

## METHODEN ZUR ANALYSE DER VOKALEN GESTALTUNG POPULÄRER MUSIK<sup>1</sup>

**Tilo Hähnel, Tobias Marx, Martin Pfeleiderer**

Nachdem Popgesang sowohl von der musikwissenschaftlichen Forschung als auch von der Gesangspädagogik lange ignoriert oder sogar pauschal als minderwertig und stimmschädigend diffamiert worden ist, zeichnet sich seit einiger Zeit ein Wandel in der Einschätzung der kulturellen Bedeutung von Popstimmen und Popgesang ab (vgl. Pfeleiderer 2009). Das wachsende Interesse an der vokalen Gestaltung populärer Musik erfordert adäquate Methoden einer analytischen Beschreibung der charakteristischen Eigenheiten von Popstimmen und deren vokalen Gestaltungsmitteln.

Ausgangspunkt der analytischen Auseinandersetzung mit Stimme und Gesang in populärer Musik sind die unzähligen Tondokumente, die seit Beginn des 20. Jahrhunderts entstanden sind. Eine herkömmliche musikwissenschaftliche Herangehensweise an Klangaufnahmen ist deren Transkription, bei der auf der Grundlage eines wiederholten minutiösen Hörprozesses ein mehr oder weniger detailgenauer Notentext erstellt wird. Dieser Notentext fungiert sodann als Gegenstand der musikalischen Analyse, oder aber er dient dazu, die bereits beim Hören – und umso mehr beim intensiven Hören während des Transkribierens – gewonnenen Einsichten zu verdeutlichen und darzustellen (vgl. Winkler 1997). In der konventionellen europäischen Notenschrift werden allerdings viele jener klanglichen Aspekte vernachlässigt, die gerade für den vokalen Ausdruck wichtig sind. So liegt eine Besonderheit des populären Stimmgebrauchs in der Annäherung an die alltägliche Sprechstimme, im Wechsel zwischen gesprochenen und gesungenen Wörtern oder

---

<sup>1</sup> Der Text basiert auf einem Vortrag, der bei der 23. Arbeitstagung des Arbeitskreis Studium Populärer Musik e.V. am 23.11.2012 in Basel gehalten wurde. Für die einzelnen Abschnitte des Textes zeichnen die Autoren Martin Pfeleiderer (Einleitung, Fazit), Tilo Hähnel (Melodischer Verlauf, Info-Boxen) und Tobias Marx (Stimmklang) verantwortlich.

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. Viele dieser Besonderheiten lassen sich in notenschriftlichen Transkriptionen nur schwer detailliert einfangen, ähnlich wie die spezifischen Klangfarben und Klangqualitäten, die so vielfältig sind wie bei der alltäglichen Stimme – vom Hauchen und Brummen bis zum Rufen und Schreien.

Um die alltäglichen, an die Sprechstimme angelehnten Qualitäten des populären Gesangs exakt zu beschreiben und so einer vergleichenden Analyse zugänglich zu machen, ist somit die Darstellung in Form eines Notentextes, der als Grundlage einer Analyse der melodischen und rhythmisch-metrischen Gestaltung durchaus sinnvoll sein kann, zumeist nicht ausreichend. Vielmehr ist es erforderlich, neue Analysemethoden und Darstellungsweisen heranzuziehen, wie sie vor allem durch die rechnergestützte Auswertung digitalisierter Klangaufnahmen möglich geworden sind. Wir möchten diese Möglichkeiten, die zum großen Teil auch auf andere Bereiche der Analyse von populärer Musik übertragbar sind, an einer Reihe von Beispielen verdeutlichen: dem Gleiten zwischen Tonhöhen, dem individuellen Vibrato, der klanglichen Gestaltung von Vokalen und der rauen Stimmgebung.

Zuvor sollen jedoch einige der Fragestellungen und theoretische Überlegungen zur Popstimme skizziert werden, die im Hintergrund der hier vorgestellten analytischen Zugänge stehen und die im Rahmen des Forschungsprojekts *Stimme und Gesang in der populären Musik der USA (1900-1960)* entwickelt worden sind. Das von der Deutschen Forschungsgemeinschaft geförderte Projekt hat sich die Analyse und Interpretation von vokalen Ausdrucksweisen und Gestaltungsmitteln populärer Musik zur Aufgabe gestellt. Wohlgermerkt sollen die Ergebnisse auch jenseits des gewählten Untersuchungszeitraums, in dem die populäre Musik der USA bekanntlich weltweit andere regionale Traditionen beeinflusst, überlagert oder sogar verdrängt hat, einer wissenschaftlichen Auseinandersetzung mit Stimme und Gesang methodologische und inhaltliche Anregungen geben. Ausgangspunkt der Untersuchungen ist die Entwicklung adäquater Analyse- und Darstellungs-

methoden sowie einer umfassenden und tragfähigen Systematik vokaler Klangtypen und Gestaltungsmittel. Übergreifendes Ziel ist die Deutung von Stimmen und Singweisen in deren sozial- und kulturgeschichtlichen Kontexten. Hierzu wurde ein theoretischer Zugang entwickelt (vgl. Pfeleiderer im Druck), der vor allem an Überlegungen von Simon Frith anknüpft und an dieser Stelle kurz angedeutet werden soll.

Nach Simon Frith (1996: 183-225) kommen bei der populären Stimm-Performance mehrere Dimensionen der Stimme in unterschiedlicher Gewichtung zur Geltung: die Stimme als Musikinstrument, die Stimme als Teil des Körpers, die Stimme als Ausdruck einer Person und schließlich die Stimme als »character« bzw. Rolle im Rahmen eines Song-Textes und dessen Interpretation oder im übergreifenden Kontext die Aufführungspersönlichkeit und Bühnenrolle eines Sängers. Friths Unterscheidung in Privatperson, Aufführungsperson und Songperson ist inzwischen von verschiedenen Autoren aufgegriffen und weiterentwickelt worden, so von Richard Middleton (2000), Philip Auslander (2009) und Allan F. Moore (2012: 179-188). Insbesondere die Aufführungsperson (»performance persona«) und ihr Verhältnis zur Privatperson und Songperson bieten Anknüpfungspunkte für eine Interpretation von Popstimmen innerhalb bestimmter Images und kultureller Ausdrucksmuster. Interessant ist hier vor allem die Frage, wie vokale Gestaltungsmittel mit anderen performativen Ausdrucksmitteln in der Gesamterscheinung eines Sängers zusammenwirken: Welche Rolle spielt die Stimme im intermedialen Gesamtkontext der Aufführungen eines Pop-Sängers, seiner Medienpersönlichkeit und seines Images? Inwiefern greift der Vokalstil auf übergreifende kulturgeschichtliche Ausdrucksmuster zurück, die er umgekehrt auch wieder prägt oder verändert? Doch auch die von Frith genannten Aspekte der Körperlichkeit der Stimme und ihrer quasi-instrumentalen Virtuosität spielen in vielen Bereichen der populären Musik eine wichtige Rolle und liefern weitere Fragestellungen für die musikalische Analyse.

Anders als im klassischen Gesang, in dem die vokalen Ausdrucksmittel durch eine »legitime« Gesangstechnik normiert werden – »legit« (von engl. »legitimate«) ist eine im Englischen weit verbreitete Bezeichnung des klassischen Gesangsideals –, gibt es in der populären Musik eine zunächst verwirrend große Vielfalt von vokalen Gestaltungsmitteln und Vokalstilen, die jedoch auch bestimmte Gemeinsamkeiten besitzen. Zentrale Fragestellungen der Analyse von Vokalaufnahmen lauten daher: Welche vokalen Gestaltungsmittel sind in den verschiedenen Genres populärer Musik anzutreffen? Welche Gestaltungsmittel sind genretypisch, welche werden genreübergreifend eingesetzt? Und schließlich: Welche Eigenheiten der vokalen Gestaltung machen die Popstimme eines Vokalistens einzigartig und unverwechsel-

bar? Voraussetzung für eine kulturelle Deutung von Vokalstilen und vokalen Ausdrucksmitteln ist dabei deren möglichst präzise analytische Beschreibung.

## Melodischer Verlauf

Bei der Analyse des melodischen Verlaufs geht es nicht um die Rekonstruktion einer vermeintlichen Spielvorlage, sondern vielmehr darum, wie ein Interpret eine Melodie im Detail ausführt bzw. ausgestaltet. Vokalist:innen verwenden eine Vielzahl verschiedener Typen des Tonhöhengleitens (Glissando), angefangen bei der expressiven Ausgestaltung von Blue Notes über das willkürliche und unwillkürliche Anschleifen und Ziehen von Tönen bis hin zur Gestaltung komplexer Gebilde, bei denen verschiedene Glissandi melismatisch verkettet werden.

Interaktive rechnergestützte Werkzeuge erlauben es, diese Tonhöhenverläufe in einem Detailgrad zu beschreiben, der weit über die Möglichkeiten der traditionellen Notenschrift hinausgeht (Senn 2007). Eine wesentliche Bereicherung der Gesangsanalyse wird die Arbeit mit den Tonhöhendaten dann, wenn sie nicht bei der bloßen »Bildinterpretation« einer melodischen Kontur stehen bleibt: Wenige Vor- und Nachverarbeitungsschritte ermöglichen beispielsweise eine kompakte Darstellung, Analyse und Auswertung von Vibrati, die nicht nur stationär als »Zustand«, sondern auch in ihren zeitlichen Verläufen erfasst werden können.

