

Klasse Formanten!

Klassifikation, Verschmelzung und klangliche Ähnlichkeit aus menschlicher und maschineller Perspektive

Christoph Reuter¹, Isabella Czedik-Eysenberg¹, Saleh Siddiq¹, Michael Oehler²

¹ Musikwissenschaftliches Institut der Universität Wien,

² Institut für Musikwissenschaft und Musikpädagogik der Universität Osnabrück

Hintergrund

Für die Beschreibung der Klangfarbe von ganzen Musikinstrumenten gibt es nur wenige Konzepte. Neben Mel Frequency Cepstral Coefficients (MFCCs, z.B. Loughran et al., 2008) und Modulation Power Spectrum (MPS, z.B. Elliott et al., 2013) haben sich für eine umfassendere Beschreibung von ganzen Musikinstrumentenklingen vor allem Formantbereiche bewährt (seit Schumann, 1929; Mertens, 1975; Meyer, 2015). Besonders für Blasinstrumente konnten sowohl ihre Entstehungsursachen geklärt werden (Fricke, 1975; Voigt, 1975) als auch ihr Beitrag zur Heraushörbarkeit und zur Verschmelzung von Musikinstrumenten im polyphonen und Unisono-Zusammenspiel (Fricke, 1986; Reuter, 1996). Auf dieser Basis und der Grundlage von 586 Klängen der gängigen Blasinstrumente in allen erreichbaren Tonhöhen und in den Dynamikstufen *pp* und *ff* wurde für eine formantbasierte Darstellung der Instrumentalklangfarbe ein interaktives Formantenfeld erstellt (Reuter et al., 2017). Die Darstellung von Blasinstrumentenklingen und -kombinationen in diesem Formantenfeld (X-Achse: Formant 1; Y-Achse: Formant 2) bietet einen generalisierbaren und intuitiven Zugang zur Instrumentenerkennung, zur möglichen Klangfarbenähnlichkeit sowie zur möglichen Verschmelzung und Separierbarkeit von Blasinstrumentenkombinationen (vor allem bei Blechblas- und Doppelrohrblattinstrumenten). Im Gegensatz zu den für die Darstellung von Klangfarben eher ungeeigneten und nicht generalisierbaren Timbre Spaces (vergl. Siddiq et al., 2014 und 2015) lassen sich im Formantenfeld die Klangfarben ganzer Musikinstrumente in verschiedenen Registern und Dynamikstufen vergleichend darstellen.

Fragestellung

Allein durch die Position des ersten und zweiten Formanten lassen sich im derzeitigen Formantenfeld besonders die Klänge von Doppelrohrblatt- und Blechblasinstrumente je nach Dynamikstufe und Register mehr oder weniger genau lokalisieren. In vielerlei Hinsicht entsprechen die ermittelten Formantpositionen den in der Literatur zu findenden Beschreibungen, so dass sich vor allem zwei Fragestellungen ergeben:

- Werden Instrumentenklänge mit den hier errechneten Formantpositionen im Paarvergleich tatsächlich als ähnlich und im Zusammenspiel als miteinander verschmelzend empfunden, wenn sie im Formantenfeld nahe beieinander liegen? Und werden sie im Paarvergleich als stark unterschiedlich und im Zusammenspiel als gut separierbar empfunden, wenn sie im Formantenfeld weit auseinander liegen?
- Bietet das Formantenfeld in der derzeitigen Form eine genügend starke Genauigkeit, um Instrumentenklänge anhand ihrer Formantpositionen nach Klangfarben zu klassifizieren und rechnerisch voneinander zu trennen? Welche zusätzlichen Timbre Features müssen eventuell mitberücksichtigt werden, um eine Verbesserung der Klassifikation zu erreichen?

