

GMTH Proceedings 2019

herausgegeben von | edited by
Florian Edler und Immanuel Ott

Notation. Schnittstelle zwischen Komposition, Interpretation und Analyse

19. Jahreskongress | 19th annual conference
Gesellschaft für Musiktheorie
Zürich 2019

herausgegeben von | edited by
Philippe Kocher



Die GMTH ist Mitglied von CrossRef.
<https://www.crossref.org>



Diese Ausgabe erscheint im Open Access und ist lizenziert unter einer
Creative Commons Namensnennung 4.0 International Lizenz.



This is an open access volume licensed under a
Creative Commons Attribution 4.0 International License.

Notation und Analyse von Tonhöhenverläufen in Sprechmelodien

ABSTRACT: Neuere Forschungsergebnisse aus dem Bereich der Musikpsychologie zeigen, dass die subjektive Beurteilung, ob jemand spricht oder singt, nicht nur von Eigenschaften des empfangenen akustischen Signals abhängt. Obwohl die Tonhöhe beim Sprechen anders moduliert wird als beim Singen, kann der Tonhöhenverlauf einer gesprochenen Äußerung vom Rezipienten als eine musikalische Tonfolge wahrgenommen werden. Dieser Eindruck brachte Musikschaffende und Sprachwissenschaftler bereits seit Ende des 18. Jahrhunderts dazu, Sprechmelodien in Form musikalischer Notenschrift aufzuzeichnen. Im aktuellen Diskurs der Linguistik setzte sich diese Art von Notation nicht durch. Für die Beschreibung der grammatikalisch relevanten Eigenschaften der Sprechmelodie in Intonationssprachen wie Deutsch und Englisch werden schematische Darstellungen eingesetzt, in denen es im Wesentlichen auf die binäre Unterscheidung zwischen Hoch- und Tieftönen ankommt. Um die Sprechmelodie aus musiktheoretischer Sicht zu betrachten, kann sich jedoch die Transkription in musikalische Notenschrift als ein wertvolles Werkzeug erweisen. Durch die Notation wird die Vielfalt der intervallischen Strukturen von Sprechmelodien erfassbar und eine Grundlage für deren musikalische Analyse geschaffen. In diesem Artikel wird zunächst ein methodischer Ansatz zur gehörmäßigen Erfassung von Sprechmelodien erprobt. Der Vergleich zwischen der exemplarisch transkribierten Sprechmelodie und den Messwerten der Grundfrequenz des Sprachsignals deutet darauf hin, dass das Gehör die verfügbare Tonhöheninformation selektiert, um daraus die Vorstellung einer musikalischen Tonfolge zu konstruieren. Im Bereich von Tonhöhenakzenten orientiert sich die Wahrnehmung scheinbar nach dem lokalen Maximum der Grundfrequenzfunktion, während sich in den übrigen Silben deren Mittelwert als ausschlaggebend herausstellt. Anhand der Transkription werden mögliche analytische Ansätze dargelegt, die strukturelle Gemeinsamkeiten zwischen sprachlichen und musikalischen Melodien veranschaulichen.

Recent research results from the field of music psychology show that the subjective assessment of whether someone speaks or sings does not only depend on the characteristics of the received acoustic signal. Although the pitch is modulated differently when speaking than when singing, the recipient can perceive the pitch of a spoken utterance as a musical tone sequence. This impression has since the end of the 18th century led musicians and linguists to record speech melodies in the form of musical notation. In the current discourse of linguistics, this kind of notation has not prevailed. For the description of the grammatically relevant characteristics of the speech melody in intonation languages such as German and English, schematic representations are used in which the main focus lies on the binary distinction between high and low tones. However, to approach speech melody from the perspective of music theory, the transcription in musical notation can be a useful tool. Notation allows to capture the variety of intervallic structures of speech melodies and provides a basis for their musical analysis. This paper introduces a methodical approach to the aural recording of speech melo-

dies. The comparison between the transcribed speech melody and the measured values of the fundamental frequency of the speech signal suggests that the auditory system selects the available pitch information in order to construct the representation of a musical tone sequence. In the area of pitch accents, the perception seems to be driven by the local maximum of the fundamental frequency function, while in the remaining syllables, its mean value appears to be determinant. Based on the transcription, possible analytical approaches are presented, illustrating structural similarities between linguistic and musical melodies.

SCHLAGWORTE/KEYWORDS: Intonation; music notation; Musiknotation; pitch perception; Prosodie; prosody; speech melody; Sprechmelodie; Tonhöhenwahrnehmung

Einleitung

Gesprochene und gesungene Sprache sind Klangerzeugnisse des menschlichen Atmungs- und Artikulationssystems. Je nachdem, ob die im Kehlkopf befindlichen Stimmbänder geöffnet bleiben oder sich quasi-periodisch öffnen und schließen, erzeugt die durchströmende Atemluft entweder stimmlose oder stimmhafte Laute. Stimmhafte Laute weisen wahrnehmbare Tonhöhen auf, die maßgeblich von der Frequenz eines solchen Öffnungs- und Schließvorgangs abhängen. Sowohl bei gesprochener als auch bei gesungener Spracherzeugung wird die Spannung der Stimmbänder gesteuert, um die Grundfrequenz des Sprachsignals zu modulieren. Diese Frequenzmodulation erzeugt ›Tonhöhenverläufe‹, die gemeinsam mit dem entsprechenden Rhythmus verschiedenartige Melodien formen. Die »melodische Gestaltung sprachlicher Äußerungen« wird in der Linguistik als ›Intonation‹ bezeichnet.¹

Trotz der grundlegenden Ähnlichkeiten zwischen den physiologischen Prozessen, die der Entstehung von Tonhöhenverläufen in gesungener und gesprochener Sprache zugrunde liegen, sind diese beiden Formen der Phonation klar unterscheidbar. Während im Gesprochenen die gleitenden und rapiden Tonhöhenbewegungen den Höreindruck prägen, wird Gesungenes meist als eine Abfolge diskreter Tonstufen gehört. Jedoch verhindert die größere Instabilität der Tonhöhe, die in Sprechmelodien feststellbar ist, nicht die Wahrnehmung der intervallischen Strukturen. Die periodischen Anteile des Sprachsignals gewährleisten die Robustheit der Tonhöhenwahrnehmung² – nur wird die Aufmerksamkeit beim

1 Peters 2014, 1.

2 Ebd., 15 f.

Hören eines Redeflusses meist nicht darauf gerichtet. Die Sprechmelodie erfüllt in Zusammenspiel mit dem Wortlaut primär kommunikative Funktionen, die in der Grammatik der jeweiligen Sprache kodiert sind. Erst wenn diese Kommunikationsebene in den Hintergrund rückt oder gar erlöscht – sei es durch gezieltes Hinhören oder wegen der Unzugänglichkeit der Bedeutung – tritt die Beschaffenheit des Tonhöhenverlaufs hervor.

Obwohl die Sprechmelodie eine zentrale Rolle im mündlichen Ausdruck einnimmt, wird sie in den meisten Sprachen nicht oder nur andeutungsweise verschriftlicht. Seit Ende des 18. Jahrhunderts wurde mangels eigener Repräsentationssysteme und wegen der Affinität zur Musik für die grafische Aufzeichnung von Sprechmelodien die musikalische Notenschrift verwendet. In der modernen Intonationsforschung, einem Teilgebiet der Phonetik und Phonologie, konnte sich jedoch diese Art der Notation nicht durchsetzen.³ Die jüngsten technischen Entwicklungen, die die Extraktion des Grundfrequenzverlaufs aus dem akustischen Signal und dessen grafische Darstellung ermöglichen, zeigen sich für die Zwecke der Intonationsforschung akkurater und geeigneter als die Notenschrift. Doch wenn die sprachliche Intonation auf ihre musikalischen Eigenschaften untersucht wird, kann ein gemeinsames Repräsentationsmodell für Musik und Sprache von Vorteil sein. Ziel dieses Beitrags ist es, zu zeigen, wie die Transkription in Form musikalischer Notation eine wertvolle Schnittstelle zur Analyse von Sprechmelodien bilden kann. Zu diesem Zweck werden zunächst frühe und moderne Formen der grafischen Repräsentation und die verschiedenen kommunikativen Funktionen der Intonation beleuchtet. Darauf aufbauend wird ein Ansatz zur Transkription in Notenschrift vorgestellt und diskutiert. Hierfür ist es unabdingbar, die neuesten Erkenntnisse im Bereich der Linguistik und der Musikpsychologie zu berücksichtigen. Zum Schluss werden am Beispiel der transkribierten Sprechmelodie mögliche analytische Ansätze dargelegt.

