

Klingt Sinus blau und Sägezahn rot? Eine Untersuchung zu Crossmodal Correspondences bei der Wahrnehmung von synthetischen Wellenformen

Gabriel Feller, Christoph Reuter

Musikwissenschaftliches Institut, Universität Wien, 1090 Wien, E-Mail: feller.gabriel@gmx.de

Hintergrund: Crossmodal Correspondences- Forschung mit Sinus-Wellen

Crossmodal Correspondences sind die Verbindung von Sinnesinformationen miteinander, die in der Allgemeinheit auftreten. Bereits frühe Forschung zeigte verschiedene solcher Crossmodal Correspondences im audiovisuellen Bereich. So etwa die Verbindung von Tonhöhe mit der Größe eines Objekts, der räumlichen Höhenanordnung, der Helligkeit, Bewegungsrichtungen und Gewicht [1]. Bereits 1971 testeten Bernstein, Eason und Schurman mithilfe von Sinuswellen und einer sehr kleinen Proband*innen-Gruppe von vier Personen mögliche Assoziationen bestimmter Farben (rot und blau) zu Tonhöhen (100 Hz und 1000 Hz). Allerdings konnten sie keine signifikanten Korrelationen finden [2]. Auch Ward, Huckstep und Tsakanikos untersuchten 2006 Crossmodal Correspondences mit Sinuswellentönen bzw. reinen Tönen im Vergleich zu anderen Klängen. Es wurde ein Zusammenhang zwischen Tonhöhe und assoziierter Farbhelligkeit bestätigt, der Farbton hingegen schien vornehmlich mit der Klangfarbe zusammenzuhängen [3]. 2012 zeigten Parise und Spence, dass unterschiedliche synthetische Wellenformen anders wahrgenommen werden. Zwar wurden diese nicht auf Farbassoziationen getestet, allerdings zeigte sich eine stärkere Assoziation zwischen Rechteckwellen und kantigeren visuellen Stimuli und zwischen Sinuswellen und runderen visuellen Stimuli [4].

Fragestellungen

Vor diesem Hintergrund soll in einem Pilotexperiment folgenden Fragestellungen nachgegangen werden:

Welche Crossmodal Correspondences treten bei synthetischen Wellenformen auf?

Können bisherige Untersuchungen mit synthetischen Wellenformen bestätigt werden?

Gibt es signifikante Unterschiede in der Beurteilung der Wellenformen?

Methoden

In einem Pilotexperiment bewerteten 44 Proband*innen ($M_{Age}=37,8$, $SD=16,5$) insgesamt 20 Klangbeispiele zweimal in randomisierter Reihenfolge. Bei den Klangbeispielen handelte es sich um vier Sekunden lange Sinus-, Dreiecks-, Rechteck- und Sägezahnwellentöne, jeweils im Oktavabstand im Bereich C2-C6.

Im ersten Durchlauf mussten die Proband*innen die Klänge in zehn Antwortskalen mit Konstrukten bestehend aus gegensätzlichen Wortpaaren einordnen. Die Wortkonstrukte hierzu stammen aus der Studie von Grill, Flexer und Cunningham (2011) zur Identifikation von Klangtexturen [5]:

hoch, hell	↔	tief, dunkel
regelmäßig, geordnet	↔	unregelmäßig, chaotisch
glatt, weich	↔	rau, rau
natürlich, analog	↔	künstlich, digital
statisch, starr	↔	dynamisch, bewegt
nahe, klar	↔	fern, verschwommen
kantig, zerrissen	↔	rund, kompakt
dicht, flächig	↔	spärlich, punktuell
homogen, gleichförmig	↔	heterogen, differenziert
tonhaltig	↔	geräuschhaft

Die Werte der Skala nahmen von den linken Wörtern hin zu den rechten Wörtern zu.

Im zweiten Durchgang mussten die Proband*innen aus der Farbtafel des Berkeley Color Projects [6] drei Farben auswählen, die sie zu dem Klang am passendsten fanden, wie bereits bei Ambros et al. [7]. Alle ausgewählten Farben wurden nach Stimmenanzahl gewichtet im RGB-Farbmodell gemittelt und zur Untersuchung in den HSV-Farbraum übertragen.

