

Implementierung eines Zwei- auf Fünfkanal Upmix- algorithmus in Pure Data

Bachelorarbeit aus Aufnahmetechnik 1, SE

Christian Neff

Betreuung: Dr. Alois Sontacchi

Graz, 01. Oktober 2012



institut für elektronische musik und akustik





CHRISTIAN NEFF
(Name in Blockbuchstaben)

0773075
(Matrikelnummer)

Erklärung

Hiermit bestätige ich, dass mir der *Leitfaden für schriftliche Arbeiten an der KUG* bekannt ist und ich diese Richtlinien eingehalten habe.

Graz, den ..01.10.2012.....

.....
Unterschrift der Verfasserin / des Verfassers

Zusammenfassung

Im Rahmen dieser Arbeit wird ein bestehender Zwei- auf Fünfkanal Upmixalgorithmus modifiziert und in der Echtzeitumgebung Pure Data implementiert. Der Algorithmus basiert auf einer Zwei- auf Dreikanal Repanoramisierungsstufe für die drei Frontkanäle einer 5.0 Lautsprecheranordnung sowie einer Hallextraktions- und -synthesestufe zur Generierung der beiden Surroundkanäle.

Zunächst werden die Anforderungen an ein Upmixsystem beschrieben und es wird ein bestehendes matrixbasiertes System vorgestellt. Das der Implementierung zugrundeliegende System und dessen Modifikationen werden eingehend behandelt und die resultierende Umsetzung wird mit diversen Signalen getestet. Abschließend werden die Programmeigenschaften dokumentiert und weitere Verbesserungskonzepte skizziert.

Abstract

In the context of this work, an existing two-to-five channel upmix algorithm is modified and implemented in the real-time programming environment Pure Data. The implemented algorithm is based on a two-to-three channel repanning stage to feed the three front channels of a 5.0 loudspeaker setup as well as an ambience extraction and synthesis stage to generate the two surround channels.

The essential requirements of an upmix system are stated and an existing matrix based upmix system is presented. The base system for the implementation and its modifications are explained in detail and the finished program is tested with various signals. Finally, the qualities of the program are documented and further suggestions are made in order to improve its functionality.

Inhaltsverzeichnis

1 Anforderungen	3
1.1 Vorteile der Surround- gegenüber der Stereowiedergabe	3
1.2 Anforderungen an ein Stereo- auf Mehrkanal Upmixsystem	4
1.3 Bestehende Upmixsysteme	5
1.3.1 Matrixbasierte Upmixsysteme	5
1.3.2 Alternative Upmixansätze	7
2 Ausgewählter Upmixansatz für die Implementierung	9
2.1 Signalmodell.....	9
2.2 Extraktion und Synthese des Hallsignals	10
2.2.1 Kurzzeitkohärenzfunktion	11
2.2.2 Hallindex	11
2.2.3 Nichtlineares Mapping	12
2.3 Repanoramisierung der Quellsignale	13
2.3.1 Panoramakoeffizient	14
2.3.2 Zerlegung des Stereosignals	14
2.3.3 Modifizierungsstufe	16
2.4 Zusammenfassung	18
3 Modifikationen und Echtzeitumsetzung in Pure Data	21
3.1 Implementierung der Hallextraktion und -synthese	21
3.1.1 Modifikationen des gewählten Ansatzes	21
3.1.2 Implementierung der Hallextraktion in Pure Data	21
3.2 Implementierung der Repanoramisierung	24
3.2.1 Modifikationen des gewählten Ansatzes	24
3.2.2 Implementierung der Repanoramisierung in Pure Data	25
4 Hörtests	33
4.1 Weißes Rauschen	33
4.2 Einzelne Instrumente	37
4.2.1 Drumloop mit Stereohall	37
4.2.2 Sologitarre	39

4.2.2 Gitarrenduo.....	42
4.3 Diverse Musikstücke	44
4.4 Filmtonauszüge.....	52
5 Zusammenfassung und Ausblick.....	53
6 Literaturverzeichnis.....	55
7 Abbildungsverzeichnis	57
8 Anhang	59
8.1 Was ist Pure Data?	59
8.2 Pure Data Patches	60
8.2.1 Überblick.....	60
8.2.2 Hallextraktion.....	61
8.2.3 Repanoramisierung	64
8.2.4 Benutzeroberfläche	69

1 Anforderungen

1.1 Vorteile der Surround- gegenüber der Stereowiedergabe¹

Einer der wichtigsten Vorteile einer mehrkanaligen Wiedergabe gegenüber der Wiedergabe über ein 2.0 System ist die größere Stabilität bei der Ortung der im Mix vorhandenen Klangquellen. Insbesondere Klangquellen, welche in die Mitte panoramisiert sind, profitieren sehr von einem zusätzlichen Mittenlautsprecher. Ein reiner Stereomix ist im Allgemeinen darauf ausgelegt, dass sich der Zuhörer im sog. Sweetspot² der Lautsprecheranordnung befindet; in diesem Fall ist die Ortung der einzelnen Klangquellen auch bei zweikanaliger Wiedergabe oft in zufriedenstellendem Maße möglich. Befindet sich der Zuhörer jedoch außerhalb dieses Punktes, wie dies beispielsweise in Kinosälen für einen Großteil des Publikums unumgänglich ist, so ändert sich auch die wahrgenommene Richtung, aus der eine Klangquelle zu kommen scheint. Außerhalb des Sweetspots herrschen andere Pegel- und Laufzeitunterschiede zwischen linkem und rechtem Kanal vor, als für die korrekte, d.h. bei der Erstellung des Mixes intendierte Ortung der Quelle notwendig wären. Ist jedoch ein Mittenlautsprecher vorhanden, so bleibt die Ortung dieser in die Mitte panoramisierten Klangquellen auch außerhalb des Sweetspots stabil, da sie nun nicht mehr vom relativen Pegel zwischen linkem und rechtem Kanal abhängig ist. Die Phantommitte des Stereomixes wird so durch eine physikalische Schallquelle ersetzt.

Auch seitlich panoramisierte Klangquellen profitieren von einem zusätzlichen Mittenlautsprecher. Soll eine Klangquelle beispielsweise halblinks wahrgenommen werden, so muss dies bei Stereowiedergabe über den relativen Pegelunterschied zwischen linkem und rechtem Kanal erreicht werden. Bei einer LCR-Lautsprecheranordnung dagegen wird dies über den relativen Pegelunterschied zwischen linkem und mittlerem Kanal erreicht; der

¹ Im Folgenden wird der Ausdruck „Stereo“ für Zweikanal-Stereo verwendet.

² Der Zuhörer befindet sich im Sweetspot, wenn er mit den beiden Lautsprechern ein gleichseitiges Dreieck bildet.

mögliche Bereich für die Ortung wird sozusagen halbiert und der rechte Lautsprecher enthält keine Anteile dieser Klangquelle.

Der zweite wichtige Aspekt beim Vergleich der Stereowiedergabe mit der Wiedergabe über ein Surroundsystem ist die akustische Umhüllung des Zuhörers. Bei der Stereowiedergabe entsteht die Umhüllung teilweise durch die Wechselwirkung des Schalls aus den Lautsprechern mit dem Raum, in dem sich die Lautsprecher und der Zuhörer befinden. Dieser Raum unterscheidet sich im Allgemeinen von dem Raum, in dem die Musikinstrumente aufgenommen wurden bzw. der durch das Hinzufügen von künstlichem Hall simuliert werden soll.

Bei der Surroundwiedergabe werden zwei grundlegende Arten der akustischen Umhüllung des Zuhörers unterschieden. Beim sogenannten „best seat in the house“-Ansatz werden über die beiden Surroundlautsprecher nur Raumklang-Anteile wiedergegeben. Auf diese Weise wird der umhüllende Klang von dem Raum entkoppelt, in dem sich der Zuhörer befindet und es wird ein in sich stimmiger Raumeindruck erzeugt. So kann für den Zuhörer beispielsweise der Eindruck erweckt werden, er befinde sich in einer Konzerthalle mittig vor der Bühne, er sitze also auf dem besten Platz des Hauses³. Der zweite Ansatz ist der „in the middle of the band“-Ansatz. Bei diesem werden alle Lautsprecher gleichberechtigt verwendet und die Instrumente über alle Lautsprecher im Raum verteilt. Hierdurch entsteht beim Zuhörer der Eindruck, er befinde sich zwischen den Instrumenten. In beiden Fällen ist durch die Surroundwiedergabe eine wesentlich bessere akustische Umhüllung des Zuhörers möglich als bei der Stereowiedergabe.

1.2 Anforderungen an ein Stereo auf Mehrkanal Upmixsystem

Die Anforderungen an ein Upmixsystem ergeben sich aus den in Kapitel 1.1 angeführten Vorteilen der Surroundwiedergabe. Zum einen sollte die Stereoabbildung der Original-Stereomischung möglichst beibehalten werden.

³ Dieser Ansatz wird auch bei Filmmischungen verfolgt, da der Zuschauer nicht durch Geräusche aus den Surroundlautsprechern vom Geschehen auf der Leinwand abgelenkt werden soll.

Gleichzeitig sollte die Lokalisation der einzelnen Klangquellen bei Bewegung des Zuhörers aus dem Sweetspot stabilisiert werden. Dies gilt insbesondere für Klangquellen, die in die Mitte panoramisiert sind. Weiters sollte ein Upmixsystem die klangliche Umhüllung des Zuhörers erhöhen und so ein dreidimensionales Hörerlebnis schaffen.

1.3 Bestehende Upmixsysteme

1.3.1 Matrixbasierte Upmixsysteme

Zu den einfachsten und am weitesten verbreiteten Upmixsystemen gehören lineare Matrixsysteme, bei denen die neuen Signale aus den Signalen des linken und des rechten Kanals gebildet werden, indem diese addiert, subtrahiert und gegebenenfalls mit einem Verstärkungsfaktor multipliziert werden. Vorgestellt werden solche Verfahren beispielsweise in [1] und [3]. Auch Dolby Pro Logic [4] ist ein Matrixverfahren, welches hier in Grundzügen erläutert werden soll.

Strenggenommen ist Dolby Pro Logic kein Upmixsystem, da hier vier Kanäle (Links, Mitte, Rechts, Surround) auf zwei Kanäle (Lt, Rt)⁴ codiert werden, welche dann als Stereomix abgespielt werden können oder aber beim Abspielen des Programms wieder zu vier Kanälen decodiert werden. Allerdings kann auch das Decodieren eines reinen Zweikanal-Stereomixes brauchbare Ergebnisse im Sinne einer Upmixanwendung liefern. Die prinzipielle Funktionsweise ist in den Abbildungen 1 und 2 dargestellt.

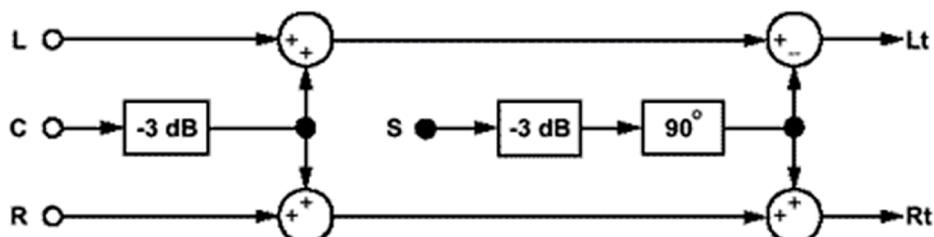


Abb.1: Schematische Darstellung der Dolby Codierungsmatrix

⁴ Left total, Right total

Bei der Codierung wird der Mittenkanal dem linken und dem rechten Kanal zu gleichen Teilen hinzugemischt. Dabei wird er zuvor im Pegel um 3 dB abgesenkt, so dass die Anteile des Mittenkanals bei der Wiedergabe von Lt und Rt als Stereomix die gleiche Energie besitzen wie vor der Codierung. Das gleiche wird mit dem Surroundkanal gemacht. Bei diesem wird aber zusätzlich die Phasenlage geändert, und zwar um $+90^\circ$ bei der Addition zum linken Kanal, sowie um -90° bei der Addition zum rechten Kanal. Insgesamt ergibt sich so bei Addition der beiden Kanäle eine Phasendifferenz der Surroundanteile von 180° und diese löschen sich bei der Rückgewinnung des Mittenkanals aus.

Der einfachste Weg zur Rückgewinnung des Surroundsignals bei der Decodierung besteht aus einem Subtrahierer. Da das Mittensignal in Phase auf beiden Kanälen vorliegt, fällt es beim Subtrahieren der beiden Kanäle heraus [13].

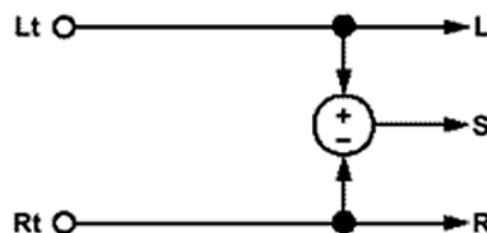


Abb.2: Funktionsweise einer einfachen Decodierungsmatrix zur Gewinnung des Surroundsignals

Die Decodierung kann folgenderweise als Matrix dargestellt werden:

Eingang \ Ausgang	Lt	Rt
(L) Links	1	0
(C) Mitte	$\sqrt{0,5}$	$\sqrt{0,5}$
(R) Rechts	0	1
(S) Surround	$\sqrt{0,5}$	$-\sqrt{0,5}$

Abb.3: Dolby Pro Logic Decodierungsmatrix

Durch diese Funktionsweise ergibt sich auch die Möglichkeit des Einsatzes der Decodierungsmatrix als einfaches Upmixsystem. Die Qualität des so gewonnenen Surroundmixes ist jedoch sehr von der Beschaffenheit des Stereo-Eingangssignales abhängig und variiert stark.

Nachteile dieses Systems sind insbesondere, dass die Stereoabbildung des Originalsignals oftmals nicht beibehalten wird, sowie die Entstehung von Artefakten. Speziell bei der Gewinnung des Surroundkanals können ungewollte Direktsignalanteile mit in den Surround wandern, wenn diese nur auf einem der beiden Stereokanäle vorhanden sind. In diesem Fall löschen sich diese Anteile bei der Subtraktion der beiden Kanäle voneinander nicht aus.

1.3.2 Alternative Upmixansätze

Neben den oben vorgestellten matrixbasierten Upmixsystemen wurde in den letzten Jahren eine Vielzahl weiterer Ansätze für Upmixverfahren entwickelt. Die hier angeführten Ansätze unterscheiden sich in erster Linie dadurch von den matrixbasierten Systemen, dass die Eingangssignale zuerst in den Frequenzbereich und die bearbeiteten Signale wieder zurück in den Zeitbereich transformiert werden. Hierdurch sind wesentlich differenziertere Signalanalysen möglich, auf denen diese Upmixansätze aufbauen. Es sei hier beispielhaft verwiesen auf die Arbeiten von Christof Faller (z.B. [5] und [6]), in denen unter anderem eine Analyse nach Frequenzgruppen vorgeschlagen wird, um die Direktkomponenten von den Hallkomponenten eines Stereosignals zu unterscheiden. Weiters zu nennen sind die Arbeiten von Oliver Thiergart (z.B. [7] und [8]), wo aus Eingangssignalen im B-Format, die zum Beispiel von einem Soundfield-Mikrofon stammen können, u.a. die Positionen einzelner Schallquellen bestimmt werden, sowie die Arbeiten von Earl Vickers der in [9] eine Optimierung für die Mittenkanalgewinnung aus einem Stereosignal vorschlägt.



2 Ausgewählter Umixansatz für die Implementierung

Das in Pure Data implementierte Programm basiert auf dem in [10] vorgestellten Ansatz von Carlos Avendado und Jean-Marc Jot. Im Folgenden wird eine detaillierte Beschreibung der darin entwickelten Stereo-auf-Mehrkanal Umixmethode gegeben.

2.1 Signalmodell

Stereoaufnahmen können in zwei Hauptkategorien unterteilt werden: *Studioaufnahmen* bzw. künstlich und *Liveaufnahmen* bzw. natürlich.

Handelt es sich um eine Studioaufnahme, so werden die verschiedenen Klangquellen bzw. Instrumente üblicherweise einzeln aufgenommen und dann zu einem Stereosignal zusammengemischt. Die räumliche Platzierung der einzelnen Quellen bei der Mischung erfolgt über Amplitudenpanoramisierung, d.h. die Richtung, aus der eine Quelle wahrgenommen wird, hängt ab vom relativen Pegelunterschied dieser Quelle zwischen dem linken und dem rechten Lautsprecher⁵. Insbesondere ist hervorzuheben, dass zeitliche Verzögerungen zwischen linkem und rechtem Kanal bei dieser Art der Panoramisierung keine Rolle spielen. Schließlich wird dem Mix bzw. den einzelnen Instrumenten künstlicher Hall hinzugefügt. Die linke und die rechte Impulsantwort eines Stereo-Hallprozessors unterscheiden sich im Allgemeinen voneinander. Sie sind nur schwach korreliert um den Eindruck von Räumlichkeit zu unterstützen.

Im Falle einer Liveaufnahme kommen mehrere räumlich verteilte Mikrofone zum Einsatz um alle Quellen aufnehmen zu können. Wird eine Stereo-Mikrofontechnik verwendet, so wird der umhüllende Schall⁶ direkt in das Stereosignal mitaufgenommen. Der umhüllende Schall setzt sich hierbei

⁵ Man bezeichnet diese Art der Stereophonie als Intensitätsstereophonie.

⁶ Der umhüllende Schall wird im Folgenden als Hall bezeichnet

zusammen aus dem Nachhall des Raumes sowie aus anderen räumlich verteilten Quellen wie z.B. Wind, Publikumsgeräuschen etc.

Das Signalmodell wird nun folgendermaßen definiert:

N Klangquellen $s_j(t)$, $j = 1, \dots, N$ sind mit Raumimpulsantworten⁷ $h_{ij}(t)$ gefaltet und bilden so den linken ($i = 1$) bzw. den rechten ($i = 2$) Kanal.

Die Impulsantworten $h_{ij}(t)$ können unterteilt werden in einen Direktpfad⁸ $d_{ij}(t)$ und eine Hallkomponente⁹ $r_{ij}(t)$; die Impulsantworten setzen sich additiv aus diesen beiden Anteilen zusammen. Das Stereosignal kann nun geschrieben werden als:

$$x_i(t) = \sum_{j=1}^N s_j(t) * d_{ij}(t) + \sum_{j=1}^N s_j(t) * r_{ij}(t) + n_i(t) \quad (1)$$

wobei $n_{ij}(t)$ Hintergrundgeräusch-Signale sind, die zur Umhüllung beitragen.

Der erste Term aus Gleichung (1) auf der rechten Seite des Gleichheitszeichens entspricht den Direktsignalen, der zweite Term entspricht dem umhüllenden Schall. Im Allgemeinen wird der umhüllende Schall auf beiden Kanälen in etwa gleiche Pegel haben, d.h. $\|r_{1j}(t)\|^2 \simeq \|r_{2j}(t)\|^2$ und $\|n_1(t)\|^2 \simeq \|n_2(t)\|^2$.

2.2 Extraktion und Synthese des Hallsignals

Die erste Stufe der benötigten Signalverarbeitungskette besteht aus einer Kurzzeit Fourier Transformations¹⁰-Analyseeinheit. Die zugrundeliegende Idee ist, die STFT des linken und die des rechten Kanals miteinander zu vergleichen, um die verschiedenen Komponenten des Stereomixes zu identifizieren. Die STFTs der beiden Stereokanäle $x_i(t)$ werden geschrieben als $X_i(m, k)$. Hier ist m der Zeitindex und k der Frequenzindex.

⁷ oder Impulsantworten eines Hallgeräts

⁸ entspricht dem direkten Weg von der Quelle zum Mikrofon

⁹ entspricht den Reflexionen

¹⁰ Im Folgenden bezeichnet mit STFT (Short-Time Fourier Transform)

2.2.1 Kurzzeitkohärenzfunktion

Die Kreuzkorrelationsfunktion für eine bestimmte Frequenz k ist definiert als

$$\phi_{ij}(k) = E\{X_i(m,k)X_j^*(m,k)\} \quad (2)$$

Mit E wird hier der Erwartungswertoperator bezeichnet.

Da Audiosignale im Allgemeinen nicht stationär sind, ändert sich der Wert der Kreuzkorrelationsfunktion mit der Zeit. Um die Änderungen des Signals zu verfolgen und um ein kausales System implementieren zu können, wird ein Faktor λ bei der Berechnung der Kreuzkorrelationsfunktion eingeführt. In der Praxis wird diese folgendermaßen berechnet:

$$\phi_{ij}(m,k) = \lambda\phi_{ij}(m-1,k) + (1-\lambda)X_i(m,k)X_j^*(m,k) \quad (3)$$

Mit Hilfe dieser statistischen Größe wird nun die Kurzzeitkohärenzfunktion definiert als:

$$\phi(m,k) = \frac{|\phi_{12}(m,k)|}{[\phi_{11}(m,k)\phi_{22}(m,k)]^{\frac{1}{2}}} \quad (4)$$

Diese ist reellwertig und nimmt Werte zwischen 0 und 1 an. In Zeit-Frequenz-Regionen mit hoher Korrelation zwischen den beiden Kanälen liegen die Werte nahe bei 1, während die Werte in Regionen mit schwacher Korrelation zwischen den Kanälen nahe bei 0 liegen.

2.2.2 Hallindex

Der Hallindex $\Phi(m,k)$ wird mit Hilfe der Kurzzeitkohärenzfunktion (4) definiert als:

$$\Phi(m,k) = 1 - \phi(m,k) \quad (5)$$

Eine einfache Methode, den Hallanteil des linken und des rechten Kanals aus einer Stereoaufnahme zu extrahieren ist nach [11] die beiden STFTs mit dem Hallindex zu multiplizieren und daraus die beiden Hallsignale $a_i(t)$ durch Rücktransformation in den Zeitbereich zu gewinnen. Ein allgemeinerer Ansatz wird in Kapitel 2.2.3 beschrieben.

