"... wenn das Gute liegt so nah"
Instrumentale Formantnähe und Klangfarbenähnlichkeit aus menschlicher und rechnerischer Perspektive

Christoph Reuter[1], Isabella Czedik-Eysenberg[1], Saleh Siddiq[1], Michael Oehler[2]

[1] *Musikwissenschaftliches Institut, Universität Wien, A-1090 Wien, E-Mail: christoph.reuter@univie.ac.at*

[2] *Musikwissenschaftliches Institut, Universität Osnabrück, D-49069 Osnabrück*

Hintergrund

In der Klangfarbenforschung ist lange bekannt, dass die Klangfarbe eines einzelnen Tons nicht vergleichbar ist mit der Instrumentalklangfarbe eines Musikinstruments an sich. Es handelt sich hier um zwei völlig verschiedene Konzepte (vergl. [1], S. 393 bis [2], S. 15). Dennoch wird bei den meisten publizierten Klangfarbenmessungen der Einzelton als prototypisch für ein ganzes Instrument behandelt, während es für die klangliche Beschreibung ganzer Musikinstrumente in allen erreichbaren Tonhöhen und Dynamikstufen nur wenige Konzepte gibt: Hier haben sich für eine umfassendere Beschreibung der Instrumentalklangfarbe neben den Mel Frequency Cepstral Coefficients (MFCCs, z.B. [3]) und dem Modulation Power Spectrum (MPS, z.B. [4]) vor allem die Formantbereiche bewährt (seit 1929 z.B. [5][6][7]). Für die Beschreibung von Instrumentalklangfarben mit Hilfe der Formantbereiche wurde auf der Grundlage von 586 Instrumentalklängen ein zweidimensionales Formantenfeld erstellt [8], in das sich besonders Doppelrohrblatt- und Blechblasinstrumente auf der Basis der ersten beiden Formanten (X-Achse: Formant 1, Y-Achse: Formant 2) nach Instrument, Dynamik und Register sehr gut voneinander getrennt darstellen lassen (z.B. Abb. 1). Die im Formantenfeld ermittelten Formantpositionen entsprechen größtenteils den in der Literatur zu findenden Beschreibungen, z.B. [1][5][6][7][9].

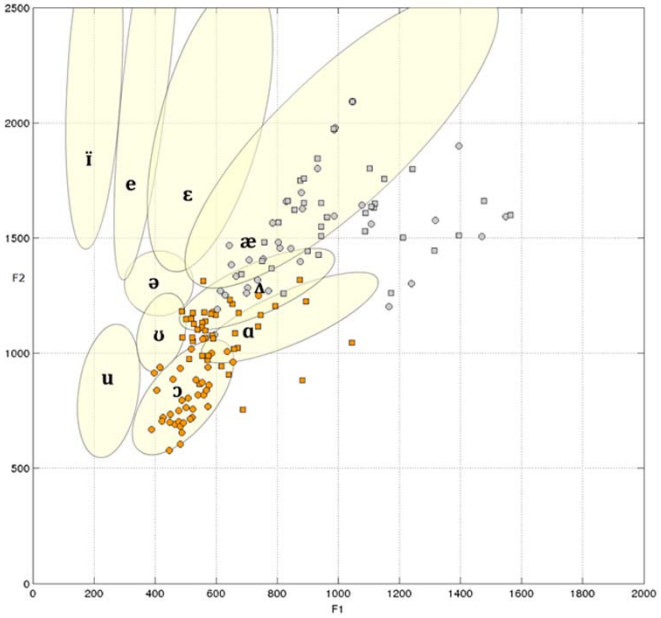

Abbildung 1: Formantenfeld mit den Klängen von Fagott (orange) und Oboe (grau) in allen erreichbaren Tonhöhen in ff und pp [8].

Fragestellung

- Werden im Formantenfeld nahe beieinanderliegende Klangfarben auch als ähnlich klingend empfunden?
- Bietet das Formantenfeld eine ausreichende Genauigkeit, um Instrumentenklänge anhand ihrer Formantpositionen rechnerisch zu unterscheiden und zu klassifizieren?
- Welche zusätzlichen Timbre Features können das Klassifikationsergebnis verbessern?