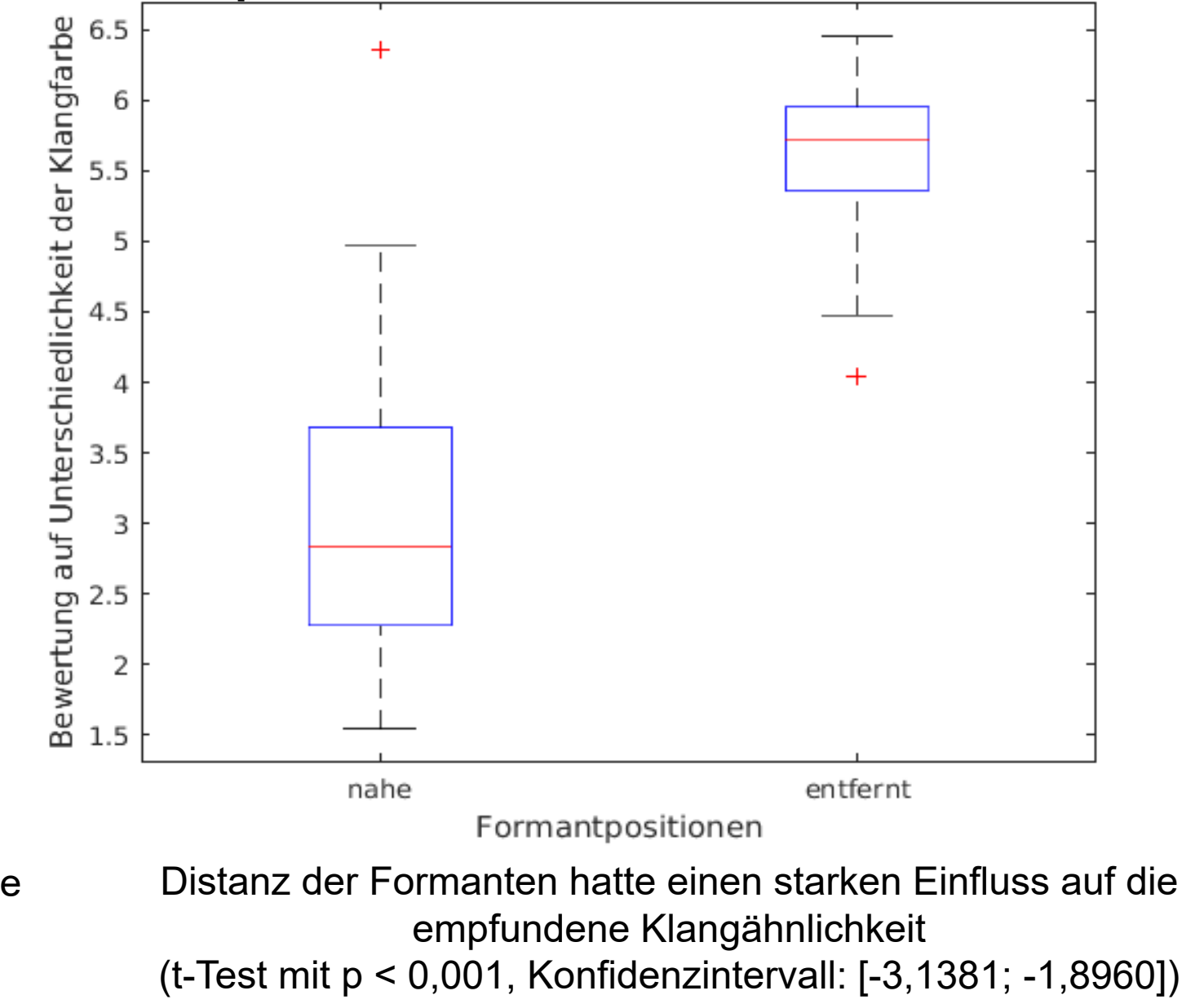
