

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	6
4	Schlussfolgerung und Ausblick	6

1 Einleitung

2 DirAC

2.1 Implementierung

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

Zu Beginn werden verschiedene Parameter und Arrays initialisiert, sowie auch eine VBAP-Matrix, je nach vorgegebener Lautsprecheranordnung berechnet. Diese Matrix wird in späterer Folge zu Dekodierung des Eingangssignals im B-Format erster Ordnung auf eine bestimmte Lautsprecheranordnung verwendet. Es handelt sich dabei also um eine nicht-parametrische Dekodierung die einer Anordnung von virtuellen Mikrofonen gleichzusetzen ist. Die Lautsprecherpositionen können dabei als sphärische Koordinaten, oder auch aus einer Textdatei mit kartesischen Koordinaten eingelesen werden. Das Einlesen von kartesischen Koordinaten ist speziell für die Dekodierung auf größere T-Designs äußerst nützlich.

Das Eingangssignal liegt als B-Format erster Ordnung vor und wird aus einer .wav-Datei eingelesen. Die vier Kanäle $w[n]$, $x[n]$, $y[n]$ und $z[n]$ werden anschließend in einer Schleife in zeitliche Blöcke von jeweils 512 Samples unterteilt und mit einer 1024-Punkt FFT in den Frequenzbereich transformiert. Die FFT-Länge ist als Doppelte Fensterlänge gewählt um Aliasing zu vermeiden und die zeitlichen Fenster werden noch mit einer Hanning-Fensterfunktion für die Resynthese beaufschlagt.

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 (Druckgradienten-Mikrofone) des B-Format Signals bestimmt, und der Schalldruck ist schlicht der omnidirektionale Anteil $W_m[k]$. Die Indizes m und k werden hier zur Kennzeichnung des zeitlichen Blockes k 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 Blindeinteile miteinbezogen würden. Der Intensitätsvektor 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 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 eine gleitende Mittelwert geglättet. Diese Glättung kann im Skript mit frequenzabhängiger Zeitkonstante bestimmt werden. Zusätzlich wird auch der Energievektor geglättet. Anschließend wird der Intensitätsvektor verwendet um die Einfallsrichtung in sphärischen Koordinaten zu bestimmen.