Die Sprechmelodie und die Geschichte ihrer grafischen Repräsentation

Die Sprechmelodie ist erstaunlich variabel: Sprachen unterscheiden sich durch ihre eigentümliche Melodik. Jeder Mensch redet anders, höher oder tiefer, in kleinerem oder größerem Tonumfang. Doch diese Variabilität unterliegt Regeln. In jeder Sprache übernimmt der Tonhöhenverlauf systematische Funktionen, die

3 Fuhrhop/Peters 2013, 130.

sich aus der jeweiligen Grammatik ergeben.⁴ In Tonsprachen – wie Mandarin-Chinesisch oder Thailändisch – sind Tonhöhenbewegungen u. a. für die Differenzierung der Bedeutung einzelner Worte zuständig. In den sogenannten Tonakzentsprachen – wie Schwedisch, Lettisch oder Kroatisch – werden Wortakzente nicht nur durch Erhöhung des Schalldrucks, sondern primär durch Variation der Tonhöhe realisiert.⁵ Auch die altgriechische Sprache besaß solche melodischen Wortakzente,⁶ deren besonderer musikalischer Charakter sich vor allem in der Redekunst bemerkbar machte. Schon Dionysios von Halikarnassos (ca. 54 v. Chr. bis 7 n. Chr.) verglich die hellenische Rhetorik mit der damaligen Vokal- und Instrumentalmusik, und beobachtete, dass sich diese einzig »in Grad« aber nicht »in ihrer Beschaffenheit« unterschieden.⁷ Um die eigentümliche Sprechmelodie des Altgriechischen schriftlich zu konservieren, begann man im 3. Jahrhundert ihre Tonakzente mittels diakritischer Zeichen zu notieren.⁸ Verschiedene Autoren sind der Auffassung, dass diese Akzentzeichen – Akut, Gravis und Zirkumflex – mögliche Vorgänger der mittelalterlichen Neumen-Notation sein können.⁹

In den Intonationssprachen, wozu Deutsch und Englisch zählen, dienen Tonhöhenverläufe primär der Realisierung syntaktischer Akzente auf Satzebene, wodurch die Informationsstruktur von Äußerungen verdeutlicht wird. Das heißt aber nicht unbedingt, dass ihr musikalischer Charakter weniger ausgeprägt ist als bei Ton- und Tonakzentsprachen. In manch deklamatorischem Vortragsstil, wie er auf Theaterbühnen bis ins 20. Jahrhundert gepflegt wurde, erscheint die Grenze zwischen Singen und Sprechen fließender. Aufgrund der erhöhten emotionalen Involviertheit des Sprechers und der Vorrangstellung der metrischen Struktur, die Gedichte und Theatertexte aufweisen, kann durch Silbendehnungen und vergrößertem tonalen Register eine singende Sprechweise entstehen.¹⁰ Diese Ähnlichkeit zwischen einer »gesteigerten Rezitation« und dem Gesang findet häufig Erwähnung in der Literatur. J.W. von Goethe macht in seinen *Regeln für Schauspieler* (1803) darauf aufmerksam, dass die Schauspielenden die Deklamation mit gewisser Zurückhaltung zu gestalten hätten: Um nicht »in das Singen« zu verfal-

4 Blühdorn 2013, 243.

5 Peters 2014, 3f.

6 Engels 2011, 13.

7 Halicarnassus 1910, 125.

8 Vendryes 1904, 6ff.

9 Haas 2016.

10 Vgl. Menninghaus et al. 2018, 1–3.

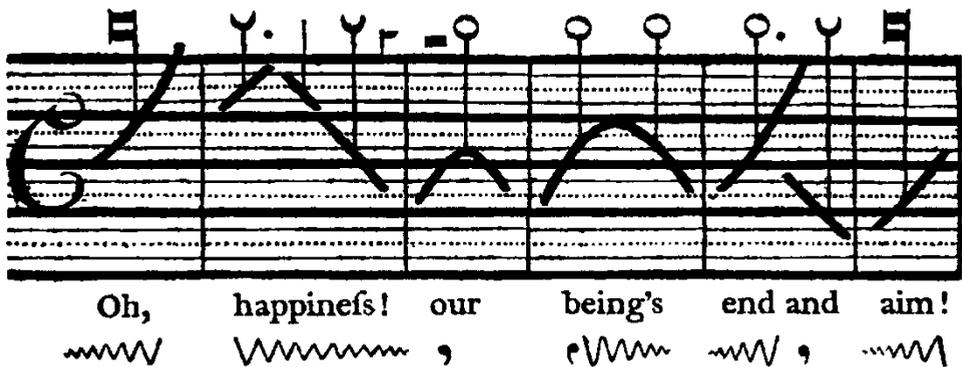


Abbildung 2: Notation der Sprechmelodie eines Verses von Alexander Pope (1688–1744). Quelle: Steele 1779, 13.

Sowohl Steele als auch Köhler notierten Sprechmelodien, indem sie selbst beim Sprechen auf die Tonbewegungen ihrer eigenen Stimme achteten.¹³ Leoš Janáček (1854–1928) seinerseits, getrieben von seinem ethnomusikologischen Interesse, lauschte auf die Sprechmelodie seiner Landsleute. Ab 1897 und bis ans Ende seines Lebens notierte Janáček Sprechmelodien (Tschechisch *nápěvy mluvy*) von Kindern, Passanten und engen Verwandten in realen Lebenssituationen. Solche »realen Motive« sind meist kurze Äußerungen, die der Komponist vor Ort und in Echtzeit mithilfe seines absoluten Gehörs notierte.¹⁴ Die meisten Aufzeichnungen, die Janáček in seinen Notizblöcken sammelte, sind in tschechischer Sprache, aber Transkriptionen aus anderen Sprachen – u. a. Französisch, Russisch und Deutsch – kommen ebenfalls vor. Der Melodieverlauf wurde in einem Notensystem im chromatischen Tonvorrat notiert und oft mit Dynamikangaben versehen. Im Regelfall wurde eine Tonhöhe pro Silbe festgelegt.¹⁵ Es war Janáček dennoch bewusst, dass die konventionelle musikalische Notation lediglich eine grobe Momentaufnahme der flüchtigen Sprechmelodien darstellen konnte.¹⁶

Auch im Bereich der Phonetik wurde seit Ende des 19. Jahrhunderts die herkömmliche musikalische Notation verwendet, um den »musikalischen Satzton« bestimmter Äußerungstypen darzustellen. Beispiel hierfür ist eine ab 1893 veröf-

13 Steele 1779, 38; Köhler 1853, 8.

14 Vainiomäki 2012, 172.

15 Pearl 2006, 135.

16 Vainiomäki 2012, 171.

fentlichte *Sammlung kurzer Grammatiken deutscher Mundarten* (Abb. 3). Die Aufzeichnungen stützten sich lediglich auf das Gehör und konnten deshalb – wie einer der Autoren selbst anmerkte – nur annähernd die »feinen Nuancierungen« der Satzintonation ausdrücken.¹⁷



Abbildung 3: »Tonbild« eines Aussagesatzes in der Nürnberger Mundart. Quelle: Gebhardt 1907, 15.

Die Rolle der Tonaufnahmetechnik in der Wahrnehmung und Repräsentation von Sprechmelodien

Erst mit der Entwicklung der Tonaufnahmetechnik im Laufe des 20. Jahrhunderts wurde es möglich, eine sprachliche Äußerung wiederholt und unverändert anzuhören. Die natürliche prosodische Variabilität, die selbst innerhalb Äußerungen desselben Sprechers auftritt, konnte damit umgangen werden.¹⁸ So waren Phonetiker in der Lage, eine exaktere Erfassung von Tonhöhenverläufen vorzunehmen.¹⁹ Doch die Erkennung der Tonhöhe erfolgte weiterhin durch das Gehör, sodass die Ergebnisse nach wie vor – auch wenn in geringerem Maße – subjektiv sind.