Für die Extraktion von Tonhöhendaten gibt es zahlreiche Ansätze.<sup>2</sup> In der Forschungspraxis hat sich in den letzten Jahren immer mehr die frei verfügbare Analyse-Software Sonic Visualiser durchgesetzt. An der Londoner Queen Mary Universität entwickelt, bietet sie den Vorteil, dass Analyse-Algorithmen in Form von Plugins selbst programmiert und eingebunden werden können. Ausgangspunkt der nachfolgend verwendeten Methoden sind Tonhöhendaten des Aubio Pitch Detection Algorithmus (Brossier 2006), der im Sonic Visualiser als Plugin zur Verfügung steht. Aufgrund einer Analyse des Frequenzspektrums in Zeitfenstern von ca. 10 ms berechnet der Algorithmus den Grundtonverlauf in einer Klangaufnahme. Die Daten können im Sonic Visualiser als Grundtonhöhenkurve angezeigt werden und lassen sich darüber hinaus als eine Reihe von Wertepaaren (Grundtonfrequenz zur jeweiligen Zeit) exportieren, was anschließend weitere Rechenoperationen ermöglicht:

---

2 Vgl. ausführlich z.B. Rossignol/Desain/Honing (2001) oder Jamaludin et al. (2012).

- Umrechnung von Hertz nach Cent. So sind Tonhöhen als musikalische Intervalle darstellbar.
- Bereinigung von Fehlern. Teilweise verrauschte Daten können wesentlich verbessert werden, indem Ausreißer eliminiert oder Oktav-Verwechslungen zurückgerechnet werden (siehe Box 1).
- Weiterverarbeitung der Tonhöhendaten. So ist es möglich, Häufigkeitsverteilungen von Tonhöhen anzeigen zu lassen oder über die Änderung der Tonhöhenkurve die Gleichmäßigkeit eines Vibratos sichtbar zu machen (siehe Abb. 4).
- Datenreduktion. Zum Beispiel muss ein Vibrato nicht mehr als eine Reihe von Tonhöhendaten im Zeitverlauf dargestellt, sondern kann direkt auf die zwei Kennwerte Frequenz (Geschwindigkeit des Vibratos) und Amplitude (Auslenkung in Cent) abgebildet werden (siehe Abb. 5); die Kennwerte der Vibrati werden dabei über die Annäherung einer Sinusfunktion geschätzt (siehe Box 2).

Für die wissenschaftliche Arbeit ist es essentiell, dass bei der Manipulation von Daten alle Parameter transparent und unter der Kontrolle des Wissenschaftlers bleiben – was bei kommerziell verbreiteter Software nicht immer der Fall ist. Die Auseinandersetzung mit den Algorithmen und ihren Möglichkeiten mag ungewohnt erscheinen, der zeitliche Aufwand der Einarbeitung ist jedoch machbar und lohnenswert.

Die graphische Darstellung von zeitabhängigen Daten, z.B. dem Tonhöhenverlauf innerhalb eines Musikstücks, erlaubt es, die bereits kurz nach dem Moment ihres Entstehens wieder vergehenden Klänge zu fixieren und so einer näheren Untersuchung zugänglich zu machen. Da sowohl in der wissenschaftlichen Analyse als auch in der Kommunikation der Forschungsergebnisse eine möglichst neutrale und zeitenthobene Darstellung angestrebt wird, leistet die graphische Repräsentation des akustischen Signals hier einen entscheidenden Mehrwert: Das Argument »Weil ich den Tonhöhenverlauf so empfinde« wird mit dem Argument »Weil die Tonhöhe so verläuft« aufgewertet und ergänzt. Voraussetzung für eine adäquate Darstellung ist dabei natürlich, dass der Analysierende Fragestellungen formuliert oder aufgrund seines Höreindrucks bereits Hypothesen zum Klanggeschehen gebildet hat. Zugleich ist nicht ausgeschlossen, dass ein subjektiver Höreindruck mit dem »akustischen Tatbestand«, wie er mit Hilfe des Computers gemessen und dargestellt wird, in Widerspruch gerät. Dies kann jedoch durchaus zu neuen Fragestellungen führen.

Letztendlich ist und bleibt die Höranalyse der entscheidende Ausgangspunkt bei der Auseinandersetzung mit dem klingenden Gegenstand. Aller-

dings war sich bereits John Fahey, der Biograph des Bluesmusikers Charley Patton, dessen bewusst, dass eine Transkription nach Gehör – und sei es noch so gut – am Ende durch eine weitere, vom analysierenden Subjekt unabhängige Analyse unterfüttert werden sollte: »Without a machine like a Melograph [...], which would adequately classify the pitches of Patton's sounds, we are left with our own presumably good ears with which to discard the autological neutral pitches« (Fahey 1970: 34).

Fahey vermutete, dass Patton die Blue Note nicht als neutrale Terz singt. Dies lässt sich heute nicht nur bestätigen, vielmehr lässt sich exakt darstellen, dass die Blues-Terz bei Charley Patton ein geradliniges Aufwärtsglissando von der kleinen zur großen Terz ist. Also nicht die Tonhöhe selbst ist hier die Blue Note, sondern die Tonhöhenbewegung (siehe Abb. 1).

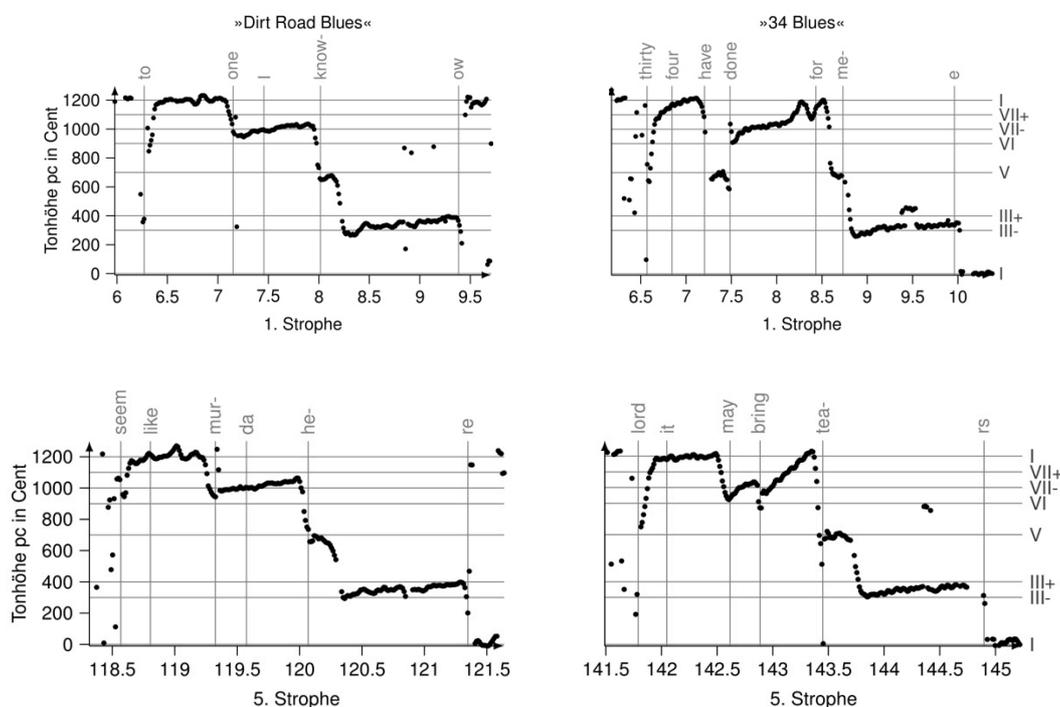
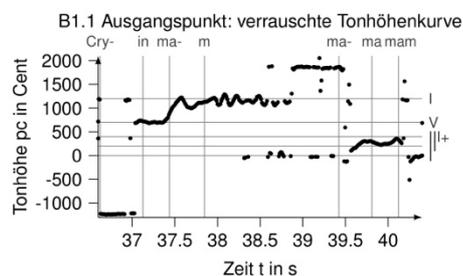


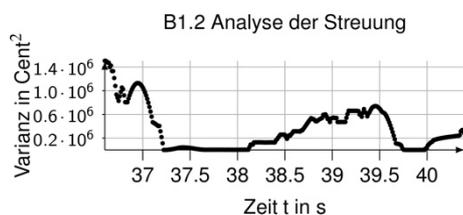
Abbildung 1: Tonhöhenverläufe in Songs von Charley Patton, links: »Down The Dirt Road Blues« (1929), rechts: »34 Blues« (1934). Die horizontale Achse zeigt den Zeitpunkt der Tonaufnahme in Sekunden. Die Bewegung der Blues-Terz bei Charley Patton ist nicht nur über einzelne Stücke, sondern sogar über mehrere Stücke stabil, wie der Vergleich der beiden Blues-Einspielungen aus Pattons erster und letzter Aufnahmesession zeigt. Die Tonstufen sind als gleichstufig temperierte Tonhöhen markiert (kleine Terz = III-; große Terz = III+).