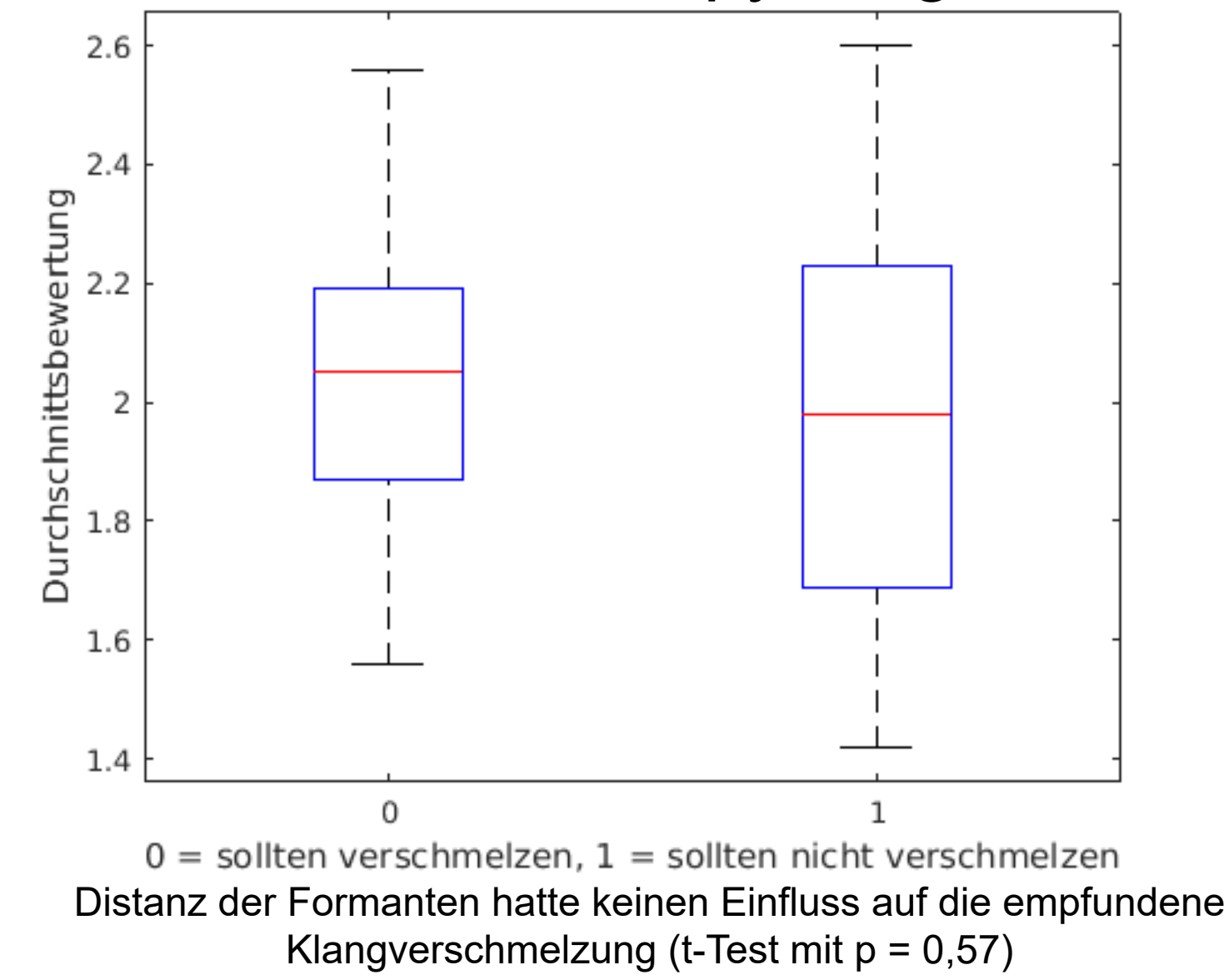
Methoden

