

Objektorientierte Audioumgebungen

Stephan Stahl

Hochschule der Medien

Stuttgart, Deutschland

stephan.stahl@gmail.com

Zusammenfassung

Die vorliegende Publikation¹ zeigt auf, wie eine maximale Kompatibilität zwischen Audioinhalten und existierenden Wiedergabeverfahren realisiert und inwiefern eine Aufwärtskompatibilität zu zukünftigen Wiedergabesystemen gewahrt werden kann. Dazu werden die Arbeitsschritte Produktion, Speicherung und Wiedergabe als Teilbereiche einer Audioumgebung definiert. An jeden Teilbereich werden Anforderungen gestellt, die durch einen objektorientierten Ansatz erfüllt werden können. Weiterhin soll die Notwendigkeit objektorientierter Lösungsansätze in Bezug auf aktuelle und zukünftige Produktionsweisen und Wiedergabeverfahren aufgezeigt werden.

1 Einleitung

Die konstante Weiterentwicklung neuer Audio-Wiedergabesysteme ermöglicht eine zunehmende qualitative Verbesserung der Ton-Produktion, der Ton-Reproduktion und somit der Klangwahrnehmung. Dabei werden entlang der neuen Wiedergabetechnologien auch neue Anforderungen an die Audioumgebungen gestellt: Angefangen bei Mono über Stereo und 5.1-Surround bis hin zu aktuellen dreidimensionalen Surround-Wiedergabeverfahren, haben sich in den letzten Jahrzehnten immer komplexer werdende Systeme entwickelt, die nicht zwingend kompatibel zueinander sind und stets eine Anpassung der Arbeitsprozesse, der Tonträgerformate und der technischen Peripherie erfordern.

¹ basiert auf der Master-Thesis des Autors „Objektorientierte Audioumgebungen für Surround-Sound“, vorgelegt im WS 2012/13 an der Hochschule der Medien, Stuttgart (Stahl 2013)

Unter Verwendung eines objektorientierten Ansatzes können diese Anpassungen vermieden werden: Die Produktion mehrkanaliger Inhalte kann beschleunigt, die Distribution von mehrkanaligen Inhalten erleichtert und die Wiedergabesituation (und somit das Klangerlebnis) beim Konsumenten verbessert werden.

2 Objektorientierung

Die Objektorientierung bezeichnet die Veranschaulichung komplexer Systeme anhand kooperierender und zueinander in Verbindung stehender Objekte und findet zunehmend in der Softwareentwicklung Verwendung. Dabei werden explizit Eigenschaften und Funktionen eines Objekts zusammengefasst und architektonische Baupläne konstruiert. Diese sind in der Regel verständlicher und einfacher auf reale Beispiele zu übertragen als in der funktionsorientierten Programmierung. Ein kurzes Beispiel soll die primäre Grundidee erläutern.

Jeder Gegenstand, jedes Medium bzw. jede konkrete Form lässt sich als Objekt bezeichnen respektive darstellen. So kann ein Stuhl ein Objekt sein, ein Schreibtisch, ein Lautsprecher, ein Mensch und so weiter. Jedes Objekt weist dabei eigene Eigenschaften und Verhaltensweisen auf: Ein Mensch kann beispielsweise gehen, stehen, sitzen, sprechen (Verhaltensweisen, genauer: Funktionen). Er kann groß, klein, dick, dünn sein (Eigenschaften, genauer: Attribute). Eine Schreibtischlampe kann dagegen nicht gehen und auch nicht sprechen – sie leuchtet (Funktion) und hat meist eine konkrete Form und Farbe (Attribute bzw. Eigenschaften). Folglich kann eine Klassifizierung des Mediums über die Funktionen und Attribute geschehen und zu einer nächstgrößeren Einheit – dem Objekt – zusammengefasst werden.

Die Theorie der Objektorientierung lässt sich nun auf Audioumgebungen übertragen. Um die Zusammenhänge genauer zu verstehen, ist jedoch ein Einblick in die bisherigen Produktionsverfahren notwendig.

3 Objektorientierte Audioproduktion

Ob Mono-, Stereo- oder 5.1-Surround-Produktion – das Endformat bzw. die Kanalanzahl entspricht immer genau der Lautsprecheranzahl des jeweiligen (Wiedergabe-) Verfahrens. Bei der Produktion werden daher Lautsprecher-signale generiert, die sich einzig auf die starre Anordnung der Wiedergabepunkte übertragen lassen. Eine solche Produktionsweise ist in Bezug auf einen

vorher definierten (und optimierten) Lautsprecheraufbau zwar zielführend, führt jedoch zu Problemen, wenn die Aufstellungsempfehlung nicht eingehalten werden kann oder sich die Lautsprecheranzahl ändert.