**Box 1. Aufbereiten der Tonhöhendaten**

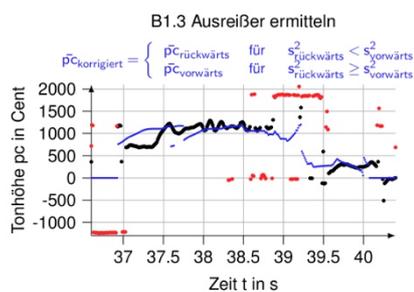
Bild B1.1 zeigt einen Ausschnitt aus Tommy Johnsons »Canned Heat Blues« (1928). Die Daten sind durch sog. Ausreißer stark verrauscht, wobei sich die gesungene Melodie durchaus erahnen lässt. Ausreißer werden über den Abstand der Tonhöhe zu einem korrigierten Mittelwert der Tonhöhe ( $\bar{pc}_{\text{korrigiert}}$  wobei pc hier »pitch cent« bedeutet) bestimmt.



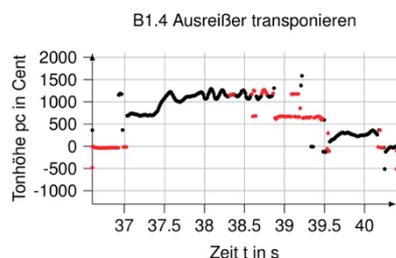
Für den Mittelwert  $\bar{pc}$  an einem beliebigen Zeitpunkt  $t$  werden zwei sich überlappende Zeitfenster gewählt, wobei ein Fenster nach vorn (entspricht einer Rechtsverschiebung auf der Zeitachse) und das zweite Fenster nach hinten verschoben wird (Linksverschiebung).



Der korrigierte Mittelwert  $\bar{pc}_{\text{korrigiert}}$  ist derjenige, der in dem Fenster liegt, in dem die Daten weniger verrauscht sind, d.h. die Varianz  $s^2$  (siehe B1.2) geringer ist (siehe B1.3). Die korrigierte Mittelwertkurve ist konservativ, d.h. sie erkennt stabile Phasen, behält sie bei und antizipiert sie. In B1.3 ist deutlich zu sehen, wie die korrigierte Mittelwertkurve (dünne blaue Linie) bei Sekunde 39 auf der Tonhöhe der vorausgehenden Phrase bleibt – da die Varianz vor Sekunde 39 geringer ist als danach (vgl. B1.2). Das Fallen der Varianz ab Sekunde 39,5 bewirkt ein »Umschalten« der korrigierten Mittelwertkurve bereits bei Sekunde 39,25, so dass die Mittelwertkurve den Phrasenanfang bei Sekunde 39,5 antizipiert.



Häufig sind Ausreißer auf Oktav-Verwechslungen zurückzuführen: Bei der Berechnung der Grundtonhöhe aus dem Frequenzspektrum des Klangs »verwechselt« der Algorithmus einen Oberton mit dem Grundton oder den Grundton mit dem ersten Oberton, der Oktave. Der Ausreißer liegt in diesen Fällen exakt 1200 Cent über oder unter dem eigentlichen Ton.



Werden gefundene Ausreißer um eine Oktave (1200 Cent) in Richtung der korrigierten Mittelwertkurve transponiert (rot in B1.4), können wichtige Informationen zurückgewonnen werden.

Ein weiteres Beispiel dafür, dass die computergestützte Darstellung Details der melodischen Gestaltung sichtbar werden lässt, die in einer Transkription notgedrungen verloren gehen, ist das Anschleifen der Töne bei der Gospel-sängerin Mahalia Jackson. In »There's Not A Friend Like Jesus« (Abb. 2) ist das Anschleifen besonders prägnant. Jackson singt wiederholt ein markant gebogenes Aufwärtsgleiten, variiert die Stärke des Bogens, wiederholt das Ausgangsmotiv und kehrt schließlich die Form des Bogens um. Hierbei drängt sich die Frage auf, ob die Wiederholung und Variation des Anschlei-fens sogar eine motivische Bedeutung hat.

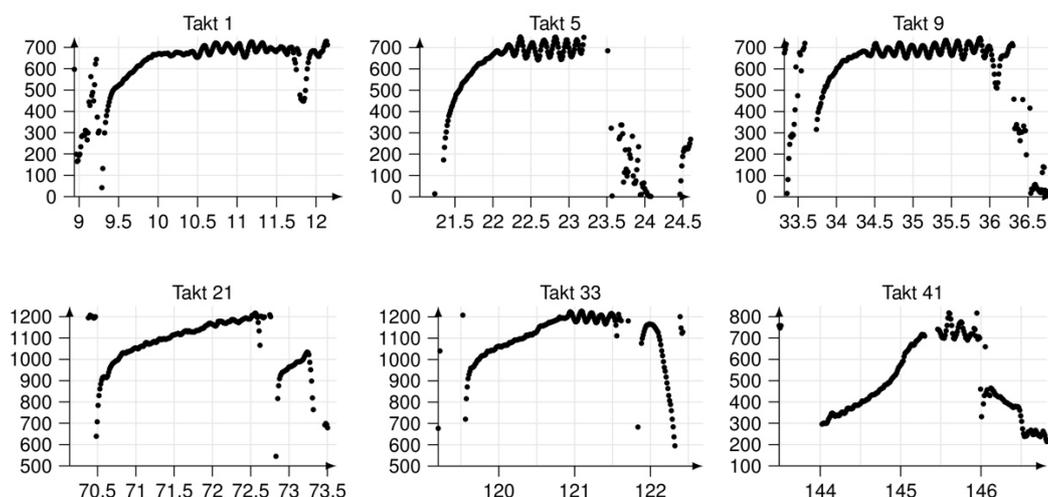


Abbildung 2: Mahalia Jacksons Schleifer in »There's Not A Friend Like Jesus« (1947). Jede Textzeile beginnt Jackson mit einem markanten Anschleifen. Dabei ist ihr Glissando anfangs bogenförmig (obere Zeile); ein bogenförmiges Anschleifen ist auch in vielen anderen Titeln zu finden und generell typisch für Gospel- und Bluesinterpretationen. Im Verlauf des Stücks variiert Jackson das Glissando, indem sie es linear singt (Takte 21 und 33) oder am Ende (Takt 41) den bogenförmigen Verlauf umkehrt.

Durch Tonhöhenkurven lassen sich auch sehr präzise die Verläufe von Bendings bestimmen. Mit Bending ist hier das Weg- und wieder Zurückziehen eines Tones von seiner Ausgangstonhöhe gemeint. Diese Technik erinnert an das Ziehen und wieder Loslassen von Gitarrensaiten. Alle Arten des Anschleifens und Gleitens zwischen oder nach Tönen können Sänger zu melismatischen Verkettungen zusammenführen, wobei einige, wie etwa Mahalia Jackson, innerhalb der Glissandobewegung sehr exakt konkrete Tonstufen ansteuern (siehe Abb. 3). In Abbildung 3 steht oben der Tonhöhenverlauf in herkömmlicher Notation, darunter der Text und Jacksons tatsächliche Aussprache laut internationalem phonetischen Alphabet (IPA 2005).

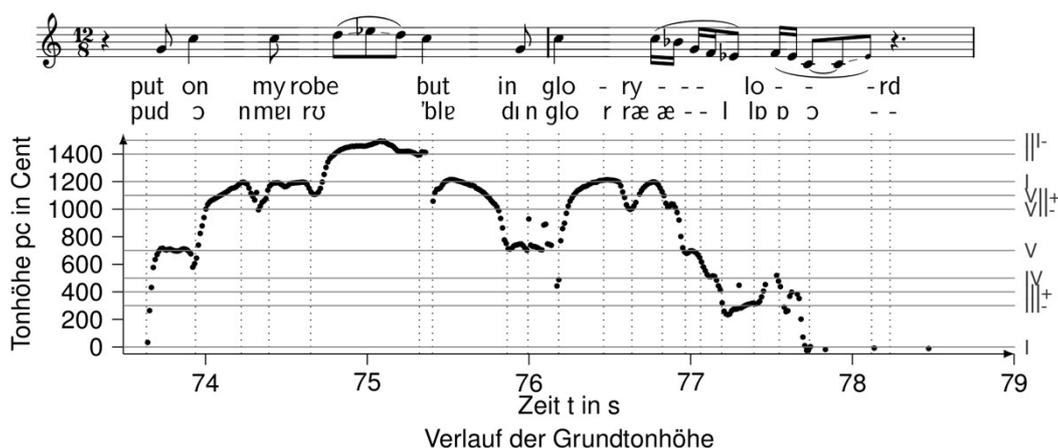


Abbildung 3: Glissandi und Bendings in »Move On Up A Little Higher« (1947) von Mahalia Jackson. Auffallend ist das Bending bei Sekunde 75 und die nachfolgende Verkettung mehrerer Glissandi zu einem Melisma (Sekunde 76.5-77.5).