Rechnerische Beschreibung von empfundener Klangfarbenähnlichkeit: Formanten und MFCCs im Vergleich

Im Hörversuch wurden 40 lautheitsangeglichene Klangpaare mit jeweils 20 sehr nahen und 20 sehr weit voneinander entfernten Formantpositionen von 22 Versuchspersonen auf einer (Un-)ähnlichkeits-Skala von 1–8 bewertet (8 = maximale Unähnlichkeit). Die Hörerurteile wurden danach mit der euklidischen Distanz de Formantpositionen im Formantenfeld auf Korrelationen geprüft. Hierbei zeigte sich eine deutliche Korrelation der Formantdistanzen mit den (Un)-Ähnlichkeitsurteilen der Hörer ($r = 0.759$, t-Test mit $p < 0.001$, Konfidenzintervall: [-3.1381; -1.8960]).

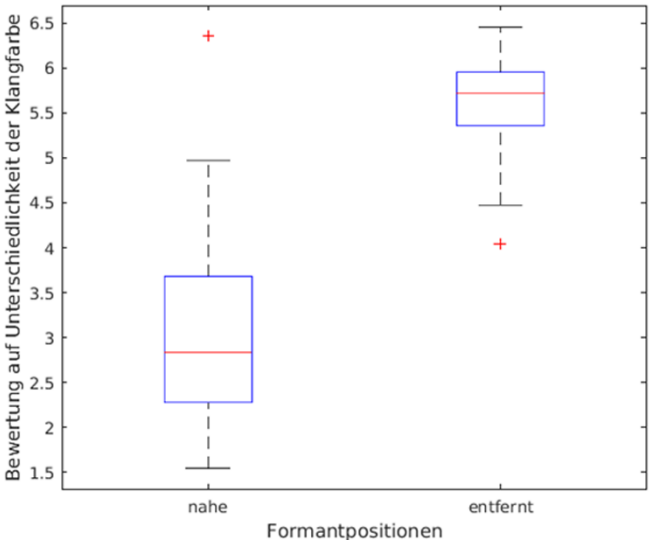

Abbildung 2: Die Distanz der Formantpositionen (X-Achse) korreliert stark mit der empfundenen Klangfarbenähnlichkeit.

Hierbei weisen die ersten beiden Formanten (F1 und F2) einen fast gleich starken linearen Zusammenhang mit den Ähnlichkeitsbewertungen auf und korrelieren auch besonders stark miteinander ($r = 0.9196$; $p < 0.0001$). Im Vergleich dazu korrelieren die ersten drei MFCCs schwächer mit den Ähnlichkeitsbewertungen der Versuchspersonen, während

die restlichen MFCCs keine Korrelationen zu den Hörerurteilen aufweisen (Formanten wurden in Praat [10] ermittelt, MFCCs mit Hilfe der MIRtoolbox in Matlab [11]).

Tabelle 1: Korrelation einzelner Formanten und MFCCs mit der empfundenen Klangfarbenähnlichkeit

Timbre feature	r	p
F1	0.7514	< 0.0001
F2	0.7477	< 0.0001
F3	0.4227	< 0.0001
MFCC1	0.6384	< 0.0001
MFCC2	0.5959	< 0.0001
MFCC3	0.3513	< 0.05

Ein ähnliches Ergebnis zeigt sich bei der Betrachtung von Formant- und MFCC-Kombinationen: In einem zwei- oder dreidimensionalen Formant- oder MFCC-Space würden die euklidischen Distanzen der Formanten 1 und 2 am stärksten mit der empfundenen Klangfarbenähnlichkeit korrelieren (r = 0.759, p < 0.0001, Abb. 3), während die euklidischen Distanzen der MFCCs 1–3 ein zwar nicht so deutliches aber dennoch vergleichbares Bild liefern (r = 0.695, p < 0.0001, Abb. 4).