Die Lautsprechersignale werden dabei wie folgt erzeugt: Bei Stereo wird eine Schallquelle anteilig auf das Lautsprecherpaar gegeben – je nachdem, ob sie sich weiter links oder weiter rechts befinden soll, ändert sich der Pegel der Signalquelle auf den Lautsprechern.² Bei 5.1-Surround ist die Vorgehensweise ähnlich: Die Schallquellen werden auf einem zweidimensionalen Feld zwischen Lautsprechern positioniert; die Lautsprecher können dabei paarweise Phantomschallquellen abbilden. Vereinfacht ausgedrückt: Links-Center, Center-Rechts und LS-RS bilden ein Stereopaar, ebenso L-LS und R-RS. Wird eine Schallquelle also auf dem zweidimensionalen Feld positioniert, werden die Signalanteile für die einzelnen Lautsprecher entsprechend berechnet.

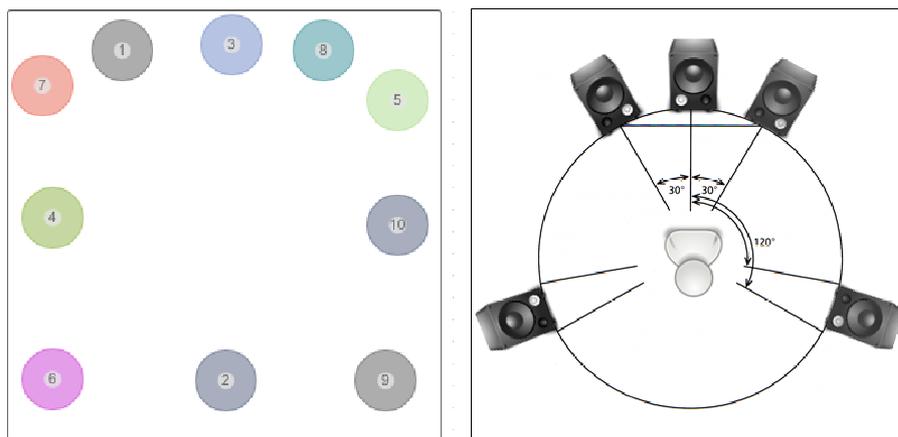


Abb. 1 links: Schallquellen mit Ihrer Position im abstrahierten virtuellen Raum; rechts: „realer“ Raum mit Lautsprechern

Die bei der Produktion durchgeführte Positionierung der Schallquellen im zweidimensionalen Feld (bzw. Raum) kann nun für ein objektorientiertes Modell genutzt werden. Jede Schallquelle besitzt an dieser Stelle bereits eine Position in einem virtuellen, gedachten Raum. Diese Rauminformationen, bzw. die Koordinaten der Schallquelle im Raum, stellen die erste Ausprägung der objektorientierten Audioumgebungen dar: Eine Schallquelle (Ob-

² Die Positionierung der Phantomschallquellen kann auch über Laufzeitdifferenzen erfolgen (vgl. dazu Stahl 2013: 21 ff.).

jekt) hat die raumbezogenen Koordinaten (Attribute) in einem virtuellen Raum (Objekt). Die anschließende Berechnung der Lautsprecher-signale und die darauffolgende Wiedergabe stellt lediglich die Manifestierung der Schallquellen an einem Ort im „realen“ Raum dar. Abbildung 1 verdeutlicht diesen Ansatz.

Der virtuelle Raum ist bei der konventionellen Produktionsweise jedoch eng an die Lautsprecheraufstellung gekoppelt: Wird eine Schallquelle virtuell positioniert, so ist sie unmittelbar an einen (bzw. als Phantomschallquelle an zwei) Lautsprecher gebunden. Verschiebt man die Lautsprecheraufstellung, so verschiebt man ebenfalls die (Phantom-) Schallquelle. Da aber ohnehin jede Schallquelle im virtuellen Raum platziert wird – im Beispiel 5.1 zwei-dimensional – können die dabei entstandenen Meta-Informationen (bzw. die Attribute) der Schallquelle für ein angepasstes Wiedergabesystem genutzt werden: Die (virtuelle) Rauminformation gibt bereits eine Aussage über die Position der Schallquelle.

