Tonmischung für Stereophonie und Wellenfeldsynthese im Vergleich

Christoph Hold[1], Hagen Wierstorf[2], Alexander Raake[2]
[1] *Assessment of IP-based Applications, TU Berlin, Deutschland, Email: Christoph.Hold@telekom.de*
[2] *Institut für Medientechnik, TU Ilmenau, Deutschland*

Einleitung

Die populäre Musikproduktion bedient bislang fast ausschließlich Wiedergabesysteme auf Basis der Stereophonie. Diese Verfahren positionieren Phantomschallquellen durch Anwendung von Pegel- und Laufzeitunterschieden zwischen den Lautsprechern. Die erzeugte Räumlichkeit, Verteilung und Klangfarbe ist jedoch stark abhängig von der Position der Zuhörer und dadurch schnell instabil. Eine andere Methode ist die (Re-)Produktion eines Schallfeldes mittels Wellenfeldsynthese. Die hierbei synthetisierten virtuellen Schallquellen sind praktisch frei im Raum positionierbar und der Sweetspot dehnt sich theoretisch auf den kompletten Zuhörerraum aus.
Beide Systeme unterscheiden sich in grundlegenden Aspekten und zeigen jeweils sehr spezielle Eigenschaften bei der Musikwiedergabe, die es bereits in der Produktionsphase zu bedenken gilt. So werden zum Beispiel in der Popularmusik die aufgenommenen Einzel-Signale stark bearbeitet um den gewünschten Klangcharakter der finalen Mischung zu erzielen. Dafür sollte das grundsätzliche Verhalten des Wiedergabesystems bereits im Vorraus bekannt sein. Entsprechend kann während der Musikproduktion dieser Charakter beachtet und die Tonmischung auf das Wiedergabesystem optimiert werden. Worin die Unterschiede für Stereophonie und Wellenfeldsynthese liegen und wie sich diese äußern, wird am Beispiel erläutert und möglicherweise limitierende Faktoren diskutiert. Anschließend werden praxisnahe Mischkonzepte sowie Klangbearbeitungen vorgeschlagen. Dabei geht es insbesondere um die Erschließung der Wellenfeldsynthese und deren Möglichkeiten in der Popularmusik.

Motivation

Als übergeordnetes Ziel gilt es, einen wissenschaftlichen Vergleich der Wiedergabesysteme zweikanal Stereophonie *(Stereo)*, 5.1 Surround Stereophonie *(Surround)* und Wellenfeldsynthese *(WFS)* zu ermöglichen. Dazu muss sichergestellt sein, dass die observierten Unterschiede durch die Systeme selbst und nicht etwa durch die abgespielten Tonmischungen verursacht sind. Folglich kommt die Forderung nach Vergleichbarkeit der Tonmischungen auf. Dafür ist es notwendig, eigenständige Mischungen herzustellen, optimiert an die Anforderungen des jeweiligen Wiedergabeverfahrens. Dabei meint eigenständig, sich nicht auf up- oder down-mixing Algorithmen zu verlassen, da diese nicht zwingend den Anforderungen genügen. Es gilt auf der anderen Seite jedoch die Möglichkeiten der Systeme aufzuzeigen. Besonders die räumlichen Eigenschaften stellen einen zentralen Bestandteil der Systemeigenschaften dar, welche bei unzureichender Tonmischung verborgen bleiben. Diese beiden Facetten können im Widerspruch stehen, eine praktische Umsetzung muss dennoch beiden folgen.
Ferner soll ein nahtloses Umschalten zwischen den Systemen einen direkten Vergleich erleichtern, was weitere Anforderungen an die Realisierung stellt.

Es gilt zu prüfen, wie die Systeme beim Abspielen komplexer Klangszenen reagieren und ob es möglich ist, moderne Mischtechniken anzuwenden und zu integrieren. Insbesondere stellt der Objektcharakter der Wellenfeldsynthese eine Herausforderung dar. Alle gängigen Konzepte und speziellere, über die letzten Jahrzehnte optimierten, Audiobearbeitungen folgen dem kanalbasierten Ansatz der Stereophonie und sind somit nicht direkt auf den Objektcharakter der Schallfeldsynthese übertragbar.