Tabelle 2: Korrelation der euklidischen Distanzen von Formant- und MFCC-Kombinationen mit der empfundenen Klangfarbenähnlichkeit

Klangdeskriptoren	r	p
F1 und F2	0.7591	< 0.0001
MFCCs 1, 2 und 3	0.6949	< 0.0001
MFCCs 1 und 2	0.6939	< 0.0001
F1, F2 und F3	0.6916	< 0.0001
MFCCs 1-13	0.6812	< 0.0001

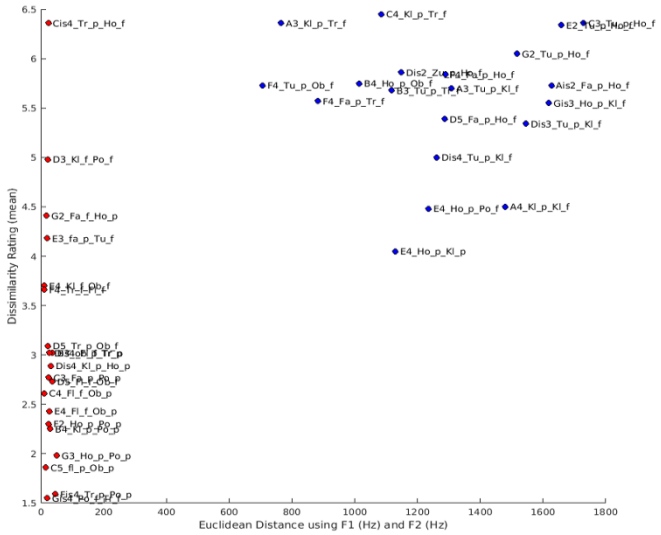

Abbildung 3: Scatterplot der euklidischen Distanzen der ersten beiden Formantpositionen (X-Achse) und der empfundenen Klangfarbenähnlichkeit (Y-Achse).

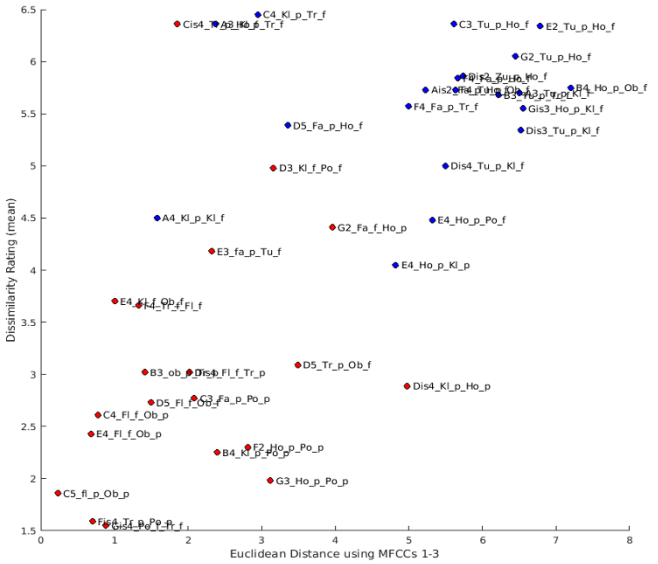

Abbildung 4: Scatterplot der euklidischen Distanzen der ersten drei MFCCs (X-Achse) und der empfundenen Klangfarben-ähnlichkeit (Y-Achse)(rot: nahe Formantpositionen, blau: entfernte Formantpositionen).

Auch in einer Spinnennetz-Darstellung mit den Achsen F1, F2, F3 bzw. MFCC1, MFCC2, MFCC3 zeigt sich in den verschiedenen Instrumentenkombinationen, dass sich die zwischen den Achsen entstehenden Dreiecke sowohl im Falle der Formanten als auch im Falle der MFCCs bei stark empfundener klanglicher Ähnlichkeit überlappen (Abb. 5) und mit ansteigend empfundener klanglicher Unähnlichkeit immer größere Abstände zueinander einnehmen (Abb. 6).