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

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

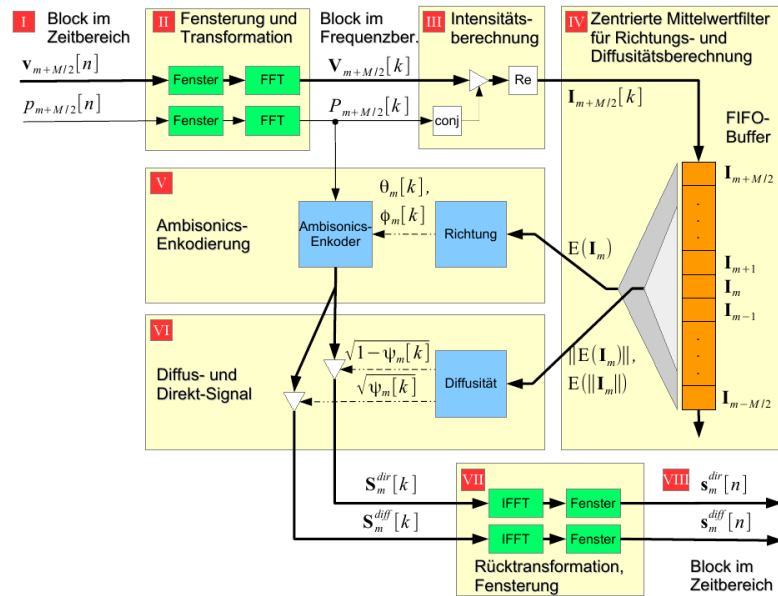


Abbildung 1 – Flussdiagramm des DirAC Algorithmus in Octave

Upmixing auf Lautsprecheranordnung Zur Dekodierung auf eine bestimmte Lautsprecheranordnung wird eine Matrix aus virtuellen Mikrofonen berechnet, woraus sich eine VBAP-Tabelle ergibt. Grundsätzlich können im Frequenzbereich dadurch einzelne Frequenzbins einer Lautsprecher zugeordnet werden. In dieser Arbeit wurden zwei Methoden für diese Dekodierung getestet: Einerseits die direkte Dekodierung auf die Lautsprecheranordnung des Widergebenden Systems, und andererseits ein eher allgemeiner Ansatz womit die Dekodierung auf ein ambisonisches B-Format Array vierter Ordnung erfolgt. Dieses Verfahren hat den Vorteil, dass die Wiedergabeanordnung nicht gegeben

sein muss und 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 "Ambisonics" nachgelesen werden. Für das B-Format vierter Ordnung wird allerdings 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.

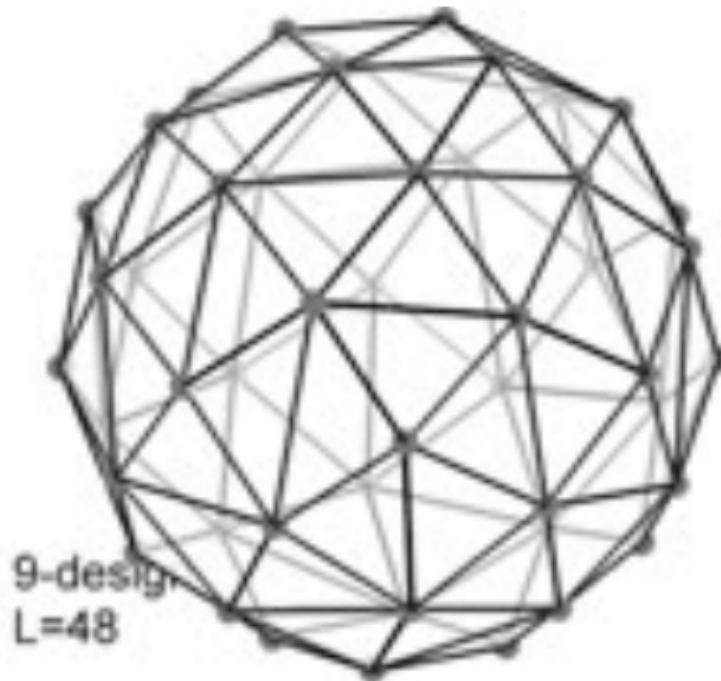


Abbildung 2 – 9-Design T-Design

Für die Dekodierung wird in Octave also eine Matrix mit Lautsprecheramplituden für Schalleinfallrichtungen erstellt. Somit kann eine gegebene Richtung aus Azimuth und Elevation eine Amplitude am entsprechenden Lautsprechersignal zugeordnet werden.

Trennung von Diffus- und Richtungsanteil Die Trennung von Diffus- und Richtungsanteil erfolgt aus dem omnidirektionalen Anteil des B-Format Signals im Frequenzbereich. Prinzipiell wird für den gerichteten Anteil ein Block im Frequenzbereich mit zwei Matrizen multipliziert. Einerseits mit der VBAP-Tabelle um die Richtung einem Lautsprecher zuzuordnen, und andererseits ebenfalls mit dem frequenzabhängigen Diffusivitätsvektor. Somit werden Frequenzbins mit frequenzabhängigen Diffusitätswerten ge-

wichtetet und Lautsprecheramplituden zugeordnet.

Der Diffusanteil entspricht lediglich dem omnidirektionalen Anteil der nicht gerichtet wurde. Somit ergeben sich an dieser Stelle zwei frequenzabhängige Signalmatrizen für Diffus- und Richtungsanteil als Arrays für alle Lautsprecherkanäle.

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. An dieser Stelle kann optional weiters die Dekorrelation des Diffusanteils durchgeführt werden.

3 Hörversuch

4 Schlussfolgerung und Ausblick