

# **Vocalmetrics. Ein Software-Tool für das Visualisieren, das Rating und den Vergleich von Musik-Datensätzen**

Axel Berndt, Felix Schönfeld, Rainer Groh, Tilo Hähnel,  
Tobias Marx & Martin Pfeleiderer

## **1 Einleitung**

In den letzten beiden Jahrzehnten sind eine Vielzahl von Anwendungen zur automatischen Analyse und Annotation von Musikdaten entwickelt worden (Casey et al., 2008). Allerdings sind die computergestützten Methoden zur Harmonie- und Melodieerkennung, zum Erkennen von Toneinsatzpunkten oder des Grundschlags und Metrums stark fehlerbehaftet und nur in Sonderfällen (Aufnahmen mit nur einem Instrument, Aufnahmen von strukturell einfacher Musik mit deutlich artikulierten Tönen ohne Tempo- und Taktänderungen etc.) annähernd korrekt. Gerade wenn es um die Beschreibung von musikalischen Details geht, etwa von Merkmalen des Interpretationsstils bestimmter Musiker oder Sänger, bleibt die Einschätzung von Expertenhörern nach wie vor unabdingbar. Allerdings geht es bei vielen Fragestellungen der Musikpsychologie, -soziologie oder -pädagogik nicht um objektiv bestimmbare Merkmale der Musik, sondern die subjektiv geprägte Wahrnehmung und Wirkung von Musik ist Forschungsgegenstand.

Im Folgenden wird mit dem Software-Tool *Vocalmetrics*<sup>1</sup> eine Möglichkeit vorgestellt, Musikdaten auf intuitive Weise hinsichtlich mehrerer Eigenschaften einzuschätzen, mit Hintergrundinformationen und Metadaten zu verknüpfen und hinsichtlich verschiedener Merkmalsdimensionen flexibel zu visualisieren. Insbesondere bei dem Prozess des Ratings von Musikmerkmalen beschreitet *Vocalmetrics* neue Wege. Außerdem bietet das Tool neue, flexible und intuitive Möglichkeiten, große Musikdatensätze adäquat zu visualisieren, um sie explorativ erkunden zu können (vgl. Berndt et al., 2014).

*Vocalmetrics* ist das Ergebnis einer Zusammenarbeit des DFG-Forschungsprojekts „Stimme und Gesang in der populären Musik der USA (1900–1960)“ am Institut für Musikwissenschaft Weimar-Jena mit dem Lehrstuhl für Mediengestaltung der Technischen Universität Dresden; ein Anwendungsbeispiel aus diesem Forschungskontext ist seit Sommer 2013 unter [www.hfm-weimar.de/popvoices/vocalmetrics](http://www.hfm-weimar.de/popvoices/vocalmetrics) online. Die Software ist so konzipiert, dass sie auf alle Arten von audiobasierten Musikdaten übertragen und sowohl in der musikpsy-

---

1 Download von *Vocalmetrics* v1.1 unter [www.hfm-weimar.de/popvoices/vocalmetrics](http://www.hfm-weimar.de/popvoices/vocalmetrics).

chologischen Forschung als auch in musikpädagogischen Kontexten angewendet werden kann.

## 2 Hintergründe zur Konzeptentwicklung

In das Konzept von *Vocalmetrics* flossen zahlreiche Fragestellungen ein, deren Beantwortung das Programm im Rahmen des DFG-Forschungsprojektes unterstützen sollte: Wie entwickelt sich die Stimme und der Gesangsstil einer Sängerin oder eines Sängers über die Zeit? Welche allgemeinen stilistischen Wandlungsprozesse lassen sich beobachten? Gibt es typische vokale Merkmalsausprägungen für einzelne Genres? Welche Vorlieben haben Plattenlabels bei der Wahl ihrer Künstler? Für diese und weitere Fragestellungen wurden überblicksartige Diagrammansichten entworfen, die einen Datensatz von über 200 Ausschnitten aus Gesangsaufnahmen verschiedener Vokalistinnen und populären Musikgenres visualisieren. Doch auch detailliertere Vergleiche einzelner Musikobjekte sollen möglich sein, was Visualisierungsmethoden für multivariate oder mehrdimensionale Daten, z. B. Star Plots oder parallele Koordinaten, nahelegt.

