

Seminararbeit aus Algorithmen in Akustik und Computermusik 2

Untersuchung verschiedener Upmixingmethoden für DirAC

Manuel Planton, BSc
Michael Hirschmugl, BSc

Betreuung: Dr. Franz Zotter, Dr. Matthias Frank

Graz, 12. Februar 2020



institut für elektronische musik und akustik



Zusammenfassung

Kurzfassungstext. Beispielzitate Buch [?], in einem Konferenzband[?], Diplomarbeit [?], Dissertation [?], Fachzeitschriftenartikel [?].

Inhaltsverzeichnis

1	Einleitung	4
2	DirAC	4
2.1	Implementierung	4
3	Hörversuch	7
4	Schlussfolgerung und Ausblick	7

1 Einleitung

2 DirAC

2.1 Implementierung

In diesem Kapitel wird beschrieben, wie der DirAC-Algorithmus in diesem Projekt implementiert wurde. Die Implementierung liegt als Octave-Skript vor und nutzt das *signal*-Package.

Zu Beginn des Skriptes wird das Eingangssignal im B-Format erster Ordnung eingelesen. Dies muss als *.wav*-Datei vorliegen und genau vier Kanäle umfassen. Die Samplerate wird aus der Datei ermittelt und als Variable *fs* im Skript gespeichert. Anschließend wird eine VBAP-Matrix, je nach vorgegebener Lautsprecheranordnung erzeugt. Diese Matrix wird in späterer Folge verwendet, um das Eingangssignal im B-Format, auf eine bestimmte Lautsprecheranordnung zu dekodieren. Es handelt sich dabei also um eine nicht-parametrische Dekodierung, die mit einer Anordnung von virtuellen Mikrofonen zu vergleichen ist. Die Lautsprecherpositionen können dabei als sphärische Koordinaten in einer Matrix mit genau Spalten vorliegen, oder auch aus einer Textdatei mit kartesischen Koordinaten eingelesen werden. Diese Textdatei muss eine Matrix mit genau drei Spalten (Raumdimensionen) enthalten, wobei diese Spalten untereinander angeordnet sein müssen. Es befinden sich demnach genau die dreifache Anzahl der Lautsprecher als Zeilen in dieser Datei. Die Koordinaten sind dabei Meterangaben als Dezimalwerte. Das Einlesen von kartesischen Koordinaten ist speziell für die Dekodierung auf größere T-Designs äußerst nützlich.

Die vier Kanäle $w[n]$, $x[n]$, $y[n]$ und $z[n]$ des B-Format Eingangssignals werden anschließend in einer Schleife in zeitlichen Blöcken (Fenster) verarbeitet. Ein Fenster besteht dabei aus jeweils 512 Samples und wird in einer 1024-Punkt Fouriertransformation mit der Hilfe der Funktion *fft()* in den Frequenzbereich transformiert. Die FFT-Länge wurde als doppelte Blocklänge gewählt um Aliasing zu vermeiden. Die zeitlichen Fenster werden noch mit einer Hanning-Fensterfunktion für die Resynthese beaufschlagt. Ein Hanning-Fenster kann in Octave mit der Funktion *hanning()* erzeugt werden und einfach als Vektor mit den Zeitsignalen multipliziert werden.

Analyse von Richtungs- und Diffusanteil Zur Bestimmung von Richtungs- und Diffusanteil werden zunächst Schallschnelle und Energie berechnet. Die Schnallschnelle wird als Vektor $\mathbf{V}_m[k] = [X_m[k], Y_m[k], Z_m[k]]$ aus den gerichteten Anteilen (der Druckgradienten-Mikrofone) des B-Format Signals bestimmt. Der Schalldruck ist schlicht der omnidirektionale Anteil $W_m[k]$. Die Indizes m und k werden hier zur Kennzeichnung des Zeitfensters n als Funktion der Frequenzzahl k verwendet.