Wird diese Rauminformation separat aufgezeichnet, kann das Lautsprecher-signal anhand dieser Meta-Informationen und den dazugehörigen Audioinhalten – in Abhängigkeit zu der Lautsprecheraufstellung im realen Raum des Wiedergabesystems – (re-)generiert werden. Die Positionierung der Schallquelle im virtuellen Raum erfolgt also nach wie vor auf Produktionsseite, die Berechnung der Lautsprecher-signale im realen Raum des Zuhörers jedoch erst bei der Wiedergabe – sie werden erst dann berechnet, wenn die Lautsprecheraufstellung bekannt ist. Somit lassen sich alternative Lautsprecheranzahlen und -positionen realisieren, da der Anteil einer Schallquelle für jeden Lautsprecher dynamisch generiert wird.

Im Zuge der Objektorientierung seien erneut die Objekte zusammengefasst, die für das Verständnis des Modells notwendig sind:

- der virtuelle Raum,
- die Schallquellen im virtuellen Raum mit den Parametern x -, y - (und z -) Position,
- der reale Raum und
- die Lautsprecher im realen Raum mit den Parametern x -, y - (und z -) Position.

Der objektorientierte Ansatz ermöglicht die Trennung zwischen virtuellem Raum (und somit den Positionen der virtuellen Schallquellen) und dem realen Raum und seinem Lautsprecheraufbau. Die Beschaffenheit des realen Raumes und der damit verbundene Lautsprecheraufbau sind für die Positionierung der Schallquellen im virtuellen Raum unerheblich. Anders ausgedrückt wird die Beziehung zwischen Schallquelle und Lautsprecher ausge-

hebelt: Es gibt keine n:2- (Stereo) oder n:5.1- (Surround) Beziehung mehr – das Schallereignis jeder Schallquelle kann dynamisch auf die jeweiligen Lautsprecher, abhängig von ihrer Position, berechnet werden.

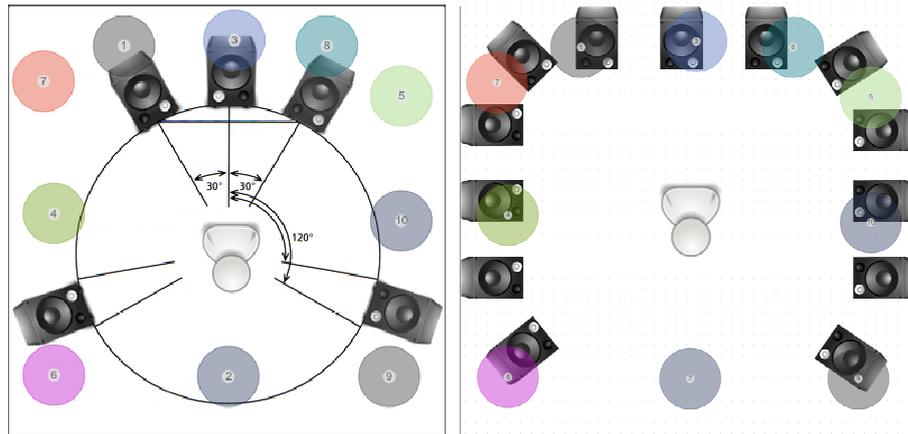


Abb. 2 Projektion des virtuellen Raums auf den realen Raum
links: fünf Lautsprecher; rechts: exemplarisch 13 Lautsprecher

4 Objektorientierte Datenspeicherung

Um die in der Produktion entstandenen Audioinhalte auch auf anderen (Konsumenten-) Systemen wiederzugeben, ist zwangsläufig ein Speicherformat notwendig. Unabhängig davon, ob zur Distribution ein haptisches oder rein digitales Trägermedium verwendet wird, müssen die produzierten Audioinhalte in einer solchen Form gespeichert werden, dass sie bei der Wiedergabe reproduziert werden können. Die Anforderungen an ein solches Speicherformat sind – rein in der Theorie – relativ simpel: Jeder Kanal muss zusammen mit seinen Panoramainformationen in ein Containerformat überführt werden.

Seit den 80er-Jahren haben sich etliche Dateiformate entwickelt und einige der Audioformate sowohl in Studioumgebungen als auch beim Endverbraucher etabliert. Abbildung 3 liefert einen Überblick über die Evolution ausgewählter Audio-Dateiformate.