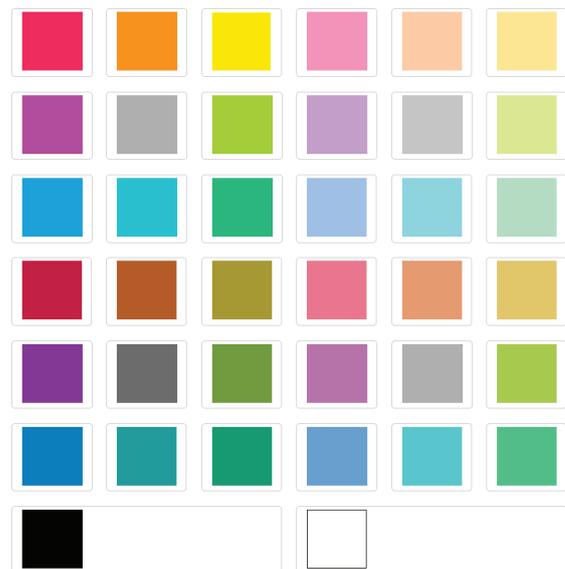


Abbildung 1: Darstellung der Farbtafel des Berkeley Color Projects im Online-Fragebogen.

Via Signalanalyse [8][9][10][11][12] wurden die Sinusklänge auf 61 berechenbare Audiomerkmale hin untersucht. Von diesen wurden fünf bereits in der Studie von Ambros et al. verwendet [7].

Über Korrelations- und Regressionsanalysen wurden Zusammenhänge zwischen den Farbdimensionen, den Einordnungen in die Wortkonstrukte und den errechneten Audiomerkmale ermittelt.

Ergebnisse

Schon bei der Aufschlüsselung der meistgewählten Farbe(n) jedes Klanges, lassen sich Gemeinsamkeiten und Unterschiede zwischen den synthetischen Wellenformen finden. Die tiefen Töne werden viel mit Schwarz und Grau assoziiert, die hohen Töne mit warmen Farben (Gelb und Rot). Bei den Sinus- und Dreieckswellentönen zeigen sich aber Tendenzen, kalte Farben (Blau, Grün) im mittleren Tonbereich zu wählen (s. Tabelle 1). Bei den gemittelten Farbwerten pro Klangbeispiel zeigt sich diese Tendenz noch deutlicher, dass teiltonarme Klänge (Sinus/Dreieckswelle) eher mit kälteren Farben assoziiert werden (blau/grün) als teiltonreiche Klänge (Rechteck/Sägezahnwelle -> braun/orange, s. Tabelle 2).

Tabelle 1: Darstellung der am häufigsten ausgewählten Farben je Klangbeispiel

Wellenform	C2	C3	C4	C5	C6
Dreieck	Black	Grey	Light Blue	Light Green / Yellow	Yellow
Rechteck	Grey	Grey	Orange	Pink	Pink
Sägezahn	Grey	Grey	Orange	Pink	Yellow
Sinus	Grey	Blue	Light Blue	Pink	Yellow

Tabelle 2: Darstellung der durchschnittlichen Farben je Klangbeispiel

Wellenform	C2	C3	C4	C5	C6
Dreieck	Dark Green	Dark Green	Dark Green	Dark Green	Orange
Rechteck	Brown	Brown	Brown	Brown	Orange
Sägezahn	Brown	Brown	Brown	Brown	Orange
Sinus	Dark Green	Dark Green	Dark Green	Dark Green	Orange

Eine Darstellung der Farbtöne (Hues) in Regenwolken-diagrammen zeigt, dass sich die assoziierten Hues von Rechteck- und Sägezahnwellen sehr auf einen Punkt konzentrieren, während sich die assoziierten Hues von Sinus- und Dreieckswellen mehr in Richtung Blautöne verteilen (s. Abb. 2).