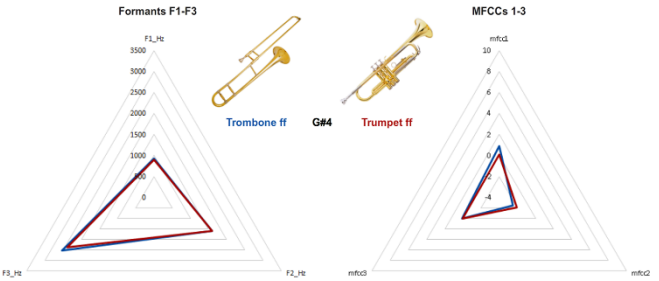

Abbildung 5: Spinnennetz-Darstellung mit den Achsen F1, F2, F3 bzw. MFCC1, MFCC2, MFCC3 des am ähnlichsten empfundenen Klangpaares (Posaune ***ff*** und Trompete ***ff*** auf gis1).

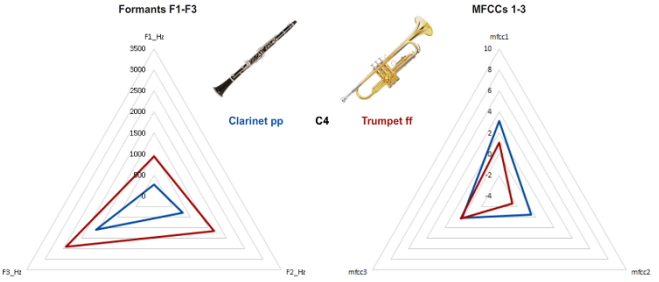

Abbildung 6: Spinnennetz-Darstellung mit den Achsen F1, F2, F3 bzw. MFCC1, MFCC2, MFCC3 des am unähnlichsten empfundenen Klangpaares (Klarinette ***pp*** und Trompete ***ff*** auf c1).

Es zeigt sich, dass Formanten und MFCCs in ähnlicher Weise für die Bestimmung von klanglicher Ähnlichkeit geeignet sind. Dies legt auch das Modell von Darch et al. 2005 nahe [12], in welchem auf der Grundlage von MFCCs auf Formantbereiche zurückgeschlossen werden kann.

Rechnerische Vorhersagbarkeit empfundener Klangfarbenähnlichkeit: Formanten und MFCCs im Vergleich

Mit Hilfe von Machine-Learning trainierten Regressions-modellen (mit fünffacher Kreuzvalidierung) lassen sich Klangfarbenähnlichkeiten bis zu einem gewissen Grad vorhersagen: Nach einem Training auf der Grundlage von Formanten zeigte sich das beste Ergebnis mit einem Bestimmtheitsmaß (R^2) von 0.53 bei der Verwendung von F1 und F2 (Abb. 7).

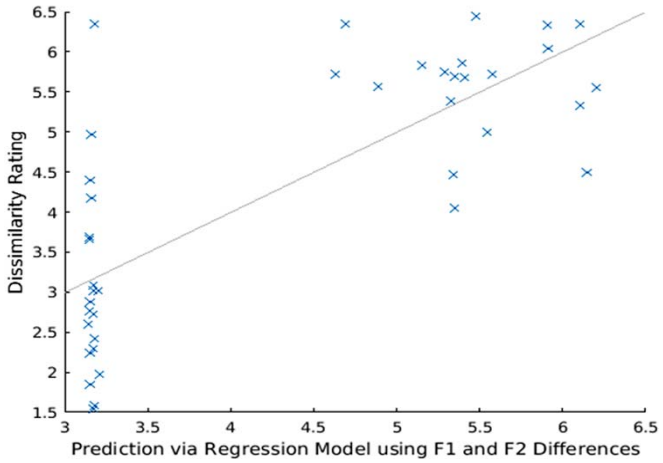

Abbildung 7: Vorhersagemodell auf der Grundlage von F1 und F2 (R^2: 0.53, RMSE: 1.08, MSE: 1.16, MAE: 0.86).

