Masterarbeit

Universität für Musik und darstellende Kunst Graz
Technische Universität Graz
Institut für Elektronische Musik und Akustik

Klangtransformationen auf Basis des Modulation Power Spectrums

Betreuer und Prüfer: O.Univ.-Prof. Mag. DI Dr. Robert Höldrich

Graz, Jänner 2017

Thomas Mayr

Eidesstattliche Erlärung

Hiermit erkläre ich an Eides statt, dass ich die vorliegende Arbeit selbstständig und ohne fremde Hilfe verfasst, andere als die angegebenen Quellen und Hilfsmittel nicht benutzt und die aus anderen Quellen entnommenen Stellen als solche gekennzeichnet habe.

Kunstuniversität Graz, am 23. Jänner 2017

Thomas Mayr

Kurzfassung

Mit dem Modulation Power Spectrum (MPS) besteht die Möglichkeit, temporale und spektrale Modulationen von Klängen sichtbar zu machen. Dazu wird eine 2D-Fouriertransformation eines Spektrogramms berechnet. In der Sprachsignalverarbeitung wurde bereits gezeigt, welchen Einfluss eine Modifikation (Filterung) in dieser Domäne auf den resynthetisierten Klang bzw. die Sprachverständlichkeit hat. In der Masterarbeit wird nun untersucht, welche Auswirkungen Transformationen in dieser MPS-Domäne auf weiteres Klangmaterial haben. Dabei kommen Transformationen wie Hoch- bzw. Tiefpassfilterung, Verschieben und Ausblenden von einzelnen Bereichen sowie spezielle Verzerrungen zum Einsatz.

Abstract

The Modulation Power Spectrum (MPS) is used to visualize temporal and spectral modulations of sounds. For this purpose the two-dimensional (2D) Fourier transformation of the spectrogram is calculated. In Speech signal processing it has already been shown how modifications (filtering) effect the resynthesized sound in this domain. This master thesis investigates how transformations in the MPS-domain effect other sound material. Transformations like high- and lowpass filtering, moving and suppressing areas and special distortions will be explored.

Inhaltsverzeichnis

Kapitel 1

Einführung

In dieser Masterarbeit sollen Modulationen von Musik und anderen Klängen untersucht werden. Dazu wird eine Methode beschrieben, anhand derer sich zeitliche bzw. spektrale Modulationen kombiniert darstellen lassen. Dies geschieht anhand der zweidimensionalen Fourier-Transformation des Spektrogramms. Mit dieser ist es möglich, in Klängen enthaltene Periodizitäten sichtbar zu machen. Im ersten Schritt wird dazu die Kurzzeitfouriertransformation (STFT) eines Klanges berechnet und anschließend die logarithmische Magnitude gebildet. Durch den Logarithmus werden die Amplituden der STFT komprimiert. Um Periodizitäten im Spektrogramm darstellen zu können, wird einerseits über die Zeitachse und andererseits über die Frequenzachse nochmals eine Fouriertransformation berechnet. Diese Darstellungsform bezeichnet man in der Literatur als Modulation Power Spectrum (MPS).

Im zweiten Teil dieser Arbeit werden Signalverarbeitungsmethoden vorgestellt, mit denen diese Darstellungsform beschrieben werden kann. Dazu wird eine Einführung in die Techniken und Eigenschaften der zweidimensionalen Fourier-Transformation gegeben. Für die spektralen Modulationen von besonderer Bedeutung ist das Cepstrum eines Klanges. Mit dem Cepstrogramm lassen sich die zeitlichen Veränderungen der Cepstren beschreiben, anhand derer es möglich ist, spektrale Periodizitäten über die Zeit zu visualisieren. Andererseits ist es auch möglich, die zeitlichen Modulationen darzustellen. Hier wird entlang der Zeitachse eine Fourier-Transformation durchgeführt, was in der Literatur als Modulation Spectrum bezeichnet wird und normalerweise für die Bestimmung des Sprachverständlichkeits-Index STI verwendet wird. Werden diese beiden Darstellungsformen nun zusammengeführt, lassen sich sowohl die zeitlichen als auch die spektralen Modulationen grafisch darstellen.

Anschließend wird anhand von einfachen und komplexeren Klängen gezeigt, wie ein MPS berechnet wird. Ein großer Vorteil dieser Signalverarbeitungskette ist ihre Umkehrbarkeit. So ist es möglich, in der MPS-Domäne Veränderungen durchzuführen und das Ergebnis wieder hörbar zu machen. Für diese Bearbeitungsschritte werden Filter für die zweidimensionale Domäne vorgestellt und anhand von Beispielen erläutert. Als ein konkretes Ereignis kann zum Beispiel das Vibrato eines Sängers unabhängig von anderen klanglichen Parametern verändert werden.

Die MPS-Berechnung und die entsprechende Signalresynthese können als Spezialfälle einer mehrstufigen Kette von Signaltransformationen, konkret FT und Logarithmierung sowie deren Inversion, angesehen werden. Ein verallgemeinertes Modell einer solchen Transformationskette, das optional auf die Betragsbildung verzichtet, auch funktionale Wurzeln von Logarithmus und Exponentiation verwendet und damit die Möglichkeit der Klangmodifikation erweitert, wird vorgestellt.

1.1 Modulationen in Klangsignalen

Modulationen treten in vielen Klängen des täglichen Lebens auf. Als Modulation bezeichnet man eine mehr oder weniger periodische Bewegung eines Parameters um einen Mittelwert. Tritt zum Beispiel eine Schwankung der Amplitude eines Klanges in periodischen Zügen auf, spricht man von **Amplitudenmodulation**. Bewegt sich hingegen die Frequenz eines Sinustons in einem periodischen Muster um eine zentrale Frequenz, handelt es sich dabei um eine **Frequenzmodulation**. Als dritte Modulationsart werden **spektrale Modulationen** bezeichnet, die periodische Muster in der Obertonstruktur eines Klanges beschreiben.

1.1.1 Wahrnehmung von Modulationen

Modulationen haben verschiedene Auswirkungen auf die Wahrnehmung von Klängen. Dies kann durch verschiedene psychoakustische Größen beschrieben werden. Dabei beschreibt die *Rauigkeit* eine Empfindung einer zeitvarianten Einhüllenden eines Klanges innerhalb einer Frequenzgruppe. Dies tritt auf, wenn beispielsweise Töne eine periodische Änderung der Amplitude oder der Frequenz aufweisen. Die Rauigkeit ist eine der elementaren Empfindungsgrößen in der Psychoakustik, die zur Beschreibung von schnellen Hüllkurven-

fluktuationen verwendet wird [1, S. 4]. Dabei werden raue Schalle oft als „missklingend“ und „unangenehm“ bezeichnet [2, S. 1]. Diese Wahrnehmung tritt allerdings erst über einer gewissen Modulationsfrequenz auf und zwar dann, wenn die Modulationsfrequenz den Bereich der Schwebung verlässt [3, S. 31ff]. Unter dieser Modulationsfrequenz im Bereich von 0 bis 20 Hz tritt ein weiteres Phänomen auf, welches als Schwankungsstärke bezeichnet wird. Dabei handelt es sich um zeitliche Lautstärkeänderungen, denen das Gehör folgen kann. Dabei weist die wahrgenommene Pegeländerung bezüglich der Modulationsfrequenz eine Bandpasscharakteristik auf, die ein Maximum bei 4 Hz aufweist [4, S. 248ff]. Diese Bandpasscharakteristik ist in Abbildung 1.1 gut zu erkennen. In Abb. 1.2 ist die Schwankungsstärke eines gesungenen Baritonvokals über der Frequenz aufgetragen. Da es sich bei der Hüllkurvenfluktuation nur um eine periodische Veränderung handelt, wird diese Modulationsfrequenz bei ca. 5 Hz erkannt. Dabei steht der Einfluss des Pegels in starkem Zusammenhang mit dem Pegel der Rauigkeit. Ein Schall mit einem Pegel von 90 dB weist in etwa die 10-fache Schwankungsstärke auf wie ein Schall mit 40 dB.

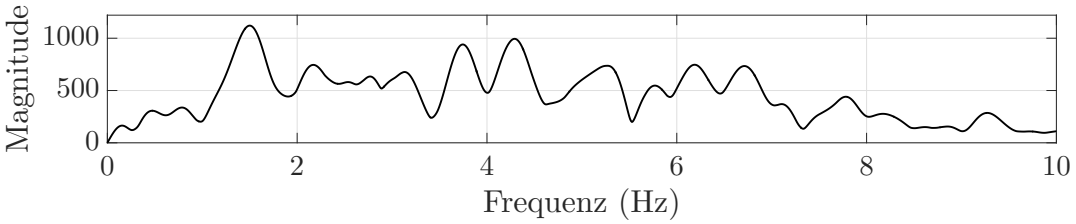


Abb. 1.1: Schwankungsstärke eines männlichen Sprechers.

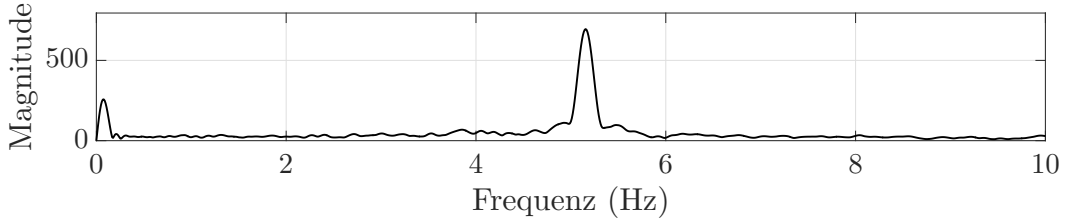


Abb. 1.2: Schwankungsstärke eines gesungenen Baritonvokals.

Dabei können die Übergänge zwischen den Bereichen als fließend angesehen werden. Als Beispiel werden zwei Töne betrachtet, die eine geringe Frequenzdifferenz aufweisen. Wird diese Frequenzdifferenz kontinuierlich erhöht und übersteigt diese Frequenzdifferenz jenen Bereich, in dem Rauigkeit wahrnehmbar ist, werden zwei getrennte Töne wahrgenommen. Dies kann damit erklärt werden, da ab diesem Abstand zwei Töne nicht mehr in eine gemeinsame Frequenzgruppe fallen.

Die Wahrnehmung der spektralen Modulationen basiert auf der mehr oder weniger periodischen Zusammensetzung der Obertöne eines Spektrums. Sind ganzzahlige Vielfache des Grundtons in einem Spektrum enthalten, wird dies als konsonant empfunden. Treten auch nicht ganzzahlige Vielfache eines Grundtons auf, wird dies als dissonant empfunden. Ist hingegen überhaupt keine harmonische Struktur in den Obertonverhältnissen auszumachen, wird dies als geräuschhaft empfunden und es können keine spektralen Modulationen erkannt werden. Mathematisch können diese Modulationsarten wie im folgenden Teil beschrieben werden.

1.1.2 Amplitudenmodulation

Ein amplitudenmoduliertes Signal lässt sich dadurch beschreiben, dass zum Beispiel ein Sinussignal, welches hier als Trägersignal bezeichnet wird, durch eine zweite Sinusschwingung moduliert wird:

$$x_{\mathrm{AM}}(t) = \big(1 + m \cdot \cos\big(2\pi f_{\mathrm{mod}} t\big)\big) \cdot \cos\big(2\pi f_0 t\big) \tag{1.1.1}$$

Dabei wird m als Modulationsgrad, f_{mod} als Modulationsfrequenz und f_0 als Trägerfrequenz bezeichnet. Nach Umformung von Gl.1.1.1 kann die Frequenzzusammensetzung für das resultierende Signal abgeleitet werden:

$$x_{\mathrm{AM}}(t) = \cos\big(2\pi f_0 t\big) + \frac{m}{2}\cos\big(2\pi (f_0 + f_{\mathrm{mod}}) t\big) - \frac{m}{2}\cos\big(2\pi (f_0 - f_{\mathrm{mod}}) t\big) \tag{1.1.2}$$

Aus Gleichung 1.1.2 wird ersichtlich, dass zusätzlich zur Trägerfrequenz f_0 noch zwei weitere Schwingungen mit den Frequenzen $f_0 - f_{\mathrm{mod}}$ und $f_0 + f_{\mathrm{mod}}$ erzeugt werden. Befindet sich diese Modulationsfrequenz unter 20 Hz, so wird das neu entstandene Signal $x_{\mathrm{AM}}(t)$ mit einer bestimmten Schwankungsstärke wahrgenommen. Übersteigt die Modulationsfrequenz 20 Hz, so wird ein Rauigkeitseindruck erzeugt, der sein Maximum bei ca. 70 Hz aufweist [4, S. 257]. Die Zahl 1 in Gl. 1.1.1 sorgt dafür, dass das Trägersignal im resultierenden Signal enthalten ist. Dies wird in der Nachrichtentechnik als *Amplitudenmodulation mit Trägersignal* bezeichnet. Würde man die 1 weglassen, entstünden nur Frequenzen mit zweifachem Abstand der Modulationsfrequenzen symmetrisch um das nicht vorhandene Trägersignal. Betrachtet man nicht wie hier als Träger ein Sinussignal, sondern ein komplexes Signal[1] und moduliert dieses mit einer Sinusschwingung, wird der Effekt einer Ringmodulation

[1] als komplex wird hier nicht die mathematische Struktur des Signals bezeichnet, sondern nur der Sachverhalt beschrieben, dass sich die Zusammensetzung des Spektrums nicht mit einer Frequenz beschreiben lässt.

erzeugt. Diese findet ihre Anwendung unter anderem in der Nachrichtentechnik und der Elektronischen Musik zur künstlerischen Klanggestaltung.

1.1.3 Frequenzmodulation

Eine weitere Klangveränderung kann durch die Frequenzmodulation beschrieben werden. Ein Sinussignal, welches seine Frequenz in einem periodischen Muster ändert, wird als frequenzmoduliert bezeichnet.
Ein mögliches Trägersignal kann beschrieben werden mit der Amplitude A_T und der Frequenz f_T:

$$x_T = A_T \cdot \cos\big(2\pi f_T t\big). \tag{1.1.3}$$

Ein Signal, welches das Trägersignal in seiner Frequenz moduliert, wird definiert mit der Amplitude A_M und der Frequenz f_M:

$$x_M = \cos\big(2\pi f_M t\big). \tag{1.1.4}$$

Die Momentanfrequenz kann aus der Phase über die Beziehung

$$f(t) = \frac{1}{2\pi} \cdot \frac{\mathrm{d}\phi(t)}{\mathrm{d}t} \tag{1.1.5}$$

abgeleitet werden. Umgekehrt ergibt sich die Phase eines Signals über die Integration der Frequenz:

$$\phi(t) = 2\pi \int f(t)\,\mathrm{d}t. \tag{1.1.6}$$

Die Momentanfrequenz eines frequenzmodulierten Trägersignals kann nun beschrieben werden durch

$$\Omega(t) = f_T + \alpha \cdot x_M(t), \tag{1.1.7}$$

bei der der konstante Faktor α den Modulationshub der Frequenzmodulation angibt. Es ist zu erkennen, dass eine sinusförmige Änderung der Frequenz um die Trägerfrequenz stattfindet.
Um die Phase des frequenzmodulierten Signals zu erhalten, wird Gl. 1.1.7 vom Zeitpunkt

0 bis t integriert:

$$\begin{aligned}\phi_{\mathrm{FM}}(t) &= 2\pi \int_0^t f_T + \alpha x_M(t)\,\mathrm{d}t \\ &= 2\pi f_T t + \alpha \int_0^t \cos\big(2\pi f_M t\big)\,\mathrm{d}t \\ &= 2\pi f_T t + \frac{\alpha}{f_M} \cdot \sin\big(2\pi f_M t\big) \end{aligned} \tag{1.1.8}$$

wobei das Verhältnis α/f_M als Modulationsindex η bezeichnet wird und eine konstante Phasenverschiebung zum Zeitpunkt 0 vernachlässigt wird. In Gl. 1.1.8 ist auch zu beobachten, dass die Phase nicht mehr konstant steigt wie bei einer konstanten Frequenz, sondern um die steigende Gerade sinusförmig schwingt.
Setzt man nun Gl. 1.1.8 in eine Cosinusfunktion, erhält man ein frequenzmoduliertes Cosinussignal:

$$\begin{aligned} x_{FM} &= \cos\big(\phi_{\mathrm{FM}}(t)\big) \\ &= \cos\big(2\pi f_T t + \frac{\alpha}{f_M} \cdot A_M \cdot \sin\big(2\pi f_M t\big)\big) \end{aligned} \tag{1.1.9}$$

Abhängig vom Modulationsindex treten symmetrisch zur Trägerfrequenz weitere Seitenbänder auf. Die Amplituden der Seitenfrequenzen können mit den Besselfunktionen erster Gattung beschrieben werden, welche sich aus der Besselschen Differentialgleichung ergeben. Dies ist in Abbildung 1.3 zu sehen.

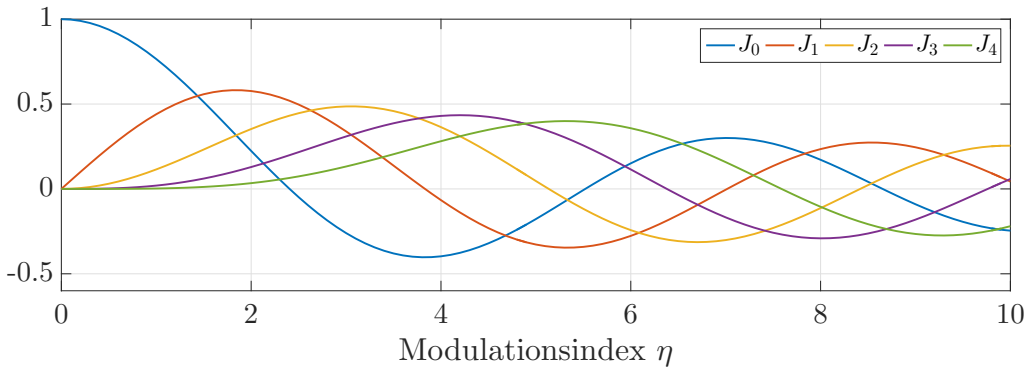


Abb. 1.3: Besselfunktionen erster Gattung. Die Amplituden der Träger- bzw. Seitenfrequenzen bei der Frequenzmodulation sind abhängig vom Modulationsindex η. Dabei bezeichnet J_0 die Amplitude der Trägerfrequenz. Die Besselfunktionen J_1 bis J_4 beschreiben die Amplituden der Seitenschwingungen symmetrisch um die Trägerfrequenz mit dem Abstand der ganzzahligen Vielfachen der Modulationsfrequenz. Es ist zu beobachten, dass zum Beispiel die Amplitude der Trägerfrequenz bei den Modulationsindices 2.4, 5.5 und 8.65 verschwindet.

Die Trägerschwingung und die Seitenbänder können über folgende Formel beschrieben werden:

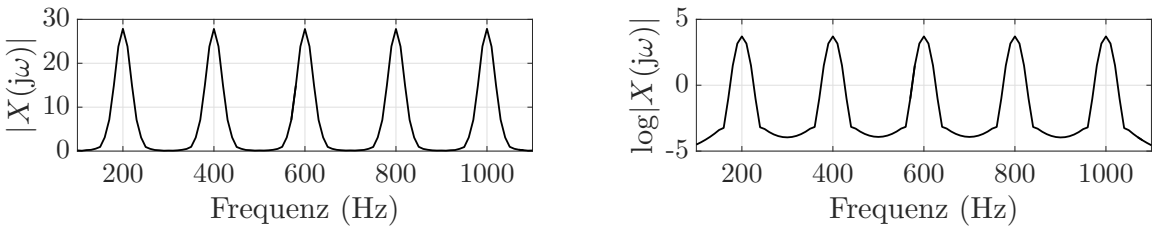
$$\begin{aligned} x_{FM}(t) =& J_0(\eta) \cdot \cos(\omega_T \cdot t) \qquad (1.1.10) \\ & - J_1(\eta) \cdot \sin((\omega_T - \omega_M) \cdot t) - J_1(\eta) \cdot \sin((\omega_T + \omega_M) \cdot t) \\ & - J_2(\eta) \cdot \sin((\omega_T - 2\omega_M) \cdot t) - J_2(\eta) \cdot \sin((\omega_T + 2\omega_M) \cdot t) \\ & - J_3(\eta) \cdot \sin((\omega_T - 3\omega_M) \cdot t) - J_3(\eta) \cdot \sin((\omega_T + 3\omega_M) \cdot t) \\ & \dots \qquad (1.1.11) \end{aligned}$$

1.1.4 Spektrale Modulationen

Um spektrale Modulationen beschreiben zu können, wird folgendes periodisches Signal mit ganzzahligen Vielfachen n der Grundfrequenz betrachtet:

$$x(t) = \sum_{n=1}^{N} A_i \cdot \cos\big(2\pi f n t\big). \qquad (1.1.12)$$

Wird dieses Signal anhand einer STFT in eine Spektraldarstellung übergeführt, so ergibt der Betragsfrequenzgang und die Logarithmierung davon nach einer Fensterung mit einem Gaußfenster folgende Darstellung (1.4):

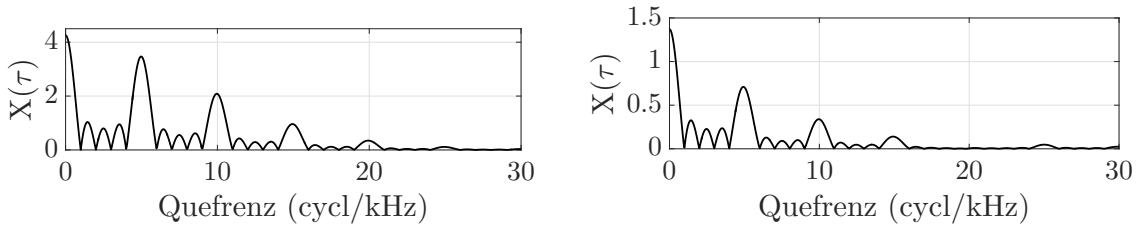


(a) Betragsfrequenzgang **(b)** logarithmierter Betragsfrequenzgang

Abb. 1.4: Darstellung von Gl. 1.1.12 mit $f = 200$ Hz, $A_i = 1$ und $N = 5$. Links der Betragsfrequenzgang mit 5 Teilschwingungen. Rechts der logarithmierte Betragsfrequenzgang. Der Logarithmus bewirkt eine Komprimierung des Dynamikumfangs des Betragsfrequenzganges. Dadurch ergibt sich aus den spitzen Nadeln (links) eine abgerundete Darstellung.