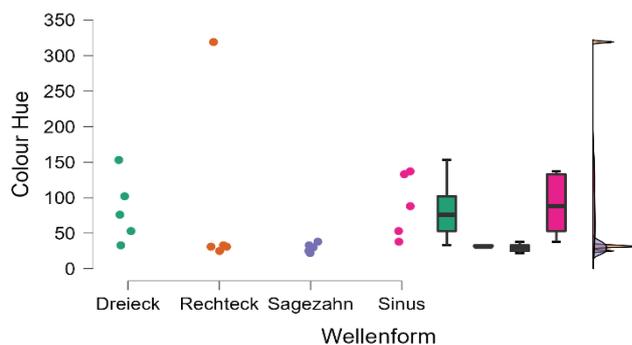


Abbildung 2: Verteilung der Color Hues der durchschnittlichen Farben nach Wellenform

Audiomerkmale und Farbdimensionen

In der Korrelationsanalyse zeigten sich folgende Zusammenhänge zwischen Audiomerkmale und Farb-Dimensionen:

Betrachtet man die Gesamtheit aller Wellenformen, so zeigen sich vor allem Korrelationen zur klanglichen Helligkeit und zur Rauigkeit, d.h.: je heller die Klangfarbe, desto stärker gesättigte Farben werden dazu assoziiert (s. Tabelle 3).

Tabelle 3: Korrelationen zwischen Farbsättigung und Klangfarbenmerkmalen

Farbsättigung korreliert mit	r	p
Spectral Centroid	0,597	0,005
Timbral Hardness	0,596	0,006
Brightness	0,591	0,006
Spectral Bandwidth	0,59	0,006
Timbral Sharpness	0,586	0,007
Timbral Brightness	0,58	0,007
Timbral Roughness	0,55	0,012
Spektrale Energie (4-20 kHz)	0,494	0,027

Einfluss auf den assoziierten Farbton (Color-Hue) haben ebenfalls klangliche Aspekte der Helligkeit (Zero-Crossing-Rate) und der klanglichen Wärme bzw. des Dröhnens (also vor allem die Ausgeprägtheit tiefer Frequenzanteile). Mit anderen Worten: je wärmer bzw. weniger hell die Klangfarbe, desto eher werden grüne/blau Farbtöne assoziiert und je heller/kälter die Klangfarbe, desto mehr werden gelbe/rote Farbtöne assoziiert (s. Tabelle 4).

Tabelle 4: Korrelationen zwischen Farbton und Klangfarbenmerkmalen

Farbton korreliert mit	r	p
Timbral Warmth	0,519	0,019
Timbral Booming	0,478	0,033
Timbral Depth	0,461	0,041
Zero-Crossing-Rate	-0,454	0,044

Ähnliches gilt für die Farbhelligkeit der je nach Wellenform ausgewählten Farben: je klanglich heller, rauer, härter die Klänge sind, je höher ihre Frequenz ist und je weniger dröhnend und warm sie sind, desto hellere Farben werden mit diesen Klängen assoziiert (s. Tabelle 5).

Tabelle 5: Korrelationen zwischen Farbhelligkeit und Klangfarbenmerkmalen

Farbhelligkeit korreliert mit	r	p
Zero-Crossing-Rate	0,584	0,007
Timbral Brightness	0,549	0,012
Spektrale Energie (4-20 kHz)	0,524	0,018
Frequenz	0,523	0,018
Timbral Sharpness	0,512	0,021
Timbral Hardness	0,479	0,033
Spektrale Energie (0,8-4 kHz)	0,472	0,035
Spectral Centroid	0,455	0,044
Timbral Depth	-0,536	0,015
Timbral Booming	-0,612	0,004
Timbral Warmth	-0,632	0,003

Auch bei den gefundenen Korrelationen der Klangfarbenmerkmale zu den einzelnen RGB-Farbwerten wird deutlich, dass der Rot-Anteil vor allem von Frequenz und klanglicher Helligkeit bestimmt wird und der Blau-Anteil besonders durch eine geringe spektrale Bandbreite (d.h. Teiltonarmut) und damit zusammenhängende Merkmale (niedriger spektraler Schwerpunkt, geringe Härte, Schärfe und Rauigkeit, s. Tabelle 5). Auch zeigt sich, dass die Timbral Warmth negativ mit dem Rot- und positiv mit dem Blau-Anteil korreliert. Wärmere Klänge werden also mit kälteren Farben assoziiert.