Eine zentrale Problemstellung bezieht sich auf den Rating-Prozess. Ein neu einzupflgendes Musikstück muss hinsichtlich mehrerer Merkmale beurteilt und in Relation zu den bereits eingepflegten Musikstücken gesetzt werden. Das erfordert die Kenntnis vergleichbarer Audio-Objekte, die ab einer gewissen Größe des Datensatzes und bei gemeinsamer Arbeit mehrerer Personen nur schwer zu realisieren ist. Im *Cantometrics*-Projekt des Musikethnologen Alan Lomax (1962, 1968), in dem Gesangsaufnahmen aus der ganzen Welt hinsichtlich ihrer Gestaltungsmerkmale auf 37 Skalen eingeschätzt werden sollten (vgl. Lomax, 1976), wurde daher ein umfangreicher Satz von Audio-Kassetten publiziert, auf denen jede Skalenausprägung anhand eines prototypischen Musikbeispiels veranschaulicht wird. Dieser Ansatz wird im Falle von *Vocalmetrics* dahingehend weiterentwickelt, dass bereits beurteilte Musikstücke als ein Referenzrahmen zum vergleichenden Gehören zur Verfügung gestellt werden. Mehr noch, die Auswahl der Referenzbeispiele soll während der Beurteilung kontinuierlich aktualisiert werden, sodass sich dieser Referenzrahmen immer enger um das neu zu beurteilende Musikstück zieht und so eine möglichst präzise und einheitliche Beurteilung aller Musikstücke unterstützt. Auf diese Weise wird eine Verschiebung zu einer relationalen Beurteilung erreicht, d. h. die Position des Musikstücks im Merkmalsraum wird nicht mehr durch absolute Koordinaten, die numerischen Skalenwerte, angezeigt (obwohl auch diese Möglichkeit bestehen bleibt), sondern ergibt sich aus der Relation zu bestehenden Musikdaten. Darüber hinaus sollen die Ratings unterschiedlicher Personen zu denselben Audio-Objekten unabhängig voneinander abrufbar sein, um sie miteinander vergleichen zu können.