Nach einem Training auf der Grundlage von MFCCs ergab die Verwendung von MFCC1 und MFCC2 als bestes Ergebnis einen vergleichbaren Wert im Bestimmtheitsmaß (R^2) von 0.56 (Abb. 8).

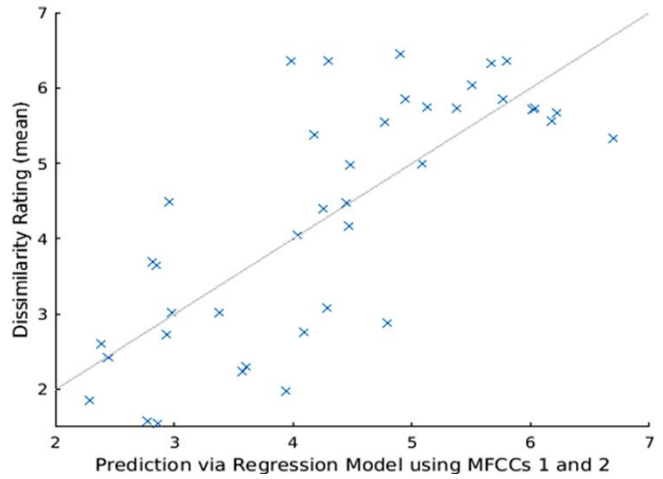

Abbildung 8: Vorhersagemodell auf der Grundlage von MFCC1 und 2 (R^2: 0.56, RMSE: 1.05, MSE: 1.10, MAE: 0.81).

Mit anderen Worten: Formanten und MFCCs sind für die Vorhersage von Klangfarbenähnlichkeiten ähnlich gut geeignet, wobei man jedoch berücksichtigen muss, dass sich für die Erstellung von Vorhersagemodellen auch andere Klangfarbenmerkmale durchaus eignen: Betrachtet man z.B. den Abstand des Spectral Centroids eines Klanges von dem eines anderen, so lässt sich auf dieser Grundlage ein Vorhersagemodell trainieren, nach dem mit wachsender Distanz zwischen den Spectral-Centroid-Werten die Klänge als immer unähnlicher empfunden werden (Bestimmtheitsmaß (R^2) von 0.78, Abb. 9).

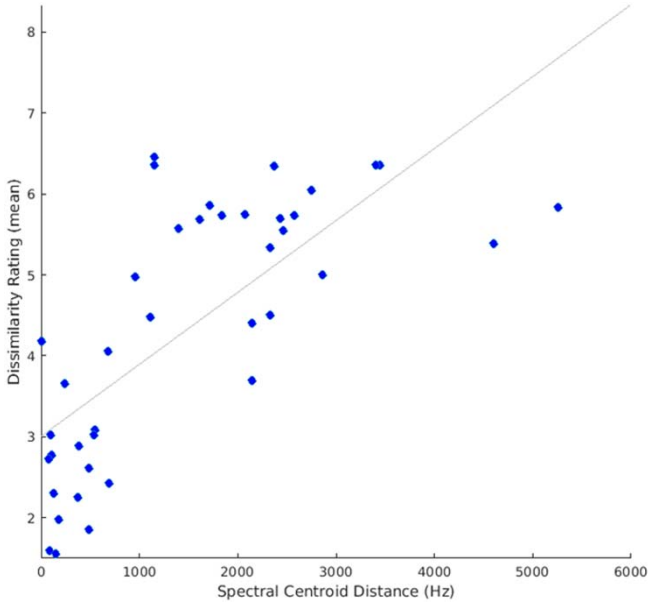

Abbildung 9: Vorhersagemodell auf der Grundlage des Spectral Centroids (R^2: 0.78, RMSE: 0.75, MSE: 0.56, MAE: 0.58).

Klangfarbenklassifikation auf der Grundlage von Formanten

Mit Hilfe von maschinellen Klassifikationsverfahren lassen sich Musikinstrumente anhand ihrer ersten drei Formanten (F1/F2/F3, kubische k-Nearest-Neighbor-Klassifikation) mit einer Präzision von 46.1% zuordnen (Abb. 10).

