Popmusik und Wellenfeldsynthese: Der Einfluss der Tonmischung

Christoph Hold[1], Hagen Wierstorf[2], Alexander Raake[3]

[1] *Assessment of IP-based Applications, Technische Universität Berlin, Deutschland, Email: Christoph.Hold@alumni.tu-berlin.de*

[2] *Filmuniversität Babelsberg KONRAD WOLF, Deutschland*

[3] *Institut für Medientechnik, TU Ilmenau, Deutschland*

Einleitung

Der Prozess der Tonmischung ist ein häufig unterschätzter Arbeitsschritt in der Produktion von Popmusik. Diese Phase prägt den Klangeindruck maßgeblich und wirkt sich auf eine Vielzahl von Faktoren aus, darunter auf die wahrgenommene Qualität [1]. Diese beeinflusst wiederum positiv die Präferenz von Musikstücken [2], welche sich schließlich mit Kaufbereitschaft verknüpfen lässt [3].

Viele neuartige Wiedergabesysteme basieren auf dem Konzept des objektbasierten Audios. Obwohl einige perzeptive Eigenschaften dieser Systeme systematisch untersucht sind [4, 5], bleibt das Zusammenspiel mit der anliegenden Tonmischung weitgehend unbetrachtet. In einer Vorstudie zeigte sich jedoch bereits, dass die Artefakte wie sie typischerweise bei Wellenfeldsynthese (WFS) auftreten durch die Tonmischung korrigiert werden konnten [6]. In dieser Vorstudie wurden für unterschiedliche Musikstücke jeweils drei Referenzmischungen für Stereo, Surround und Wellenfeldsynthese erzeugt und in einem Paarvergleichstest die Präferenz der ZuhörerInnen ermittelt, wobei die Wellenfeldsynthese für alle Musikstücke das präferierte Wiedergabesystem darstellte.

Dieser Beitrag rückt den Fokus auf die Tonmischung und untersucht dafür verschiedene übliche Bearbeitungen wie Entzerrung, Kompression, Hall und Positionierung einzelner Quellen bezüglich ihrer quantitativen Auswirkung auf Präferenz. Ausgehend von der Referenzmischung für die Wellenfeldsynthese, dienen die zugrundeliegenden Einzelsignale des Pop-Stückes als Ausgangspunkt für die folgenden skalierten Parameteränderungen.

Ähnlichen Studien im Stereokontext [7, 8] zeigen signifikante Auswirkungen des Mixingprozesses auf Präferenz auf. Wir gehen davon aus, diese auch bei der Wiedergabe durch Wellenfeldsynthese festzustellen. Es stellt sich jedoch zusätzlich die Frage, wie groß der Einfluss der unterschiedlichen Wiedergabesysteme verglichen mit den Parameteränderungen in der Tonmischung ausfällt.

Methoden

Aufbau

Der Versuch fand im Raum Pinta des Telefunkengebäudes der Technischen Universität Berlin statt. Dieser ist ausgestattet mit einem runden 56 Kanal Lautsprecherarray (ELAC 301) sowie einem Subwoofer (Genelec 7060A). Der Raum ist akustisch durch einige Absorber optimiert. Die WFS Lautsprechersignale werden von einer modifizerten Variante des SoundScape Renderers [9] erzeugt. Die verwendete Version[1] [6] kompensiert den distanzabhängigen Amplitudenabfall und sorgt somit für positionsunabhängige Amplitude aller virtuellen Quellen. Für die Stereo- und Surroundwiedergabe spielen nur die entsprechenden Lautsprecher des Arrays zusammen mit dem Subwoofer, eingebunden per Bassmanagement.

Stimuli

Dieser Versuch übernimmt die Referenzmischungen aus der Vorstudie [6]. Der grundlegende Ablauf zur Stimulierzeugung und Tonmischung ist in [10] beschrieben. Ausgangspunkt bilden Einzelsignale[2] des Pop-Stückes *Lighthouse*. Der Tonschaffende kontrolliert dabei sowohl den Mix, als auch das folgende Wiedergabesystem (siehe Abb. 1, links). Es entsteht eine gemeinsame Tonmischung, deren Summenkanäle die drei Systeme speist. Nahtloses Umschalten der Systeme ermöglicht den Mix den Systemansprüchen entsprechend anzupassen.