Dabei ist zu erkennen, dass die Oberschwingungen des Grundtons in einem ganzzahligen Verhältnis zu diesem stehen. Anders ausgedrückt kann die Ganzzahligkeit auch als etwas Periodisches beschrieben werden, wenn der Betragsfrequenzgang wieder als neues Signal

angesehen wird. Um Periodizitäten in diesem Signal detektieren zu können, wird wieder die Fourier-Transformation berechnet.



(a) Spektrum des Betragsfrequenzganges

(b) Spektrum des logarithmierten Betragsfrequenzganges

Abb. 1.5: Fourier-Transformation des Betragsfrequenzganges (a) und Fourier-Transformation des logarithmierten Betragsfrequenzganges (b), was auch als Cepstrum bezeichnet wird. Das Cepstrum weist eine geringere Anzahl an Rahmonischen auf im Gegensatz zur Fourier-Transformation des Betragsfrequenzganges, da der Logarithmus den Betragsfrequenzgang komprimiert und so ein sinusförmigeres Signal erzeugt.

In Abbildung 1.5 ist die Fourier-Transformation des jeweiligen Betragsfrequenzganges und des logarithmierten Betragsfrequenzganges zu sehen. Die zweite Darstellung wird als Cepstrum bezeichnet, in dem die Abszisse die Zyklen pro kHz beschreibt. Der Unterschied zwischen den beiden Abbildungen ergibt sich dadurch, dass die logarithmierte Betragsfrequenzgangsdarstellung eher eine sinusförmige Gestalt aufweist als der reine Betragsfrequenzgang. Dies ist an den Harmonischen im Cepstrum zu erkennen, welche in der Literatur auch als „Rahmonische“ bezeichnet werden. Das erste Maximum zeigt die Grundperiode der spektralen Darstellung an. In Abbildung 1.4 (a) befinden sich fünf spektrale Maxima innerhalb von einem kHz. Dies ergibt im Cepstrum (b) ein erstes Maximum bei fünf Zyklen/kHz. Die Rahmonischen ergeben sich dadurch, dass der Betragsfrequenzgang bzw. der logarithmierte Betragsfrequenzgang nicht sinusförmig ausgeprägt ist. Dies kann mit einer Fourierreihendarstellung verglichen werden. Um ein einfaches Signal mit Teilsignalen beschreiben zu können, wird nur eine geringe Anzahl an Teilsignalen benötigt. Dies ist in Abbildung 1.4 sehr gut zu sehen: Der Logarithmus bewirkt eine Komprimierung des Betragsfrequenzganges hin zu einem annähernd sinusförmigen Signal. Deswegen auch der geringere Anteil an Rahmonischen im rechten Teil von Abbildung 1.5.

1.2 Modulationen in Sprache

Modulationen spielen bei Sprache eine wichtige Rolle. Bei gesprochener Sprache beträgt die Anzahl der Silben pro Sekunde ungefähr 4. Das bedeutet, dass Sprache ungefähr mit 4 Hz moduliert ist [5, S. 70]. Damit Sprache eine gute Sprachverständlichkeit aufweist, sollte der Modulationsgrad groß sein. In [6, S. 1] wurde gezeigt, dass die Sprachverständlichkeit erheblich sinkt, wenn zeitliche Modulationen unter 12 Hz und spektrale Modulationen unter 4 Zyklen/kHz entfernt wurden. Dabei wurde auch gezeigt, dass die wichtigsten zeitlichen Modulationen zwischen 1 und 7 Hz, und die wichtigsten spektralen Modulationen unter 1 Zyklus/kHz unerlässlich für die Sprachverständlichkeit sind. Weiters wurde gezeigt, dass, falls der Bereich von 3 - 7 Zyklen/kHz weggefiltert wird, eine Erkennung des Geschlechts des Sprechers oder der Sprecherin erheblich erschwert wird. Zum Beispiel sinkt in halligen Räumen die Sprachverständlichkeit deshalb, weil der Modulationsgrad sinkt. Als Maß für die Qualität einer Übertragungsstrecke von einem Sprecher zu einem Zuhörer wird beispielsweise der Sprachübertragungsfaktor, kurz STI, verwendet. Dazu wird eine Raumimpulsantwort in einzelne Oktavbänder unterteilt. Von diesen Oktavbändern wird das Spektrum analysiert, welches anschließend angibt, welche zeitlichen Modulationen in diesem Frequenzband enthalten sind. Dies wird als Modulationsübertragungsfunktion (MTF) bezeichnet. Diese einzelnen MTFs werden zu einem gemeinsamen Sprachverständlichkeitsindex STI zusammengefasst, der die Qualität einer Übertragungsstrecke angibt.

1.3 Modulationen in Musik und anderen Klängen

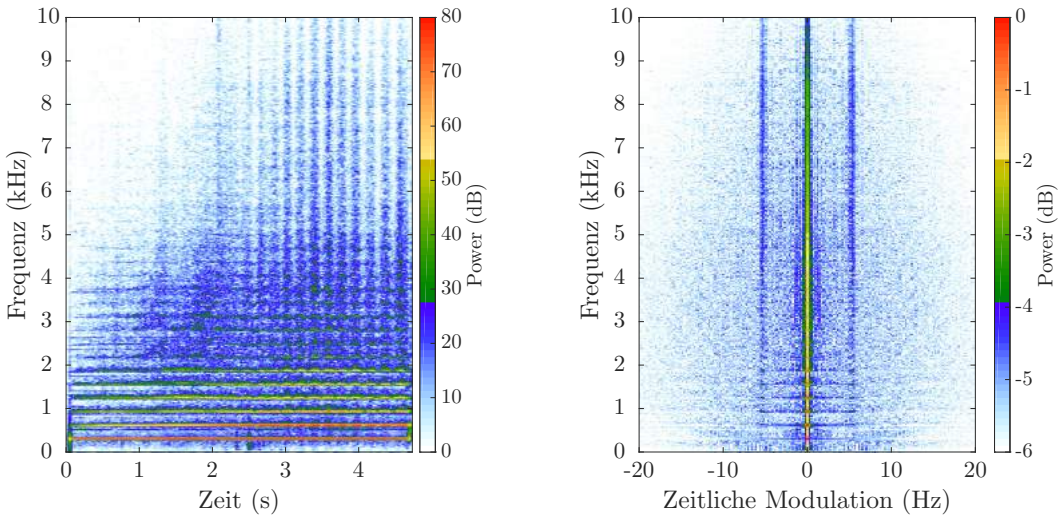
In musikalischen Aufführungen werden Schwebungen zum Beispiel zum Stimmen von Instrumenten verwendet. Spielen zwei Instrumente ungefähr den selben Ton, lässt sich der Abstand der Grundtöne an der Schwebungsfrequenz ausfindig machen. Bewegt sich die Schwebungsfrequenz hin zu kleineren Werten, erreichen die beiden Grundtöne immer mehr die selbe Frequenz [7, 577ff]. Diese Schwebungsfrequenz kann mit dem psychoakustischen Parameter der Schwankungsstärke beschrieben werden. Erklingen zwei Instrumente zusammen und sind die Frequenzen der Obertöne der beiden Instrumente nur sehr gering von einander entfernt, wird ein Rauigkeitseindruck wahrgenommen. Modulationen kommen aber auch auf natürliche Weise zum Ausdruck. So unterliegen praktisch alle Instrumentalklän-

ge statistischen Schwankungen. Diese Schwankungen werden in der Musik üblicherweise in Cent angegeben. Diese lassen sich aus einem beliebigen Frequenzverhältnis folgendermaßen berechnen:

$$x_{\text{cent}} = 1200 \cdot \log_2\left(\frac{f_1}{f_2}\right). \tag{1.3.1}$$

Die Schwankungen befinden sich bei Streichinstrumenten bei etwa ±4 Cent und bei großem Bogendruck sogar bei ±20 Cent [8, S. 151]. Eine periodische Änderung der Frequenz eines Tones in der Musik wird als Vibrato bezeichnet. Als optimal gilt eine Vibratofrequenz von $6-8$ Hz [8, S. 151], da in diesem Bereich die Modulationsfrequenz als besonders belebend spürbar wird. Eine eindeutige Tonhöhenwahrnehmung bleibt aber auch mit solchen Modulationsfrequenzen erhalten. Physikalisch wird durch ein Vibrato bei Streichinstrumenten eine periodische Längenänderung der Saite und dadurch eine Frequenzmodulation erzeugt. Jens Meyer schreibt dazu »*Bedingt durch die scharfen Resonanzen des Instrumentenkorpus bewirkt diese Frequenzmodulation zusätzlich eine Amplitudenmodulation des abgestrahlten Klanges. Diese Amplitudenmodulation ist für die einzelnen Teiltöne unterschiedlich stark und nicht gleichphasig*« [8, S. 151ff]. Wie man hier sieht, spielen Amplituden- und Frequenzmodulation immer zusammen eine Rolle, wenn es um zeitliche Modulationen in einem Klang geht. Zum Beispiel kann bei einer Gesangsstimme ein Vibrato von ca. ±40 bis ±80 Cent ausgemacht werden, welches auch als angenehm empfunden wird. Steigert sich diese Abweichung auf bis zu ±200 Cent, bildet sich zusätzlich zur Frequenzmodulation auch eine Amplitudenmodulation aus, welche der Stimme eine besondere Auffälligkeit verleiht [8, S. 152].

Bei Blasinstrumenten wie der Flöte treten Frequenzschwankungen von ±10 bis ±15 Cent auf, welche schon als ein sehr starkes Vibrato angesehen werden können. Dabei treten fast keine Amplitudenschwankungen bei den unteren Teiltönen auf. Diese sind erst ab ca 2000 Hz ausfindig zu machen. Wird einem Flötenklang zusätzlich noch eine Amplitudenschwankung aufgeprägt, spricht man von einem Tremolo, das in Abbildung 1.6 zu sehen ist.



(a) Spektrogramm **(b)** Zeitliches Modulationsspektrum

Abb. 1.6: Spektrogramm und zeitliches Modulationsspektrum eines Flötentons, dem ein Tremolo aufgeprägt ist. In (b) ist gut zu sehen, dass sich die Frequenz der Amplitudenschwankung bei ca. 5 Hz befindet und erst über 2000 Hz gut ausgeprägt ist bzw. dort wahrgenommen werden kann.

1.4 Aufbau der Arbeit

All diese beschriebenen Modulationen können in einem Klang auftreten bzw. an der Zusammensetzung beteiligt sein. Amplitudenmodulation und Frequenzmodulation können in einem Klang gemeinsam auftreten und beeinflussen jeweils auch die zeitliche Hüllkurve eines Signals. Diese beiden Modulationen werden im weiteren Verlauf als zeitliche Modulationen bezeichnet und gemeinsam mit den spektralen Modulationen in eine gemeinsame Darstellung überführt - dem Modulation Power Spectrum. Im nächsten Kapitel werden die signalverarbeitungstechnischen Grundlagen erläutert, welche zur Herleitung des Modulation Power Spectrums in Kapitel 3 benötigt werden. In Kapitel 4 werden Filtermethoden präsentiert, welche es erlauben, Veränderungen in der MPS-Domäne vorzunehmen und diese über eine Signalresynthese wieder hörbar zu machen. Abschließend wird in Kapitel 5 ein verallgemeinertes Modell dieser Signalverarbeitungskette beschrieben, welches es ebenfalls ermöglicht, Klangmanipulationen vorzunehmen.

Kapitel 2

Signalverarbeitungsgrundlagen

Dieses Kapitel soll eine Einführung in die in dieser Arbeit verwendeten Signalverarbeitungsmethoden darstellen. Im ersten Schritt wird die Fourier-Transformation um eine Dimension erweitert, um eine Spektraldarstellung von einem Spektrogramm zu erstellen. Um zur Spektrogrammdarstellung und deren Eigenschaften zu gelangen, wird eine Kurzzeit-Fourier-Transformation (STFT) verwendet, die hier im zeitkontinuierlichen Fall beschrieben wird. Ein wichtiges Werkzeug für die Analyse von spektralen Modulationen ist das Cepstrum. Dazu wird ein kurzer Abriss gegeben, wie dieses entstanden ist und wie es in dieser Arbeit verwendet wird. Eine Signalresynthese wird nach der Transformation in die MPS-Domäne wieder angestrebt. Deswegen wird kurz die Signalresynthese in Zusammenhang mit der Spektrogramminversion erläutert. Ein wichtiges Werkzeug, um Bereiche in einer zweidimensionalen Fläche anheben oder absenken zu können, ist die zweidimensionale Gaußverteilung, welche am Ende dieses Kapitels beschrieben wird.

2.1 Kurzzeit-Fourier-Transformation

Während es die Fourier-Transformation nicht erlaubt, Aussagen über die zeitliche Änderung des Spektrums zu geben, ist die STFT auch für nichtstationäre Signale geeignet. Bei nichtstationären Signalen ist das Spektrum über die Zeit veränderlich, wonach es nicht mehr genügt, ein Spektrum des gesamten Signals zu berechnen. Sollen auch Spektren von kurzen Passagen in einem Signal ermittelt werden, muss das Signal in mehrere Teile unterteilt werden. Dies geschieht anhand der Fensterfunktion, mit welcher die so extrahierten Signalanteile in den Spektralbereich übergeführt werden können. Eine zeitliche Aneinan-

derreihung dieser Fenster ergibt die STFT, welche in folgender Form definiert ist:

$$\mathbf{STFT}\{x(t)\}(t', \mathrm{j}\omega) = X(t', \mathrm{j}\omega) = \int_{-\infty}^{+\infty} x(t) w(t - t') e^{-\mathrm{j}\omega t} \mathrm{d}t \tag{2.1.1}$$

wonach $w(t - t')$ die Fensterfunktion beschreibt. Durch die Fensterfunktion wird das Signal $x(t)$ außerhalb des Fensters unterdrückt. Dadurch wird es möglich, ein nichtstationäres Signal in Teilsignale zu unterteilen, in denen Quasistationarität[1] herrscht. Wird als Fensterfunktion ein Gaußfenster gewählt, spricht man auch von einer Gabor-Transformation [10]. Der Vorteil eines Gaußfensters ist, dass die Fourier-Transformation einer Gaußkurve wieder eine Gaußkurve ergibt:

$$\mathrm{e}^{-\alpha t^2} \quad \circ\!\!-\!\!\bullet \quad \sqrt{\frac{\pi}{\alpha}}\, \mathrm{e}^{-\frac{\omega^2}{\alpha}}. \tag{2.1.2}$$

wobei mit α die Breite der Gaußkurve gesteuert werden kann. Es ist auch zu sehen, dass die Fourier-Beziehung in Gl. 2.1.3 gültig ist: Eine Zeitdehnung ($|a| < 1$) entspricht einer Frequenzpressung und umgekehrt, eine Zeitpressung ($|a| > 1$) entspricht einer Frequenzdehnung.

$$x(at) \quad \circ\!\!-\!\!\bullet \quad \frac{1}{|a|} X\left(\mathrm{j}\frac{\omega}{a}\right) \tag{2.1.3}$$

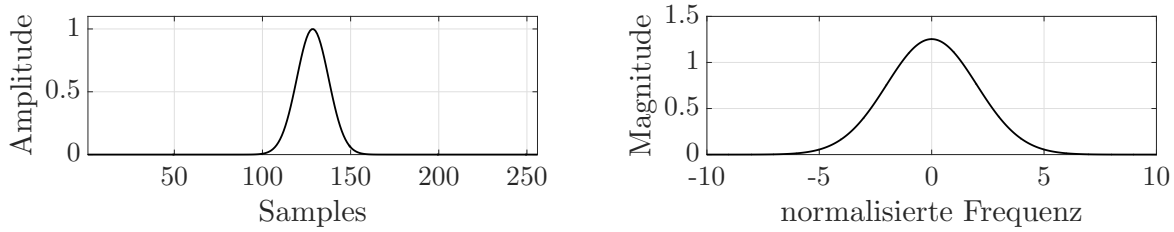
Dies steht unmittelbar in Zusammenhang mit dem Zeit-Bandbreite-Produkt [11, S. 467ff], welches besagt, dass die Zeitdauer eines Gaußfensters mit der Bandbreite seines Spektrums gekoppelt ist. Es gibt mehrere Definitionen der Zeitdauer und der Bandbreite eines Signals. Eine Möglichkeit besteht darin, ein Rechteck zu finden, welches die selbe maximale Amplitude und die gleiche Energie wie das Gauß-Signal aufweist. Ein weiteres Kriterium ist die -3dB-Dauer, welche jene Punkte angibt, bei der die Energie um die Hälfte der maximalen Energie gefallen ist. Für die -3dB-Dauer ergibt das Zeit-Bandbreite Produkt (ZBP) [12]:

$$T \cdot B = \frac{\ln(4)}{\pi} = 0.441. \tag{2.1.4}$$

Dieser Wert wird in der Literatur als optimal [11, S. 267ff] bezeichnet, wonach das Gaußfenster das kleinste Zeit-Bandbreite-Produkt besitzt.
Außerdem ist zu beachten, dass Gaußfenster eine unendliche Ausdehnung hin zu positiven und negativen Zeiten besitzen [13]. Um es dennoch einsetzen zu können, muss auch dieses

[1] Die Statistik bleibt nur innerhalb von kurzer Abschnitt gleich [9].

Fenster gefenstert oder abgeschnitten werden, was im Spektralbereich einer Faltung mit der Sinc-Funktion[2] entspricht. Üblicherweise wird das Fenster erst dort abgeschnitten, damit die Nebenkeulen im Spektrum keine hörbaren Artefakte mehr erzeugen. Ein Gaußfenster mit der Länge von 256 Samples ist in Abbildung 2.1 zu sehen. Dabei wird ersichtlich, was Gl. 2.1.2 ausdrückt: Die Fourier-Transformation eines Gaußsignals ergibt wieder ein Gaußsignal.



(a) Zeit-Bereich **(b)** Frequenzbereich

Abb. 2.1: Darstellung eines Gaußfensters nach Gl. 2.1.2 mit einer Fensterlänge von 256 Samples und einem Faktor $\alpha = 2$ im Zeit- und Frequenzbereich. Es ist zu beobachten, dass das Gaußfenster im Frequenzbereich die selbe Form aufweist.

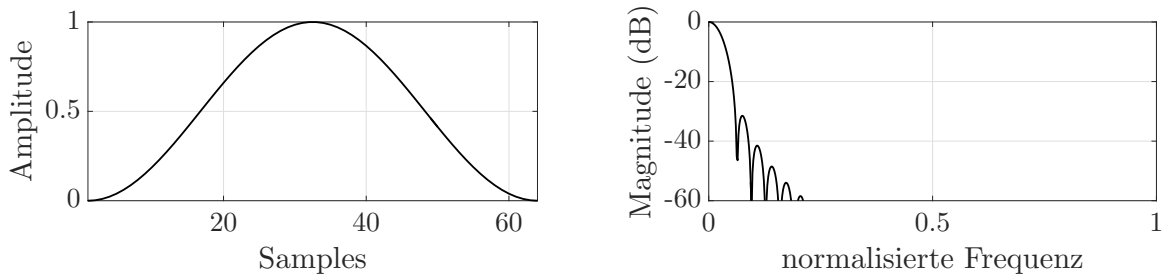
In der zeitdiskreten Signalverarbeitung werden üblicherweise die Fenster der STFT überlappend angeordnet. Aufgrund der Fensterung wird erreicht, dass keine Signalanteile ausgelassen werden und damit ein kontinuierlicher Übergang zwischen den Blöcken entsteht. Dabei spielt die *Constant Overlap-Add (COLA)* Eigenschaft eine wichtige Rolle. Diese ist im zeitdiskreten wie folgt definiert:

$$\sum_{m=-\infty}^{\infty} w(n - mR) = 1, \forall \in \mathbb{Z}. \tag{2.1.5}$$

Hierbei bedeutet m das m-te Fenster und R die Hopsize[3]. Für Gaußfenster wird diese Eigenschaft ungefähr dann erreicht, wenn die Hopsize sehr gering gewählt wird. Es ist nun ein Fenster zu wählen, bei dem es möglich ist, eine größere Hopsize zu verwenden, aber trotzdem Gl. 2.1.5 nicht zu verletzen. Weiter unten wird noch erläutert, wie die Hopsize mit der Abtastung des Spektrogramms zusammenhängt, was für die Grenzfrequenz der zeitlichen Modulationen von Bedeutung ist. Als Lösung für dieses Problem bietet sich das Hann-Fenster an, welches ebenfalls einen schnellen Abfall der Nebenkeulen aufweist und mit dem es möglich ist, mit vierfacher Überlappung Gl. 2.1.5 zu erfüllen (2.2).