Setup

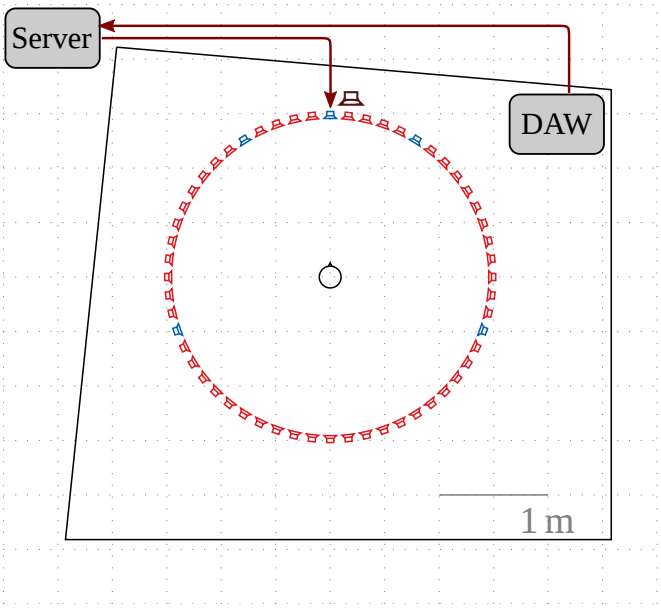


Abbildung 1: Raumskizze mit der verwendeten Lautsprecheranordnung. Die Wellenfeldsynthese spielt auf alle 56 Lautsprecher aus, in blau sind die entsprechenden Lautsprecher der stereophonen Systeme markiert. Ein Subwoofer befindet sich im vorderen Bereich der Anordung und die Hörposition befindet sich stets in der Mitte.

Das gewählte Setup nach Abbildung 1 enthält neben einer *Digital Audio Workstation (DAW)*, einen Audioserver auf dem der *SoundScape Renderer (SSR)* [1] läuft und sämtliches Audiorouting inklusive Bassmanagement geschieht. Dieser spielt die Audiosignale an eine kreisförmige Lautsprecheranordnung aus. Die Trennung