2.2.3 Nichtlineares Mapping

Um bessere Ergebnisse bei der Hallextraktion zu erzielen, wird eine nichtlineare Funktion des Hallindexes eingeführt. Ziel ist es, Zeit-Frequenz-Regionen mit hohem Hallindex unverändert zu lassen, Regionen mit niedrigem Hallindex jedoch stark zu dämpfen. Hierdurch werden die in diesen Regionen vorherrschenden Direktsignalanteile aus dem resultierenden Hallsignal entfernt. Eine geeignete Funktion ist der Tangens hyperbolicus.

$$\Gamma[\Phi] = \left(\frac{\mu_1 - \mu_0}{2} \right) \tanh\{\sigma\pi\hat{\Phi}\} + \left(\frac{\mu_1 + \mu_0}{2} \right) \quad (6)$$

In dieser Gleichung ist $\hat{\Phi} = \Phi - \Phi_0$, wobei Φ_0 ein Schwellwert ist. Die Parameter μ_1 und μ_0 definieren den Wertebereich der Funktion und über σ kann die Steilheit der Funktion kontrolliert werden.

Multipliziert man nun die STFTs des linken und des rechten Kanals mit dieser nichtlinearen Funktion, so erhält man die modifizierten oder auch Hall-Transformationen:

$$A_i(m,k) = X_i(m,k)\Gamma[\Phi(m,k)] \quad (7)$$

Durch Rücktransformation in den Zeitbereich werden aus $A_i(m, k)$ die beiden Hallsignale $a_i(t)$ gewonnen. In Abbildung 4 ist die Vorgangsweise bei der Hallextraktion schematisch dargestellt:

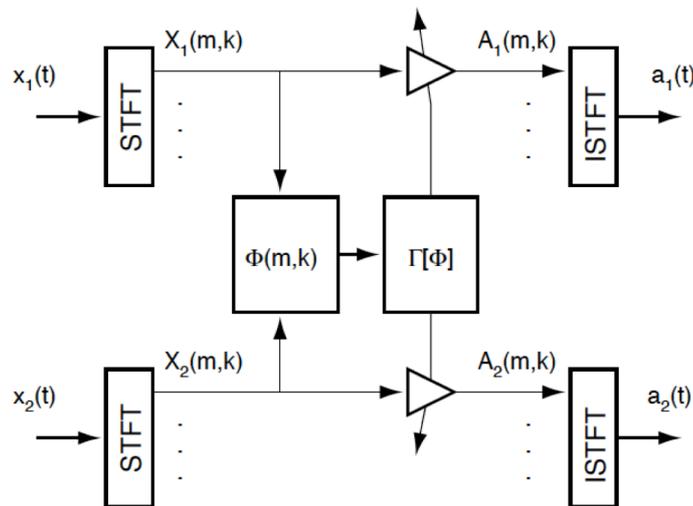


Abb.4: Schematische Darstellung der Hallextraktion und -synthese

2.3 Repanoramisierung der Quellsignale

Um ein Stereosignal über ein 5.1 System wiederzugeben, wird symmetrisch zwischen den linken und den rechten Lautsprecher ein zusätzlicher Mittenlautsprecher eingefügt. Anschließend wird eine Zwei- auf Dreikanal-konversion des Stereosignals durchgeführt. Ziel bei dieser Konversion ist es, trotz des zusätzlichen Lautsprechers die Richtung, aus der eine Klangquelle vom Zuhörer wahrgenommen wird, beizubehalten. Hierzu müssen die Panoramapositionen der einzelnen Quellen berechnet und für die dreikanalige Wiedergabe modifiziert werden.

2.3.1 Panoramakoeffizient

Die Ausgangsbasis für die Repanoramisierung stellt die Bestimmung des Panoramakoeffizienten dar. Für die Berechnung des Panoramakoeffizienten wird zur Vereinfachung von einem Mix ausgegangen, in dem nur eine einzelne amplitudenpanoramisierte Quelle $s_j(t)$ vorhanden ist und der keine Hallkomponenten enthält. Gemäß dem Signalmodell in Gleichung (1) können das linke und das rechte Signal dann geschrieben werden als

$$x_1(t) = (1 - \alpha)s_j(t) \text{ und } x_2(t) = \alpha s_j(t) \quad (8)$$

mit $d_{10}(t) = 1 - \alpha$ und $d_{20}(t) = \alpha$.

Der Panoramakoeffizient α kann Werte zwischen 0 und 1 annehmen. Er ist 0, wenn die Quelle nach links panoramisiert ist und er wird 1, wenn die Quelle nach rechts panoramisiert ist. Für eine Quelle, die in die Mitte panoramisiert ist, ist $\alpha = 0,5$.

Avendado und Jot machen keine Angabe dazu, wie α direkt aus einem gegebenen Stereosignal berechnet werden kann. In Kapitel 3.2.2.3 wird die gewählte Berechnung im Rahmen der Implementierung gezeigt.

2.3.2 Zerlegung des Stereosignals

Der nächste Schritt bei der Zwei- auf Dreikanal-Konversion besteht darin, zwei neue Signalpaare aus dem Stereosignal zu generieren, wie in der "unmix-" Stufe von Abbildung 5 dargestellt.

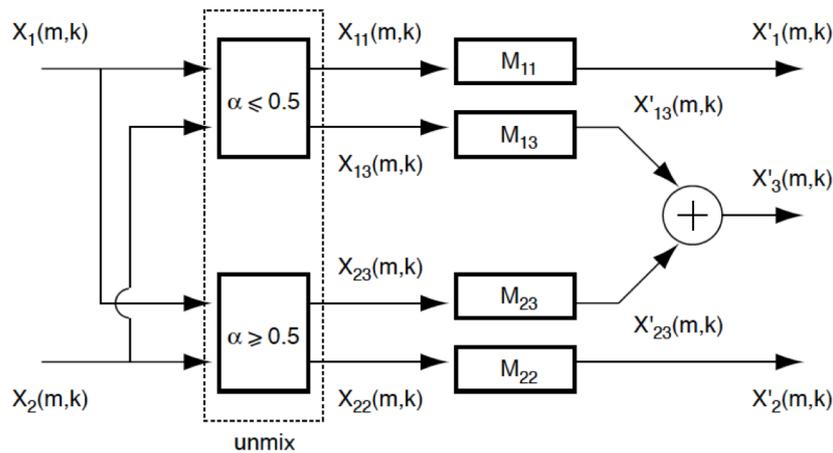


Abb.5: Funktionsweise des Zwei- auf Dreikanal Upmix

Die beiden STFTs $X_1(m, k)$ und $X_2(m, k)$ durchlaufen eine Zerlegungs-Stufe, welche die Signale abhängig vom Wert des Panoramakoeffizienten α in vier neue Signale aufsplittet. Konkret werden STFT-Regionen identifiziert, in welchen $\alpha \leq 0,5$ ist und solche, in denen $\alpha \geq 0,5$ ist. Dieser Prozess generiert vier neue Signale:

- $X_{11}(m, k)$ enthält nach links panoramisierte Komponenten des linken Kanals.
- $X_{13}(m, k)$ enthält nach links und in die Mitte panoramisierte Komponenten des rechten Kanals.
- $X_{23}(m, k)$ enthält nach rechts und in die Mitte panoramisierte Komponenten des linken Kanals.
- $X_{22}(m, k)$ enthält nach rechts panoramisierte Komponenten des rechten Kanals.

Um die Stereoabbildung des Originalsignals beizubehalten, werden diese vier Signale in der nächsten Stufe mittels Mapping-Funktionen modifiziert.

2.3.3 Modifizierungsstufe

Wie in Abbildung 5 dargestellt, durchlaufen die vier aus der Zerlegung des Stereosignals resultierenden Signale eine weitere Modifizierungsstufe M_{ij} . Diese stellt sicher, dass die Stereoabbildung der Originalmischung erhalten bleibt.

Angenommen eine Klangquelle ist so panoramisiert dass sie aus einem bestimmten Winkel θ_o wahrgenommen wird, bei einem Winkel θ zwischen den beiden Lautsprechern L und R (s. Abbildung 6 a)). Die Modifikation des Panoramawertes dieser Klangquelle muss dann so erfolgen, dass der wahrgenommene Winkel sich nicht ändert, bei einer Halbierung des Winkels θ zwischen den Lautsprechern (s. Abbildung 6 b)).

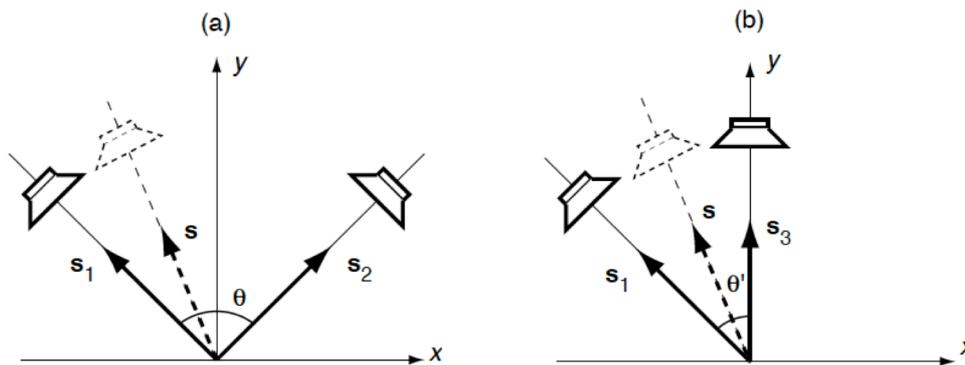


Abb.6: Koordinatensystem für Lautsprecheranordnung. (a) Original-Anordnung, (b) Links und Center Lautsprecher-Anordnung.

Nach M. Gerzon [2] wird in einem zweidimensionalen Koordinatensystem die wahrgenommene Richtung der Quelle $s = [x, y]^T$ einerseits bestimmt vom Panoramakoeffizientenvektor $a = [a_1, a_2]^T$, wobei $a_1 = 1 - \alpha$ und $a_2 = \alpha$. Andererseits bestimmen die Positionen der Lautsprecher relativ zum Zuhörer die Richtung, aus der die Quelle wahrgenommen wird. Diese können durch die Vektoren $s_1 = [x_1, y_1]^T$ und $s_2 = [x_2, y_2]^T$ dargestellt werden.

Für tiefe Frequenzen ($f < 700$ Hz) erhält man die wahrgenommene Richtung durch Vektoraddition:

$$s = [s_1 s_2]^T a \quad (9)$$

Bei hohen Frequenzen ($f > 700$ Hz) wird die wahrgenommene Richtung der Klangquelle durch Addition der Intensitätsvektoren bestimmt, die von jedem Lautsprecher abgestrahlt werden:

$$\mathbf{s} = [\mathbf{s}_1 \mathbf{s}_2]^T \mathbf{p} \quad (10)$$

wobei $\mathbf{p} = [a_1^2 a_2^2]^T$.

Diese Klangquelle nun zu zwischen L und C zu repanoramisieren ist gleichbedeutend mit einer Kompensation für die Reduzierung des Winkels zwischen linkem und rechtem Lautsprecher. Aus diesem Grund müssen die Panoramakoeffizienten neu berechnet werden.

Wird die Position des eingefügten Mittenlautsprechers durch den Vektor $\mathbf{s}_3 = [x_3, y_3]^T$ dargestellt, so kann der neue Koeffizientenvektor \mathbf{a}' durch Lösen der folgenden Gleichungen gefunden werden.

Für $f < 700$ Hz:

$$[\mathbf{s}_1 \mathbf{s}_2]^T \mathbf{a} = [\mathbf{s}_1 \mathbf{s}_3]^T \mathbf{a}' \quad (11)$$

und für $f > 700$ Hz:

$$[\mathbf{s}_1 \mathbf{s}_2]^T \mathbf{p} = [\mathbf{s}_1 \mathbf{s}_3]^T \mathbf{p}' \quad (12)$$

wobei \mathbf{p}' der neue Intensitätsvektor ist. Für die numerische Lösung dieser Gleichungen siehe Kapitel 3.2.2.5.

Um die Stereoabbildung des Originalsignals bei Änderung der Lautsprecherkonfiguration beizubehalten, wird das Signal mit dem Verhältnis der neuen zu den alten Koeffizienten multipliziert. Dieses Verhältnis ergibt also die gesuchten Modifikationsfunktionen M_{11} und M_{13} aus Abbildung 5.

Zur Berechnung der Funktionen M_{22} und M_{23} , müssen die Gleichungen (11) und (12) mit den entsprechenden Vektoren zwischen Mittenlautsprecher und

rechtem Lautsprecher gelöst werden. Sie sind bei symmetrischer Lautsprecheraufstellung spiegelbildlich zu M_{11} und M_{13} .

Die beschriebene Vorgangsweise kann grundsätzlich auf jede beliebige Anzahl von Lautsprechern erweitert werden.

2.4 Zusammenfassung

Mit den in diesem Kapitel beschriebenen Techniken kann ein Zwei- auf Fünfkanal Upmixsystem realisiert werden, wie in Abbildung 7 dargestellt.

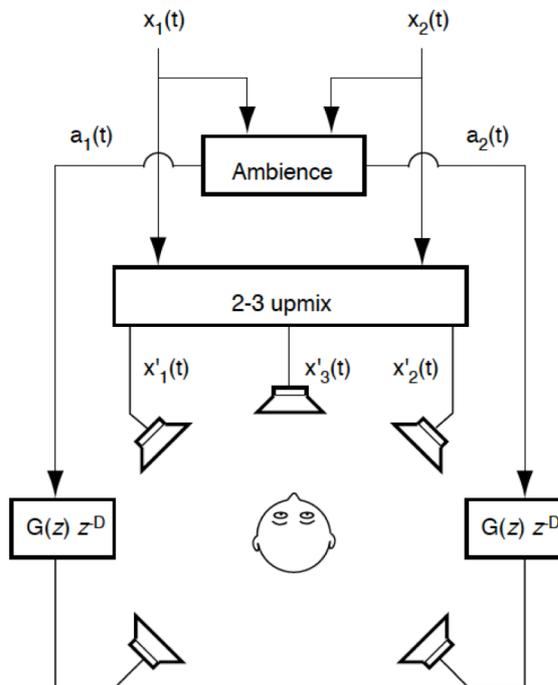


Abb. 7: Blockdiagramm Zwei- auf Fünfkanal Upmixsystem

Die Surroundsignale werden mittels der Hallextraktionsmethode aus Kapitel 2.2 generiert. Um sie von den Hallsignalen aus den Frontlautsprechern zu dekorrelieren, sollten sie außerdem eine Allpass-Filterstufe durchlaufen. Weiters werden sie idealerweise um einen Betrag D zeitlich gegenüber den Signalen der Frontlautsprecher verzögert, um unerwünschte Lokalisierungseffekte in den Surroundlautsprechern zu vermeiden.

Zur Gewinnung der drei Signale für die Frontlautsprecher wird die Repanoramisierungstechnik aus Kapitel 2.3 verwendet. Die nach der Modifikationsstufe erhaltenen STFT-Signale $X_1'(m, k)$, $X_2'(m, k)$ und $X_3'(m, k)$ werden schließlich zurück in den Zeitbereich transformiert, um $x_1'(t)$, $x_2'(t)$ und $x_3'(t)$ zu erhalten.



3 Modifikationen und Echtzeitumsetzung in Pure Data

3.1 Implementierung der Hallextraktion & -synthese

3.1.1 Modifikationen des gewählten Ansatzes

Um das Eingangs-Zeitsignal in den Frequenzbereich zu transformieren schlagen die Autoren in [11] vor, eine 2048-Punkt-FFT¹¹ mit einem Hamming-Fenster einer Breite von 1024 Samples durchzuführen. Hierdurch ergäbe sich eine Frequenzauflösung von $\frac{44100}{2048} \approx 22$ Hz. Die derzeit implementierte Version arbeitet stattdessen nur mit einer 64-Punkt-FFT, was einer wesentlich schlechteren Frequenzauflösung von 689 Hz entspricht. Dies führt wie wir sehen werden dazu, dass der Extraktionsalgorithmus bei etwas komplexerem Audiomaterial relativ schlecht arbeitet, da es kaum Zeit-Frequenz-Regionen gibt, in denen vorwiegend Hallkomponenten vorhanden sind.

Andererseits korrespondiert die gröbere Frequenzauflösung bei der FFT mit einer feineren Zeitauflösung. Dies resultiert in einer wesentlich höheren Empfindlichkeit des Algorithmus auf kurzzeitige Änderungen der Korrelation. Um trotzdem ein stabiles Verhalten der Hallextraktion zu erreichen, wird eine zeitliche Glättung der Kohärenzfunktion durchgeführt. Dies erfolgt durch die Wahl eines höheren Wertes für den Faktor λ aus Gleichung (3). In [11] wird für den Faktor λ ein Wert von 0,85 für alle Frequenzen angegeben. Demgegenüber führen bei der implementierten Version erst Werte ab 0,95 zu brauchbaren Ergebnissen.

3.1.2 Implementierung der Hallextraktion in Pure Data

Für die Durchführung der Fourier-Transformation wurde das in Pure Data integrierte „rfft“-Objekt verwendet. Dieses liefert die Real- und Imaginärteile

¹¹ Fast Fourier Transformation

der Fast-Fourier-Transformation eines eingehenden Audiosignals. Bei der derzeit verwendeten Blockgröße von 64 Samples erhalten wir so 64 Real- und Imaginärteile. Allerdings sind die Amplitudenwerte für alle Frequenzen oberhalb der halben Abtastrate auf Null gesetzt (Nyquist-Frequenz), so dass für die Weiterverarbeitung nur die ersten 32 Werte jedes Blocks von Interesse sind.

Nach der FFT werden die Real- und Imaginärteile mittels des „pack~“-Objektes in Listen umgewandelt. Alle weiteren Berechnungen werden mit diesen Listen durchgeführt. Die Fourier-transformierten Audiosignale werden dann mit dem Ergebnis der Berechnungen multipliziert und mittels des „riff~“-Objektes zurück in den Zeitbereich transformiert. Der Datenfluss innerhalb des Programms spaltet sich also nach der Fourier-Transformation auf in einen Audiosignalfad und einen Berechnungspfad, welcher mit Kontrolldaten arbeitet.

3.1.2.1 Berechnung des Hallindex

Zur Erinnerung: Die Kreuzkorrelation berechnet sich in der Praxis gemäß Gleichung (3) zu

$$\phi_{ij}(m, k) = (1 - \lambda)\phi_{ij}(m - 1, k) + \lambda X_i(m, k)X_j^*(m, k).$$

Somit erhalten wir z.B.

$$\phi_{12}(m, k) = (1 - \lambda)\phi_{12}(m - 1, k) + \lambda X_1(m, k)X_2^*(m, k).$$

Schreiben wir nun Real- und Imaginärteil getrennt als

$$X_1(m, k) = (a + ib)(m, k) \text{ und } X_2(m, k) = (c + id)(m, k),$$

so erhalten wir

$$\phi_{12}(m, k) = ((ac + bd) + i(bc - ad))(m - 1, k) + (1 - \lambda)((ac + bd) + i(bc - ad))(m, k).$$

Bei der Berechnung von ϕ_{11} sowie ϕ_{22} wird jeweils das Signal mit seiner konjugiert komplexen Größe multipliziert. Dadurch werden die Imaginärteile Null und wir erhalten

$$\phi_{11}(m, k) = \lambda(a^2 + b^2)(m - 1, k) + (1 - \lambda)(a^2 + b^2)(m, k)$$

sowie

$$\phi_{22}(m, k) = \lambda(c^2 + d^2)(m - 1, k) + (1 - \lambda)(c^2 + d^2)(m, k).$$

Dies sind einfach zu implementierende Funktionen in Pure Data. Zu beachten ist, dass diese Berechnungen zu jedem Zeitindex m für alle k Frequenzbereiche parallel ausgeführt werden müssen. Dabei können prinzipiell in unterschiedlichen Frequenzbereichen auch verschiedene Werte für λ verwendet werden [11].

Die Berechnungen von Kurzzeitkohärenzfunktion (4) und Hallindex (5) wurden in Pure Data schließlich durch Expressions realisiert.

3.1.2.2 Implementierung des nichtlinearen Mappings

Das Programm lässt den Benutzer auswählen, ob dieser für die Hallextraktion den einfachen Hallindex verwenden will, oder die nichtlineare Funktion des Hallindex $\Gamma[\Phi]$ (siehe Gleichung (6)). Diese soll, wie in Kap. 2.2.3 beschrieben, dafür sorgen, dass Zeit-Frequenz-Bereiche mit niedrigem Hallindex stark gedämpft werden, während Zeit-Frequenz-Bereiche mit hohem Hallindex unverändert bleiben. Realisiert wurde diese Funktion über ein Lookup-Table. Dies ist eine Tabelle, aus der für jeden eingehenden Hallindex-Wert der entsprechende Funktionswert ausgelesen wird. Diese Art der Implementierung hat den Vorteil, dass die benötigte Rechenleistung des Programms geringer ist als bei direkter Berechnung zur Laufzeit. Der Nachteil

ist allerdings, dass die Tabelle jedes Mal neu geschrieben werden muss, wenn ein Parameter der Funktion geändert werden soll.

3.2 Implementierung der Repanoramisierung

3.2.1 Modifikationen des gewählten Ansatzes

Eine wesentliche Modifikation bei der Implementierung der Repanoramisierung ist, dass sie im Gegensatz zur in Kapitel 2.3 beschriebenen Methode mit Frequenzgruppen arbeitet. Hierzu werden jeweils die STFT-Regionen zusammengefasst, die innerhalb einer Frequenzgruppe liegen. Nachfolgende Definition der Frequenzgruppen findet sich in [16]:

„Mit Frequenzgruppen (Plural), engl. „critical bands“, bezeichnet man beim menschlichen Gehör Frequenzbereiche, die gemeinsam ausgewertet werden“.

„Eine Auswertung in Frequenzgruppen erfolgt z. B. bei der Bestimmung der Lautstärke, des Klangs oder der Richtung des Schalls.“ [ebd.]