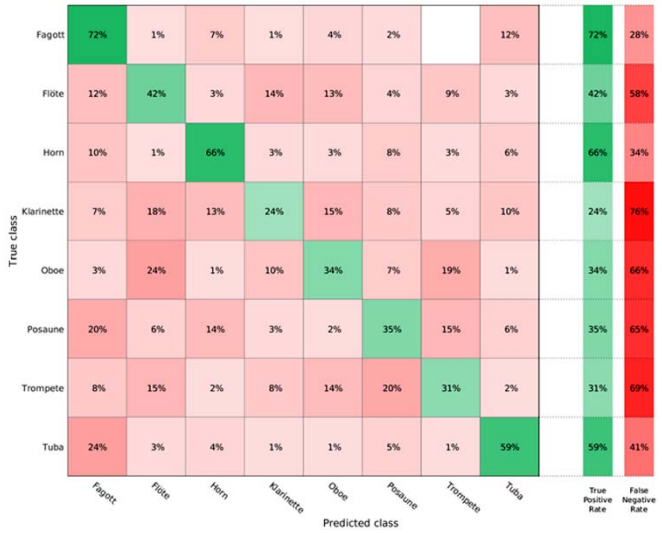

Abbildung 10: Verwechslungsmatrix auf der Grundlage von F1, F2 und F3 (cubic KNN, Klassifikationspräzision von 46.1%).

Kombiniert man die Formanten mit zusätzlichen Klangeigenschaften (Attack Time, Spectral Flux, Roughness, Brightness, Spectral Entropy, Maximaler RMS-Energie-Wert, Spectral Centroid und Unpleasantness; letztere implementiert auf der Basis von [13], übrige ermittelt mit [11]), so lässt sich die Präzision der Instrumentenerkennung auf 84.6% erhöhen. Die so entstandene Verwechslungsmatrix zeigt plausible Parallelen zur menschlichen Wahrnehmung, da im menschlichen Alltag häufig vorkommende Instrumenten-verwechslungen sich auch auf dem rechnerischen Weg ergeben, wie z.B. bei den Klängen von Flöte, Klarinette und Oboe, bei denen von Trompete und Posaune oder bei denen von Tuba und Fagott (Abb. 11).

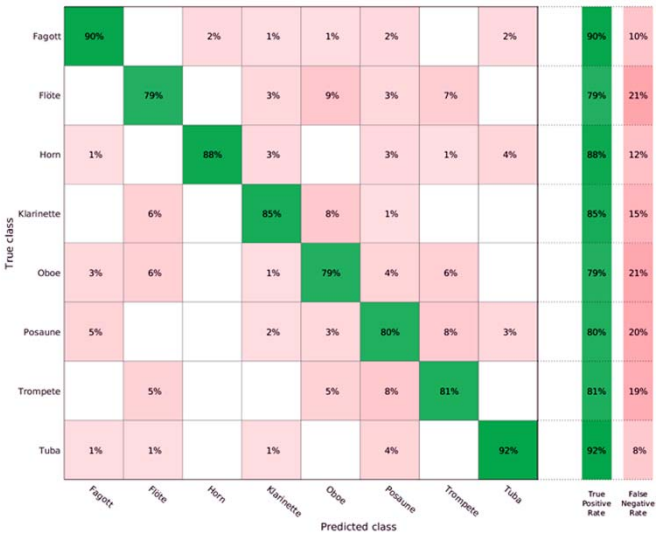

Abbildung 11: Verwechslungsmatrix auf der Grundlage von F1/F2/F3, Attack Time, Spectral Flux, Roughness, Brightness, Spectral Entropy, Maximaler RMS-Energie-Wert, Spectral Centroid und Unpleasantness (quadratic SVM, Klassifikationspräzision von 84.6%).

Zusammenfassung

Eine Instrumentenklassifikation durch Formanten allein auf der Grundlage von Machine-Learning ist zwar möglich, führt jedoch nicht zu besonders hohen Erkennungsraten. Es reicht hier jedoch schon eine geringe Anzahl an zusätzlichen Klangdeskriptoren aus, um zu einem sehr aussagekräftigen und auch der menschlichen Wahrnehmung entsprechenden Klassifikationsmodell zu kommen.