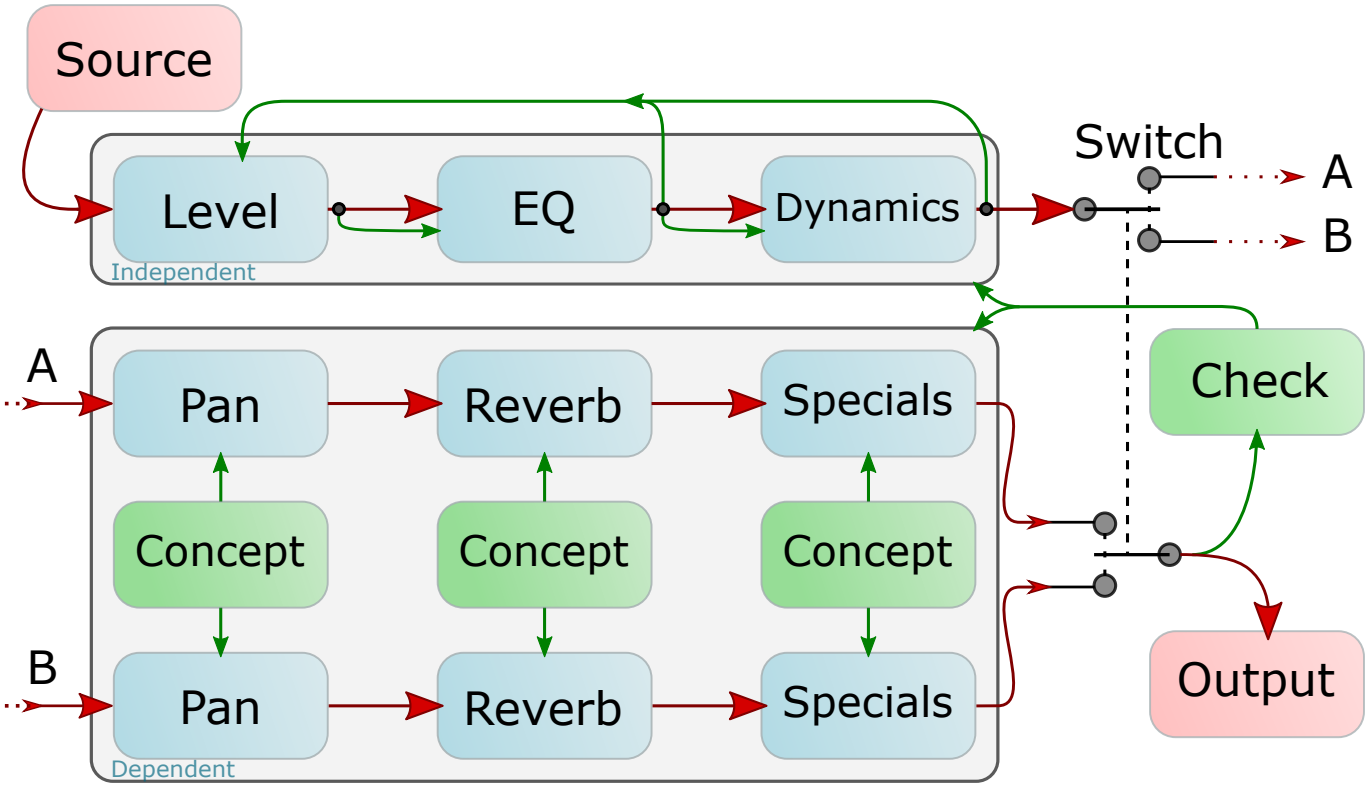


Abbildung 2: Blockschaltbild zur Veranschaulichung des vorgeschlagenen Tonmischungsprozesses, vom Rohmaterial zum Ausgangssignal. In rot sind Audiosignale, in blau Bearbeitungsstufen und in grün wichtige Abhängigkeiten sowie Einflüsse des Tonschaffenden dargestellt.

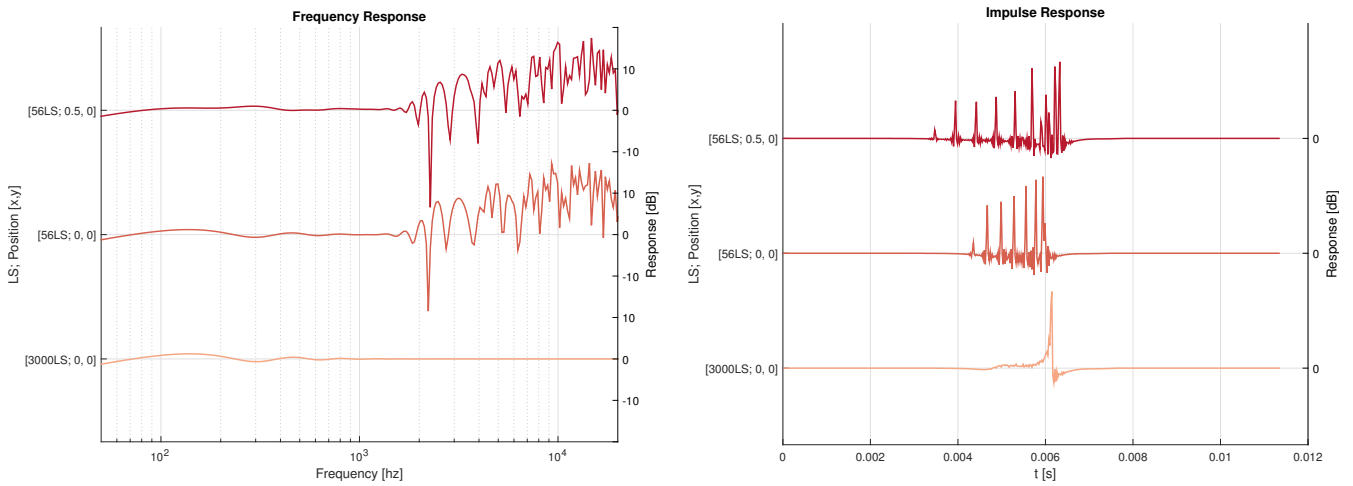
von Schallfeldberechnung durch den SSR und Audiobearbeitung erlaubt, dass sich die DAW für die Audiobearbeitung im Abhörraum befinden kann. Der SSR führt die Berechnung der auszugebenden Signale für die Wellenfeldsynthese durch, wobei dessen Quellcode angepasst wurde um die enthaltene räumlich abhängige Amplitudenabschwächung zu kompensieren. Die Positionierung von Quellen führt dadurch nicht mehr zu einer Änderung des Quellenpegels, was sich für die Anwendung innerhalb einer komplexeren Klangmischung als praktisch herausstellt. Der Durchmesser der Lautsprecheranordnung im Abhörraum beträgt 3 m und enthält 56 Lautsprecher, von welchen der Wellenfeldsynthese alle 56 zur Verfügung stehen, während Surround und Stereo nur die entsprechenden Lautsprecher anspricht, wie in Abbildung 1 skizziert. Allen Systemen steht ein Subwoofer zur Verfügung, welcher als Tiefton-Quelle im SSR initiiert und bei den stereophonen Systemen durch Bassmanagement involviert ist. Daraus folgt, dass die selben Lautsprecher sowohl ein natives Stereo, Surround nach ITU-R [2], als auch die Wellenfeldsynthese realisieren.
Unabhängig der untersuchten Systeme steht ein Referenzsystem für die Tonmischungen zur Verfügung, welches aus zwei hochwertigen Studiolautsprechern und Subwoofer besteht.