Nummer des Bandes in Bark	Untere Frequenzgrenze In Hz	Obere Frequenzgrenze In Hz	Bandbreite In Hz
1	0	100	100
2	100	200	100
3	200	300	100
4	300	400	100
5	400	510	110
6	510	630	120
7	630	770	140
8	770	920	150
9	920	1080	160
10	1080	1270	190
11	1270	1480	210
12	1480	1720	240
13	1720	2000	280
14	2000	2320	320
15	2320	2700	380
16	2700	3150	450

17	3150	3700	550
18	3700	4400	700
19	4400	5300	900
20	5300	6400	1100
21	6400	7700	1300
22	7700	9500	1800
23	9500	12000	2500
24	12000	15500	3500

Tabelle 1: Aufteilung und Eckfrequenzen der 24 Frequenzgruppen nach [16]

Statt also alle STFT-Regionen getrennt zu analysieren und zu repanoramisieren, werden diese gemäß der Tabelle 1 in 24 Gruppen zusammengefasst. Durch diese Zusammenfassung soll die Repanoramisierung einerseits dem menschlichen Hörempfinden besser nachempfunden werden. Andererseits wird auch der Rechenaufwand zur Programmlaufzeit erheblich reduziert.

3.2.2 Implementierung der Repanoramisierung in Pure Data

Um das eingehende Audiosignal vom Zeit- in den Frequenzbereich zu überführen, wird in Pure Data zunächst eine 1024-Punkt FFT durchgeführt. Dies geschieht wieder mit dem „rfft“-Objekt bei einer Blockgröße von 1024. Durch diese Transformation erhalten wir 1024 FFT-Bins, was einer Frequenzauflösung von $\frac{44100}{1024} \approx 43,1$ Hz entspricht. Auch hier sind für die Weiterverarbeitung nur die ersten 512 Bins von Interesse, da alle Bins darüber Null gesetzt sind.

3.2.2.1 Aufteilung in Frequenzgruppen

Die resultierenden Real- und Imaginärteile der Fourier-Transformation werden zunächst umgerechnet in das Betrags- und das Phasenspektrum. Dies geschieht auf Audiosignalebene durch Implementierung der Funktionen $|X_i(m, k)| = \sqrt{\text{Re}\{X_i(m, k)\}^2 + \text{Im}\{X_i(m, k)\}^2}$ für das Betragsspektrum und

$\angle\{X_i(m, k)\} = \tan^{-1} \frac{\text{Im}\{X_i(m, k)\}}{\text{Re}\{X_i(m, k)\}}$ für das Phasenspektrum. Anschließend werden die Audiosignale mittels des „pack~“-Objektes in Listen umgewandelt. Wie bei der Hallextraktion werden alle weiteren Berechnungen mit diesen Listen, also auf Kontrollsignalebene, durchgeführt.

Aus dem Betragsspektrum werden nun durch Zusammenfassung und Mittelwertbildung mehrerer Bins die Frequenzgruppen modelliert. Hierfür werden jeweils die Amplituden jener Bins, die innerhalb einer Frequenzgruppe liegen, addiert und das Ergebnis durch die Anzahl dieser Bins dividiert. Da die Frequenzauflösung bei der 1024-Punkt-FFT ca. 43,1 Hz beträgt, können die exakten Grenzfrequenzen aus Tabelle 1 nicht eingehalten werden. Die folgende Tabelle 2 gibt die Grenzfrequenzen der implementierten Frequenzgruppen an. Außerdem kann darin die Anzahl der FFT-Bins abgelesen werden, die innerhalb der jeweiligen Frequenzgruppe liegen.

Nummer des Bandes in Bark	Untere Frequenzgrenze In Hz	Obere Frequenzgrenze In Hz	Bandbreite In Hz	Anzahl FFT-Bins (1024-Punkt-FFT)
1	0	86,1	86,1	2
2	86,1	172,3	86,1	2
3	172,3	301,5	129,2	3
4	301,5	387,6	86,1	2
5	387,6	516,8	129,2	3
6	516,8	646,0	129,2	3
7	646,0	775,2	129,2	3
8	775,2	904,4	129,2	3
9	904,4	1076,7	172,3	4
10	1076,7	1248,9	172,3	4
11	1248,9	1464,3	215,4	5
12	1464,3	1722,7	258,4	6
13	1722,7	1981,1	258,4	6
14	1981,1	2325,6	344,5	8
15	2325,6	2713,2	387,6	9
16	2713,2	3143,8	430,6	10
17	3143,8	3703,7	559,9	13
18	3703,7	4392,8	689,1	16
19	4392,8	5297,2	904,4	21
20	5297,2	6416,9	1119,7	26
21	6416,9	7708,9	1292,0	30
22	7708,9	9517,7	1808,8	42

23	9517,7	12015,5	2497,8	58
24	12015,5	15503,9	3488,4	81

Tabelle 2: Grenzfrequenzen der implementierten Frequenzgruppen

3.2.2.2 Glättung des zeitlichen Verlaufs

Nach der Bildung der Frequenzgruppen erhalten wir jeweils eine Liste mit den Amplituden der 24 Frequenzbänder für den linken und für den rechten Kanal. Um zu schnelle Änderungen der Amplitude eines Frequenzbandes und damit einhergehende ungewollte klangliche Artefakte zu verhindern, wird der zeitliche Verlauf der Amplituden zusätzlich geglättet. Dies geschieht über Implementierung der Funktion $Y_i(m, k) = \varepsilon Y_i(m - 1, k) + (1 - \varepsilon) X_i(m, k)$. Der Wert für den Faktor ε ist derzeit für alle 24 Frequenzbänder gleich und kann manuell eingegeben werden.

3.2.2.3 Berechnung des Panoramakoeffizienten

Wie in Kapitel 2.3.1 bereits erwähnt, wird der Panoramakoeffizient α in [10] nicht explizit berechnet. Er wird lediglich beschrieben als ein Parameter, der abhängig von der Panoramaposition eines Quellsignals zwischen linkem und rechtem Lautsprecher Werte zwischen Null und Eins annimmt. Implementiert wurde die Berechnung des Panoramakoeffizienten durch die Funktion

$$\alpha(m, k) = \frac{|X_2(m, k)|}{|X_1(m, k)| + |X_2(m, k)|} \quad (13)$$

Diese Funktion nimmt für das Frequenzband k den Wert Null an, wenn $|X_2(m, k)| = 0$ ist, wenn also die Amplitude des k -ten Frequenzbandes im rechten Kanal zum Zeitindex m Null ist. Dies ist gleichbedeutend mit der Panoramisierung dieses Frequenzbandes nach links. Ist $|X_1(m, k)| = 0$, ist die Amplitude eines Frequenzbandes im linken Kanal also Null, so ist dieses Frequenzband nach rechts panoramisiert und $\alpha(m, k)$ nimmt den Wert Eins an. Entsprechend nimmt $\alpha(m, k)$ den Wert 0,5 an, wenn die Amplituden eines

Frequenzbandes in beiden Kanälen gleich groß sind und dieses Frequenzband somit in die Mitte panoramisiert ist.

Die so erhaltenen Werte für α werden für alle 24 Frequenzbänder in ein Array geschrieben, aus welchem sie für die weiteren Berechnungen ausgelesen werden.

3.2.2.4 Zerlegung des Stereosignals

Abhängig vom Wert des Panoramakoeffizienten werden aus dem Stereosignal nun, wie in Kapitel 2.3.2 beschrieben, die vier neuen Signale $X_{11}(m, k)$, $X_{13}(m, k)$, $X_{23}(m, k)$ und $X_{22}(m, k)$ erzeugt. Realisiert wird dies in Pure Data durch eine Funktion, welche die 24 Panoramakoeffizientenwerte ausliest und zwei Ausgänge besitzt. Ausgang 1 liefert den Wert 1 für Zeit-Frequenz-Bereiche, in denen $\alpha(m, k) \leq 0,5$ ist. Ansonsten liefert dieser Ausgang den Wert 0. Ausgang 2 liefert den Wert 1 für Zeit-Frequenz-Bereiche, in denen $\alpha(m, k) \geq 0,5$ ist und sonst den Wert 0. Somit liefern beide Ausgänge den Wert 1 für Regionen, in denen $\alpha(m, k) = 0,5$ ist.

Auf der Audiosignalebene werden nun die STFTs der beiden Kanäle des Stereosignals mit den so erhaltenen Werten multipliziert. Durch Multiplikation von $X_1(m, k)$ mit den Werten von Ausgang 1 erhält man $X_{11}(m, k)$. Multipliziert man $X_2(m, k)$ mit den Werten von Ausgang 1, so erhält man $X_{13}(m, k)$. Entsprechend ergibt Multiplikation von $X_1(m, k)$ mit den Werten von Ausgang 2 das neue Signal $X_{23}(m, k)$ und Multiplikation von $X_2(m, k)$ mit den Werten von Ausgang 2 liefert $X_{22}(m, k)$.

3.2.2.5 Modifizierungsstufe

Die in Kapitel 2.3.3 beschriebene Modifizierungsstufe wird in Pure Data mittels Lookup-Tables realisiert. Diese liefern abhängig vom Eingangswert des Panoramakoeffizienten $\alpha(m, k)$ für jedes Frequenzband den entsprechenden Wert der Modifikationsfunktionen.

Um die Lookup-Tables zu erstellen, müssen die Gleichungen (11) und (12) aus Kapitel 2.3.3 gelöst werden. Im Folgenden wird die Lösung für Werte des

Panoramakoeffizienten $\alpha \leq 0,5$ gezeigt. Die Berechnung für $\alpha \geq 0,5$ erfolgt spiegelbildlich dazu.

Für die Berechnung wird von einer Lautsprecheranordnung wie in Abbildung 6 ausgegangen. Der Öffnungswinkel θ zwischen dem linken und dem rechten Lautsprecher betrage 60° . Der zusätzliche Mittenlautsprecher wird dann symmetrisch auf einer gedachten Kreislinie mit dem Radius r zwischen den linken und den rechten Lautsprecher eingefügt. Hierdurch ergibt sich der Winkel θ' zu 30° . Für die Berechnung wird die 0° -Richtung in Richtung der x-Achse des Koordinatensystems angenommen.

Die Positionen der Lautsprecher relativ zum Zuhörer, welcher sich im Nullpunkt des Koordinatensystems befindet, können dann in Vektorschreibweise für linken, rechten und mittleren Lautsprecher wie folgt dargestellt werden:

$$\text{Links:} \quad s_1 = [x_1, y_1]^T = [r \cos(120^\circ), r \sin(120^\circ)]^T \quad (14)$$

$$\text{Rechts:} \quad s_2 = [x_2, y_2]^T = [r \cos(60^\circ), r \sin(60^\circ)]^T \quad (15)$$

$$\text{Mitte:} \quad s_3 = [x_3, y_3]^T = [r \cos(90^\circ), r \sin(90^\circ)]^T \quad (16)$$

Bei tiefen Frequenzen ($f < 700\text{Hz}$) ergibt sich nach Gleichung (11) folgendes Gleichungssystem für die Repanoramisierung zwischen dem linken und dem mittleren Lautsprecher:

$$\begin{pmatrix} r \cos(120^\circ) & r \cos(60^\circ) \\ r \sin(120^\circ) & r \sin(60^\circ) \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} = \begin{pmatrix} r \cos(120^\circ) & r \cos(90^\circ) \\ r \sin(120^\circ) & r \sin(90^\circ) \end{pmatrix} \begin{pmatrix} a'_1 \\ a'_2 \end{pmatrix} \quad (17)$$

wobei $a_1 = 1 - \alpha$ und $a_2 = \alpha$.

Löst man dieses Gleichungssystem, so erhält man den neuen Panoramakoeffizientenvektor:

$$\begin{pmatrix} a'_1 \\ a'_2 \end{pmatrix} = \begin{pmatrix} 1 - \alpha' \\ \alpha' \end{pmatrix} = \begin{pmatrix} 1 - 2\alpha \\ 2 \sin(120^\circ) \alpha \end{pmatrix}. \quad (18)$$

Für hohe Frequenzen ($f > 700$ Hz) ergibt sich nach Gleichung (12) folgendes Gleichungssystem:

$$\begin{pmatrix} r \cos(120^\circ) & r \cos(60^\circ) \\ r \sin(120^\circ) & r \sin(60^\circ) \end{pmatrix} \begin{pmatrix} p_1 \\ p_2 \end{pmatrix} = \begin{pmatrix} r \cos(120^\circ) & r \cos(90^\circ) \\ r \sin(120^\circ) & r \sin(90^\circ) \end{pmatrix} \begin{pmatrix} p'_1 \\ p'_2 \end{pmatrix} \quad (19)$$

wobei $p_1 = a_1^2 = (1 - \alpha)^2$ und $p_2 = a_2^2 = \alpha^2$.

Für den modifizierten Panoramakoeffizientenvektor erhält man somit für Frequenzen > 700 Hz:

$$\begin{pmatrix} a'_1 \\ a'_2 \end{pmatrix} = \begin{pmatrix} \sqrt{p'_1} \\ \sqrt{p'_2} \end{pmatrix} = \begin{pmatrix} \sqrt{1-2\alpha} \\ \sqrt{2\sin(120^\circ)} \end{pmatrix}. \quad (20)$$

Um die Modifikationsfunktionen zu erhalten, wird nun der Quotient aus den alten und den neuen Koeffizienten gebildet. Es ergeben sich die folgenden Funktionen:

- Für tiefe Frequenzen unterhalb von 700 Hz:

$$M_{11,\text{low}} = \frac{a'_1}{a_1} = \frac{1-2\alpha}{1-\alpha} \quad (21)$$

$$M_{13,\text{low}} = \frac{a'_2}{a_1} = \frac{2\sin(120^\circ)\alpha}{\alpha} = 2\sin(120^\circ) \quad (22)$$

- Für Frequenzen oberhalb von 700 Hz ergeben sich entsprechend:

$$M_{11,\text{hi}} = \frac{a'_1}{a_1} = \frac{\sqrt{1-2\alpha}}{1-\alpha} \quad (23)$$

$$M_{13,\text{hi}} = \frac{a'_2}{a_1} = \frac{\alpha\sqrt{2\sin(120^\circ)}}{\alpha} = \sqrt{2\sin(120^\circ)} \quad (24)$$

Um die Modifikationsfunktionen für Werte des Panoramakoeffizienten $\alpha \geq 0,5$ zu erhalten, werden die entsprechenden Gleichungssysteme für die Repanoramisierung zwischen dem mittleren und dem rechten Lautsprecher aufgestellt und gelöst. Man erhält:

- Unterhalb von 700 Hz:

$$M_{22,\text{low}} = \frac{2\alpha - 1}{\alpha} \quad (25)$$

$$M_{23,\text{low}} = \frac{2\sin(120^\circ)(1-\alpha)}{1-\alpha} = 2\sin(120^\circ) \quad (26)$$

- Sowie oberhalb von 700 Hz:

$$M_{22,\text{hi}} = \frac{\sqrt{2\alpha - 1}}{\alpha} \quad (27)$$

$$M_{23,\text{hi}} = \frac{(1-\alpha)\sqrt{2\sin(120^\circ)}}{1-\alpha} = \sqrt{2\sin(120^\circ)} \quad (28)$$

Um die wahrgenommene Richtung der Schallquellen bei Wiedergabe des Stereosignals über drei Lautsprecher beizubehalten, werden die aus der Zerlegung des Stereosignals erhaltenen Signale $X_{11}(m, k)$, $X_{13}(m, k)$, $X_{23}(m, k)$ und $X_{22}(m, k)$ mit den entsprechenden Modifikationsfunktionen multipliziert. Schaubilder zu den Modifikationsfunktionen finden sich in Form der Lookup-Tables im Anhang (Abbildung 35).



4 Hörtests

Um die Qualität des implementierten Upmixalgorithmus einschätzen zu können, wurden Upmixes von diversen Stereosignalen durchgeführt und die so erhaltenen Fünfkkanalmixes mit den Originalsignalen verglichen. Die Ergebnisse dieser Tests beruhen auf rein subjektiven Hörempfindungen. Es wurde jedoch versucht, diese subjektiven Höreindrücke mittels Spektrogrammdarstellungen¹² der einzelnen Kanäle zu stützen. Durch eine solche Darstellung kann insbesondere die Qualität der Repanoramisierung gezeigt werden während sich die Halleextraktion visuell nur schlecht darstellen lässt. In den folgenden Abbildungen ist die Frequenzachse logarithmisch aufgetragen. Die Pegelangabe erfolgt in dBFS und ist unbewertet.

4.1 Weißes Rauschen, linker und rechter Kanal unkorreliert

DVD Ordner: 03_AudioExamples -> 01_WhiteNoise

2.0: Original 2-Kanal Signal, 5.0: Upmixed 5-Kanal Signal

Als erstes Quellsignal diente weißes Rauschen, wobei zwei verschiedene Rauschquellen für den linken und für den rechten Kanal verwendet wurden, so dass es zwischen den beiden Kanälen des Originalsignals nur schwache Korrelation gibt. Hierdurch sollte insbesondere die Halleextraktion getestet werden, da diese auf der Kohärenz der einzelnen Frequenzbänder zwischen linkem und rechtem Kanal beruht.

Ergebnis 2-auf-3-Kanal Upmix:

Das resultierende 3-Kanal-Signal erscheint verglichen mit dem 2-Kanal-Signal viel mehr aus der Mitte zu kommen. Mit einem Peakmeter wurden ca. 15 dB weniger Pegel auf dem linken und dem rechten Kanal gemessen als auf dem Mittenkanal.

¹² Als Spektrogramm bezeichnet man die Darstellung von Lautstärkepegel über der Frequenz über der Zeit in einem dreidimensionalen Diagramm, wobei die Lautstärke farblich codiert ist.

Es wurden außerdem verschiedene Einstellungen für den zeitlichen Glättungsfaktor ε ausprobiert. Dabei hat sich gezeigt, dass bei niedrigeren Werten für ε die Stereoabbildung etwas breiter wird, da der Repanoramisierungsalgorithmus dann schneller auf Änderungen des Eingangssignals reagiert. Ein ungewollter Nebeneffekt sind jedoch Artefakte die hierbei entstehen, wenn die Panoramapositionen für einzelne Frequenzbänder zu schnell wechseln. Als guter Wert für ε hat sich ein Wert von 0,9 erwiesen, welcher auch für die weiteren Tests beibehalten wurde. Bei Verwendung dieser Einstellung sind immer noch deutliche Artefakte hörbar, wenn nur der linke und der rechte Kanal abgehört werden. Im Zusammenklang aller drei Frontkanäle werden diese aber verdeckt und treten nicht mehr störend in Erscheinung. Wie erwähnt wird die Stereoabbildung bei Verwendung dieses Wertes jedoch deutlich schmaler.

Ergebnis Hallextraktion:

Da der implementierte Hallextraktionsalgorithmus auf der Kohärenz zwischen linkem und rechtem Kanal basiert, war zu erwarten dass bei zwei unkorrelierten Rauschsignalen für den linken und rechten Kanal das gesamte Eingangssignal als Hallsignal interpretiert wird. Diese Erwartung hat sich auch bestätigt; das Eingangssignal wird bis auf einen Verstärkungsfaktor unverändert auf die beiden Surroundkanäle weitergeleitet. Dabei macht es keinen Unterschied, ob zur Hallextraktion der Hallindex Φ oder die nichtlineare Funktion des Hallindex $\Gamma[\Phi]$ verwendet wird.