Die Schallintensität $I_m[k]$ wird aus dem Schnellevektor und dem konjugiert komplexen Schalldruck in Gleichung 1 abgeleitet, wobei hier rein der Realteil herangezogen wird, da sonst auch die Blindeinteile des Schnellevektors miteinbezogen würden. Der Intensitäts-

vektor stellt bereits die Schalleinfallrichtung für alle Frequenzbins einzeln dar, jedoch entgegengesetzt der Einfallrichtung $\mathbf{D}_m[k]$:

$$-\mathbf{D}_m[k] = \mathbf{I}_m[k] = \Re(W_m[k]^* \cdot \mathbf{V}_m[k]) \quad (1)$$

Die Schallenergie $E_m[k]$ wird mithilfe von Gleichung 2 ausgewertet:

$$E_m[k] = \frac{|W_m[k]|^2 + ||\mathbf{V}_m[k]||^2}{2} \quad (2)$$

Um Sprünge in der Lautsprecherzuordnung von gerichteten Signalen bei raschen Bewegungen zu vermeiden, wird der Intensitätsvektor zusätzlich durch einen Mittelwertbildung geglättet. Diese Glättung kann im Skript mit frequenzabhängiger Zeitkonstante vorgegeben werde. Zusätzlich wird auch der Energievektor geglättet. Anschließend wird der Intensitätsvektor verwendet um die Einfallrichtung in sphärischen Koordinaten zu bestimmen.

Die Diffusheit $\psi_m[k]$ kann schlussendlich aus dem Vergleich des Betrags des Intensitätsvektors mit der Energie ermittelt werden (Glg. 3). Der Erwartungswert entspricht hier den (zeitlich) gemittelten Vektoren.

$$\psi_m[k] = \sqrt{1 - \frac{||\mathbb{E}(\mathbf{I}_m[k])||}{\mathbb{E}(E_m[k])}} \quad (3)$$

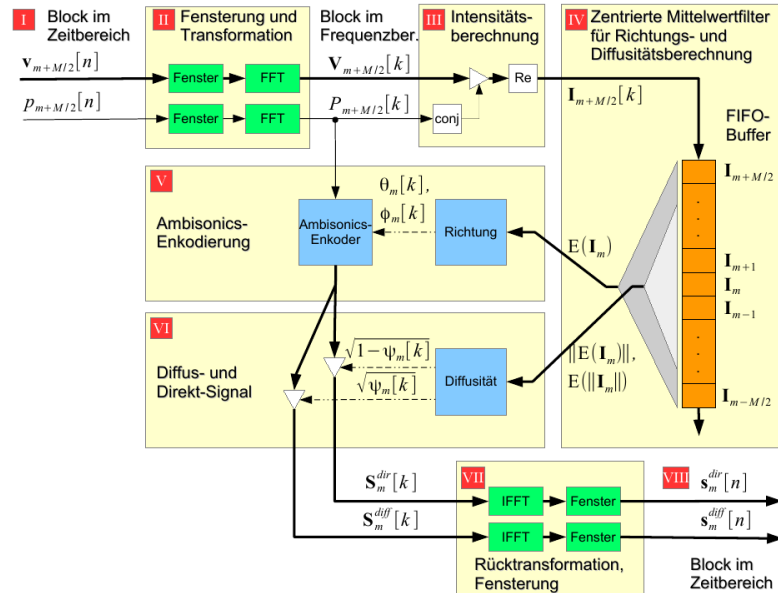


Abbildung 1 – Flussdiagramm des DirAC Algorithmus in Octave

Upmixing auf Lautsprecheranordnung Zur Dekodierung auf eine bestimmte Lautsprecheranordnung wird eine Matrix aus virtuellen Mikrofonen berechnet, woraus sich eine Tabelle mit Gain-Werte ergibt. Diese Tabelle kann die einzelne Frequenzbins in einem

VBAP-Verfahren den entsprechenden Lautsprechern zuordnen. In dieser Seminararbeit wurden dabei zwei Ansätze für die Dekodierung verglichen: Einerseits die direkte Dekodierung auf die Lautsprecheranordnung des wiedergebenden Systems, und andererseits ein eher allgemeiner Ansatz womit die Dekodierung auf ein ambisonisches B-Format Array vierter Ordnung erfolgt. Dieser Dekodierungsansatz hat den Vorteil, dass die Wiedergabeordnung zum Zeitpunkt des Upmixings nicht vorgegeben sein muss. Auch für die Klangqualität stellen sich hierdurch Vorteile heraus welche im Hörversuch besprochen werden.