Die für die Referenzmischung gewählten Einstellungen stellen übliche Praxis dar und fallen bereits recht stark aus, häufig mit mehreren Entzerrer- und Kompressorinstanzen pro Spur. Diese Referenzeinstellungen bilden den Ausgangspunkt für die folgenden skalierten Parameteränderungen, sodass die Versuchsteilnehmer dieser Studie neben den Systemen auch verschiedene Mixvarianten bewerten (siehe Abb. 1, rechts). Die Testpersonen hörten dafür immer den selben 30 Sekunden Ausschnitt des Stückes.

Die Parameteränderungen des Entzerrers (EQ) beziehen sich auf die für die Referenzmischung (REF) gefundenden Filterverstärkungen (Gain). Dabei wurde in jeder Equalizerinstanz und jedem Band die Verstärkung halbiert ($-$) oder verdoppelt ($+$), sowie der Equalizer komplett umgangen ($--$).

Ähnlich sind die Varianten für Kompression entstanden – dieses mal bezogen auf die Gain Reduction (GR). Der Threshold jedes Kompressors wurde variiert bis sich die Gain Reduction halbiert ($-$), verdoppelt ($+$) oder entsprechend umgangen ($--$).

Der künstliche Nachhall (Hall) wurde anhand des Return-busses jeder Effekteinheit verändert. Absenken

[1] `https://github.com/chris-hld/ssr/tree/without_amplitude`

[2] Einzelsignale online verfügbar; `https://doi.org/10.5281/zenodo.55718`

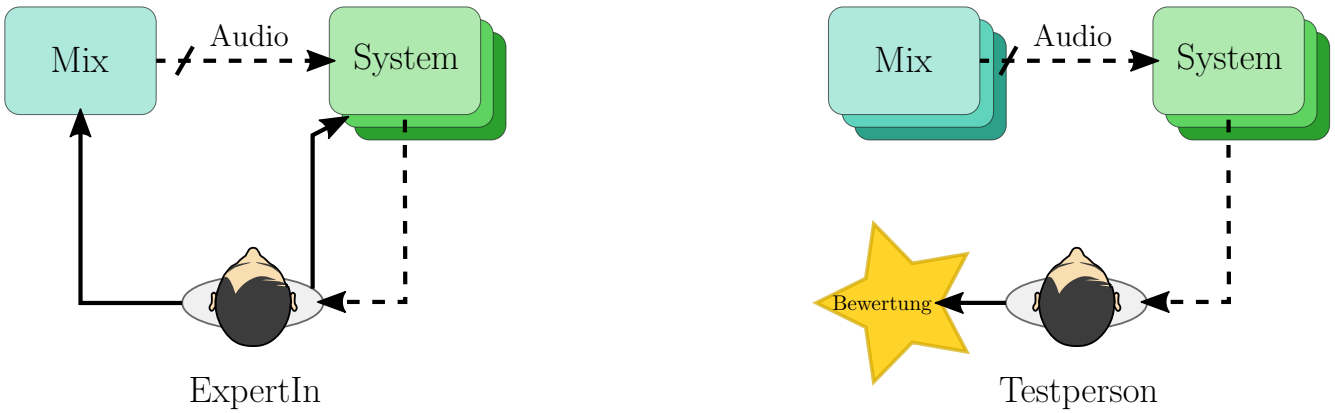

Abbildung 1: Schematischer Ablauf des Versuchs. Durchgezogene Linien markieren den direkten Einfluss der Personen, gestrichelte stellen den Audiopfad dar. Zuerst wird eine Referenzmischung erstellt, die verschiedene Wiedergabesysteme speist (links). Anschließend bewerten Testpersonen neben den Wiedergabesystemen auch verschiedene Mixvarianten (rechts).

um −6 dB (−), Anheben um +6 dB (+) oder stumm schalten (−−) stellen die Varianten dar.