[2] $\mathrm{sinc}(x) = \sin(x)/x$

[3] jene Anzahl an Samples, um die ein darauffolgendes Fenster verschoben wird.



(a) Zeit-Bereich **(b)** Frequenzbereich

Abb. 2.2: Darstellung eines Hann-Fensters mit einer Fensterlänge von 64 Samples im Zeit- und Frequenzbereich.

2.2 Spektrogramm

Ein Spektrogramm ermöglicht es, ein Signal $x(t)$ durch die Dekomposition in einzelne Frames und deren Transformation in den Spektralbereich über die Zeit darzustellen. Überlicherweise erfolgt die Spektrogrammdarstellung mit dem Quadrat der Magnitude der STFT von Gl. 2.1.1:

$$\textbf{SPEKTROGRAMM}\{x(t)\}(t', \omega) = X_s(t', \omega)) = |X(t', \mathrm{j}\omega)|^2 \tag{2.2.1}$$

In der Audiosignalverarbeitung wird allerdings die logarithmierte Darstellung verwendet. Um ein Spektrogramm in dB anzugeben, wird folgendes berechnet:

$$\textbf{SPEKTROGRAMM}\{x(t)\}(t', \omega) = X_s(t', \omega)) = 20 \cdot \lg|X(t', \mathrm{j}\omega)|. \tag{2.2.2}$$

Eine Berechnung nach Gl. 2.2.2 ist in Abbildung 2.3 zu sehen.

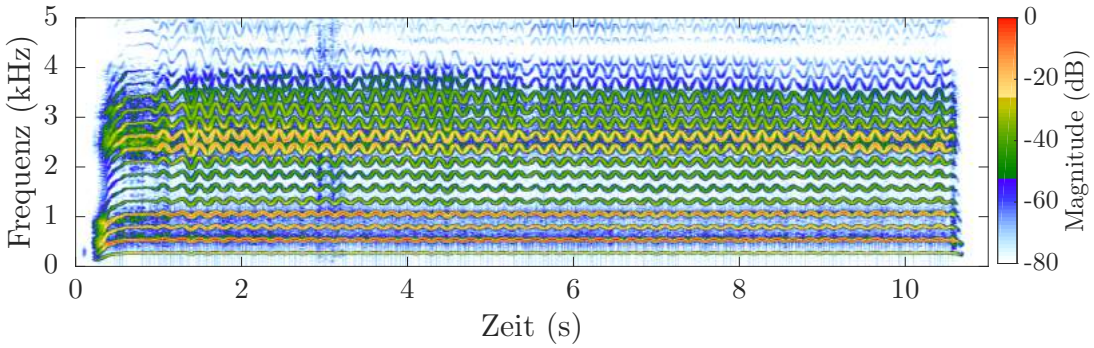


Abb. 2.3: Spektrogramm eines Baritonvokals. Dabei wurde die Frequenz-Achse linear dargestellt um den konstanten Abstand zwischen den Teiltönen zu verdeutlichen. Parameter für die Berechnung: Fenster: Gauß, Fensterlänge: 100 ms, Overlap-Faktor: 100

Dabei ist zu beobachten, dass sich die Frequenzen der Obertöne beim dargestellten Baritonvokal in einem ganzzahligen Verhältnis zum Grundton befinden. Durch die Grundfrequenz von ca. 250 Hz befinden sich ca. vier Obertöne innerhalb des Bereichs von einem kHz, das Auskunft über die spektralen Modulationen gibt. Um diese Periodizität zu bestimmen, wird für jedes Frame das Cepstrum berechnet, was im nächsten Abschnitt beschrieben wird. Es ist aber schon vorab festzuhalten, dass sich eine spektrale Modulation mit vier Zyklen/kHz einstellt.

2.2.1 Zeitliche und spektrale Modulationen

Im Bezug auf die zeitliche Veränderung des Spektrogramms lässt sich feststellen, dass die Frequenzen der Obertöne frequenzmoduliert sind und zwar mit einem steigenden Frequenzhub für höhere Frequenzen der Obertöne. Die Modulationsfrequenz der Frequenzmodulation der einzelnen Obertöne lässt sich ebenfalls über eine Spektraldarstellung bestimmen. Dies wird im nächsten Kapitel beschrieben, in dem das zeitliche Modulationsspektrum definiert wird. Im Abschnitt der zweidimensionalen Fourier-Transformation wird beschrieben, welche Auswirkungen Änderungen im Spektrogramm sowohl in der vertikalen Richtung (spektrale Modulationen), als auch in der horizontalen Richtung (zeitliche Modulationen) haben. Bei genauerer Betrachtung des Spektrogramms wird ebenfalls ersichtlich, dass auch eine Zwischenform spektraler und zeitlicher Modulationen vorkommen kann. Und zwar wenn beide gleichzeitig auftreten. Dies wird bei Glissandi nach oben und unten erreicht, welche in den Obertönen bei der Auf- und Abbewegung vorkommen. Diese Bewegungen nach oben und unten werden in der MPS-Domäne getrennt voneinander abgebildet. Demnach besteht ein Spektrogramm aus einer Komposition von jeweils zeitlichen und spektralen Modulationen und auch Kombinationen aus beiden. Die Obertöne des Baritonvokals in Abb. 2.3 ohne die Frequenzmodulationen kann als eine Welle mit der spektralen Modulation $\tau_i = 4$ Zyklen/kHz, und nicht als zeitliche Modulation gesehen werden. Andersherum kann die Modulation der Frequenzmodulation als eine Welle mit zeitlicher Modulation $f_{tmod} = 5$ Hz und keiner spektralen Modulation gedeutet werden[4]. Bewegungen nach oben und unten ergeben sich dadurch, dass Wellen aus spektralen und zeitlichen Modulationen erzeugt werden. Das gesamte Spektrogramm ergibt sich schlussendlich aus einer Summierung dieser Wellen. Da sich spektrale Modulationen über eine Fourier-Transformation der Frequenz-Achse ergeben, wird die Einheit als 1/Hz, oder um die Perioden in einem an-

[4] Dieser Wert wird im nächsten Kapitel ermittelt.

schaulicheren Bereich wiedergeben zu können, in 1/kHz angegeben, was auch als Quefrenz bezeichnet wird. Die Einheit der zeitlichen Modulationen wird in Hz angegeben, da diese das Ergebnis einer Fourier-Transformation über die Zeit sind.

2.2.2 Abtastung des Spektrogramms

Die Auflösung eines Spektrogramms ergibt sich einerseits über die Fensterlänge und andererseits über die Hopsize. Da ein Spektrum eines Signals wieder die selbe Länge wie der Fensterausschnitt hat, ergibt sich die Frequenzauflösung eines Spektrums mit der Abtastrate f_s und der Fensterlänge im Zeitbereich N_t mit:

$$f_\Delta = \frac{f_s}{N_t} \tag{2.2.3}$$

In diesem Fall werden von der STFT-Berechnung nur die positiven Frequenzen und der Gleichanteil verwendet, was eine Länge des Spektrums von $N_f = N_t/2 + 1$ Punkten ergibt. Dadurch ergibt sich entlang dieser Achse ein reellwertiges aber unsymmetrisches Signal. Das Ergebnis einer Fourier-Transformation über ein solches Signal ergibt ein komplexwertiges Cepstrum mit einem geraden Betrag und schiefsymmetrischen Phasengang. Bewegungen im Spektrogramm nach oben oder unten werden über die Phase bestimmt. Dies wird im weiteren Verlauf der Arbeit noch genauer erläutert. Bei Verwendung von positiven und negativen Frequenzen würde sich ein rein reelles Cepstrum ergeben, wonach nicht mehr zwischen Auf- und Abwärtsbewegungen unterschieden werden kann. Da vom berechneten Cepstrum nur der Betrag betrachtet wird und dieser gerade ist, können die negativen Quefrenzen vernachlässigt werden. Deshalb ergibt sich eine maximale Quefrenz in Abhängigkeit der Frequenzauflösung und der Anzahl der betrachteten Frequenzbins zu:

$$\max\{\tau\} = \left\lceil \frac{(N_f + 1)}{2} \right\rceil \cdot \frac{1}{f_\Delta N_f} \tag{2.2.4}$$

Bei einer Abtastrate von $f_s = 44100$ Hz ergibt sich bei einer Fensterlänge von 100 ms eine Sampleanzahl von $N_t = 4410$ im Zeitbereich. Für die betrachteten Samples im Frequenzbereich ergeben sich $N_f = 2206$ Samples und eine Frequenzauflösung über Gl. 2.2.3 von $f_\Delta = 10Hz$. Dies in Gl. 2.2.5 eingesetzt, ergibt eine maximale Quefrenz von 0.050 1/Hz oder 50 Zyklen/kHz.

Die maximale zeitliche Modulationsfrequenz ergibt sich über die Hopsize R und die Abta-

strate f_s:

$$\max\{\mathrm{f_{tmod}}\} = \frac{f_s}{2R} \tag{2.2.5}$$

Zum Beispiel ergibt sich bei einer Hopsize von $R = 40$ Samples eine Abtastrate von ca. 1000 Hz, die einer maximalen zeitlichen Modulationsfrequenz von 500 Hz entspricht. Da in natürlichen Klängen eine so große maximale Modulationsfrequenz nicht benötigt wird, ist es möglich, die Hopsize zu vergrößern. Da bei Hann-Fenstern schon eine vierfache Überlappung genügt, würde dies bei einer Fensterlänge von $N = 4410$ Samples eine Hopsize von 1102 Samples und über Gl. 2.2.5 eine maximale zeitliche Frequenz von 40 Hz ergeben, was für diese Zwecke als ausreichend angesehen werden kann. Weiters spielt auch die Fensterlänge eine wichtige Rolle bei der Bestimmung der zeitlichen Modulationen. Wird das Fenster zu lang gewählt, wird die Quasistationarität nicht mehr erreicht und es kann zum Beispiel die Frequenzmodulation nicht mehr richtig bestimmt werden, da über einen zu langen Zeitraum gemittelt wird. Wird hingegen eine zu große Hopsize verwendet, kann es zu Aliasing-Effekten aufgrund der Unterabtastung der Frequenzmodulation kommen.

2.3 Cepstrum

Die cepstrale Analyse von Signalen gehört zur Klasse der nichtlinearen Techniken und wurde 1963 von Bogert, Healy und Tukey definiert [14]. Der Begriff Cepstrum ergibt sich, wenn die ersten vier Buchstaben des Wortes Spectrum umgedreht werden. Damals wurde der Begriff des Power Cepstrums eines Signals definiert:

$$x_{pc} = \left|\mathcal{F}^{-1}\Big\{\log(|\mathcal{F}\{x(t)\}|^2)\Big\}\right|^2. \tag{2.3.1}$$

Mit dieser Berechnung wird der logarithmierte Betragsfrequenzgang eines Signals $x(t)$ anhand der Fourierrücktransformation wieder in den Zeitbereich gebracht und dann die quadrierte Magnitude berechnet. Allerdings handelt es sich dabei nicht mehr um den normalen Zeitbereich, sondern um den Quefrenzbereich [15]. In dieser Domäne werden andere Begriffe für die Signalverarbeitung verwendet, welche in Tabelle 2.1 angeführt sind.

Tab. 2.1: Definition der Bezeichnungen für das Cepstrum

normal	cepstrum
frequency	quefrency
spectrum	cepstrum
phase	saphe
amplitude	gamnitude
filtering	liftering
harmonic	rahmonic
period	repiod

Diese Begriffe bezeichnen Methoden, welche das selbe wie der Originalbegriff meinen, aber in der Quefrenzdomäne ausgeführt werden. Das Cepstrum wurde am Anfang verwendet, um Echos detektieren zu können. Dazu wird Abbildung 2.4 betrachtet. Ein Signal $x(t)$ und dessen verzögerte Version werden anschließend wieder zusammengeführt. Dieses Szenario kann zum Beispiel bei Echos oder Kammfiltern beobachtet werden.

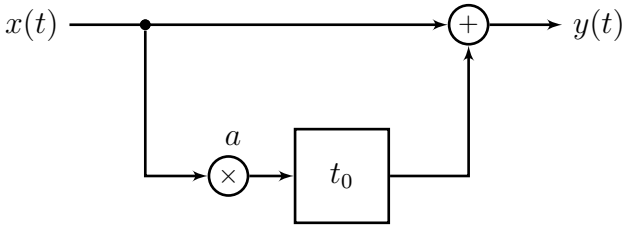
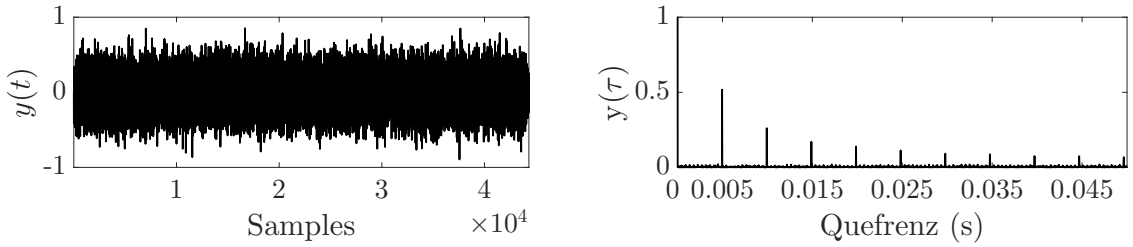


Abb. 2.4: Kammfilter

Das Cepstrum kann in diesem Fall dazu dienen, die Signallaufzeit des Systems zu detektieren. Als Beispiel wird normalverteiltes Rauschen mit einer um $t_0 = 5$ ms verzögerten Version überlagert. Das resultierende Signal $y(t)$ ergibt sich aus der Faltung des Eingangssignals mit der Impulsantwort des Systems. Eine Faltung im Zeitbereich entspricht einer Multiplikation im Frequenzbereich. Anhand der homomorphen Signalverarbeitung ist es nun möglich, Periodizitäten im Spektrum, die sich durch Einbrüche aufgrund des Kammfilters ergeben, detektieren zu können. Eine Anwendung des Logarithmus auf die multiplikative Verbindung des Eingangssignals mit der Impulsantwort des Systems resultiert in einer additiven Aufspaltung der beiden Spektren.
Nach anschließender Fourierrücktransformation werden die periodischen Anteile im Cepstrum als wiederkehrende Peaks dargestellt, die die Zeitverzögerung des Systems wiedergeben.



(a) Ausgangssignal **(b)** Power Cepstrum

Abb. 2.5: Darstellung des Ausgangssignals $y(t)$ in (a) und dessen Cepstrum (b).

2.4 Zweidimensionale Fourier-Transformation

Ähnlich, wie ein eindimensionales Signal aus Basissignalen besteht, besteht ein Bild aus Basisbildern [17, S. 44ff]. Um ein Spektrogramm in seine Basisbilder zerlegen zu können, werden Grundlagen der Bildverarbeitung benötigt. Diese sollen hier kurz erläutert werden, um in die Eigenschaften und Methoden der zweidimensionalen Fourier-Transformation überleiten zu können. Mit diesen Basisbildern ist es möglich, jedes Bild zu erzeugen. Dabei spannen diese Basisbilder eine orthonormale Basis auf. Dieser Sachverhalt kann anhand des Skalarproduktes analysiert werden: Bei zwei verschiedenen Basisbildern ergibt das Skalarprodukt null, da diese eine orthonormale Basis bilden. Ein Skalarprodukt eines Bildes mit sich selbst ergibt eins. Ein Bild mit einer Dimension von M × N Punkten besitzt demnach M × N Basisbilder. Diese spannen einen M × N-dimensionalen Vektorraum über dem Körper der reellen Zahlen auf. Demnach repräsentiert ein M × N Bild einen Punkt im M × N-dimensionalen Vektorraum. Genau wie in der Fourieranalyse von eindimensionalen Signalen ist es nun möglich, die Intensität der Basisbilder, die in einem Bild vorkommen, zu beschreiben. Dazu wird Abbildung 2.6 betrachtet.

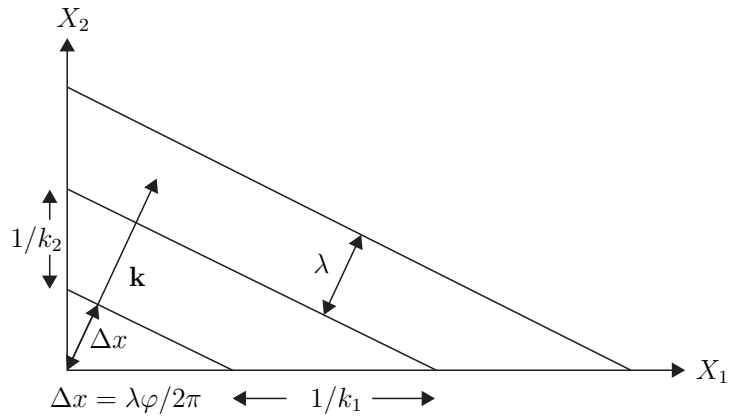


Abb. 2.6: Beschreibung eines Basisbildes. Der Vektor $\mathbf{k}$ liegt im zweidimensionalen Wellenzahlraum und gibt die Richtung des Basisbildes an. Seine Komponenten k_1 und k_2 beschreiben die Anzahl der Wellenlängen in die jeweilige Richtung.

Ein Basisbild kann durch ein periodisches Muster mit dem Wellenzahlvektor $\mathbf{k}$, der die Richtung angibt, beschrieben werden. Dieser Vektor steht normal auf die Linien konstanter Breite und beinhaltet die Richtungsinformation des Musters. Dabei besitzt der Vektor die Länge $|k| = 1/\lambda$. Der Vektor $\mathbf{k}$ liegt im Wellenzahlraum der Dimension zwei. Seine Komponenten geben die Anzahl der Wellenlängen pro Einheitslänge in die jeweilige Richtung des Wellenzahlraumes an: $\mathbf{k} = [k_1, k_2]^T$. Seine Phase wird über die Beziehung $\varphi = 2\pi\Delta x/\lambda = 2\pi\mathbf{k}\Delta x$ beschrieben. Der Wellenzahlvektor $\mathbf{k}$ kann beliebig auf zusätzliche Dimensionen erweitert werden. Um dieses Muster mathematisch ausdrücken zu können, wird ähnlich einer ebenen Welle ohne der zeitlichen Komponente der Ausdruck in komplexer Form angeschrieben:

$$\Re\left(g\exp(\mathrm{j}2\pi\mathbf{k}^T\mathbf{x} - \Phi)\right) = g\cos(2\pi\mathbf{k}^T\mathbf{x} - \Phi) \tag{2.4.1}$$

Dabei besitzt ein reelles Bild keinen Imaginärteil. Um ein Bild in seine Basisbilder zerlegen zu können, wird die zweidimensionale Fourier-Transformation eingeführt:

$$\mathcal{F}[f(m,n)] = F(u,v) = \frac{1}{\sqrt{MN}} \sum_{m=0}^{M-1}\sum_{n=0}^{N-1} f(m,n)\exp\Big[-\mathrm{j}2\pi\Big(\frac{um}{M} + \frac{vn}{N}\Big)\Big] \tag{2.4.2}$$

$$\mathcal{F}^{-1}[F(u,v)] = f(m,n) = \frac{1}{\sqrt{MN}} \sum_{u=0}^{M-1}\sum_{v=0}^{N-1} F(u,v)\exp\Big[\mathrm{j}2\pi\Big(\frac{um}{M} + \frac{vn}{N}\Big)\Big] \tag{2.4.3}$$

Dabei wird ein Bild $f(m,n)$ mit allen Basisbildern multipliziert und bis zu den Werten

M und N summiert und normiert. In dieser Domäne kann nun die Intensität einzelner Basisbilder, welche im Bild enthalten sind, visualisiert werden. Dabei wird im Fall, dass es sich beim transformierten Bild um ein Spektrogramm mit logarithmierten Magnituden handelt, die x-Achse als die Achse der zeitlichen Modulationen $\mathrm{f_{tmod}}$ mit der Einheit Hz bezeichnet. Auf der y-Achse werden die spektralen Modulationen τ mit der Einheit 1/kHz dargestellt, da es sich bei dieser Achse um eine cepstrale Darstellung der Modulationen handelt. Dabei sind Basisbilder mit einer geringeren Wellenlänge oder einer höheren Frequenz weiter vom Ursprung entfernt als Basisbilder mit geringeren Frequenzen. Es können nun einzelne Bereiche vom Ausgangsbild, dem Spektrogramm, einzeln analysiert und diskutiert werden. Zum Beispiel werden Muster in einem Bild, die einem Basisbild entsprechen, bei denen der Vektor $\mathbf{k}$ nur eine y-Komponente aufweist, auf der y-Achse mit einer bestimmten Entfernung vom Ursprung aufgetragen (Abbildung 2.7). Die Distanz vom Ursprung ergibt sich durch die Anzahl der Perioden innerhalb von einem kHz im Spektrogramm. Demnach erfolgt eine Verteilung wie beim Cepstrum: besitzt ein Klang einen Grundton bei 200 Hz und Obertöne im Abstand von ganzzahligen Vielfachen, so entsteht eine spektrale Modulation bei 5 Zyklen/kHz. Bewegt sich der Grundton zu tieferen Frequenzen, fallen mehr Obertöne innerhalb von einem kHz an und die spektrale Modualtion bewegt sich auf der τ-Achse nach oben. Besitzt hingegen ein Klang eine Amplituden- oder Frequenzmodulation und weist so periodische Muster entlang der Zeit-Achse auf, werden damit zeitliche Modulationen beschrieben, die anhand der zweidimensionalen Fourier-Transformation auf der $\mathrm{f_{tmod}}$-Achse aufgetragen werden. Klänge, bei denen sich der Grundton und seine Obertöne entweder nach oben oder unten bewegen, werden in der Darstellung entweder links oder rechts abgebildet. Der Bereich, in dem eine solche Bewegung abgebildet wird, ergibt sich einerseits durch die Anzahl an Obertönen innerhalb von einem kHz und andererseits durch die Anzahl an zeitlichen Modulationen innerhalb einer Sekunde.

