\chi^2(7) = 3,484$, $p = 0,837$ mit Nagelkerke $R^2 = 0,004$). Das bedeutet, dass ein Geschlechterunterschied in dieser Untersuchung nicht weiter beachtet werden muss.

6.2.7.4 Gefallen allgemeiner Gesangsstile

Die Abfrage des Gefallens verschiedener Gesangsstile zielte darauf ab, ein Bild von den Musikpräferenzen der Teilnehmer zu bekommen. Da die ausgewählten Tonbeispiele nur aus dem popular-musikalischen Bereich waren, sollte geschaut werden, ob die Teilnehmer grundsätzlich diese Gesangsstile mögen oder ablehnen. Im Mittel neutrale Ratings zeigten Country und Rap sowie Musical, Metal gefällt als einziger eher nicht.

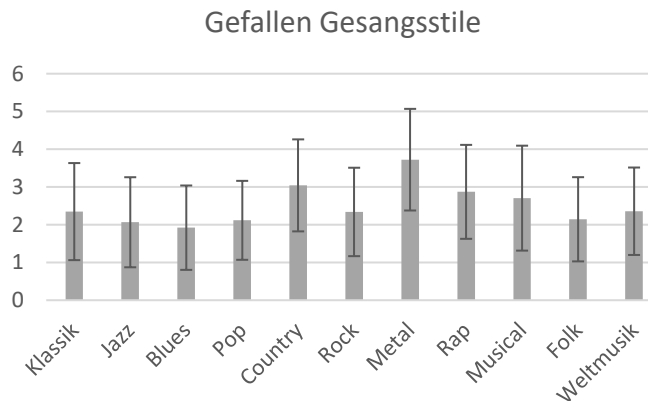


Abbildung 14: Mittelwerte der Ratings zum Gefallen der Gesangsstile (von 0 = gefällt mir bis 6 = gefällt mir nicht) mit Standardabweichung.

Auch wurden die Ratings mit den *MUSE Questionnaire*-Daten verglichen, um Zusammenhänge zwischen den Hörgewohnheiten und der musikalischen Vorbildung zu erfahren. Lineare Regressionen wurden für die drei Indizes mit den Gesangsstil-Ratings gerechnet.

- IML: kein sign. Regressionsmodell
- IMIP: sign. Modell ($F(11, 204) = 2,221, p = 0,015, R^2 = 0,107$)
 - Sign. Merkmal: Jazz (Beta = -11,950, SEM = 4,284, $p = 0,006$)
- IMT: sign. Modell ($F(11, 204) = 4,389, p < 0,001, R^2 = 0,191$)
 - Sign. Merkmale: Klassik (Beta = -0,276, SEM = 0,083, $p = 0,001$)
 - Jazz (Beta = -0,330, SEM = 0,119, $p = 0,006$)
 - Rock (Beta = -0,266, SEM = 0,110, $p = 0,017$)

Die Daten zeigen dabei deutlich, dass die Teilnehmer mit höherem Index in ‚instrument playing‘ und ‚training‘ ein stärkeres Gefallen an klassischem Gesang, Jazz-Gesang und Rock-Gesang haben. Dieses Ergebnis stimmt mit Elvers, Omigie, Fuhrmann und Fischinger (2015) überein, die zeigten, dass Studierende der Musikwissenschaft eher „sophisticated styles“ (Elvers et al., 2015, S. 1), d.h. Klassik und Jazz, hören.

Allerdings sind die Ergebnisse des *MUSE Questionnaire* mit Vorsicht zu genießen. Ein Problem mit den Indizes wird deutlich, wenn man sich die Verteilungen der Daten ansieht, die sehr ungleichmäßig sind.

So fällt z.B. der Großteil der Werte bei IML auf 10 und 15, bei IMIP auf 0. Die Streuung der Daten ist also nicht ausreichend, um sinnvolle Berechnungen und Vergleiche anzustellen. Es kann an dieser Stelle vermutet werden, dass die Indexe des *MUSE Questionnaire* aufgrund der zu erwartenden Verteilungen in einer Gruppe kaum verwertbar sind (wie auch in Merrill, 2017 bereits ansatzweise vermutet).

6.2.7.5 Gefallen und Nicht-Gefallen einzelner Sänger

Die Teilnehmer konnten bis zu drei SängerInnen angeben, die ihnen gefallen und nicht gefallen. Insgesamt wurden 429 verschiedene SängerInnen genannt, was bedeutet, dass sich kaum Überschneidungen fanden. Die Tabelle 33 zeigt die SängerInnen, die 10mal und mehr genannt wurden: Adele und Amy Winehouse gefielen, während Rihanna, Jan Delay, Anastacia, Herbert Grönemeyer und Helene Fischer eher nicht gemocht wurden. Dass die Liste insgesamt so viele verschiedene Namen enthält, zeigt, dass sich keine Stimmen finden, wo sich das Gefallen oder Nicht-Gefallen polarisiert.

Interpret	Gefällt	Gefällt nicht
Adele	33	2
Amy Winehouse	11	0
Rihanna	5	13
Jan Delay	0	12
Anastacia	3	10
Herbert Grönemeyer	1	10
Helene Fischer	0	10

Tabelle 33: SängerInnen, die gefallen und nicht gefallen.

6.2.8 Benötigte Stichprobengröße

Für die Fragestellung wesentliche Tests wurden noch einmal mit einer kleineren Stichprobe durchgeführt, um für zukünftige Studien eine Stichprobengröße zu definieren, die zu aussagekräftigen Ergebnissen führen kann. Es ist an dieser Stelle darauf hinzuweisen, dass eine Poweranalyse (oder Fallzahlenanalyse) aufgrund der vorliegenden Verteilung der Daten nicht sinnvoll ist. Deshalb wurde eine Zufallsauswahl getroffen, die 25 % der Teilnehmer umfasst. Zum Vergleich wurden noch einmal eine MDA zur Diskriminanz der Sänger auf Basis der Merkmale sowie Chi-Quadrat-Tests der Verteilungen für die Merkmale für Dylan, bei dem es weniger Einigkeit gab, und Presley, bei dem es mehr Einigkeit gab, berechnet.

Die MDA wurde über alle Stimmen zusammen durchgeführt (N = 329, d.h. zwischen 42 und 65 Teilnehmer pro Stimme) und zeigte vergleichbare Ergebnisse (vgl. Kapitel 6.2.4.1): Chi-Quadrat-Tests mit Wilks' Lambda zeigten drei (nicht wie vorher fünf) Funktionen, die einen statistisch signifikanten Einfluss auf die Diskrimination haben (Funktionen 1 bis 3: $p < 0,001$). Nur die erste Funktion hat einen Eigenwert > 1 (3,481).

- Die erste Funktion erklärt 74,9 % der Varianz und darauf laden die Merkmale Geräusch und Spannung sowie Tonhöhenveränderungen (letzteres kommt hinzu im Vergleich zur größeren Stichprobe).

- Die zweite Funktion erklärt 17,8 % der Varianz und darauf laden die Artikulationsbesonderheiten (nicht Artikulationspräzision, Phonationsform).
- Die dritte Funktion erklärt 6,1 % und darauf laden Timbre und Stimmlage. (identisch)
- Die vierte Funktion erklärt 0,7 % und darauf lädt das Merkmal Ausdruck. (identisch)
- Die fünfte Funktion erklärt 0,4 % und drauf laden die Merkmale Phonationsform und Artikulationspräzision (nicht das Merkmal Tonhöhenveränderungen).

66,3 % der kreuzvalidierten gruppierten Fälle wurden korrekt klassifiziert. Dylan wurde in 59,4 (67,1 %) der Fälle (kreuzvalidiert) korrekt klassifiziert, Presley in 90,3 (81,5 %), Brown 34,1 (47,4), Turner 57,1 (54,2), Jackson 64,6 (52,8) und Houston 82,7 (73,1 %).

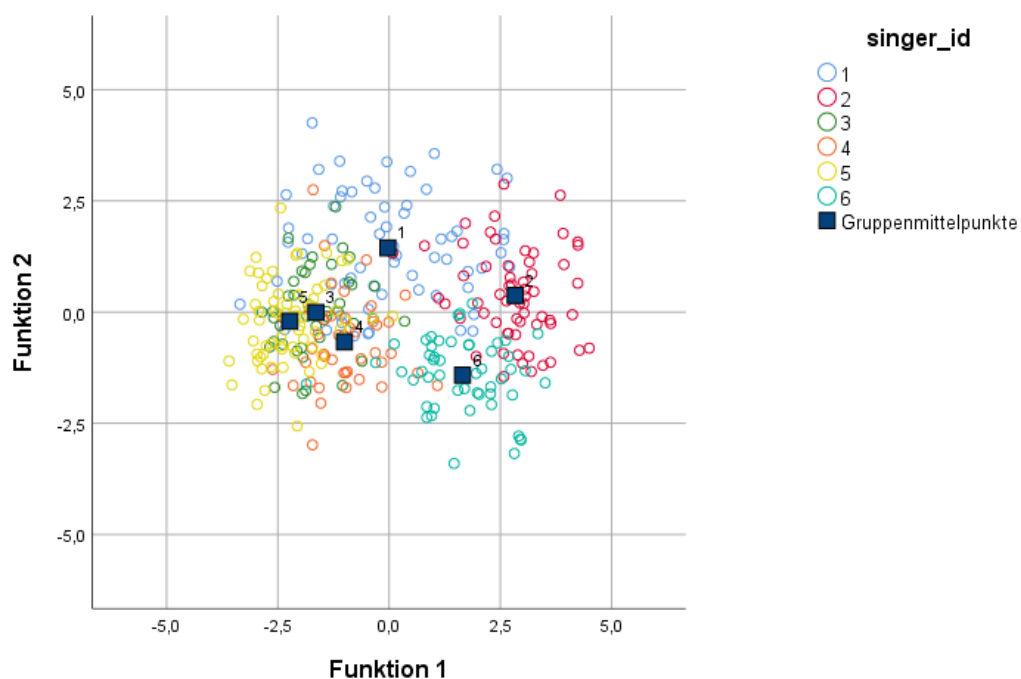


Abbildung 15: Streudiagramm der MDA mit 25 % der Teilnehmer. Sänger und Merkmale, Funktion 1 und 2 siehe Text. Singer_id: 1 = Dylan, 2 = Presley, 3 = Brown, 4 = Turner, 5 = Jackson, 6 = Houston.

Auch in den beispielhaften Chi-Quadrat-Tests auf Basis von 25 % der Daten (Kapitel 6.2.3) unterscheiden sich die folgenden Merkmale bei Dylan signifikant zwischen der beobachteten und der hypothetischen Verteilung in den Merkmalen Artikulationsbesonderheiten, Artikulationspräzision, Ausdruck, Timbre und Tonhöhenveränderungen. Dies bedeutet zugleich, dass die in den Chi-Quadrat-Tests über alle Teilnehmer festgestellten kleinen/mittleren Effektgrößen hier nicht signifikant werden, d.h. Stimmlage, Geräusch, Phonationsform und Spannung fallen heraus. Damit sind die Ergebnisse der größeren Stichprobe in den relevanten Effekten weit genug identisch, um eine kleinere Stichprobe zu verwenden. Es fällt sogar leichter, die Ergebnisse mit der kleineren Stichprobe zu interpretieren, da keine Effektgrößen zur Interpretation hinzugezogen werden müssen. Für Presley werden alle Chi-

Quadrat-Tests mit $N = 45$ signifikant. Auch die Ergebnisse dieser Tests sind somit mit denen der großen Stichprobe vergleichbar.

Daraus kann geschlossen werden, dass eine Stichprobe von $N = 45$ ausreicht, um die Ergebnisse, die mit 216 Teilnehmern erzielt wurden, zu erreichen. Während sich die Ergebnisse von Dylan und Presley ohne Verluste darstellen ließen, gab es bei der Diskriminanz der Sänger leichte Unterschiede in der Zusammensetzung der Funktionen. Die Distanzen zwischen den Sängern (siehe Abbildung 15) bleiben aber gleich und die Merkmale Spannung und Geräusch sowie artikulatorische Merkmale tragen am stärksten zur Diskriminanz bei.

7 Gesamtbetrachtung der Ergebnisse

Die Entwicklung eines Inventars zur Evaluation von Singstimmen, das auch von Nicht-Experten genutzt werden kann, war das vorrangige Ziel dieser Untersuchungen. Das Inventar sollte Merkmale des stimmlich-artikulatorischen Ausdrucks umfassen, die eine Stimme möglichst charakteristisch und eindeutig darstellen können. Es wurde gleichzeitig in jedem Untersuchungsschritt dazu genutzt, die vokalen Merkmale in Bezug zu musikalischen und stimmlichen Präferenzen sowie ästhetischen Emotionen zu setzen, wodurch die Beschreibungsebene mit Bewertungsebenen in Verbindung gebracht werden konnte.

Über die Studien hinweg wurde dem Feedback der Teilnehmer viel Raum gegeben, so dass ihre Anmerkungen einen direkten Einfluss auf die Entwicklung des Inventars hatten. So wurden die Musikstücke/Stimmen in der Interview-Studie auch nicht von der Studienleiterin, sondern von den Teilnehmern ausgewählt. Damit wurde eine Auswahl an Stimmen generiert, die aus dem musikalischen Alltag der Teilnehmer stammten, und es kristallisierten sich wichtige zu bewertende Merkmale heraus, die in der freien Beschreibung und somit im regulären Vokabular der Nicht-Experten vorkamen.

Der Ausgangspunkt der Untersuchungen war die Ablehnung bestimmter Stimmen aufgrund ihres stimmlich-artikulatorischen Ausdrucks. Dieser Ansatz wurde gewählt, da vermutet wurde, dass besonders die abgelehnten Stimmen oft einen auffälligen Stimmklang aufweisen und somit zu einer charakteristischeren Beschreibung führen, was von Vorteil für die Evaluation verschiedener Stimmerkmale ist. Dass sich, wie erwartet, die individuelle Ablehnung nicht in einer größeren Gruppe widerspiegelt, konnte in der folgenden Gruppen-Studie gezeigt werden. Da die Musikauswahl in der Ausgangsstudie den Teilnehmern überlassen war und die Ablehnung am selben Set überprüft werden musste, basierte die Entwicklung des Inventars auf Stimmen der populären Musik: Kein Teilnehmer brachte eine klassische Stimme aus dem Genre Oper oder Kunstlied mit (Sarah Brightman war mit einem Musical-Titel vertreten), weswegen auch in den Folgestudien mit Stimmen der populären Musik gearbeitet wurde. Der Wunsch, dass man gerne auch klassische Stimmen bewertet hätte, wurde auch nur wenige Male in der Online-Studie geäußert, außerdem fiel auf, dass das Gefallen von Klassik (und Jazz) in Verbindung zur musikalischen Vorbildung steht. Da klassische Stimmen vorwiegend im Konzert oder an Musikhochschulen beschrieben und evaluiert werden, handelt es sich um eine bestimmte und gesellschaftlich gesehen eher kleine Gruppe, die sich damit auseinandersetzt. Das heißt, der Umgang mit klassischen Stimmen scheint auf bestimmte Hörer begrenzt zu sein, für die es bereits andere Inventare zur Evaluation gibt (z.B. Henrich et al., 2008). Klassische Musik/Stimmen spielen offenbar in der Lebenswelt der Teilnehmer keine so große Rolle wie bestimmte Stile der populären Musik und wurden möglicherweise auch deshalb nicht mitgebracht. Es lässt sich zudem die Vermutung anstellen, dass eine charakteristische Beschreibung klassischer Stimmen schwieriger ist, da es sich um einen sehr

normierten Gesangsstil handelt (vgl. die Studie in Kapitel 3). Bevor die Merkmale des Inventars und die ästhetischen Urteile der Teilnehmer der vorliegenden Studienreihe abschließend diskutiert werden, soll auf einige methodische Aspekte eingegangen werden.

Die Bedeutung des Expertenurteils veränderte sich über die Studien hinweg. Da der Ausgangsfragebogen Fachbegriffe enthielt, wurden in Interview- und Gruppen-Studie das Expertenurteil als eine Korrektorebene herangezogen, an dem das Urteil der Nicht-Experten auf seine Korrektheit überprüft wurde. Diese Vorgehensweise wurde für die Online-Studie dahingehend geändert, dass nun Vergleiche zwischen unterschiedlich geschulten Gruppen mit verschiedenen Erfahrungswerten im Umgang mit Stimmen angestellt wurden. Es zeigten sich kaum Unterschiede zwischen den Gruppen und eine große Einigkeit unter den Teilnehmern, was bedeutet, dass alle Gruppen gleichermaßen gut mit dem Inventar umgehen konnten. Auch konnten die Sängerstimmen auf Basis der gegebenen Merkmale von Experten und Nicht-Experten gut eingeschätzt und beschrieben werden. Diese Ergebnisse lassen den Schluss zu, dass die ausgewählten Stimmen in ihren Eigenheiten dargestellt werden konnten und somit das Inventar sein Ziel erreicht hat, Gemeinsamkeiten und Unterschiede von Stimmen abbilden zu können.