Mixing

Der Mischprozess für Pop-Musik beinhaltet Klangbearbeitungen, welche weitestgehend unabhängig des Wiedergabesystems einzustellen sind und den Grundklangcharakter des Musikstückes definieren. Auf der anderen Seite sollen Prozesse wie die Positionierung in der Klangszene oder die Einbindung von Hall auf die jeweiligen Wiedergabesysteme angepasst sein, um diese möglichst optimal einzusetzen. Abbildung 2 illustriert den vorgeschlagenen Mischprozess für mehrere parallele Systeme im Sinne der geforderten Ziele. Die roten Pfeile markieren den Audio-Fluss, blaue Blöcke die Audio-Bearbeitungsstufen und grün zeigt wichtige Abhängigkeiten sowie Einflüsse des Tonschaffenden. Dabei umfasst der obere graue Block alle Stufen unabhängig des Wiedergabesystems und der untere graue Block alle systemabhängigen Prozesse.

Systemunabhängige Klangbearbeitung

Der Großteil der Klangbearbeitung findet auf dem bereits beschriebenen Referenzsystem statt, unabhängig der untersuchten Wiedergabesysteme. Diese umfasst Pegeljustierungen, sowie den Einsatz von Equalizern und Dynamikbearbeitungsstufen. Das Rohmaterial besteht aus Multitrackaufnahmen der Popstücke, d. h. jedes Instrument ist einzeln und häufig mit mehreren Mikrofonen und Durchgängen aufgenommen. Zuerst gilt es diese vielen Tonspuren in ihren Pegeln abzustimmen. Anschließend sorgen Equalizer für eine tonale Balance und helfen einzelne Instrumente zu separieren. Da akustische Instrumente und insbesondere menschliche Stimmen drastisch in ihren Pegeln variieren können, manipulieren Dynamikprozessoren deren Dynamikumfang. Gerade Kompressoren stellen einen entscheidenden Teil des weit verbreite-



(a) Frequenzgang

(b) Impulsantwort fokussierte Quelle

Abbildung 3: Simulation von Frequenzgang und Impulsverhalten. Der jeweils unterste Verlauf zeigt das optimale Verhalten, der mittlere das mit 56 Lautsprechern ausgespielte im Mittelpunkt der Lautsprecheranordnung, der oberste den um 50 cm seitlich verschobenen Abhörpunkt. Es zeigen sich die spektralen und temporalen sowie ortsabhängigen Artefakte der Wellenfeldsynthese.

ten und gewünschten modernen Klangcharakters her.
Dabei beeinflussen sich die einzelnen Stufen gegenseitig. Erfolgt beispielsweise eine deutliche Anhebung der Höhen mittels Equalizer, verringert sich häufig der zu wählende Pegel des Instrumentes. Die Einschränkung der Dynamik wirkt sich offensichtlich auf den Pegel eines Signals aus, allerdings verändert ein Kompressor auch häufig dessen Klangcharakter.
Passende Einstellungen auf dieser ersten Ebene resultieren in einer guten Gesamtbalance der Mischung, unabhängig des wiedergebenden Systems. Somit steht der Grundcharakter fest und es folgt die Verteilung der Mischungen auf die zu untersuchenden Systeme.

Systemabhängige Klangbearbeitung

Die systemabhängig eingestellten Klangbearbeitungen bestehen aus Panning/Positionierung, Hall/Delay, sowie einigen speziellen Effekten wie beispielsweise modulationsbasierenden Effekten. Um Vergleichbarkeit sicherzustellen folgen diese Schritte alle einem gemeinsamen Konzept, welches grundsätzlich unterschiedliche Ergebnisse zwischen den Systemen unterbindet. Dieses entspricht einer eher konservativen Herangehensweise an die einzelnen Stufen, d. h. alle Hauptkomponenten der Musikstücke befinden sich in der frontalen Klangebene und bewegte Quellen kommen nicht vor. Insbesondere Hall und Delays bieten sich auch für die hintere Klangebene an. Alle weiteren spezielleren Effekte sind zwischen den Systemen so ähnlich wie möglich eingebunden.
Es folgt abschließend das Ausspielen auf die jeweiligen Lautsprecher sowie das Gegenhören. Dieser Kontrollschritt wirkt auf alle vorhandenen Bearbeitungsstufen zurück und schließt damit einen Kreis, welches mit dem häufig sehr zeitaufwändigen Mischprozess korrespondiert.
Falls eine geeignete Bearbeitung nur auf stereophonen Spuren erfolgen kann, so werden diese beiden Kanäle mittels *Virtual Panning Spots* [3] in die objektbasierte WFS Umgebung eingebunden. Solche Stereobusbearbeitung umfasst besonders *Bus-Kompression*, welche das Erzeugen eines sehr dichten Klangs ermöglicht. Die Kombination beider getrennt zu betrachtender Wiedergabeverfahren kann dadurch durchaus ansprechende Resultate erzielen.