## Vibratoanalyse

Abbildung 4 zeigt zwei Vibrati als rotierende Verläufe von Tonhöhe und Tonhöhenänderung. Der Verlauf des Vibratos in der Zeit ist durch Pfeile markiert, wobei jeder Pfeil die Bewegung zwischen zwei aufeinander folgenden Messzeitpunkten repräsentiert. Die Zeit verläuft in Pfeilrichtung, die Pfeile werden zudem dunkler und dicker mit der Zeit und scheinen dadurch »auf den Betrachter zuzurollen«. Die unweigerlichen Überdeckungen im Graph sind in diesem Falle erwünscht, denn je deutlicher sich ein gleichmäßiges Oval abzeichnet, desto gleichmäßiger verläuft auch das Vibrato. Zusätzlich zur Gleichmäßigkeit erfasst diese Darstellung den Ambitus des Vibratos, also den Abstand vom oberen und unteren Umkehrpunkt der Tonhöhenschwingung. Vor allem in der vergleichenden Analyse ist diese Form der Darstellung aussagekräftig, wie der Vergleich zweier Vibrati von Elvis Presley und Mahalia Jackson zeigt (s. Abb. 4).

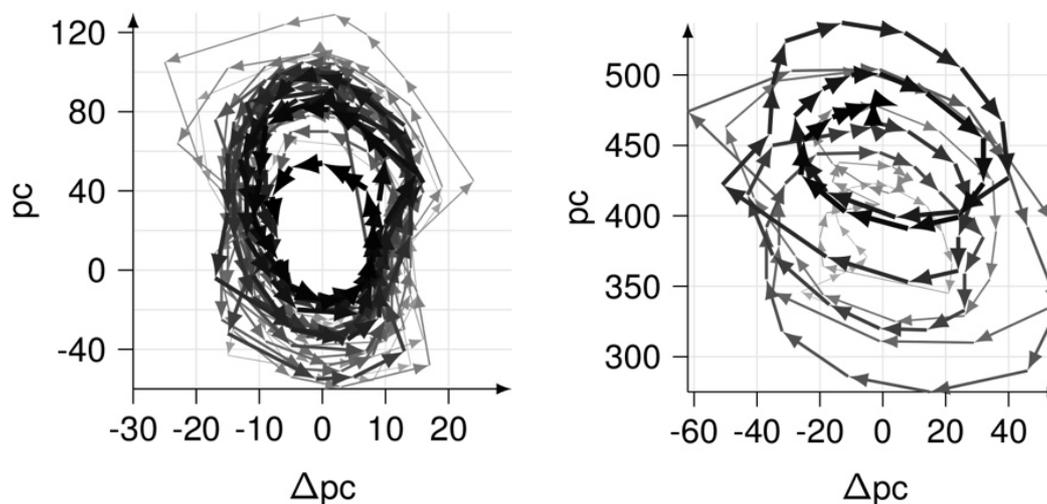


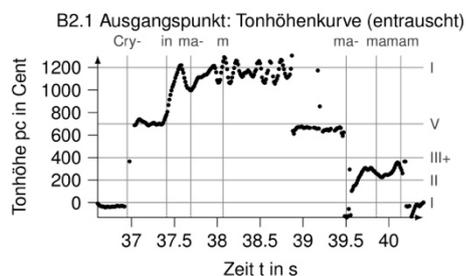
Abbildung 4: Links: Mahalia Jackson zeigt kaum Abweichungen während eines viersekündigen Vibratos in »Amazing Grace« (1947). Rechts: Elvis Presleys Vibrato in »Love Me Tender« (1956) ist dagegen ungleichmäßiger, was (zusammen mit einer behauchten Stimmgebung und geringer Stimmintensität) zum Eindruck der Intimität beiträgt.

Eine noch komprimiertere Darstellungsweise bietet sich an, wenn es gilt, viele Vibrati über große Strecken oder gar ganze Titel zu betrachten. Dies ist möglich, wenn man die Vibrati auf die zwei wesentlichen Parameter reduziert: Frequenz und Amplitude. Um diese beiden zunächst unbekannt Parameter zu ermitteln, wird abschnittsweise eine Sinusfunktion an das Signal angepasst. Das Prinzip der Anpassung ist in Box 2 skizziert.

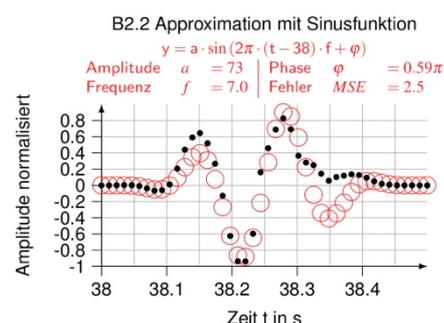
Mit dieser Methode lassen sich z.B. wichtige Hinweise auf den Personalstil der Jazzsängerin Ethel Waters finden. In »Maybe Not At All« (1925) unterbricht Waters nach der ersten Strophe den Titel und kündigt mit den Worten »Now, if Miss Clara Smith would sing the same song« an, die Bluessängerin Clara Smith zu imitieren. Anschließend gibt sie in ähnlicher Weise vor, die letzte Strophe im Stil von Bessie Smith zu singen (»I'm gettin' ready for the Empress, Miss Bessie Smith, Lord!«). Waters ändert in den verschiedenen Interpretationsansätzen die Intensität ihres Vibratos, aber nicht die Frequenz (s. Abb. 5). Das lässt vermuten, dass die Tonhöhenauslenkung im Vibrato – weil sie diese nach Belieben ändern kann – eine bewusste ästhetische Entscheidung von Waters ist. Dagegen scheint die stabile Geschwindigkeit des Vibratos ein unwillkürliches Merkmal ihres Gesangsstils zu sein.

**Box 2. Automatische Vibrato-Extraktion**

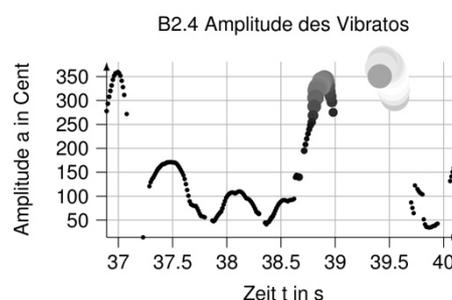
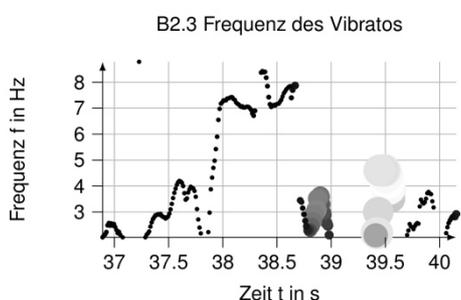
B2.1 zeigt den entrauschten Ausschnitt aus dem »Canned Heat Blues« von Tommy Johnson (siehe Box 1). Um das Vibrato automatisch zu bestimmen, wird für jeden Datenpunkt  $t$  in einer Umgebung von 500ms jene Sinusfunktion gesucht, die dem Tonhöhenverlauf am nächsten kommt.



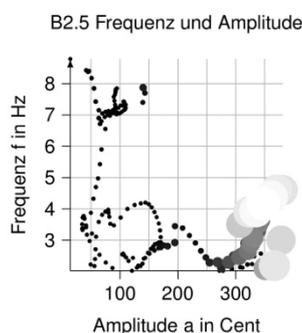
B2.2 zeigt eine entsprechende Annäherung einer Sinusfunktion (rote Kreise) an die Daten (schwarze Punkte) für den Zeitraum 38,0-38,5 Sekunden. Das Signal und die anzupassende Sinusfunktion werden an den Rändern ausgeblendet (mit einem Hanning-Fenster gefaltet). Dadurch wirken sich die Ränder weniger auf die Annäherung aus als der gesuchte Bereich in der Mitte des Fensters. Der Algorithmus gibt für jedes Fenster die bestmögliche Sinusfunktion aus, egal wie gut oder schlecht sie passt – also selbst dort, wo kein Vibrato vorliegt.



Deshalb wird zum Erkennen von Vibrati die Güte der Anpassung ermittelt, und zwar als durchschnittlicher quadrierter Abstand der Messpunkte zur Funktion (Mean Squared Error, siehe B2.2). Zusätzlich können Bereiche für Frequenzen und Amplituden definiert werden, innerhalb derer ein Vibrato gilt (z.B. eine Frequenz von 4-9 Hz).



B2.3, B2.4 und B2.5 zeigen verschiedene Darstellungsformen für Vibrati. Die Güte der Anpassung (die »Vibratohaftigkeit«) wird über die Grauwerte und Punktgrößen angezeigt: Je besser die Daten einer Sinusfunktion entsprechen, desto kleiner und dunkler (schärfer) werden sie dargestellt. Jedes Ergebnis einer Approximation entspricht hier genau einem Punkt im Graph.