Das Inventar ist nicht als standardisiertes Instrument zu verstehen, sondern als offener Fragebogen, d.h. es kann der Forschungsfrage angepasst werden. Als Beispiel sei die Abfrage nach der Phonationsform gegeben: Hier können die Pole Singen und Sprechen ebenso wie Rufen oder Schreien gewählt werden, womit Gesangstechniken der populären Musik (z.B. Shouting und Belting) weiter differenziert werden können. Generell wird für alle Merkmale die Nutzung einer Mittelkategorie empfohlen, damit auch Sänger, die ein Merkmal nicht eindeutig aufweisen (wie Phonationsform bei Bob Dylan oder Timbre bei Whitney Houston) von den Teilnehmern zufriedenstellend bewertet werden können.

Bei der Auswahl der zu bewertenden Stimmen ist zu berücksichtigen, dass die gewählten Ausschnitte nicht zu lang sein sollten. Das Merkmal ‚Ausdruck‘ ist zwar dafür bestimmt, Veränderungen über einen gewissen Zeitverlauf hinweg zu betrachten, allerdings ist davon auszugehen, dass die zu bewertenden Merkmale bei mehr Variation in ihren Ausprägungen zu mehr mittleren Bewertungen führen, auch weil sich die Teilnehmer zunehmend in ihren Urteilen unsicher werden. Wichtig ist die Anweisung, dass ein Gesamteindruck wiedergegeben werden soll, also eine Art Durchschnitt oder Mittelwert.

Das resultierende Inventar (Tabelle 34) ist für die Anwendung in der Forschung entwickelt worden. Ein Gesamturteil über mehrere Personen hinweg ist möglich und die resultierenden Profile verschiedener Stimmen können als Grundlage weiterer Untersuchungen verwendet werden. Zu einer hinreichenden Diskriminanz der Sänger und einer statistisch überprüften Einigkeit der Teilnehmer führen in der vorliegenden Studie bereits 45 Teilnehmer.

7.1 EVer – Ein neues Inventar zur Beschreibung von Stimmen in der (populären) Musik

EVer – Evaluation of Vocal Expression

Inventar zur Beschreibung von Stimmen in der populären Musik

Merrill, J. (2018)

Bitte beschreiben Sie den Gesamteindruck der Stimme anhand der vorgegebenen Merkmale.

Die Stimme klingt eher...						
tief	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	hoch
rau/kratzig	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	weich
gepresst	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	drucklos
quäkig/nasal	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	dunkel/dumpf
Die Veränderungen von Ton zu Ton sind eher...						
sprunghaft	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	gleitend
Die Artikulation ist eher...						
präzise	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	unpräzise
eigentümlich	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	gewöhnlich
Die Vortragsform ist eher...						
gesungen	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	gesprochen
Der stimmliche Ausdruck ist eher...						
wechselhaft	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	gleichförmig

Tabelle 34: Inventar vokaler Merkmale zur Beschreibung von Stimmen in der populären Musik für Stimm-Experten und Nicht-Experten.

Im Folgenden werden die Merkmale des vorliegenden Inventars beschrieben und in ihrer Entwicklung über die Studienreihe hinweg zusammengefasst.

Die Stimmlage kann auf zwei Ebenen verstanden werden: Zum einen kann jemand eine höhere oder tiefere Stimme haben, zum anderen aber auch höher oder tiefer sprechen (Eckert & Laver, 1994, S. 32ff.). Während ersteres einer physiologischen Konstitution entspricht, ist letzteres eher ein situationsabhängiges Phänomen. Menschen sprechen oft höher, wenn sie lauter sprechen, da die gesamte muskuläre Spannung im Stimmapparat mit gesteigerter Lautstärke zunimmt. Bezogen auf die Sprechstimme kann von einer „mittleren Sprechstimmlage“ gesprochen werden, einem Ton, um den herum das ‚normale‘ Sprechen stattfindet (vgl. Anders 2013).

Singstimmen werden nach wie vor primär in die klassischen Stimmgattungen eingeteilt, wonach mit Sopran und Tenor in höhere Stimmen und Alt und Bass in tiefere Stimmen unterschieden wird. Typischerweise sind auch die Kompositionen diesen Stimmgattungen angepasst, d.h. ein Sopran singt in höheren Lagen als ein Alt und hat einen anderen Stimmumfang. In der populären Musik ist dies weniger eindeutig zu kategorisieren, wobei auch hier die klassische Einteilung mit Einschränkungen anwendbar ist: Annett Louisan ist eher ein Sopran als ein Alt und Udo Lindenberg ließe sich eher als Bass klassifizieren. Auch in der populären Musik wählen Sänger die zu singenden Tonumfänge passend zu ihrer Stimme. Bestimmte Techniken des populären Gesangs wie das Belting überdecken allerdings

die klare Dichotomie ‚hoch-tief‘. Auch wenn Sänger einen großen Stimmumfang haben und diesen in ihren Songs nutzen, wie zum Beispiel bei Whitney Houston, wird die klare Einordnung in ‚hoch-tief‘ erschwert. Zudem erschweren einzelne tiefe Töne (zum Beispiel bei Bob Dylan) eine eindeutige Bewertung. Dies zeigt, dass der Umfang der Melodie einen Einfluss auf die Stimmlage hat und auch in dieser Abhängigkeit bewertet wird. Bei Gitte Haenning wurde in den Kommentaren angemerkt, dass das Lied für sie sowohl zu hoch als auch zu tief sei, was bedeutet, dass die Stimmlage Haennings in der Bewertung in Bezug zur Melodie gesetzt wurde.

Das Merkmal ‚Stimmlage‘ wurde vor allem mit dem Merkmal ‚Klangfarbe‘ (hell-dunkel) verwechselt. Nicht-Experten konnten diese beiden Merkmale offenbar nicht gut voneinander unterscheiden, was auch daran liegen kann, dass beide miteinander verbunden sind: Eine hohe Stimme geht oft mit einem hellen und eine tiefe Stimme mit einem dunklen Stimmklang einher. Dies spiegelt sich auch in der Anwendung des aktuellen Inventars, da die tiefen Stimmen mit einem dunkel/dumpfen Klang und die hohen mit einem quäkig/nasalen korrelieren. Bei den ausgewählten Stimmen ist diese Kombination auch zutreffend (Wanda Jackson und James Brown). Gerade bei Whitney Houston wird der Unterschied innerhalb des Songs deutlich, da ihr Stimmklang in der tieferen Lage zunehmend dunkler wird. In der populären Musik wird sich diese Überschneidung häufiger finden lassen als im klassischen Gesang, in dem ein klanglicher Ausgleich aller Lagen angestrebt wird (vgl. Freytag, 2003; Seidner & Wendler, 2010).

An dieser Stelle sei angemerkt, dass mit hohen Tönen und der damit verbundenen erhöhten körperlichen Spannung eine Art auditive Illusion einhergeht. Ein Ton mittlerer Lage wirkt anders, je nachdem, ob er mit Rand- oder Vollstimme bzw. im Kopf- oder Bruststimmregister gesungen wird, wodurch die Einschätzung von ‚hoch‘ und ‚tief‘ abhängig ist – dies zeigt sich im Belting (vgl. Freytag, 2003).

Das Merkmal ‚Geräusch‘ (rau/kratzig-weich) hat über die Studien hinweg einige Änderungen erfahren. Während die Merkmalsausprägung ‚rau‘ unproblematisch war, war die Bestimmung des Gegenpols durchaus schwierig. Das Merkmal ‚rau‘ ist ein prominentes Merkmal zur Stimmbeschreibung in der klinischen Forschung, da es als Bestandteil des Komplexes Heiserkeit ein Symptom einer Stimmstörung ist (vgl. Kapitel 2.2.1). Auch wird zusätzlich zwischen Rauigkeit und Behauchtheit unterschieden. In der vorliegenden Studie wurde ‚behaucht‘ nicht weiter berücksichtigt, da es von den Nicht-Experten nicht zur Beschreibung der Stimme herangezogen und in der Interview-Studie als ‚unklar‘ beschrieben wurde. Auch zeigen andere Untersuchungen, dass Nicht-Experten diese Geräuschanteile nur schlecht (Kreiman & Gerratt, 1998) bzw. nur nach Training (Anders et al., 1988) unterscheiden können (vgl. Kapitel 2.2.3). Der Gegenpol ‚weich‘ wurde von Teilnehmern in der Gruppen-Studie vorgeschlagen und kann aus Sicht der Experten eine Behauchtheit einschließen. Besonders raue Stimmeinsätze können

als hart wahrgenommen werden, dem ein weicher Einsatz gegenübersteht (vgl. Kapitel 2.1.1). Deshalb wurde die Kombination der Merkmale ‚hart-weich‘ und ‚rau‘ als sinnvoll angesehen, was zu ‚rau/kratzig-weich‘ führte. ‚Kratzig‘ wurde als ein besonders gut verständliches Merkmal für Nicht-Experten beibehalten.

Das Merkmal ‚Geräusch‘ korreliert mit ‚Spannung‘, d.h. eine raue Stimme geht mit einem gepressten Eindruck von der Stimme einher, der weiche mit einem drucklosen. Die Merkmale wurden getrennt voneinander beibehalten, da ein gepresster Stimmklang nicht rau sein muss; er zeigt eine (zu) hohe Spannung an, die sich auch im quäkigen Stimmklang findet. Die Merkmale ‚Timbre‘, ‚Geräusch‘ und ‚Spannung‘ korrelieren auch deshalb miteinander, da die untersuchten Stimmen diese Kombinationen auch zeigen (wie für einige Gesangstile der populären Musik üblich; vgl. Hähnel, 2015; Freytag, 2003). ‚Geräusch‘ und ‚Spannung‘ sind ebenso bestimmende Merkmale für die Diskriminanz der Sänger.

Das Merkmal ‚Spannung‘ (gepresst-drucklos) war in seiner konzeptionellen Bedeutung den Teilnehmern von Anfang an gut verständlich. Der Druck beim Singen (d.h. der subglottische Luftdruck) ist ein wesentliches Merkmal des Singens und sollte unbedingt abgefragt werden. Schwierig war allerdings die Wahl der konkreten Pole: Während ‚gepresst‘ oft als Begriff zur Beschreibung von Stimmen genutzt wurde, ist der Gegenpol, also die Unterspannung, schwierig ohne negative Konnotation zu beschreiben. Während Bose (2001) die Begriffe ‚gespannt-ungespannt‘ nutzt, ist diese Wortwahl für die Nicht-Experten schwer nachvollziehbar. In der Gruppen-Studie kamen Rückfragen zum Thema Stütze (ein gesangstechnischer Begriff), welches das Konzept durchaus trifft, aber ein Fachwort darstellt, welches vermieden werden sollte. Der im Verlauf der Untersuchung gewählte Begriff ‚drucklos‘ führte zu guten Ergebnissen, da er auch positiv verstanden werden kann und nicht nur die negative Unterspannung (als ein ‚zu wenig‘) darstellt. Der Begriff ‚gespannt‘ wurde im Verlauf der Studien durch den Begriff ‚gepresst‘ ersetzt, da er als eine extreme Ausprägung der Spannung verstanden werden kann. Damit wird ein Begriff verwendet, den auch die Nicht-Experten aktiv benutzt haben. Auch konnte gezeigt werden, dass sich ‚gespannt‘ und ‚gepresst‘ ähnlich sind (Kapitel 5.2.3).

Der Merkmalskomplex zum Timbre (quäkig/nasal-dunkel/dumpf) hat im Verlauf der Studien die meisten Änderungen erfahren und umfasst mehrere Dimensionen der Stimmbeschreibung, die Experten voneinander unterscheiden (was auch entsprechend kommentiert wurde). Trotzdem war das Merkmal für die Experten verständlich. Durch die Kombination verliert es zwar an Exaktheit, die aber von den Nicht-Experten nicht ohne Training beurteilt werden könnte: Während ‚nasal‘ bei ihnen ein durchaus geläufiger Term war, wurde gezeigt, dass er oft inkorrekt verwendet wurde (Kapitel 4.2.4.6) und eher eine Enge im Rachenbereich beschreibt (= Twang, vgl. Kapitel 2.3.1). Der Begriff ‚quäkig‘ wurde im Interview mehrfach verwendet und in der Gruppen-Studie nicht als schwierig kommentiert. Aus diesen Gründen sowie durch seine onomatopoetische Wirkung wurde er übernommen, auch wenn

er kein Fachterminus ist. Die Kombination mit ‚nasal‘ entstand durch die Fehlnutzung der Teilnehmer, aber auch, um das Merkmal weiterhin im Fragebogen zu behalten, da es eine weitere Beschreibungsebene eröffnet.

Auch der Gegenpol zu ‚quäkig/nasal‘ wurde mehrfach angepasst, wobei hier ebenfalls andere Merkmale wie die Kombination ‚hell-dunkel‘ einfließen. ‚Dumpf‘ ergab sich aus der erprobten Kombination mit ‚schrill‘, welches häufiger genannt wurde, allerdings dann durch den Begriff ‚quäkig‘ ersetzt wurde. ‚Dumpf‘ verweist auf gedeckte höhere Frequenzen, die im ‚Quäkigen‘ als Gegensatz vorhanden sind. Helligkeit und „formantenbedingte Qualität“ bilden die zwei wichtigsten Teil- oder Unteraspekte der stimmlichen Qualität (Neppert, 1999, S. 65ff.), nach der Schallqualitäten auf einer Skala von dunkel bis hell eingeordnet werden können. Neppert benennt als zusätzliche Beschreibungen zu ‚dunkel‘ noch „stumpf, dumpf, warm, o.ä.“ und für ‚hell‘ noch „scharf, spitz, o.ä.“ (Neppert, 1999, S. 65). Der spektrale Energieschwerpunkt der dunklen Klänge und Rauschschalle liegt im Bereich relativ niedriger Frequenzen und bei helleren in relativ höheren Frequenzen. Die Kombination der Begriffe ist also sinnvoll und hat sich in der vorliegenden Studie bewährt. Allerdings verhindert sie eine von den Experten gewünschte begriffliche Exaktheit in der Beschreibung.

Die Form der Tonhöhenveränderungen ‚sprunghaft-gleitend‘ wurde beibehalten, da diese Veränderungen ein Ausdrucksmittel des Gesangs sind, und beschreiben den Einsatz eines Glissando oder Portamento. Zur „Ästhetik“ des Portamento schreibt Lehmann, dass ein stilsicherer Gebrauch „ein Kennzeichen sängerischer Meisterschaft“ sei, hingegen ein übertriebener Gebrauch („Geheule“) und ein „wahlloser, gewohnheitsmäßiger Einsatz [...] abzulehnen“ sei (Lehmann, 2016, S. 482; vgl. auch Merrill, 2017). Beide Ausdrucksmittel (die definitorisch schwer voneinander zu trennen sind) werden sowohl für die Beschreibung der Stimme in der klassischen wie der populären Musik verwendet, so z.B. im Jazz-Gesang (vgl. Seedorf, 2016b, S. 268; Merrill, 2019).

Die Artikulation wird in der finalen Version des Inventars auf zwei Ebenen abgefragt: auf der Ebene der Präzision und auf der Ebene der Besonderheiten bzw. Eigenheiten der Artikulation. Die Präzision war nicht von Beginn der Untersuchung an von Interesse, da sie in der Interview-Studie keine bedeutende Rolle spielte, wohl aber die Eigenheiten der Artikulation wie Dehnungen von Lauten oder Vokalfarben. Es stellte sich jedoch in den folgenden Untersuchungen heraus, dass die alleinige Beschreibung der Besonderheiten zu Irritationen führte, da die Teilnehmer bei einer Abfrage der Artikulation zuerst an die Präzision (i.S. von Verständlichkeit) dachten und diese mit dem Merkmal ‚Besonderheiten‘ nicht beschreiben konnten. Deshalb wurde das Artikulationsmerkmal um eine Abfrage der Präzision erweitert. In der Online-Studie wurden nur englische Liedtexte untersucht, was die Einschätzung der deutschen Muttersprachler hinsichtlich der Artikulation erschwerte. Manche Teilnehmer waren im Singen populärer Musik ausgebildet, wodurch ersichtlich wurde, dass (Pop-)Sänger hier wohl einen

anderen Maßstab als die Sprechstimm-Experten oder Nicht-Experten anlegten, was durch ihre Erfahrungen im Singen englischer Texte begründet werden kann. Nach der Ergänzung der Abfrage zur Artikulationspräzision können die Teilnehmer nun die Artikulation erschöpfend und korrekt beschreiben.