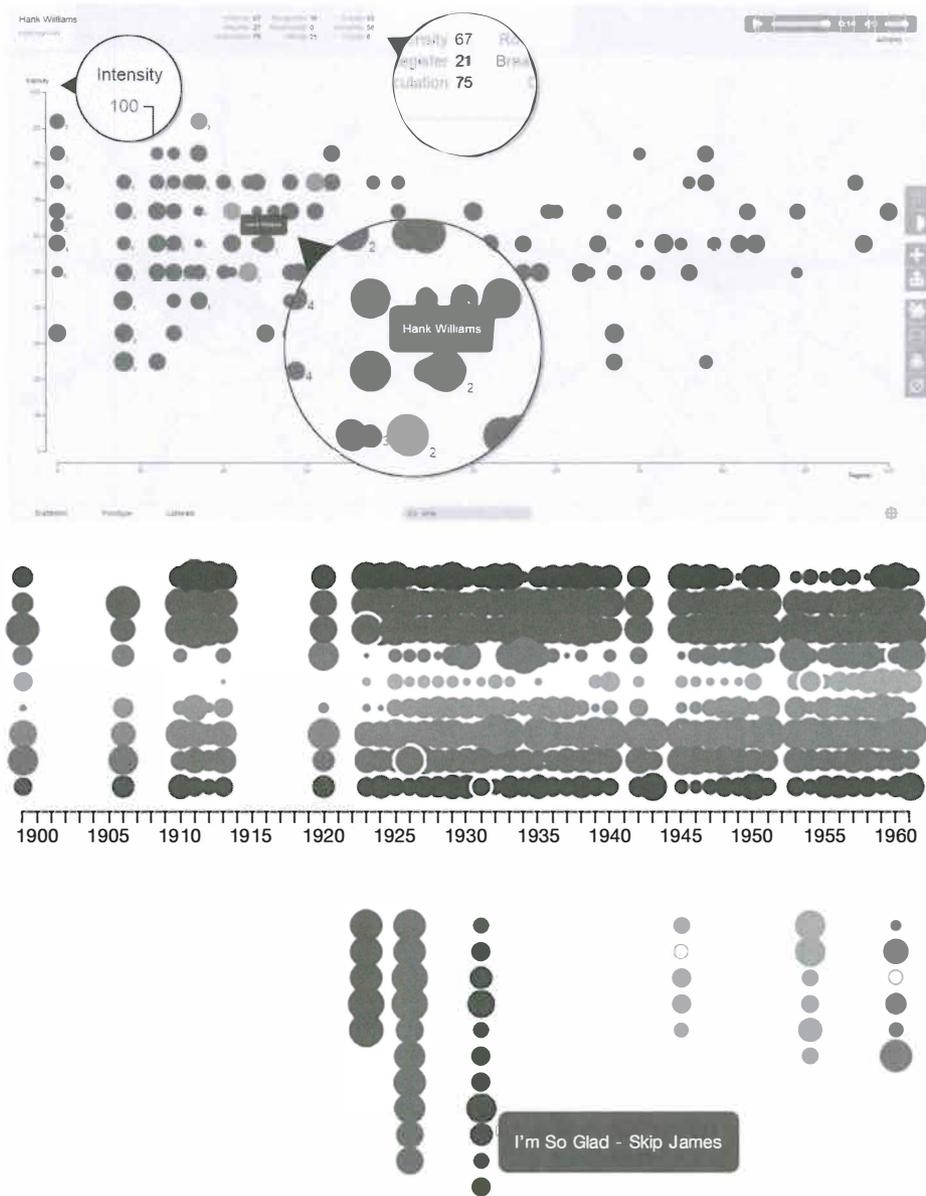
**Tab. 1:**

Datenmodell eines Datenobjektes (einer einzelnen Aufnahme im Anwendungsfall des „Stimme und Gesang“-Projekts) im Datensatz

	Feature	Datentyp	Beispiel
Metadaten	Künstler	String	Little Richard
	Titel	String	rip it up
	Geschlecht	[w,m]	m
	Plattenlabel	String	Specialty
	Genre	String	Rock'n'Roll
Dateien	audio sample	Audiodatei	LR3.ogg
	spectrogram	Datei	LR3.png
	pdf	Datei	LR3.pdf
Rating-Dimensionen	vibrato	[0% ... 100%]	0
	glissando	[0% ... 100%]	32
	intensity	[0% ... 100%]	67
	roughness	[0% ... 100%]	95
	breathiness	[0% ... 100%]	25
	register	[0% ... 100%]	31
	articulation	[0% ... 100%]	67
	rubato	[0% ... 100%]	8
	off-beat	[0% ... 100%]	58