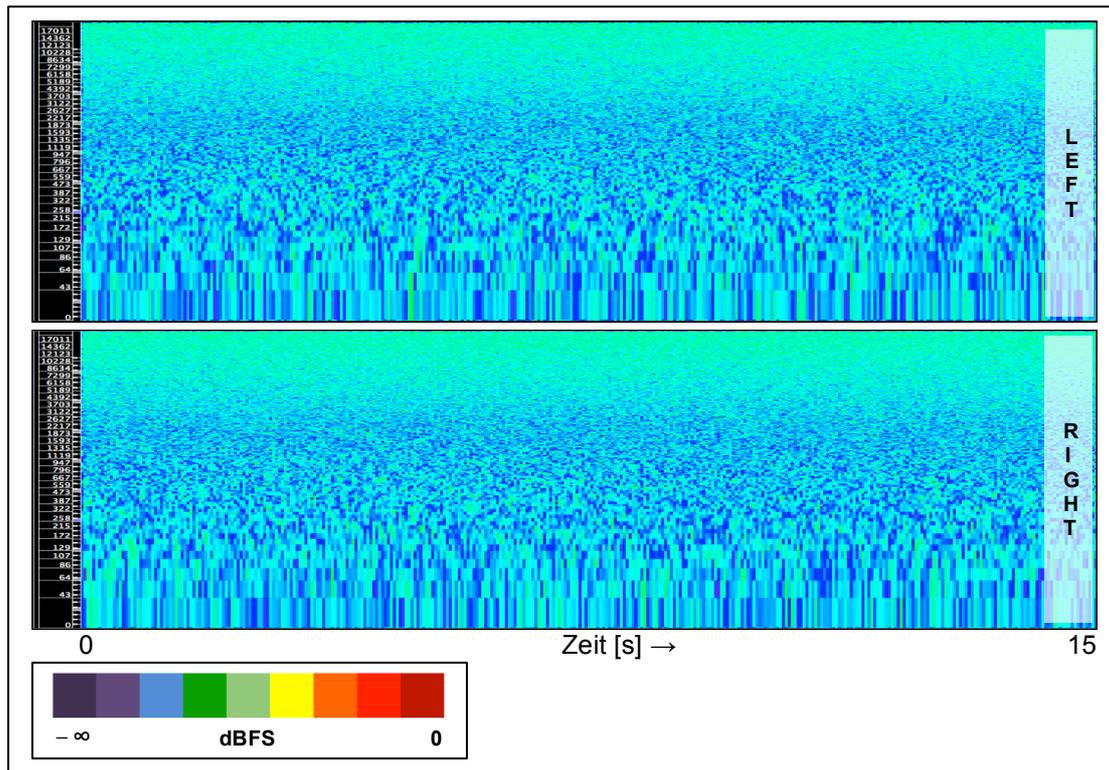


Abb.8: Weißes Rauschen (15 Sekunden), Originalsignal

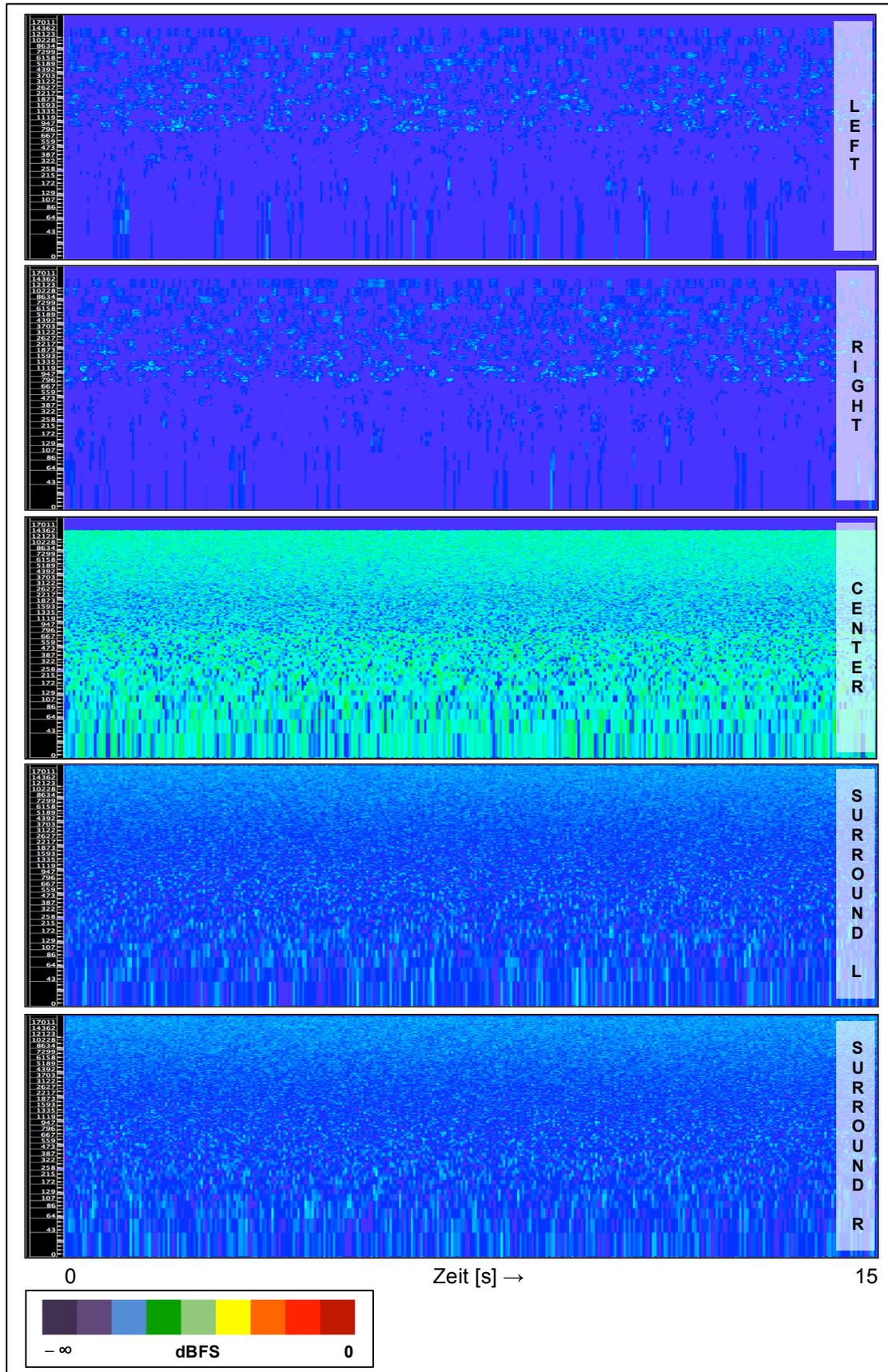


Abb.9: Weißes Rauschen nach Upmixprozess

4.2 Einzelne Instrumente

4.2.1 Drumloop mit Stereohall

DVD Ordner: 03_AudioExamples -> 02_Drumloop

2.0: Original 2-Kanal Signal, 5.0: Upmixed 5-Kanal Signal

Ergebnis 2-auf-3-Kanal Upmix:

Der Upmix eines einfachen Drumloops mit Hall bestätigt und erweitert die Erkenntnisse aus dem Hörtest mit weißem Rauschen. Durch den 2-auf-3-Kanal Upmix wird die Stereoabbildung erheblich schmaler. Gleichzeitig verbessert sich hierdurch jedoch die Ortung der einzelnen Elemente. Besonders die Snare, deren Position im Zweikanalmix aufgrund des Hallanteils nicht eindeutig bestimmt werden kann, erhält durch den Upmix einen festen Platz im Stereobild.

Ergebnis Hallextraktion:

Bei Verwendung des Hallindex Φ ergibt sich ein relativ großer Surroundanteil aufgrund des Stereohalls im Originalsignal. Es wurde außerdem mit verschiedenen Werten für den Faktor λ experimentiert, wobei sich für dieses Klangbeispiel ein Wert von $\lambda = 0,9$ als guter Kompromiss zwischen Direktsignalanteilen im Surround einerseits und klanglichen Artefakten durch zu schnelle Änderungen des Hallindex erwiesen hat.

Bei Verwendung der Mappingfunktion $\Phi[\lambda]$ ergibt sich generell ein wesentlich geringerer Surroundanteil. Allerdings sind bei Verwendung dieser Funktion sehr viele störende Artefakte wahrnehmbar, die aus zu großen Amplitudendifferenzen von benachbarten Bins resultieren. Auch bei den weiteren Tests hat sich herausgestellt, dass die nichtlineare Mappingfunktion erhebliche Artefakte hervorruft, die auch im Zusammenklang aller fünf Lautsprecher noch störend hervortreten. Deshalb wurde für die folgenden Tests nur noch der Hall-Index verwendet.

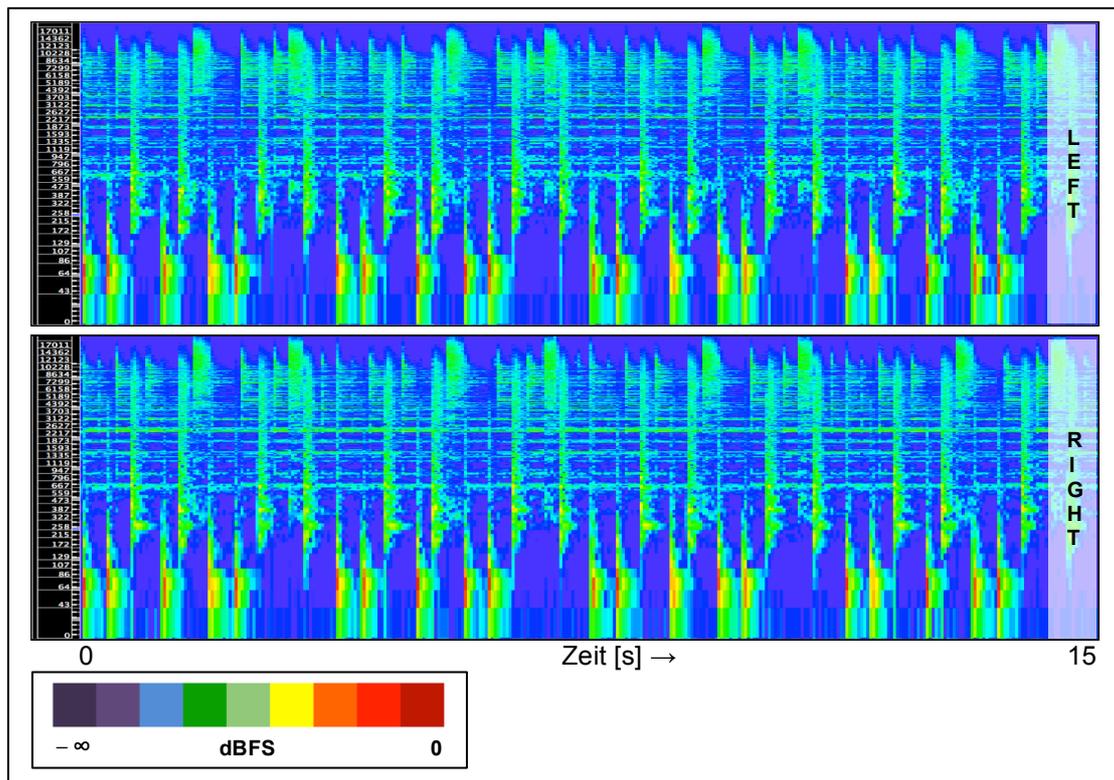
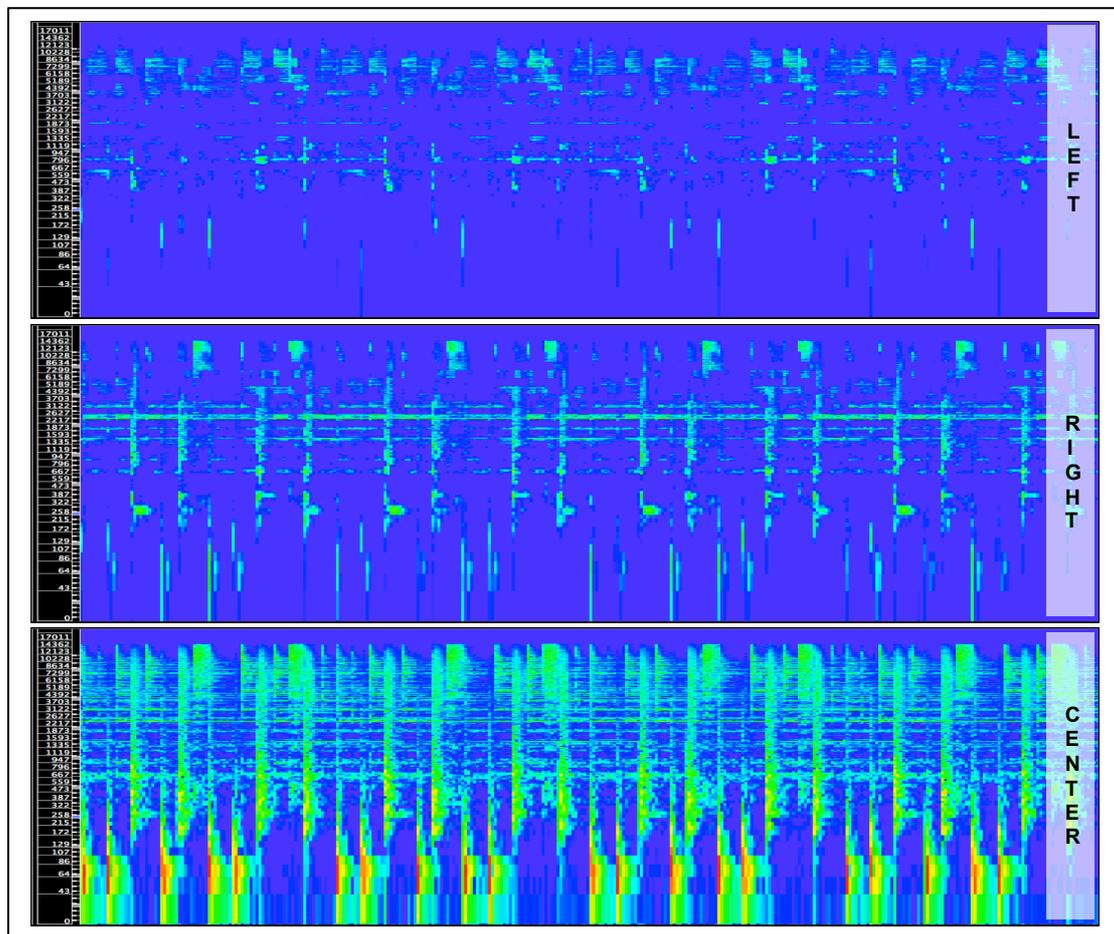


Abb.10: Drumloop (15 Sekunden), Originalsignal



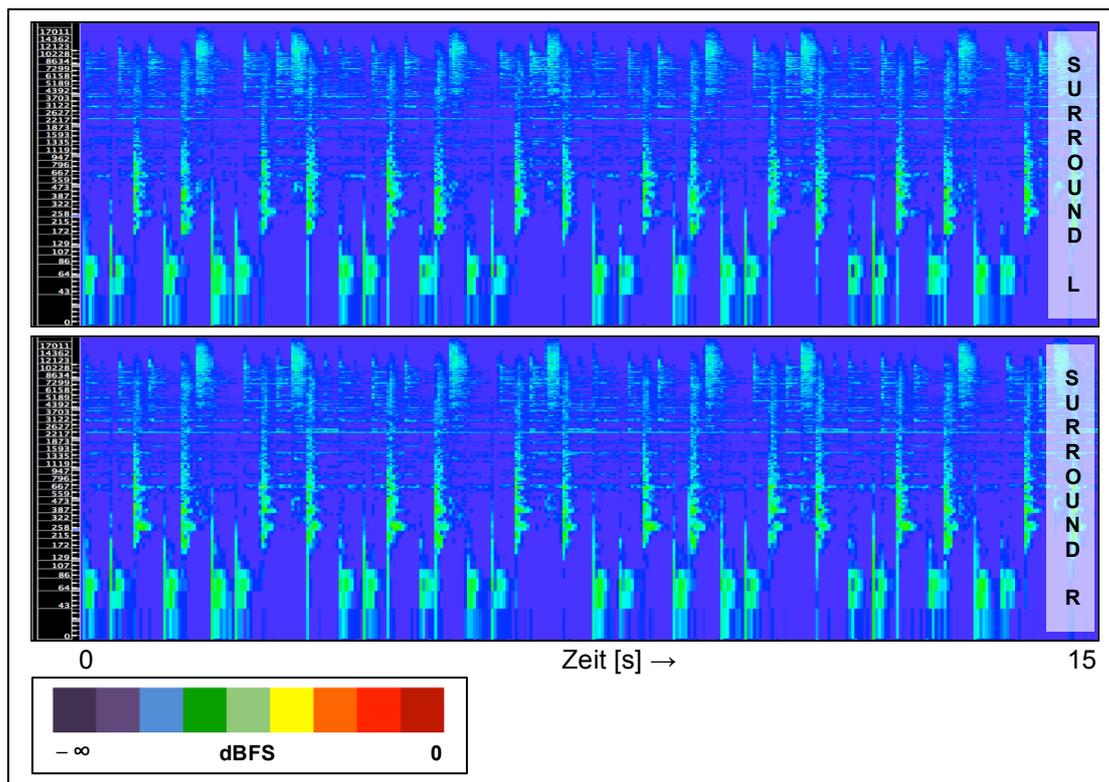


Abb.11: (mit S. 38 unten): Drumloop nach Upmixprozess

4.2.2 Sologitarre

DVD Ordner: 03_AudioExamples -> 03_GuitarSolo

2.0: Original 2-Kanal Signal, 5.0: Upmixed 5-Kanal Signal

Panoramakoeffizient $\alpha = 0,25$

Ergebnis 2-auf-3-Kanal Upmix:

Die wahrgenommene Panoramaposition beim Upmix einer einzelnen Klangquelle wie in diesem Beispiel bleibt erhalten, wenn sich der Zuhörer im Sweetspot der Lautsprecheranordnung befindet. Bewegt sich der Zuhörer seitlich aus dem Sweetspot heraus, so wird die Ortung der Klangquelle durch den Upmix erheblich verbessert. Damit ist eine der in Kapitel 1.2 gestellten Anforderungen an ein Upmixsystem bei Vorhandensein einer einzelnen Klangquelle erfüllt.

Weiters wurde bei diesem Test festgestellt, dass der Klang durch den Upmixvorgang etwas an Höhen und tiefen Mitten verliert. Dies kann jedoch

durch einen Equalizer im Mittenkanal ausgeglichen werden. Folgende Einstellungen für den Equalizer werden vorgeschlagen:

+2,5 dB @ 140 Hz, Q = 1,0 sowie +3,5 dB @ 6,3 kHz, Q = 1,1

Es hat sich gezeigt, dass der wahrgenommene Mitten- und Höhenverlust durch diese Einstellungen kompensiert werden kann.

Ergebnis Hallextraktion:

Da bei diesem Beispiel eine einzelne Monoklangquelle verwendet wurde, ist die Kohärenz zwischen linkem und rechtem Kanal maximal und der extrahierte Hallanteil ist null.

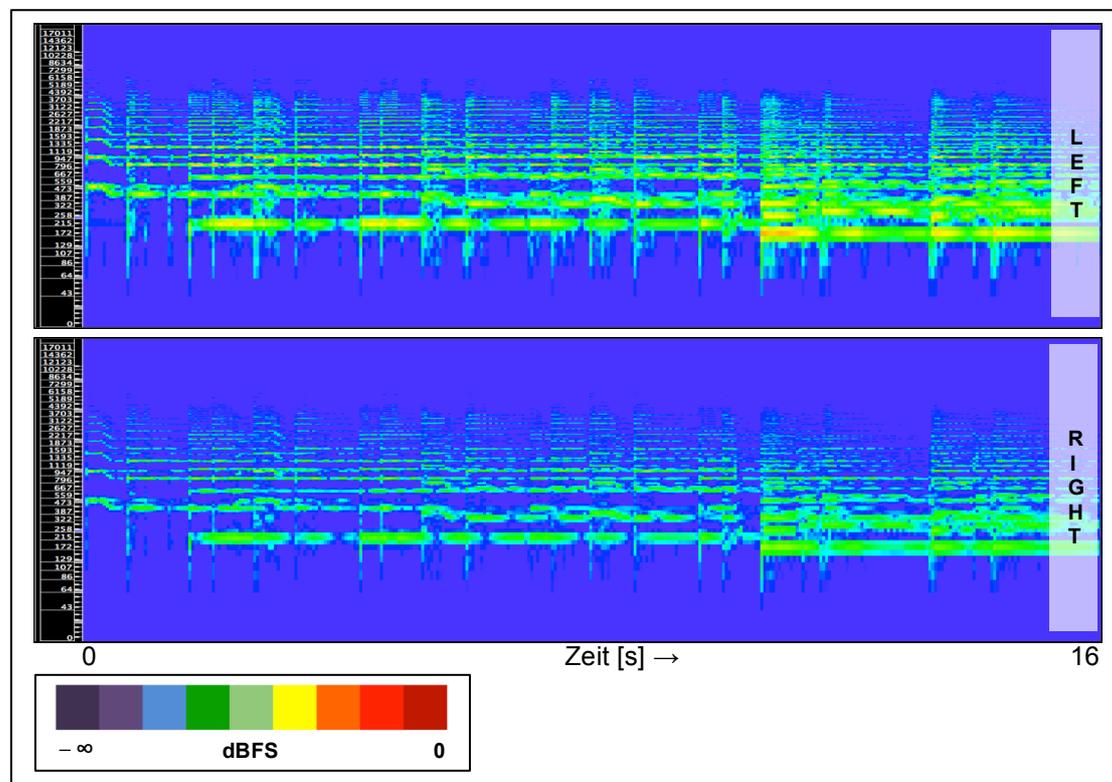


Abb.12: Einzelne Gitarre halblinks, Originalsignal

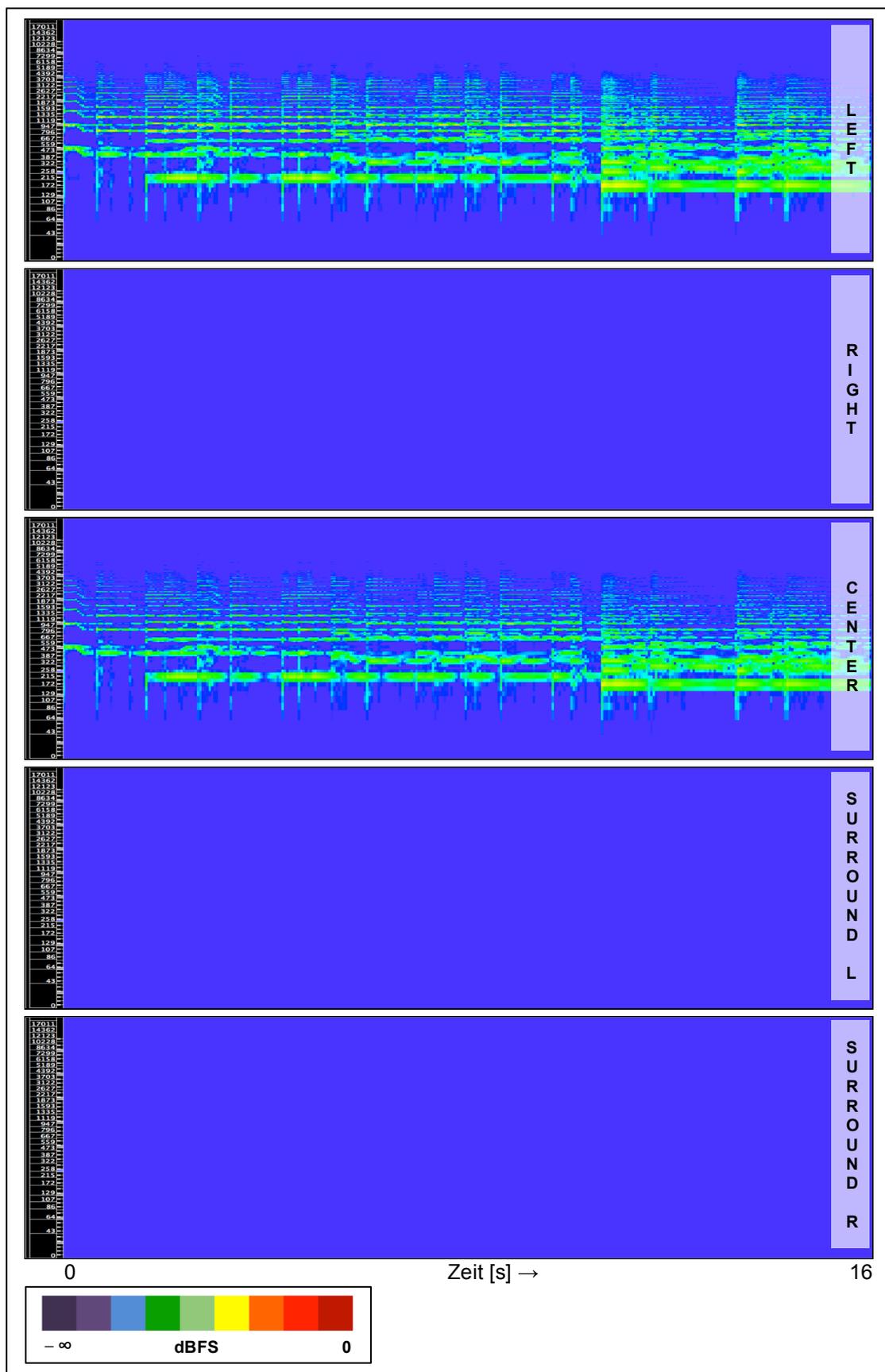


Abb.13: Einzelne Gitarre halblinks, Signal nach Upmix

4.2.3 Gitarrenduo

DVD Ordner: 03_AudioExamples -> 04_GuitarDuo

2.0: Original 2-Kanal Signal, 5.0: Upmixed 5-Kanal Signal

Panoramakoeffizient Gitarre 1: $\alpha_1 = 0,25$

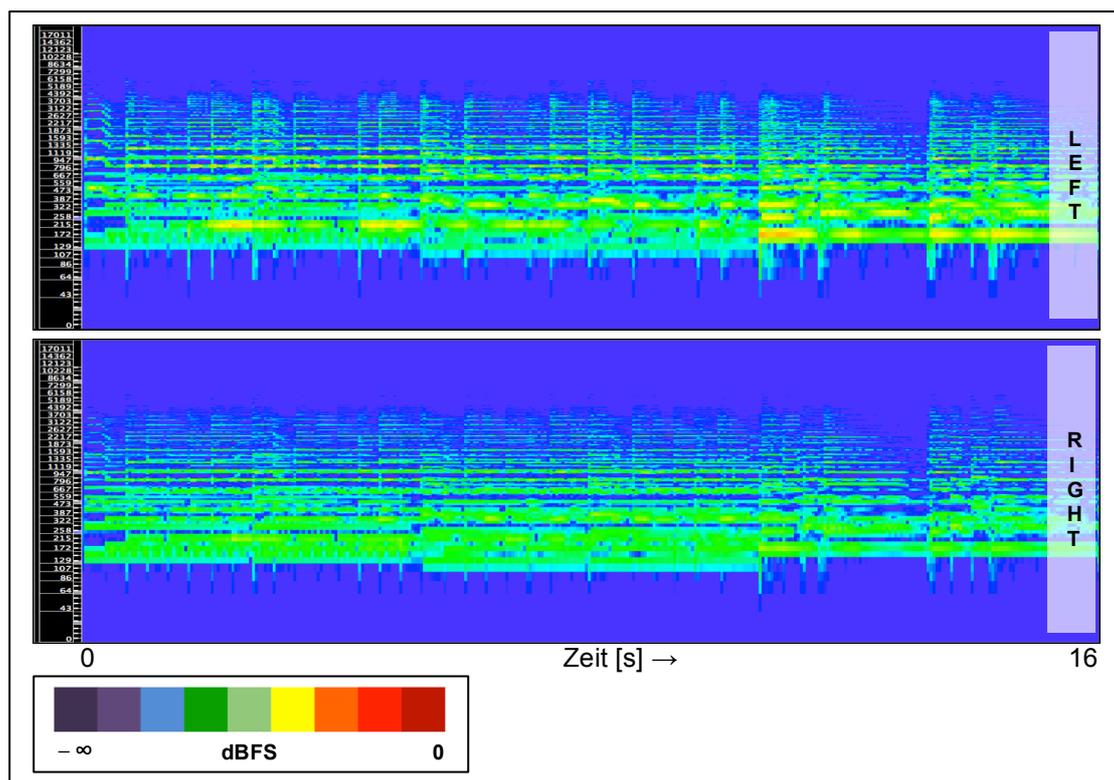
Panoramakoeffizient Gitarre 2: $\alpha_2 = 0,6$