Um Positionierungsänderungen zu untersuchen wurde nur der Vordergrund variiert, da ein Verändern der Gesamtbreite aus Mischungssicht weniger geeignet erschien. Die Wahl der Vordergrundelemente – Gesang, Snare, Bassdrum, Gitarren – fiel für das vorliegende Musikstück eindeutig aus und deckt sich den am häufigsten genannten Elementen einer verwandten Studie [2]. Die Verschiebung reicht von sehr eng (−−), über etwa die Breite von Stereo (−), bis hin zu breit (+) und sehr breit (++). Bei letzterem sind bereits die ersten Vordergrundelemente hinter der Hörposition platziert. Für eine detaillierte Beschreibung der Positionierung der Vordergrundquellen sei auf [11] verwiesen.

Tabelle 1: Stimuli – Varianten der Tonmischung für Wellenfeldsynthese

Kennung	Bearbeitung			
	EQ	Kompression	Hall	Positionierung
	Gain	Gain Reduction	Pegel	Vordergrund
−−	aus	aus	aus	eng
−	0.5	0.5	−6 dB	∼ Stereo
REF	1	1	0 dB	üblich
+	2	2	+6 dB	breiter
++				sehr breit

Alle präsentierten Stimuli sind in ihrer Lautheit angeglichen. Zunächst erfolgte der Abgleich durch den Versuchsleiter. Anschließend nahm ein KEMAR Kunstkopf das Ergebnis im Hörpunkt, sowie zusätzlich Varianten mit ±3 dB Signalpegel auf. Ein Lautheitsmodell (nonstationary Zwicker Funktion aus der Genesis loudness toolbox 1.2) analysierte nun diese drei Kunstkopfaufnahmen und lieferte jeweils einen Lautheitswert L_x zurück. Die drei ermittelten Lautheiten dienten als Stützpunkte zur linearen Interpolation der Lautheit über den anliegenden Signalpegel. Der jeweilige Punkt gleicher Lautheit zur Referenzmischung zeigt den entsprechenden Signalpegel für die Mixvarianten an.

Versuch

Die Versuchsteilnehmer saßen in der Mitte des Lautsprecherarrays. Nach einer kurzen Einweisung mit betreuter Trainingsrunde wählten sie jeweils aus zwei dargebotenen, 30 Sekunden langen Varianten die präferierten Variante aus. Der Versuch gliederte sich in mehrere Durchgänge, wobei in einem Durchgang jeweils alle Varianten einer Bearbeitungsmethode des WFS-Mixes und die drei Referenzmischungen für WFS, Stereo und Surround dargeboten wurden. Zwischen den einzelnen Durchgängen gab es jeweils eine etwa einminütige Pause. Randomisiert wurden alle Durchgänge als auch die Stimuli innerhalb eines Durchganges. Das gesamte Experiment umfasste 107 Paare, was eine Gesamtdauer von unter 45 Minuten ergab. An dem Versuch nahmen insgesamt 41 Personen ohne besondere Vorselektion teil. Diese stuften sich selbst als normalhörend ein und wurden finanziell für ihren Aufwand entschädigt. Die Studie entsprach den ethischen Richtlinien der Technischen Universität Berlin (RA-01-20140422).

Auswertung

Zwar bietet der Paarvergleich den Vorteil eines hoch sensitiven Tests, allerdings bedarf die Auswertung im vorliegenden Fall weiterer statistischer Methoden. Um eine kontinuierliche Skala aus den paarweise ermittelten Präferenzurteilen abzuleiten eignet sich das Bradley-Terry-Luce Modell [12]. Dieses fordert allerdings, dass die Versuchspersonen die getesteten Stimuli auf einer (gemeinsamen) perzeptiven Dimension unterscheiden und anordnen können [13, 14] – was im vorliegendem Versuch nicht ausreichend gegeben sein könnte. Um dies zu testen, liefert die Implementation [15] (in *R* 3.2.3) als Teststatistik den p-Wert zu *Goodness of Fit* des entwickelten (restriktierten) Modells zu dem perfekt angepassten (saturierten) Modell. Als hinreichend gilt ein p-Wert über 0.1 [15], welcher in allen gezeigten Fällen vorliegt. Durch Normierung repräsentiert die ermittelte *Präferenzwertung* einer Kondition die Wahrscheinlich-