B2.5 zeigt eine Zusammenfassung aller Ergebnisse über die komplette Tonhöhenkurve in B2.1. Dabei werden Frequenz und Amplitude als Achsen definiert. Die für den Gesang typischen Vibrati liegen bei 5-8 Hz. Tommy Johnson zeigt in diesem Ausschnitt mit ca. 7 Hz und ca. 100 Cent ein eher schnelles Vibrato mit kleiner Amplitude.

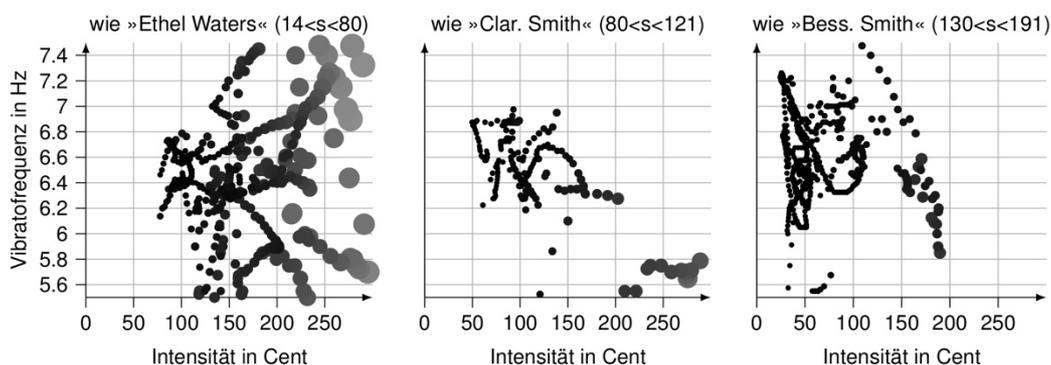


Abbildung 5: Ethel Waters Vibrati in »Maybe Not At All« (1925). Im ersten Teil zeigt Waters viele Vibrati mit starker Auslenkung ( $\pm 150$  Cent entspricht einem Vibratoumfang einer kleinen Terz). Singt Waters im Stil von Clara Smith, so sinkt die Auslenkung. Im Stil von Bessie Smith ist die Tonhöhenänderung des Vibratos stark zurückgenommen. Die mittlere Vibratofrequenz bleibt jedoch über die verschiedenen Interpretationen stabil.

## Stimmklang

Worüber die Analyse des melodischen Verlaufs noch nichts verrät, ist die vokale Klangfarbe. Spektrogramme visualisieren die Obertöne eines Klages und somit dessen Klangfarbe. Bei einem Spektrogramm wird auf der vertikalen Achse die Frequenz abgetragen, auf der horizontalen Achse die Zeit. Die im Signal enthaltene Energie pro Zeit und pro Frequenzbereich wird durch unterschiedliche Farben oder verschieden starke Grauwerte dargestellt. Ein Spektrogramm wird aus mehreren aufeinander folgenden Spektren berechnet. Die Länge des Zeitfensters für die Berechnung der einzelnen Spektren lässt sich festlegen. Von ihr hängt die Genauigkeit der zeitlichen bzw. frequenzbezogenen Auflösung der Spektraldarstellung ab. Durch eine Überlappung der Zeitfenster wird die Qualität der Darstellung weiter erhöht.

Abbildung 6 zeigt drei mögliche Spektraldarstellungen eines Vibratos, die im Sonic Visualiser mit unterschiedlichen Fensterlängen berechnet wurden. Kleine Fenster (oben) führen zu einer hohen Auflösung der Zeit, erlauben jedoch keine adäquate Frequenzauflösung. Große Fenster (Mitte) führen zwar zu einer sehr hohen Auflösung des Frequenzspektrums, jedoch können zeitliche Veränderungen der Tonhöhe, die sich innerhalb einer Fensterdauer ereignen, nicht erfasst werden. Ein Vibrato mit schnellen Tonhöhenänderungen kann daher auf diese Weise nicht visualisiert werden. Erst mit einer mittleren Fensterlänge wird der zeitliche Verlauf der Veränderung

von Grundtonhöhe und Klangspektrum eines Vibratos sichtbar (siehe Abb. 6).

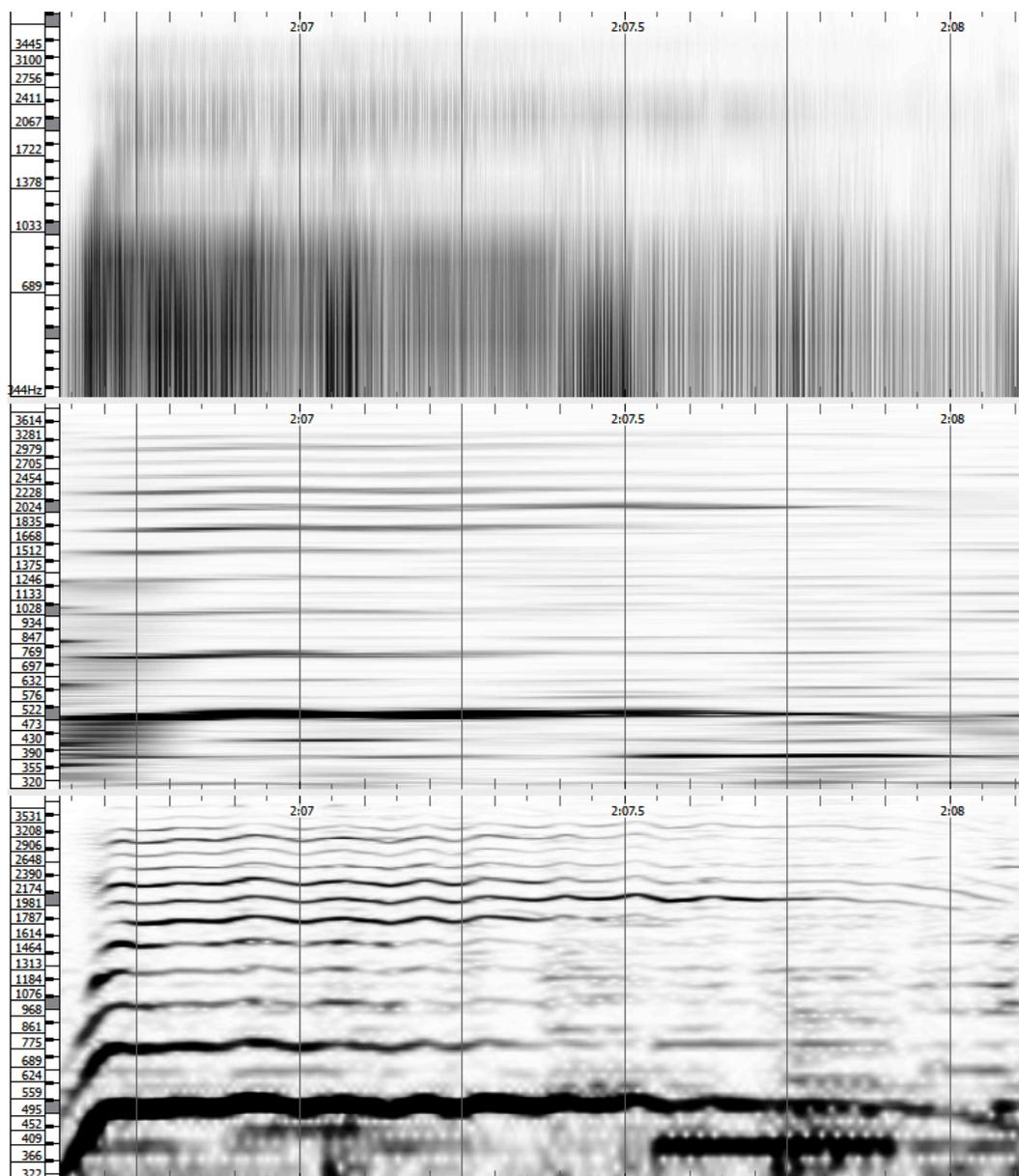


Abbildung 6: Drei Spektraldarstellungen desselben Audio-Ausschnittes, eines Gesangsvibrato von Roy Acuff: »Freight Train Blues« (1949), ca. 2:06 bis 2:08 berechnet mit unterschiedlichen Fensterlängen: oben 2,9 ms oder 128 Samples, in der Mitte 371,52 ms oder 16384 Samples, unten 46,44 ms oder 2048 Samples, jeweils mit einer Überlappung der Fenster von 93,75%.

## Vokale

Die Klangfarbe kommt u.a. durch Energiezentren im Spektrum zustande, also Frequenzbereiche mit relativ hoher Intensität. Diese Energiezentren werden als Resonanzspitzen oder Formanten bezeichnet. Ihre Lage ist abhängig von den Eigenschaften des Klangerzeugers und des Resonanzkörpers. Im Unterschied zu Resonanzkörpern von Instrumenten ist der Mundraum des Menschen flexibel formbar, sodass die Frequenzlage einiger der Formanten bewusst verändert werden kann. Wandert z.B. die Artikulation des vorderen Vokals /i/ langsam in Richtung des hinteren /a/, verändert sich die Form der Mundhöhle und somit der Resonanzraum, in dem die Formanten gebildet werden. Aus der Lage der ersten beiden Formanten ergeben sich die verschiedenen Vokale.