Die Phonationsform wird hier auf Singen und Sprechen reduziert, da ‚Rufen‘ in der Gruppenstudie nicht als besonders eindeutig empfunden wurde. Die gewählten Pole hängen bei diesem Merkmal stark von der Forschungsfrage sowie den Audiobeispielen ab. Weiterhin wird im Inventar nach der Vortragsform gefragt, da ‚Phonationsform‘ ein Fachterminus ist, der auch in den jeweiligen wissenschaftlichen Disziplinen nicht einheitlich verwendet wird (z.B. ‚Phonationsart‘ bei Bose, 2001; vgl. Merrill, 2019).

Der stimmliche Ausdruck wird mit ‚wechselhaft-gleichförmig‘ beschrieben. Hier waren sich die Teilnehmer der Online-Studie bei der Bewertung der Stimmen einig. Die Interview-Studie hatte gezeigt, dass die Nicht-Experten nicht zwischen den Modulationen im Klang und den Tönen (der Melodie) unterscheiden, d.h. dass hier nicht immer davon ausgegangen werden konnte, dass auch wirklich nur die stimmlichen Modulationen beurteilt wurden. Der Ausdruck ist ein wichtiges Merkmal, da er einen Komplex an vokalen Merkmalen umfasst, von denen einige im Inventar enthalten sind. Somit könnte die Merkmalsausprägung ‚wechselhaft‘ zum Beispiel erklären, warum bei anderen Merkmalen des Inventars eine mittlere Bewertung angegeben wurde.

Bei der Entwicklung der Merkmale stand das Bemühen um konnotationsfreie Begriffe im Mittelpunkt, damit eine Beurteilung des Gefallens nicht durch eine positive oder negative Konnotation gelenkt wird. Während dies in der Interview-Studie noch durchwachsen war, d.h. dass eher negativ konnotierte Begriffe gewählt und die Ablehnung beschrieben wurde, wurden die Begriffe in den folgenden Studien verändert und alle Merkmalsausprägungen so neutral wie möglich formuliert. Dies ist besonders wichtig, da das Gefallen eines Sängers (indirekt) auf Basis der Beschreibung vokaler Merkmale eingeschätzt werden sollte.

7.2 Ästhetische Emotionen und Werturteile

Ein weiteres Ziel der Studien war es, ästhetische Urteile zu den vorliegenden Stimmen zu untersuchen. So wurden die Teilnehmer nach dem Gefallen der Stimmen sowie der Songs befragt. Da die Studien zeigten, dass es für die Teilnehmer schwierig war, Stimmen nur hinsichtlich ihrer Merkmale und nicht hinsichtlich der von ihnen ausgelösten Emotionen zu beschreiben und zu bewerten, wurde in der Online-Studie die Abfrage nach der Wirkung der Stimmen in Hinsicht auf ästhetische Emotionen hinzugefügt. Leider ist, da die vorliegenden Musikstücke die Stimmen in Zusammenhang mit Instrumenten zeigten, eine Trennung zwischen der Beurteilung der Stimme und der Begleitung kaum möglich, auch wenn es im Interesse der Auswertung wünschenswert wäre. Die Verwendung von A-cappella-Stücken wäre jedoch artifiziell, da im Alltag im Allgemeinen Instrumente und Gesang

zusammen rezipiert werden. Außerdem ist, ganz generell, der Gesang nicht von dem gesungenen Song bzw. der Melodie zu trennen. Durch die getrennte Abfrage nach dem Gefallen von Stimme und Song wurde versucht, zumindest einen indirekten Vergleich zu ermöglichen, auch wenn die Bewertungen große Überschneidungen aufwiesen.

Die in der Interview-Studie gewonnenen Informationen zu abgelehnten Merkmalen von Stimmen wurden in der folgenden Studie einer größeren Gruppe zur Bewertung gestellt, um zu überprüfen, ob die zumeist sehr auffälligen und charakteristischen Stimmen generell ein gesteigertes Potential zur Polarisierung aufweisen, also vielleicht auch bei einer größeren Gruppe Ablehnung erfahren. Die Gruppen-Studie ergab, dass sich die Ablehnung aus den Einzel-Interviews trotz der stimmlichen Auffälligkeiten nicht im Gruppenurteil widerspiegelte, die Bewertungen waren im Mittel neutral. Allerdings zeigen die Ergebnisse der (Korrelations-) Analysen in der Gruppen-Studie Übereinstimmungen mit den in der Interview-Studie berichteten abgelehnten Merkmalen: Nicht gemocht wurden die vokalen Merkmale ‚gepresst‘, ‚quäkig/nasal‘ und ‚sprunghaft‘ in den Tonhöhenveränderungen. Es finden sich aber auch Merkmale, deren Ausprägungen nicht eindeutig einer Richtung zuzuordnen sind, da sowohl ein rauher Stimmklang als auch eine gesprochene Phonationsform manchmal gefallen, aber manchmal auch nicht gefallen können. In der Online-Studie sollte der Frage nachgegangen werden, ob diese Merkmale vom Sänger bzw. dem Song abhängen.

Rauigkeit im Kontext künstlerischer Forschung ist im Vergleich zur Rauigkeit als Symptom einer Stimmstörung (Kapitel 2.2.1) relativ wenig betrachtet worden (Kapitel 2.3.1) und wird auch als ästhetisches Merkmal aus Sicht der Phoniater eher negativ betrachtet (Kapitel 2.5.2). Die vorliegende Untersuchung hat gezeigt, dass ein rauher/kratziger Stimmklang in Kombination mit dem Song energetisierend und ein weicher entspannend wirken kann. Der rauhe/kratzige Stimmklang kann sowohl zum Gefallen als auch zur Ablehnung einer Stimme beitragen. Er scheint zum Song und zum Gesangsstil passen zu müssen – wie es bei James Brown der Fall ist, bei Casper fällt er hingegen negativ auf – was zum Ergebnis der Interview-Studie passt, die als häufigsten Grund für die Ablehnung auf sachlicher Ebene zeigte, dass die Stimme nicht zum Lied passe.

Diese Ergebnisse zeigen, dass es nur wenige prominente Merkmale gibt, die einigermaßen eindeutig mit Ablehnung oder Gefallen korrelieren. Am stärksten ist diese Verbindung bei ‚quäkig/nasal‘ ausgeprägt, die eher abgelehnt werden, also ein nasaler Stimmklang und ein verengter Rachenbereich, der in der klanglichen Qualität des im Amerikanischen bekannten Twang zur Geltung kommt. Die undeutliche Aussprache scheint auch eher ein nicht gemochtes Merkmal zu sein, während Besonderheiten in der Artikulation wieder verschieden beurteilt werden.

Insgesamt sind die vokalen Merkmale keine besonders guten Prädiktoren des Gefallens: Nur 23,4 % der Varianz der Daten wurden in der Gruppen-Studie erklärt und nur 17 % in der Online-Studie. Dies

hatten die Teilnehmer bereits indirekt geäußert, indem sie anmerkten, nicht nur einen Ist-Zustand der Stimme, sondern auch die ausgelösten Gefühle als relevant für die Wertung beurteilen zu wollen. Im Vergleich dazu ist die Varianzaufklärung des Gefallens durch die Emotionen wesentlich höher, bei 58 % für die Stimme, bei 57 % für den Song. Dies bestätigt die Annahme, dass sachliche, deskriptive Bewertungen zumindest für die Urteilsbildung in Hinsicht auf den Geschmack bei Nicht-Experten-Hörern weniger wichtig zu sein scheinen als die durch eine Stimme ausgelösten Gefühle.

Weitere Studien zeigten, dass mithilfe von Diskriminanzanalysen verschiedenen Emotionen (Bänziger et al., 2014) oder Intentionen (Hellbernd & Sammler, 2016) in der Sprache bestimmte vokale Merkmale zugeordnet werden können. Anhand vokaler Merkmale versuchte Ketzmerick (2007), Sprecher zu klassifizieren, allerdings nur mit Korrelationsanalysen. Der Versuch einer Klassifikation von Sängern auf Basis vokaler Merkmale sowie den im Hörer ausgelösten Gefühlen scheint bisher nicht unternommen worden zu sein. In der aktuellen Untersuchung zeigen die Diskriminanzanalysen, dass eine Trennung der Sänger auf Basis der Emotionen nicht eindeutig möglich ist, allerdings die Unterscheidung der Stimmen auf Basis der Merkmale gelingt. Ästhetische Emotionen als Reaktion auf die Stimmen gleichen sich wesentlich stärker als die charakteristischen Merkmale der Stimmen. Die ausgelösten Emotionen sind schlussendlich nicht charakteristisch für einzelne Stimmen – es sei denn, es liegt eine Ablehnung vor. Dann führen sie zu negativen emotionalen wie körperlichen Abwehrreaktionen, wie die Interview-Studie zeigte (ob im Fall einer negativen Bewertung der Stimme die berichteten Emotionen jedoch in einer Diskriminanzanalyse besser unterschieden würden, ist bislang nicht untersucht).

Da die abgefragten Merkmale für die Beschreibung des stimmlich-artikulatorischen Ausdrucks entwickelt worden sind, sind sie zur Beurteilung des Instrumentalspiels nicht geeignet. Das bedeutet, dass es bei der Beurteilung der Stimme anhand dieser Merkmale keine Konfusion mit der Beschreibung der begleitenden Instrumente geben sollte. Allerdings kann es vorkommen, dass stimmliche Merkmale nicht gut davon differenziert werden können, wenn die Begleitung zu sehr im Vordergrund ist. Dies ist bei der Beurteilung der ästhetischen Emotionen etwas anders, da sich diese auf verschiedene künstlerische Gegenstände beziehen können, was Begleitmusik und Gesang einschließt (und auch Literatur und visuelle Kunst; z.B. AESTHEMOS, Schindler et al., 2017). Deutlich wird, dass sich in der vorliegenden Untersuchung keine ästhetischen Emotionen zeigten, die sich spezifisch auf die Stimme beziehen. Energetisierung scheint sogar nur spezifisch für Songs zu sein (es wird nicht direkt von einer Stimme evoziert). Für alle Songs und Stimmen fanden sich die Emotionen Faszination und (Nicht-) Gereiztheit und viele Songs und Stimmen machten zusätzlich fröhlich und berührten die Teilnehmer. Entspannung und (Nicht-) Traurigkeit wird bei der gewählten Stück- und Stimmenauswahl fast gar nicht ausgewählt. Für weitere Forschung wäre eine grundlegende Arbeit von Interesse, mit der spezifisch den ausgelösten Emotionen beim Gesang (womöglich im Vergleich zum sängerischen Ausdruck) nachgegangen wird.

Die Online-Studie zeigt, dass das Gefallen der Stimme und des Songs gleichermaßen durch die ausgelösten Emotionen vorhergesagt werden können. Es finden sich keine Emotionen, die spezifisch durch die Stimme verursacht werden, trotz der expliziten Frage nach der Beurteilung der Stimme und nicht der Musik. Dies zeigt, dass eine Beurteilung der Stimme losgelöst von der Musik für die Teilnehmer nicht möglich oder nicht relevant ist – es sei denn, eines von beiden wird, wie in der Interview-Studie gezeigt, abgelehnt. Dort gab es mehrere Beschreibungen der Ablehnung einer Stimme bei gleichzeitiger Akzeptanz oder sogar Präferenz des Liedes bzw. des Musikstils. Dabei nannten die Teilnehmer als wichtigen sachlichen Grund der Ablehnung, dass die Stimme nicht zum Lied passe. Bei den dezidiert abgelehnten Stücken konnten die Teilnehmer also zwischen Stimme und Musik differenzieren, was bei den größtenteils neutral bis positiv bewerteten Beispielstücken der beiden Folgestudien nicht zu beobachten war.

7.3 Fazit und Ausblick

Das vorliegende Inventar ist vermutlich das erste, mit dem nicht speziell geschulte Hörer Singstimmen der populären Musik mit aussagekräftigen vokalen Merkmalen beschreiben können. Es umfasst eine Kombination an Merkmalen aus wesentlichen Kategorien der Stimmbeschreibung wie Klang, Töne, Artikulation und Ausdruck, mit denen ein möglichst umfassendes Profil einer Stimme erfasst werden kann. Damit können charakteristische Eigenschaften von Stimmen dargestellt und Ähnlichkeiten genauso wie Unterschiede zwischen Stimmen erfasst werden. Für zukünftige Untersuchungen kann also auch das Laien-Urteil zur Bewertung und Beschreibung von Stimmen herangezogen werden.

Dieses neue Inventar eignet sich für eine Vielzahl an Anwendungsmöglichkeiten. Unter anderem lassen sich damit die Zusammenhänge zwischen stimmlichen Merkmalen und bestimmten Reaktionen und Empfindungen untersuchen, die beim Hören von Singstimmen auftreten, wie in dieser Arbeit in Form von Präferenzen sowie ästhetischen Emotionen bereits geschehen ist. Auch wenn diese Untersuchung damit erste Ergebnisse darüber liefert, welche stimmlichen und klanglichen Merkmale zu welcher Reaktion beim Hörer führen, sind weitere und tiefer gehende Studien nötig. Aufgestellte Stimmprofile können dann die Wahrnehmung der Hörer ergänzen und als Erklärung für solche Bewertungen herangezogen werden (zum Beispiel, in welcher Situation ein wahrgenommener weicher Stimmklang eher entspannt und präferiert wird).

Es ist davon auszugehen, dass das Inventar Stimmen besser abbilden kann, je ausgefallener und individueller sie sind. Je ähnlicher sich die Stimmen werden, desto weniger wird das Inventar die Stimmen trennscharf und eindeutig beschreiben können. Das kann vor allem innerhalb eines Gesangsstils, der hoch standardisiert ist wie zum Beispiel der Operngesang, zu Schwierigkeiten führen, aber auch innerhalb besonders ausgefallener Stile wie beispielsweise dem Metal-Gesang, der

technische Mittel nutzt, die das Inventar nur bedingt abbilden kann und die hier auch nicht evaluiert wurden.

Abschließend kann gesagt werden, dass mittels EVEx die Merkmale von Singstimmen untersucht werden können, die Experten als auch Nicht-Experten bei Singstimmen wahrnehmen und bewerten können. Damit ist dieses Inventar eine wertvolle Ergänzung bereits bestehender Inventare, da es keinerlei Vorbildung oder Training zur Anwendung erfordert. Die mittels des Inventars erhobenen Beschreibungen des Hörers können dann als Grundlage für eine Vielzahl weiterer Forschungsfragen unterschiedlicher Disziplinen herangezogen werden.

8 Anhang

8.1 Studie Kapitel 3

8.1.1 Texte

Die folgende Aufstellung entspricht dem Format, wie es die Teilnehmer vorliegen hatten. Das heißt, es sind keine Auslassungen aus der originalen Textvorlage markiert. Bei der Auswahl wurde versucht, den zusammenhängenden Text beizubehalten. Es sind zum Teil jedoch einzelne Sätze oder Teilsätze ausgelassen, die den Namen des Sängers oder zu eindeutige Aussagen zur Identifikation des Sängers enthielten. Die hochgestellten Zahlen in den Texten zeigen den Code für die jeweiligen Texte (A-F) an, die Zahl in Klammern dahinter die Anzahl der Markierungen. Einmalnennungen sind nicht angemerkt. Auf die Codes wird in den folgenden Tabellen wieder verwiesen.

8.1.1.1 Beispiel A – Luciano Pavarotti

Mit ihm war der romantische Tenor¹ (2x) auf die Bühne zurückgekehrt – ein Sänger mit einer hellen, brillanten² (5x) Stimme ohne baritonale Tinta³ (3x) und mit der sexuellen Ambiguität⁴ (2x) einer Knabenstimme⁵ (5x). Die Reinheit des Klanges⁶ (2x) und die Deutlichkeit der Diktion⁷ (4x) sind außerordentlich, mit viel artikulatorischer Prägnanz⁸ (3x) und zugleich karessierender Innigkeit⁹. In der Höhe strahlt die Stimme mit jenem Glanz¹⁰, den die Italiener als *squillante* bewundern. Frühe Aufnahmen verblüffen durch die Verbindung von Lyrismus und Brillanz. Das „hohe c“¹¹ (3x) hielt er aus bis in alle Ewigkeit¹². Doch klingen die hohen Töne wegen der extremen Anspannung nicht frei¹³ (10x). Das bisweilen angespannte Singen¹⁴ (2x) ist der *Mezza-voce* wenig förderlich und macht ein Messa di voce – das An- und Abschwollen – fast unmöglich¹⁵ (3x); auch die Bildung von Sfumature¹⁶, von sanft und elegisch endenden Phrasen¹⁷, ist seine Sache nicht. Dass er kein wiedererstandener Rubini war – würde das heutige Publikum dessen androgyn-helle Stimme¹⁸ wirklich akzeptieren? -, verrät die 1980 veröffentlichte Aufnahme von *La Sonnambula*. Er ist zwar berühmt seine Stimme aufzulichten¹⁹ und schmelzen zu lassen, doch sind die Arien um einen Ton transponiert²⁰ (3x). Trotz der niedriger gelegten *Tessitura* klingen die Sostenuti nicht so sanft und weich²¹ (3x) wie etwa bei anderen Sängern⁹⁹.