Für die Fusion der Audioinhalte und Panoramainformation in einer Datei existieren dahingehend verschiedene Strategien: Einerseits könnte ein bestehendes Containerformat zur Vereinigung verwendet werden, andererseits könnten die Panoramainformationen auch in ein bestehendes (mehrkanales)

Dateiformat eingeflochten werden. Für die objektorientierte Datenspeicherung geeignete Containerformate sind bspw. MPEG-4 ALS, MPEG-4 SA und MPEG-4 BIFS, da innerhalb der Container eine große Anzahl Audioinhalte (bzw. Kanäle) samt zusätzlicher Meta-Information (bzw. Panorama-Information) gespeichert werden kann (vgl. Stahl 2013: 75 ff.).

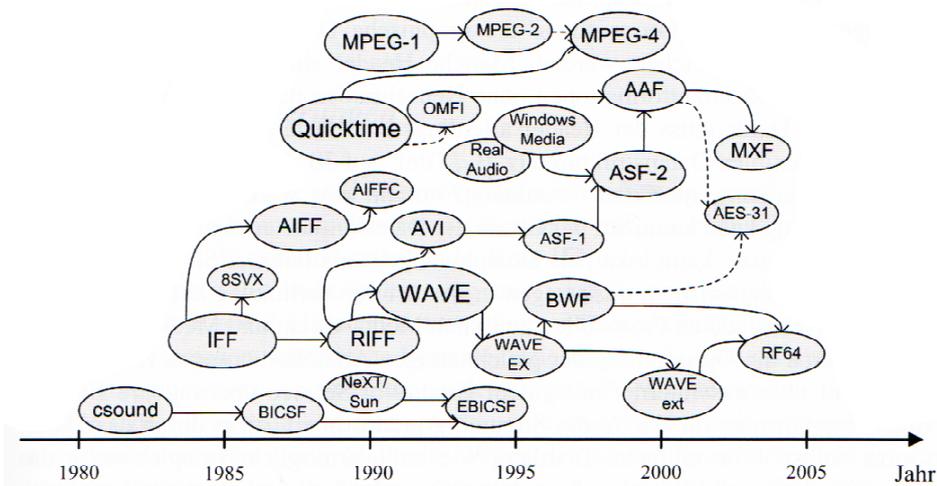


Abb. 3 Evolution der Audio-Dateiformate (Auszug) (Quelle: Petermichel 2008: 691)

5 Objektorientierte Wiedergabeverfahren

Um die in der objektorientierten Datenspeicherung gekapselten Daten wieder aufzubereiten – oder auch, um der objektorientierten Audioproduktion eine unmittelbare Abhörmöglichkeit bereitzustellen –, müssen die Audioinhalte wieder mit den Panoramadaten zu einer sogenannten Audioszene zusammengeführt werden. Das Darstellen einer Audioszene wird im folgenden Rendering genannt, das komplette Wiedergabesystem wird als Renderer bezeichnet. Zwischen den Schallquellen bzw. den Produktionskanälen und den Lautsprechersignalen entsteht dabei eine $n:m$ -Beziehung: Jede Schallquelle kann durch eine beliebige Anzahl Lautsprecher repräsentiert werden und jeder Lautsprecher kann eine beliebige Anzahl an Kanälen wiedergeben. Dies lässt sich anhand Abbildung 4 verdeutlichen.

Für die Berechnung der Lautsprechersignale ist es dabei zwingend notwendig, dass der Renderer die tatsächliche Raumposition der Lautsprecher kennt. Nur so können die Kanäle auch entlang ihrer Panoramainformation im (realen) Raum positioniert werden.

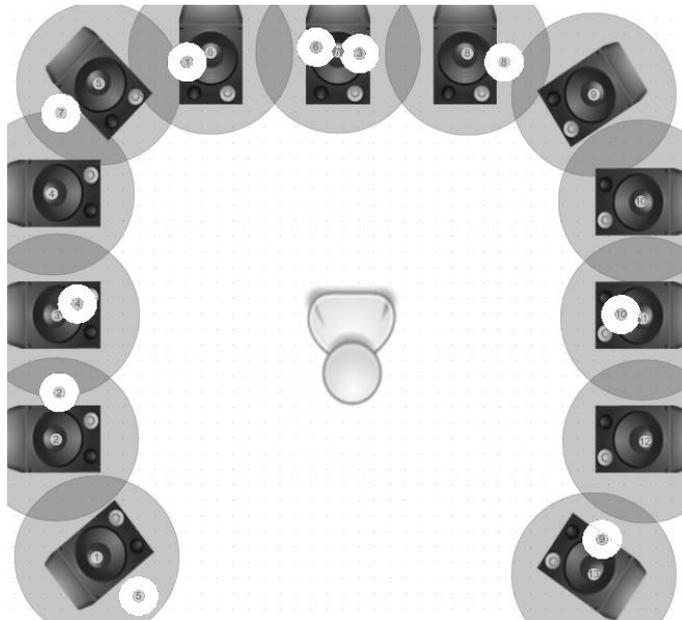


Abb. 4 n:m-Beziehung zwischen Schallquellen (kleine weiße Kreise) und Lautsprecher