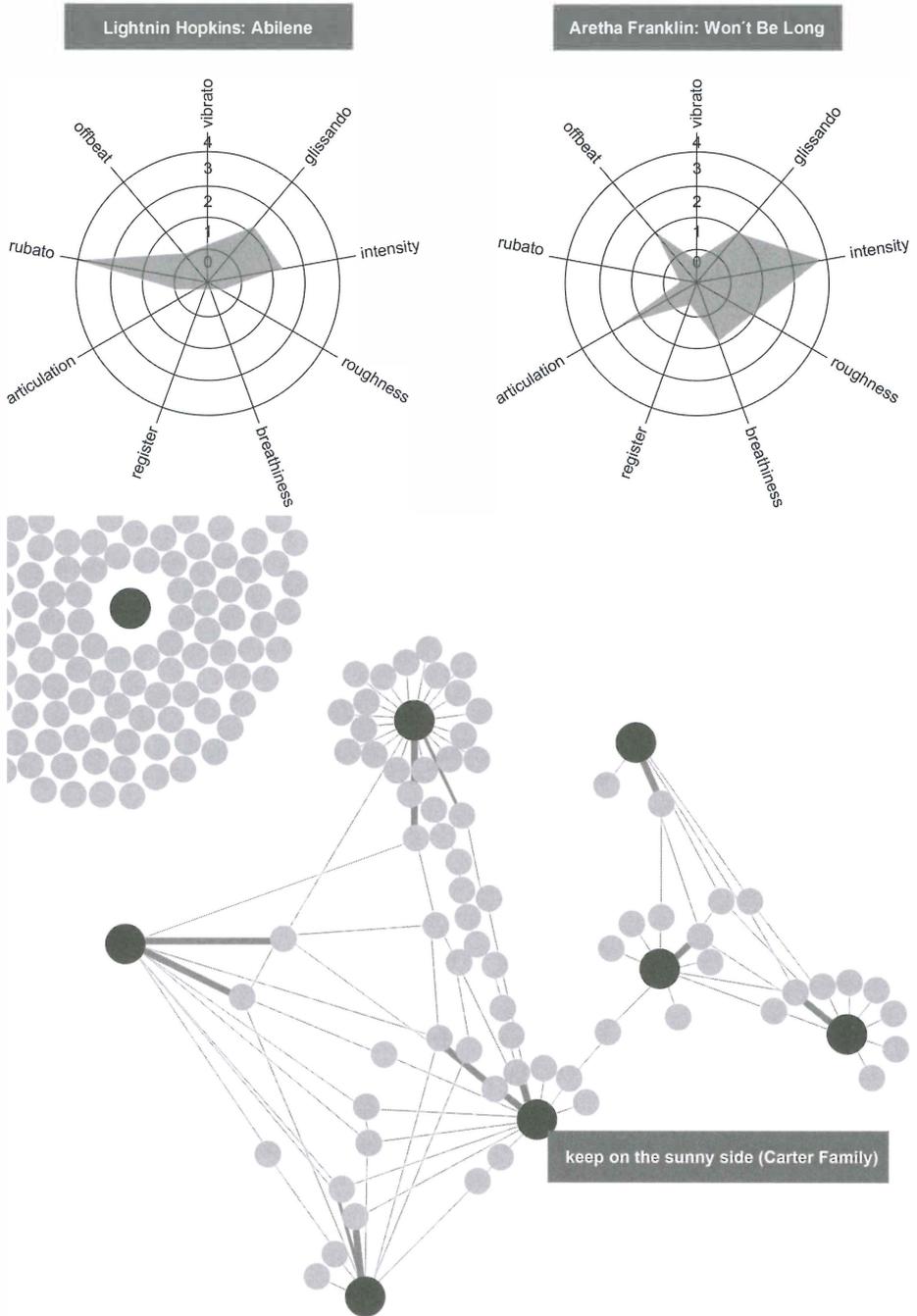
### 3 Die Ansichten in *Vocalmetrics*

Die Software *Vocalmetrics* visualisiert beliebige Musikdatensätze. Das Datenmodell eines Eintrages, wie er in *Vocalmetrics* verwendet wird, ist exemplarisch in Tabelle 1 verdeutlicht. Jeder Audiodatei sind hier fünf Metadaten und neun Rating-Dimensionen einer Gesangsaufnahme zugeordnet, die auf der Grundlage von musikalischen Analysen und phoniatischen Differenzierungen (Bergan & Titze, 2001; Dejonckere et al., 2001; Omori et al., 1997) von Experten eingeschätzt wurden. *Vocalmetrics* visualisiert jedes Musikbeispiel als einen Kreis, der nach dem Anklicken alle Metadaten (Genre, Entstehungsjahr etc.) und Eigenschaften (Vibrato, Rauheit etc.) darstellt; das Musikbeispiel selbst lässt sich über einen integrierten Audio-Player abspielen. Zudem können zusätzliche Daten wie Bilder (z. B. Spektrogrammansichten oder Transkriptionen) oder PDFs (z. B. Hintergrundinformationen zum Klangbeispiel) durch Mausclick angezeigt werden. Über die Anordnung aller Kreise auf dem Bildschirm ist es möglich, Zusammenhänge und Verteilungen über mehrere Merkmalsdimensionen hinweg