8.1.1.2 Beispiel B – Placido Domingo

Er besaß eine unverkennbar timbrierte¹ (2x), weiche² (8x), volle³ (8x) und sinnlich einschmeichelnde⁴ (2x) Stimme. Sie war nicht so hell und schlank wie die italienischen Stimmen von anderen Sängern⁵ (8x). Ihr fehlte jene durchdringende Intensität⁶ (5x), die als *Squillo* bezeichnet wird. Ihre Farbe war die eines dunkelroten Vino de Rioja⁷. Es war eine Stimme mit einem unverhohlenen Macho-Klang⁸ (9x). Vor allem in seinen jüngeren Jahren war die Stimme nicht frei von kehligen Einfärbungen⁹ (5x). Zur Höhe hin war die Stimme begrenzt¹⁰ (3x). Selbst in besten Momenten schlug er die Töne der Vollhöhe nicht con attacca⁹⁹ an; er brauchte einen längeren Einschwingvorgang¹¹ (6x) oder den Bogen einer aufsteigenden Kantilene. Selbst dann klang es angestrengt¹² (8x) oder klanglich isoliert¹³ (6x). Die kurze

Höhe war der Tribut dafür, dass die Bruststimme mit großem Atemdruck¹⁴ (Pressphonation¹⁵) (6x) in die Höhe getrieben wurde. Erstaunlicherweise hat die Stimme im Verlauf der späteren Jahre an Festigkeit¹⁶ und an Konzentration¹⁷ (3x) gewonnen, insbesondere im Bereich vom Ende der ersten Oktave bis zum hohen b¹⁸ – ein Ton aus gut gehärtetem Metall¹⁹ (5x). Die untere Quinte der ersten Oktave klang⁹⁹ hingegen matter und, je nach Disposition, resonanzärmer²⁰ (2x).

8.1.1.3 Beispiel C – Jussi Björling

Die Stimme war von überragender Qualität¹. Sie besaß den Eigenreiz eines unverkennbaren silbrigen Timbres⁹ und, bei mittlerem Volumen³ (3x), große Schallkraft⁴ (4x) und eine wundervoll leicht ansprechende Höhe⁵ (7x). Er konnte h, c, und selbst das cis² ansatzlos attackieren⁶ (8x): Die Töne kamen wie von einem Sprungbrett abgewippt⁷. [...] Er verfügte über eine Tenorstimme⁸ mit silbrigem Klang⁹ (6x) und brillanter Höhe¹⁰ (bis cis²) (5x) sowie über eine nahezu perfekte Gesangstechnik¹¹ (3x). Präziser Stimmsitz¹² (5x), ausgezeichnete Atemkontrolle¹³ (2x) und eine schlanke Stimmführung¹⁴ gaben seiner Stimme eine Durchschlagskraft¹⁵ (2x), die sich auch in großen Häusern wie der Metropolitan Opera zu entfalten vermochte. Ohne die Robustheit eines *Tenore di forza* zu besitzen, gelang es ihm dank der Strahlkraft seiner Stimme¹⁶ (2x) – wohl Resultat eines ausgeprägten Sängerformantclusters¹⁷ -, sich auch in Heldenpartien (z.B. bei Verdi) zu behaupten. Er war ein Meister der klugen Dosierung¹⁸ (2x), in der Behandlung seiner Stimme ebenso wie in der Anwendung gesanglicher Stilmittel. Durch seinen schnörkellosen Gesangsstil¹⁹ (3x) erwies er sich als idealer Operninterpret für Aufnahmen.

8.1.1.4 Beispiel D – Aretha Franklin

Geprägt vom biographischen Hintergrund in der Gospel Music singt sie in vielen Songs zahlreiche Melismen¹ (7x) und dehnt viele Vokale und Konsonanten² (8x) und steigert durch ein sukzessives Ansteigen der Tonhöhe³ (8x) und Lautstärke⁴ (8x) die emotionale Spannung und Intensität ihrer Gesangsdarbietung⁵ (2x). Weiteres Kennzeichen ihres Gesangsstils ist ein vielgestaltiges Gleiten zwischen präzise intonierten Zieltönen (Glissando)⁶ (10x), zum Teil über größere Intervalle hinweg⁷ (3x). Ihr Stimmumfang reicht über zweieinhalb Oktaven⁸ (3x), ca. von e bis b², erscheint allerdings oftmals größer, da sie ihre Bruststimme weit nach oben ausdehnt (Belt)⁹ (11x). So singt sie auf den emotionalen Höhepunkten vieler Songs im Belt bis zum g² ¹⁰ (2x) und darüber hinaus, wobei ihr ansonsten voller Stimmklang¹¹ (2x) etwas enger¹² (2x) und manchmal nasal¹³ (11x) wird.

8.1.1.5 Beispiel E – Mahalia Jackson

Sie ist stark vom Blues¹ (2x) und den sog. Gospel-Shouters² (4x) beeinflusst. Ihrer Stimme gibt sie mit gerundeten Lippen³ (2x) eine Präsenz⁴ (2x), die sie bisweilen gezielt mit Rauheit⁵ (15x) anreichert. Bei aller Variabilität behält sie oft eine Ausdrucksweise⁶ (4x) über einen ganzen Song bei. Längere Töne singt sie mit einem konstanten⁷ (7x) und regelmäßigen Vibrato⁸ (15x). Ihr Stimmumfang bewegt sich selten über eine Dezime⁹ (5x) hinaus. Ihre Stimmlage ist tief¹⁰ (16x), ihr Gesang ist von Verzerrungen

und Improvisationen¹¹ (3x) geprägt. Häufig schleift sie Töne von unten an¹² (10x) oder erreicht sie über einen oder wenige Nachbartöne¹³. Übergänge zwischen Tönen gestaltet sie über Glissandi oder Melismen¹⁴, die sich aus Ab- und Aufwärtsläufen zusammensetzen.

8.1.1.6 Beispiel F – Bessie Smith

Sie singt mit kräftiger Stimme¹ (7x), sie gilt als typische Blues Shouterin² (4x). Ihr Gesang hat einen Tonumfang von anderthalb Oktaven (ca. von g bis d²)³ (2x), wobei sie in vielen Liedern nur einen Ambitus von einer Oktave (c¹ bis c²)⁴ (3x) oder weniger einsetzt. Ihre melodische Gestaltung zeichnet sich durch differenziertes Gleiten zwischen Tonhöhen¹⁰ (5x) und durch vielfältige Variationen von Tonhöhe, Klangfarben und Melodiephrasen⁵ (4x) aus. Charakteristisch für ihren Gesangsstil ist ein häufiges Aufwärtsgleiten⁶ (11x) zu Beginn von Tönen sowie ein Abwärtsgleiten⁷ (7x) zwischen Tönen innerhalb einer Phrase (Glissando)⁷; die meisten längeren Töne an Phrasenenden⁸ singt sie mit einem schnellen Vibrato⁹ (7x).

8.1.2 Tabellen

Befragung 2: Die folgenden zwei Tabellen zeigen die Anzahl der markierten Textstellen, aufgeteilt auf die vermuteten Sänger/innen. Beispielsweise stammen die Codes A1-A21 aus der Beschreibung der Stimme A (Pavarotti) und, wie in der Tabelle eingesehen werden kann, wurden die Beschreibungen fast gleich oft der Stimme Domingos sowie Pavarottis zugeordnet und nur sehr selten der Stimme Björlings. Die Codes sind in den Texten (A-C) zu ersehen, nur die häufiger genannten wurden hier in der Spalte „Beschreibung“ extra benannt. Zwei Codes wurden verbunden, wenn die Unterstreichungen ineinander übergangen oder die Beschreibungen sich inhaltlich eindeutig entsprachen.

Tabelle 35: Kodierung und Anzahl an Zuordnungen mit zusätzlichen Beschreibungen der klassischen Sänger.

Kodierung	1C_Björling	2B_Domingo	3A_Pavarotti	Beschreibung
A1		1	1	
A2		2	3	hell, brillant
A3		2	1	
A4			1	
A5		2	2	
A6		1	1	
A7	1	2	1	
A8		2	1	
A9			1	
A10		1		
A11/A12		1	2	
A13	6	1	5	Anspannung
A14				
A15	1	1	1	
A16			1	
A17	1			
A18		1		
A19		1		
A20		1	2	
A21	2		1	
B1		1	1	
B2	4	3	3	weich
B3		2	4	voll
B4			2	sinnlich
B5	2	4	2	nicht so hell und schlank
B6		3		durchdringende Intensität
B7	1			
B8	7	2		Macho
B9	3	1	1	kehlig
B10	1	1	1	
B11	2	2	2	längeren Einschwingvorgang
B12	5	1	2	Anstrengung
B13	4		2	klanglich isoliert + angestrengt
B14			2	
B15		1		
B16		1	1	
B17		1		
B18		1	1	
B19		2		
B20	1		1	
C1	1			
C3	1		2	
C4	3	1	1	Schallkraft
C5	2	3	2	leichte Höhe
C6	3	2	3	ansatzlos attackieren
C7			1	
C9	4		2	silbriger Klang
C10	3	2		brillante Höhe
C11		1	2	
C12	2	1	2	
C13	2			
C14		1		
C15	2			Durchschlagskraft
C16	1	1		
C17	1			
C18		2		
C19	1	1	1	

Tabelle 36: Kodierung und Anzahl an Zuordnungen mit zusätzlichen Beschreibungen der popularmusikalischen Sängerinnen.

Kodierung	1F_Smith	2D_Franklin	3E_Jackson	Beschreibung
D1	1	4	2	zahlreiche Melismen
D2	2	5	1	dehnt viele Vokale und Konsonanten
D3		7	1	sukzessives Ansteigen der Tonhöhe
D4		6	2	sukzessives Ansteigen der Lautstärke
D5		1	1	
D6	1	6	3	Glissando
D7		3	1	Glissandi über größere Intervalle hinweg
D8		3		Stimmumfang
D9	1	9	1	Bruststimme/Belt
D10		2		bis zum g2
D11		1	1	
D12		1	1	
D13	1	7	3	nasal
E1			2	Blues
E2	1	1	2	
E3	2			
E4	1		1	
E5	4	3	8	Rauheit
E6	1	1	2	
E7	3	2	2	
E8	7	3	5	regelmäßiges Vibrato
E9	1	1	3	
E10	4	1	11	Stimmlage ist tief
E11		1	2	
E12/E13	2	2	6	schleift die Töne von unten an
E14	1		2	
E15			1	
F1	4		3	kräftige Stimme
F2	1		3	
F3	1		1	
F4	1		2	
F5	1	2	1	
F6	8	1	2	häufiges Aufwärtsgleiten
F7	6	1		Abwärtsgleiten/Glissando
F8	4	1	2	Phrasenenden
F9	12	1	5	schnelles Vibrato
F10	3	1	1	differenziertes Gleiten

8.2 Interview-Studie Kapitel 4

Tabelle 37: Übersicht der abgelehnten Merkmale im Interview.

DV01	unsauber	lang gezogen	gequält	kratzen
DV02	unnatürlicher Akzent	nasal	monton	kindlich
DV03	hält keinen Ton länger	weinerlich		
DV04	Tonlage passt nicht (höher oder tiefer)	schreibelig	flach	singt nur in einer Tonlage wie aus dem Eimer
DV05	gruselig	tief	monton	langweilig
DV06	unsauber	unnatürlich	hoch	spitz Alt-Sänger
DV07_1	verschnupft	nicht klar	Betonung	Spannung anstrengend
DV07_2	rau	angestrengt	Unlust	
DV08	knödelt	nicht ton für ton	Betonung	schmalzig
DV09	blöd	zugehaltene Nase	angestrengt	kindlich
DV10_1	abgehackt	hoch und runter	weite Spanne	
DV10_2	rauspressen	flapsig	alt	Unlust
DV11	quäkig	unmännlich	nicht klar	kratzend
DV12	hingerotzt	rezitiert	gesprochen	nutzt Tonhöhen und Stimmvolumen nicht Unlust
DV13	piepsig	quietschend	weiblich	doof unreif
DV14	vokal i u	scharrend	singt vokale nicht aus	arrogant genäsel
DV15	quaken	Ton kommt von oben im Hals	maskulin	
DV16	hoch und runter	schreit	wellenförmig	Laut aus der Kehle raus wechselhaft
		Ansingen Ton ng	rauspressen	
DV17	hoch	gepresst	dünn	hervorgestoßen
DV18_1	quakt	schwache Stimme presst	nicht offen wehleidig	Druck jammert
DV18_2	substanzlos	kraftlos	kein Klangbild dünn	nichtssagend schwach
DV19	blechern	keine Tiefe	flach jugendlich	nicht weich unmännlich
DV20	quäkig	gepresst	unnatürlich	rau Wechsel von hoch quäkig zu rau

Interview-Leitfaden

1. Haben Sie eine oder zwei Einspielungen dabei?
(Die Aufnahme wurde sich gemeinsam über Lautsprecher angehört. Bei zwei mitgebrachten Aufnahmen wurde zuerst eine angehört und der Fragenkatalog durchgegangen und dann die zweite, nach der alle Fragen und der Katalog wiederholt wurden.)
2. Was ist das für eine Aufnahme? Wer ist der Sänger / die Sängerin?
3. Was mögen Sie an dieser Stimme nicht?
4. Warum mögen Sie (genannte Eigenschaft) nicht? (Alle Eigenschaften nacheinander durchgehen.)
5. Welche dieser Eigenschaften ist für Sie die Wichtigste? Was mögen Sie am wenigsten?
6. Wie klingt diese Stimme für Sie? Können Sie sie beschreiben?

7. Was löst diese Stimme in Ihnen aus?
8. Wie geht es Ihnen, wenn Sie diese Aufnahme hören? (falls nicht genannt: Wie reagiert Ihr Körper darauf? Was für Gefühle haben Sie beim Hören?)
9. Erinnert Sie der Klang dieser Stimme an etwas? (Assoziationen)
10. Verbinden Sie Erinnerungen mit dieser Stimme oder dem Stück?

11. Mögen Sie die Stimme dieses Sängers / dieser Sängerin in anderen Stücken?
12. Mögen Sie diese Art von Musik?
 - a. Würden Sie das Lied vielleicht eher mögen, wenn es von einem anderen Sänger gesungen würde?
 - b. Mögen Sie den Musikstil grundsätzlich (nicht)?
13. Was denken Sie über den Sänger / die Sängerin? Wie ist er / sie?
14. Was denken Sie über den Text dieses Stückes? Können Sie sagen, worum es in dem Stück geht?
15. Haben Sie noch etwas hinzuzufügen?

Tabelle 38: Fragebogen mit standardisierten Erklärungen bei Nachfragen für den Interviewer (in Klammern).

Die Stimme ist zu...	<input type="checkbox"/> tief	<input type="checkbox"/> hoch
	<input type="checkbox"/> dunkel	<input type="checkbox"/> hell
	<input type="checkbox"/> laut	<input type="checkbox"/> leise
	<input type="checkbox"/> gespannt (überspannt)	<input type="checkbox"/> ungespannt (unterspannt)
Die Laute sind zu...	<input type="checkbox"/> gedehnt	<input type="checkbox"/> verkürzt
Die Tonveränderungen sind zu...	<input type="checkbox"/> sprunghaft (abgesetzt, abgehackt)	<input type="checkbox"/> gleitend (Glissando; Schmieren)
Das Singen erinnert mehr an	<input type="checkbox"/> Sprechen	<input type="checkbox"/> Rufen
Die Stimme ist instabil bzgl.	<input type="checkbox"/> Vibrato	<input type="checkbox"/> Zittern
Die Stimme klingt zu...	<input type="checkbox"/> klangvoll (dick, kräftig)	<input type="checkbox"/> klangarm (dünn, schwach)
	<input type="checkbox"/> weit (hohl)	<input type="checkbox"/> eng (Knödel, twang)
	<input type="checkbox"/> weich	<input type="checkbox"/> hart
	<input type="checkbox"/> dumpf	<input type="checkbox"/> schrill
	<input type="checkbox"/> behaucht	
	<input type="checkbox"/> rau	
	<input type="checkbox"/> geknarrt	
	<input type="checkbox"/> nasal (Näseln)	
Die Stimmeinsätze sind zu...	<input type="checkbox"/> weich	<input type="checkbox"/> hart
Die Artikulation ist zu...	<input type="checkbox"/> präzise (überartikuliert)	<input type="checkbox"/> unpräzise (schwer zu verstehen, nuscheln)
Die Stimme moduliert im Klang zu...	<input type="checkbox"/> wenig (eintönig)	<input type="checkbox"/> viel (übertrieben)
Die Stimme moduliert die Töne zu...	<input type="checkbox"/> wenig (eintönig)	<input type="checkbox"/> viel (übertrieben)

8.2.1 Anweisungen für die Experten

Aufgabe: Beschreibe anhand der aufgelisteten Merkmale den stimmlichen Ausdruck. Es sollte immer eine Entscheidung getroffen werden, in welche Richtung ein Merkmal tendenziell ausgeprägt ist.