In zwei Hörversuchen wurden jeweils 40 in ihrer Lautheit angeglichenen Klangpaare mit 20 nahe beieinander liegenden und 20 weit voneinander entfernten Formantpositionen im Formantenfeld hinsichtlich ihrer Ähnlichkeit und ihrer Verschmelzungsfähigkeit bewertet: Im Hörversuch zur Klangähnlichkeit bewerteten 22 Versuchspersonen ($\sigma^2=13/\sigma=9$, Alter: 20-48, Ø 25 Jahre, Instrumentalpraxis: Ø 11,5 Jahre) die sukzessiv erklingenden Klangpaare auf einer Skala von 1-8 auf ihre Unähnlichkeit (8 = maximale Unähnlichkeit), während im Hörversuch zur Verschmelzungsfähigkeit 25 Versuchspersonen ($\sigma^2=14/\sigma=11$, Alter: 20-48, Ø 25 Jahre, Instrumentalpraxis Ø 13 Jahre) die simultan erklingenden Klangpaare daraufhin beurteilten, ob sie zwei verschiedene, zwei gleiche oder nur ein einziges Instrument hörten. Wurden zwei gleiche oder nur ein einziges Instrument wahrgenommen, so wurde dies als klangliche Verschmelzung gewertet.

Parallel dazu wurde mit Hilfe von Support Vector Machines (SVM), k-Nearest-Neighbor (KNN) und anderen Klassifikationsverfahren mit fünffacher Kreuzvalidierung ermittelt, mit welcher Genauigkeit die Instrumentalklänge anhand ihrer Formantbereiche (F1/F2/F3) erkannt werden können und sich in sinnvolle Gruppen unterteilen lassen. Darüber hinaus wurde untersucht, mit Hilfe welcher zusätzlichen Timbre-Deskriptoren die Klassifikationspräzision erhöht werden kann.

Ergebnisse

Während die Versuchspersonen nahezu alle simultan erklingenden Klangpaare unabhängig von ihrer Position im Formantenfeld als verschmelzend wahrnahmen, zeigte sich die empfundene klangliche Ähnlichkeit bei den sukzessiv erklingenden Paaren eindeutig als abhängig von der Distanz zwischen den Formantbereichen. Die euklidische Distanz zwischen F1 und F2 im Formantenfeld weist eine Korrelation von $r = 0,759$ ($p < 0,001$) mit den (Un-)Ähnlichkeitsurteilen auf. Bei 89 getesteten Merkmalen ergaben sich hier nur für Entropy, Brightness und Unpleasantness höhere Werte.