Tabelle 5: Korrelationen zwischen RGB-Farbwerten und Klangfarbenmerkmalen (**= $p < 0,001$, *= $p < 0,01$, = $p < 0,05$)

RGB-Farbwerte korrelieren mit	Rot-Anteil (r)	Grün-Anteil (r)	Blau-Anteil (r)
Frequenz	0,902***	0,533*	
Zero-Crossing-Rate	0,933***	0,491*	
Spectral Centroid	0,472*	-0,46*	-0,82***
Spectral Strongpeak	0,796***	0,638**	
Spectral Bandwidth		-0,531*	-0,819***
Spectral Complexity		-0,79***	-0,512*
Spectral Energy	-0,474*		
Spectral Energy (20-150 Hz)	-0,492*		
Spectral Energy (150-800 Hz)			0,468*
Spectral Energy (0,8-4 kHz)	0,737***		
Spectral Energy (4-20 kHz)	0,576**		0,682***
Timbral Hardness	0,549*		-0,845***
Timbral Brightness	0,648**		-0,803***
Timbral Sharpness	0,589**		-0,907***
Timbral Warmth	-0,904***		0,525**
Timbral Booming	-0,912***		0,574*
Timbral Roughness		-0,709***	-0,837***
Timbral Depth		-0,543*	
Hardness		-0,508*	

Dies zeigte sich auch im t-Test (Sinus/Dreieck vs. Rechteck/Sägezahn): Blau- und Grün-Anteile (RGB-System) sind signifikant geringer bei Farben vertreten, die bei Wellenformen mit einem höheren Teiltonreichtum assoziiert werden (blau: $t = -5,804$, $p < 0,001$; grün: $t = -3,627$, $p = 0,002$).

Audiomerkmale und Wortkonstrukte

Auch zwischen den Einordnungen in die Wortkonstrukte und den Audiomerkmalen scheint es Zusammenhänge zu geben.

Erwartbar waren die Korrelationen des Konstrukts „hoch, hell ↔ tief, dunkel“ mit der Tonhöhe ($r = -0,912$, $p < 0,001$), der Helligkeit ($r = 0,562$, $p = 0,010$), der spektralen Energie in den Bereichen 0,8-4 kHz ($r = -0,693$, $p < 0,001$) und 4-20 kHz ($r = 0,478$, $p = 0,033$).

Besonders großen Einfluss scheint die Höhe des Spectral Centroids zu haben. Dieser korreliert mit acht der zehn Wortkonstrukte (s. Tabelle 6):

Tabelle 6: Korrelationen zwischen Spectral Centroid und Wortkonstrukten (**= $p < 0,001$, *= $p < 0,01$, = $p < 0,05$)

Spectral Centroid korreliert mit	r
natürlich, analog ↔ künstlich, digital	0,903***
glatt, weich ↔ rau, rau	0,901***
tonhaltig ↔ geräuschhaft	0,893***
homogen, gleichförmig ↔ heterogen, differenziert	0,824***
dicht, flächig ↔ spärlich, punktuell	0,729***
regelmäßig, geordnet ↔ unregelmäßig, ungeordnet	0,723***
nahe, klar ↔ fern, verschwommen	-0,503*
kantig, zerrissen ↔ rund, kompakt	-0,934***

Dagegen korreliert die Inharmonizität lediglich mit dem Konstrukt „nahe, klar ↔ fern, verschwommen“ ($r = 0,540$, $p = 0,017$).

Mit Hilfe eines t-Tests lassen sich die schon bei den Farbassoziationen beobachteten Unterschiede zwischen teiltonarmen (Sinus/Dreieckswelle) und teiltonreichen Klängen (Rechteck/Sägezahnwelle) deutlich in den damit assoziierten Wortkonstruktionen finden (s. Tabelle 7, Abb. 3).