Die Stimmen in den Aufnahmen sind als Ganzes zu bewerten, stimmliche ‚Ausreißer‘ (einmalige, kurzzeitig auftretende Merkmale) sollen nicht berücksichtigt werden. Es sind auch bearbeitete Stimmen dabei, was sich im populärmusikalischen Bereich nicht vermeiden lässt. Hier kann i.d.R. keine Unterscheidung getroffen werden, welcher Effekt technisch und welcher natürlich produziert ist.

Fast alle Merkmale sind dem Sprechausdrucks katalog von Ines Bose entnommen und werden auch so definiert. Zusätzliche Hinweise zu den einzelnen Merkmalen:

- Das Merkmal weich-hart tritt absichtlich doppelt auf, einmal als generelle Beschreibung, einmal bezogen auf die Einsätze.
- Gespannt-ungespannt kann mit dem Kommentar versehen werden, ob es sich z.B. eher um artikulatorische Spannung oder einen gepressten Stimmklang handelt etc.
- Dumpf-schrill bleiben absichtlich undefiniert; Kommentare erbeten, wenn es angekreuzt wird, d.h. wie es in dem speziellen Fall verstanden wird
- die Geräuschanteile können mit zusätzlichen Kommentaren versehen werden, falls das Merkmal nicht durchgängig, sondern z.B. bei Ein- und Absätzen auftritt
- Das Merkmal Sprechen und Rufen kann auch gleichermaßen auftreten.
- bei Artikulation kann das Auffällige kommentiert werden, z.B. Dialekt, besondere Ausspracheauffälligkeiten etc.; präzise-unpräzise trifft hier oft das Markante nicht
- Die Modulationen des Klanges beschreiben, ob sich der Stimmklang im Lied verändert, hinsichtlich seiner Häufigkeit in der Zeit, seines Umfanges und seiner Form (d.h. der ‚künstlerische‘ Umgang mit dem Stimmklang wird beschrieben)
- Die Modulationen der Töne beziehen sich auf den zeitlichen Tonhöhenverlauf *und* den Umfang; es wird eigentlich die Melodie bewertet, d.h. ob diese abwechslungsreich und/oder anspruchsvoll hinsichtlich sängerisch-technischen Ansprüchen ist. (Die Tonhöhenveränderungen sprunghaft-gleitend werden gesondert abgefragt.)

NN (Spalte): repräsentiert eine Ausweichkategorie, die gewählt werden kann, wenn eine Entscheidung nicht möglich ist, z.B. wenn dieses Merkmal nicht auffällig oder charakteristisch ist, es sich im neutralen/mittleren Raum bewegt, es ungeeignet zur Beschreibung der Stimme ist etc. (Gründe können bei Bedarf angegeben werden.)

Kommentar (Spalte): Kann enthalten, ob ein Merkmal evtl. nur manchmal auftritt oder wechselt; weiteres wurde bereits oben bei den Merkmalen benannt.

Kommentar (Generell): Hier bitte angeben, ob Merkmale fehlen, oder im spezifischen Fall anders hätten benannt werden müssen. Auch auffällige spezifische gesangstechnische Kommentare sind willkommen, wie Register (Belting etc.) oder Effekte (Twang, Distortion etc.).

Tabelle 39: Liste der Merkmale mit Prozent Nennungen, davon korrekten Antworten und Nachfragen (= unklar).

Merkmal	Nennungen %	korrekt %	unklar %
geknarrt	21,7391304	20	8,69565217
behaucht	21,7391304	40	13,0434783
Vibrato	26,0869565	83,3333333	13,0434783
Tonhöhenveränderungen	30,4347826	71,4285714	4,34782609
rau	30,4347826	85,7142857	0
Lautheit	34,7826087	75	4,34782609
Einsätze	34,7826087	75	21,7391304
Faukale Distanz	43,4782609	63,6363636	34,7826087
Lautdauer	43,4782609	90	0
Klangfarbe	47,826087	81,8181818	0
Resonanz	47,826087	81,8181818	0
Phonationsform	52,173913	66,6666667	0
Stimmage	52,173913	75	0
nasal	56,5217391	46,1538462	4,34782609
Artikulation	56,5217391	69,2307692	0
Klangfülle	60,8695652	57,1428571	0
Stimmklang	60,8695652	57,1428571	0
Modulationen Töne	65,2173913	64,2857143	21,7391304
Spannung	69,5652174	75	17,3913043
Modulationen Klang	73,9130435	75	17,3913043

Tabelle 40: Rangfolgen der abgelehnten Merkmale.

	Rang 1	Rang 2	Rang 3
Stimmage	8	2	2
Klangfarbe	9	0	2
Lautheit	0	0	8
Spannung	4	6	6
Klangfülle	8	5	1
Faukale Distanz	2	6	2
Stimmklang	2	9	3
Resonanz	8	2	0
behaucht	3	1	1
rau	3	1	3
geknarrt	3	1	1
nasal	11	2	0
Lautdauer	3	4	3
Tonhöhenveränderungen	1	6	0
Phonationsform	1	2	9
Vibrato	3	3	0
Einsätze	2	5	1
Artikulation	3	6	3
Modulationen Klang	9	7	1
Modulationen Töne	8	4	3

Tabelle 41: Experteneinschätzung (Median der drei Experten).

	Stimmlage	Klangfarbe	Lautheit	Spannung	Klangfülle	Faukale Distanz	Stimmklang	Resonanz	behaucht	rau
01_med	2	2	1	1	1	2	2	2		1
02_med	2	2	1	1	2	2	2	2		
03_med	2	2	1	2	2	2	1	2		1
04_med	2	2	1	1	1	1,5	2	2	1	
05_med	1	1	2	2	1	1	1	1		1
06_med	2	1	2	1	2	2	1	1	1	
07_1_med	2	2	1	2	2	2	2	1,5	1	1
07_2_med	2	2	1	1	2	2	2	2		1
08_med	2	2	1	1	1	1	1	2		
09_med	2	2	2	2	2	2	1	2	1	1
10_1_med	2	2	1	1	1	2	2	2		1
10_2_med	1	1	1	2	1	1	1	1		1
11_med	2	2	1	1	2	2	2	2		1
12_med	2	2	1	2	1	2	1	2		1
13_med	2	2	1	1	1	2	2	2	1	
14_med	2	2	1	1	1	1	1	2	1	
15_med	2	1	1	1	1	1,5	2	2		1
16_med	1,5	1	2	1	1	1	1	1	1	1
17_med	2	2	1	1	2	2	2	2		1
18_1_med	2	2	2	2	2	1,5	1	1,5		
18_2_med	2	2	1	1	2	2	2	2	1	
19_med	2	2	2	2	1	2	1	2	1	1
20_med	2	2	1	1	2	2	2	2	1	1
	geknarrt	nasal	Lautdauer	Tonhöhen- veränderungen	Phonationsform	Vibrato	Einsätze	Artikulation	Modulationen Klang	Modulationen Töne
01_med	1		1	1	2	1	2	2	1	2
02_med			1,5	1	2	1	2	1	1	1
03_med	1	1	2	2	1	2	1	2	1	2
04_med			1	1	2		2	1	2	2
05_med			1	2	1	1,5	1	1	1	2
06_med		1	1,5	1,5	1,5	2	1	2	1	1
07_1_med	1	1	1	2	1	1	2	2	1	1
07_2_med	1		1	2	2		2	1	1	1
08_med			1	1	1,5	1	1	1	2	2
09_med			1	2	1	1	1	1	1	1
10_1_med	1		1	1	1		1	1	1	2
10_2_med	1	1	1	2	1		1	2	1	2
11_med			1	1	2		2	1	1	2
12_med		1	1	2	1	1	1	1	1	2
13_med	1	1	1	1	2		2	1	2	2
14_med		1	1	1	2	2	1	1	2	2
15_med	1		1	1	2	1	2	1	2	2
16_med		1	1	2	2	2	1	2	2	2
17_med	1		2	1	2	1	2	1	2	2
18_1_med		1	1	2	1		1	2	1	2
18_2_med			1	1	2		2	2	1	2
19_med		1	1	1	1	2	1	1	1	1,5
20_med	1	1	2	2	1,5		2	1,5	2	2

In 27 % der zu bewertenden Merkmale (insges. N = 460) stimmten nicht alle drei Experten überein, entweder durch die abweichende Bewertung eines Experten oder durch die mögliche Angabe, dass das Merkmal bei diesem Sänger schwer zu bewerten war. In nur 15 Merkmalen waren nur zwei Ratings vorhanden, die sich zudem widersprachen (in der Tabelle am Wert 1,5 zu erkennen). Die fehlenden Werte der unipolaren Merkmale zeigen, dass das Merkmal laut Experten nicht in der Stimme vorhanden war.

8.3 Gruppen-Studie Kapitel 5

Tabelle 42: Fragebogen-Rückseite mit zusätzlichen Fragen und Merkmalen.

	sehr gut		gar nicht	
Wie gut konnten Sie Ihren Eindruck von der Stimme mit den gegebenen Merkmalen wiedergeben?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Fehlte Ihnen etwas?	<input type="checkbox"/> nein	<input type="checkbox"/> ja, und zwar _____		
<hr/>				
Kennen Sie das Lied?	<input type="checkbox"/> nein	<input type="checkbox"/> ja		
Kennen Sie den Sänger/ die Sängerin?	<input type="checkbox"/> nein	<input type="checkbox"/> ja, er/sie heißt _____		
<hr/>				
	gefällt mir		gefällt mir nicht	
Wie gefällt Ihnen das Lied?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Wie gefällt Ihnen der/die Sänger/in im Ganzen?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Wie gefällt Ihnen der Text?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Wie gefällt Ihnen der Musikstil generell?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

8.3.1 Abschlussbefragung

Wie empfanden Sie die Beschreibung der Stimmen anhand der einzelnen Merkmale?
Bitte geben Sie die Schwierigkeit für jedes Merkmal an.

Das Merkmal ist...	einfach		schwierig		Warum fiel Ihnen die Beschreibung schwer?
tief - hoch	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
laut - leise	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
klangvoll - klangarm	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
dunkel - hell	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
weich - hart	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
gespannt - ungespannt	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
rau/kratzig	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
glatt	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
dumpf/weit	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
quäkig/nasal	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
locker - gepresst	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
Tonhöhenverlauf: sprunghaft - gleitend	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
Artikulation: markant - schlicht	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
Singen erinnert an: Sprechen - Rufen	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
Ausdruck: wechselhaft - gleichförmig	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	

8.3.2 Übersicht der Kommentare zu den Schwierigkeiten mit einzelnen Merkmalen

Spannung und gepresst: 13 Kommentare weisen direkt darauf hin, dass die Begriffe gespannt-ungespannt unklar seien. Rückfragen umfassen vergleichbare Begriffe wie 'angespannt' oder 'entspannt'. In acht Fällen war vor allem der Unterschied zu gepresst-locker nicht klar. Bei gepresst-locker wird hingegen nur zweimal angemerkt, dass der Gegensatz zu gespannt-ungespannt nicht klar sei. 'Locker' als Gegenpol zu gepresst wird nicht weiter kommentiert.

Rauigkeit und Stimmklang: Während rau/kratzig keine Schwierigkeiten macht (auch die Kombination nicht), wird glatt neunmal kritisiert, vorwiegend aus dem Grund, dass diese Kombination nicht verständlich sei bzw. keinen Höreindruck widergebe. Der Stimmklang führt zwar zu keinen weiteren Problemen, allerdings wird angemerkt, dass die Beurteilung von hart-weich schwerfällt und innerhalb eines Liedes eine Stimme auch beides haben könne.

Phonationsform: Hier wird oft ein 'weder noch' angegeben, dass die Übergänge fließend seien, sie sich gegenseitig nicht ausschließen würden und der Gegensatz nicht eindeutig sei. Es wird auch ein besseres Attribut als 'Rufen' verlangt, interessant ist auch die Anmerkung 'Sprechgesang ist Gesang, Rufen ist kein Gesang'.

Artikulation: Die meisten Kommentare sagen, dass markant und schlicht keine Gegensätze seien und beide gleichzeitig zutreffen können. Auch wird darauf hingewiesen, dass eine Kategorie 'Genauigkeit der Artikulation' gefehlt habe und die Verständlichkeit des Textes evaluiert werden will (z.B. mit dem Merkmal 'vernuschelt').

Resonanz: Hier kritisieren einige Teilnehmer eine mangelnde Übereinstimmung der Begriffe: dumpf und weit gehören für einige zu verschiedenen Kategorien, außerdem sei die Definition der einzelnen Begriffe unklar. 'Weit' sei kein Begriff, mit dem eine Stimme beschrieben würde, oder man wisse nicht, wie eine 'dumpfe' Stimme klänge. Allerdings ist anzumerken, dass es nur vier direkte Kommentare zu dem Begriff dumpf gibt. Im Vergleich dazu funktionieren quäkig/nasal sehr gut, hier wurden keine Probleme in der Kombination gesehen.

Klangfarbe: Das Merkmal Klangfarbe wird sehr oft in Bezug zur Stimmlage gesetzt und gesagt, es sei schwierig davon abzugrenzen, da wie erwartet hoch mit hell und tief mit dunkel in Verbindung gebracht werden.

Klangfülle: Hier fehlte den Teilnehmern manchmal die Vergleichsebene, d.h. sie waren sich unsicher, ob es im Vergleich zu den Instrumenten gesehen werden solle, oder zu anderen Stimmen. Klangarm wurde als geschmacksabhängig bezeichnet, ein interessanter, aber in der Häufigkeit zu vernachlässigender Kommentar.

Lautheit: Obwohl die Schwierigkeit sich hier in Grenzen hielt, wurde wie erwartet angemerkt, dass dies von der Aufnahmequalität abhinge. Allerdings wurde zusätzlich gesagt, dass dies in Relation zu den Instrumenten oder ähnlichem eingeschätzt wird (vgl. Klangfülle). Das heißt, ein Vergleichspunkt (ein Ankerpunkt) ist hier anscheinend sehr wichtig.

Stimmlage: Aus den Anmerkungen der Teilnehmer wird ersichtlich, dass deutlich gemacht werden muss, ob es sich allgemein auf eine hohe Stimme bezieht, oder auf die Stimmlage, die im Moment genutzt wird, z.B. 'die singt mir zu hoch'. Auch wird angemerkt, dass diese in einem Stück variiert. Aus den Kommentaren wird deutlich, dass eher die Tonlage eingeschätzt wird, auf der die Person singt, was auch leichter einzuschätzen ist. Für eine allgemeine Einschätzung der Stimmlage wurden auch die Ankerpunkte vermisst, d.h. eine Vergleichsebene.

Ausdruck: Auch hier wurde angemerkt, dass der Ausdruck in einem Stück wechselt.

Es gab kaum Anmerkungen zu den Tonhöhenveränderungen.

8.4 Online-Studie Kapitel 6

Einführungstext

Online-Studie zur Stimmästhetik: Vielen Dank für Ihr Interesse an dieser Online-Studie, mit der wir die Wirkung von Singstimmen untersuchen. Sie werden im Folgenden Pop-Stimmen hören und nach Ihrem Gefallen sowie deren Wirkung auf Sie befragt. Außerdem sollen Sie die Stimmen anhand vorgegebener Merkmale beschreiben. Es ist kein Vorwissen nötig - es geht um Ihre Wahrnehmung und Bewertung. Außerdem werden Ihnen zu Beginn ein paar Fragen zu Ihrer Person und Ihrer musikalischen Vorbildung gestellt. Sie werden gebeten, insgesamt sechs Stimmen zu bewerten, was ungefähr 25 Minuten dauert. Achten Sie auf eine gute Sound-Wiedergabe, zum Beispiel über Kopfhörer.