Der Einfluss der technisch erzeugten und exakten Wiederholung auf die subjektive Einschätzung, ob eine Äußerung gesungen oder gesprochen wird, war in den letzten Jahren Gegenstand musikpsychologischer Forschung. Das Phänomen der perceptiven Verwechslung beider Arten der Klangerzeugung zeigt sich in der von der englischen Musikpsychologin Diana Deutsch dokumentierten *Speech-to-Song Illusion*.²⁰ Dabei wurde experimentell festgestellt, dass eine kurze gesprochene Äußerung den meisten Zuhörern mit musikalischer Ausbildung als gesun-

17 Heilig 1898, 7.

18 Peters 2014, 36.

19 Kiefer 2016.

20 Deutsch et al. 2011.

gen vorkommt, wenn sie mehrmals unverändert wiederholt wird. Die Wiederholungen begünstigen offenbar die Entbindung der Sprechmelodie von ihrer dienenden Rolle als Werkzeug der Informationsübermittlung. Es setzt eine Stilisierung des Tonhöhenverlaufs ein, die der melodischen Gestaltung einen Eigenwert verleiht. Laut Deutsch et al. wird bei diesem Phänomen die individuelle Fähigkeit gefordert, regelbasierte Eigenschaften des musikalischen Tonsystems, die im Langzeitgedächtnis verankert sind, abzurufen. Damit wird der akustische Reiz so verarbeitet, dass er sich der Vorstellung einer präzisen Melodie annähert. Neuere Studien zeigen, dass dies auch von Musikhörern ohne besondere musikalische Bildung bewerkstelligt wird.²¹ Dennoch ist zu beobachten, dass die Illusion nicht bei jeder gesprochenen Äußerung in gleicher Ausprägung eintritt. Die Untersuchungen von Tierney et al. liefern hierfür den statistischen Beleg.²² Manche Äußerungen stuften die Probanden schon nach wenigen Wiederholungen tendenziell als musikähnlich ein, andere wurden unter vergleichbaren Bedingungen weiterhin eher als gesprochen empfunden. Die Experimente von Tierney et al. zeigen eine Reihe von Faktoren auf, die die perzeptorische Transformation begünstigen und bestätigen die Beobachtung, dass Gesprochenes als Musik gewertet werden kann, auch wenn es nicht mit der Absicht produziert wurde, als Musik zu gelten.²³

Obwohl Gesprochenes wie eine musikalische Umsetzung des Textes empfunden werden kann, wird im modernen Diskurs der Linguistik der musikalischen Beschaffenheit von Tonhöhenverläufen wenig Aufmerksamkeit geschenkt. Nach dem Ansatz der ›Autosegmental-Metrischen Phonologie‹, auf dem die meisten Systeme der Intonationsanalyse beruhen, sind in Tonhöhenverläufen nur diejenigen Eigenschaften relevant, die eine sprachliche Funktion erfüllen. Hauptziel der aktuellen Intonationsforschung ist also, aus der Menge von Tonhöhenverläufen, die beim Sprechen erklingen, die kommunikativ relevanten Konturen zu identifizieren.²⁴ Hierfür ist in erster Linie die binäre Unterscheidung zwischen Hoch- und Tieftönen ausschlaggebend. Für die Beschreibung der Intonation des Deutschen, sowie anderer westgermanischer Sprachen, werden Annotationssysteme verwendet, die die Hoch- und Tieftöne mit den entsprechenden Silben assoziieren (Abb. 4). In den 1990er Jahren wurden Algorithmen entwickelt, die die Bestim-

21 Vanden Bosch der Nederlanden et al. 2015.

22 Tierney et al. 2018.

23 Ebd., 902.

24 Peters 2014, 2.

mung des Grundfrequenzverlaufs aus dem digitalen Sprachsignal mit großer Zuverlässigkeit ermöglichen. Mit Computerprogrammen für phonetische Analyse wie *Praat*²⁵ können grafische Darstellungen des Grundfrequenzverlaufs erzeugt werden, welche in der Intonationsforschung breite Anwendung finden.

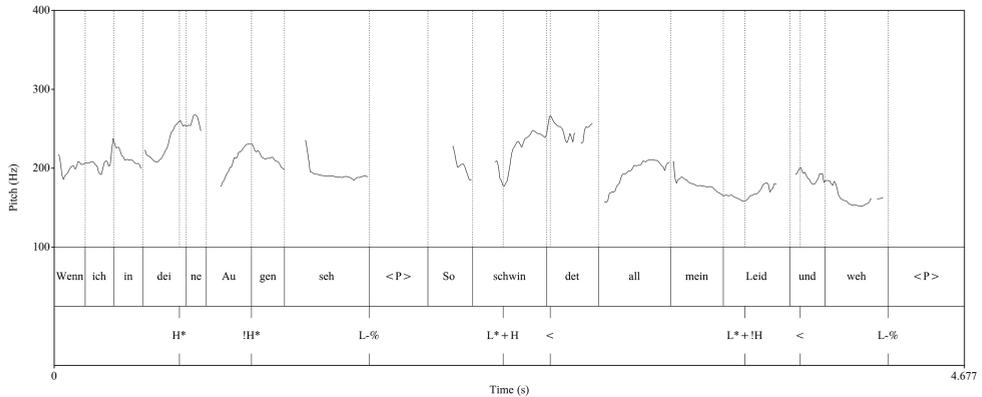


Abbildung 4: Grundfrequenzverlauf (Hz) einer phonetischen Realisierung der ersten zwei Verse eines Gedichtes von Heinrich Heine. Die Extraktion des Grundfrequenzverlaufs (y-Achse) erfolgte mithilfe des Computerprogramms *Praat*. Für die prosodische Annotation wurde das Beschreibungsmodell *German Tones and Break Indices – GToBI* verwendet (Grice/Baumann 2002).

Eine Form der grafischen Darstellung des Grundfrequenzverlaufs anhand von algorithmischen Messungen, die Analogien zu Musik erweckt, liefert das von Piet Mertens entwickelte *Prosogram*-Modell.²⁶ Dieses basiert auf der Beobachtung, dass der Umfang der Tonhöhenschwankungen innerhalb einer Silbe im Verhältnis zur Silbendauer meist klein genug ist, um beim Rezipienten die mentale Repräsentation einer stabilen Tonhöhe auszulösen. Sprechmelodien werden deswegen beim *Prosogram*-Modell weitgehend als Abfolge diskreter Tonhöhen dargestellt. Hierbei wird jeder Silbe einer Äußerung entweder eine Tonhöhenstufe oder eine Gleitbewegung (*Glissando*) zugewiesen.

Abseits der Hauptströmung der Intonationsforschung wurde 2017 im Rahmen einer Konferenz der *International Speech Communication Association* eine »neue Methode zur Transkription von sprachlicher Prosodie« in Form musikalischer

25 Die letzte Version der von Paul Boersma und David Weenink entwickelten Anwendung *Praat* kann kostenfrei auf <http://www.praat.org/> (13.03.2023) heruntergeladen werden.

26 Mertens 2004.

Notation vorgestellt. Bei dieser Methode wird das akustische Sprachsignal – ähnlich wie beim *Prosogram*-Modell – in Vokaleinheiten (*vowel-to-vowel units*) zergliedert.²⁷ Für jeden tonhöhentragenden Silbenkern wird der Mittelwert der Grundfrequenz algorithmisch ermittelt und damit die auftretenden Schwankungen, unabhängig davon, wie groß sie ausfallen, geglättet. Die ermittelten Frequenzwerte werden mittels eines automatisierten Verfahrens in Tonhöhen umgerechnet und mit der nächstliegenden Stufe einer gleichschwebenden Vierteltonskala in Übereinstimmung gebracht. In Anlehnung an die bereits erwähnte Arbeit von Steele wird das Ergebnis in gewöhnlicher Notenschrift mithilfe von Vierteltonversetzungszeichen notiert. Ein ähnliches automatisiertes Prozedere wurde von Menninghaus et al. vorgeschlagen, um großen Datenmengen zu sammeln. Auch hier wird pauschal der Mittelwert der Grundfrequenz auf Silbenebene herangezogen. Diese diskrete Werte werden anschließend in Halbtonstufen konvertiert und in MIDI-Daten umgewandelt, die wiederum in Notenschrift dargestellt werden.²⁸

Notation von Sprechmelodien

Die Transkription in Form von Notenschrift mag in der Intonationsforschung nicht zielführend sein, aber sie kann sich als nützliches Medium erweisen, intervallische Strukturen des Tonhöhenverlaufs offenzulegen, die mit Mitteln der sprachwissenschaftlichen Intonationsanalyse nicht darstellbar sind. Im Folgenden wird ein methodischer Ansatz zur Notation von Tonhöhenverläufen skizziert, der – anders als bei den bereits zitierten Arbeiten von Meireles et al. und Menninghaus et al. – auf dem subjektiven Tonhöheneindruck basiert. Der Sprachrhythmus bleibt dabei unbeachtet. Diese Vorgehensweise ist in erster Linie zur Anwendung auf die deutsche Sprache konzipiert.

Als Quelle für die nachfolgende Transkription dienten unkomprimierte digitale Audioaufnahmen, die die technischen Empfehlungen im Bereich der phonetischen Analyse befolgen.²⁹ Die benutzten Daten wurden mit 44100 Hz Abtastrate in 16bit-Quantisierung aufgenommen und im WAV-Format gespeichert. Als Vorlage wurde ein Gedicht von Heinrich Heine verwendet. Eine Deutschmutter-

²⁷ Meireles et al. 2017.

²⁸ Menninghaus et al. 2018, 6 f.