Instrumentalklangfarbenähnlichkeit lässt sich tatsächlich gut durch die Nähe der ersten beiden Formantbereiche beschreiben und auch (im Ergebnis vergleichbar mit MFCCs) vorhersagen. Zwar sind die einzelnen MFCCs als Klangdeskriptoren linear unabhängiger voneinander (während die Formanten stark miteinander korrelieren), jedoch ist der Vorteil von Formanten gegenüber MFCCs, dass Formanten mit nur zwei Werten sehr kompakt einen konkreten hörbaren und dadurch intuitiv zugänglichen spektralen Inhalt beschreiben und durch eine mehr als 100jährige Forschungsgeschichte eine sichere Grundlage bieten.

Aus diesen Gründen sollten Formanten (insbesondere für Blasinstrumente) als Alternativen zu den MFCCs im Bereich des Music Information Retrievals eingesetzt werden.

Die vorliegende Arbeit wurde vom Jubiläumsfonds der Österreichischen Nationalbank gefördert (OeNB Projekt 16473) sowie von der Vienna Symphonic Library (VSL) mit zwei Super Packages unterstützt.

Literatur

[1] Stumpf, C.: Die Sprachlaute. Springer, Berlin, 1926.

[2] Siedenburg, K.; Jones-Mollerup, K.; McAdams S.: Acoustic and categorical dissimilarity of musical timbre: Evidence from asymmetries between acoustic and chimeric sounds. Frontiers in Psychology, 6:1977 (2016), doi: 10.3389/fpsyg.2015.01977.

[3] Loughran, R.; Walker, J.; O'Neill, M.; O'Farrell, M.: Musical instrument identification using principal component analysis and multi-layered perceptrons. International Conference on Audio, Language and Image Processing ICALIP 2008, S. 643–648.

[4] Elliott, T.; Hamilton, L.; Theunissen, F.: Acoustic Structure of the five Perceptual Dimensions of Timbre in Orchestral Instrument Tones. JASA 133:1 (2013), S. 389–404.

[5] Schumann. K.E.: Physik der Klangfarben. Berlin 1929.

[6] Mertens, P.-H.: Die Schumannschen Klangfarbengesetze und ihre Bedeutung für die Übertragung von Sprache und Musik. Bochinsky, Frankfurt 1975.

[7] Meyer, J.: Akustik und Musikalische Aufführungspraxis. PPV Medien, Bergkirchen, 2015.

[8] Reuter, C.; Czedik-Eysenberg, I.; Siddiq, S.; Oehler, M.: Formanten als hilfreiche Timbre-Deskriptoren für die Darstellung von Blasinstrumentenklängen. Fortschritte der Akustik. DAGA 2017. 43. Deutsche Jahrestagung für Akustik. Kiel 2017, S. 190–193.

[9] Reuter, C.: Die auditive Diskrimination von Orchesterinstrumenten. Lang, Frankfurt, 1996.

[10] Boersma, P.; Weenink, D. Praat: doing phonetics by computer [Computer program], Version 5.3.51, retrieved 2 June 2013 URL: http://www.praat.org/

[11] Lartillot, O.; Toiviainen, P.: A Matlab toolbox for musical feature extraction from audio. International Conference on Digital Audio Effects, September 2007, S. 237–244.

[12] Darch, J.; Milner, B.; Shao, X.; Vaseghi, S.; Yan, Q.: Predicting formant frequencies from MFCC vectors. Acoustics, Speech, and Signal Processing, ICASSP'05 Proceedings of the IEEE International Conference 2005, Vol. 1, S. I/941–I/944.

[13] Reuter, C.; Oehler, M.; Mühlhans, J.: Physiological and acoustical correlates of unpleasant sounds. In: Proceedings of the Joint Conference ICMPC13-APSCOM5, August 4–8, 2014, Yonsei University, Seoul, Korea, S. 97.