Tabelle 43: Deskriptive Statistik für die Ratings Gefallen Stimme (gef_st), Gefallen Song (gef_so) und die Emotionsratings berührt, fasziniert, entspannt, energetisiert, frohlich, traurig, gereizt für alle Sänger (Anfangsbuchstabe, d_ = Dylan).

Statistiken für Dylan										
		d_gef_st	d_gef_so	d_e_ber	d_e_fasz	d_e_ent	d_e_ener	d_e_froe	d_e_trau	d_e_ger
N	Gültig	216	216	216	216	216	216	216	216	216
	Fehlend	0	0	0	0	0	0	0	0	0
Mittelwert		2,76	2,69	3,07	2,75	3,07	2,01	2,27	2,68	1,95
Standardfehler des Mittelwerts		0,084	0,088	0,081	0,078	0,08	0,062	0,062	0,076	0,074
Median		3	2	3	3	3	2	2	3	2
Modus		2	2	4	2	4	2	2	3	1
Std.-Abweichung		1,241	1,295	1,185	1,143	1,169	0,912	0,917	1,111	1,09
Minimum		1	1	1	1	1	1	1	1	1
Maximum		5	5	5	5	5	5	5	5	5
Perzentile	25	2	2	2	2	2	1	2	2	1
	50	3	2	3	3	3	2	2	3	2
	75	4	4	4	4	4	3	3	4	3
Statistiken für Presley										
		p_gef_st	p_gef_so	p_e_ber	p_e_fasz	p_e_ent	p_e_energ	p_e_froe	p_e_trau	p_e_ger
N	Gültig	216	216	216	216	216	216	216	216	216
	Fehlend	0	0	0	0	0	0	0	0	0
Mittelwert		2,06	2,28	3,49	3,09	3,63	1,89	2,2	2,83	1,62
Standardfehler des Mittelwerts		0,067	0,077	0,069	0,076	0,067	0,066	0,066	0,073	0,067
Median		2	2	4	3	4	2	2	3	1
Modus		2	2	3	3	4	1	2	3	1
Std.-Abweichung		0,987	1,137	1,007	1,119	0,985	0,966	0,966	1,07	0,986
Minimum		1	1	1	1	1	1	1	1	1
Maximum		5	5	5	5	5	5	5	5	5
Perzentile	25	1	1	3	2	3	1	1	2	1
	50	2	2	4	3	4	2	2	3	1
	75	3	3	4	4	4	2	3	4	2

Statistiken für Brown										
		b_gef_st	b_gef_so	b_e_ber	b_e_fasz	b_e_ent	b_e_ener	b_e_froe	b_e_trau	b_e_ger
N	Gültig	216	216	216	216	216	216	216	216	216
	Fehlend	0	0	0	0	0	0	0	0	0
Mittelwert		2,27	1,95	2,70	3,55	2,09	3,81	3,61	1,35	2,08
Standardfehler des Mittelwerts		0,079	0,076	0,078	0,075	0,069	0,071	0,076	0,044	0,075
Median		2,00	2,00	3,00	4,00	2,00	4,00	4,00	1,00	2,00
Modus		2	1	3	4	2	4	4	1	1
Std.-Abweichung		1,159	1,124	1,143	1,103	1,016	1,042	1,124	0,643	1,097
Minimum		1	1	1	1	1	1	1	1	1
Maximum		5	5	5	5	5	5	5	5	5
Perzentile	25	1,00	1,00	2,00	3,00	1,00	3,00	3,00	1,00	1,00
	50	2,00	2,00	3,00	4,00	2,00	4,00	4,00	1,00	2,00
	75	3,00	2,00	4,00	4,00	3,00	5,00	4,00	2,00	3,00
Statistiken für Turner										
		t_gef_st	t_gef_so	t_e_ber	t_e_fasz	t_e_ent	t_e_ener	t_e_froe	t_e_trau	t_e_ger
N	Gültig	216	216	216	216	216	216	216	216	216
	Fehlend	0	0	0	0	0	0	0	0	0
Mittelwert		2,24	2,44	3,02	3,31	2,35	3,24	2,96	1,85	1,89
Standardfehler des Mittelwerts		0,073	0,085	0,081	0,080	0,068	0,074	0,070	0,063	0,068
Median		2,00	2,00	3,00	4,00	2,00	3,00	3,00	2,00	2,00
Modus		2	1	4	4	2	4	3	1	1
Std.-Abweichung		1,073	1,248	1,187	1,178	0,996	1,085	1,031	0,928	1,004
Minimum		1	1	1	1	1	1	1	1	1
Maximum		5	5	5	5	5	5	5	5	5
Perzentile	25	1,00	1,00	2,00	2,00	2,00	2,00	2,00	1,00	1,00
	50	2,00	2,00	3,00	4,00	2,00	3,00	3,00	2,00	2,00
	75	3,00	3,00	4,00	4,00	3,00	4,00	4,00	2,00	2,75

Statistiken für Jackson										
		j_gef_st	j_gef_so	j_e_ber	j_e_fasz	j_e_ent	j_e_ener	j_e_froe	j_e_trau	j_e_ger
N	Gültig	216	216	216	216	216	216	216	216	216
	Fehlend	0	0	0	0	0	0	0	0	0
Mittelwert		3,05	2,48	2,13	3,09	1,65	3,56	3,41	1,22	2,56
Standardfehler des Mittelwerts		0,087	0,084	0,069	0,080	0,060	0,083	0,078	0,036	0,084
Median		3,00	2,00	2,00	3,00	1,00	4,00	4,00	1,00	2,00
Modus		4	2	1	4	1	4	4	1	1 ^a
Std.-Abweichung		1,279	1,239	1,014	1,176	0,886	1,226	1,150	0,525	1,237
Minimum		1	1	1	1	1	1	1	1	1
Maximum		5	5	5	5	5	5	5	5	5
Perzentile	25	2,00	1,00	1,00	2,00	1,00	3,00	3,00	1,00	1,00
	50	3,00	2,00	2,00	3,00	1,00	4,00	4,00	1,00	2,00
	75	4,00	3,75	3,00	4,00	2,00	4,75	4,00	1,00	4,00
a. Mehrere Modi vorhanden. Der kleinste Wert wird angezeigt.										
Statistiken für Houston										
		h_gef_st	h_gef_so	h_e_ber	h_e_fasz	h_e_ent	h_e_ener	h_e_froe	h_e_trau	h_e_ger
N	Gültig	216	216	216	216	216	216	216	216	216
	Fehlend	0	0	0	0	0	0	0	0	0
Mittelwert		2,16	2,55	3,31	3,25	2,75	2,81	2,62	2,37	1,83
Standardfehler des Mittelwerts		0,075	0,090	0,080	0,083	0,068	0,079	0,068	0,070	0,074
Median		2,00	2,00	4,00	3,00	3,00	3,00	2,50	2,00	1,00
Modus		1	1	4	4	2	2	2	2 ^a	1
Std.-Abweichung		1,105	1,318	1,177	1,220	1,000	1,161	1,005	1,035	1,083
Minimum		1	1	1	1	1	1	1	1	1
Maximum		5	5	5	5	5	5	5	5	5
Perzentile	25	1,00	1,00	2,00	2,00	2,00	2,00	2,00	2,00	1,00
	50	2,00	2,00	4,00	3,00	3,00	3,00	2,50	2,00	1,00
	75	3,00	4,00	4,00	4,00	3,00	4,00	3,00	3,00	2,00
a. Mehrere Modi vorhanden. Der kleinste Wert wird angezeigt.										

Tabelle 44: Deskriptive Statistik der Ratings für die vokalen Merkmale Stimmlage, Rauigkeit, Spannung, Timbre, Tonhöhenveränderungen, Artikulationspräzision, Artikulationsbesonderheiten, Phonationsform, Ausdruck für alle Sänger.

Statistiken für Dylan										
		d_b_lage	d_b_rau	d_b_span	d_b_timb	d_b_tonv	d_b_artp	d_b_artg	d_b_sng	d_b_ausd
N	Gültig	216	216	216	216	216	216	216	216	216
	Fehlend	0	0	0	0	0	0	0	0	0
Mittelwert		3,22	3,13	3,35	2,64	3,65	4,38	2,52	3,45	4,11
Standardfehler des Mittelwerts		0,079	0,087	0,091	0,093	0,101	0,088	0,089	0,096	0,096
Median		3	3	3	2	4	5	2	3	5
Modus		3	3	3	2	5	5	2	3	5
Std.-Abweichung		1,155	1,28	1,339	1,364	1,484	1,3	1,305	1,407	1,416
Minimum		1	1	1	1	1	1	1	1	1
Maximum		6	6	6	6	6	6	6	6	6
Perzentile	25	2	2	2	2	2	4	2	2	3
	50	3	3	3	2	4	5	2	3	5
	75	4	4	4	4	5	5	3	5	5
Statistiken für Presley										
		p_b_lage	p_b_rau	p_b_span	p_b_timb	p_b_tonv	p_b_artp	p_b_artg	p_b_sng	p_b_ausd
N	Gültig	216	216	216	216	216	216	216	216	216
	Fehlend	0	0	0	0	0	0	0	0	0
Mittelwert		2	4,9	4,62	4,76	5,37	2,48	4,11	1,96	4,84
Standardfehler des Mittelwerts		0,057	0,079	0,069	0,056	0,053	0,084	0,089	0,078	0,08
Median		2	5	5	5	5	2	4	2	5
Modus		2	5	5	5	6	2	5	1	5
Std.-Abweichung		0,832	1,163	1,019	0,821	0,778	1,227	1,306	1,142	1,18
Minimum		1	1	2	1	2	1	1	1	1
Maximum		5	6	6	6	6	6	6	6	6
Perzentile	25	1	4,25	4	4	5	2	3	1	4
	50	2	5	5	5	5	2	4	2	5
	75	2	6	5	5	6	3	5	2	6
Statistiken für Brown										
		b_b_lage	b_b_rau	b_b_span	b_b_timb	b_b_tonv	b_b_artp	b_b_artg	b_b_sng	b_b_ausd
N	Gültig	216	216	216	216	216	216	216	216	216
	Fehlend	0	0	0	0	0	0	0	0	0
Mittelwert		4,07	1,56	2,15	2,91	2,82	3,59	2,63	2,72	2,31
Standardfehler des Mittelwerts		0,083	0,054	0,064	0,076	0,097	0,087	0,084	0,086	0,081
Median		4,00	1,00	2,00	3,00	2,00	4,00	2,00	3,00	2,00
Modus		5	1	2	3	2	5	2	2	2
Std.-Abweichung		1,225	0,793	0,940	1,111	1,424	1,283	1,229	1,261	1,197
Minimum		1	1	1	1	1	1	1	1	1
Maximum		6	5	5	6	6	6	6	6	6
Perzentile	25	3,00	1,00	2,00	2,00	2,00	3,00	2,00	2,00	1,00
	50	4,00	1,00	2,00	3,00	2,00	4,00	2,00	3,00	2,00
	75	5,00	2,00	3,00	4,00	4,00	5,00	3,00	4,00	3,00

Statistiken für Turner										
		t_b_lage	t_b_rau	t_b_span	t_b_timb	t_b_tonv	t_b_artp	t_b_artg	t_b_sng	t_b_ausd
N	Gültig	216	216	216	216	216	216	216	216	216
	Fehlend	0	0	0	0	0	0	0	0	0
Mittelwert		3,42	2,07	2,43	3,03	3,09	2,87	3,22	2,17	2,81
Standardfehler des Mittelwerts		0,085	0,066	0,059	0,082	0,082	0,076	0,088	0,072	0,081
Median		3,00	2,00	2,00	3,00	3,00	3,00	3,00	2,00	3,00
Modus		3	2	2	2	2	2	3	2	2
Std.-Abweichung		1,248	0,974	0,870	1,210	1,211	1,121	1,288	1,065	1,184
Minimum		1	1	1	1	1	1	1	1	1
Maximum		6	6	5	6	6	6	6	6	6
Perzentile	25	2,00	1,00	2,00	2,00	2,00	2,00	2,00	1,00	2,00
	50	3,00	2,00	2,00	3,00	3,00	3,00	3,00	2,00	3,00
	75	4,00	2,00	3,00	4,00	4,00	4,00	4,00	3,00	3,00
Statistiken für Jackson										
		j_b_lage	j_b_rau	j_b_span	j_b_timb	j_b_tonv	j_b_artp	j_b_artg	j_b_sng	j_b_ausd
N	Gültig	216	216	216	216	216	216	216	216	216
	Fehlend	0	0	0	0	0	0	0	0	0
Mittelwert		4,31	1,39	1,97	2,15	2,67	3,36	2,86	2,84	2,82
Standardfehler des Mittelwerts		0,086	0,043	0,062	0,076	0,089	0,090	0,091	0,089	0,099
Median		5,00	1,00	2,00	2,00	2,00	3,00	3,00	3,00	2,00
Modus		5	1	2	2	2	2	2	2	2
Std.-Abweichung		1,261	0,637	0,907	1,119	1,301	1,329	1,330	1,307	1,450
Minimum		1	1	1	1	1	1	1	1	1
Maximum		6	4	5	6	6	6	6	6	6
Perzentile	25	3,25	1,00	1,00	1,00	2,00	2,00	2,00	2,00	2,00
	50	5,00	1,00	2,00	2,00	2,00	3,00	3,00	3,00	2,00
	75	5,00	2,00	2,75	3,00	3,00	4,00	4,00	4,00	4,00
Statistiken für Houston										
		h_b_lage	h_b_rau	h_b_span	h_b_timb	h_b_tonv	h_b_artp	h_b_artg	h_b_sng	h_b_ausd
N	Gültig	216	216	216	216	216	216	216	216	216
	Fehlend	0	0	0	0	0	0	0	0	0
Mittelwert		3,85	4,73	3,44	3,47	4,43	2,09	4,29	1,30	4,15
Standardfehler des Mittelwerts		0,077	0,068	0,083	0,072	0,087	0,072	0,090	0,039	0,097
Median		4,00	5,00	3,00	4,00	5,00	2,00	5,00	1,00	5,00
Modus		4	5	3	4	5	2	5	1	5
Std.-Abweichung		1,136	0,994	1,222	1,056	1,277	1,057	1,316	0,568	1,424
Minimum		1	2	1	1	1	1	1	1	1
Maximum		6	6	6	6	6	6	6	4	6
Perzentile	25	3,00	4,00	3,00	3,00	4,00	1,00	3,00	1,00	3,00
	50	4,00	5,00	3,00	4,00	5,00	2,00	5,00	1,00	5,00
	75	5,00	5,00	4,00	4,00	5,00	2,00	5,00	2,00	5,00

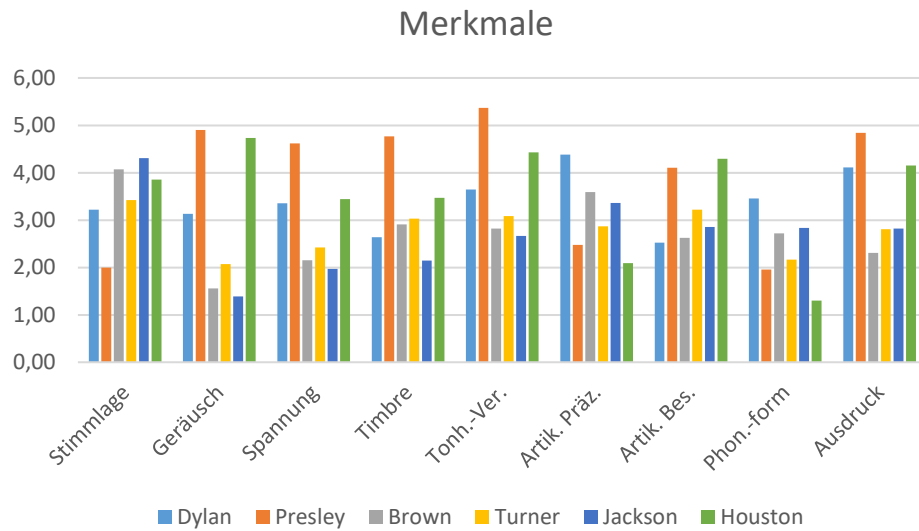


Abbildung 16: Balkendiagramm der Mittelwerte der Merkmale für jeden Sänger, bewertet auf einer 6-stufigen Skala.