Im Gesang kommt der Vokalartikulation oder Vokalbetonung (vgl. Freitag 2003: 61) besondere Bedeutung sowohl in Hinblick auf die Textverständlichkeit als auch auf die klanglich-musikalische Gestaltung zu. Ein Beispiel für deutliche Vokalartikulation ist »The Prisoner's Song« von Vernon Dalhart aus dem Jahr 1924. Dalhart war ein klassisch ausgebildeter Operettensänger, der mit dem sehr erfolgreichen »The Prisoner's Song« zu einem der zentralen Vorläufer der Hillbilly Music wurde. Besonders deutlich sind die Unterschiede zwischen den Vokalen /a/ und /i/ zu sehen und zu hören.

Dalhart klingt weniger wie ein Hillbilly-Sänger, er hat eher einen klassischen Stimmklang. Wie in der Abbildung zu erkennen ist, hebt er willkürlich die Intensität der Frequenzen zwischen 2,5 und 3,3 kHz an, dem Bereich des sogenannten Sängerformanten. Der Sängerformant wird in der klassischen Gesangsausbildung trainiert und sorgt für eine große Durchsetzungskraft der Stimme gegenüber einem Orchester (Sundberg 2003: 11), senkt jedoch die Sprachverständlichkeit (Cleveland et. al 2001: 59). Vokalisten populärer Musik haben den Sängerformanten meist nicht ausgebildet. Auch bei Dalhart ist dieses Energiezentrum (in Abb. 7 mit einem Pfeil markiert) weniger deutlich ausgeprägt als der Sängerformant trainierter Opernsänger. Insofern gibt es ein klares Unterscheidungsmerkmal zwischen klassischem und populärem Stimmklang, das im Spektrogramm sichtbar gemacht werden kann.

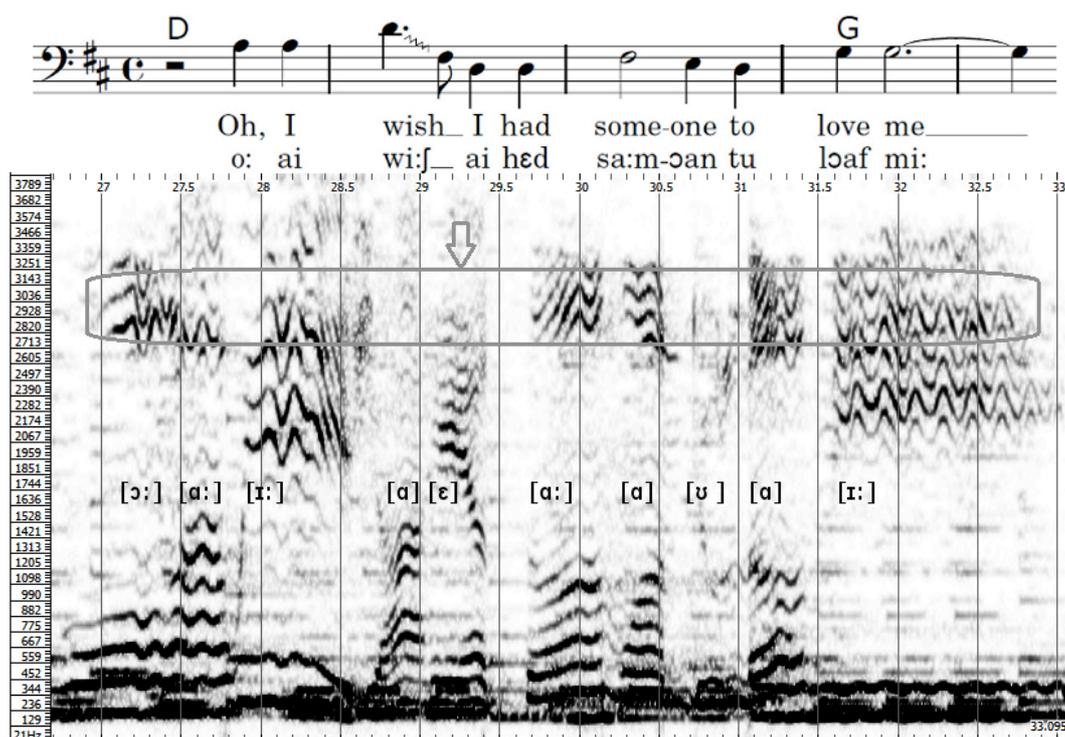


Abbildung 7: Notation, Text, Aussprache und Spektrogramm eines Ausschnitts aus Vernon Dalhart: »The Prisoner's Song« (1924).

Ein Beispiel für sprechnahen Gesang ist der Song »Can The Circle Be Unbroken« der Carter Family aus dem Jahr 1935. Sara Carter nutzt hier die Möglichkeit, Vokale klar voneinander abzugrenzen und generiert so klangliche Kontraste, die im Spektrum sichtbar werden. In Abbildung 8 sind jeweils die energiereichsten Frequenzbereiche der Vokale markiert. Optisch und klanglich ist der Kontrast zwischen /i/ und /œ/ besonders gut auszumachen. Das Auftreten solcher Vokalkontraste – und damit die Möglichkeit ihrer künstlerischen Gestaltung – hängt natürlich von der Reihenfolge der silbenkonstituierenden Vokale im Songtext ab.

Im Gegensatz zu einem bewussten Umgang mit der Vokalartikulation gibt es auch Beispiele für geringe Vokaldistinktion. Ein Beispiel ist der Song »Big Road Blues« des Bluesängers Tommy Johnson aus dem Jahr 1928. Johnson tendiert in dieser Aufnahme zu ein und derselben Vokalartikulation. Dadurch sind die Vokale weniger gut unterscheidbar und die Worte weniger deutlich zu verstehen. Es gibt nur wenige Stellen bei Johnson, die mit den deutlich unterschiedlichen Vokalartikulationen von Carter vergleichbar sind. Eine davon ist in Abbildung 9 mit einem Pfeil markiert, Johnson wechselt auf dem Wort »Lord« vom Vokal /o/ zum Vokal /u/.

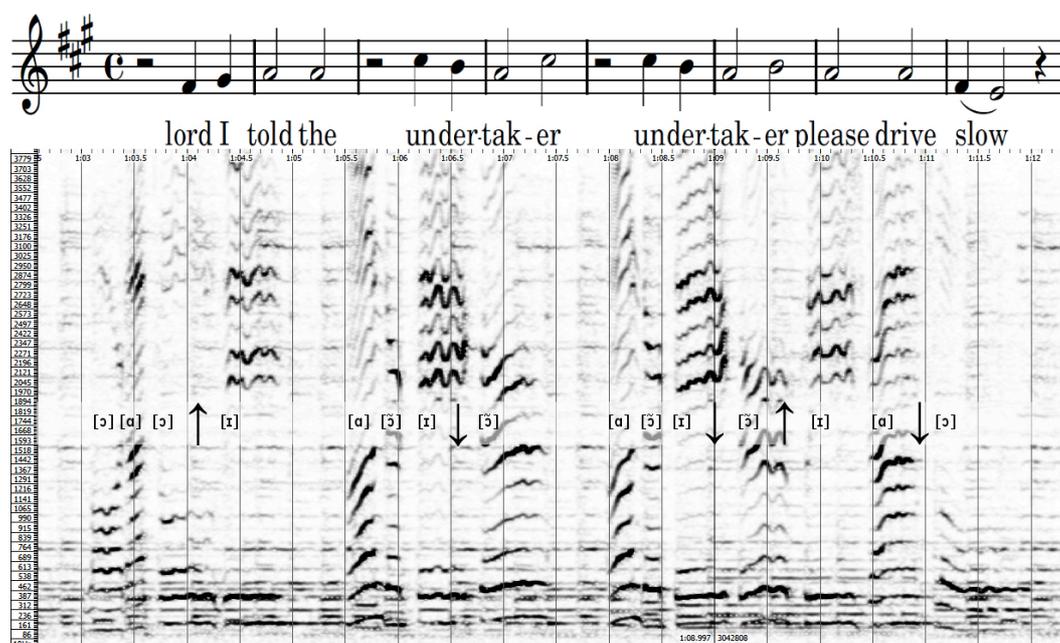


Abbildung 8: Ausschnitt aus The Carter Family: »Can The Circle Be Unbroken« (1935). Pfeile weisen auf Vokalkontraste hin.