²⁹ Deutsche Forschungsgesellschaft 2013.

prachlerin mit baden-württembergischen Wurzeln wurde aufgefordert, die folgende Strophe des Gedichtes ausdrucksvoll vorzulesen:

Wenn ich in deine Augen seh,
So schwindet all mein Leid und Weh;
Doch wenn ich küsse deinen Mund,
So werd ich ganz und gar gesund.³⁰



https://storage.gmth.de/proceedings/articles/259/attachments/259_audio_01.wav

Audiobeispiel 1: Tonaufnahme der phonetischen Umsetzung des Gedichtes, deren Sprechmelodie transkribiert wurde.

Gehörmäßige Erfassung des Tonhöhenverlaufs

Zunächst wurde die entstandene Tonaufnahme in kleinere Abschnitte geteilt. In diesem Fall war es naheliegend, aus jedem Vers einen Abschnitt zu bilden. Diese Abschnitte fallen jeweils mit einer Intonationsphrase zusammen, die von der nächsten entweder durch eine akustische Pause oder eine Verlangsamung der Sprechgeschwindigkeit getrennt ist. Durch wiederholtes Abspielen jedes einzelnen Verses wurde die Gelegenheit geschaffen, die bereits beschriebene *Speech-to-Song Illusion* auszulösen, um dadurch den wahrgenommenen Tonhöhenverlauf vom Gehör zu erfassen. Da wo die perzeptorische Transformation sich nicht mit der nötigen Robustheit einstellte, wurde die Tonaufnahme silbenweise abgespielt. Somit konnte für jede Silbe die Tonhöhe bestimmt werden, die die Intervallqualität zu den benachbarten Silben am besten darstellte. Schließlich wurde die Tonfolge per Annäherungsprinzip im chromatischen Tonvorrat notiert (Abb. 5). Diese Aufgabe wird vom Gehör womöglich mithilfe der bereits 1738 vom Mathematiker Leonhard Euler beschriebenen Fähigkeit des ›Zurechthörens‹ bewerkstelligt.³¹ Auf das komplexe Phänomen, welches das tonhöhenkorrigierende Hören ermöglicht, kann im begrenzten Umfang dieser Arbeit jedoch nicht eingegangen werden.³²

30 Quelle: Heinrich Heine. *Buch der Lieder. Lyrisches Intermezzo. IV.* (1822–23).

31 Debbeler 2007, 77 f.

32 Näheres über das Zurechthören bei Fricke 2012.



https://storage.gmth.de/proceedings/articles/259/attachments/259_audio_02.wav

Audiobeispiel 2: *Speech-to-Song Illusion*, mehrfache Wiederholung der phonetischen Umsetzung des Gedichtes, zuerst in originaler Abspielgeschwindigkeit und danach verlangsamt.



Abbildung 5: Rohtranskript der Tonaufnahme ($a^1 = 440$ Hz); die Tonhöhenfassung erfolgte vom Gehör nach wiederholtem Anhören des Signalabschnitts in verschiedenen Abspielgeschwindigkeiten.

Messtechnischer Abgleich

Das Rohtranskript (Abb. 5) ist Ergebnis einer subjektiven Tonhöhenfassung, die nun einem messtechnischen Abgleich unterworfen wird. Dabei soll untersucht werden, welche Eigenschaften des Sprachsignals für den subjektiven Tonhöheindruck, die im Rohtranskript repräsentiert ist, maßgeblich sind. Um das Ergebnis mit dem algorithmisch ermittelten Grundfrequenzverlauf (f_0 -Verlauf) zu vergleichen, wird die Audiodatei im Phonetik-Programm *Praat* analysiert. Zur Analyse gehört die Aufteilung des Signals in Segmente (Silben), die orthografische Transkription sowie die Annotation der Tonhöhenakzente und Grenztöne am f_0 -Verlauf nach dem Modell *German Tones and Break Indices*, kurz GToBI (Abb. 4).³³ Die isolierten Silben werden anschließend einzeln analysiert, um die jeweiligen Tonhöhen untereinander zu vergleichen. Dieser Vergleich ist dadurch erschwert, dass die Tonhöhe innerhalb der Silbe tatsächlich deutlichen Schwankungen unterliegt. Für jede Silbe wird deswegen das Minimum, das Maximum und der Mittelwert³⁴ des f_0 -Verlaufs sowie der f_0 -Wert am Punkt maximaler Schallintensität in der betreffenden Zeitspanne gemessen. Dadurch wird einerseits untersucht, ob Übereinstimmungen zwischen dem melodischen Höreindruck und den gemessenen f_0 -Werten zu finden sind. Andererseits kann aus eventuellen Übereinstimmungen abgeleitet werden, welcher gemessene Wert den Höreindruck in diesem konkreten Beispiel geprägt hat.

33 Das GToBI ist ein Transkriptionssystem für die Prosodie des Deutschen nach dem Ansatz der ›Autosegmental-Metrischen Phonologie‹. Für eine Entschlüsselung der benutzen Diakritika s. Grice et al. o.J. unter http://www.gtobi.uni-koeln.de/gm_gtobi_modell.html (4.10.2023).

34 Als Mittelwert des f_0 -Verlaufs wird hier der Mittelwert der Funktion der Frequenz in einem gegebenen Zeitintervall verstanden. Dies wird von der Software *Praat* automatisch errechnet.

Bei der Auswertung des gemessenen f_0 -Verlaufs darf nicht ignoriert werden, dass nicht alle Silben in gleichem Umfang zur melodischen Gestaltung einer Äußerung beitragen.³⁵ Nach dem Ansatz der ›Autosegmental-Metrischen Phonologie‹ werden Tonhöhenverläufe als Sequenzen von phonetischen Zielpunkten und deren linearer Interpolation repräsentiert. Diese Zielpunkte befinden sich am Tonhöhengipfel der Akzentsilbe und deren Begleittöne, sowie an Phrasengrenzen. Tonakzente sind meist mit schnellen, gleitenden Tonhöhenbewegungen verknüpft. Diese werden im Notenbild mit Slide-Symbolen angedeutet, wie sie in der Jazz-Notation gebräuchlich sind (Abb. 6).



Abbildung 6: Zieltöne des Tonhöhenverlaufs.

Mithilfe eines manipulierten f_0 -Verlaufs lässt sich feststellen, dass die Reduktion des ursprünglichen Verlaufs auf eine lineare Tonsequenz keine signifikante Änderung der wahrgenommenen Sprechmelodie verursacht (Abb. 7).³⁶ Der Tonhöhenverlauf kann also als eine lineare Sequenz beschrieben werden: Der initiale tiefe Grenzton (›Wenn‹) breitet sich im leichten Tonhöhenanstieg zum Anfang des Tonakzents auf der Silbe ›dei(ne)‹ aus. Von dort an folgt ein steiler Anstieg bis zum Gipfelton der Intonationsphrase. Der Gipfelton wird am Punkt des maximalen f_0 -Wertes innerhalb der betonten Silbe ›dei(ne)‹ erreicht. Danach folgt ein herabgesetzter Tonhöhenakzent auf der Silbe ›Au(gen)‹ und schließlich fällt die Melodie bis zum finalen Grenzton, der sich über die Länge der Silbe ›seh‹ erstreckt.

In dieser Tonsequenz (Abb. 6) entstehen zwischen den fünf festgelegten Zielpunkten vier Intervalle, deren Größe nun mit dem Verhältnis zwischen den algorithmisch errechneten f_0 -Werten verglichen wird (Abb. 9). Das gemessene Intervall vom initialen Grenzton (›Wenn‹) zum Anstiegsbeginn des Akzents (›deine‹) beträgt 105 Cent, und stimmt somit mit der wahrgenommenen Intervallqualität überein. Das gleiche gilt für die Intervalle vom Hochtton ›dei(ne)‹ zum nächsten Hochtton ›Au(gen)‹, und von diesem zum finalen Grenzton. Doch im Tonakzent-

35 Fuhrhop/Peters 2013.

36 Eine ausführliche experimentelle Untersuchung zur Wahrnehmung von stilisierten Sprechmelodien liefert die Arbeit von Hart et al. (1990).

bereich ›deine‹ erklingt die Terz (213,2:260,7 Hz \approx 9:11) mit 348 Cent etwa einen Viertelton höher als eine temperierte kleine Terz. Nichtsdestotrotz befindet sich dieser Wert näher an der reinen kleinen Terz (5:6), deren Größe 316 Cent beträgt, als an der reinen großen Terz (4:5), die mit 386 Cent ein wenig entfernter liegt. Die Tonhöhen *a* und *c*¹ bringen somit annäherungsweise die Intervallqualität zum Ausdruck. Folglich bilden die notierten Zieltöne ein robustes Gerüst der wahrgenommenen Sprechmelodie.