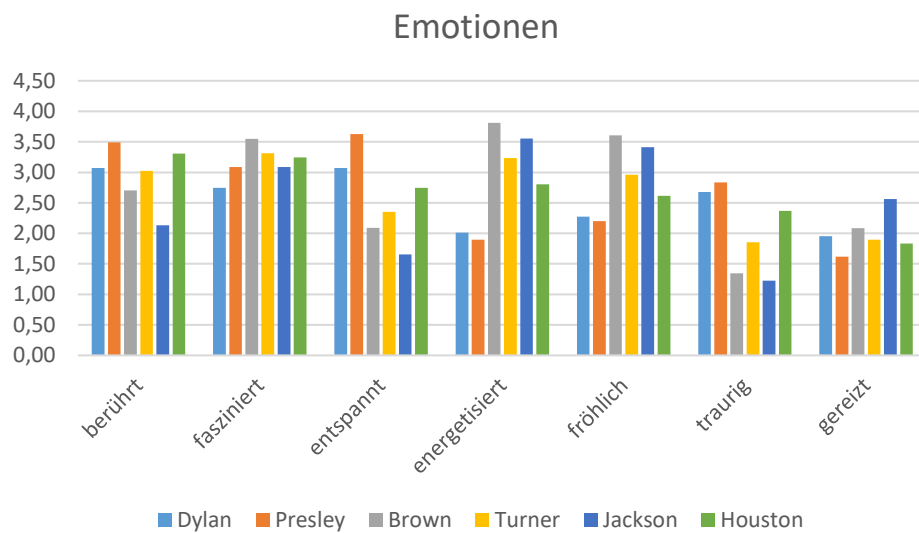


Abbildung 17: Balkendiagramm der Mittelwerte der Emotionen für jeden Sänger, bewertet auf einer 5-stufigen Skala.

9 Literaturverzeichnis

- Ackermann, T. (2014, 4. September). *Die Bedeutung des Musikhörens bei der Identitätskonstruktion im Jugendalter. Masterarbeit für den Master of Arts*. Master Thesis. Westfälische Wilhelms-Universität, Münster. Zugriff am 02.05.2019. Verfügbar unter <https://www.uni-muenster.de/Musiktherapie/Literaturdienst/bestellservice.html>
- Ackermann, T. (2018, 26. September). „Disliked Music“. *Merkmale, Gründe und Funktionen abgelehnter Musik*. Dissertation. Universität Kassel, Kassel.
- Anders, L. C. (2012). Perzeptive Beurteilung der Stimme. In W. Seidner & T. Nawka (Hrsg.), *Handreichungen zur Stimmdiagnostik. Aus der Praxis für die Praxis* (S. 29–57). Berlin: XION GmbH.
- Anders, L. C., Hollien, H., Hurme, P., Sonninen, A. & Wendler, J. (1988). Perception of Hoarseness by Several Classes of Listeners. *Folia Phoniatrica et Logopaedica*, 40 (2), 91–100. <https://doi.org/10.1159/000265889>
- Babel, M., McGuire, G. & King, J. (2014). Towards a more nuanced view of vocal attractiveness. *PLoS one*, 9 (2), e88616. <https://doi.org/10.1371/journal.pone.0088616>
- Bacilly, B. d. (1668). *Remarques curieuses sur l'art de bien chanter*. Paris: Robert III Ballard et Pierre Bienfait.
- Ball, M. J., Esling, J. & Dickson, C. (1995). The VoQS System for the Transcription of Voice Quality. *Journal of the International Phonetic Association*, 25 (2), 71–80. <https://doi.org/10.1017/S0025100300005181>
- Banse, R. & Scherer, K. R. (1996). Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology*, 70 (3), 614–636. <https://doi.org/10.1037//0022-3514.70.3.614>
- Bänziger, T., Patel, S. & Scherer, K. R. (2014). The Role of Perceived Voice and Speech Characteristics in Vocal Emotion Communication. *Journal of Nonverbal Behavior*, 38 (1), 31–52. <https://doi.org/10.1007/s10919-013-0165-x>
- Bassich, C. J. & Ludlow, C. L. (1986). The use of perceptual methods by new clinicians for assessing voice quality. *The Journal of speech and hearing disorders*, 51 (2), 125–133.
- Berg, M., Fuchs, M., Wirkner, K., Loeffler, M., Engel, C. & Berger, T. (2017). The Speaking Voice in the General Population. Normative Data and Associations to Sociodemographic and Lifestyle Factors. *Journal of voice : official journal of the Voice Foundation*, 31 (2), 257.e13-257.e24. <https://doi.org/10.1016/j.jvoice.2016.06.001>
- Berli, O. (2014). *Grenzenlos guter Geschmack. Die feinen Unterschiede des Musikhörens* (Kultur und soziale Praxis). Bielefeld: Transcript.
- Biehle, H. (1955). *Stimmkunde. Für Beruf, Kunst und Heilzwecke* (Sammlung Göschen, Bd. 60). Berlin: Walter de Gruyter.
- Boersma, P. & Weenink, D. (2017). Praat. doing phonetics by computer (Version 6.0.36) [Computer software]. Verfügbar unter <http://www.praat.org/>
- Bojack-Weber, R. (2012). *Singen in der Grundschule. Eine Untersuchung zur Singfähigkeit und zum Singverhalten von Grundschulkindern* (Augsburger Schriften, Bd. 104). Zugl.: Freiburg, Pädag. Hochsch., Diss., 2011. Augsburg: Wißner.
- Bose, I. (2001). Methoden der Sprechausdrucksbeschreibung am Beispiel kindlicher Spielkommunikation. *Gesprächsforschung - Online-Zeitschrift zur verbalen Interaktion*, 2, 262–303. Zugriff am 10.08.2016. Verfügbar unter <http://www.gespraechsforschung-ozs.de/heft2001/ga-bose.pdf>
- Bose, I. (2003). *dóch da sín ja ' nur müster //*. *Kindlicher Sprechdruck im sozialen Rollenspiel* (Hallesche Schriften zur Sprechwissenschaft und Phonetik, Bd. 9). Frankfurt am Main: Peter Lang.
- Bose, I. (2010). Stimmlich-artikulatorischer Ausdruck und Sprache. In Deppermann, A., Linke, A. (Hrsg.), *Sprache intermedial. Stimme und Schrift, Bild und Ton* (S. 29–68). Berlin, New York: De Gruyter.
- Bose, I., Hirschfeld, U. & Neuber, B. (2016). *Einführung in die Sprechwissenschaft. Phonetik, Rhetorik, Sprechkunst* (Narr Studienbücher, 2., überarbeitete und erweiterte Auflage). Tübingen: Narr Francke Attempto.
- Brandenburg, D. (2016). Björling, Jussi. In A.-C. Mecke, M. Pfeleiderer, B. Richter & T. Seedorf (Hrsg.), *Lexikon der Gesangsstimme. Geschichte, wissenschaftliche Grundlagen, Gesangstechniken, Interpreten* (S. 86–87). Laaber: Laaber-Verlag.
- Bryson, B. (1996). "Anything But Heavy Metal". Symbolic Exclusion and Musical Dislikes. *American Sociological Review*, 61 (5), 884–899.

- Chapman, J. L. (2017). *Singing and teaching singing. A holistic approach to classical voice* (Third edition). San Diego, CA: Plural Publishing.
- Chin, T. & Rickard, N. S. (2012). The Music USE (MUSE) Questionnaire. An Instrument to Measure Engagement in Music. *Music perception*, 29 (4), 429–446. <https://doi.org/10.1525/mp.2012.29.4.429>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum. Retrieved from <http://search.ebscohost.com/login.aspx?direct=true&scope=site&db=nlebk&db=nlabk&AN=582094>
- Coutinho, E. & Scherer, K. R. (2017). Introducing the GENEVA Music-Induced Affect Checklist (GEMIAC). A brief instrument for the rapid assessment of musically induced emotions. *Music perception*, 33 (4), 371–386. Verfügbar unter <https://doi.org/10.1525/MP.2017.34.4.371>
- Denzin, N. K. (1994). *Handbook of qualitative research* (4. Aufl.). Thousand Oaks, CA: Sage Publ.
- Eckert, H. & Laver, J. (1994). *Menschen und ihre Stimmen. Aspekte der vokalen Kommunikation*. Weinheim: Beltz/Psychologie Verlags Union.
- Ekholm, E., Papagiannis, G. C. & Chagnon, F. P. (1998). Relating objective measurements to expert evaluation of voice quality in Western classical singing. Critical perceptual parameters. *Journal of voice : official journal of the Voice Foundation*, 12 (2), 182–196.
- Elvers, P., Omigie, D., Fuhrmann, W. & Fischinger, T. (2015). Exploring the musical taste of expert listeners. Musicology students reveal tendency toward omnivorous taste. *Frontiers in psychology*, 6, 1252. <https://doi.org/10.3389/fpsyg.2015.01252>
- Evans, P. & Schubert, E. (2008). Relationships between expressed and felt emotions in music. *Musicae Scientiae*, 12 (1), 75–99. <https://doi.org/10.1177/102986490801200105>
- Flick, U. (2007). *Qualitative Sozialforschung. Eine Einführung* (Rororo, Bd. 55694, Vollst. überarb. und erw. Neuausg.). Reinbek bei Hamburg: Rowohlt Taschenbuch.
- Foehr-Janssens, Y. (2008). Über die "körperliche Beredsamkeit" in der römischen Rhetorik. In H. K. Geissner (Hrsg.), *Das Phänomen Stimme. Natürliche Veranlagung oder kulturelle Formung* (S. 125–131). 6. Internationale Stuttgarter Stimmtage 2006. St. Ingbert: Röhrig Univ.-Verl.
- Freytag, M. (2003). *Stimmbildung in der Populärmusik. Ein Lehr- und Arbeitsbuch mit CD*. Berlin: Henschel.
- Friedrich, G. & Dejonckere, P. H. (2005). The voice evaluation protocol of the European Laryngological Society (ELS) - first results of a multicenter study. *Laryngo-Rhino-Otologie*, 84 (10), 744–752. <https://doi.org/10.1055/s-2005-861450>
- Gabrielsson, A. (2002). Emotion perceived and emotion felt. Same or different? *Musicae Scientiae*, 5 (1_suppl), 123–147. <https://doi.org/10.1177/102986490200505105>
- Gerratt, B. R., Kreiman, J., Antonanzas-Barroso, N. & Berke, G. S. (1993). Comparing internal and external standards in voice quality judgments. *Journal of speech and hearing research*, 36 (1), 14–20.
- Gescheider, G. A. (1997). *Psychophysics: The Fundamentals*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Goy, H., Fernandes, D. N., Pichora-Fuller, M. K. & van Lieshout, P. (2013). Normative voice data for younger and older adults. *Journal of voice : official journal of the Voice Foundation*, 27 (5), 545–555. <https://doi.org/10.1016/j.jvoice.2013.03.002>
- Goy, H., Pichora-Fuller, M. K. & van Lieshout, P. (2016). Effects of age on speech and voice quality ratings. *The Journal of the Acoustical Society of America*, 139 (4), 1648. <https://doi.org/10.1121/1.4945094>
- Grawunder, S., Bose, I., Hertha, B., Trauselt, F. & Anders, L. C. (2006). Perceptive and acoustic measurement of average speaking pitch of female and male speakers in German radio news. *INTERSPEECH*, 885–888.
- Greasley, A., Lamont, A. & Sloboda, J. (2013). Exploring Musical Preferences. An In-Depth Qualitative Study of Adults' Liking for Music in Their Personal Collections. *Qualitative Research in Psychology*, 10 (4), 402–427. <https://doi.org/10.1080/14780887.2011.647259>
- Hähnel, T. (2015). Was ist populärer Gesang? Zur Terminologie vokaler Gestaltungsmittel in populärer Musik. In M. Pfeleiderer, T. Hähnel, K. Horn & C. Bielefeldt (Hrsg.), *Stimme, Kultur, Identität. Vokaler Ausdruck in der populären Musik der USA, 1900 - 1960* (Texte zur populären Musik, Bd. 8, S. 53–74). Bielefeld: Transcript.
- Hähnel, T. (2016). Jackson, Mahalia. In A.-C. Mecke, M. Pfeleiderer, B. Richter & T. Seedorf (Hrsg.), *Lexikon der Gesangsstimme. Geschichte, wissenschaftliche Grundlagen, Gesangstechniken, Interpretieren* (S. 302–303). Laaber: Laaber-Verlag.
- Hähnel, T., Marx, T. & Pfeleiderer, M. (2014). Methoden zur Analyse der vokalen Gestaltung populärer Musik. *Samples - Online-Publikationen der Gesellschaft für Populärmusikforschung*, 12. Zugriff am 29.12.2017. Verfügbar unter www.gfpm-samples.de/Samples12/haehneletal.pdf

- Hancock, A. B., Krissing, J. & Owen, K. (2011). Voice perceptions and quality of life of transgender people. *Journal of voice : official journal of the Voice Foundation*, 25 (5), 553–558.
<https://doi.org/10.1016/j.jvoice.2010.07.013>
- Hellbernd, N. & Sammler, D. (2016). Prosody conveys speaker's intentions. Acoustic cues for speech act perception. *Journal of Memory and Language*, 88, 70–86. <https://doi.org/10.1016/j.jml.2016.01.001>
- Henrich, N., Bezard, P., Expert, R., Garnier, M., Guerin, C., Pillot, C. et al. (2008). Towards a common terminology to describe voice quality in western lyrical singing: Contribution of a multidisciplinary research group. *Journal of interdisciplinary music studies*, 2 (1&2), 71–93.
- Himonides, E. (2009). Mapping a beautiful voice. Theoretical considerations. *Journal of Music, Technology and Education*, 2 (1), 25–54. <https://doi.org/10.1386/jmte.2.1.25/1>
- Hirano, M. (Hrsg.). (1981). *Clinical examination of voice* (Disorders of human communication, Bd. 5). Wien: Springer.
- Hollien, H., Gelfer, M. P. & Carlson, T. (1991). Listening preferences for voice types as a function of age. *Journal of communication disorders*, 24 (2), 157–171.
- Hollien, H. (2000). The concept of ideal voice quality. In R. D. Kent & M. J. Ball (Hrsg.), *Voice quality measurement* (S. 13–24). San Diego, CA: Singular Publ. Group.
- Juslin, P. N. & Laukka, P. (2003). Communication of emotions in vocal expression and music performance. Different channels, same code? *Psychological Bulletin*, 129 (5), 770–814. <https://doi.org/10.1037/0033-2909.129.5.770>
- Karpf, A. (2006a). *Frauen reden anders, Männer auch. Was die Stimme über unsere wahren Gefühle verrät*. Bergisch Gladbach: Ehrenwirth.
- Karpf, A. (2006b). *The human voice. The story of a remarkable talent*. London: Bloomsbury.
- Kempster, G. B., Gerratt, B. R., Verdolini Abbott, K., Barkmeier-Kraemer, J. & Hillman, R. E. (2009). Consensus Auditory-Perceptual Evaluation of Voice. Development of a Standardized Clinical Protocol. *American Journal of Speech-Language Pathology*, 18 (2), 124. [https://doi.org/10.1044/1058-0360\(2008/08-0017\)](https://doi.org/10.1044/1058-0360(2008/08-0017))
- Kesting, J. (Hrsg.). (2008a). *Die großen Sänger* (4 Bände). Hamburg: Hoffmann und Campe.
- Kesting, J. (2008b). Domingo, Plácido. In J. Kesting (Hrsg.), *Die großen Sänger* (Bd. 4, S. 1929–1941). Hamburg: Hoffmann und Campe.
- Kesting, J. (2008c). Pavarotti, Luciano. In J. Kesting (Hrsg.), *Die großen Sänger* (Bd. 4, S. 1920–1929). Hamburg: Hoffmann und Campe.
- Kesting, J. (11.10.2016). *Skizze zur Beurteilung der Gesangsstimme* (Email).
- Ketzmerick, B. (2007). *Zur auditiven und apparativen Charakterisierung von Stimmen* (Studientexte zur Sprachkommunikation, Bd. 47). Zugl.: Cottbus, Techn. Univ., Diss., 2007. Dresden: TUDpress Verl. der Wiss. Verfügbar unter http://deposit.d-nb.de/cgi-bin/dokserv?id=3058478&prov=M&dok_var=1&dok_ext=htm
- Klasmeyer, G. (1999). *Akustische Korrelate des stimmlich emotionalen Ausdrucks in der Lautsprache* (Forum phoneticum, Bd. 67). Zugl.: Berlin, Techn. Univ., Diss., 1998. Frankfurt am Main: Wiss. Buchh. Hector.
- Klasmeyer, G. & Sendlmeier, W. F. (2000). Voice and Emotional States. In R. D. Kent & M. J. Ball (Hrsg.), *Voice quality measurement* (S. 339–357). San Diego, CA: Singular Publ. Group.
- Koelsch, S. (2014). Brain correlates of music-evoked emotions. *Nature reviews. Neuroscience*, 15 (3), 170–180. <https://doi.org/10.1038/nrn3666>
- Kranich, W. (2014a). Überlegungen zur Beschreibung der Klangqualität gesprochener Sprache. *Sprechen*, 30 (57), 36–48.
- Kranich, W. (2014b). Untersuchungen zur Beschreibung der Klangqualität gesprochener Sprache. *Sprechen*, 31 (58), 21–38.
- Kranich, W. (2016). Die Betrachtung stimmlicher Merkmale in der antiken Rhetorik. *Sprechen*, 33 (62), 28–41.
- Kreiman, J. & Gerratt, B. R. (1998). Validity of rating scale measures of voice quality. *The Journal of the Acoustical Society of America*, 104 (3 Pt 1), 1598–1608.
- Kreiman, J., Gerratt, B. R., Kempster, G. B., Erman, A. & Berke, G. S. (1993). Perceptual evaluation of voice quality. Review, tutorial, and a framework for future research. *Journal of speech and hearing research*, 36 (1), 21–40.
- Kreiman, J., Gerratt, B. R. & Ito, M. (2007). When and why listeners disagree in voice quality assessment tasks. *The Journal of the Acoustical Society of America*, 122 (4), 2354–2364. <https://doi.org/10.1121/1.2770547>
- Latimerlo, G. & Popeil, L. (2012). *Sing anything. Mastering vocal styles*: G. Latimerlo and L. Popeil.