Tabelle 7: Ergebnisse des t-Tests zu den Unterschieden zwischen Sinus/Dreieckswelle vs. Rechteck/Sägezahnwelle und deren Entsprechungen in den Wortkonstrukten (**= $p < 0,001$, *= $p < 0,01$, = $p < 0,05$)

Unterschiede (Sinus/Dreieck ↔ Rechteck/Sägezahn)	t
glatt, weich ↔ rau, rau	11,61***
tonhaltig ↔ geräuschhaft	10,174***
homogen, gleichförmig ↔ heterogen, differenziert	8,185***
natürlich, analog ↔ künstlich, digital	6,771***
regelmäßig, geordnet ↔ unregelmäßig, chaotisch	6,675***
dicht, flächig ↔ spärlich, punktuell	3,234*
kantig, zerrissen ↔ rund, kompakt	-12,422***

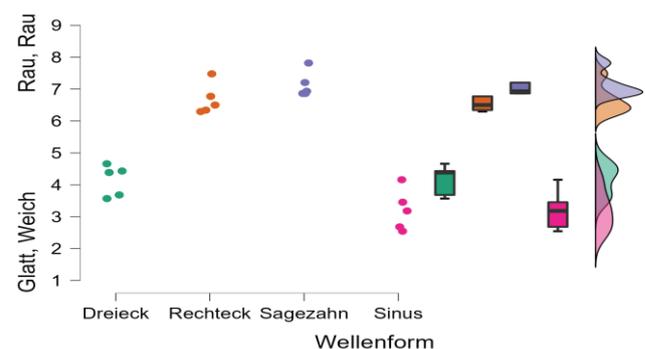


Abbildung 3: Verteilung der Klänge im Konstrukt „glatt, weich ↔ rau, rau“ nach Wellenform

Farbdimensionen und Wortkonstrukte

Betrachtet man alle Wellenformen gemeinsam ergeben sich auch zwischen den Dimensionen des HSV-Farbraums und den Wortkonstrukten Korrelationen (s. Tabelle 8).

Tabelle 8: Korrelationen zwischen HSV-Farbwerten der assoziierten Farben und Wortkonstrukten (**= $p < 0,001$, **= $p < 0,01$, *= $p < 0,05$)

Wortkonstrukte korrelieren mit Farbassoziationen aus dem HSV-Farbraum	H (Hue, r)	S (Sättigung, r)	V (Helligkeit, r)
glatt, weich ↔ rau, rau		0,565**	
kantig, zerrissen ↔ rund, kompakt		-0,586**	
tonhaltig ↔ geräuschhaft		0,553*	
natürlich, analog ↔ künstlich digital		0,631**	0,567**
hoch, hell ↔ tief, dunkel			-0,597**
nahe, klar ↔ fern, verschwommen			-0,566**
dicht, flächig ↔ spärlich, punktuell	-0,649**		0,559*

So werden assoziierte Farben mit ansteigender Sättigung mit rauheren, kantigeren, geräuschhaften und eher künstlich/digitalen Wortkonstrukten verbunden, während sich assoziierte Farben mit ansteigender Helligkeit eher höheren/helleren, näheren/klareren, punktuelleren/spärlicheren und ebenfalls eher künstlich/digitalen Wortkonstrukten zur Beschreibung von Klangfarben zuordnen lassen. Mit absteigendem assoziierten Farbton hingegen (Hue, d.h. je mehr die Farbe von Blau- zu Rottönen wandert) werden den Klängen spärlichere/punktuellere Wortkonstrukte zugeordnet.