Die Berechnung der Signale kann – je nach Aufstellung der Lautsprecher – über verschiedene Verfahren erfolgen. Exemplarisch seien an dieser Stelle Vector Base Amplitude Panning (vgl. Pulkki 1997), Ambisonics (vgl. Slavik/Weinzierl 2008) oder auch die Wellenfeldsynthese (vgl. Theile 2008) aufgeführt (vgl. dazu auch Stahl 2013: 83 ff.).

6 Vorteile der objektorientierten Audioumgebung

6.1 Produktion

Durch die Entkopplung der Produktion vom letztendlichen Lautsprecheraufbau können endformatunabhängige Produktionen durchgeführt werden. Weiterhin können bereits während der Produktion verschiedene Lautsprecher-Setups evaluiert werden – sofern ein entsprechender Renderer zur Wiedergabe bereitsteht. Denkbar wäre bspw. auf einem zwei- oder sogar dreidimensionalen Wiedergabesystem zu produzieren und – durch Ändern des Renderer-Systems – unmittelbar zwischen verschiedenen (Abhör-) Verfahren zu wechseln. So kann ein unmittelbarer Vergleich zwischen 5.1, Auro 3D, 22.2 oder sogar Verfahren wie der Wellenfeldsynthese realisiert werden, ohne

erneut Einfluss auf die Produktion und das Routing der Lautsprecherkanäle zu nehmen. Weiterhin können nicht-ideale Lautsprechersituationen untersucht werden, um auch explizit auf Konsumenten-Bedingungen einzugehen.

6.2 *Distribution*

Durch die objektorientierte Audioproduktion können zudem verschiedene Panorama-Versionen – sprich verschiedene Varianten der Positionierungen der Schallquellen – angefertigt und bei der Distribution integriert werden. Da die Panoramadaten getrennt von den Audioinhalten aufgezeichnet werden und im Hinblick auf Speicherbedarf unbedenklich sind, können verschiedene Varianten – jeweils angepasst auf spezielle Abhörbedingungen – angefertigt werden, ohne die Audioinhalte bearbeiten zu müssen. Somit stehen den Produzenten verschiedene gestalterische Möglichkeiten offen.

6.3 *Rezeption*

Umkehrt könnte der Konsument bei der Wiedergabe zwischen verschiedenen Panorama-Varianten wählen und so eventuell neue Eindrücke über die wiedergegebenen Audioinhalte gewinnen. Beispielsweise könnte eine Version mit statischen Positionierungen von Instrumenten einer dynamischen bewegten Positionierung gegenübergestellt werden. Weiterhin könnte der Zuhörer bei einer Musikwiedergabe exemplarisch einzelne Instrumente gesondert betrachten – bspw. durch Lautstärke-Absenkung oder -Anhebung – oder sie sogar anders im Raum positionieren und somit eine alternative Darstellungsmethode des Gesamtwerks schaffen. Es wäre zudem möglich, alle Schallquellen – bzw. im Falle eines Musikstücks das gesamte Ensemble – zu verschieben und damit die eigene Position im Raum zu verändern – obwohl die Hörposition dabei dieselbe bleibt. Das bisher passive Musikhören könnte durch solche Möglichkeiten durchaus zu einer aktiven, fast spielerischen Herangehensweise an Musik bzw. Audioinhalte werden.

6.4 *Wiedergabe (konsumentenseitig)*

Die Probleme der Verschiebung der Phantomschallquellen, wenn die Lautsprecher nicht gemäß Aufstellungsempfehlung (bspw. ITU-R BS. 775) angeordnet sind, wird durch den Renderer ausgehebelt: Wird die Lautsprecherposition im Raum verändert, kann der Renderer explizit auf die neue Position reagieren und somit die Phantomschallquellen berechnen. Dies kann bereits

einige Darstellungsprobleme bei unausgewogenen Abhörräumlichkeiten und damit verbundenen Lautsprecherpositionen lösen. Dieses Verfahren ermöglicht zudem die Annäherung an weitere Surround-Standards, ohne das Trägermedium zu ändern: Beispielsweise kann ein dynamischer 7.1-Aufbau realisiert werden, indem zwischen L und LS bzw. R und RS weitere Lautsprecher positioniert werden – die Signale für diese Lautsprecher können vom Renderer unmittelbar berechnet werden.