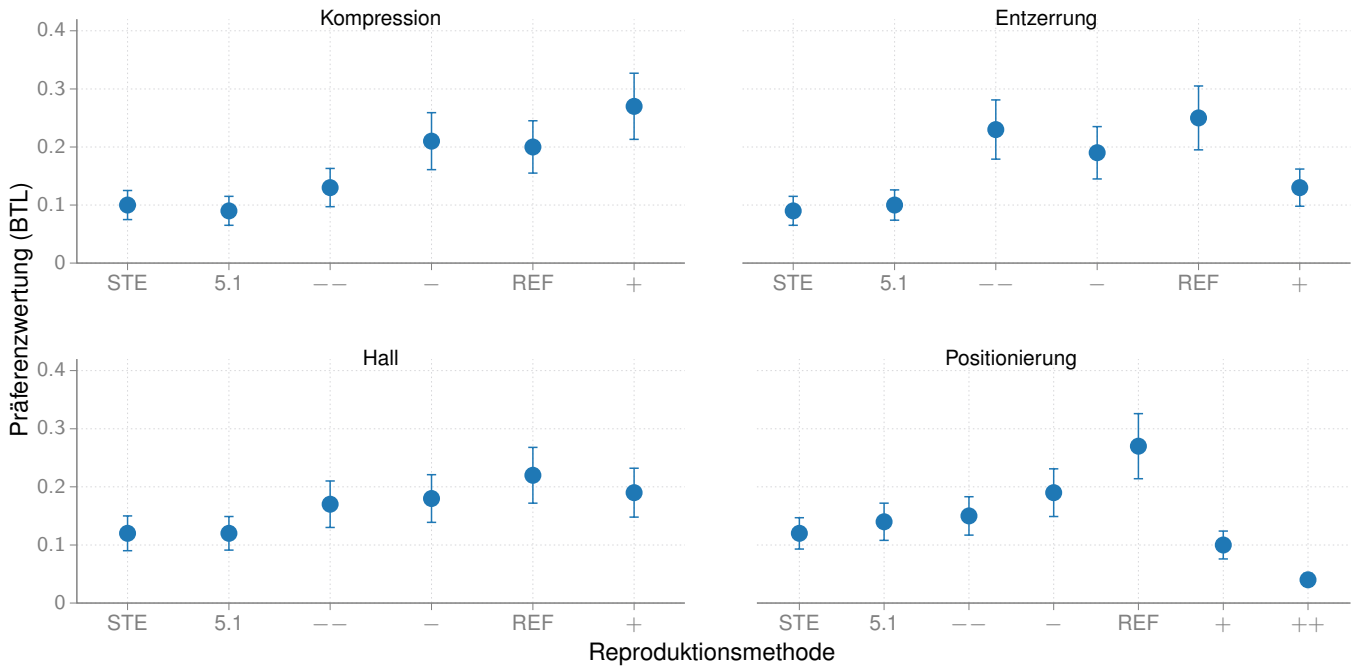

Abbildung 2: Präferenzwertungen des aus Paarvergleich abgeleiteten Bradley-Terry-Luce Modells. Wellenfeldsynthese (REF), Stereo (STE) und Surround (5.1) gaben die Referenzmischung wieder. Von dieser ausgehend wurden verschiedene Mixparameter skaliert.

keit, mit der diese Kondition innerhalb der analysierten Gruppe präferiert wird.

Ergebnisse

Versuchspersonen bewerteten in einem Paarvergleich verschiedene Mixvarianten bezüglich Präferenz, ausgehend von der Referenz (REF). Abbildung 2 zeigt alle durch das Bradley-Terry-Luce Modell ermittelten Präferenzwertungen im Vergleich.