- Laver, J. (1980). *The phonetic description of voice quality* (Cambridge studies in linguistics, Bd. 31). Cambridge: Cambridge Univ. Press.
- Laver, J. (1994). *Principles of phonetics* (Cambridge textbooks in linguistics). Cambridge: Cambridge University Press.
- Laver, J., Wirz, S., Mackenzie, J. & Hiller, S. (1981). A perceptual protocol for the analysis of vocal profiles. *Edinburgh University Department of Linguistics Work in Progress*, 14, 139–155.
- Lehmann, C. (2016). Portamento. In A.-C. Mecke, M. Pfeleiderer, B. Richter & T. Seedorf (Hrsg.), *Lexikon der Gesangsstimme. Geschichte, wissenschaftliche Grundlagen, Gesangstechniken, Interpreten* (S. 481–483). Laaber: Laaber-Verlag.
- Mathelitsch, L. & Friedrich, G. (1995). *Die Stimme. Instrument für Sprache, Gesang und Gefühl*. Berlin, Heidelberg: Springer.
- Mathieson, L. (2001). *Greene and Mathieson's the Voice and its Disorders* (6. Aufl.). London, England and Philadelphia, PA: Whurr Publishers Ltd.
- Mayring, P. (2010). *Qualitative Inhaltsanalyse. Grundlagen und Techniken* (Pädagogik, 11., akt. und überarb. Aufl.). Weinheim u.a.: Beltz.
- Mayring, P. (2012). Qualitative Inhaltsanalyse. In U. Flick (Hrsg.), *Qualitative Forschung. Ein Handbuch* (Rororo, Bd. 55628, Orig.-Ausg., 9. Aufl., S. 468–475). Reinbek bei Hamburg: Rowohlt.
- McGlashan, J., Thuesen, M. A. & Sadolin, C. (2017). Overdrive and Edge as Refiners of "Belting"? An Empirical Study Qualifying and Categorizing "Belting" Based on Audio Perception, Laryngostroboscopic Imaging, Acoustics, LTAS, and EGG. *Journal of voice : official journal of the Voice Foundation*, 31 (3), 385.e11-385.e22. <https://doi.org/10.1016/j.jvoice.2016.09.006>
- Mecke, A.-C., Pfeleiderer, M., Richter, B. & Seedorf, T. (Hrsg.). (2016). *Lexikon der Gesangsstimme. Geschichte, wissenschaftliche Grundlagen, Gesangstechniken, Interpreten* (Instrumenten-Lexika, 5). Laaber: Laaber-Verlag.
- Merrill, J. (2017). Schoenberg's Pierrot Lunaire Revisited. Acceptance of Vocal Expression. *Acta Musicologica*, 89 (1), 95–117.
- Merrill, J. (2019). Perception und Rezeption des vokalen Ausdrucks im Grenzbereich von Singen und Sprechen. *Sprechen* (68).
- Merrill, J. & Larrouy-Maestri, P. (2017). Vocal Features of Song and Speech: Insights from Schoenberg's Pierrot Lunaire. *Frontiers in psychology*, 8, 1108. <https://doi.org/10.3389/fpsyg.2017.01108>
- Merton, R. K. & Kendall, P. L. (1979). Das fokussierte Interview. In C. Hopf, A. H. Barton & F. Büchner (Hrsg.), *Qualitative Sozialforschung* (Sozialwissenschaften, 1. Aufl., S. 171–204). Stuttgart: Klett-Cotta.
- Misoch, S. (2015). *Qualitative Interviews*. Berlin: De Gruyter Oldenbourg. <https://doi.org/10.1515/9783110354614>
- Mittelstenscheidt, P. (2009). Bob Dylan. Stimme und Gesicht. *Musik und Ästhetik*, 13 (51), 31–40.
- Nawka, T. & Anders, L. C. (Autor). (1996). *Die auditive Beurteilung heiserer Stimmen nach dem RBH-System. Doppel-Audio CD mit Stimmbeispielen und Begleitheft*. Stuttgart, New York: Thieme.
- Nawka, T., Anders, L. C., Cebulla, M. & Zurakowski, D. (1997). The speaker's formant in male voices. *Journal of voice : official journal of the Voice Foundation*, 11 (4), 422–428. [https://doi.org/10.1016/S0892-1997\(97\)80038-0](https://doi.org/10.1016/S0892-1997(97)80038-0)
- Nawka, T., Anders, L. C. & Wendler, J. (1994). Die auditive Beurteilung heiserer Stimmen nach dem RBH-System. *Sprache, Stimme, Gehör*, 18, 130–133.
- Nawka, T. & Wirth, G. (2008). *Stimmstörungen. Für Ärzte, Logopäden, Sprachheilpädagogen und Sprechwissenschaftler* (5., völlig überarb. Aufl.). Köln: Dt. Ärzte-Verl.
- Neppert, J. M. H. (1999). *Elemente einer akustischen Phonetik* (4., vollst. neu bearb. Aufl.). Hamburg: Buske.
- Nespital, U. (2013). *Wirkungen des funktionellen Nachvollzugs physiologischer Gesangsstimmen auf die Qualität der Sprechstimme* (Hallesche Schriften zur Sprechwissenschaft und Phonetik). Frankfurt am Main: Peter Lang. Verfügbar unter <http://search.ebscohost.com/login.aspx?direct=true&scope=site&db=e000xat&AN=665141>
- Nessel, E. (1960). Über das Tonfrequenzspektrum der pathologisch veränderten Stimme. *Acta Oto-Laryngologica, Suppl.* 157, 3–45.
- Panconcelli-Calzia, G. (1947). *Das Als Ob in der Phonetik. Eine erkenntniskritische Studie*. Hamburg-Bergedorf: Stromverlag.

- Panconcelli-Calzia, G. (1956). *Die Stimmatmung. Das Neue - Das Alte* (Nova Acta Leopoldina, 18/123). Leipzig: Johann Ambrosius Barth.
- Panconcelli-Calzia, G. (1961). *3000 Jahre Stimmforschung. Die Wiederkehr des Gleichen*. Marburg: N.G. Elwert.
- Parada-Cabaleiro, E., Baird, A., Batliner, A., Cummins, N., Hantke, S. & Schuller, B. W. (2017). The Perception of Emotion in the Singing Voice. In K. Page (Hrsg.), *Proceedings of the 4th International Workshop on Digital Libraries for Musicology - DLFM '17* (S. 29–36). New York: ACM Press.
- Pease, A. & Pease, B. (2001). *Warum Männer nicht zuhören und Frauen schlecht einparken. Ganz natürliche Erklärungen für eigentlich unerklärliche Schwächen* (Ullstein, Bd. 35969, Dt. Erstausg., 10. Aufl.). München: Ullstein.
- Pfleiderer, M. (2016a). Franklin, Aretha. In A.-C. Mecke, M. Pfeleiderer, B. Richter & T. Seedorf (Hrsg.), *Lexikon der Gesangsstimme. Geschichte, wissenschaftliche Grundlagen, Gesangstechniken, Interpretieren* (S. 233–235). Laaber: Laaber-Verlag.
- Pfleiderer, M. (2016b). Smith, Bessie. In A.-C. Mecke, M. Pfeleiderer, B. Richter & T. Seedorf (Hrsg.), *Lexikon der Gesangsstimme. Geschichte, wissenschaftliche Grundlagen, Gesangstechniken, Interpretieren* (S. 559–561). Laaber: Laaber-Verlag.
- Pfleiderer, M., Hänel, T., Horn, K. & Bielefeldt, C. (Hrsg.). (2015). *Stimme, Kultur, Identität. Vokaler Ausdruck in der populären Musik der USA, 1900 - 1960* (Texte zur populären Musik, Bd. 8). Bielefeld: Transcript.
- Rapoport, E. (1996). Emotional expression code in opera and Lied singing. *Journal of New Music Research*, 25 (2), 109–149. <https://doi.org/10.1080/09298219608570700>
- Re, D. E., O'Connor, J. J. M., Bennett, P. J. & Feinberg, D. R. (2012). Preferences for very low and very high voice pitch in humans. *PLoS one*, 7 (3), e32719. <https://doi.org/10.1371/journal.pone.0032719>
- Reddy, A. A. & Subramanian, U. (2015). Singers' and Nonsingers' Perception of Vocal Vibrato. *Journal of voice : official journal of the Voice Foundation*, 29 (5), 603–610. <https://doi.org/10.1016/j.jvoice.2014.09.022>
- Richter, B. (Hrsg.). (2014). *Die Stimme. Grundlagen, künstlerische Praxis, Gesunderhaltung* (2., durchges. Aufl.). Leipzig: Henschel.
- Sadolin, C. (2009). *Komplette Gesangstechnik*. Copenhagen, DK: Shout Publications ApS.
- Schindler, I., Hosoya, G., Menninghaus, W., Beermann, U., Wagner, V., Eid, M. et al. (2017). Measuring aesthetic emotions. A review of the literature and a new assessment tool. *PLoS one*, 12 (6), e0178899. <https://doi.org/10.1371/journal.pone.0178899>
- Schubert, E. (2013). Emotion felt by the listener and expressed by the music. Literature review and theoretical perspectives. *Frontiers in psychology*, 4, 837. <https://doi.org/10.3389/fpsyg.2013.00837>
- Seedorf, T. (2016a). Belcanto / bel canto. In A.-C. Mecke, M. Pfeleiderer, B. Richter & T. Seedorf (Hrsg.), *Lexikon der Gesangsstimme. Geschichte, wissenschaftliche Grundlagen, Gesangstechniken, Interpretieren* (S. 74–77). Laaber: Laaber-Verlag.
- Seedorf, T. (2016b). Glissando. In A.-C. Mecke, M. Pfeleiderer, B. Richter & T. Seedorf (Hrsg.), *Lexikon der Gesangsstimme. Geschichte, wissenschaftliche Grundlagen, Gesangstechniken, Interpretieren* (S. 266–268). Laaber: Laaber-Verlag.
- Seedorf, T., Pfeleiderer, M., Richter, B. & Mecke, A.-C. (2016). Sprechen und Singen. In A.-C. Mecke, M. Pfeleiderer, B. Richter & T. Seedorf (Hrsg.), *Lexikon der Gesangsstimme. Geschichte, wissenschaftliche Grundlagen, Gesangstechniken, Interpretieren* (S. 573–578). Laaber: Laaber-Verlag.
- Seidner, W. (2012). Zur Ästhetik der heiseren Singstimme - grausig oder großartig? In Deutsche Gesellschaft für Phoniatrie und Pädaudiologie (Hrsg.), *Tagungsband*. German Medical Science GMS Publishing House. Zugriff am 25.01.2018. Verfügbar unter <http://www.egms.de/static/de/meetings/dgpp2012/12dgpp97.shtml>
- Seidner, W. & Büttner, M. (1998). Zur Ästhetik der heiseren Singstimme. In H. Gundermann (Hrsg.), *Die Ausdruckswelt der Stimme. Kongressband 1. Stuttgarter Stimmtage*. Heidelberg: Hüthig.
- Seidner, W., Dippold, S. & Fuchs, M. (2009). "Stimmklanglauschen" und Hörtraining mit Kinder- und Jugendstimmen. In M. Fuchs (Hrsg.), *Hören Wahrnehmen (Aus-)Üben* (Kinder- und Jugendstimme, Bd. 3, S. 133–148). Berlin: Logos.
- Seidner, W. & Wendler, J. (2010). *Die Sängerstimme. Phoniatrie Grundlagen des Gesangs*. Leipzig: Henschel.
- Sendlmeier, W. F. (2002). Stimmliche und phonetische Manifestation emotionaler Sprechweise. In H. Geißner (Hrsg.), *Stimmkulturen. 3. Stuttgarter Stimmtage 2000* (S. 39–47). St. Ingbert: Röhrig.
- Sendlmeier, W. F. (2016). *Sprechwirkungsforschung. Grundlagen und Anwendungen mündlicher Kommunikation*. Berlin: Logos.

- Shewell, C. (1998). The Effect of Perceptual Training on Ability to Use the Vocal Profile Analysis Scheme. *International Journal of Language & Communication Disorders*, 33 (S1), 322–326. <https://doi.org/10.3109/13682829809179444>
- Shewell, C. (2009). *Voice work: art and science in changing voices*: John Wiley & Sons.
- Stumpf, C. (1926). *Die Sprachlaute*. Berlin, Heidelberg: Springer. <https://doi.org/10.1007/978-3-662-33112-5>
- Sundberg, J. (1974). Articulatory interpretation of the "singing formant". *The Journal of the Acoustical Society of America*, 55 (4), 838–844.
- Sundberg, J. (1987). *The science of the singing voice*. DeKalb, Ill.: Northern Illinois University Press.
- Thuesen, M. A., McGlashan, J. & Sadolin, C. (2017). Curbing - The Metallic Mode In-between. An empirical study qualifying and categorizing restrained sounds known as Curbing based on audio perception, laryngostroboscopic imaging, acoustics, LTAS, and EGG. *Journal of voice : official journal of the Voice Foundation*, 31 (5), 644.e1-644.e10. <https://doi.org/10.1016/j.jvoice.2017.01.010>
- Vieregge, W. H., Pahn, J. & Schutte, H. K. (1996). *Patho-Symbolphonetik. Auditive Deskription pathologischer Sprache* (Zeitschrift für Dialektologie und Linguistik, Heft 100). Stuttgart: F. Steiner.
- Von Appen, R. (2005). Die Wertungskriterien der Deutschland sucht den Superstar-Jury vor dem Hintergrund sozialer Milieus und kulturindustrieller Strategien. In D. Helms & T. Phleps (Hrsg.), *Keiner wird gewinnen. Populäre Musik im Wettbewerb* (Beiträge zur Populärmusikforschung, Bd. 33, S. 187–208). Bielefeld: Transcript.
- Von Appen, R. (2007). *Der Wert der Musik. Zur Ästhetik des Populären* (Texte zur populären Musik, Bd. 4). Zugl.: Gießen, Univ., Diss., 2007. Bielefeld: Transcript. Verfügbar unter http://deposit.d-nb.de/cgi-bin/dokserv?id=2960350&prov=M&dok_var=1&dok_ext=htm
- Webb, A. L., Carding, P. N., Deary, I. J., MacKenzie, K., Steen, N. & Wilson, J. A. (2004). The reliability of three perceptual evaluation scales for dysphonia. *European archives of oto-rhino-laryngology : official journal of the European Federation of Oto-Rhino-Laryngological Societies*, 261 (8), 429–434. <https://doi.org/10.1007/s00405-003-0707-7>
- Wendler, J. & Anders, L. C. (1986). Hoarse voices—on the reliability of acoustic and auditory classifications. *Proceedings of the 20th congress of the International Association of Logopedics and Phoniatrics, Tokyo, Japan*, 438–439.
- Zelcer, S., Henri, C., Tewfik, T. L. & Mazer, B. (2002). Multidimensional voice program analysis (MDVP) and the diagnosis of pediatric vocal cord dysfunction. *Annals of Allergy, Asthma & Immunology*, 88 (6), 601–608. [https://doi.org/10.1016/S1081-1206\(10\)61892-3](https://doi.org/10.1016/S1081-1206(10)61892-3)
- Zentner, M., Grandjean, D. & Scherer, K. R. (2008). Emotions evoked by the sound of music. Characterization, classification, and measurement. *Emotion*, 8 (4), 494–521. <https://doi.org/10.1037/1528-3542.8.4.494>