Ergebnis 2-auf-3-Kanal Upmix:

Bei diesem Test gelangt der implementierte Algorithmus an seine Grenzen, da eine der Grundvoraussetzungen für seine Funktion nicht gegeben ist. Sind in den beiden Eingangskanälen sich überlappende Zeit-Frequenz-Regionen zweier verschiedener Klangquellen vorhanden, so ist die resultierende Repanoramisierung unbefriedigend. So verschiebt sich beispielsweise eine Klangquelle im Panorama, wenn eine weitere Klangquelle mit anderem Panoramakoeffizienten im gleichen Frequenzbereich hinzukommt. Durch diese plötzliche Verschiebung im Panorama entstehen Artefakte wie z.B. Knackser.

Ergebnis Hallextraktion:

In den Surroundkanälen entstehen bei diesem Beispiel störende Artefakte.



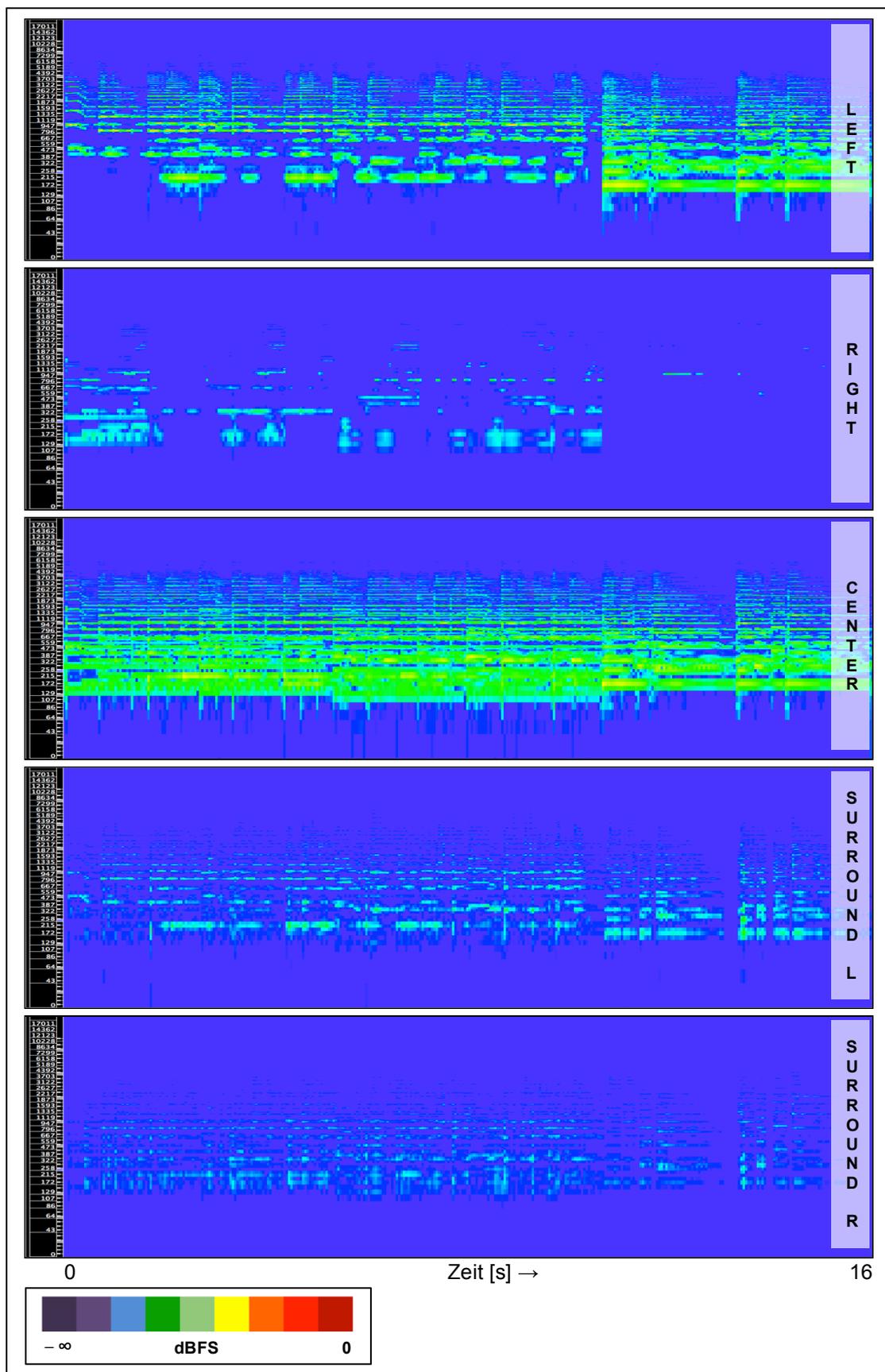


Abb.15: Gitarrenduo, nach Upmix

4.3 Diverse Musikstücke

Die folgenden Beispiele auf der beiliegenden DVD sollen dazu dienen, dem Leser einen Eindruck der Vor- und Nachteile des implementierten Upmixalgorithmus zu verschaffen. Es sind vier Musikbeispiele vorhanden sowie drei charakteristische Filmtönebeispiele für die Bereiche Dialog, Action und Musik.

4.3.1 Pink Floyd - Wearing the Inside Out

DVD Ordner: 03_AudioExamples -> 05_PinkFloyd

2.0: Original 2-Kanal Signal, 5.0: Upmixed 5-Kanal Signal

Sehr räumlicher Mix mit breiter Stereoabbildung und ausgeprägter Tiefenstaffelung.

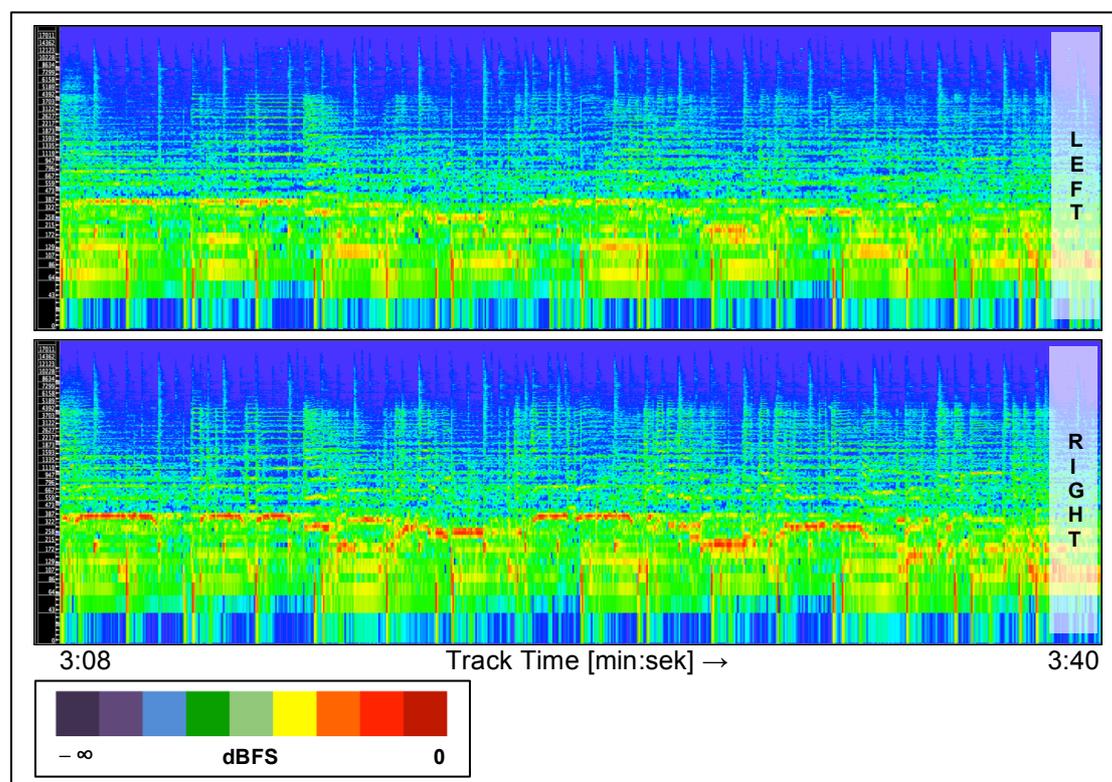


Abb.16: Auszug aus "Wearing the Inside Out", Originalsignal

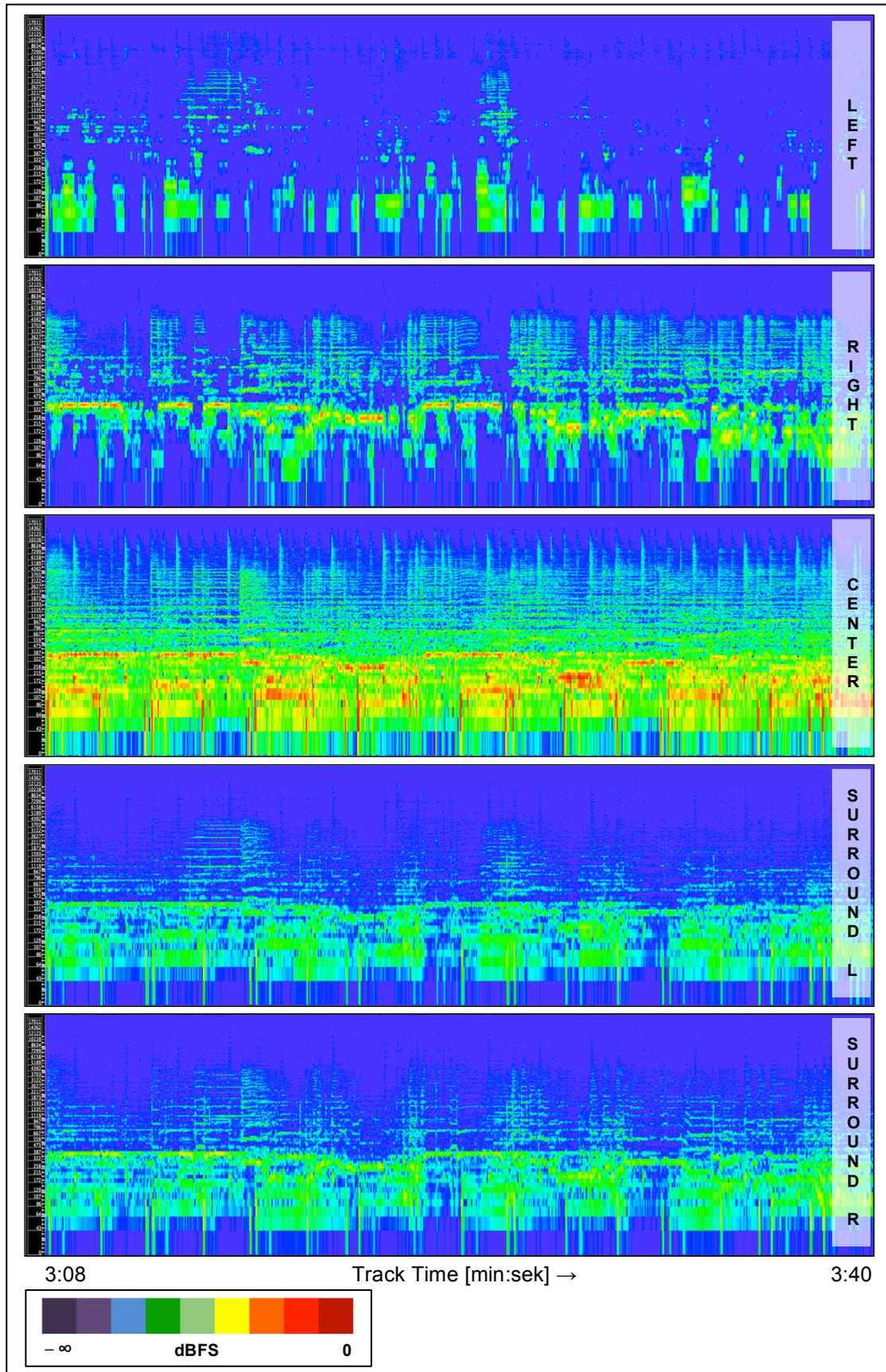


Abb.17: Auszug aus "Wearing the Inside Out", Signal nach Upmix

4.3.2 Steve Reich - Music for Mallet Instruments, Voices and Organ

DVD Ordner: 03_AudioExamples -> 06_SteveReich

2.0: Original 2-Kanal Signal, 5.0: Upmixed 5-Kanal Signal

Ein Stück mit vielen perkussiven Elementen, die weit im Stereopanorama verteilt sind.

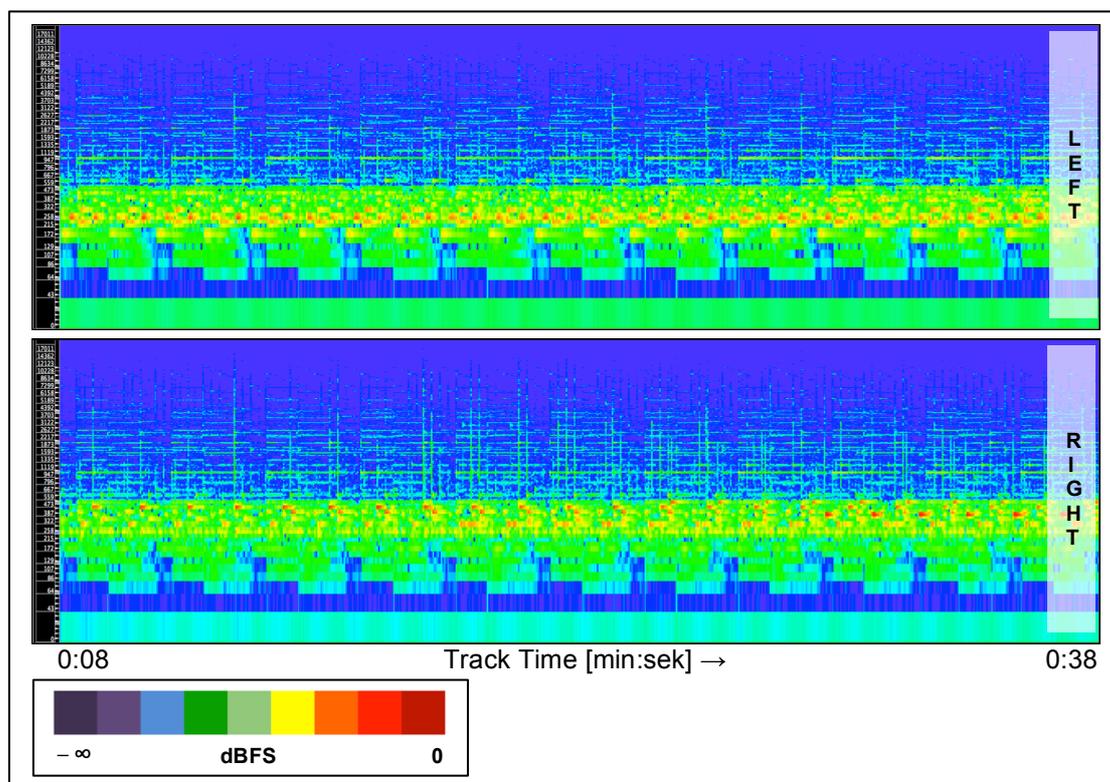


Abb.18: Auszug aus "Music for Mallet Instruments...", Originalsignal

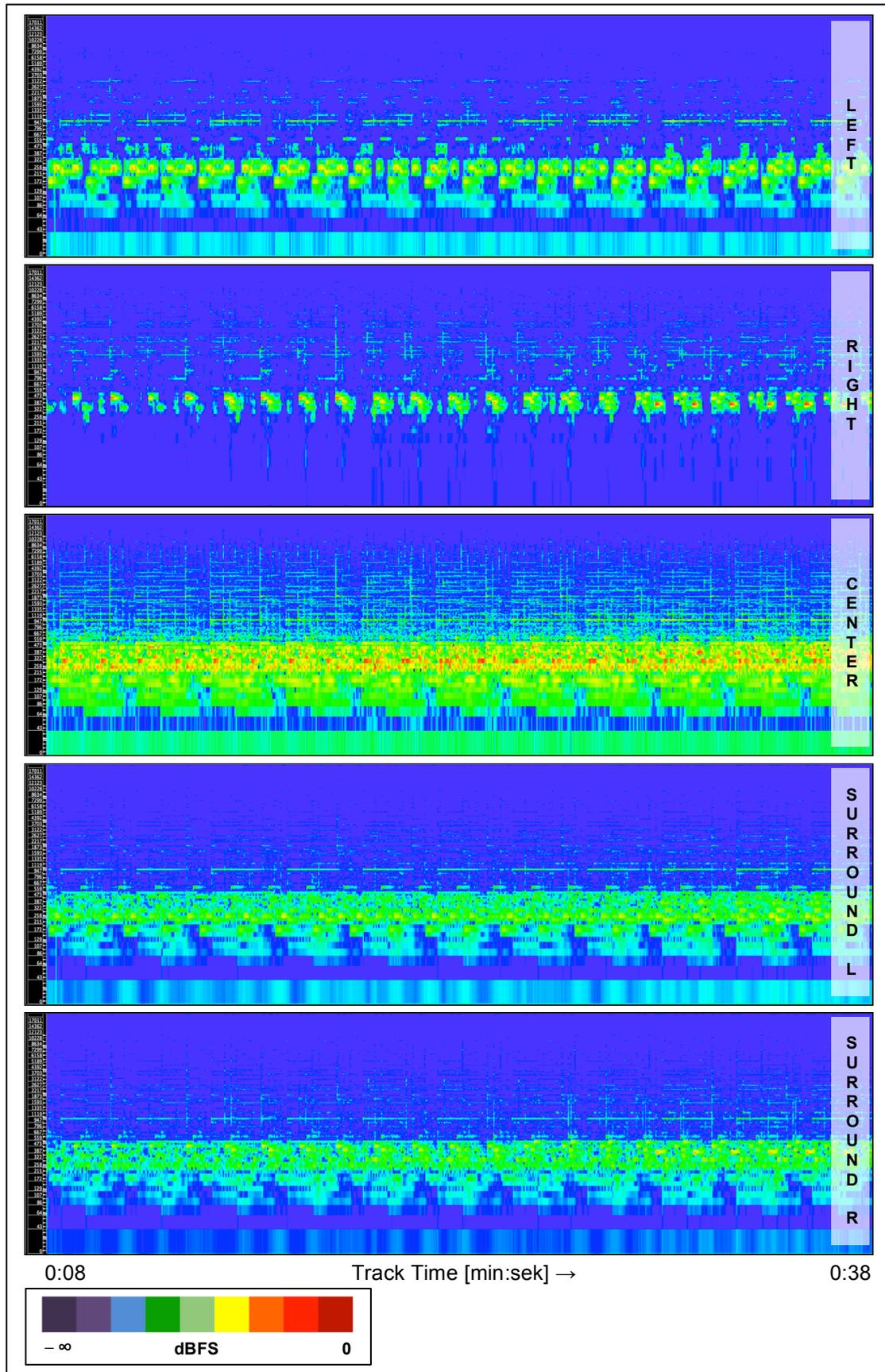


Abb.19: Auszug aus "Music for Mallet Instruments...", Signal nach Upmix

4.3.3 Charles Mingus - Solo Dancer, Stop! Look! And Listen, Sinner Jim Whitney!

DVD Ordner: 03_AudioExamples -> 07_CharlesMingus

2.0: Original 2-Kanal Signal, 5.0: Upmixed 5-Kanal Signal

Einzelne Instrumente sind in diesem Mix weit im Stereopanorama verteilt. Bass und Schlagzeug sind in die Mitte panoramisiert.