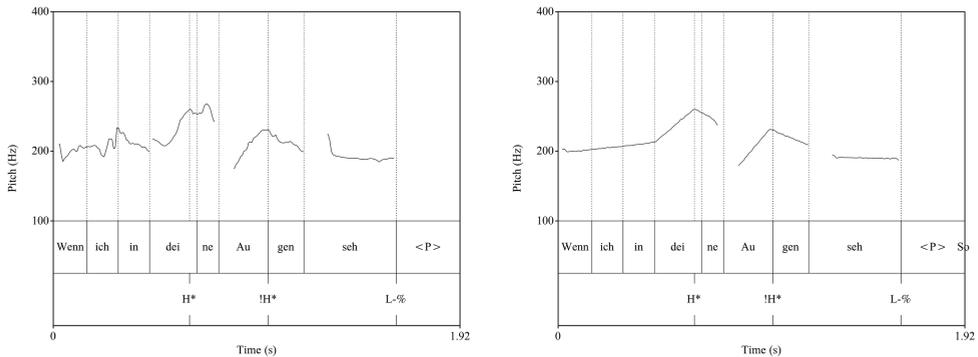


Abbildung 7: Grafische Darstellung des algorithmisch ermittelten f_0 -Verlaufs im Original (links) und in Form einer reduzierten linearen Sequenz (rechts). Als Grundlage für die Erstellung der Reduktion werden in der Regel Mittelwerte der f_0 in der betreffenden Silbe (›wenn‹, ›ich‹ und ›seh‹) verwendet. Ausnahme hierzu bildet der Gipfel im Tonakzentbereich (›dei‹ und ›Au‹), wo der Maximalwert ausschlaggebend ist.

Silbe	Zielton	Frequenz (Hz)		Intervall (Bruch)	Intervall (Halbtöne, Cent)	Intervallqualität
Wenn	(%L)	200,7	Mittelwert			
ich						
in		213,2	Mittelwert	16/17	1,05	kl. Sekunde
dei-	H*	260,7	Maximum	9/11	3,48	kl./gr. Terz
ne						
Au-	!H*	231,1	Maximum	8/9	-2,09	gr. Sekunde
gen						
seh	L-%	191,6	Mittelwert	5/6	-3,25	kl. Terz

Abbildung 8: Messwerte der Grundfrequenz für die mit Zieltönen verknüpften Silben. Die Intervalle werden aus dem Verhältnis des Frequenz-Wertes der betroffenen Silbe zu dem darauffolgenden Zielton errechnet. Die für die Beschriftung der Zieltöne benutzten Diakritika stammen aus dem GToBI-Modell (Grice et al. o.J.).

🔊 https://storage.gmth.de/proceedings/articles/259/attachments/259_audio_03.wav

Audiobeispiel 3: Isolierte Silben, die einen Zielton der Tonsequenz tragen und, zum Vergleich, die entsprechenden generierten Sinustöne mit gleichbleibender Frequenz (200,7 Hz; 213,2 Hz; 260,7 Hz; 231,1 Hz und 191,6 Hz).

Die Töne für die restlichen drei Silben (>ich<, >ne< und >gen<) liegen auf Übergängen zwischen den notierten Zielönen (Abb. 7), und besitzen deswegen nach den Modellen der Intonationsanalyse geringere funktionale Relevanz.³⁷ Damit ist gemeint, dass sie für die Wahrnehmung der Intonationskontur eine untergeordnete Rolle spielen. Dies lässt vermuten, dass der perzeptive Eindruck über die Tonhöhe solcher >intermediären Silben< unter Einfluss der benachbarten Zielöne steht. Beim Vergleich der gehörmäßig erfassten Tonfolge mit den entsprechenden algorithmischen Messwerten in den Silben, die sich in der Nähe des Akzenttons befinden, kann eine deutliche Abweichung zwischen der wahrgenommenen Tonhöhe und dem f_0 -Mittelwert beobachtet werden (Abb. 9). So orientiert sich anscheinend das Gehör im Akzentton-Bereich eher am f_0 -Maximum, während in der Nähe von Phrasengrenzen der f_0 -Mittelwert und der f_0 -Wert am Punkt maximaler Schallintensität maßgebend sind. Werden diese unterschiedlichen Referenzwerte herangezogen, so liegen die notierten Tonhöhen und die entsprechenden Messwerte kaum mehr als 50 Cent auseinander.

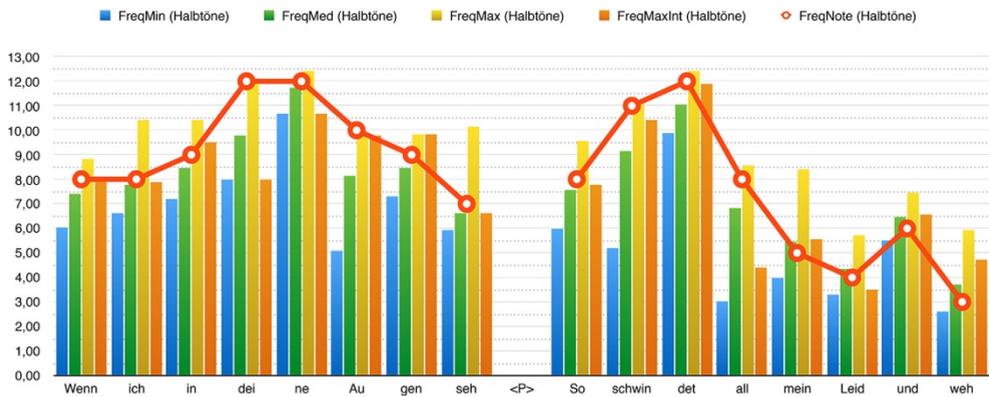


Abbildung 9: Algorithmische Messwerte des f_0 für jede Silbe der Äußerung. Die Balken repräsentieren jeweils Minimum, Mittelwert und Maximum der f_0 -Messwerte, sowie die Frequenz am Punkt maximaler Schallintensität. Die rote interpolierte Linie stellt die Tonhöhen der gehörmäßigen Tonhöhenenerfassung (wie in Abb. 6) dar. Die Frequenzwerte wurden hier in Halbtöne und Cent umgerechnet, wobei 0 der Tonhöhe c (130,813 Hz) entspricht. Die Akzenttöne sind mit den Diakritika H* und L* markiert.

37 Peters 2014, 26.



https://storage.gmth.de/proceedings/articles/259/attachments/259_audio_04.wav

Audiobeispiel 4: Phonetische Umsetzung des Gedichtes mit einem stilisierten Tonhöhenverlauf, in dem die Tonhöhen – in Entsprechung zur rot interpolierten Linie (Abb. 9) – in gleichschwebender Temperatur erklingen.

Bei der Transkription des Tonhöhenverlaufs wird die unterschiedliche Funktion von Zieltönen und Übergängen berücksichtigt: Tonhöhen von Silben, die Zieltöne beinhalten, werden mit größeren Notenköpfen notiert, als diejenigen, die auf Übergängen liegen. Gleitende Tonhöhenbewegungen an Tonakzenten werden mit Slide-Symbolen angedeutet. Zusätzlich wird die musikalische Darstellung mit einer prosodischen Annotation nach dem GToBI-Modell versehen. Die Abbildung 11 zeigt nun eine fertige Transkription für die ersten zwei Verse des Gedichtes, die nach der beschriebenen Vorgehensweise erstellt wurde.

The image shows two lines of musical notation in G-clef, 2/4 time, with a key signature of one flat (B-flat). The notes are stylized with varying stem directions and note heads to represent pitch contours. Above the notes are prosodic annotations: H* (high target), !H* (high accent), L-% (low boundary), L*+H (low-to-high glide), L*+!H (low-to-high accent), and L-% (low boundary).

Line 1: H* !H* L-% L*+H L*+!H L-%
{Wenn ich in **dei** - ne Au - gen seh,} {So **schwin**-det all mein **Leid** und weh}

Line 2: H* L- H% H* L*+!H L-%
{Doch wenn ich **küs** - se dei - nen Mund,} {so werd ich **ganz** und **gar** ge - sund.}

Abbildung 10: Fertige Transkription eines Tonhöhenverlaufs einer phonetischen Umsetzung des Gedichtes von Henrich Heine.