**Abb. 1:**  
Beispiele von Ausschnitten aus der Overview-Ansicht (oben) und  
der Timeline-Ansicht (unten)

sichtbar zu machen. Hierzu sind in *Vocalmetrics* vier verschiedene Ansichten des Datensatzes implementiert: eine *Overview*-, eine *Timeline*-, eine *Comparison*- und eine *Prototype*-Ansicht.



**Abb. 2:**  
Beispiele von Ausschnitten aus der Comparison-Ansicht (oben) und  
der Prototype-Ansicht (unten)

Die *Overview-Ansicht* bietet die Möglichkeit, alle Datenobjekte als Kreise in einem Streudiagramm anzuzeigen. Über die horizontale und vertikale Position, die Kreisgröße und Kreistransparenz können vier Rating-Dimensionen gleichzeitig dargestellt werden. Die Belegung der Dimensionen ist frei wählbar. Zudem ist es möglich, Mittelwerte der Metadaten als Kreise anzeigen zu lassen. Abbildung 1 (oben) zeigt einen Ausschnitt der Darstellung, bei der die Mittelwerte der Genres eingblendet wurden.

Die *Timeline-Ansicht* (siehe Abb. 1, unten) zeigt die Mittelwerte aller Rating-Dimensionen und aller Objekte im Datensatz für jedes Jahr über den Zeitraum des Datensatzes. Hiermit können zeitliche Entwicklungen sichtbar gemacht werden.

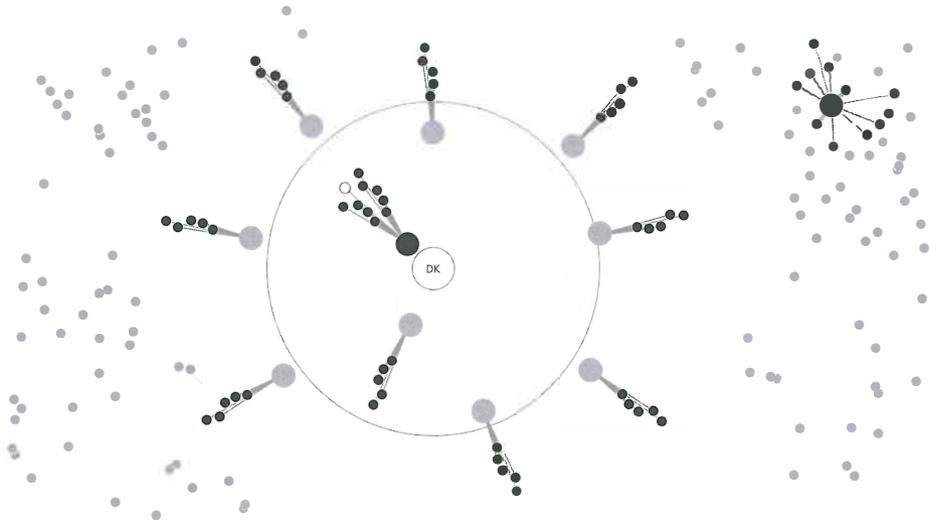
Die *Comparison-Ansicht* zeigt alle Ausprägungen der Rating-Dimensionen jedes einzelnen angewählten Datenobjektes als Star Plot (siehe Abb. 2, oben). Durch die Gegenüberstellung mehrerer Star Plots können Ähnlichkeiten und Unterschiede der ausgewählten Beispiele schnell erfasst werden.