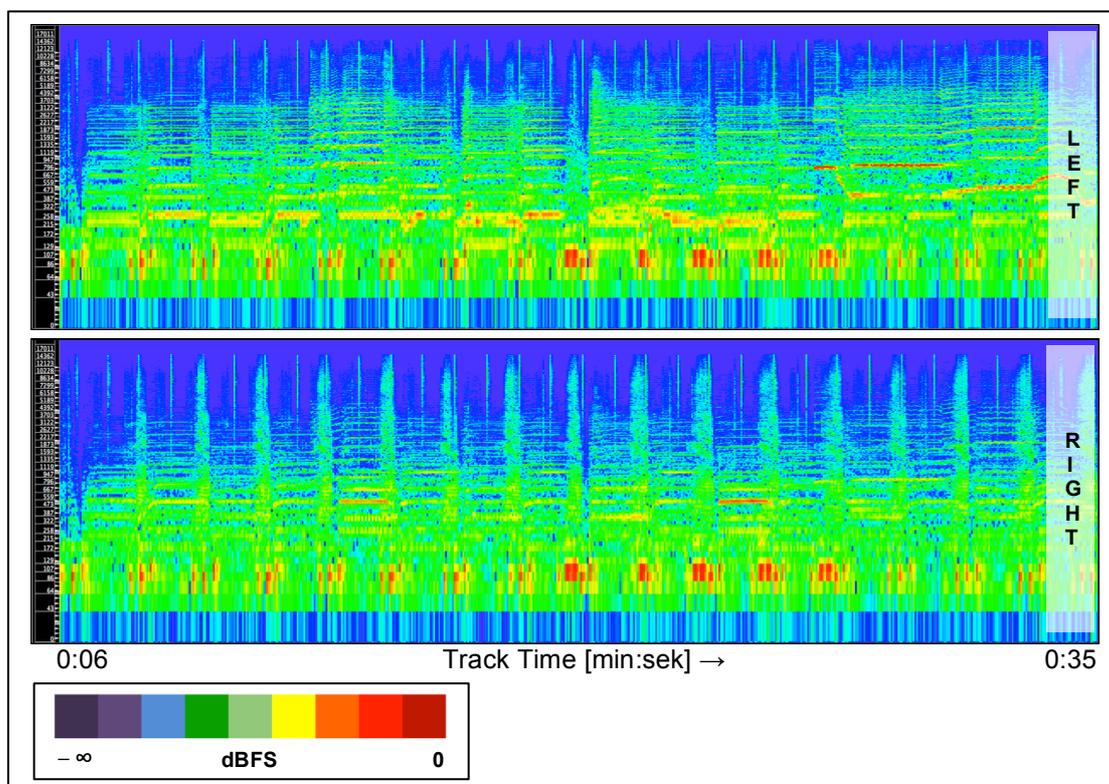


Abb.20: Auszug aus "Solo Dancer...", Originalsignal

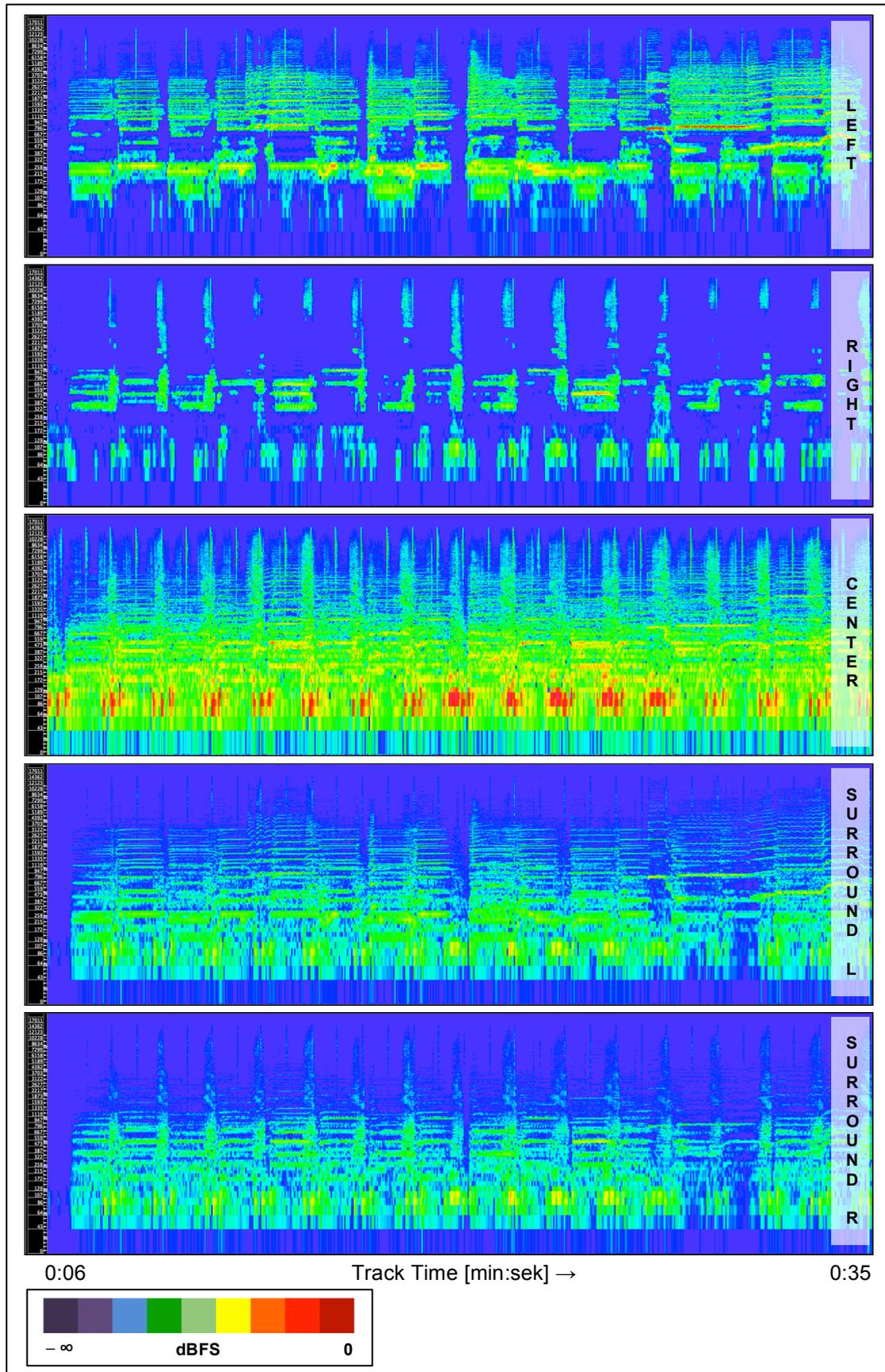


Abb.21: Auszug aus "Solo Dancer...", Signal nach Upmix

4.3.4 Dusty Springfield – Son of a Preacher Man

DVD Ordner: 03_AudioExamples -> 08_DustySpringfield

2.0: Original 2-Kanal Signal, 5.0: Upmixed 5-Kanal Signal

Sogenanntes Beatles-Stereo mit sehr weit nach außen panoramisierten Instrumenten. Bass und Gitarre sind im Originalmix fast nur im linken, Schlagzeug und Blasinstrumente fast nur im rechten Lautsprecher zu hören. Einzig die Hauptgesangsstimme ist in die Mitte panoramisiert.

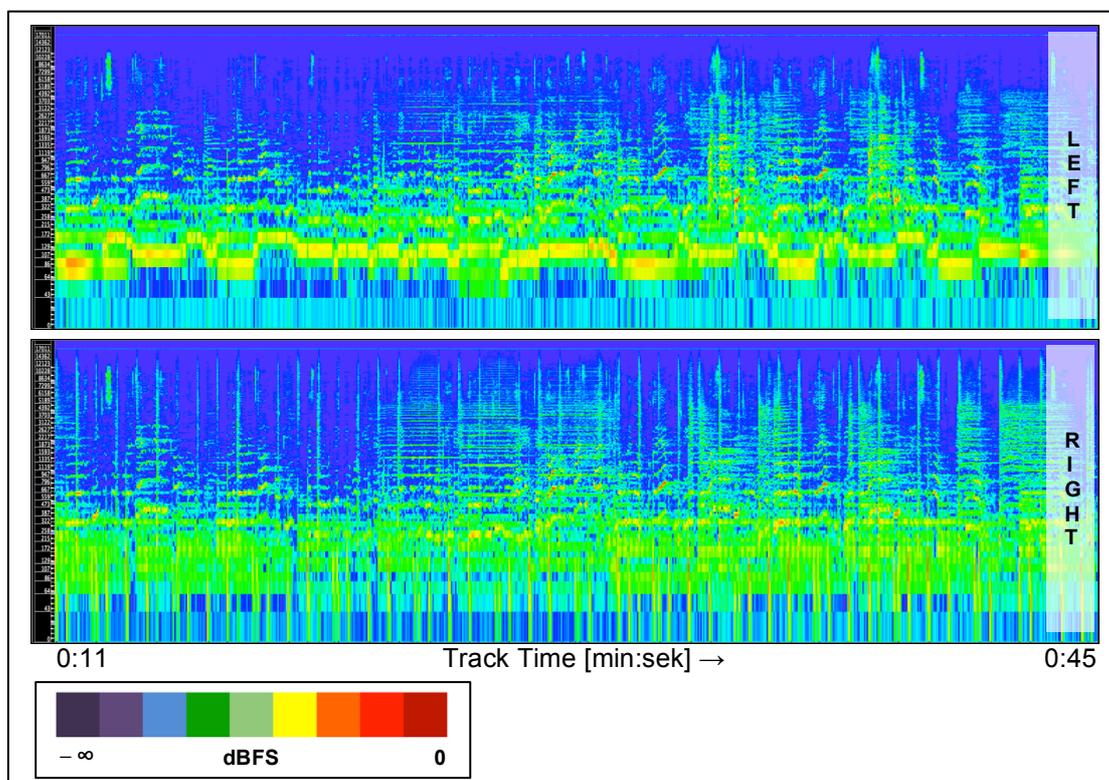


Abb.22: Auszug aus "Son of a Preacher Man", Originalsignal

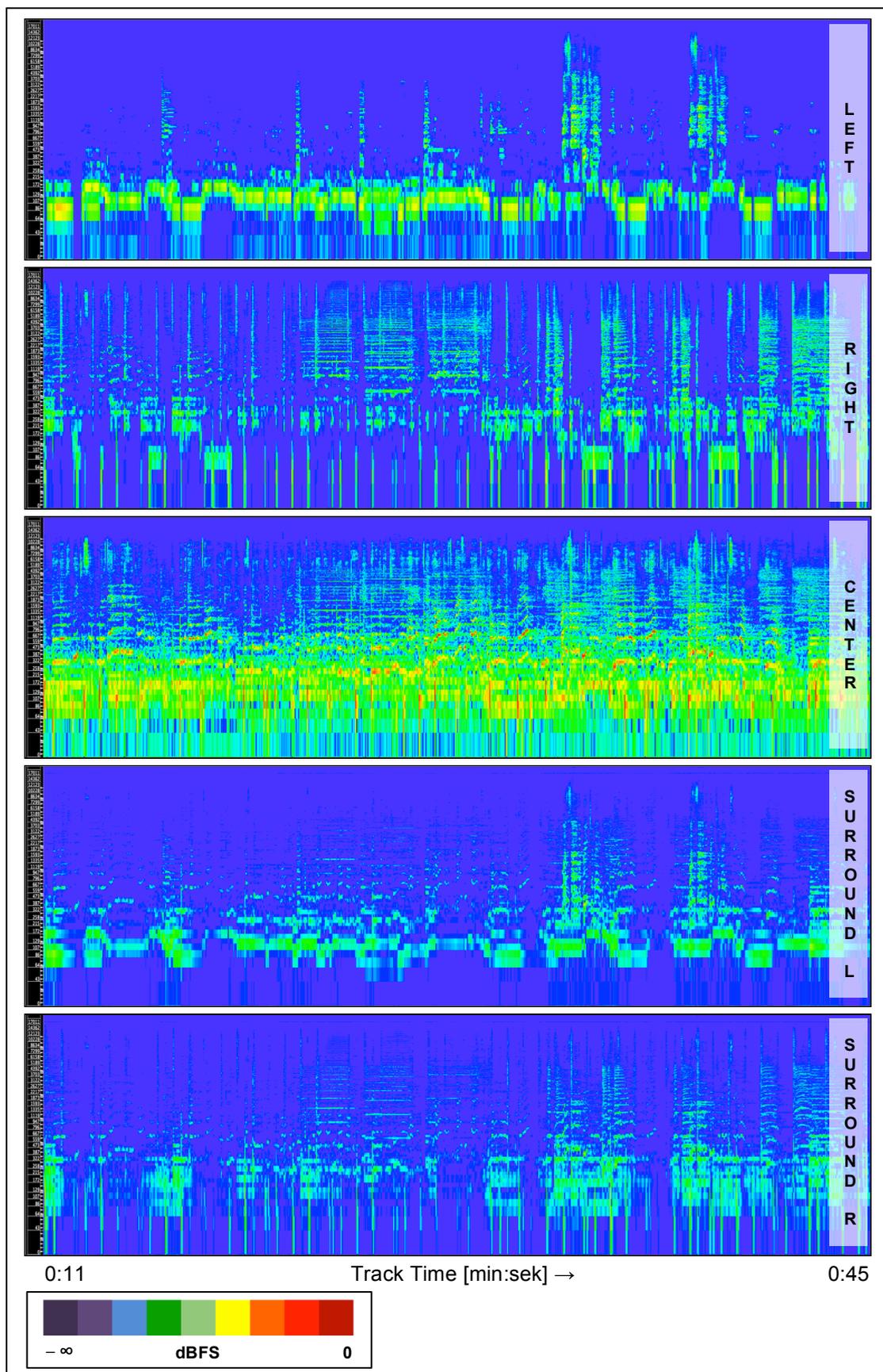


Abb.23: Auszug aus "Son of a Preacher Man", Signal nach Upmix

4.4 Filmtonauszüge

4.4.1 Musik und Voice Over (aus: The Big Lebowski)

DVD Ordner: 03_AudioExamples -> 09_TheBigLebowski

2.0: Original 2-Kanal Signal, 5.0: Upmixed 5-Kanal Signal

4.4.2 Dialog, Atmo, Geräusche (aus: Wer früher stirbt ist länger tot)

DVD Ordner: 03_AudioExamples -> 10_WerFrueherStirbtIstLaengerTot

2.0: Original 2-Kanal Signal, 5.0: Upmixed 5-Kanal Signal

4.4.3 Action und Musik (aus: Inception)

DVD Ordner: 03_AudioExamples -> 11_Inception

2.0: Original 2-Kanal Signal, 5.0: Upmixed 5-Kanal Signal

5 Zusammenfassung und Ausblick

Die Hörversuche haben große Schwachstellen des implementierten Programms vor allem im Bereich der Hallextraktion und –synthese von komplexem Audiomaterial aufgezeigt. Um diese Schwachstellen zu beseitigen sollte als erstes die Frequenzauflösung der STFT erhöht werden. Die derzeit verwendete FFT-Blockgröße von 64 ist zu klein, benachbarte Frequenzbins weisen, abhängig vom Eingangssignal, teilweise sehr große Differenzen in der Amplitude auf. Dies führt zu deutlich hörbaren und störenden Artefakten. Ergänzend zur Erhöhung der Frequenzauflösung kann außerdem eine Interpolation der Amplituden von benachbarten Bins durchgeführt werden. Hier würde sich beispielsweise die Spline-Interpolation anbieten.

Die Zwei- auf Dreikanal-Repanoramisierung funktioniert im Vergleich dazu gut; ungewollte Artefakte ergeben sich im Wesentlichen bei Überlappung von Zeit-Frequenz-Regionen verschiedener Klangquellen im zu bearbeitenden Zweikanalmix. Dies ist aber eine prinzipbedingte Schwäche des implementierten Systems, da Frequenzbereiche unabhängig von deren Zugehörigkeit zu einer bestimmten Klangquelle repanoramisiert werden. Es existieren bereits Upmixsysteme, die auf der Erkennung der einzelnen Klangquellen eines Stereomixes basieren (auch ein solches System wird in [10] skizziert). Dies ist jedoch ein grundlegend anderer Ansatz als der dieser Arbeit zugrundeliegende und kann somit nur bedingt für die Optimierung des implementierten Algorithmus herangezogen werden.



6 Literaturverzeichnis

- [1] M. A. Gerzon, "Optimum Reproduction Matrices for Multispeaker Stereo", *AES 91st Convention*, New York 1991
- [2] M. A. Gerzon, "General Metatheory of Auditory Localisation", *AES 92nd Convention*, Vienna 1992
- [3] M. T. Miles, "An Optimum Linear-Matrix Stereo Imaging System", *AES 101st Convention*, preprint 4364(J-4), 1996
- [4] R. Dressler, "Dolby Surround Pro Logic Decoder Principles of Operation", <ftp://ftp.ema.edu.eg/pub/Priit/5.1/Dolby%20Papers/1%20Dolby%20Pro-Logic/Dolby%20Pro-Logic%20Decoder%20Principles.pdf>
- [5] C. Faller, "Multiple-Loudspeaker Playback of Stereo Signals", *Journal of the Audio Engineering Society*, Vol. 54, No. 11, November 2006
- [6] A. Walther und C. Faller, "Direct-Ambient Decomposition and Upmix of Surround Signals", *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, New Paltz October 2011
- [7] O. Thiergart et al., "Localization of Sound Sources in Reverberant Environments Based on Directional Audio Coding Parameters", *AES 127th Convention*, New York 2009
- [8] G. Del Galdo, O. Thiergart et al., "Optimized Parameter Estimation in Directional Audio Coding Using Nested Microphone Arrays", *AES 127th Convention*, New York 2009

- [9] E. Vickers, "Frequency-Domain Two- to Three-Channel Upmix for Center Channel Derivation and Speech Enhancement", *AES 127th Convention*, New York 2009
- [10] C. Avendado, J. M. Jot, "Frequency Domain Techniques for Stereo to Multichannel Upmix", *AES 22nd International Conference on Virtual, Synthetic and Entertainment Audio*, Espoo, Finland 2002
- [11] C. Avendado, J. M. Jot, "Ambience Extraction and Synthesis from Stereo Signals for Multi-Channel Up-Mix", *IEEE International Conference on Acoustics, Speech and Signal Processing*, ICASSP'02, Orlando, FL 2002
- [12] J. Kreidler, "Loadbang", *Wolke Verlagsgesellschaft*, 2009
- [13] http://de.wikipedia.org/wiki/Dolby_Pro_Logic, Abrufdatum 15.07.2012
- [14] http://en.wikipedia.org/wiki/Pure_data, Abrufdatum 02.06.2012
- [15] <http://en.flossmanuals.net/puredata/>, Abrufdatum 02.06.2012
- [16] <http://de.wikipedia.org/wiki/Frequenzgruppe>, Abrufdatum 09.06.2012

7 Abbildungsverzeichnis

*Abb.1: Schematische Darstellung der Dolby Codierungsmatrix (Seite 5),
Quelle: [13]*

*Abb.2: Funktionsweise einer einfachen Decodierungsmatrix zur
Gewinnung des Surroundsignals (Seite 6), Quelle: [13]*

*Abb.4: Schematische Darstellung der Hallextraktion und –synthese (Seite 13),
Quelle: [10], S.5*

*Abb.5: Funktionsweise des Zwei- auf Dreikanal Upmix (Seite 15),
Quelle: [10], S.7*

*Abb.6: Koordinatensystem für Lautsprecheranordnung (Seite 16),
Quelle: [10], S.8*

*Abb.7: Blockdiagramm Zwei- auf Fünfkanal Upmixsystem (Seite 18),
Quelle: [10], S.9*



8 Anhang

8.1 Was ist Pure Data?

Pure Data, oder kurz Pd, ist eine grafische Echtzeitprogrammiersprache für Audio-, Midi-, Video- und grafische Anwendungen. Sie wurde in den 1990er Jahren von Miller Puckette entwickelt, um interaktive Computermusik und Multimediaprojekte zu realisieren [13]. Da es sich um ein Open Source Projekt handelt, existieren zahlreiche Erweiterungen, sog. „Externals“, welche die Funktionalität der Originalversion, genannt „Vanilla“, erheblich ausbauen. Bei der für diese Bachelorarbeit verwendeten Version, Pd 0.42.5-extended, sind viele der verfügbaren Externals bereits standardmäßig enthalten, so z.B. die extensiv verwendete zexy-Bibliothek.

Die Programmierung in Pd ist sehr intuitiv. Die einzelnen Objekte werden mittels virtuellen Kabeln verbunden und es entstehen sog. „Patches“. Dies sind Diagramme, die gleichzeitig den Datenfluss visuell darstellen und die dargestellten Funktionen auch ausführen. Dabei gibt es keinen Unterschied zwischen der Programmiersituation und der Ausführung des Programms. Alle Änderungen innerhalb eines Patches sind sofort aktiv und müssen nicht erst kompiliert werden oder Ähnliches [14].

Kontrolldaten und Audiosignale

Grundsätzlich werden alle Daten innerhalb von Pd als 32-Bit Fließkommazahlen dargestellt und verarbeitet. Eine wichtige Unterscheidung bei der Arbeit mit Pd ist jedoch die zwischen Kontrolldaten und Audiosignalen. Kontrolldaten sind einzelne Datenblöcke, die manuell zwischen den Objekten verschickt werden können. Es können natürlich Strukturen programmiert werden, welche Kontrolldaten in regelmäßigen Zeitabständen verschicken, jedoch muss dieser Vorgang jedes Mal explizit ausgelöst werden. Demgegenüber werden Audiodaten als konstanter Datenstrom verarbeitet.