Die Dekodierung auf eine gegebene Lautsprecheranordnung durch eine VBAP-Tabelle wird hier nicht weiter ausgeführt und kann zum Beispiel im Buch *Ämbisonics*?? nachgelesen werden. Für das B-Format vierter Ordnung wird ein sogenanntes T-Design als virtuelle Anordnung von Lautsprechern verwendet. Ein T-Design ist grundsätzlich zur Darstellung der Maxima von sphärischen Harmonischen notwendig. Wird ein T-Design neunter Ordnung eingesetzt ("9-Design", Abb. 2.1) ergibt sich daraus eine sphärische Anordnung von 48 Lautsprechern, welche die verlustlose Konvertierung zwischen B-Format und einer solchen Lautsprecheranordnung bietet. Es lässt sich sagen, dass ein 9-Design dem ambisonischen B-Format vierter Ordnung entspricht.

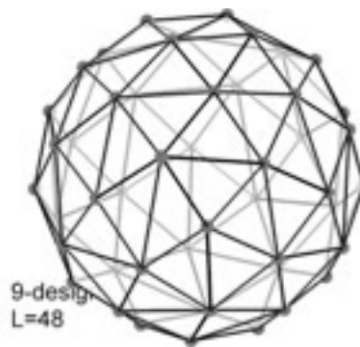


Abbildung 2 – 9-Design T-Design

Zur Dekodierung wird in Octave also eine Matrix mit Lautsprecheramplituden für Schalleinfallrichtungen erstellt. Somit kann einer gegebenen Richtung (bestehend aus Azimuth und Elevation) eine Amplitude am entsprechenden Lautsprecherkanal zugeordnet werden.

Trennung von Diffus- und Richtungsanteil Die Trennung von Diffus- und Richtungsanteil erfolgt im Frequenzbereich, wobei beide Anteile aus dem omnidirektionalen Anteil des B-Format Signals durch gewichtete Filterung erzeugt werden. Prinzipiell wird für den gerichteten Anteil ein spektraler Block im Frequenzbereich mit zwei Matrizen multipliziert: Einerseits mit der VBAP-Tabelle um die Richtung einem Lautsprecher zuzuordnen, und andererseits ebenfalls mit dem frequenzabhängigen Diffusitätsvektor. Somit werden Frequenzbins mit frequenzabhängigen Diffusitätswerten gewichtet und diesen entsprechende Lautsprecheramplituden zugeordnet.

Nach der Berechnung des gerichteten Signalanteils kann auch der Diffusanteil aus dem omnidirektionalen Signal erzeugt werden. Dieser entspricht lediglich dem nicht-gerichteten Anteil des omnidirektionalen Signals und wird daher aus der Filterkurve des Direktanteils bestimmt. Somit ergeben sich an dieser Stelle zwei frequenzabhängige Signalmatrizen für Diffus- und Richtungsanteil. Beide Matrizen besitzen eine Spalte für jeden Lautsprecherkanal und die Zeilenzahl entspricht der FFT-Länge.

Resynthese Die Resynthese entspricht einem gewöhnlichem Overlap-Add Verfahren und erzeugt mittel inverser Fast Fourier Transformation wieder Zeitsignale aus den Matrizen für Richtungs- und Diffusanteil. Im Zeitbereich kann an dieser Stelle weiters auch (optional) die Dekorrelation des Diffusanteils durchgeführt werden. In unserem Skript wurde zur Dekorrelation in Octave ein Random Phase Algorithmus verwendet, welcher im Kapitel ?? näher erläutert wird.

3 Hörversuch

4 Schlussfolgerung und Ausblick