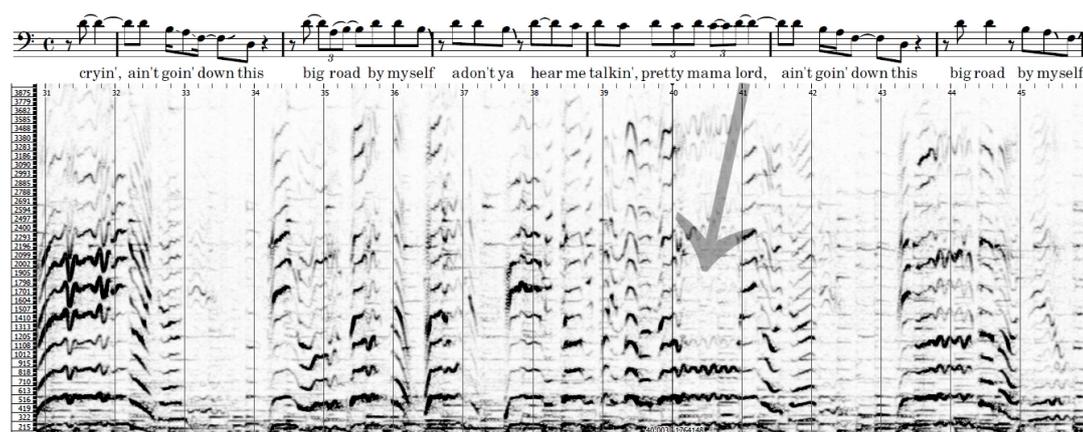


Abbildung 9: Ausschnitt aus Tommy Johnsons »Big Road Blues« (1928) mit minimaler Vokaldistinktion. Der Pfeil markiert einen der wenigen eindeutig identifizierbaren Vokalwechsel (von /o/ zu /u/ auf dem Wort »Lord«).

Artikulationsverschiebungen nennt man Settings, wenn sie zu einer überdauernden, habituellen Eigenschaft werden. Da bei Sprechern und Sängern aus den Südstaaten der USA häufig eine Vokalverschiebung in Richtung des hinteren /a/ vorliegt, wird der entsprechende Sprachklang auch mit Musikstilen wie Country Music oder Blues aus den Südstaaten assoziiert. Die Tatsache, dass auch Sara Carter aus dem Süden der USA stammt, zeigt aller-

dings, dass mit der regionalen Herkunft nicht automatisch eine bestimmte Vokalverschiebung und geringere Sprachverständlichkeit verbunden ist.

Der unterschiedliche Helligkeitsgrad verschiedener Vokale kann auch gezielt als musikalisches Gestaltungsmittel eingesetzt werden. Ein Beispiel dafür ist eine Scat-Passage aus dem Song »Guess Who's In Town« von Ethel Waters aus dem Jahr 1928. Waters erzeugt durch die Abfolge der gesungenen Vokale im Bereich der Vokalformanten eine kontinuierliche Veränderung der Klangfarbe, die als Abwärtsbewegung empfunden werden kann. Der Anfang der Passage in Abbildung 10 ist obertonreich, im Bereich von 1 bis 2,5 kHz ist wenig Energie enthalten. Das Ende der Passage ist obertonarm, der höher liegende Bereich von 2,5 bis 4 kHz enthält wenig Energie. Die formale Funktion dieser Passage als Abschluss des Scat-Solos wird von Waters somit gleich zweifach durch eine Abwärtsbewegung realisiert: sowohl über den Melodieverlauf als auch über die abnehmende Helligkeit des Vokalklangs.

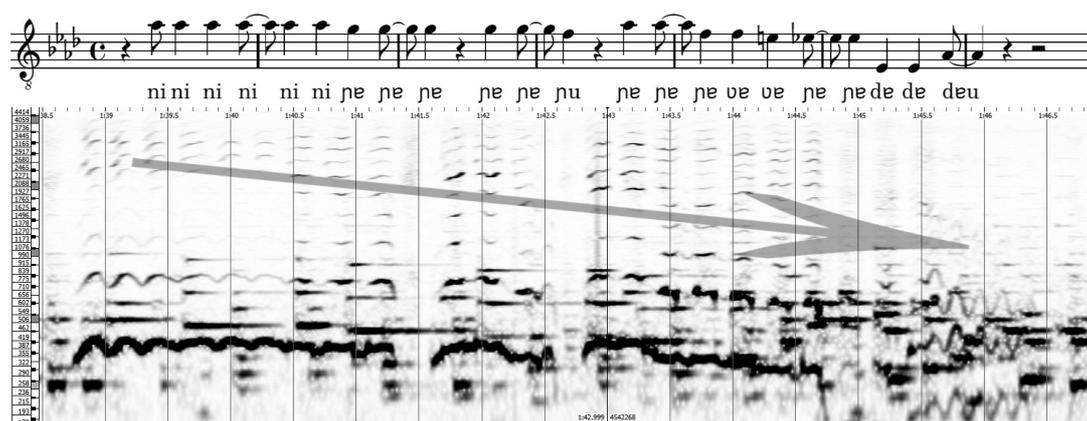


Abbildung 10: Ausschnitt aus Ethel Waters: »Guess Who's In Town« (1928).

An den Beispielen wird deutlich, dass Vokalformanten auf unterschiedliche Weise als klangliche Gestaltungsmittel eingesetzt werden können. Personalstile können Settings beinhalten, die den Stimmklang mitbestimmen. Vokalkontraste und Vokalbetonung können als gezieltes Gestaltungsmittel eingesetzt werden oder auch unwillkürlich einen Personalstil mitbestimmen.

## Rauheit

Ein weiteres Merkmal des Stimmklangs, das im Spektrogramm sichtbar gemacht werden kann, ist Rauheit. Der Jazzpianist und Sänger Fats Waller nutzt eine raue Stimmgebung häufig als Steigerungsmittel im Laufe eines

Songs. Ein Beispiel hierfür findet sich in »Dinah« (1935). Waller singt die zweite Silbe des Wortes »Carolina« betont rau. In Abbildung 11 ist das Spektrum des Wortes vergrößert dargestellt. An der rauhen Stelle wird eine Verdichtung im Spektrum sichtbar. Bei dem hier auftretenden Phänomen handelt es sich um sogenannte Subharmonics, die Folgen einer Amplitudenmodulation sind (vgl. Omori et. al. 1990: 47). Zwischen den Frequenzen der Partialschwingungen des gesungenen Tones treten zusätzliche Schwingungen auf, die als Rauheit wahrgenommen werden. Eine Ursache könnte sein, dass sich die Stimmlippen nicht bei jedem Bewegungszyklus vollständig schließen, sodass eine zweite überlagernde Schwingung entsteht. Eine weitere mögliche Ursache könnte in dem Umstand liegen, dass außer den Stimmlippen noch mehr mitschwingt, wie zum Beispiel die falschen Stimmlippen oder Schleim im Rachenraum. Ein Rauheitsempfinden entsteht, wenn der Abstand zweier Töne zueinander innerhalb der sogenannten kritischen Bandbreite liegt. Innerhalb dieses Frequenzbandes ist das Gehör nicht mehr in der Lage, nah beieinander liegende Töne als zwei voneinander getrennte Töne aufzulösen oder zu einem einzigen Ton zu verschmelzen. Die kritische Bandbreite weist unterhalb von 500 Hz eine konstante Breite von 100 Hz auf, ab 500 Hz aufwärts beträgt die Breite 20% der mittleren Frequenz des Bandes, sie steigt also linear zur Frequenz an (vgl. Fastl/Zwicker 2007: 150-173).

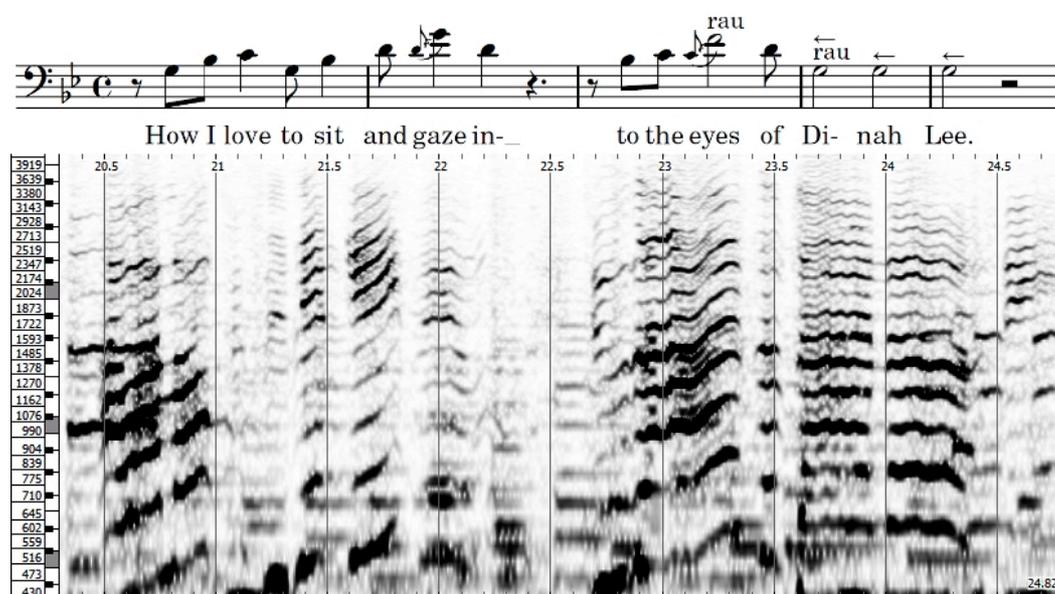


Abbildung 11: Raue Stimmgebung durch Subharmonics auf dem Wort »eyes« und der ersten Silbe von »Dinah« (ab Sekunde 23 bzw. 23.6) in Fats Waller: »Dinah« (1935), erste Strophe.