Standardmäßig arbeitet Pd mit einer Abtastrate von 44,1 kHz und einer Blockgröße von 64 Samples. Unter Beibehaltung dieser Einstellungen werden Audiodaten in Blocks von 64 Samples verarbeitet, d.h. es werden pro Sekunde konstant $\frac{44100}{64} \approx 689$ Blöcke mit je 64 32-Bit-Zahlen zwischen den Audioobjekten verschickt.

Die oben bereits erwähnte zexy-Bibliothek stellt Objekte bereit, mit denen Audiosignale in Kontrolldaten umgewandelt werden können und umgekehrt.

8.2 Pure Data Patches

Um dem Leser einen Überblick des implementierten Programms zu geben, finden sich im Folgenden Abbildungen der wesentlichen Module sowie der Benutzeroberfläche.

8.2.1 Überblick

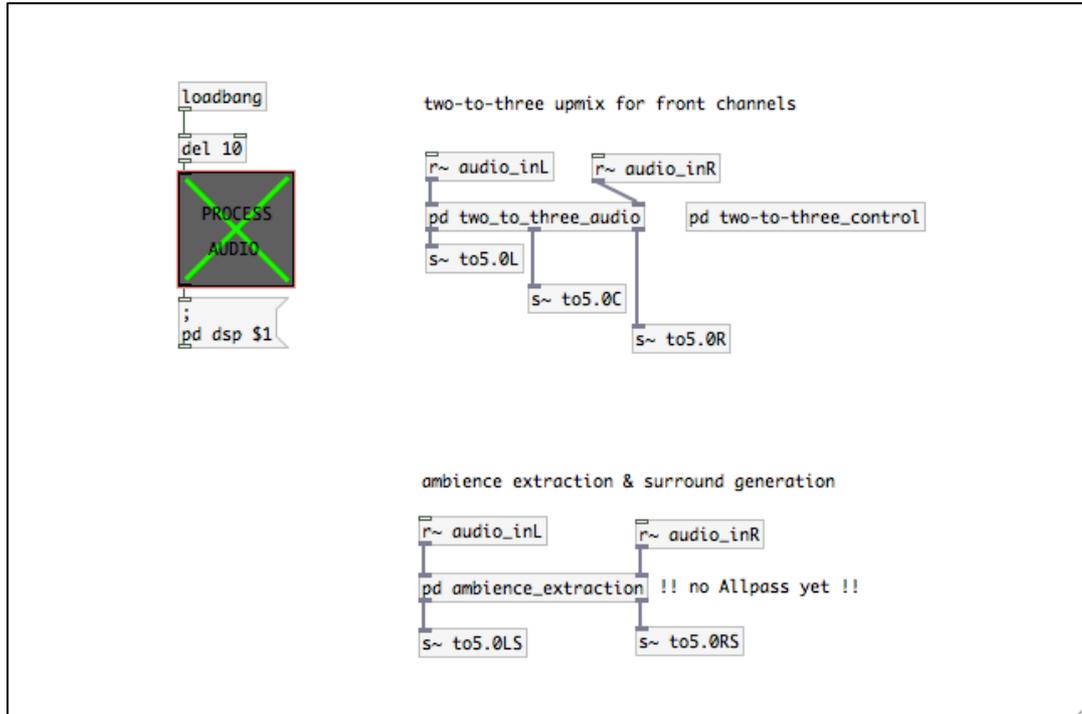


Abb.24: Überblick über das Programm mit Repanoramisierungsstufe (oben) und Hallextraktionsstufe (unten)

8.2.2 Halleextraktion

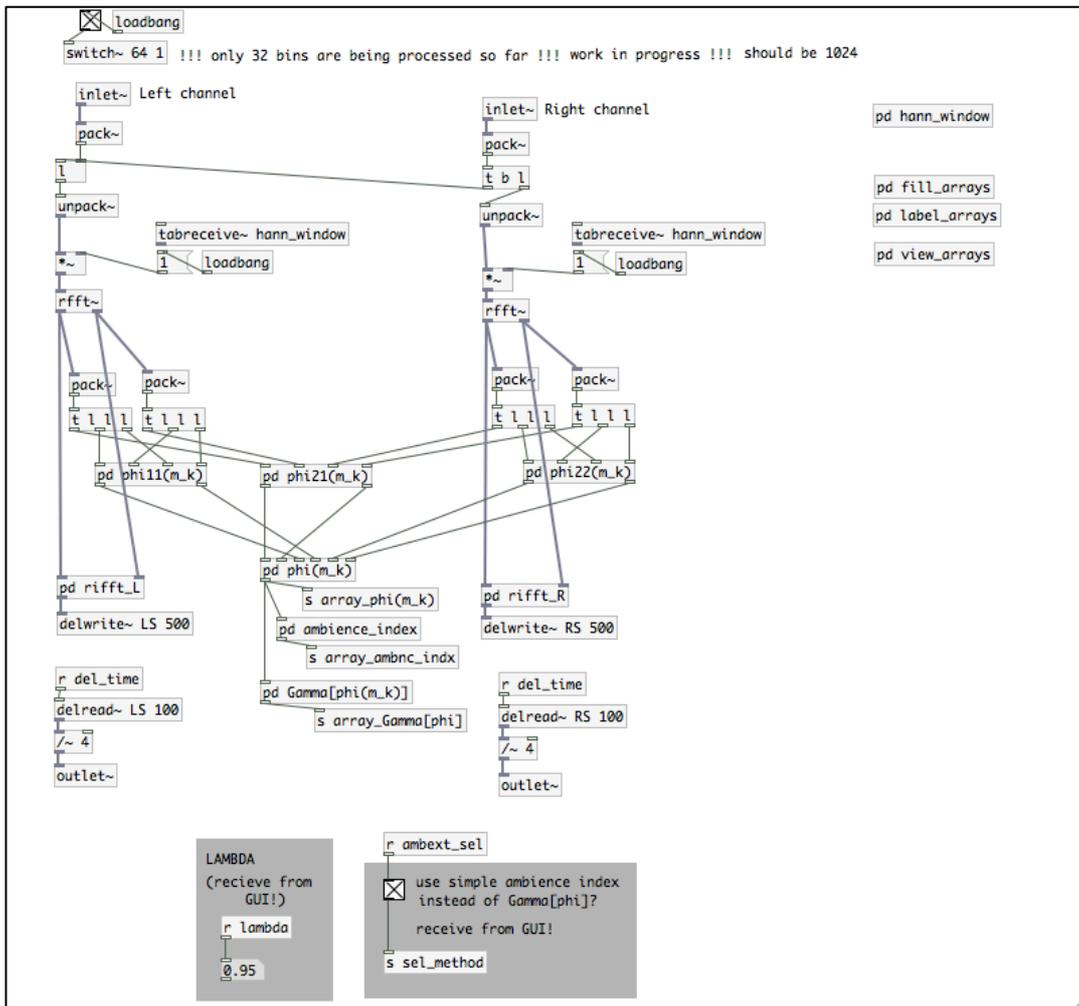


Abb.25: Halleextraktionsstufe mit Audioverarbeitung und Berechnungsmodulen

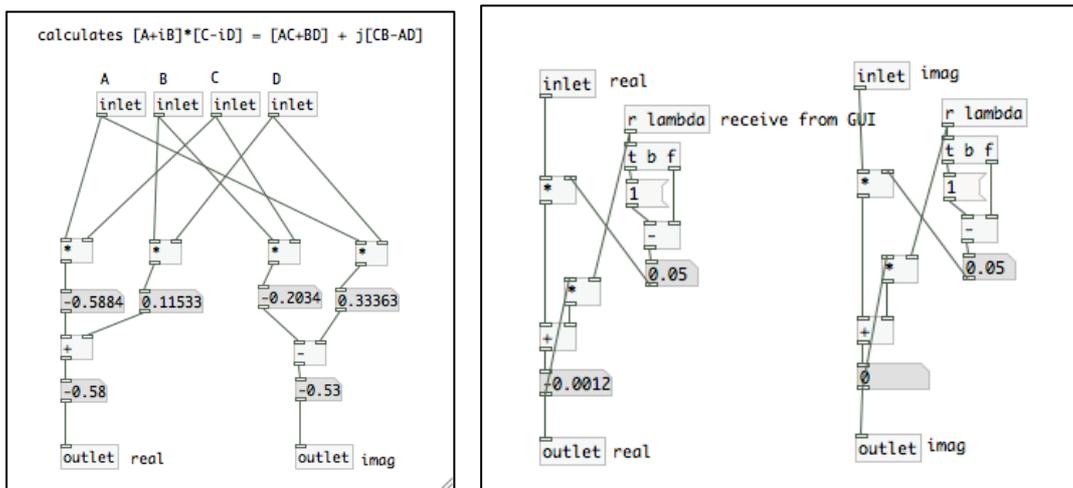


Abb.26: Berechnung der Kreuzkorrelation mit konjugiert komplexer Multiplikation (L) und Glättungsfaktor λ (R)

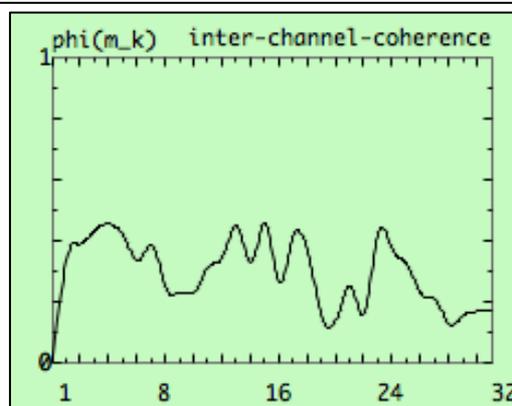
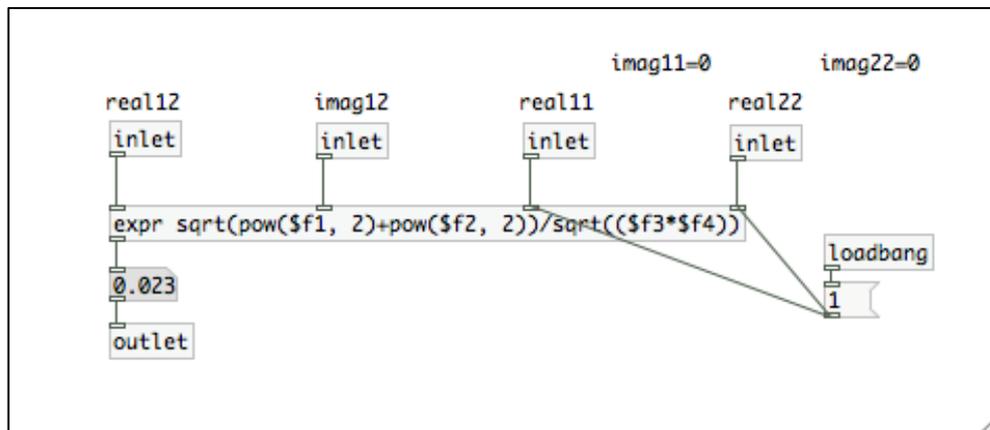


Abb.27: Berechnung der Kurzzeitkohärenz (oben) und Array mit berechneten Werten für alle Frequenzbänder (unten)

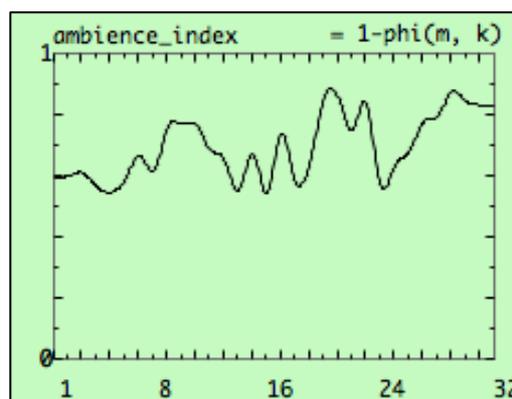


Abb.28: Array mit Hallindexwerten für alle Frequenzbänder

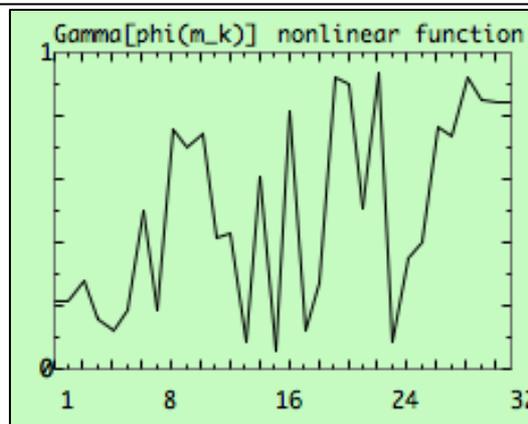
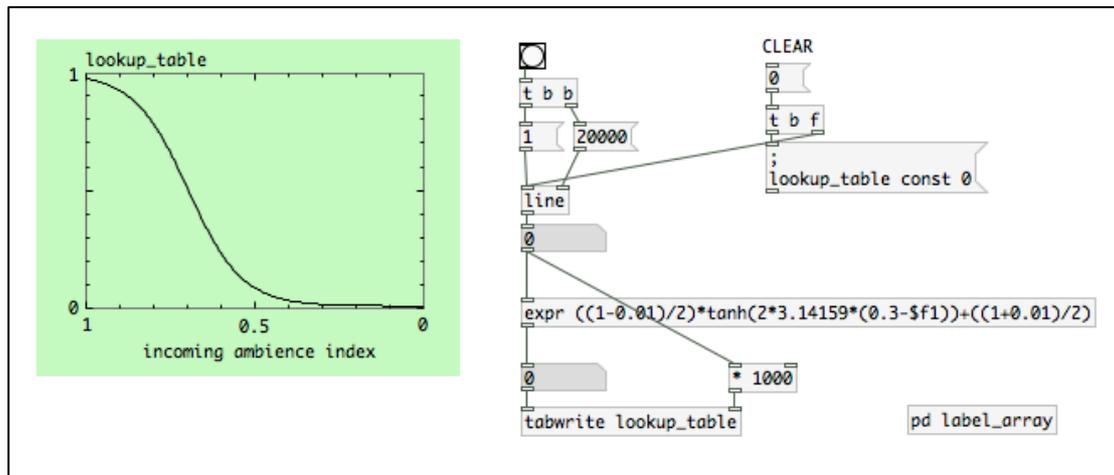


Abb.29: Lookup-Table für nichtlineare Funktion des Hallindex und Array mit ausgelesenen Werten für alle Frequenzen

8.2.3 Repanoramisierung

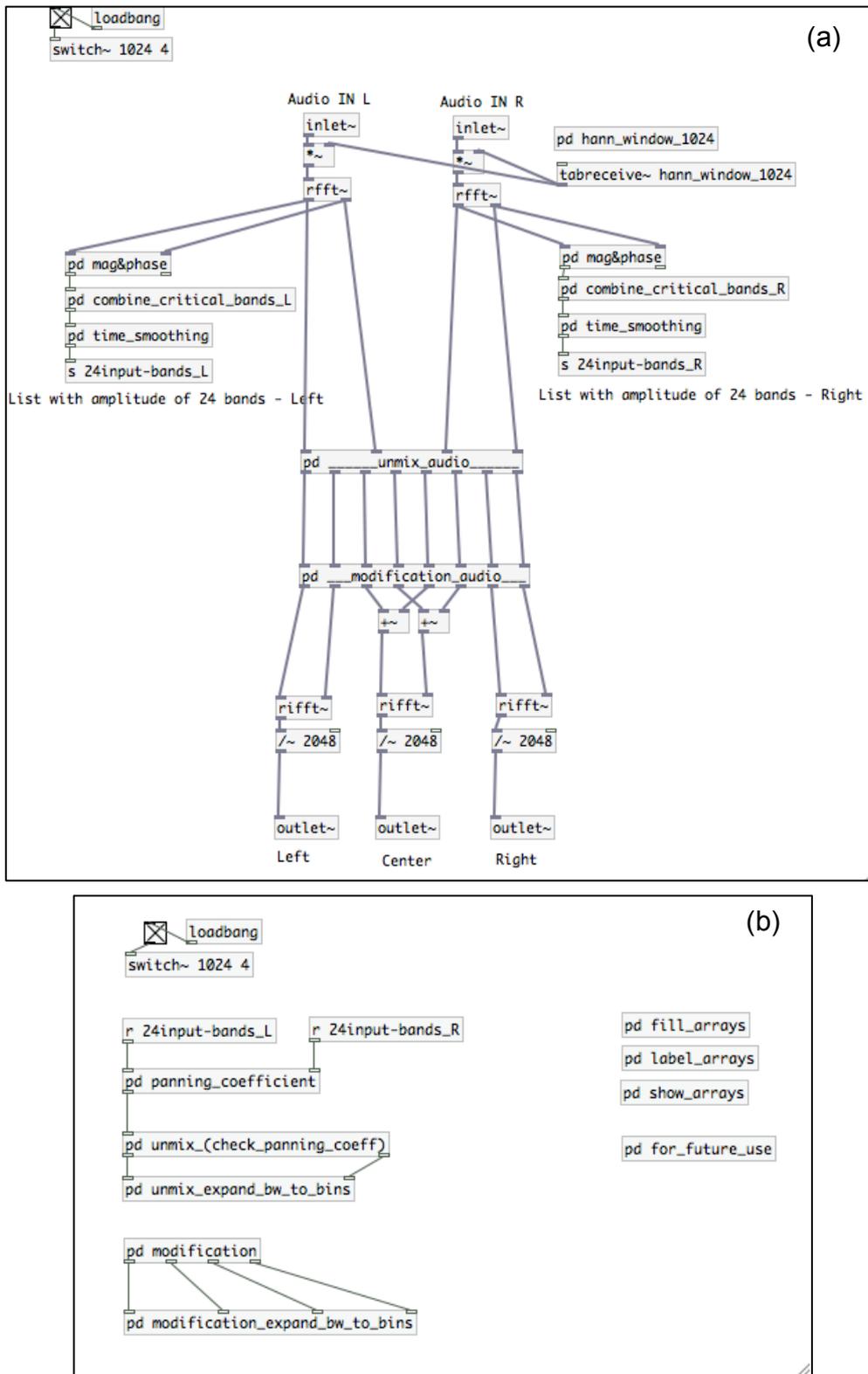


Abb.30: Repanoramisierungsstufe, (a) Audiosignalebene, (b) Kontrollsignalebene

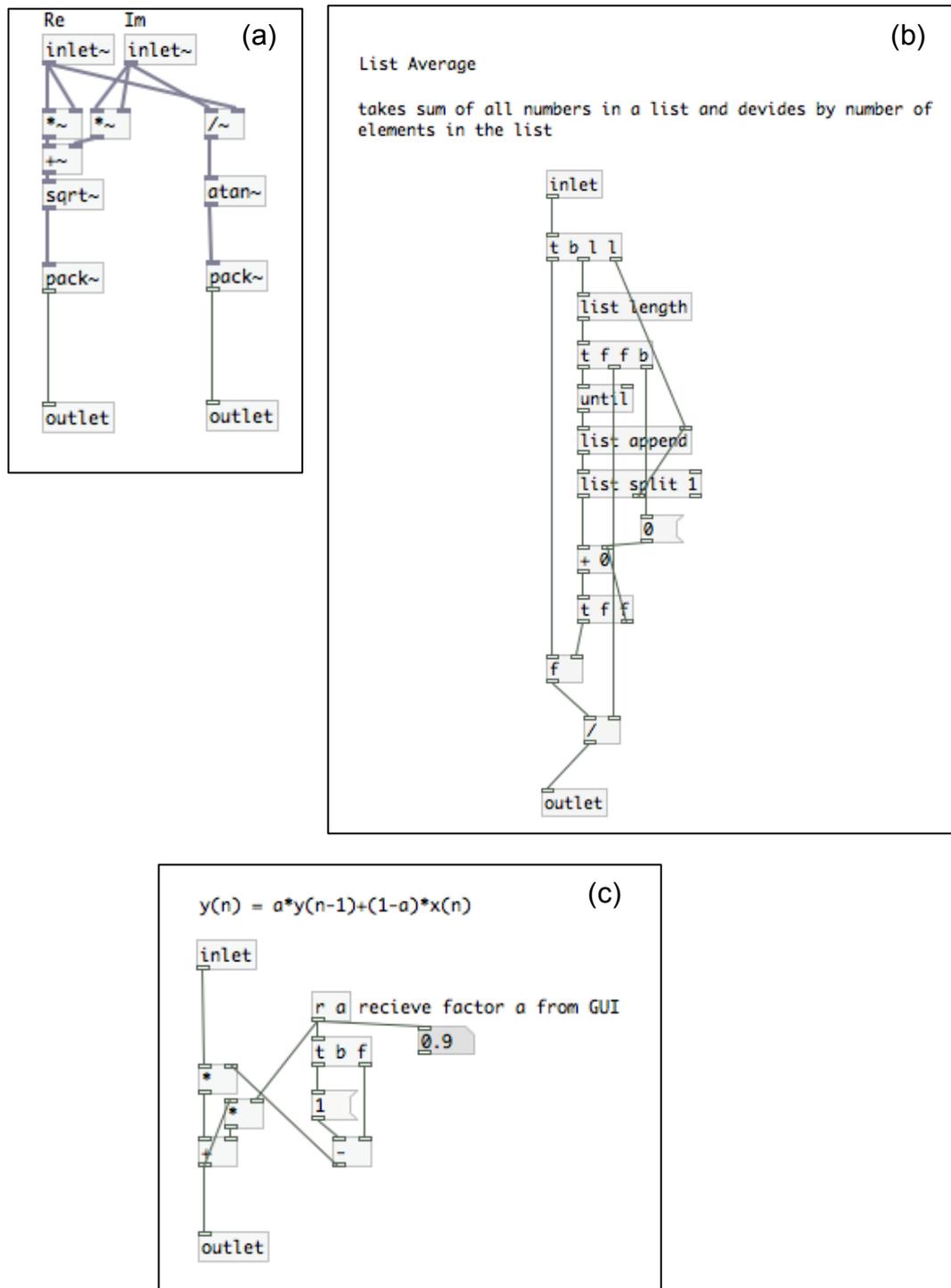


Abb.31: (a) Berechnung des Betrags- und Phasenspektrums,
 (b) Zusammenfassung mehrerer Bins zu Frequenzgruppen,
 (c) Zeitliche Glättung des Amplitudenverlaufs