Diskussion

Der vorgestellte Ansatz zur Transkription von Sprechmelodien ist in der Annahme begründet, dass Tonhöhenverläufe gesprochener Äußerungen als Abfolge diskreter Tonhöhen aufgefasst werden können. In diesem konkreten Fall zeigen die f_0 -Messwerte, dass die Tonhöhe deutlichen Schwankungen unterliegt, und dies sowohl auf Phrasenebene als auch innerhalb der Silbe. Manche lokalen Schwankungen sind auf ›mikroprosodische Variationen‹ zurückzuführen, die in der Regel nur als »Variation der Stimmqualität« wahrgenommen werden.³⁸ Deutlich wahrnehmbar sind allerdings die Tonhöhenbewegungen im Bereich der Tonakzente. Dort

38 Peters 2014, 44 f.

liegen die gemessenen f_0 -Extremwerte weit auseinander. Aus den erhobenen Daten ist herauszulesen, dass sich die in diesem konkreten Experiment wahrgenommene Tonhöhe im Bereich der Tonakzente dem f_0 -Maximum nähert. In den restlichen Silben am Rande der Intonationsphrase ist wiederum eine Annäherung zwischen dem f_0 -Mittelwert bzw. dem f_0 -Wert am Punkt maximaler Schallintensität und dem Höreindruck festzustellen. Das lässt zweierlei Vermutungen zu: Erstens, dass das Gehör in der Lage ist, aus einem schwankenden f_0 -Verlauf einer Silbe den Eindruck einer diskreten Tonhöhe zu bilden; zweitens, dass dieser Eindruck auf unterschiedliche Merkmale der Grundfrequenzkurve zurückführen ist, je nachdem ob sich die Silbe in der Nähe eines Tonhöhenakzents oder in einer anderen wenig prominente Position innerhalb der Intonationskontur befindet. Das deutet darauf hin, dass Methoden, wie die von Meireles et al. und Menninghaus et al., die pauschal nur den Mittelwert der Grundfrequenz heranziehen, die wahrgenommene Intervallqualität nicht zuverlässig abbilden können. Diese verallgemeinernden Hypothesen müssten anhand umfangreicherer empirischer Versuche geprüft werden. Belastbare Resultate könnten nur dann erzielt werden, wenn man mehrere Tonaufnahmen verschiedener Versuchspersonen heranziehen würde.

Eine weitere Voraussetzung für die Zuverlässigkeit der Transkriptionsmethode ist, dass die wahrgenommenen diskreten Tonhöhen der Sprachmelodie im chromatischen Tonvorrat dargestellt werden können. Es ist evident, dass diese Tonhöhen in keinem unmittelbaren Gleichklang mit dem Halbton-Raster der gleichschwebenden Stimmung stehen. Vielmehr ist die Transkription das Ergebnis einer Simulation der musikalischen Wahrnehmung, bei der die Intervallqualität in den Vordergrund rückt. Die Annahme liegt nahe, dass dieser Prozess aus der in Zusammenhang mit der *Speech-to-Song Illusion* beschriebenen Fähigkeit hervorgeht, die gehörte Tonhöhenfolge perzeptorisch so anzupassen, dass sie der Vorstellung einer tonalen Melodie entspricht. Diese Anpassung kommt vermutlich durch das Phänomen des ›Zurechthörens‹ zustande. Um den Tonhöhenverlauf einer gesprochenen Äußerung in eine abstrakte Tonfolge umzuwandeln, muss das Gehör in der Lage sein, Intervallqualitäten zu erkennen, auch wenn deren Frequenzverhältnisse nicht präzise eingehalten werden. Wie von Deutsch et al. betont, beinhaltet der Prozess, der der perzeptiven Transformation einer gesprochenen Phrase in eine wohlgeformte Melodie zugrunde liegt, notwendigerweise verschiedene Abstraktionsebenen.³⁹ Nur eine vertiefte Beschäftigung mit den Erkenntnissen musikpsychologischer Forschung in diesem Gebiet kann in diesen Aspekten zu mehr Klarheit führen.

39 Deutsch et al. 2011, 2251.

Analytische Ansätze

Unter der Annahme, dass die festgehaltene Tonfolge (Abb. 10) eine gültige Annäherung an den Tonhöhenverlauf des gesprochenen Gedichtes darstellt, kann sie in Hinsicht auf ihre musikalischen Eigenschaften analysiert werden. Die folgenden Beobachtungen folgen nicht dem Anspruch, allgemeingültige musikalische Eigenschaften der Prosodie aufzudecken, sondern lediglich mögliche Anwendungen der Transkriptionsmethode aufzuzeigen. Verallgemeinerungen wären nur möglich, wenn Tonaufnahmen von verschiedenen Versuchspersonen zu verschiedenen Textvorlagen herangezogen werden würden. Insbesondere müsste bei der Auswahl der Probanden eventueller dialektaler Färbungen Rechnung getragen werden, denn diese haben einen starken Einfluss auf die Beschaffenheit der Sprechmelodie.⁴⁰ Eine weitere Bemerkung gilt der verwendeten Terminologie. Für vergleichbare Phänomene, die sich auf die Beschaffenheit von Tonhöhenverläufen beziehen, werden in der Phonologie und Musiktheorie nicht selten unterschiedliche Fachbegriffe verwendet. Im Folgenden werden nur Begriffe der Phonologie benutzt, wenn kein geeignetes musiktheoretisches Pendant zur Verfügung steht.

Was die Form anbetrifft, ist die Melodie zweiteilig, wobei der zweite Teil – bestehend aus den letzten zwei Versen – eine Variation des ersten Teils darstellt. Diese Gliederung ist nicht nur an die wiederkehrende Schlusswendung *ges-es*, sondern am sogenannten ›Pitch reset‹ am Anfang des dritten Verses zu erkennen. Unter ›Pitch reset‹ versteht man in der Phonetik die »Anhebung der Tonhöhe am Beginn einer Phrase, mit der der Abwärtstrend der vorhergehenden Phrase aufgehoben wird«. ⁴¹ Außerdem kann man am Verlauf der Grundlinie des ersten Teils erkennen, dass die Melodie eine übergeordnete fallende Tendenz aufweist (Abb. 11). Die Grundlinie besteht aus der Fortschreitung eines tiefen Tons zum nächsten tieferen Ton des Melodieverlaufs. Diese tiefen Hauptpunkte bilden einen glatten Sekundgang,⁴² der der Melodie der zwei ersten Verse eine einheitliche Richtung verleiht. Auch im zweiten Teil (Verse 3 und 4) kann diese abfallende Tendenz beobachtet werden. Allerdings laufen dort zwei Sekundgänge nebeneinander: Das *f* zu Beginn der Phrase findet seine Fortsetzung in der Silbe ›gar‹, während bei ›küsse‹ ein neuer fallender Stufengang einsetzt.

40 Peters 2014, 87 ff.

41 Ebd., 41.

42 Vgl. Hindemith 1940, 228 ff.

Abbildung 11: Analytische Darstellung des transkribierten Tonhöhenverlaufs. Im obersten System wird die Dachlinie dargestellt, im untersten System die Grundlinie, die wegen des Deklinationstrends jeweils zwei Intonationsphrasen zu einem Sinnzusammenhang verbindet.

Die Dachlinie, die die Hochtöne verbindet, weist im Gegensatz zur Grundlinie keinen Sekundgang auf. Stattdessen bilden die Gipfeltöne der vorderen Teilhälften (α , γ) und die jeweiligen finalen Grenztöne reine Quartan in Abwärtsbewegung (c^1-g und e^1-h). In den hinteren Teilhälften (β , β') ist demgegenüber die Wiederkehr der übermäßigen Quarte (c^1-ges) zwischen den intonatorischen Gipfeltönen zu beobachten. Diese unterschiedlichen Intervallqualitäten der Dachlinie tragen zur Kontrastbildung zwischen den vorderen und den hinteren Teilhälften bei. In Übereinstimmung mit der Versmetrik und den Phrasengrenzen, kann demnach jeder Teil in zwei kontrastierende musikalische Phrasen untergliedert werden. Darüber hinaus lässt sich die Phrase γ als eine variierte Wiederholung von α analysieren. Beibehalten wird bei der Wiederholung neben der fallenden Quarte der Dachlinie auch der Tonakzent, dessen Gipfel bei γ allerdings durch das Phänomen des ›Upstep‹ modifiziert wird. Diese Heraufstufung des Akzenttons um eine große Terz führt zur Vergrößerung des Tonumfangs. Hinzu kommt ein Wiederanstieg der Tonhöhe am Ende der Phrase γ , wodurch der Grenztöne als verhältnismäßig hoch gehört wird, was eine melodische Fort-

setzung erwarten lässt.⁴³ Trotz der Unterschiede zwischen α und γ , bilden beide jeweils kontrastierende Einheiten zu β und β' . Letztere unterscheiden sich von α und γ vor allem durch ihre Schlusswirkung und gleichbleibenden Stimmumfang. Einen Vergleich mit dem motivischen Aufbau einer Periode liegt nahe: Die Sprechmelodie gliedert sich in einen Vordersatz, der eine Phrase und eine Gegenphrase umfasst, und einen Nachsatz, der aus einer veränderten Wiederholung besteht.⁴⁴