Auffällig ist zunächst die recht niedrige Wertung beider stereophoner Systeme, obwohl diese die Referenzmischung wiedergegeben haben. Nur die sehr breite, seitlich ausgelenkte Positionierung von Vordergrundelementen in der WFS-Mischung unterbietet diesen Wert. Davon abgesehen liegen selbst die stärkeren Eingriffe auf Mixebene, wiedergegeben über Wellenfeldsynthese, grundsätzlich darüber. Weiterhin ist die WFS-Referenzmischung auch als höchstes bewertet, ausgenommen im Fall Kompression.

Die durch die Variation des Mixes hervorgerufene Änderung der observierten Präferenzwertung ist pro Bearbeitungsart unterschiedlich. Während diese für Entzerrung und Kompression in einem ähnlichen Bereich liegt, scheint sie für Nachhall schwächer und für Positionierung stärker ausgeprägt.

Diskussion

Die Varianten wiedergegeben durch Wellenfeldsynthese weisen gegenüber stereophoner Wiedergabe tendenziell einen erhöhten Präferenzwert auf. Dies lässt vermuten, dass der Wellenfeldsynthese inhärente Eigenschaften diese auslösen. Es scheinen allerdings nicht nur rein spektrale Eigenheiten relevant, wie sie beispielsweise durch *spatial aliasing* auftreten. Selbst ein komplettes Umgehen aller Filter im Mix lässt den Präferenzwert nicht unterhalb die stereophone Wiedergabe sinken. Größer scheint im vorliegendem Fall der Einfluss der Szenenbreite und der Lokalisierung zu sein. Genau heir spielt die Wellenfeldsynthese auch ihre Vorteile aus, da ein stabiles Positionieren auch neben und hinter der Hörposition möglich ist. Die leicht erhöhte Breite der Referenzszene, genauer der Vordergundelemente, wirkt sich positiv auf die Präferenz aus. Hier scheint in der Referenzmischung auch tatsächlich ein Optimum getroffen zu sein. Im Fall der Filterausprägung ist dieses schon weniger offensichtlich und das Optimum für Kompression scheint für dieses Stück leicht über der Wahl für die Referenzmischung zu liegen. Zwar stellt auch der gewählte Nachhallpegel der Referenzmischung die am stärksten präferierte Variante dar, der Präferenzunterschied zwischen den einzelnen Ausprägungen ist allerdings deutlich geringer. Dies lässt vermuten, dass sich künstlicher (stereophon erzeugter) Nachhall weniger kritisch verhält. Wahrscheinlich überwiegt der ohnehin sehr räumliche Eindruck der Präsentation mittels Wellenfeldsynthese. Erstaunlich groß fällt die Spanne der Präferenzbewertung für Variation der Kompression einzelner Signale aus. Den Effekt beschreiben viele, gerade ungeübte Hörer als sehr subtil – und doch zeigt sich im Paarvergleich, dass Kompression ein häufig unterschätzer Bearbeitungsschritt in Hinblick auf die Auswirkung auf Präferenz ist. Das zeitliche ist im Vergleich zu dem spektralen Verhalten meist weniger ausführlich untersucht.

Die Wahl der Parameter während der Tonmischung erweist sich als etwa so einflussreich wie ein Umschalten

des Wiedergabesystems zwischen Stereophonie und Wellenfeldsynthese. Während des Experiments veränderte sich jedoch immer nur eine einzelne Bearbeitungsmethode. Beachtet man, dass jede Methode weitestgehend unabhängig arbeitet, kann eine Kombination einen erheblich größeren Bereich abdecken. Der vermutlich einflussreichste Parameter, der Pegel eines Instrumentes im Mix, ist hier sogar noch unbetrachtet.

Interessanterweise fällt die Bewertung von Surround deutlich niedriger als in der Vorstudie [6] aus. Da in dieser allerdings nur die Systeme verglichen wurden, gilt es zu klären, inwieweit die häufigere Präsentation von Stimuli wiedergegeben durch Wellenfeldsynthese einen systematischen Einfluss auf die Präferenzbewertung aufweist.