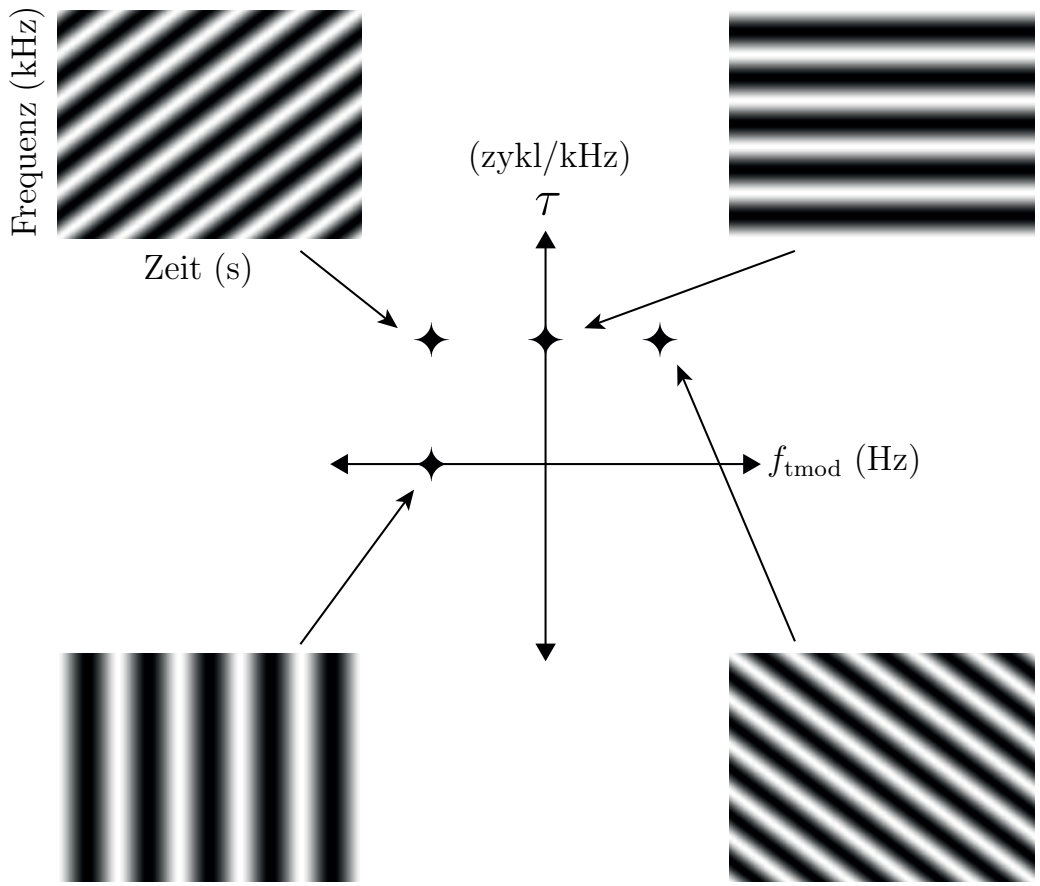


Abb. 2.7: Veranschaulichung der zweidimensionalen Fourier-Transformation. Wellen in einem Spektrogramm, die periodische Muster nur in y-Richtung aufweisen und somit ein konstantes spektrales Verhalten über die Zeit aufweisen, werden entlang der τ-Achse aufgetragen. Der Abstand vom Ursprung ergibt sich über die cepstrale Berechnung: Besitzt ein Klang einen Grundton bei 200 Hz und Obertöne im Abstand von ganzzahligen Vielfachen, so entsteht eine spektrale Modulation bei 5 Zyklen/kHz. Bewegt sich der Grundton zu tieferen Frequenzen, fallen mehr Obertöne innerhalb eines kHz und die spektrale Modulation bewegt sich auf der τ-Achse nach oben. Besitzt hingegen ein Klang eine Amplituden- oder Frequenzmodulation und weist so periodische Muster entlang der Zeit-Achse auf, werden damit zeitliche Modulationen beschrieben, die anhand der zweidimensionalen Fourier-Transformation auf der f_{tmod}-Achse aufgetragen werden. Klänge, bei denen sich der Grundton und seine Obertöne entweder nach oben oder unten bewegen, werden in der Darstellung entweder links oder rechts abgebildet. Der Bereich, in dem eine solche Bewegung abgebildet wird, ergibt sich einerseits durch die Anzahl an Obertönen innerhalb von einem kHz und andererseits durch die Anzahl an zeitlichen Modulationen innerhalb einer Sekunde.

Es werden im nächsten Kapitel noch zwei weitere Darstellungsweisen präsentiert, die es erlauben, zuerst nur eine Fourier-Transformation entlang der Zeit-Achse und erst dann über die Frequenz-Achse durchzuführen. Ersteres wird als zeitliches Modulationsspektrum und Zweiteres als Cepstrogramm bezeichnet. Das Ergebnis einer Fourier-Transformation über

beide Achsen wird als Modulation Power Spectrum bezeichnet. Um diese verschiedenen Transformationen in kompakter Form ausdrücken zu können, wird folgende Nomenklatur definiert. Die Berechnung des Modulation Power Spectrums erfolgt anhand von:

$$\mathbf{MPS}(\mathrm{f_{tmod}}, \tau) = X_{\mathrm{MPS}}(\mathrm{f_{tmod}}, \tau) = \left| \mathcal{F}_{1,1} \left[\log|X(t', \mathrm{j}\omega)| \right] \right|. \tag{2.4.4}$$

Dabei bezeichnet der erste Index eine Fourier-Transformation entlang der Zeit-Achse und der zweite Index entlang der Frequenz-Achse. Eine Transformation in beide Richtungen wird als MPS bezeichnet. Dies soll Tabelle 2.2 verdeutlichen.

Tab. 2.2: Indizes für die Fourier-Transformation

$\mathrm{t_{mod}}$	$\mathrm{f_{mod}}$	Transformation
1	0	Fourier-Transformation entlang der Zeit-Achse (zeitliches Modulationsspektrum)
0	1	Fourier-Transformation entlang der Frequenz-Achse (Cepstrogramm)
1	1	2D Fourier-Transformation (MPS)
-1	0	inverse Fourier-Transformation entlang der Zeit-Achse
0	-1	inverse Fourier-Transformation entlang der Frequenz-Achse
-1	-1	inverse 2D Fourier-Transformation (Spektrogramm)

Aufgrund der Linearität der Fourier-Transformation können Transformationen zuerst entlang der einen Achse und dann entlang der anderen Achse ohne weiteres vertauscht oder hintereinander ausgeführt werden:

$$\mathbf{MPS}(\mathrm{f_{tmod}}, \tau) = X_{\mathrm{MPS}}(\mathrm{f_{tmod}}, \tau) = \left| \mathcal{F}_{1,1} \left[\log|X(t', \mathrm{j}\omega)| \right] \right| = \left| \mathcal{F}_{1,0} \left[\mathcal{F}_{0,1} \left[\log|X(t', \mathrm{j}\omega)| \right] \right] \right| \tag{2.4.5}$$

2.5 Spektrogramm Inversion

Da im weiteren Verlauf dieser Arbeit von einem Spektrogramm wieder auf ein Audiosignal zurückgeführt werden soll, wird dies kurz anhand vom Griffin und Lim-Algorithmus [18] erläutert. Dies ist zum Beispiel dann nötig, wenn im Spektrogramm Manipulationen vorgenommen wurden oder einzelne STFT-Frames bearbeitet wurden. Dabei wird auch unterschieden, ob eine Phase vorhanden ist, oder ob nur die Betragsspektren der STFT zur Verfügung stehen. Ist eine Phase vorhanden, können die einzelnen Frames der STFT über die Transformation auf Polarkoordinaten wieder auf die komplexe Form $|| \cdot || \, e^{\mathrm{j}\phi}$ gebracht werden. Dies wird als Y_w bezeichnet. Ist hingegen keine Phase vorhanden, werden nur die Magnituden zur Signalresynthese herangezogen. Diese Frames werden anschließend über die Overlap-Add-Methode und anschließender Fensterung mit dem gewählten Fenster, auf ein Signal $x(n)$ rücktransformiert. Von diesem Signal wird wieder ein Spektrogramm berechnet, welches als X_w bezeichnet wird. Um nun die Qualität der Resynthese bewerten zu können, wird folgende Distanzmessung [18] eingeführt:

$$D[x(n), Y_w(mS, \omega)] = \sum_{m=-\infty}^{\infty} \frac{1}{2\pi} \int_{\omega=-\pi}^{\pi} |X_w(mS, \omega) - Y_w(mR, \omega|^2 \mathrm{d}\omega. \tag{2.5.1}$$

Dabei wird R als die Hopsize bezeichnet, die die Abtastung des Spektrogramms entlang der Zeit-Achse angibt. Hier wird zuerst der quadrierte Fehler zwischen X_w und Y_w berechnet und über alle ω integriert. Weiters wird über alle Frames summiert. Der Fehler wurde in Abhängigkeit von $x(n)$ und Y_w angegeben, da X_w eine gültige STFT darstellt, wohingegen Y_w nicht zwingend eine solche ergibt. Da das Theorem von Parseval gültig sein muss, kann Gl. 2.5.1 auch im Zeitbereich beschrieben werden:

$$D[x(n), Y_w(mR, \omega)] = \sum_{m=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} (x_w(mR, l) - y_w(mR, l))^2. \tag{2.5.2}$$

Da Gl. 2.5.2 eine quadratische Form von $x(n)$ beschreibt, kann die Minimierung von $D[x(n), Y_w(mR, \omega)]$ über eine Nullsetzung des Gradienten und einer Auflösung nach $x(n)$ erreicht werden:

$$x(n) = \frac{\sum_{m=-\infty}^{\infty} w(mR - n) y_w(mR, n)}{\sum_{m=-\infty}^{\infty} w^2(mR - n)}. \tag{2.5.3}$$

Dabei bezeichnet $w(mR - n)$ ein um mR verschobenes Fenster.

Der Algorithmus, in dem eine optimale Schätzung des Signals berechnet wird, geht davon aus, dass der quadratische Fehler zwischen X_w und Y_w bei jeder Iteration sinkt. Dabei wird $x^i(n)$ als das geschätzte Signal nach der i-ten Iteration bezeichnet. Die i + erste Schätzung wird über die STFT-Berechnung von $x^i(n)$ erreicht, wohingegen die Magnitude von $X_w^i(mR, \omega)$ durch die Magnitude von $Y_w(mR, \omega)$ ersetzt wird und dann der geringste Abstand mit Gl. 2.5.2 gefunden wird. Diese Berechnung ist in Abb. 2.8 zu sehen.

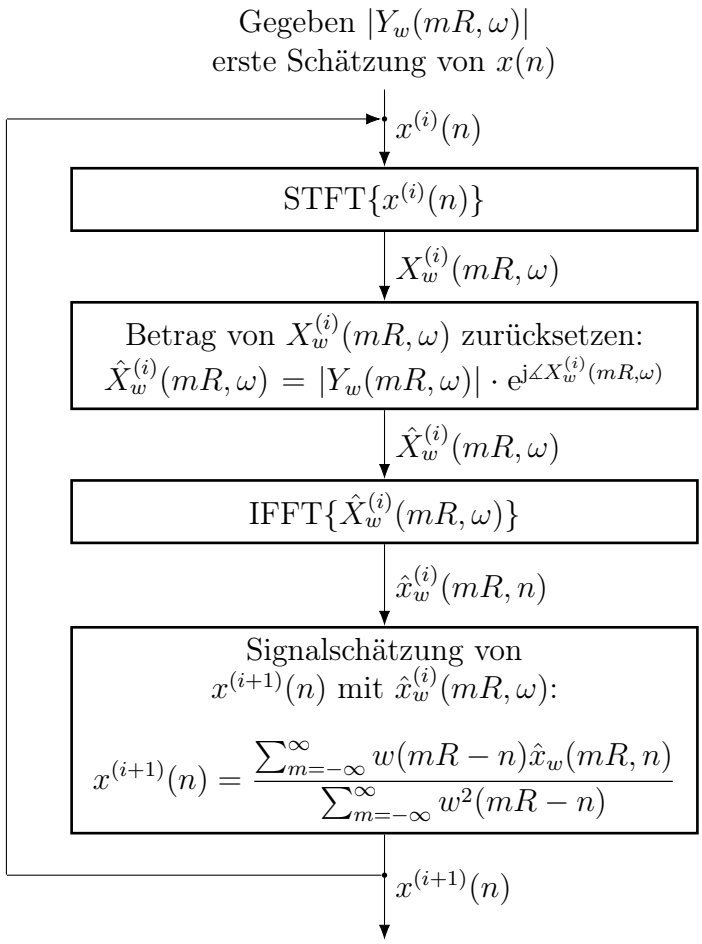


Abb. 2.8: Flussdiagramm des Algorithmus von Griffin und Lim.

Die Anzahl der Iterationen kann vor der Resynthese bestimmt werden. Wenn nun der quadratische Fehler des gesamten Spektrums und des geschätzten Spektrums einen Wert unterschreitet und die maximale Anzahl an Iterationen noch nicht erreicht wurde, wird die Resynthese abgebrochen.

2.6 Mehrdimensionale Gaußverteilungen

Die multivariate Gauß- oder Normalverteilung stellt eine Verallgemeinerung der eindimensionalen Gaußverteilung dar. Diese wird auch bivariate Gaußverteilung genannt. Eine vektorbasierte Zufallsvariable $X = [X_1 \dots X_n]^T$ hat eine multivariate Gaußverteilung mit Mittelwert $\mu \in \mathbb{R}^n$ und der Kovarianzmatrix $\Sigma \in \mathbb{R}^{n \times n}$, wenn die Dichteverteilungsfunktion wie folgt lautet:

$$p(\mathbf{x}; \mu, \Sigma) = \frac{1}{(2\pi)^{n/2}|\mathbf{\Sigma}|^{1/2}} \exp\Big(-\frac{1}{2}(\mathbf{x} - \mu)^T \Sigma^{-1} (\mathbf{x} - \mu)\Big) \tag{2.6.1}$$

Dies kann auch als $X \sim \mathcal{N}(\mu, \Sigma)$ notiert werden. Für den Zweck dieser Arbeit wird der Normierungsterm weggelassen und mit einem Verstärkungsfaktor β ersetzt:

$$p_G(\mathbf{x}; \mu, \Sigma) = \beta \cdot \exp\Big(-\frac{1}{2}(\mathbf{x} - \mu)^T \Sigma^{-1} (\mathbf{x} - \mu)\Big) \tag{2.6.2}$$

Dabei bezeichnet μ_1 den Mittelwert für die zeitlichen Modulationen und μ_2 den Mittelwert für die spektralen Modulationen. Weiters sind x_1 und x_2 Variablen, die die Dimension der zweidimensionalen Filterfunktion angeben.

$$p_G(x_1, x_2; \mu_1, \mu_2, \Sigma) = \beta \cdot \exp\left(-\frac{1}{2}\begin{pmatrix} x_1 - \mu_1 \\ x_2 - \mu_2 \end{pmatrix}^T \Sigma^{-1} \begin{pmatrix} x_1 - \mu_1 \\ x_2 - \mu_2 \end{pmatrix}\right) \tag{2.6.3}$$

Für den Fall $n = 2$ und nichtkorrelierter Zufallsvariablen wird die Kovarianzmatrix zu einer 2×2 Matrix mit den Einträgen auf der Nebendiagonalen gleich 0.

$$\begin{aligned} p_G(x_1, x_2; \mu_1, \mu_2, \Sigma) &= \beta \cdot \exp\left(-\frac{1}{2}\begin{pmatrix} x_1 - \mu_1 \\ x_2 - \mu_2 \end{pmatrix}^T \begin{bmatrix} 1/\sigma_1^2 & 0 \\ 0 & 1/\sigma_2^2 \end{bmatrix} \begin{pmatrix} x_1 - \mu_1 \\ x_2 - \mu_2 \end{pmatrix}\right) \\ &= \beta \cdot \exp\left\{ -\frac{1}{2}\left[\left(\frac{x_1 - \mu_1}{\sigma_1}\right)^2 + \left(\frac{x_2 - \mu_2}{\sigma_2}\right)^2 \right]\right\} \end{aligned} \tag{2.6.4}$$

Für $\sigma_1 = 2$, $\sigma_2 = 2$ und $\beta = 1$ ist dies in Abbildung 2.9 zu sehen.

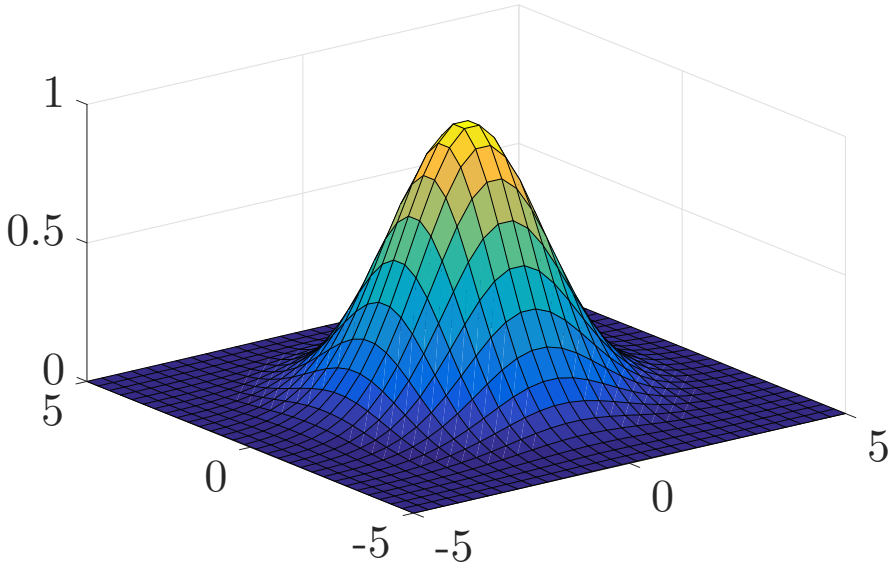


Abb. 2.9: Multivariate Gaußverteilung mit $n = 2$, $\sigma_1 = 2$, $\sigma_2 = 2$, $\Sigma_{12} = \Sigma_{21} = 0$ und $\beta = 1$. Darstellung ohne Normierungsterm.

2.7 Tetration

Schon zu Anfang des 20. Jahrhunderts überlegte sich Wilhelm Ackermann [19], wie es möglich ist, Funktionen zu bilden, die stärker steigen als herkömmliche Funktionen. 1926 definierte Ackermann Funktionen, die nicht primitiv-rekursiv[5] sind, aber von einem Computer in endlicher Zeit ausgewertet werden können. Dazu wird die Folge $a + b, a \cdot b, a^b, \ldots$ betrachtet. Es ist zu sehen, dass bei jedem Folgeglied die Operation des vorigen Folgengliedes $(b-1)$-mal auf a angewandt wird. Zum Beispiel ist $a \cdot b = a + a + a + \cdots + a$ wobei die Variable a b-mal vorkommt. Ackermann versuchte nun diese Idee als Funktion aufzufassen. Setzt man in obiger Folge für $a = 2$ und für $b = 4$ ein, so erhält man die Folge:

[5] sind Funktionen, die aus einfachen Grundfunktionen (Addition, Multiplikation, Potenz, ...) gebildet werden können

$6, 8, 16, 65536$. Ackermann notierte dies folgendermaßen:

$$\begin{aligned}\varphi(a,b,0) &= a+b\\ \varphi(a,b,1) &= a\cdot b\\ \varphi(a,b,2) &= a^b\\ &\dots\end{aligned} \tag{2.7.1}$$

Ab der vierten Zeile ist es nicht mehr möglich, dies mit herkömmlichen Operatoren[6] durchzuführen. Deswegen wurde der Hyperoperator eingeführt, der es erlaubt, auch Operationen jenseits der Potenz zu notieren. Dies wird allgemein folgendermaßen als dreistelliger[7] Operator (mit $(a, b, n > 0)$) notiert:

$$a^{(n)}b := \begin{cases} a, & \text{wenn} = 1, b = 0\\ 0, & \text{wenn} n = 2, b = 0\\ 1, & \text{wenn} n > 2, b = 0\\ a^{(n-1)}(a^{(n)}(b-1)), & \text{sonst}\end{cases}$$

was zu $\text{hyper}(a, n, b) = a^{(n)}b$ führt.
Zum Beispiel bedeutet $3^{(3)}3 = 3^3 = 3\cdot 3\cdot 3 = 27$. Das n gibt die Operation an, die ausgeführt wird. Dabei steht 1 für die Addition, 2 für die Multiplikation, 3 für die Potenz und 4 für die Tetration. Für die Tetration gibt es nun mehrere Schreibweisen. Dabei sind $a^{(4)}b$ und ${}^b a$ äquivalent.