In den bisherigen Ausführungen wurde der Tonhöhenverlauf auf seine formgebenden Aspekte untersucht. Ein weiterer Punkt, der nähere Betrachtung verdient, ist die Entfaltung von Tonbeziehungen in der inneren Intervallstruktur der Melodie. Da die vorliegende Transkription Ergebnis eines approximativen Verfahrens ist, sind Tonbeziehungen nicht im Sinne eines Tonsystems – wie etwa in einer Tonalität – aufzufassen. Vielmehr geht es darum, intervallische Strukturen aufzudecken, die wegen ihres wiederholten Auftretens die melodische Gestaltung prägen. Eine solche Struktur entfaltet sich in Zusammenhang mit den Gipfeltönen der Melodie. Mit Ausnahme der Silbe ›kü(s)e‹ werden die gipfelnden Hochtöne mit der Tonhöhe c^1 realisiert. Beim Vergleichen der Messwerte kann bestätigt werden, dass das f_0 -Maximum in den betreffenden Hochtönen nur in geringem Maße voneinander abweicht (Abb. 12). Der wiederholte Einsatz des Tones c^1 als Gipfel der Intonationsphrase gewinnt an Bedeutung, wenn man die Tonhöhen in ihrer melodischen Umgebung betrachtet. In den Phrasen β und β' bildet das c^1 die Oberquinte zum darauffolgenden f . Dasselbe kann in der Phrase γ beobachtet werden, wenn der Akzentton durch ›Upstet‹ auf e^1 heraufgestuft wird: Dort bildet das e^1 ebenfalls eine Quinte zu den Tönen in seiner unmittelbaren Umgebung. Der Akzent scheint hier mit dem Intervall der Quinte in Beziehung zu stehen. Wegen möglicher Analogien zur Musik – z.B. zum Reperkussionston in der Psalmodie – verdient dieser Aspekt in zukünftigen Analysen sicherlich intensivere Betrachtung.

	FreqMax in Hz	Intervall zum Notierten Referenzton c^1 (261,626 Hz) in Cent
dei(ne)	260,7	-6
(schwin)det	267,1	36
küs(se)	333,7	421
ganz	256,5	-34

Abbildung 12: Messwerte (f_0 -Maximum) am Gipfelton der jeweiligen Intonationsphrase; die Gipfeltöne werden – mit Ausnahme der dritten Phrase (›küsse‹) – auf etwa gleicher Tonhöhe realisiert.

43 Peters 2014, 58.

44 Hier wird die von Ulrich Kaiser verwendete Terminologie übernommen, vgl. Kaiser 2015.

Zusammenfassung und Ausblick

Sowohl beim Singen als auch beim Sprechen unterliegt die Sprache einer melodischen Gestaltung. Doch während beim Singen die Melodie im Vordergrund steht, ist sie im Gesprochenen nicht ohne Weiteres vom Wortlaut zu entkoppeln. Bei größerer emotionaler Beteiligung des Sprechenden, wie etwa bei der Deklamation von Gedichten, kommt die Sprechmelodie deutlicher hervor. Aber selbst wenn die Sprechmelodie ausdrucksvoller gestaltet wird, erfordert die Bewusstmachung des Tonhöhenverlaufs ein gezieltes Hinhören, was meist nur durch Wiederholung der Äußerung ermöglicht wird. Beim Sprechen verweilt die Stimme oft nicht wie beim Singen auf stabilen Tonhöhen, sondern sie gleitet von einem Zielpunkt zum nächsten. Unter den oben beschriebenen Umständen können diese gleitenden Tonhöhenbewegungen als Abfolge diskreter Tonhöhen gehört werden und damit den perzeptiven Eindruck einer tonalen Melodie erwecken.

Die vorgestellten Messungen zeigen, dass sich die vom Gehör zurechtgebogene Melodie mit ziemlicher Genauigkeit an dem Grundfrequenzverlauf des Sprechsignals ausrichtet. Die nach dem dargelegten Verfahren erstellte Transkription bildet somit eine approximative Darstellung des Tonhöhenverlaufs, die die erfassten Intervallqualitäten abgebildet. Obwohl der messtechnische Abgleich weitreichende Übereinstimmung zwischen der vom Gehör erfassten Tonhöhe und dem Grundfrequenzverlauf zeigt, kann nicht über die Subjektivität der geschilderten Transkriptionsmethode hinweggesehen werden. Die transkribierten Melodien sind kein bloßes Abbild des Grundfrequenzverlaufs, sondern vielmehr eine Repräsentation einer individuellen Klangvorstellung, die bei der Auseinandersetzung mit dem Phänomen entsteht. Eine weiterführende Untersuchung könnte also darin bestehen, Transkriptionen mehrerer Hörer*innen mit Messwerten des Grundfrequenzverlaufs zu vergleichen. Der Vergleich würde zum einen dazu dienen, den Grad der Subjektivität des Tonhöhereindrucks zu ermitteln, zum anderen die Zuverlässigkeit der Transkriptionsmethode zu prüfen.

Durch ein Repräsentationsmodell, das die Sprechmelodie als ein musikalisches Phänomen auffasst, können Aspekte beleuchtet werden, die über den Erkenntnisbereich der Linguistik hinausgehen. Darunter fallen die intervallische Struktur der Melodie und die damit verbundenen Regelmäßigkeiten. Die vorgestellten analytischen Ansätze könnten als Ausgangspunkt für zukünftige Untersuchungen dienen. Dabei wäre der Frage nachzugehen, inwiefern die in der exemplarisch transkribierten Sprechmelodie festgestellten musikalischen Eigenschaften bloßer Zufall oder doch Manifestation der Regelmäßigkeit der Prosodie – in diesem Fall

der deutschen Sprache – sind. Konkret wäre zu erörtern, wie Kontrast und Similarität auf motivischer Ebene entstehen, oder inwiefern Sekundgänge der intervallischen Struktur der Sprechmelodie eine latente Linearität verleihen. Ferner könnte der Frage nachgegangen werden, ob bestimmte Intervallqualitäten mit bestimmten sprachlichen Funktionen (z.B. Akzentsetzung) assoziiert sind, und inwiefern sich in Sprechmelodien tonale Zentren befestigen. Allenfalls wäre es unabdingbar ein umfangreicheres Korpus an Transkriptionen zu benutzen, das auf Tonaufnahmen verschiedener Sprecher und verschiedener Textvorlagen basiert.

Aus der Analyse der intervallischen Struktur der Sprechmelodie ergibt sich zudem die Möglichkeit, den Tonhöhenverlauf des Gesprochenen mit komponierten Vertonungen zu vergleichen. Es wäre als nächster Schritt naheliegend, die transkribierte Sprechmelodie zu Heinrich Heines Gedicht auf Übereinstimmungen mit der berühmten Vertonung von Robert Schumann (*Dichterliebe*, op.48, Nr.4) zu untersuchen. Diese Art vergleichender Analyse, wenn sie sich auf andere Sprecher und andere Vertonungen ausweitet, könnte Aufschluss darüber geben, inwiefern sprechmelodische Gestaltungsprinzipien in komponierter Musik zur Geltung kommen. In einer größer angelegten Studie könnte der Frage nachgegangen werden, wie die melodische Kreativität von Komponist*innen von den prosodischen Regelmäßigkeiten ihrer Sozialisationsprache beeinflusst wird, wenn sie diese vertonen. Untersuchungen aus dem Gebiet der Musikpsychologie zeigen, dass solche Interferenzen durchaus statistisch nachweisbar sind.⁴⁵ Es läge im Bereich der Musiktheorie zu erörtern, wie Melodik und Prosodie in konkreten Musikwerken zusammenhängen. Hierfür ist es unabdingbar, musikalische Melodien unter Einbeziehung sprachwissenschaftlicher Erkenntnisse zu betrachten. Das Musikalische an der Sprache zu enthüllen, nicht zuletzt in ihrem Tonhöhenverlauf, ist der erste Schritt in Richtung einer aufgewerteten ›Melodielehre‹, in der die Analogien zwischen sprachlichen und musikalischen Melodien im Fokus stehen.

Neben dem Einsatz in der musikalischen Analyse kann der vorgestellte methodische Ansatz auch in der künstlerischen Praxis angewandt werden. Komponist*innen, die eine sprachgerechte Vertonung beabsichtigen, jedoch die zu vertonende Sprache nicht beherrschen, können sich anhand von Sprachaufnahmen die natürliche Prosodie aneignen und bessere Textverständlichkeit erzielen. Au-

45 Vgl. Zusammenfassung von Forschungsergebnissen in Dombrowski 1995, 130; sowie die Untersuchungen von Patel et al. 2006.