Bedeutung für Crossmodal Correspondences

Die Pilotstudie zeigte, dass bekannte Crossmodal Correspondences auch bei synthetischen Wellenformen anwendbar sind, wie Tonhöhe und Farbhelligkeit [1][7], sowie dass die Einordnung der Klänge in Wortkonstrukte auch mit bisherigen visuellen Crossmodal Correspondences übereinstimmt. So werden Rechteckwellentöne eher als kantig und zerrissen beschrieben, während Sinuswellentöne als runder und kompakter wahrgenommen werden. Dies passt zu Parise und Spences Ergebnissen einer Untersuchung mithilfe von impliziter Assoziation zwischen genannten Wellenformen und den Bildern einer Rundung und eines spitzen Winkels [4]. Visuelle Crossmodal Correspondences lassen sich also auch in ihrer deskriptiven Form wiederfinden. Die Einordnung in „hoch, hell ↔ tief, dunkel“ stimmt ebenfalls mit visuellen Crossmodal Correspondences überein. Hier ist allerdings zu beachten, dass das Wortkonstrukt die Helligkeit bereits mit der Höhe verbindet.

Zusammenfassung

Mehrere Crossmodal Correspondences wie zwischen Farbhelligkeit und Tonhöhe oder Wellenform und Kantigkeit konnten bestätigt werden. Auch für bereits verworfene Ideen wie ein Zusammenhang zwischen Tonhöhe und Farbton

konnten signifikante Ergebnisse erzielt werden. Es ist allerdings zu beachten, dass sich neben der Tonhöhe noch viele weitere Variablen, beispielsweise der Teiltonreichtum, auf die Auswahl der assoziierten Farbe auswirken. Hinzu kommt, dass – obwohl es Konstrukte gibt, die durchaus dabei helfen können, Wellenformen voneinander zu unterscheiden – assoziierte Farbtöne nicht in diese Kategorie fallen, da auch diese sich innerhalb einer Wellenform stark verändern können. Wortkonstrukte wie jene von Grill, Fletcher und Cunningham [5] eignen sich besser, um zwischen verschiedenen synthetischen Wellenformen zu unterscheiden.

Literatur

- [1] Louden, S. J.: Cross-Sensory Perception in Music and Visual Media. A Neurogenetic Approach to Multimedia Counterpoint. Buffalo (2018).
- [2] Bernstein, I. H., Eason, T. R.: Hue-Tone Sensory Interaction. A Negative Result. In: Perceptual and Motor Skills 33/3 (1971), 1327-1330.
- [3] Ward, J., Huckstep, B., Tsakanikos, E.: Sound-Colour Synaesthesia. To What Extent Does It Use Cross-Modal Mechanisms Common to Us All? In: Cortex 42 (2006), 264-280.
- [4] Parise, C. V., Spence, C.: Audiovisual crossmodal correspondences and sound symbolism. A study using implicit association test. In: Experimental Brain Research 220/3-4 (2012), 319-333.
- [5] Gill, T., Flexer, A., Cunningham, S.: Identification of perceptual qualities in textual sounds using the repertory grid method. In: Proceedings of the 6th Audio Mostly Conference (2011), 67-74.
- [6] Palmer, S. E. et al.: Music-color associations are mediated by emotion. In: PNAS 110/22 (2013), 8836-8841.
- [7] Ambros, S. et al.: Sind crossmodal correspondences kulturell erlernt? Zuordnung von Farben und Tönen in Madagaskar. In: Fortschritte der Akustik – DAGA 2021. Deutsche Jahrestagung für Akustik (2021), 1076-1079.
- [8] Lartillot, O. et al. (2008). A Matlab Toolbox for Music Information Retrieval. In: C. Preisach et al. (Hrsg.). Data Analysis, Machine Learning and Applications (S. 261–268). Berlin, Heidelberg: Springer.
- [9] Lartillot, O. (2019). Miningsuite: A comprehensive Matlab framework for signal, audio and music analysis. Proceedings of the 16th SMC (S. 489-489), Málaga.
- [10] Bogdanov, D et al. (2013). Essentia: An Audio Analysis Library for Music Information Retrieval. Proceedings of the 14th ISMIR Conference (S. 493–498), Curitiba.
- [11] Pearce, A., et al. (2017). Timbral attributes for sound effect library searching. AES Proceedings, Erlangen.
- [12] McFee, B., et al. (2015). Librosa: Audio and music signal analysis in python. Proceedings of the 14th python in science conference (S. 18–25), Austin.