Allgemein kann dies wie folgt geschrieben werden:

$$\exp_b^c(z) = \underbrace{\exp_b\Big(\exp_b(\dots\exp_b(z)\dots)\Big)}_{\text{c mal}}. \tag{2.7.2}$$

Dies wird ebenfalls als Tetration[8] bezeichnet und dabei steht b für die Basis der Exponentiation und c gibt an, wie oft die Basis b potenziert wird. Dabei wird die Tetration

[6] Addition, Subtraktion, Multiplikation, Division.
[7] Ein zweistelliger Operator (Addition, Multiplikation, etc.) benötigt zwei Operanden und wird demnach als zweistelliger Operator bezeichnet, hingegen wird ein Hyperoperator in diesem Fall als dreistellig bezeichnet, da n angibt, welche Operation ausgeführt werden soll.
[8] andere Bezeichnungen dafür sind: generalized exponential function, ultraexponentiation oder super exponentiation.

in der Hierachie der Operatoren nach der Addition, Multiplikation und Potenz als vierter Operator angeführt. Im Gegensatz zu den Elementarfunktionen, die auf die ersten drei Operatoren zutreffen, wurde bis vor kurzem noch keine Definition für reelle und komplexe Zahlen gefunden. Erst 2010 definierte der russische Mathematiker Kouznetsov [20] [21] ein Set von Funktionen, mit denen die Tetration auch auf die komplexen Zahlen erweitert wurde. In Abbildung 2.10 ist die Tetration gemeinsam mit der Exponentialfunktion dargestellt.

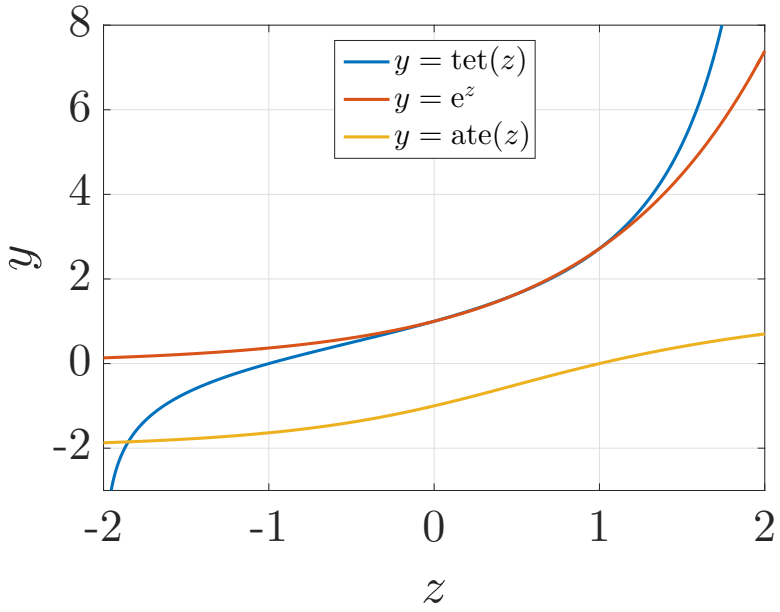


Abb. 2.10: Tetrations-, Exponential, und Arctetrationsfunktion. Wie zu sehen ist, steigt die Tetrationsfunktion ab dem Wert $x = 1$ stärker an wie die Exponentialfunktion.

Tetration wird dazu verwendet, um die Exponentialfunktion ($b = \mathrm{e}$) hintereinander ausführen zu können. Dies wird als natürliche Tetration bezeichnet. Für nicht ganzzahlige Werte von c wird die verallgemeinerte Exponentialfunktion folgendermaßen definiert[9]:

$$\exp^c(z) = \mathrm{tet}\Big(c + \mathrm{tet}^{-1}(z)\Big). \tag{2.7.3}$$

Hierbei wird tet^{-1} als die inverse Tetration bezeichnet. Für $c = 1$ ergibt sich die normale Exponentialfunktion und für $c = -1$ die Umkehrung davon, die Logarithmusfunktion. Wird für $c = 0$ eingesetzt, ergibt sich über 2.7.3 $\exp^0(z) = z$.

[9] dabei ist $\exp_b^1 = \exp_b$ und wenn der Index ausgelassen wird, spricht man von der Basis e, also $\exp^c = \exp_{\mathrm{e}}^c$.

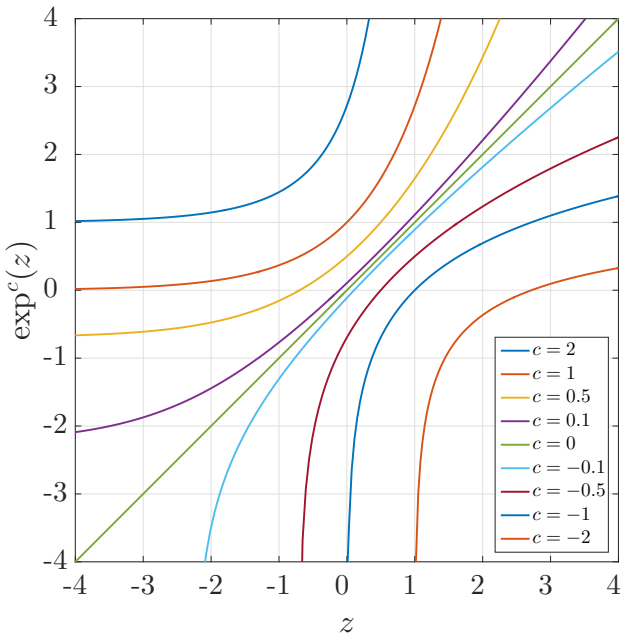


Abb. 2.11: Darstellung von Gl. 2.7.3 für verschiedene c-Werte. Es ist zu beobachten, dass der Fall $c = 1$ der Exponentialfunktion und $c = -1$ der Logarithmusfunktion entspricht.

Aufgrund von 2.7.3 ist es möglich, auch von komplexen Zahlen Zwischenformen von e^x bis $\log(x)$ zu berechnen. Dies findet sich im letzten Kapitel, wo die Tetration verwendet wird, um Signalverarbeitungsketten und deren Inversion zu erzeugen, die mit der Verwendung der Standardfunktionen e^x und $\log(x)$ nicht exakt möglich wären. Im Zuge dieser Arbeit wurde die Implementierung der C-Funktionen von [22] und [23], welche auf [20] basieren, in MATLAB durchgeführt.

Kapitel 3

Modulation Power Spectrum

In diesem Abschnitt wird die Berechnung des Modulation Power Spectrums beschrieben. Dazu werden die Signalverarbeitungsgrundlagen aus Abschnitt zwei benötigt. Im weiteren Verlauf dieses Kapitels werden zwei weitere Darstellungsmöglichkeiten präsentiert, mit denen es möglich ist, entweder spektrale oder zeitliche Modulationen darstellen zu können. Um dies anhand von einem Beispiel zu erläutern, wird eine geschlossene Lösung anhand eines Gauß-Chirps erläutert. Weitere Beispiele sollen den Sachverhalt verdeutlichen.

3.1 Cepstrogramm

Die STFT eines Zeitsignals $x(t)$ berechnet sich mit Gl. 2.1.1. Um daraus das Spektrogramm für akustische Zwecke zu berechnen, wird Gl. 2.2.2 verwendet. Das Cepstrum der einzelnen Frames ergibt sich nun über die Fourierrücktransformation und anschließender Betragsbildung. Werden diese Cepstren über die Zeit aufgetragen, erhält man das Cepstrogramm.

$$\textbf{CEPSTROGRAM}\{x(t)\}(t',\tau) = X_c(t',\tau) = \left|\mathcal{F}_{0,1}\left[\log|X(t',\mathrm{j}\omega)|\right]\right| \tag{3.1.1}$$

Dieses gibt den zeitlichen Verlauf der Cepstren über die Zeit an. Die Abszisse bleibt bestehen und gibt nach wie vor die Zeitpunkte der einzelnen Frames an. Aufgrund der cepstralen Berechnung ergibt sich für die neue Frequenz-Achse eine Quefrenzachse in Zyklen pro kHz. Anhand dieser Darstellung ist es nun möglich, spektrale Modulationen über die Zeit zu betrachten, die Aufschluss darüber geben, wie sich die Obertonzusammensetzung eines Klanges über die Zeit verändert.

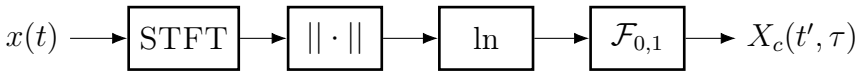
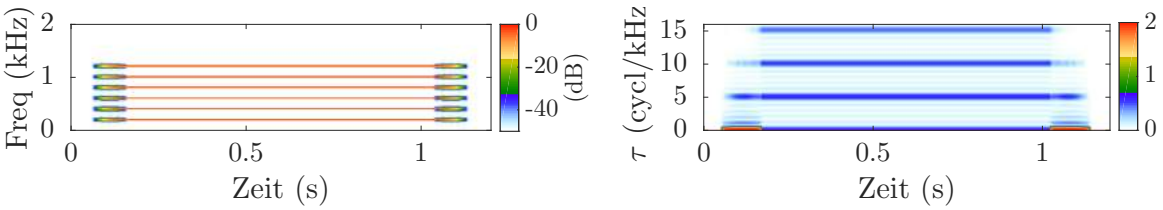


Abb. 3.1: Zur Berechnung des Cepstrogramms.

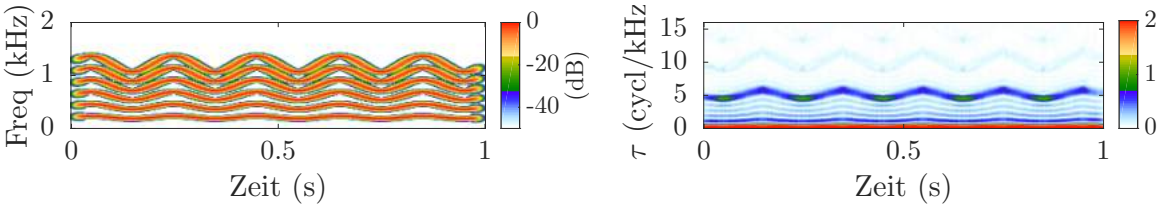
Verändert sich zum Beispiel der Grundton eines Klanges über die Zeit nicht und besitzt dieser Obertöne, die im ganzzahligen Verhältnis zu ihm stehen, ergibt auch das Cepstrogramm Linien parallel zur Zeit-Achse, die sich nicht verändern (Abb. 3.2). Die vertikale Position der ersten Linie im Cepstrogramm ergibt sich aus der Anzahl an Wellenbergen im Spektrogramm innerhalb von einem kHz. Die Rahmonischen parallel zur ersten Linie stehen auch zu dieser in einem ganzzahligen Verhältnis. Dabei ist zu erkennen, dass die Gamnitude bei höheren Quefrenzen abnimmt. Dies resultiert daraus, dass höhere Rahmonische weniger zur Formung des Spektrums durch die Cepstren beitragen als der erste Rahmonische.

Werden hingegen der Grundton und seine Obertöne frequenzmoduliert, ergibt sich je nach Position des minimalen und maximalen Modulationshubes auch im Cepstrogramm eine Modulation der Rahmonischen (3.3). Die Modulation entsteht dadurch, dass sich die Dichte der spektralen Obertöne je nach Modulationshub vergrößert (höhere spektrale Modulation) oder verkleinert (niedrigere spektrale Modulation). Dieser Fall der Abhängigkeit des Modulationshubes von der Tonhöhe ist auch in musikalischen Klängen zu finden. Beim Vibrato eines Sängers oder einer Sängerin ist dies ebenfalls zu beobachten, was in Abbildung 3.5 zu sehen ist. Der Fall, dass sich der Modulationshub bei steigenden Obertönen nicht ändert, erzeugt im Cepstrogramm das selbe Bild wie nicht frequenzmodulierte Obertöne (Abb. 3.4). Da sich auch hier die Dichte der Obertöne über die Zeit nicht verändert, verändern sich auch die Rahmonischen im Cepstrogramm über die Zeit nicht.



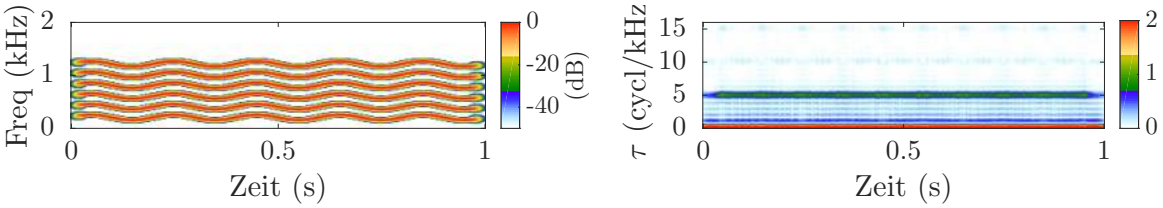
(a) Spektrogramm **(b)** Cepstrogramm

Abb. 3.2: Darstellung eines Spektrogramms und dessen Cepstrogramm eines Sinustons, der seine Frequenz nicht ändert und fünf Obertöne besitzt. Parameter: $f_0 = 200$ Hz.



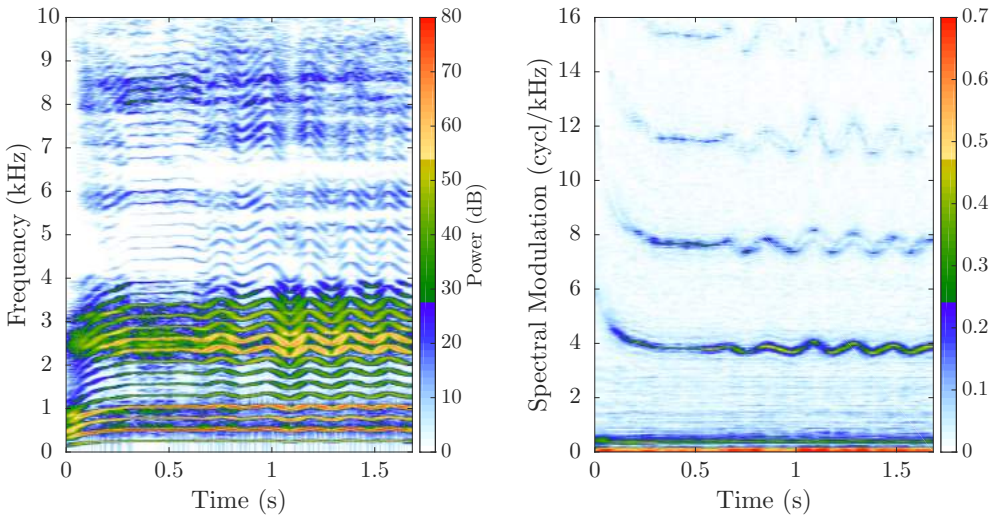
(a) Spektrogramm **(b)** Cepstrogramm

Abb. 3.3: Darstellung eines Spektrogramms und dessen Cepstrogramm, bei dem ein Sinuston und seine fünf Obertöne ($N = 6$) frequenzmoduliert sind. Für steigende Tonhöhen ergibt sich auch ein steigender Frequenzhub. Parameter: $f_0 = 200$ Hz, $f_{\text{mod}} = 5$ Hz, der Modulationshub verhält sich zum Modulationshub ohne Obertöne mit $m_n = m \cdot n/2$.



(a) Spektrogramm **(b)** Cepstrogramm

Abb. 3.4: Darstellung eines Spektrogramms und dessen Cepstrogramm, bei dem ein Sinuston und seine fünf Obertöne ($N = 6$) frequenzmoduliert sind. Diese Obertöne weisen aber zu jedem Zeitpunkt den selben Frequenzabstand zueinander auf, wonach es sich nicht um in der Natur vokommende Obertöne handelt. Ebenso verändert sich für höhere Obertöne der Modulationshub nicht. Parameter: $f_0 = 200$ Hz, $f_{\text{mod}} = 5$ Hz.



(a) Spektrogramm **(b)** Cepstrogramm

Abb. 3.5: Darstellung eines Spektrogramms und dessen Cepstrogramm bei einem gesungenen Baritonvokal. Im Cepstrogramm (b) ist zu sehen, dass die höheren Rahmonischen einen größeren cepstralen Hub besitzen als die erste Rahmonische. Dies ergibt sich dadurch, dass bei der Synthese eines Spektrums durch Quefrenzen auch höhere spektrale Modulationen beteiligt sind als nur jene, die dem Frequenzabstand zwischen den Teiltönen entsprechen.

Mit dem Cepstrogramm ist es also möglich, Modulationen entlang der Frequenz-Achse zu visualisieren. Die Modulationsfrequenz der Frequenzmodulation in den Beispielen (3.2), (3.3), (3.4) und (3.5) wird aus dem Cepstrogramm allerdings nicht ersichtlich. Deshalb liegt es nahe, auch die Fouriertransformation entlang der zeitlichen Achse zu untersuchen, was im nächsten Abschnitt beschrieben wird.

3.2 Zeitliches Modulationsspektrum

Um das zeitliche Modulationsspektrum berechnen zu können, wird ähnlich wie beim Cepstrogramm, zuerst die STFT berechnet. Nach anschließender Betragsbildung und Logarithmierung wird wieder eine Fouriertransformation gebildet. Jetzt allerdings entlang der Zeit-Achse. Diese Vorgehensweise wird in der Literatur [24] verwendet, um den Sprachverständlichkeitsindex STI zu bestimmen. Dabei wird das Spektrogramm in Frequenzbänder der Breite einer Oktave unterteilt und von jedem Frequenzband die Hüllkurve berechnet. Anhand dieser Hüllkurven wird anschließend der STI bestimmt.

$$\mathbf{MODSPEC}\{x(t)\}(\mathrm{f_{tmod}}, \omega) = X_m(\mathrm{f_{tmod}}, \omega) = \left|\mathcal{F}_{1,0}\Big[\log|X(t', \mathrm{j}\omega)|\Big]\right| \tag{3.2.1}$$

Die Vorgehensweise, um das zeitliche Modulationsspektrum zu erhalten, ist ganz ähnlich, nur, dass die Bandbreite eines Frequenzbandes von der Fensterlänge der STFT abhängig ist. Die Bandbreite ergibt sich über Gl. 2.2.3. Es wird nun von jedem dieser Frequenzbänder eine Fouriertransformation entlang der Zeit-Achse berechnet, um die spektrale Zusammensetzung in diesem Frequenzband zu erhalten.

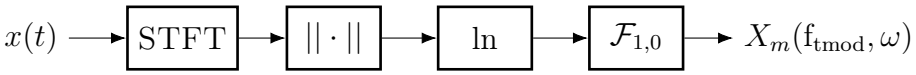
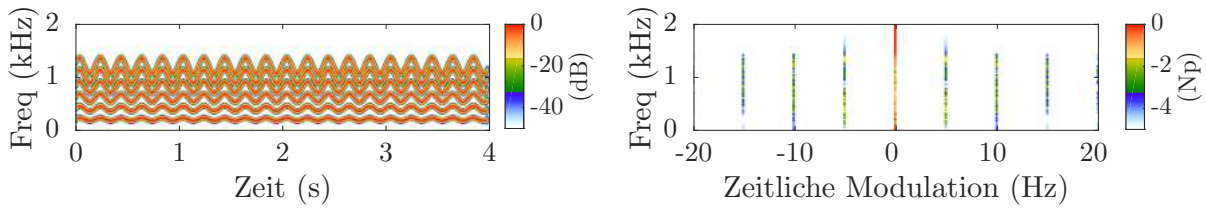


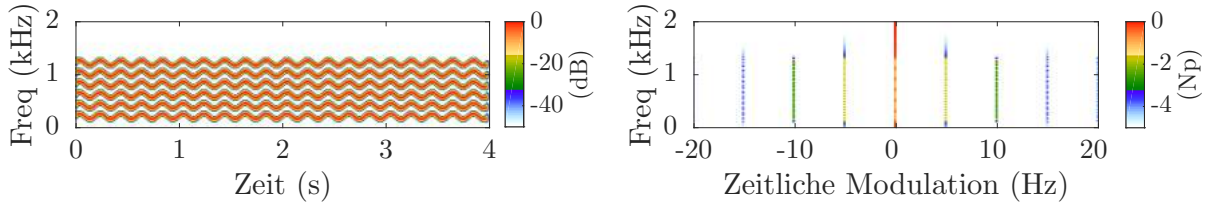
Abb. 3.6: Berechnung des zeitlichen Modulationsspektrums.