Systembedingte Unterschiede

In erster Linie nimmt die Räumlichkeit hin zu den höher kanaligen Systemen zu. Dabei erscheint der Schritt von Stereo auf Surround größer als von Surround auf WFS. Die virtuellen Schallquellen der WFS lassen sich im gesamten Hörbereich stabil positionieren, während seitliche Phantomschallquellen eher unstabil wirken. Phantomschallquellen in der hinteren Ebene sind nicht zuverlässig möglich.

Die einzelnen Elemente komplexerer Klangszenen klingen in der WFS separierter, schärfer Umrissen, ein Interagieren und Verschmelzen der einzelnen Elemente zu einer gemeinsamen Klangszene ist allerdings schwieriger zu erreichen. Dagegen erscheint der Grundcharakter der stereophonen Systeme etwas undurchsichtiger und schwammiger, allerdings besser verwoben.

Neben der Veränderung des Frequenzgangs beeinflussen die *spatial aliasing* Artefakte der WFS das zeitliche Verhalten, wie in Abbildung 3 für den simulierten Fall erkennbar. Die ausgespielten Ergebnisse variieren dadurch zwischen Stereophonie und Schallfeldsynthese, folglich gilt es diese Unterschiede im Mischprozess zu berücksichtigen. Gerade Bearbeitungen im höheren Frequenzbereich ab 1.5 kHz unterschieden sich im Resultat. Impulse verhalten sich ebenfalls in der WFS anders als in der gewohnten stereophonen Umgebung und das Positionieren von impulshaftem Audiomaterial als fokussierte Quelle gilt es zu vermeiden, da hier die temporalen Artefakte besonders ausgeprägt sind und zusätzlich dem

Nutzsignal voreilen (vgl. Abbildung 3).

Resultate

Die Wellenfeldsynthese kann mit geeigneter Tonmischung das Hörerlebnis gegenüber stereophonen Systemen verbessern.

Es stellt sich heraus, dass die bereits betrachteten Artefakte [4] der Wellenfeldsynthese zwar vorhanden, die Auswirkungen dieser allerdings weniger drastisch und beeinträchtigend als erwartet ausfallen. Aus Sicht des Tonschaffenden erweist sich die Wellenfeldsynthese als sehr ansprechendes Wiedergabesystem. Die Tonmischung selbst bleibt das entscheidenste Merkmal im Vergleich. Mit den erstellten, die systeminduzierten Unterschiede aufzeigenden, Musikmischungen ist es möglich Hörversuche durchzuführen. Diese sollen die Frage nach dem bevorzugtem Wiedergabeverfahren beantworten und überdies klären, ob die Wellenfeldsynthese damit als ein geeignetes Wiedergabesystem im Popmusik Kontext zählt.

Danksagung

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

Literatur

[1] Geier, M., Ahrens, J., und Spors, S., "The SoundScape Renderer: A Unified Spatial Audio Reproduction Framework for Arbitrary Rendering Methods," in Proceedings of the 124th Convention of the Audio Engineering Society, Paper 7330, 2008.

[2] ITU-R, "Multichannel stereophonic sound system with and without accompanying picture BS Series," International Telcommunication Union Reccomendations BS. 775-3, 3, p. 23, 2012.

[3] Theile, G., Wittek, H., und Reisinger, M., "Potential wavefield synthesis applications in the multichannel stereophonic world," in Proceedings of the 24th Conference of the Audio Engineering Society, Paper 35, 2003.

[4] Wierstorf, H., Hohnerlein, C., Spors, S., und Raake, A., "Coloration in Wave Field Synthesis," in Proceedings of the 55th International Conference of the Audio Engineering Society, Paper 5–3, 2014.