Die *Prototype-Ansicht* basiert auf der Prototypensemantik, die Rosch (1983) beschreibt. Alle Datenobjekte werden als Kreise in der Bildfläche dargestellt. Sie können ebenso angeklickt werden, wie in den übrigen Ansichten, um alle Informationen anzuzeigen. Zusätzlich können einzelne Datenobjekte als Prototypen deklariert werden, die dann ähnliche Datenobjekte im Datensatz anziehen. Hier kommt eine Magnet-Metapher zum Einsatz. Wird ein Datenobjekt von mehreren Prototypen angezogen, wird es entweder dem stärksten und folglich ähnlichsten zugeordnet oder nimmt eine relative Position zwischen ihnen ein, welche die Ähnlichkeitsrelation widerspiegelt. Die Ähnlichkeit basiert auf der euklidischen Distanz der Datenobjekte im Merkmalsraum, wofür alle Rating-Dimensionen herangezogen werden. Indem der Datensatz auf diese Weise sehr flexibel und intuitiv strukturiert werden kann, wird eine gezielte Exploration möglich.

## 4 Das Rating mit *Vocalmetrics*

Das Einpflegen neuer Datenobjekte geschieht durch die Eingabe der Metadaten, die Verlinkung mit der Audiodatei und zusätzlichen Dateien sowie durch das Rating der Audiodatei hinsichtlich der für ein Projekt definierten Rating-Dimensionen (vgl. Tab. 1).

Neben der herkömmlichen Eingabe von Skalenwerten bietet *Vocalmetrics* eine neue intuitive Rating-Möglichkeit, die auf der Prototypensemantik der Prototype-Ansicht aufbaut und sich sowohl einer Magnet- als auch einer Eizellen-Metapher bedient (siehe Abb. 3). Das neu zu beurteilende Musikstück wird zunächst in der Bildschirmmitte als Zellkern platziert. Die übrigen Musikstücke im Datensatz werden um die Zellhülle herum angeordnet, ebenso die sog. Attribut-Magneten, die je ein zu beurteilendes Attribut (eine Rating-Dimension) repräsentieren. Die Attribut-Magneten sind mit einem möglichst schwachen und einem möglichst starken Vertreter versehen, die im Rating-Prozess als Referenzobjekte dienen und vergleichend angehört werden können.



**Abb. 3:**  
Ausschnitt aus einer Rating-Ansicht (Beispiel)

Auch im Datensatz bereits enthaltene Musikstücke können in dieser Ansicht wieder als Prototypen deklariert werden. Prototyp-Magneten und Attribut-Magneten können in die Zellhülle hineingezogen werden. Das zu bewertende Musikstück (im Zellkern) „erbt“ nun die Merkmalsausprägungen der Prototypen in dem Maße, wie sie in die Nähe des „Zellkerns“ gezogen werden (gewichtete Interpolation). Je näher ein Prototyp dem Zellkern kommt, desto stärker prägt seine Merkmalsausprägung diejenigen des Musikstücks im Zellkern. Hier wird die Beziehung von Ähnlichkeit und räumlicher Nähe für ein intuitives, interaktives Rating genutzt. Dabei genießen die Attribut-Magneten eine gesonderte Behandlung. Wird ein Attribut-Magnet in die Zellhülle gezogen, so dominiert er hinsichtlich seines Attributs über die Prototyp-Magneten. Das Attribut wird dann wie mit einem Slider von 0% am Zellrand bis 100% direkt am Zellkern über die Positionierung des Attribut-Magneten eingestellt. Der gewichtete Mittelwert der Prototyp-Magneten wird für dieses Attribut ignoriert.

Eine Vorgehensweise beim Rating könnte folgendermaßen aussehen: Zunächst werden ähnliche Musikstücke aus dem Datensatz gesucht und als Prototyp-Magneten in die Zelle hineingezogen. So entsteht schnell ein initiales Rating, das sodann mittels einzelner Attribut-Magneten verfeinert werden kann. Unterscheidet sich das Musikstück bspw. nur hinsichtlich eines einzigen Attributs von den Prototypen (z. B. „der gleiche Gesangsstil nur mit mehr Vibrato“), so wird dieses mittels des zugehörigen Attribut-Magneten eingestellt. Da die Prototyp-Magneten selbst Musikstücke sind und ihrerseits auf weitere ähnliche Musikstücke verweisen, sind die zum vergleichenden Ge-

gehören relevanten Musikstücke stets präsent und direkt anwähl- und anhörbar.