Besitzt ein Signal eine periodische Hüllkurvenfluktuation in einem Frequenzband, ist dies im zeitlichen Modulationsspektrum sichtbar. Für die frequenzmodulierten Sinustöne ergibt sich ein zeitliches Modulationsspektrum wie in Abbildung 3.7. Die Modulationsfrequenz von 5 Hz wird damit als erstes lokales Maximum sichtbar. Die vertikalen Linien werden nur in dem Frequenzbereich erzeugt, in dem auch eine Modulation stattgefunden hat. Diese Modulationsfrequenz im Frequenzbereich ändert sich auch dann nicht, wenn der Modulationshub nicht proportional zur Frequenz des Obertones im Spektrogramm verläuft (Abb. 3.8). Auch dann ist im zeitlichen Modulationsspektrum eine vertikale Linie genau im selben Frequenzbereich zu erkennen, allerdings mit leichten Magnitudenschwankungen im Gegensatz zu dem Beispiel mit proportionalem Modulationshub.



(a) Spektrogramm **(b)** Zeitliches Modulationsspektrum

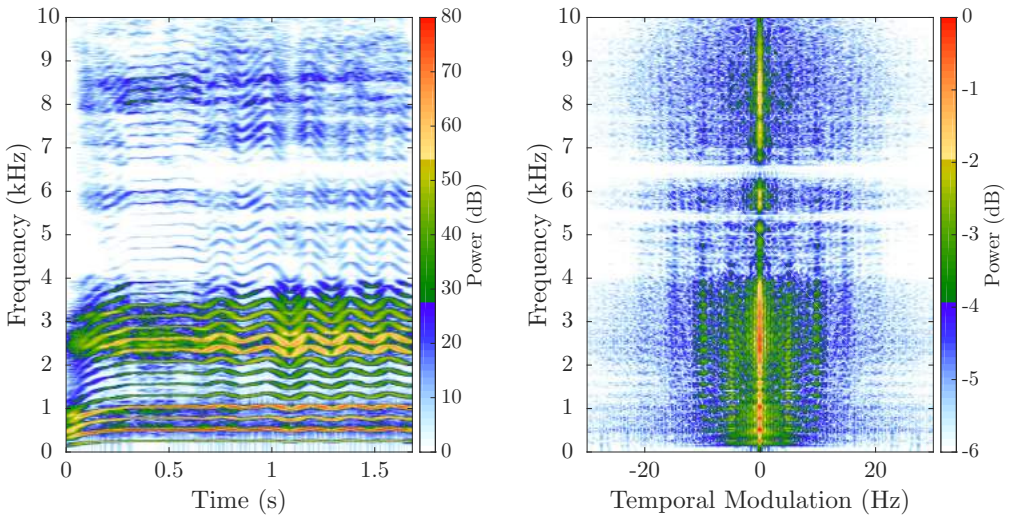
Abb. 3.7: Darstellung eines Spektrogramms und dessen zeitlichen Modulationsspektrum, bei dem ein Sinuston und seine fünf Obertöne ($N = 6$) frequenzmoduliert sind. Für steigende Tonhöhen ergibt auch ein steigender Frequenzhub. Parameter: $f_0 = 200$ Hz, $f_{\text{mod}} = 5$ Hz, der Modulationshub verhält sich zum Modulationshub ohne Obertöne mit $m_n = m \cdot n/2$.



(a) Spektrogramm **(b)** Zeitliches Modulationsspektrum

Abb. 3.8: Darstellung eines Spektrogramms und dessen zeitliches Modulationsspektrum, bei dem ein Sinuston und seine fünf Obertöne ($N = 6$) frequenzmoduliert sind. Für steigende Tonhöhen verändert sich der Modulationhub nicht. Parameter: $f_0 = 200$ Hz, $f_{\text{mod}} = 5$ Hz.

Um das zeitliche Modulationsspektrum anhand eines konkreten musikalischen Beispiel veranschaulichen zu können, wird Abbildung 3.9 betrachtet. Es ist zu beobachten, dass zeitliche Modulationen eine Tiefpasscharakteristik [25] aufweisen. Dies resultiert einerseits am Tiefpassverhalten der Fensterung bei der STFT. Wird das Fenster zu lang gewählt, wird über einen zu großen Bereich von zeitlichen Modulationen gemittelt und somit ist es nicht mehr möglich, höhere zeitliche Modulationen zu bestimmen. Längere Fenster ergeben dadurch eine geringere maximale zeitliche Modulationsfrequenz. Andererseits tritt bei der Klangerzeugung durch die menschliche Stimme und durch Instrumente ebenso ein Tiefpassverhalten auf. Dies resultiert aus der Trägheit des menschlichen Stimmapparats und der mechanischen Klangerzeugung hin zu höheren Modulationsfrequenzen. Wie im Cepstrogramm sind auch im zeitlichen Modulationsspektrum Vielfache der Modulationsfrequenz ausfindig zu machen. Das ist dadurch zu erklären, dass eine periodische Hüllkurve nicht genau der einer Sinusschwingung entspricht. Wie im Spektrogramm schon ersichtlich, kommt es im gesamten Frequenzbereich zu Fluktuationen der zeitlichen Amplitude. In diesem Beispiel ist bei ca. 5 Hz ein erstes lokales Maximum genau dort zu finden, wo auch im Spektrogramm eine annähernd sinusförmige Bewegung der Teiltöne zu finden ist.



(a) Spektrogramm **(b)** zeitliches Modulationsspektrum

Abb. 3.9: Darstellung eines Spektrogramms und dessen zeitlichen Modulationsspektrum bei einem gesungenen Baritonvokal.

3.3 Modulation Power Spectrum

Um sowohl zeitliche als auch spektrale Modulationen sichtbar machen zu können, wurde das Modulation Power Spectrum definiert [25]. Wie in Abb. 3.10 zu sehen ist, wird dazu nach der Betragsbildung der STFT die Magnitude logarithmiert und anschließend über eine zweidimensionale Fouriertransformation in eine zweidimensionale Darstellung übergeführt.

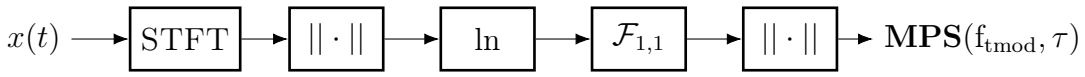


Abb. 3.10: Berechnung des Modulation Power Spectrum.

Dabei bezeichnet die Abszisse die Achse der positiven und negativen zeitlichen Modulationen. Auf der Ordinate werden die positiven spektralen Modulationen aufgetragen. Die zweidimensionale Fouriertransformation liefert als Ergebnis eine Matrix mit vier Quadranten. Durch die MATLAB-Funktion `fftshift` werden im 1D-Fall die positiven und negativen Frequenzen so angeordnet, dass sich der Gleichanteil in der Mitte befindet. Im 2D-Fall, wie er hier benötigt wird, werden die Quadranten diagonal getauscht. Damit befindet sich der Gleichanteil in der Mitte des Bildes und die positiven Frequenzen für steigende Werte der zeitlichen Modulationen werden in der rechten Hälfte angeordnet. Negative zeitliche Modulationsfrequenzen werden demnach in der linken Halbebene abgebildet. Aufgrund der Punktsymmetrie der zweidimensionalen Fouriertransformation ergibt sich eine Redundanz in der Darstellung. Demnach enthält der erste und dritte Quadrant die selbe Information des Betragsspektrums. Somit reduziert sich die Darstellungsweise des MPS im weiteren Verlauf auf die Quadranten 1 und 2. Der Unterschied zwischen diesen beiden Quadranten wird durch die zeitlichen Modulationen beschrieben. Findet eine Bewegung der Teiltöne im Spektrogramm nach oben statt, was einem Glissando nach oben entspricht, wird dies im MPS in der linken Hälfte abgebildet. Eine Bewegung nach unten wird in der rechten Hälfte dargestellt.

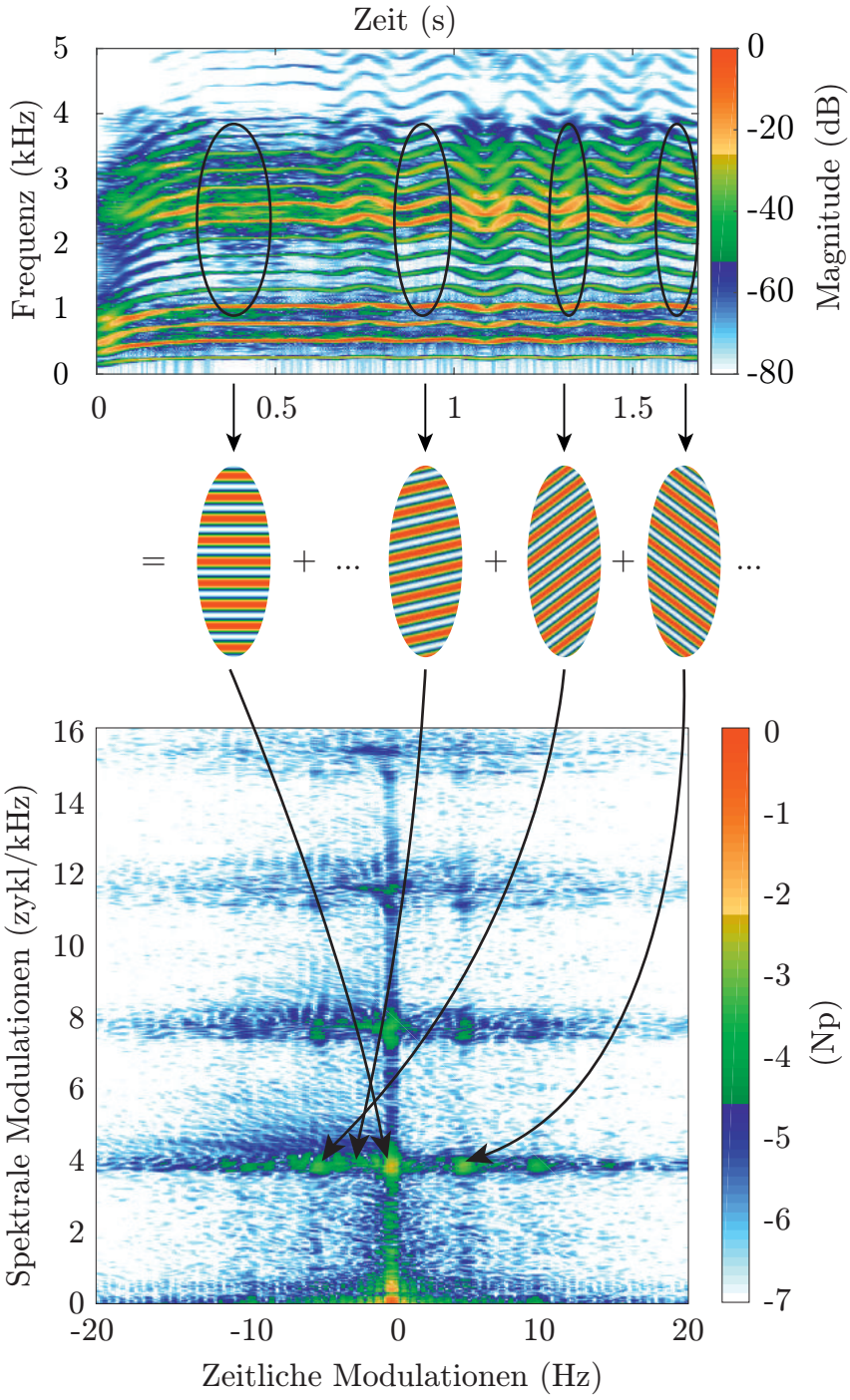


Abb. 3.11: Darstellung eines Spektrogramms eines gesungenen Baritonvokals und dessen MPS. Bereiche im Spektrogramm mit konstanter Frequenz und Obertonstruktur werden im MPS entlang der τ-Achse abgebildet. Glissandi nach oben (Sekunde 0.9) werden in der linken Hälfte abgebildet, wohingegen sich Glissandi nach unten in der rechten Hälfte befinden. Der Formantbereich des Klanges wird um den Bereich des Ursprungs abgebildet. Die zeitlichen Modulationen des Formantbereichs bilden sich bei niedrigen spektralen Modulationen entlang der zeitlichen Modulationsachse ab. Die zeitlichen Modulationen der Teiltonstruktur werden auf der jeweiligen Höhe der spektralen Modulationen ebenfalls entlang der zeitlichen Modulationsachse abgebildet.

Mit Abbildung 3.11 wird verdeutlicht, wie ein MPS aus einem Spektrogramm gebildet wird. Es ist zu sehen, dass bei Sekunde 0.4 eine Obertonstruktur zu finden ist, bei der sich die Frequenzen der Teiltöne nicht verändern. Ein solches Muster wird im MPS entlang der spektralen Modulationsachse (τ-Achse) dargestellt. Befinden sich mehr Obertöne im Bereich von einem kHz, entspricht dies einer Grundtonverschiebung nach unten und somit einer Bewegung auf der τ-Achse nach oben. Wird dieser Anteil der Welle entlang der τ-Achse nun festgehalten und kommt auch ein Anteil entlang der f_{tmod}-Achse hinzu, wird diese Welle links oder rechts von der τ-Achse abgebildet, aber auf der selben Höhe, die dem Anteil in τ-Richtung entspricht. Die genaue Position in der linken oder rechten Hälfte ergibt sich anhand der Steilheit der Bewegung nach oben oder unten. Weiters ist zu beobachten, dass auch noch spektrale bzw. zeitliche Modulationen mit Vielfachen der ersten spektralen und zeitlichen Modulation an der Zusammensetzung des Spektrums beteiligt sind.
Bei höheren spektralen Modulationen sind geringe zeitliche Modulationen zu finden und es ist auch die höchste zeitliche Modulationsenergie bei niedrigeren spektralen Modulationen zu erkennen. Der größte Energieanteil an Modulationen ist im Bereich des Ursprungs des MPS zu erkennen. In diesem Bereich befindet sich die Energie der Formanten des Klanges. Unter einem zykl/kHz befinden sich Formanten, die mindestens eine Breite von einem kHz besitzen und sich somit über größere Bereiche im Spektrum bewegen und das Spektrum im groben formen. In Kapitel 4 wird untersucht, welche klanglichen Auswirkungen sich bei einer Unterdrückung der Formanten ergeben.
Anhand von einfachen und komplexeren Klängen soll im letzten Abschnitt dieses Kapitels das MPS dieser Klänge erläutert werden. Als mathematische Herangehensweise wird anhand eines Gauß-Chirps eine geschlossene Lösung präsentiert.

3.4 Geschlossene Lösung anhand eines Gauß-Chirp

Ein Gauß-Chirp mit einer Mittenfrequenz gleich null kann geschrieben werden als:

$$g(t) = \exp\Big[-\alpha t^2\Big] \exp\Big[\mathrm{j}\pi S_0 t^2\Big] = \exp\Big[-(a-\mathrm{j}b)t^2\Big], \tag{3.4.1}$$

wobei S_0 die Sweeprate in Hz/s angibt und α zur Formung der Gaußschen Hüllkurve beiträgt. Die Parameter a and b werden somit definiert mit $a \equiv \alpha$ und $b \equiv \pi S_0$.

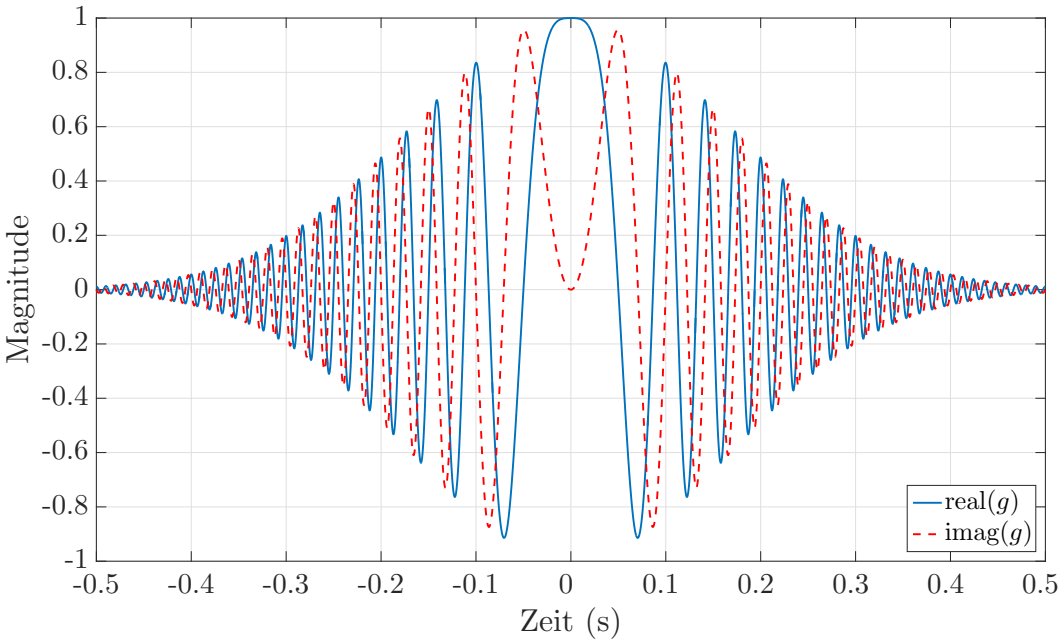


Abb. 3.12: Beispiel eines komplexen Gauß-Chirps. Parameter für diese Darstellung: $S_0 = 200$ Hz/s, $\alpha = 3$

Spektrale Darstellung

Die Fourier-Transformation eines Gaußsignals ergibt im Spektrum wieder die Form einer Gaußkurve:

$$\mathrm{e}^{-ct^2} \quad \circ\!\!-\!\!\bullet \quad \sqrt{\frac{\pi}{c}}\mathrm{e}^{-\frac{\omega^2}{c}}. \tag{3.4.2}$$

Das Signal aus 3.4.1 kann über $c = (a - \mathrm{j}b) = r \cdot \mathrm{e}^{\mathrm{j}\varphi}$ mit $r = \sqrt{a^2 + b^2}$ und $\varphi = -\arctan(b/a)$ wie folgt angeschrieben werden:

$$\mathrm{e}^{-(r\cdot\mathrm{e}^{\mathrm{j}\varphi})t^2} \quad \circ\!\!-\!\!\bullet \quad \sqrt{\frac{\pi}{r}} \cdot \mathrm{e}^{-\mathrm{j}\frac{\varphi}{2}} \cdot \exp\left[-\frac{\omega^2}{r} \cdot \mathrm{e}^{-\mathrm{j}\varphi}\right]. \tag{3.4.3}$$

wobei $\omega = 2\pi f$ der Kreisfrequenz im Spektrum entspricht.
Das komplexe Spektrum ergibt sich nach einer Aufspaltung der Eulerformel in Real- und Imaginärteil:

$$\begin{aligned} G_{\mathrm{comp}}(\mathrm{j}\omega) &= \sqrt{\frac{\pi}{r}} \cdot \mathrm{e}^{-\mathrm{j}\frac{\varphi}{2}} \cdot \exp\left[-\frac{\omega^2}{r} \cdot (\cos(\varphi) - \mathrm{j}\sin(\varphi))\right] \\ &= \sqrt{\frac{\pi}{r}} \cdot \mathrm{e}^{-\mathrm{j}\frac{\varphi}{2}} \cdot \exp\left[-\frac{\omega^2}{r} \cdot \cos(\varphi) + \mathrm{j}\frac{\omega^2}{r} \cdot \sin(\varphi)\right] \\ &= \sqrt{\frac{\pi}{r}} \cdot \mathrm{e}^{-\mathrm{j}\frac{\varphi}{2}} \cdot \exp\left[-\frac{\omega^2}{r} \cdot \cos(\varphi)\right] \cdot \exp\left[+\mathrm{j}\frac{\omega^2}{r} \cdot \sin(\varphi)\right] \end{aligned} \tag{3.4.4}$$

Bei der Betragsbildung fallen alle Terme mit $\mathrm{e}^{\mathrm{j}\cdot}$ weg:

$$\begin{aligned} |G_{\mathrm{comp}}(\mathrm{j}\omega)| &= \left|\sqrt{\frac{\pi}{r}}\right| \cdot \underbrace{\left|\mathrm{e}^{-\mathrm{j}\frac{\varphi}{2}}\right|}_{=1} \cdot \left|\exp\left[-\frac{\omega^2}{r} \cdot \cos(\varphi)\right]\right| \cdot \underbrace{\left|\exp\left[+\mathrm{j}\frac{\omega^2}{r} \cdot \sin(\varphi)\right]\right|}_{=1} \\ &= \sqrt{\frac{\pi}{r}} \cdot \exp\left[-\frac{\omega^2}{r} \cdot \cos(\varphi)\right] \end{aligned} \tag{3.4.5}$$

Einsetzen von r und φ ergibt den fertigen Betragsfrequenzgang:

$$
\begin{aligned}
|G_{\text{comp}}(\omega)| &= \sqrt{\frac{\pi}{\sqrt{a^2+b^2}}} \cdot \exp\left[-\frac{\omega^2}{\sqrt{a^2+b^2}} \cdot \cos(-\arctan(b/a))\right] \\
&= \sqrt{\frac{\pi}{\sqrt{a^2+b^2}}} \cdot \exp\left[-\frac{\omega^2}{\sqrt{a^2+b^2}} \cdot \frac{a}{\sqrt{a^2+b^2}}\right]
\end{aligned} \tag{3.4.6}
$$