## Fazit und Ausblick

Die Visualisierungen von Audio-Daten durch Tonhöhenkurven und Spektrogramme ermöglichen einen hohen Grad an Exaktheit bei der Untersuchung und Darstellung musikalischer Sachverhalte. Mit Hilfe der hier vorgestellten Analysetools lassen sich individuelle Höreindrücke bestätigen (oder verwerfen), veranschaulichen und kommunizieren. Dies gilt nicht nur für die Analyse von Vokalaufnahmen, sondern selbstverständlich auch für Instrumentalmusik, wo es ebenfalls sinnvoll sein kann, auf Darstellungs- und Vermittlungsformen zurückzugreifen, die über die europäische Notenschrift oder eine sprachliche Charakterisierung des Klanggeschehens hinausgehen. Durch eine Messung und Visualisierung etwa von verschiedenen Typen des Tonhöhengleitens (Patton, Jackson), der Vibratogestaltung (Waters), der Vokalartikulation (Dalhart, Carter, Johnson) sowie einer gezielt rauhen Stimmgebung (Waller) wird es möglich, individuelle oder genretypische Gestaltungsmittel exakt zu beschreiben. Durch einen Vergleich der vokalen Gestaltungsmittel bei verschiedenen Aufnahmen können Personalstile und musikalische Genres eingegrenzt und voneinander abgegrenzt werden. Auf dieser Grundlage wird es möglich, stil- und kulturgeschichtliche Interpretationen der entsprechenden Musikaufnahmen und Vokalisten durch präzise Hinweise auf musikalische Gestaltungsmittel zu belegen und durch deren Visualisierung zu verdeutlichen.

## Literatur

- Auslander, Philip (2009). »Musical Persona. The Physical Performance of Popular Music.« In: *The Ashgate Research Companion to Popular Musicology*. Hg. v. Derek B. Scott. Farnham: Ashgate, S. 303-315.
- Brossier, Paul (2006). *Automatic Annotation of Musical Audio for Interactive Applications*. PhD thesis. London: Queen Mary University of London.
- Cleveland, Thomas F. / Sundberg, Johan / Stone, R. E. (2001). »Long Term-Average Spectrum Characteristics of Country Singers During Speaking and Singing.« In: *Journal of Voice* 15, S. 54-60.
- Fahey, John (1970). *Charley Patton*. London: Studio Vista.
- Fastl, Hugo / Zwicker, Eberhard (2007). *Psychoacoustics. Facts and Models*. Berlin: Springer.
- Freytag, Martina (2003). *Stimmausbildung in der Populärmusik. Ein Lehr- und Arbeitsbuch*. Berlin: Henschel-Verlag.
- Frith, Simon (1996). *Performing Rites. Evaluating Popular Music*. Oxford: Oxford University Press.

- IPA (2005). »Reproduction of The International Phonetic Alphabet (Revised to 2005).« In: *The International Phonetic Association*, <http://www.langsci.ucl.ac.uk/ipa>, Zugriff: 11.01.2013.
- Jamaludin, Mohd R. / Salleh, Sheikh H. S. / Swee, Tan T. / Ahmad, Kartini / Ibrahim, Ahmad K. A. / Ismail, Kamarulafizam (2012). »An Improved Time Domain Pitch Detection Algorithm for Pathological Voice.« In: *American Journal of Applied Sciences* 9, Nr. 1, S. 93-102.
- Middleton, Richard (2000). »Rock Singing.« In: *The Cambridge Companion to Singing*. Hg. v. John Potter. New York: Cambridge University Press, S. 28-41.
- Moore, Allan F. (2012). *Song Means. Analysing and Interpreting Recorded Popular Song*. Farnham: Ashgate.
- Omori, Kiochi / Kojima, Hisajoshi / Kakani, Rajesh / Slavid, David H. / Blaugrund, Stanley M. (1997). »Acoustic Characteristics of Rough Voice: Subharmonics.« In: *Journal of Voice* 11, Nr. 1, S. 40-47.
- Pfleiderer, Martin (2009). »Stimmen populärer Musik. Vokale Gestaltungsmittel und Aspekte der Rezeption.« In: *Musical Acoustics, Neurocognition, and Psychology of Music / Musikalische Akustik, Neurokognition und Musikpsychologie*. Hg. v. Rolf Bader (= Hamburger Jahrbuch für Musikwissenschaft 25). Frankfurt: Peter Lang, S. 233-274.
- Pfleiderer, Martin (im Druck). »Popstimmen. Theoretische und methodologische Überlegungen zur vokalen Gestaltung populärer Musik.« In: *Singstimmen. Ästhetik – Geschlecht – Vokalprofil*. Hg. v. Stephan Mösch, Anno Mungen und Saskia Woyke (= Thurnauer Schriften zum Musiktheater 24). Würzburg: Königshausen & Neumann.
- Rossignol, Stéphane / Desain, Peter / Honing, Henkjan (2001). »State-of-the-art in fundamental frequency tracking.« In: *Proceedings of Workshop on Current Research Directions in Computer Music*, Barcelona: UPF, S. 244-254.
- Senn, Olivier (2007). *Die Analyse von Tonaufnahmen. Konzepte und Methoden zur musikwissenschaftlichen Analyse von Tonaufnahmen – dargestellt an Sarah Vaughans Einspielung des Musicalhits ›My Favorite Things‹ von 1961*. Zürich: Studentendruckerei.
- Sundberg, Johan (2003). »Research on the singing voice in retrospect.« In: *Quarterly Progress and Status Report TMH-QPSR* 45, Nr. 1, S. 11-14, [http://www.speech.kth.se/prod/publications/files/qpsr/2003/2003\\_45\\_1\\_011-022.pdf](http://www.speech.kth.se/prod/publications/files/qpsr/2003/2003_45_1_011-022.pdf), Zugriff: 11.01.2013.
- Winkler Peter (1997). »Writing Ghost Notes. The Poetics and Politics of Transcription.« In: *Keeping Score. Music, Disciplinarity, Culture*. Hg. v. David Schwarz, Anahid Kassabian und Lawrence Siegel. Charlottesville/London: University Press of Virginia, S. 169-203.

## Diskographie

- Carter Family (1935). »Can The Circle Be Unbroken.« Auf: *The Original Carter Family. Country Music's First Family*. Sony Music Entertainment/Columbia 37660.
- Charley Patton (1929). »Down The Dirt Road Blues.« Auf: *Charley Patton, Complete Recorded Works, Vol. 1 (1929)*. Document Records DOCD 5009.
- Charley Patton (1934). »34 Blues.« Auf: *Charley Patton, Complete Recorded Works, Vol. 3 (1929-1934)*. Document Records DOCD 5011.

- Elvis Presley (1956). »Love Me Tender.« Auf: *The King of Rock 'N' Roll - The Complete 50's Masters* [Box-Set, Original Recording Remastered]. Sony LC00316.
- Ethel Waters (1928). »Guess Who's In Town.« Auf: *The Chronological Ethel Waters (1926-1929)*. Classics 688.
- Ethel Waters (1925). »Maybe Not At All.« Auf: *The Chronological Ethel Waters (1925-1926)*. Classics 672.
- Fats Waller (1935). »Dinah.« Auf: *Fats Waller and his Rhythms »Ain't Misbehavin'« 1934-1943*. Giants of Jazz 53078 AAD.
- Mahalia Jackson (1947). »Amazing Grace.« Auf: *How I Got It Over. The Apollo Sessions 1946-1947*. Westside (Edel) WESX 303.
- Mahalia Jackson (1947). »Move On Up A Little Higher (Part 1).« Auf: *How I Got It Over. The Apollo Sessions 1946-1947*. Westside (Edel) WESX 303.
- Mahalia Jackson (1947). »There's Not A Friend Like Jesus.« Auf: *Complete Mahalia Jackson Vol. 1 (1937-1946)*. Frémaux & Associés FA 1311.
- Roy Acuff (1947). »Freight Train Blues.« Auf: *The Essential Roy Acuff 1936-1949*. Columbia 48956.
- Tommy Johnson (1928). »Big Road Blues.« Auf: *Tommy Johnson 1928-1929*. Document Records DOCD 5001.
- Tommy Johnson (1928). »Canned Heat.« Auf: *Tommy Johnson 1928-1929*. Document Records DOCD 5001.
- Vernon Dalhart (1924). »The Prisoner's Song.« Auf: *The History of Pop Radio: 1920-1951 [OSA/Radio History]*, Disc 1. The International Music Company AG 205515.

## Abstract

Although voice and singing play a crucial role in many genres of popular music, to date there are only few approaches to an in-depth exploration of vocal expression. The paper aims at presenting new ways for describing, analysing and visualizing several aspects of singing using computer-based tools. After outlining a theoretical framework for the study of voice and singing in popular music, some of those tools are introduced and exemplified by vocal recordings from various genres (blues, gospel music, country music, jazz). Firstly, pitch gliding (slurs, slides, bends, melismas) and vibrato are discussed referring to a computer-based visualization of pitch contour. Secondly, vocal timbre and phonation (e.g. vocal roughness) are explored and visualized using spectrograms.