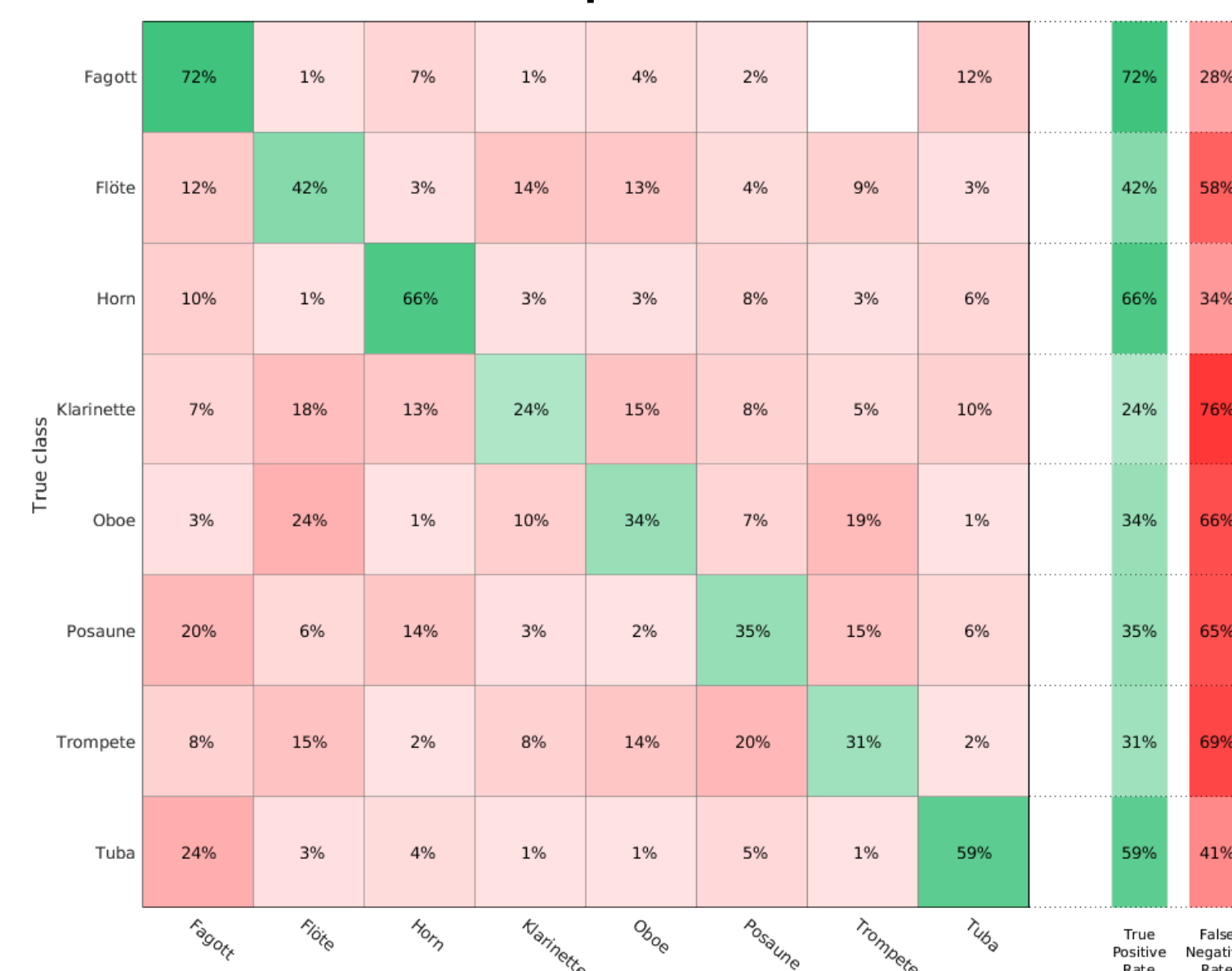


Klänge mit nahe beieinander liegenden Formanten wurden als sehr ähnlich empfunden und Klänge mit weit voneinander entfernten Formantpositionen wurden als sehr unähnlich empfunden.

feature combination	recognition	classification model
F1/F2/F3	46,1%	Cubic KNN
F1/F2/F3 + F0	48,5%	Cubic SVM
F1/F2/F3 + brightness	49,7%	Medium KNN
F1/F2/F3 + roughness	50,5%	Bagged Trees
F1/F2/F3 + firstAttackTime	50,7%	Cubic SVM
F1/F2/F3 + entropy	51,5%	Weighted KNN
F1/F2/F3 + spectralFluxMean	51,7%	Cubic SVM
F1/F2/F3 + spectralCentroid	52,6%	Cubic SVM
F1/F2/F3 + unpleasantness	53,1%	Cubic SVM
F1/F2/F3 + MaxRMSValue	53,8%	Weighted KNN
F1/F2/F3 + MaxRMSValue + entropy	61,3%	Cubic SVM
F1/F2/F3 + firstAttackTime + spectralFluxMean + roughness + brightness + entropy + maxRmsValue + unpleasantness	84,6%	Quadratic SVM

Mit Hilfe von maschinellen Klassifikationsverfahren lassen sich Musikinstrumente anhand ihrer ersten drei Formanten (F1/F2/F3) mit einer Präzision von 46,1% zuordnen (mittels Cubic KNN).

Kombiniert man die Formanten mit weiteren Timbre Features (siehe Tabelle), so lässt sich die Präzision der Instrumentenerkennung auf 84,6% erhöhen. Auch die so erhältliche Verwechslungsmatrix entspricht menschlichen Instrumentenverwechslungen recht gut, indem z.B. die Klänge von Flöte, Klarinette und Oboe häufig miteinander verwechselt werden, ebenso wie die von Trompete und Posaune oder die von Tuba und Fagott.