ßerdem liefert die vorgestellte Methode Tonfolgen, die als *objects trouvés* kompositorisch verarbeitet werden können, auch wenn der textliche Inhalt ausbleibt.

Die Betrachtung von Sprechmelodien, so man sie Textvertونungen gleichsetzt, führt schließlich zur Frage: Sprechen wir noch oder singen wir schon?

Literatur

- Blühdorn, Hardarik (2013), »Intonation im Deutschen – nur eine Frage des schönen Klangs?«, *Pandaemonium*, 16/22, 242–278. <https://doi.org/10.1590/S1982-88372013000200013>
- Debbeler, Judith (2007), *Harmonie und Perspektive. Die Entstehung des neuzeitlichen abendländischen Kunstmusiksystems*, München: Epodium.
- Deutsch, Diana / Trevor Henthorn / Rachael Lapidis (2011), »Illusory transformation from speech to song«, *Journal of the Acoustical Society of America*, 129, 2245–2252. <https://doi.org/10.1121/1.3562174>
- Deutsche Forschungsgesellschaft (hg.) (2013), *Empfehlungen zu datentechnischen Standards und Tools bei der Erhebung von Sprachkorpora*. https://www.dfg.de/download/pdf/foerderung/grundlagen_dfg_foerderung/informationen_fachwissenschaften/geisteswissenschaften/standards_sprachkorpora.pdf (17.9.2019).
- Dombrowski, Ernst (1995), »Über strukturelle Gemeinsamkeiten zwischen sprachlichen und musikalischen Melodien«, in: *Musikpsychologie. Jahrbuch der Deutschen Gesellschaft für Musikpsychologie* 12, 110–133.
- Engels, Inga (2011), *Vergleichende Prosodie Lettisch-Deutsch*, Mannheim: Institut für Deutsche Sprache. https://ids-pub.bsz-bw.de/frontdoor/deliver/index/docId/23/file/Engels_Vergleichende_Prosodie_Lettisch_Deutsch_2011.pdf (19.9.2019)
- Fricke, Jobst Peter (2012), *Intonation und musikalisches Hören*, Osnabrück: Electronic Publishing Osnabrück. <https://www.epos.uni-osnabrueck.de/books/f/frij012/OnlineBook/> (28.10.2019)
- Fuhrhop, Nanna / Jörg Peters (2013), *Einführung in die Phonologie und Graphematik*, Stuttgart: Metzler. <https://doi.org/10.1007/978-3-476-00597-7>
- Gebhardt, August (1907), *Grammatik der Nürnberger Mundart*, hg. von Otto Brenner, Leipzig: Breitkopf & Härtel.
- Goethe, Johann Wolfgang von (1833), »Regeln für Schauspieler« [1803], in: *Goethes Werke. Vollständige Ausgabe letzter Hand*, 44. Bd., Stuttgart, Tübingen: J.G. Cotta, 293–315.
- Grice, Martine / Stefan Baumann (2002), »Deutsche Intonation und GToBI«, *Linguistische Berichte* 191, 267–298.
- Grice, Martine / Stefan Baumann / Simon Ritter / Christine Röhr (o.J.), *Übungsmaterialien zur deutschen Intonation und GToBI*. Universität Köln. <http://www.gtobi.uni-koeln.de> (4.10.2023)
- Haas, Max (2016), »Notation, Neumen, Über Ursprung und Gebrauch von Neumen, Der Ursprung der Neumen« [1997], in: *MGG Online*, hg. von Laurenz Lütteken, Kassel: Bärenreiter. <https://www.mgg-online.com/mgg/stable/51366>

- Halicarnassus, Dionysus of (1910), *On Literary Composition. Being the Greek Text of the De Compositione Verborum. Edited with introduction, translation, notes, glossary, and appendices*, hg. und übers. von W. Rhys Roberts, London: Macmillan and co. <https://archive.org/details/cu31924026465165> (4.10.2023)
- Hart, Johan t / René Collier / Antonie Cohen (1990), *A perceptual study of intonation. An experimental-phonetic approach to speech melody*, Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511627743>
- Heilig, Otto (1898), *Grammatik der Ostfränkischen Mundart des Taibergrundes und der Nachbarmundarten*, hg. von Otto Brenner, Leipzig: Breitkopf & Härtel.
- Hindemith, Paul (1940), *Unterweisung im Tonsatz*, Bd. 1, Mainz: B. Schott's Söhne.
- Kaiser, Ulrich (2015), *Formenlehre IV – Periode und Satz*. <http://www.musikanalyse.net/tutorials/periode-und-satz/> (4.10.2023).
- Kiefer, Reinhard (2016), »Sprachmelodie. Experimentelle Erfassung« [1998], in: *MGG Online*, hg. von Laurenz Lütteken, Kassel: Bärenreiter. <https://www.mgg-online.com/mgg/stable/51380>
- Köhler, Louis (1853), *Melodie der Sprache in ihrer Anwendung besonders auf das Lied und die Oper mit Berührung verwandter Kunstfragen*, Leipzig: J. J. Weber.
- Meireles, Alexsandro R. / Antônio R. M. Simões / Antonio Celso Ribeiro / Beatriz Raposo Meideiros (2017), »Musical Speech: A New Methodology for Transcribing Speech Prosody«, in: *Proceedings of Interspeech 2017*. <http://dx.doi.org/10.21437/Interspeech.2017-316>
- Menninghaus, Winfried / Valentin Wagner / Christine A. Knoop / Mathias Scharinger (2018), »Poetic speech melody: A crucial link between music and language«, *PLoS ONE* 13(11), Artikel e0205980. <https://doi.org/10.1371/journal.pone.0205980>
- Mertens, Piet (2004), »The Prosogram: Semi-Automatic Transcription of Prosody Based on a Tonal perception Model«, *Proceedings of Speech Prosody 2004*. https://www.isca-speech.org/archive/pdfs/speechprosody_2004/mertens04_speechprosody.pdf (4.10.2023)
- Patel, Aniruddh D. / John R. Iversen / Jason C. Rosenberg (2006), »Comparing the rhythm and melody of speech and music: The case of British English and French«, *Journal of the Acoustical Society of America* 119/5, 3034–3047, <https://doi.org/10.1121/1.2179657>
- Pearl, Jonathan Geoffrey Secora (2006), »Eavesdropping with a Master: Leo Janáček and the Music of Speech«, in: *Empirical Musicology Review*, Vol. 1, Nr. 3, 131–165. <https://doi.org/10.18061/1811/24010>
- Peters, Jörg (2014), *Intonation*, Heidelberg: Winter.
- Steele, Joshua (1779), *Prosodia Rationalis: or, an essay towards establishing the melody and measure of speech, to be expressed and perpetuated by peculiar symbols*, London: J. Nichols. <https://archive.org/details/prosodiarationa01steegoog/page/n10> (4.10.2023)
- Tierney, Adam / Aniruddh Patel / Maren Breen (2018), »Acoustic foundations of the speech-to-song illusion«, *Journal of Experimental Psychology: General* 147(6), 888–904. <https://doi.org/10.1037/xge0000455>
- Vainiomäki, Tiina (2012), *The Musical Realism of Leoš Janáček. From Speech Melodies to a Theory of Composition*, Phil. Diss., University of Helsinki. <https://helda.helsinki.fi/handle/10138/36087> (4.10.2023)

Vanden Bosch der Nederlanden, Christina / Erin Hannon / Joel Snyder (2015), »Everyday Musical Experience is Sufficient to Perceive the Speech-to-Song Illusion«, *Journal of Experimental Psychology: General* 144(2), e43–e49. <https://doi.org/10.1037/xge0000056>
Vendryes, Joseph (1904), *Traité d'accentuation grecque*, Paris: Librairie C. Klincksieck.

© 2023 Manuel Durão (manueldurao@hotmail.com, ORCID iD: 0009-0008-8565-5722)

Staatliche Hochschule für Musik und Darstellende Kunst Mannheim, Staatliche Hochschule für Musik Trossingen [University of music and performing arts Mannheim, University of Music Trossingen]

Durão, Manuel (2023), »Notation und Analyse von Tonhöhenverläufen in Sprechmelodien« [Notation and Analysis of Pitch Tracks in Speech Melodies], in: *Notation. Schnittstelle zwischen Komposition, Interpretation und Analyse. 19. Jahreskongress der Gesellschaft für Musiktheorie* (GMTH Proceedings 2019), hg. von Philippe Kocher, 115–139. <https://doi.org/10.31751/p.259>

eingereicht / submitted: 15/06/2020

angenommen / accepted: 30/05/2021

veröffentlicht / first published: 20/12/2023

zuletzt geändert / last updated: 20/12/2023