Sofern bei der objektorientierten Audioproduktion auch dreidimensionale Panoramadaten aufgezeichnet wurden, können zudem Höhenlautsprecher angebracht und direkt angesteuert werden. Die Lautsprecheranzahl kann somit dynamisch an die eigenen Räumlichkeiten und Bedürfnisse angepasst werden. Ein System kann daher zunächst mit wenigen Lautsprechern (bspw. angefangen mit einem einzigen Lautsprecherpaar) zu einem großen Aufbau erweitert werden. Das Hinzufügen neuer Lautsprecher erfordert dabei keine Änderung der Produktionsweise, der Dateispeicherung oder der Wiedergabe, sondern lediglich eine Erweiterung der Lautsprecheranzahl.

7 Perspektiven

Die vorhergehenden Kapitel zeigen, dass objektorientierte Audioumgebungen sowohl theoretisch als auch praktisch realisierbar sind. Die dazu notwendige Theorie und die ebenfalls dazu benötigte Technik steht bereits seit Jahren zur Verfügung. Da für einen objektorientierten Ansatz allerdings Änderungen in jedem Teilbereich vorgenommen werden müssen – sowohl bei der Produktion, der Speicherung, als auch der Wiedergabe –, müssen alle Stakeholder aktiv werden: Zunächst müssen Produzenten (objektorientierte) Audioinhalte erschaffen und diese über geeignete Distributionswege zum Kunde transportieren. Solange allerdings keine objektorientierten Wiedergabesysteme bereitstehen, ist es fragwürdig, ein solches Format zu produzieren. Umgekehrt ist es risikoreich, ein objektorientiertes Wiedergabesystem zu entwickeln, wenn dafür noch keine Inhalte produziert wurden. Zudem muss eine Einigung auf ein einheitliches Speicherformat getroffen werden, das sowohl bei der Produktion, als auch bei potenziellen Wiedergabesystemen, zum Einsatz kommen kann. Schlussendlich können Konsumenten aber erst dann auf objektorientierte Audioinhalte zurückgreifen, wenn diese sowohl produziert wurden, ein Distributionskanal existiert und auch ein geeignetes Wiedergabesystem zur Verfügung steht. Dieses Phänomen lässt sich mit der Entwicklung und Distribution von HD-Videoinhalten und den dafür benötigten Wiedergabedisplayen vergleichen.

Hier ist gewissermaßen eine Parallelentwicklung der Inhalte und Wiedergabesysteme notwendig, um den Markt zeitgleich durchdringen zu können. Ob sich ein solches System auf dem Konsumentenmarkt durchsetzen kann und sich somit eine Relevanz bzw. Nachfrage seitens der Produktion bildet, bleibt abzuwarten. Auch wenn durch objektorientierte Audioumgebungen technische und klangliche Vorteile entstehen, müssen diese nicht über den Erfolg des Systems entscheiden.

Literaturverzeichnis

- Petermichel (2008): Dateiformate für Audio. In: Weinzierl, S. (Hrsg.): *Handbuch der Audiotechnik*. Berlin/Heidelberg: Springer, S. 687–718.
- Pulkki, V. (1997): Virtual Sound Source Positioning Using Vector Base Amplitude Panning. In: *J. Audio Eng. Soc.* 45 (6), 456–466.
- Stahl, S. (2013): *Objektorientierte Audioumgebungen für Surround-Sound* (Master-Thesis, Hochschule der Medien, Stuttgart); online: <http://www.hdm-stuttgart.de/~curdt/Stahl.pdf>
- Theile, G. (2008): Wellenfeldsynthese. In: Dickreiter, M.; Dittel, V.; Hoeg, W.; Wöhr, M. (Hrsg.): *Handbuch der Tonstudioteknik*. München: Saur, Bd. 1, S. 304–319.
- Slavik, K. M.; Weinzierl, S. (2008): Wiedergabeverfahren. In: Weinzierl, S. (Hrsg.): *Handbuch der Audiotechnik*. Berlin/Heidelberg: Springer, S. 609–682.