Die Logarithmierung des Betragsfrequenzganges ergibt schließlich das Spektrum in logarithmierter Darstellung:

$$
\begin{aligned}
\log|G_{\text{comp}}(\omega)| &= \log\left(\sqrt{\frac{\pi}{\sqrt{a^2+b^2}}} \cdot \exp\left[-\omega^2 \cdot \frac{a}{a^2+b^2}\right]\right) \\
G_{\text{spec}}(\omega) &= \log\left(\sqrt{\frac{\pi}{\sqrt{a^2+b^2}}}\right) \cdot \left[-\omega^2 \cdot \frac{a}{a^2+b^2}\right]
\end{aligned} \tag{3.4.7}
$$

Darstellung eines Gaußschen Chirpsignals in der MPS-Domäne

Als Gauß-Chirps werden kurze gaußgefensterte Sweeps bezeichnet. Um N Sinussignale mit deren Grundton darstellen zu können, wird folgende Form verwendet:

$$
g_h(t) = \sum_{n=1}^{N} g_n(t) = \sum_{n=1}^{N} e^{\mathrm{j}2\pi n(f_m + \frac{S_0}{2}t)t} \tag{3.4.8}
$$

Wobei n den jeweiligen Teilton beschreibt, f_m für die Mittenfrequenz des Grundtons steht und S_0 die Sweeprate des Grundtons in Hz/s angibt.
Die STFT dieser Summe ergibt eine Matrix:

$$
\mathbf{STFT}\{g_h(t)\}(t', \mathrm{j}\omega) = G_h(t', \mathrm{j}\omega) = \int_{-\infty}^{+\infty} g_h(t) w(t-t') e^{-\mathrm{j}\omega t} \mathrm{d}t \tag{3.4.9}
$$

Wobei $w(t-t')$ ein Gaußfenster bezeichnet. Zusammen mit dem Verschiebungssatz der Fourier-Transformation, 3.4.4 und A.1.1 ergibt sich:

$$
G_h(t', \mathrm{j}\omega) = \sum_{n=1}^{N} e^{\mathrm{j}\pi n(2f_m + S_0 t')t' - \mathrm{j}\omega t'} \; G^n_{\text{comp}}(\mathrm{j}\omega - 2\pi(f_m n + S_0 n t')) \tag{3.4.10}
$$

Das Spektrogramm der STFT-Matrix wird über die logarithmierte Magnitude berechnet:

$$\textbf{SPECTROGRAM}\{g_h(t)\}(t',\omega) = G_{\mathrm{s}}(t',\omega) = \log|G_h(t',\mathrm{j}\omega)| \tag{3.4.11}$$

Wird 3.4.10 in 3.4.11 eingesetzt, ergibt sich:

$$G_{\mathrm{s}}(t',\omega) = \log\Big|\sum_{n=1}^{N} e^{\mathrm{j}\pi n(2f_m+S_0t')t'-\mathrm{j}\omega t'}\; G^n_{\mathrm{comp}}(\mathrm{j}\omega - 2\pi(f_m n + S_0 n t'))\Big| \tag{3.4.12}$$

Unter der Annahme, dass individuelle Teile von G^n_{comp} im Spektrum nicht überlappen, kann 3.4.12 umgeformt werden:

$$G_{\mathrm{s}}(t',\omega) = \log\sum_{n=1}^{N} \Big|e^{\mathrm{j}\pi n(2f_m+S_0t')t'-\mathrm{j}\omega t'}\; G^n_{\mathrm{comp}}(\mathrm{j}\omega - 2\pi(f_m n + S_0 n t'))\Big| \tag{3.4.13}$$

Um eine Berechnung von $\log(0)$ zu verhindern, wird ein Rauschteppich ε addiert:

$$G_{\mathrm{s}}(t',\omega) \approx \log\Big[\sum_{n=1}^{N} \Big|G^n_{\mathrm{comp}}(\mathrm{j}\omega - 2\pi(f_m n + S_0 n t'))\Big| + \varepsilon\Big] \tag{3.4.14}$$

Mit $\log\sum_{n=1}^{N} \mathrm{e}^{x_n} = \max\{x_n\}$, wo $x_n = \ln|G^n_{\mathrm{comp}}|$ ist und wieder unter der Annahme, dass die Spektren der einzelnen Sweeps nicht überlappen, kann 3.4.14 wie folgt berechnet werden:

$$\begin{aligned} G_{\mathrm{s}}(t',\omega) &\approx \max\{G^1_{\mathrm{spec}}, G^2_{\mathrm{spec}}, G^3_{\mathrm{spec}}, \ldots G^N_{\mathrm{spec}}, \ln\varepsilon\} \\ &\approx \sum_{i=1}^{N} \max\{G^n_{\mathrm{spec}}, \ln\varepsilon\} \end{aligned} \tag{3.4.15}$$

Für einen Sweep mit 5 Obertönen ergibt sich für ein Frame Abb. 3.13 und als Spektrogramm Abb. 3.14.

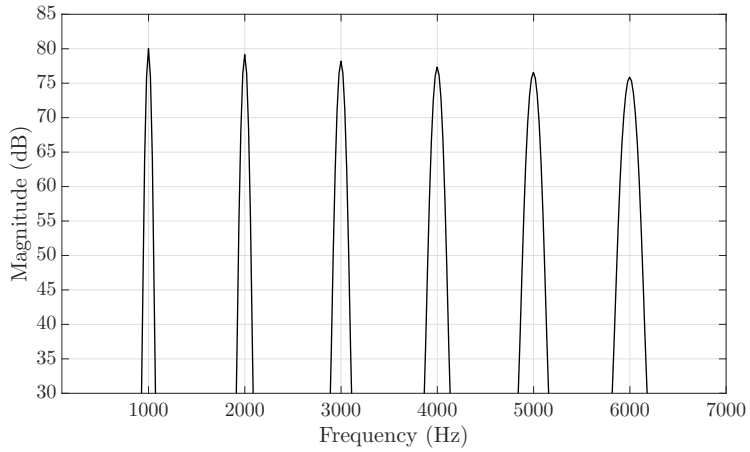


Abb. 3.13: Spektrum eines Frames für ein Sweepsignal mit 5 Obertönen. Es ist zu beobachten, dass die Obertöne ganzzahlige Vielfache des Grundtons sind. Obertöne weisen aufgrund der höheren Sweeprate auch ein breiteres Spektrum auf, da in der selben Zeit ein größerer Frequenzbereich durchschritten wird.

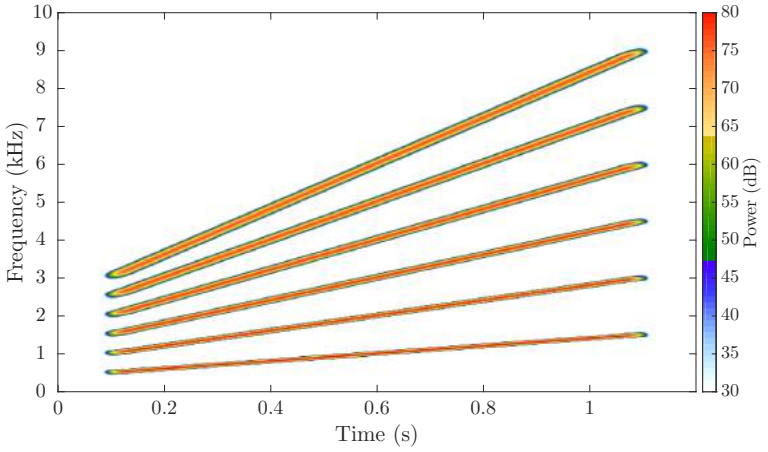


Abb. 3.14: Spektrogram eines Sweepsignals mit 5 Obertönen.

Zur Vereinfachung wird zuerst ein Sweep mit nur einem Teilton betrachtet. Im Spektrogramm ergibt dies eine Linie mit einer Steigung der Sweeprate. Eine Fourier-Transformation entlang der Frequenz-Achse für jedes Frame ergibt das Cepstrogramm. Da ein Sinuston im Spektrum keine Periodizität aufweist, ergibt auch das Ceptrum nur einen Gleichanteil. Die Fourier-Transformation von Gl. 3.4.7 ergibt[1]

$$G_p(\tau) = -c\frac{1}{2\pi}\Big(-{\omega_b}^2\,\frac{\sin\left(\frac{\omega_b}{2}\tau\right)}{2\tau} - 2\omega_b\,\frac{\cos\left(\frac{\omega_b}{2}\tau\right)}{\tau^2} + 4\frac{\sin\left(\frac{\omega_b}{2}\tau\right)}{\tau^3}\Big) \tag{3.4.16}$$

[1] Für die Herleitung siehe Anhang A.2.

das in Abb. 3.15 zu sehen ist.

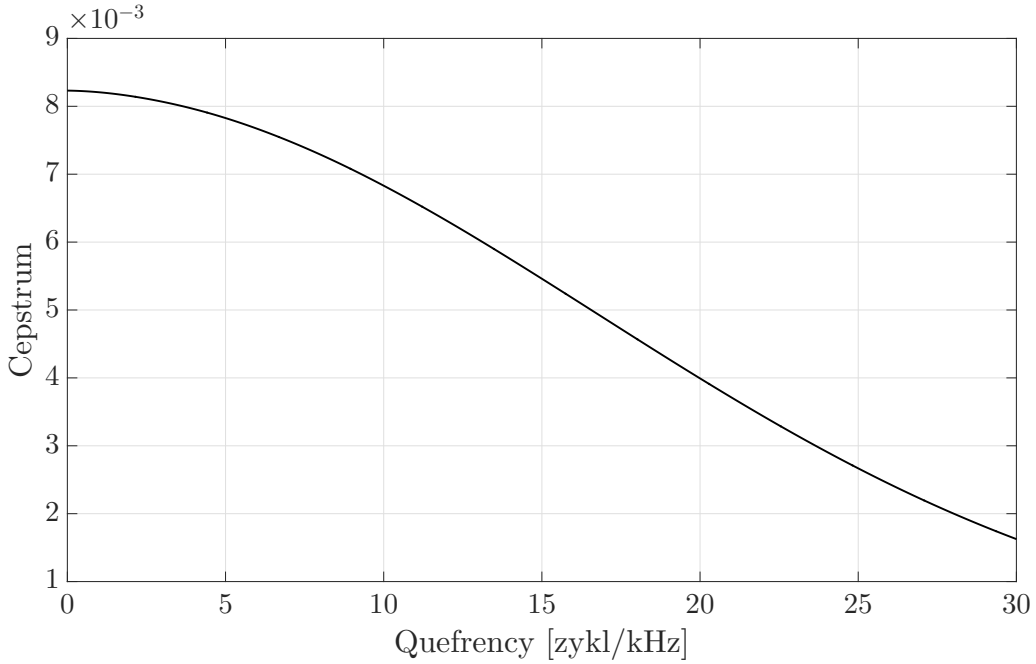


Abb. 3.15: Betrag des Cepstrums eines Sweepsignals mit einem Teilton in einem Frame. Da das Spektrum nur einen Teilton enthält, ergibt sich im Cepstrum nur ein Gleichanteil. Dieser Gleichanteil bleibt für jedes Frame gleich. Einzig in der Phase ist der Unterschied zwischen den Frames zu erkennen.

Eine Erhöhung der Sweeprate erzeugt eine breitere Parabel im Spektrum und somit ein schmäleres Cepstrum. Für verschiedene Bandbreiten ergeben sich unterschiedliche Breiten der Kurve im Cepstrum (Abb. 3.16).

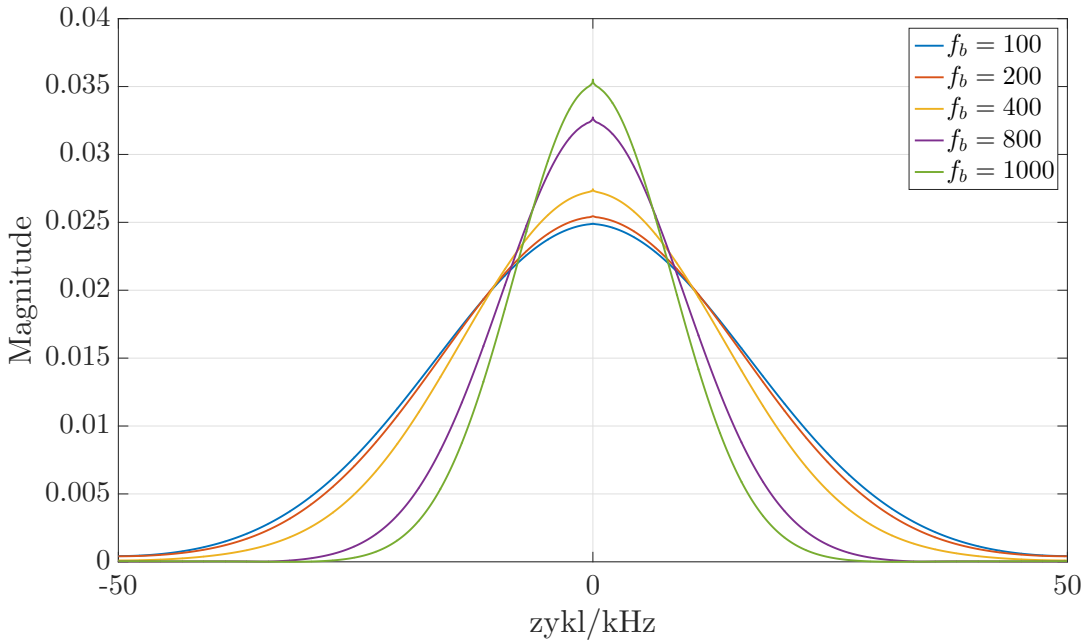


Abb. 3.16: Cepstrum eines einzelnen Chirps mit verschiedenen Sweepraten. Es ist zu beobachten, dass, falls die Sweeprate S_0 steigt, sich die Breite des Cepstrums verkleinert.

Fourier-Transformation entlang der Zeit-Achse

Um im letzten Schritt das MPS berechnen zu können, fehlt noch die Fourier-Transformation entlang der Zeit-Achse. Um dies verstehen zu können, wirde eine Quefrenz des Cepstrogramms betrachtet (Abb. 3.17). Dabei ist zu sehen, dass es sich um ein annähernd rechteckförmiges Gebilde handelt.

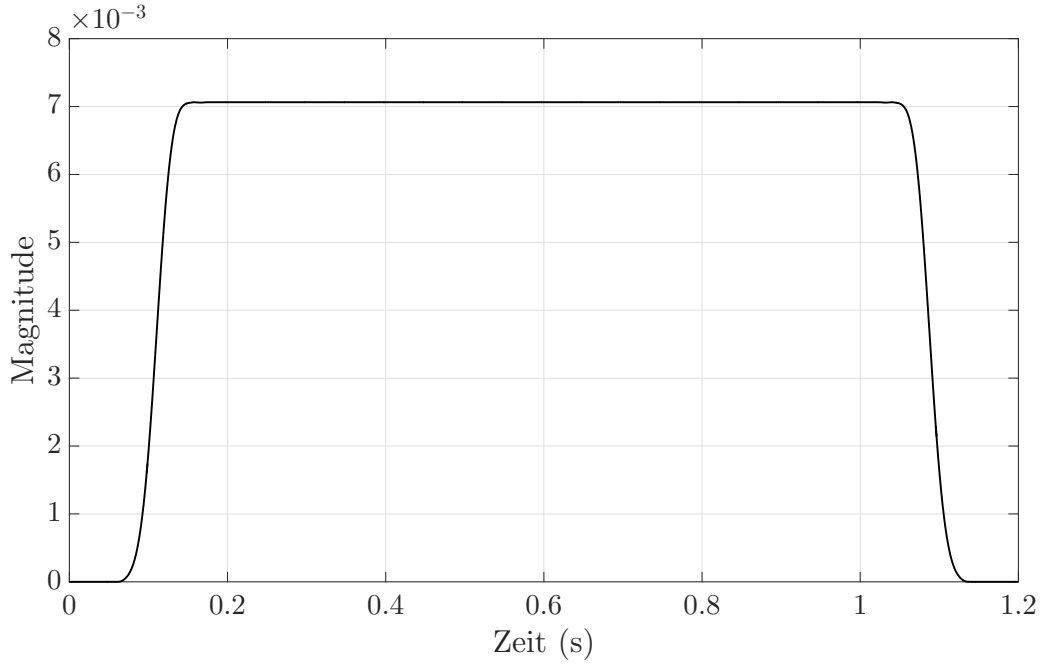


Abb. 3.17: Horizontale Linie des Cepstrogramms, welche einer Quefrenz bei ca. 9 zykl/kHz entspricht.

Unter der Berücksichtigung der Phase des Cepstrogramms wird ersichtlich, dass diese mit einer höheren negativen Steigung für höhere Mittenfrequenzen und höhere Sweepraten fortschreitet:

$$G_c(t', \tau) = \mathrm{e}^{-\mathrm{j}2\pi\left(f_m + S_0 t'\right)\tau} G_p(\tau); \tag{3.4.17}$$

wobei t' den Zeitpunkt des Frames im Spektrogramm angibt und τ für die Variable der Quefrenzen entlang der vertikalen Achse steht. Mit der Betragsbildung des Cepstrogramms ist die Phase nicht mehr sichtbar. Sie gibt aber an, ob ein Sweep nach oben oder unten verläuft. Falls S_0 positiv ist, fällt die Phase noch stärker als linear zu negativen Werten ab. Ein negativer Wert von S_0 ergibt aber eine Bewegung der Phase zu positiven Werten (Abb. 3.18).

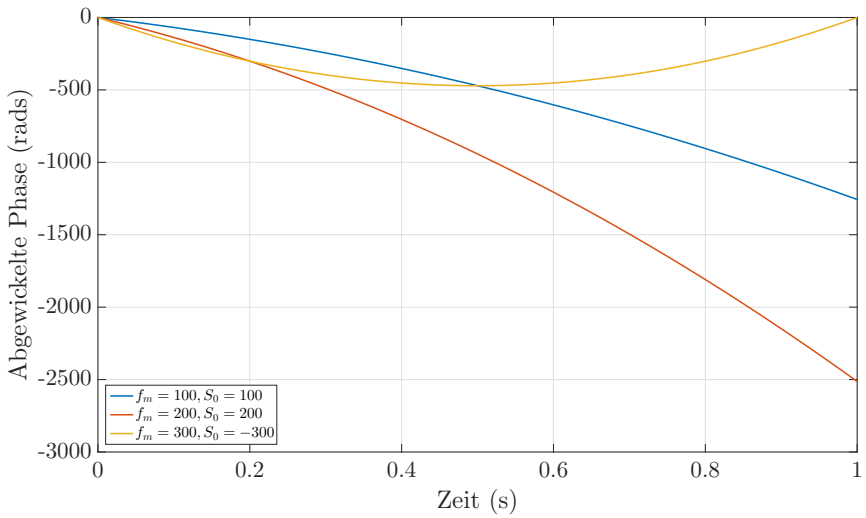


Abb. 3.18: Abgewickelte Phase in dem Phasenterm in 3.4.17. Ein positiver Wert von S_0 resultiert in einer schneller abfallenden Phase hin zu negativen Werten. Ein negativer Wert von S_0 erzeugt eine Phasenbewegung hin zu positiven Bewegungen.

Eine Fourier-Transformation entlang der Zeit-Achse berücksichtigt auch diese Phase. Die Phase wird nur aus Darstellungsgründen nicht sichtbar. Zusammen mit dem Phasenterm und dem Betrag des Cepstrogramms ergibt sich das MPS:

$$\begin{aligned} G_{\mathrm{MPS}}(\mathrm{f_{tmod}}, \tau) &= \left| \mathcal{F}_{1,0}\Big[G_c(t', \tau) \Big] \right| \\ &= \int_{-\infty}^{+\infty} \mathrm{e}^{-\mathrm{j}2\pi \left(f_m + S_0 t' \right) \tau} \, G_c(t', \tau) \, e^{-\mathrm{j}\omega t} \mathrm{d}t' \end{aligned} \tag{3.4.18}$$

Wird $G_c(t', \tau)$ als das zeitliche Frame betrachtet und dies als eine annähernd rechteckige Funktion gesehen, ergibt sich über die Phase eine Verschiebung der Sincfunktion hin zu positven (negatives S_0) oder negativen (positives S_0) zeitlichen Modulationen. Dies ist in Abbildung 3.19 zu sehen.

$$G_{\mathrm{MPS}}(\mathrm{f_{tmod}}, \tau) = G_{\mathrm{sinc}}(\mathrm{f_{tmod}} - \left(f_m + S_0 t' \right) \tau, \tau) \tag{3.4.19}$$

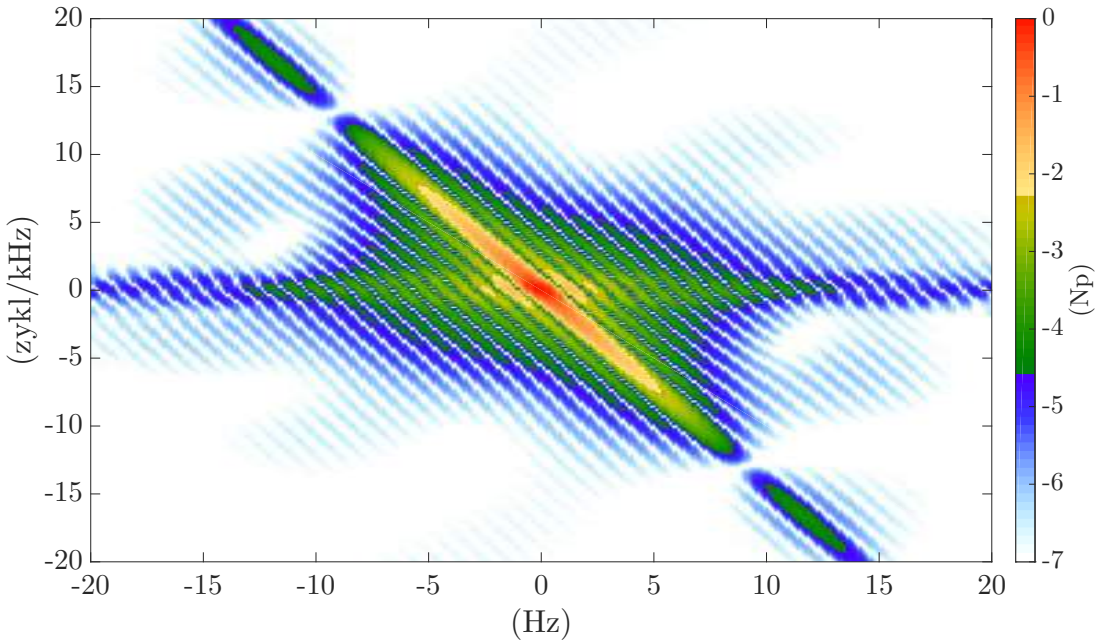


Abb. 3.19: Modulation Power Spectrum eines Chirps mit einem Teilton. Parameter: $f_m = 2000$ Hz, $S_0 = 700$ Hz/s.