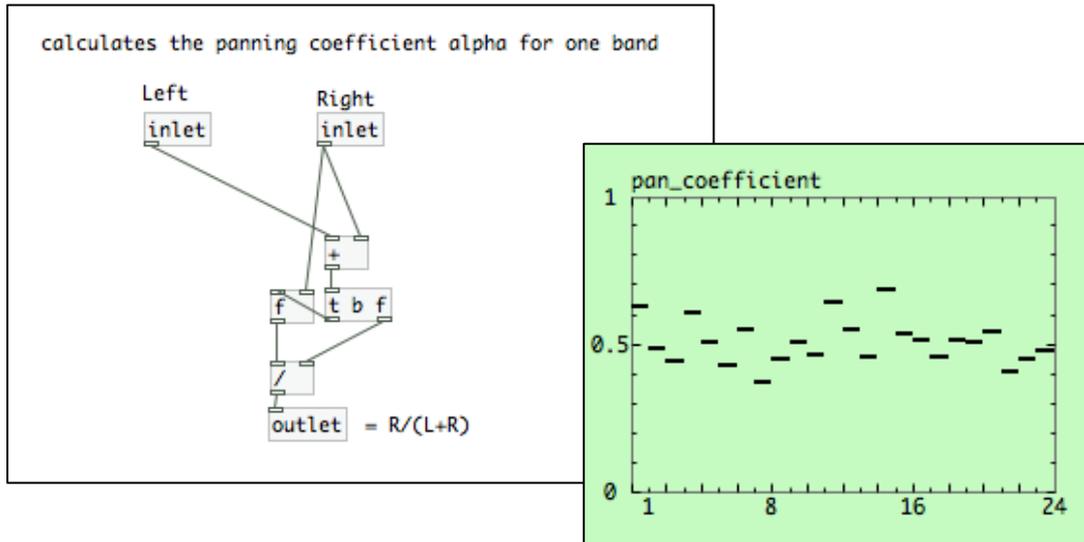


Abb. 32: Berechnung des Panoramakoeffizienten und Array mit Koeffizienten für alle 24 Frequenzgruppen

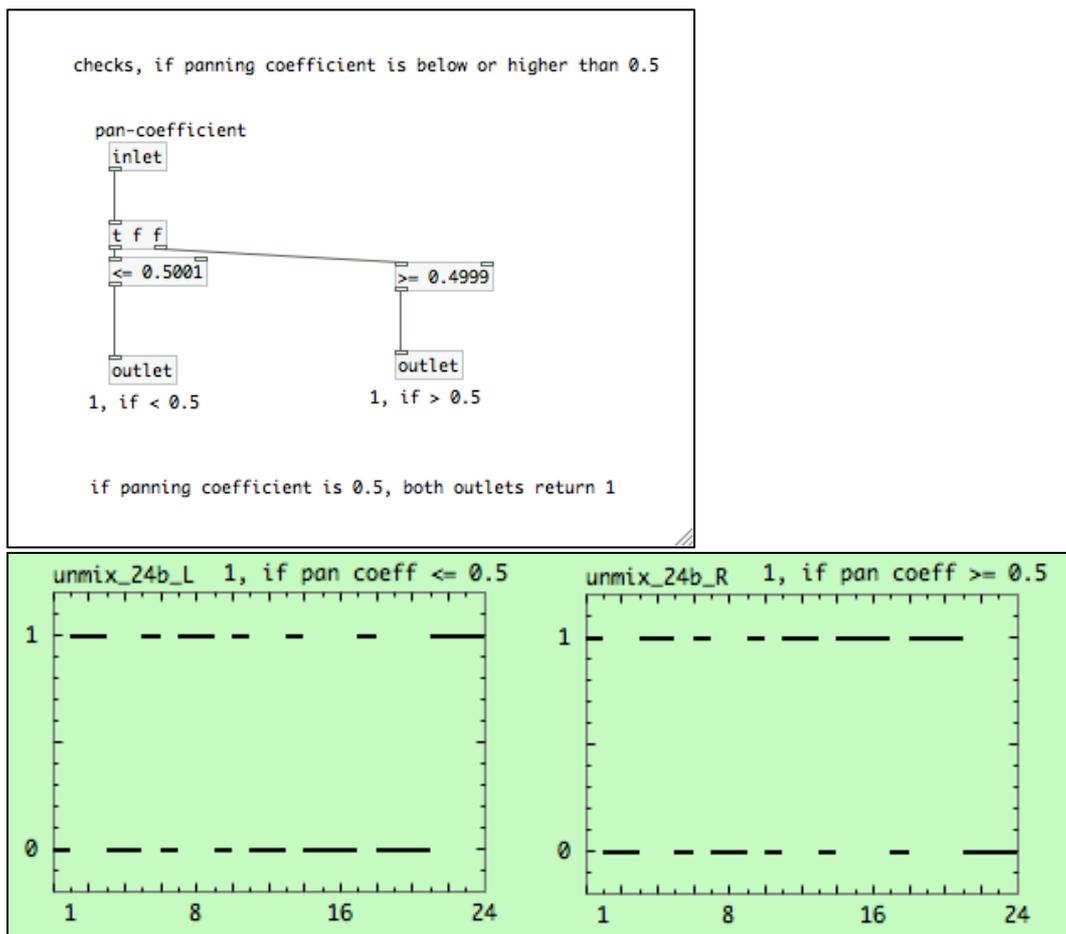


Abb.33: Zerlegungsstufe, Berechnung (oben) und Arrays mit berechneten Werten für die 24 Frequenzgruppen (unten)

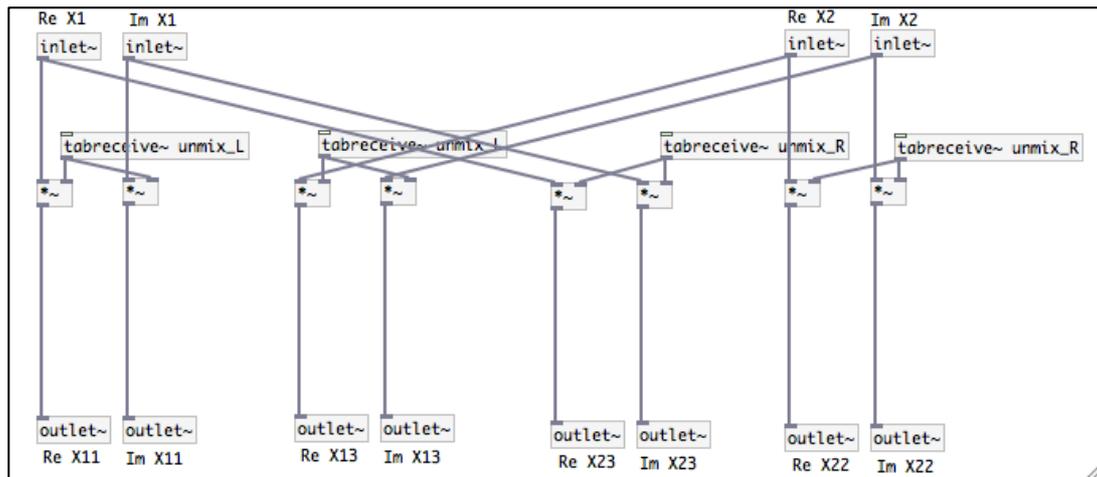


Abb.34: Zerlegungsstufe, Audiosignalebene

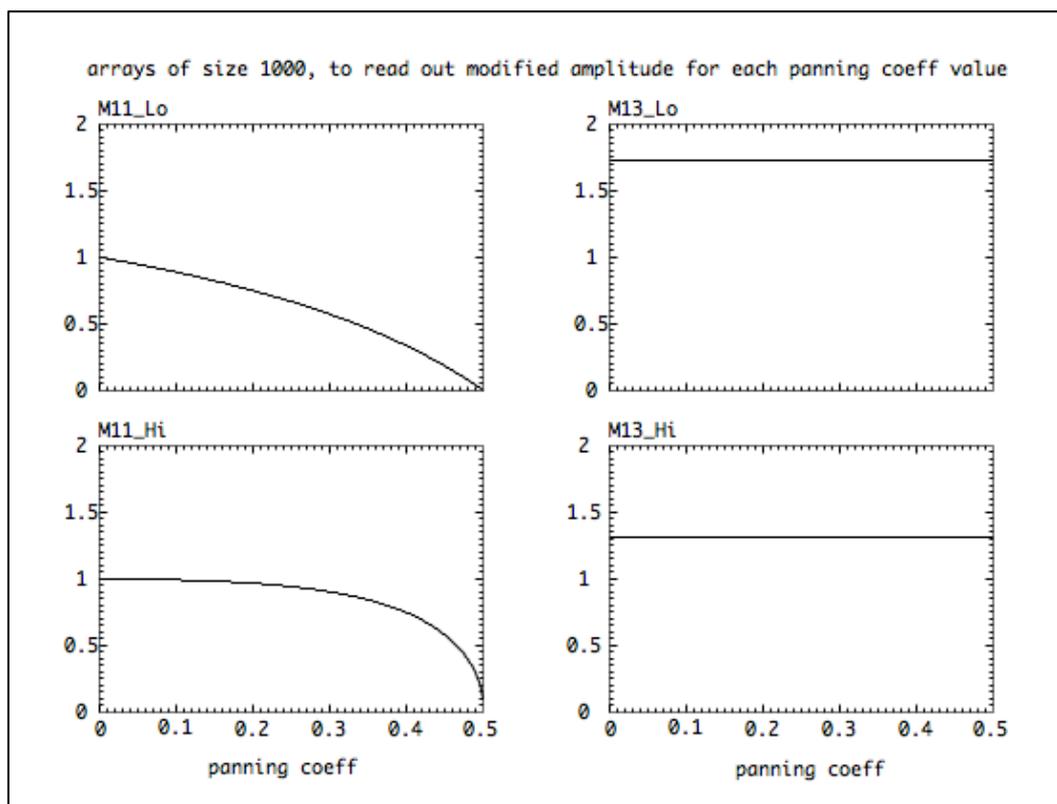


Abb.35: Lookup-Tables für Modifikationsfunktionen zwischen linkem und mittlerem Lautsprecher

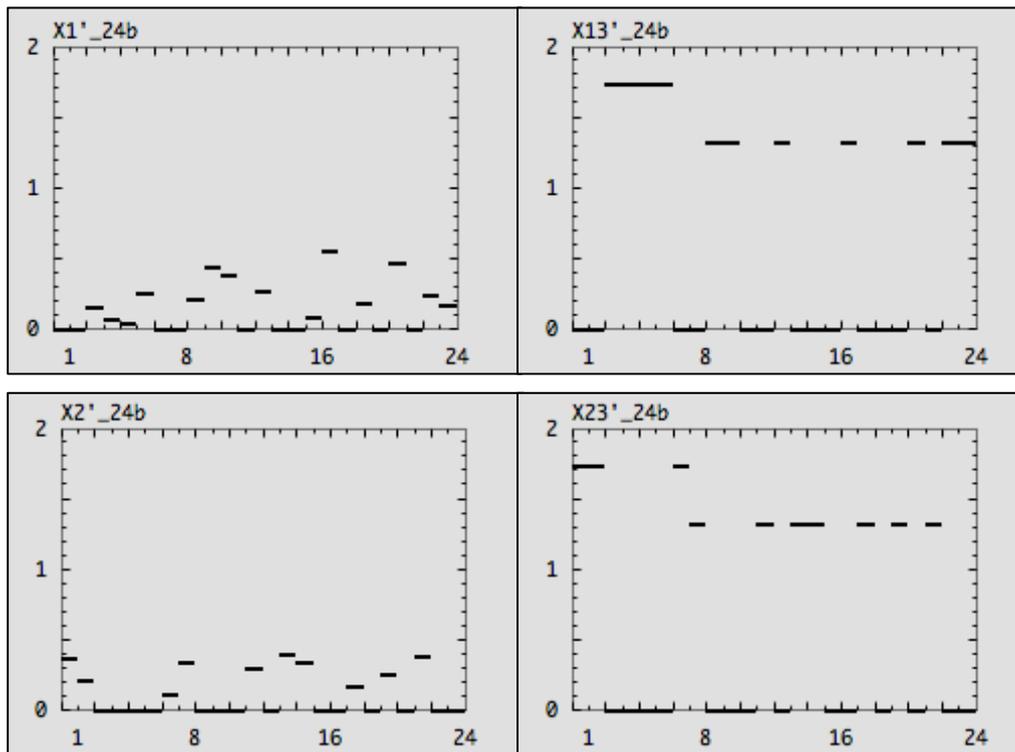


Abb.36: Arrays mit ausgelesenen Modifikationswerten für alle 24 Frequenzgruppen

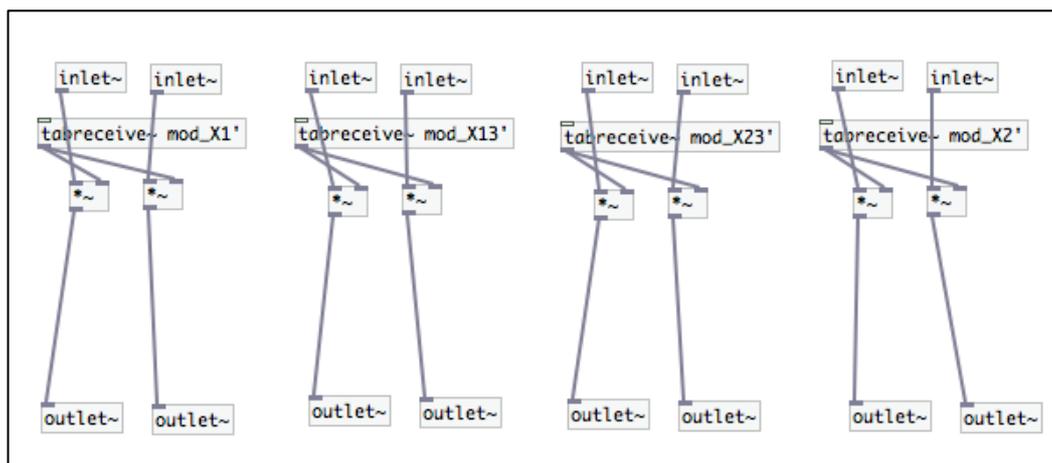


Abb.37: Modifikationsstufe, Audiosignalebene

8.2.4 Benutzeroberfläche

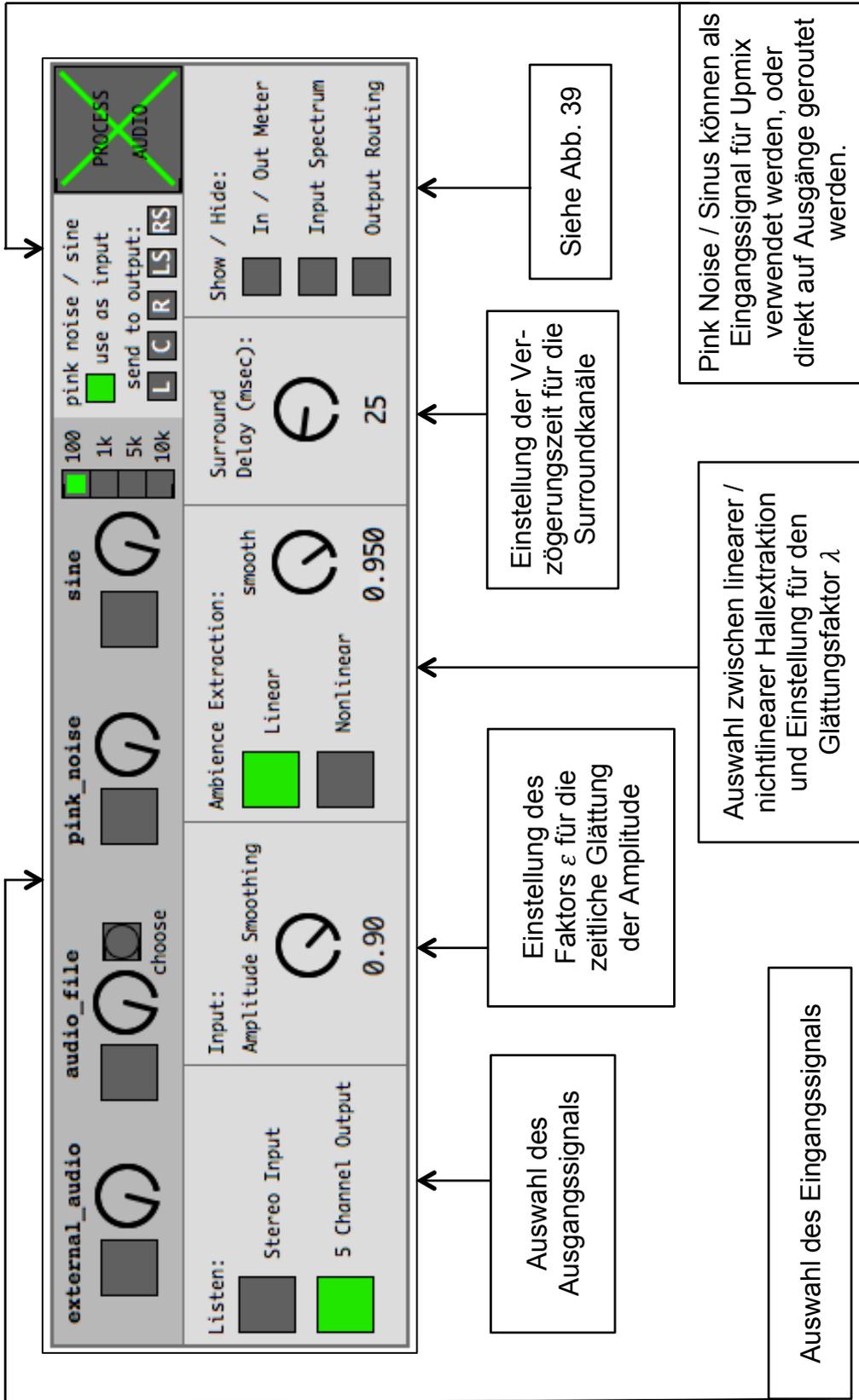


Abb.38: Elemente der Benutzeroberfläche

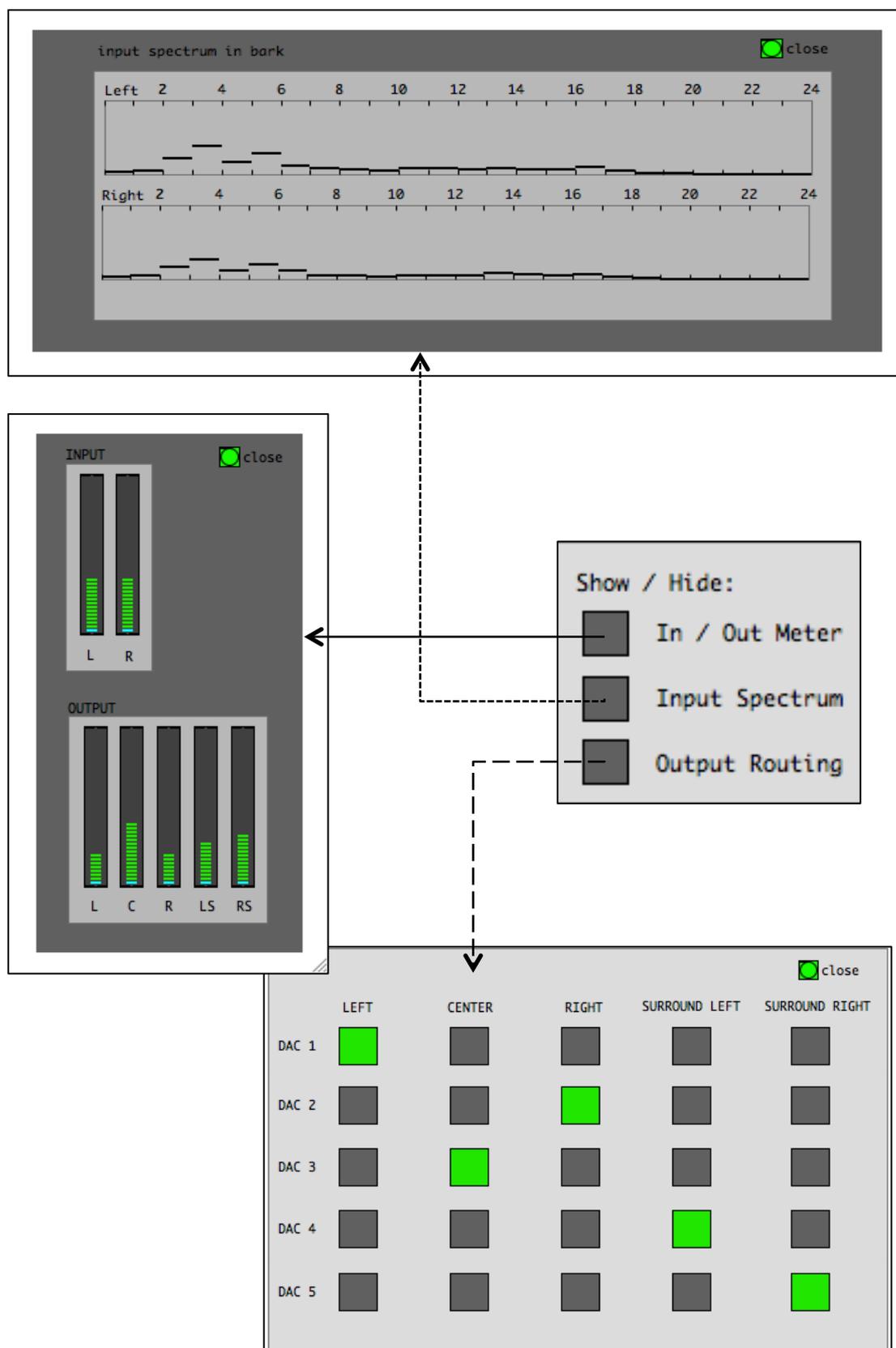


Abb.39: Weitere Elemente der Benutzeroberfläche mit Spektrumanalyse des Eingangssignals in Bark (oben), Eingangs- und Ausgangspegelanzeige (Mitte), Routing Matrix für Ausgangssignale (unten)