Zusammenfassung

Wie erwartet zeigen sich auch in WFS signifikante Auswirkungen des Mixingprozesses. Dabei erzeugen allerdings die jeweiligen Bearbeitungen eine unterschiedliche Effektstärke auf Präferenz und auch eher subtil scheinende Eingriffe, wie unter Umständen Kompression, erweisen sich als sehr wirksam. Der Effekt auf die Präferenzbewertung hervorgerufen durch die getesteten Tonmischungen ist vergleichbar mit jenem hervorgerufen durch die unterschiedlichen Wiedergabesysteme.

Danksagung

Dieses Projekt ist gefördert durch EU FET Grant TWO!EARS, ICT-618075.

Alle Stimuli öffentlich online verfügbar;

`https://doi.org/10.5281/zenodo.61000`

Literatur

[1] A. Wilson and B. M. Fazenda, "Perception of audio quality in productions of popular music," *AES: Journal of the Audio Engineering Society*, vol. 64, no. 1-2, pp. 23–34, 2016.

[2] A. Wilson and B. Fazenda, "Relationship Between Hedonic Preference and Audio Quality in Tests of Music Production Quality," in *8th International Conference on Quality of Multimedia Experience (QoMEX)*, 2016, p. 6.

[3] H. Maempel, *Klanggestaltung und Popmusik: Eine experimentelle Untersuchung*, ser. Labor Synchron. Synchron, 2001.

[4] H. Wierstorf, "Perceptual Assessment of sound field synthesis," Ph.D. dissertation, Technische Universität Berlin, 2014.

[5] N. Zacharov, C. Pike, and F. Melchior, "Next generation audio system assessement using the multiple stimulus ideal profile method," *8th International Conference on Quality of Multimedia Experience (QoMEX)*, 2016.

[6] C. Hold, H. Wierstorf, and A. Raake, "The Difference Between Stereophony and Wave Field Synthesis in the Context of Popular Music," in *Proceedings of the 140th Convention of the Audio Engineering Society*, 2016, p. 8.

[7] A. Wilson and B. Fazenda, "101 Mixes: A statistical analysis of mix-variation in a dataset of multitrack music mixes," *Proceedings of the 139th Convention of the Audio Engineering Society*, pp. 1–10, 2015.

[8] B. De Man, K. McNally, and J. D. Reiss, "Perceptual Evaluation and Analysis of Reverberation in Multitrack Music Production," *AES: Journal of the Audio Engineering Society*, vol. 65, no. 1, 2017.

[9] M. Geier, J. Ahrens, and S. Spors, "The SoundScape Renderer : A Unified Spatial Rendering Methods," in *Proceedings of the 124th Convention of the Audio Engineering Society*, 2008.

[10] C. Hold, H. Wierstorf, and A. Raake, "Tonmischung für Stereophonie und Wellenfeldsynthese im Vergleich," *Fortschritte der Akustik - DAGA*, 2016.

[11] C. Hold, L. Nagel, H. Wierstorf, and A. Raake, "Positioning of Musical Foreground Parts in Surrounding Sound Stages," *AES Conference on Audio for Virtual and Augmented Reality*, pp. 1–7, 2016.

[12] R. Bradley and M. Terry, "Rank analysis of incomplete block designs: I. The method of paired comparisons," *Biometrika*, vol. 39, no. 3/4, pp. 324–345, 1952.

[13] M. G. Kendall and B. B. Smith, "On the method of paired comparisons." *Biometrika*, vol. 34, no. Pt 3-4, pp. 324–345, 1947.

[14] S. Choisel and F. Wickelmaier, "Evaluation of multichannel reproduced sound: scaling auditory attributes underlying listener preference." *The Journal of the Acoustical Society of America*, vol. 121, no. 1, pp. 388–400, 2007.

[15] F. Wickelmaier and C. Schmid, "A Matlab function to estimate choice model parameters from paired-comparison data." *Behavior research methods, instruments, & computers*, vol. 36, no. 1, pp. 29–40, 2004.