Verwechslungsmatrix auf der Grundlage von F1, F2 und F3 (cubic KNN, Klassifikationspräzision von 46,1%)



Verwechslungsmatrix auf der Grundlage von F1/F2/F3, firstAttackTime, spectralFluxMean, roughness, brightness, entropy, maxRmsValue, spectralCentroid, unpleasantness (quadratic SVM, Klassifikationspräzision von 84,6%)

Auffallend ist, dass auch eine Kombination von Brightness, Entropy und Unpleasantness mit 59,9% (mit Boosted Trees) schon eine größere Erkennungsrate ergibt als die ersten drei Formanten alleine.

Zusammenfassung

Insgesamt zeigen die Hörversuche, dass sich die empfundene klangliche Ähnlichkeit von sukzessiv gespielten Instrumentalklingen gemäß der Formanttheorie tatsächlich durch die Nähe ihrer ersten beiden Formanten beschreiben lässt. Die Empfindung von klanglicher Verschmelzung und Separierbarkeit bei simultan erklingenden Musikinstrumenten ließ sich jedoch nicht anhand der Formantpositionen vorhersagen. Dieses Ergebnis steht im völligen Widerspruch zu vorangegangenen Untersuchungen. Es kann damit zusammenhängen, dass im vorliegenden Hörversuch die Klänge lautheitsangeglichen gemischt wurden (sonst pegelangepasst) und dass hier Mischungen von jeweils zwei Einzeltönen beurteilt wurden im Gegensatz zu sonst simultan erklingenden Tonleiterausschnitten (vergl. Reuter, 1996 & 2000). Formanten mögen nach diesen Ergebnissen vielleicht nicht die effizienteste Methode für eine automatische Instrumenten-Klassifikation sein, jedoch bilden sie ein recht gutes (und theoretisch untermauertes) Kriterium für die Wahrnehmung von klanglicher (Un-)Ähnlichkeit.

Literatur

Elliott, T.M.; Hamblin, L.S.; Theunissen, F.E. (2013). Acoustic structure of the five perceptual dimensions of timbre in orchestral instrument tones. In: JASA 133(1), S. 389-404. → Fricke, J.P. (1975). Formantbilde Impulsfolgen bei Blasinstrumenten. In: Fortschritte der Akustik 4, DAGA75 (pp. 407-411). Braunschweig. → Fricke, J.P. (1986). Prose, J.P. (1986). Zur Anwendung optischer Klangfarbenkarten bei Aufnahme und Wiedergabe. In: Bericht über die 14. Tagung der Akustik, München. → Loughran, D.; Toivanen, P.; Ervasti, T. (2007). A Matlab Toolbox for Music Information Retrieval. In: Data Analysis, Machine Learning and Applications, London. Springer, 281-288. → Loughran, D.; Walker, J.; O'Neill, M. (2008). The Use of Mel-Frequency Cepstral Coefficients in Musical Instrument Identification. In: Proceedings of the International Computer Music Conference, 24-29 August, 2008, Belfast, Northern Ireland. → Mertens, P. et al. (1975). Die Schwingenformen, Klangfarben und ihre Bedeutung für die Übertragung von Sprache und Musik. Frankfurt. → Meyer, J. (2015). Akustik und musikalische Aufführungspraxis. Bergheim: PPK-Verlag. → Reuter, C. (1996). Die auditive Dimension von Orchesterinstrumenten. Frankfurt: Lang. → Reuter, C. (2000). Verschmelzung und partielle Verschmelzung. In: Fortschritte der Akustik 2000 (pp. 175-177). Osnabrück. → Reuter, C.; Czedik-Eysenberg, I.; Siddiq, S.; Oehler, M. (2017). Formanten zur Beschreibung der Klangfarbe von Blasinstrumenten. In: Fortschritte der Akustik 43, DAGA 2017 (pp. 180-183). Kiel. → Siddiq, S.; Reuter, C. (2015). Towards the comparability and generality of timbre spaces. In: Proceedings of the 15th Triennial ISMIR (pp. 222-233), 16-19 Sept. 2015, University of Music and Performing Arts, Vienna. → Siddiq, S.; Reuter, C. (2016). Untersuchungen zur Formantdarstellung in Klängen von Fagott und Oboe. Bergheim: PPK-Verlag.