## 5 Zusammenfassung und Ausblick

Das Software-Tool *Vocalmetrics* vermag einen Überblick über die musikalischen Merkmale und Metaeigenschaften von großen Audio-Datensätzen zu liefern. Es ermöglicht einen differenzierten Vergleich zwischen diesen Merkmalen, die innerhalb eines Projektes frei definiert werden können. Neben musikstrukturellen oder klanglichen Eigenschaften der Audio-Objekte können so auch die individuellen emotionalen Wirkungen von den am Ratingprozess beteiligten Personen eingegeben und anschließend untersucht werden. Der Prozess des Ratings folgt einem intuitiven und interaktiven Ansatz, der auch für Schüler oder musikalische Laien gut nachvollziehbar ist. Da die Ratings unterschiedlicher Personen getrennt eingegeben, dargestellt und als Rohdaten exportiert werden können, eignet sich das Tool auch als Messinstrument in der Rezeptions- und Wirkungsforschung. Darüber hinaus ist ein eher spielerischer Umgang denkbar, etwa im Rahmen des schulischen Musikunterrichts, bei dem Schüler ihre Lieblingsstücke hinsichtlich bestimmter Rating-Dimensionen (Gefallen, Stimmung, emotionale Wirkung etc.) bewerten. Dies kann als Ausgangspunkt einer aktiven und reflektierten Auseinandersetzung mit der gehörten Musik, den eigenen Musikpräferenzen, der eigenen Musikrezeption sowie mit Musikpräferenzen und Musikrezeption der Mitschüler dienen.

## Literatur

- Bergan, C. C. & Titze, I. (2001). Perception of pitch and roughness in vocal signals with subharmonics. *Journal of Voice*, 15, 165–175. [http://doi.org/1.1016/S0892-1997\(01\)00018-2](http://doi.org/1.1016/S0892-1997(01)00018-2)
- Berndt, A., Groh, R., Hähnel, T., Pfeleiderer, M. & Schönfeld, F. (2014, October). *Vocalmetrics. An interactive software for visualization and classification of music*. Proceedings of the Audio Mostly Conference 2014, Aalborg, Denmark.
- Casey, M. A., Veltkamp, R., Goto, M., Leman, M., Rhodes, C. & Slaney, M. (2008). Content-based music information retrieval: Current directions and future challenges. *Proceedings of the IEEE*, 96(4), 668–696. <http://doi.org/1.1109/JPROC.2008.916370>
- Dejonckere, P. H., Bradley, P., Clemente, P., Cornut, G., Crevier-Buchman, L., Friedrich, G. et al. (2001). A basic protocol for functional assessment of voice pathology, especially for investigating the efficacy of (phonosurgical) treatments and evaluating new assessment techniques. *European Archives of Oto-Rhino-Laryngology*, 258(2), 77–82. <http://doi.org/1.1007/s004050000299>
- Lomax, A. (1962). Song structure and social structure. *Ethnology*, 1, 425–451. <http://doi.org/1.2307/3772850>
- Lomax, A. (1968). *Folk song style and culture*. Washington, DC: American Association for the Advancement of Science.
- Lomax, A. (1976). *Cantometrics. An approach to the anthropology of music*. Berkeley, CA: University of California Extension Media Center.

- Omori, K., Kojima, H., Kakani, R., Slavid, D.H. & Blaugrund, S. M. (1997). Acoustic characteristics of rough voice: Subharmonics. *Journal of Voice*, *11*(1), 40–47. [http://doi.org/1.1016/S0892-1997\(97\)80022-7](http://doi.org/1.1016/S0892-1997(97)80022-7)
- Rosch, E. (1983). Prototype classification and logical classification: The two systems. In E. F. Scholnick (Ed.), *New trends in conceptual representation: Challenges to Piaget's theory?* (pp. 73–86). Hillsdale, NJ: Erlbaum.