Berechnung des MPS eines Sweeps mit mehreren Teiltönen

Aufgrund der Linearität der Fourier-Transformation können verschiedene spektrale Anteile in einem Spektrum als getrennte Bereiche angesehen und getrennt berechnet werden.

$$\mathcal{F}[c_1 x_1(t) + c_2 x_2(t)] = c_1 \mathcal{F}[x_1(t)] + c_2 \mathcal{F}[x_2(t)]. \tag{3.4.20}$$

Für jeden Teil im Spektrum ergibt sich ein unterschiedliches Cepstrum (breiter für schmälere Bereiche im Spektrum und vice versa) mit einer unterschiedlichen Phase. Eine Überlagerung dieser Anteile resultiert im Cepstrum nicht nur in einem Gleichanteil, sondern in lokalen Maxima bei jenen spektralen Modulationen, die dem ganzzahligen Vielfachen des Frequenzabstandes im Spektrum entsprechen. Ein Sweep mit drei Teiltönen ergibt im Cepstrogramm folgende Darstellung:

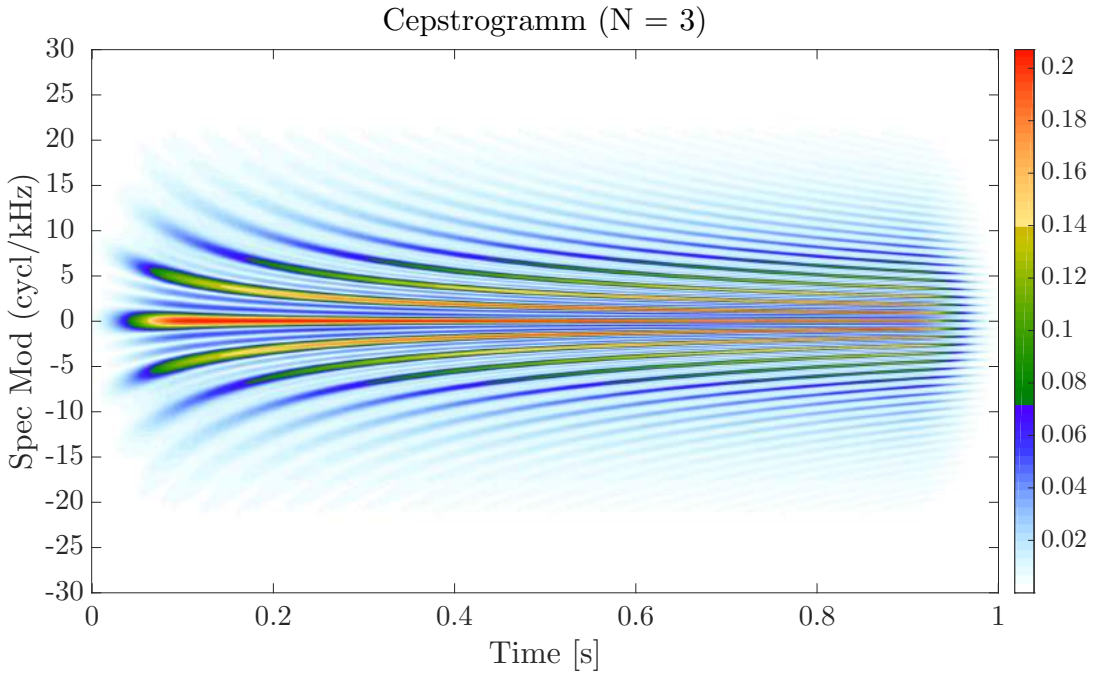


Abb. 3.20: Cepstrogramm eines Sweeps mit drei Teiltönen. Die breiteren Verläufe ergeben sich aus der geringen Anzahl an spektralen Peaks.

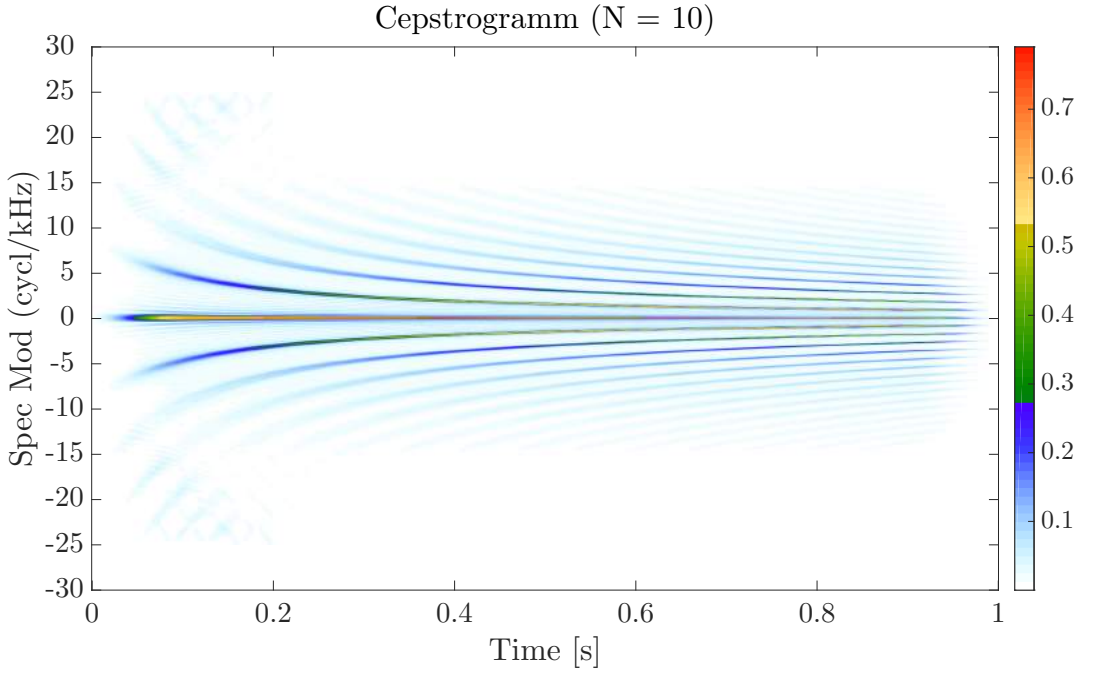


Abb. 3.21: Cepstrogramm eines Sweeps mit zehn Teiltönen.

Eine Fourier-Transformation von Sweeps mit mehreren Teiltönen entlang der Zeit-Achse

mündet im MPS. Die verschiedenen Sweepraten werden im MPS mit der Steilheit der Linien abgebildet (Abb. 3.22).

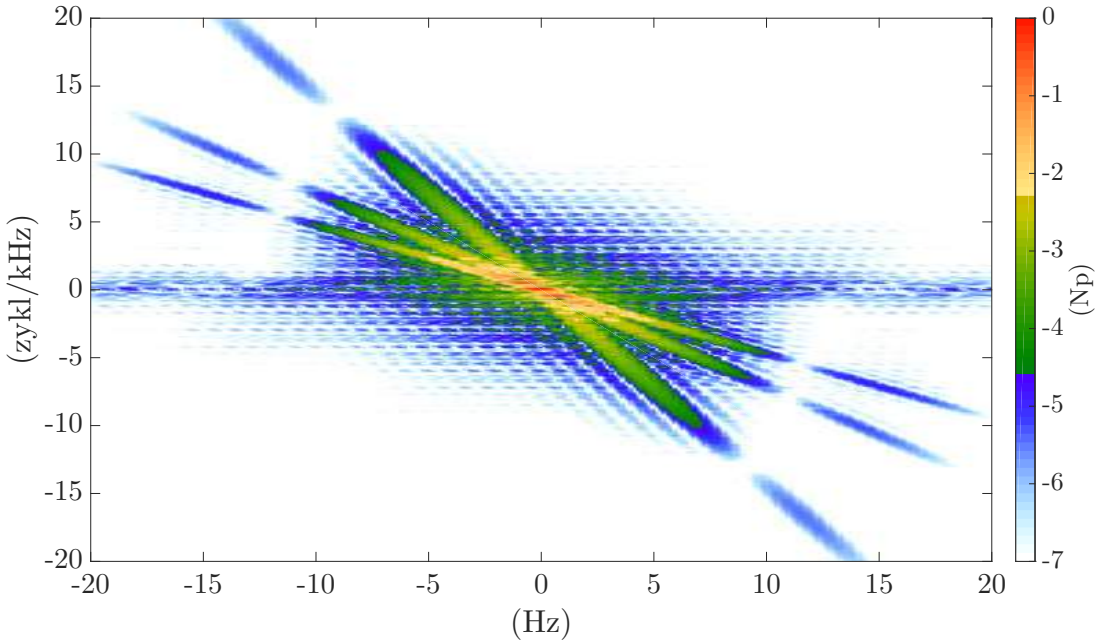


Abb. 3.22: Modulation Power Spectrum eines Sweepsignals mit 3 Teiltönen. Parameter: $f_m = 2000$ Hz, $S_0 = 700$ Hz/s.

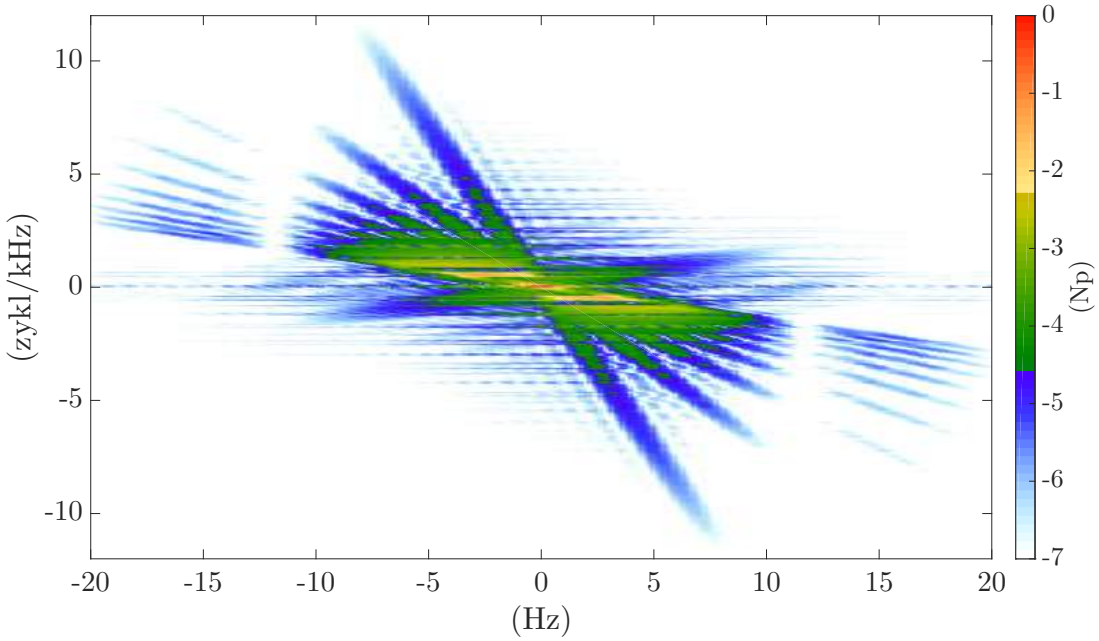


Abb. 3.23: Modulation Power Spectrum eines Sweepsignals mit 10 Teiltönen. Parameter: $f_m = 2000$ Hz, $S_0 = 700$ Hz/s.

3.5 Beispiele

Als Abschluss sollen Klangaufnahmen anhand der oben präsentierten Darstellungsweisen analysiert werden. Dazu wurden kurze Phrasen oder Klänge zur Analyse herangezogen, in denen einerseits eine Modulation zu hören ist, andererseits dies auch in den Darstellungsweisen sichtbar wird. Alle Beispiele wurden mit einem Hann-Fenster und einer Fensterlänge von 30 ms berechnet. Als Hopsize wurde eine vierfache Überlappung gewählt, welche 330 Sampeln entspricht. Um eine bessere Auflösung der Modulationen entlang der beiden Achsen zu erhalten, wurde ein Oversampling-Faktor von 10 verwendet.

3.5.1 Haydn Streichquartett in C-Dur op. 76, Nr. 3

Als erste kurze Phrase werden die ersten 5 Zusammenklänge zu Beginn des Stücks analysiert. Der Auftakt geht über in eine Kadenz (C-d-G-C). Die Töne des C-Dur Akkords zu Beginn des ersten Taktes sind im Cepstrogramm zu erkennen (Abb. 3.25 (c)).

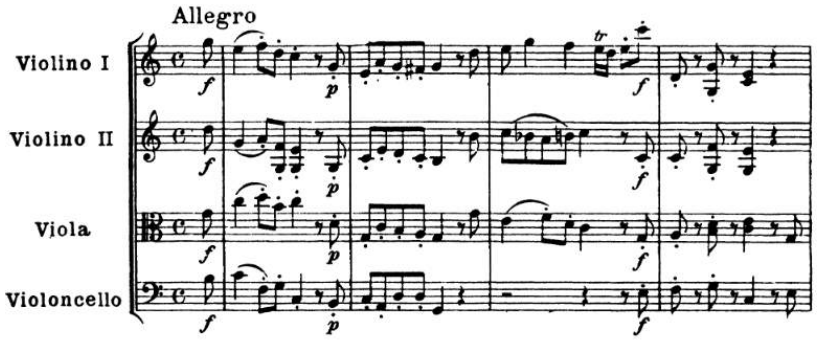
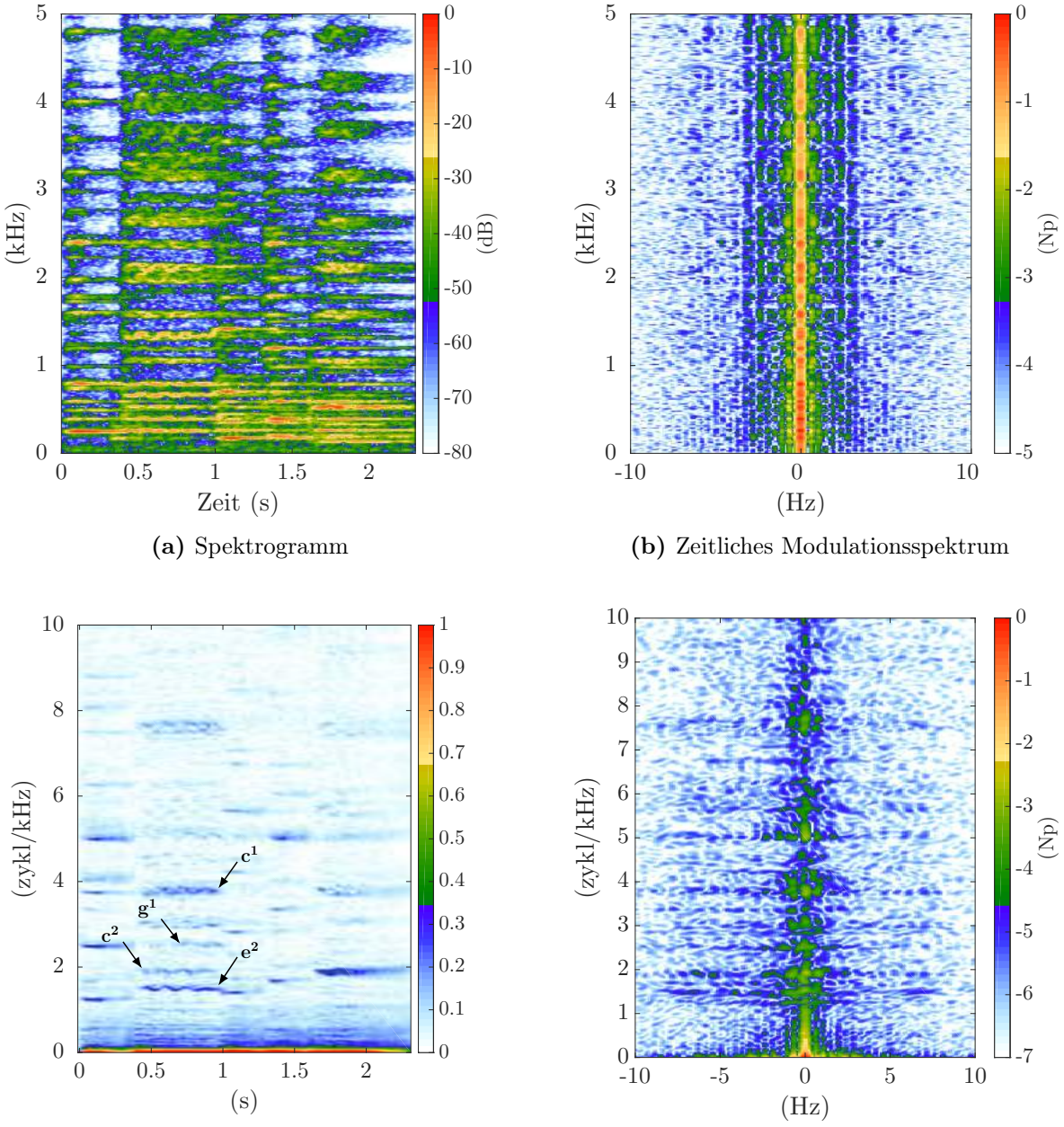


Abb. 3.24: Die ersten vier Takte des Haydn Streichquartetts in C-Dur op. 76, Nr.3. Die Phrase zu Beginn vor der Achtelpause wird zur Erstellung der Diagramme verwendet.

Das Cepstrogramm visualisiert die einzelnen Periodizitäten der Obertöne der vier Töne der Instrumente. Der höchste Ton im C-Dur Akkord (1. Violine e^2 659 Hz) besitzt die wenigsten Perioden innerhalb von einem kHz und erscheint dadurch im Cepstrogramm als tiefste Rahmonische bei ca. 1.51 zykl/kHz. Der tiefste Ton des Akkords (Cello c^1 261 Hz) besitzt hingegen mehr Perioden innerhalb von einem kHz und erscheint bei ca. 3.8 zykl/kHz. Die Grundfrequenzen der zweiten Violine (g^1) bei ca. 391 Hz (2.55 zykl/kHz) und der Viola (c^2) bei ca. 522 Hz sind ebenfalls erkennbar.
Im zeitlichen Modulationsspektrum in (b) ist zu erkennen, dass nahezu alle Frequenzbänder eine Hüllkurvenfluktuation bis ca. 7 Hz aufweisen. Das bedeutet, dass all diese

Frequenzbänder Modulationsenergie bei tiefen Modulationsfrequenzen aufweisen. Dieses Tiefpassverhalten von zeitlichen Modulationen ist generell bei Instrumentalklängen zu finden. Um im nächsten Schritt das MPS zu berechnen, wird entweder das zeitliche Modulationsspektrum entlang der Frequenz-Achse oder das Cepstrogramm entlang der Zeit-Achse fouriertransformiert. Beide Berechnungen münden im MPS. In diesem Fall ist das Ergebnis des MPS einfacher anhand des Cepstrogramms zu interpretieren. Im Cepstrogramm ist beim höchsten Ton (e^2) zu erkennen, dass diesem ein leichtes Vibrato aufgeprägt wurde. Der Frequenzhub der Obertöne in diesem Bereich ist auch im Spektrogramm ersichtlich. Eine Fouriertransformation entlang der Zeit-Achse in diesem Quefrenzbereich ermittelt den spektralen Gehalt dieses Vibratos. Im MPS (d) ist dies bei den zeitlichen Modulationen auf Höhe von ca. 1.5 zyklen/kHz zu sehen. Im Bereich von 15 Hz und unter 9 Hz ist die Modulationsenergie des Vibratos zu erkennen. Dies ist außerdem auch beim (c^1) und beim (c^2) im MPS zu erkennen. Die Modulationsenergie der restlichen Töne konzentriert sich um den Bereich kleiner als 9 Hz, wobei dies sowohl für positive als auch negative zeitliche Modulationen gilt, da bei einer sinusförmigen Bewegung im Cepstrogramm Bewegungen nach oben und nach unten ungefähr gleich stark ausgeprägt sind.



(a) Spektrogramm

(b) Zeitliches Modulationsspektrum

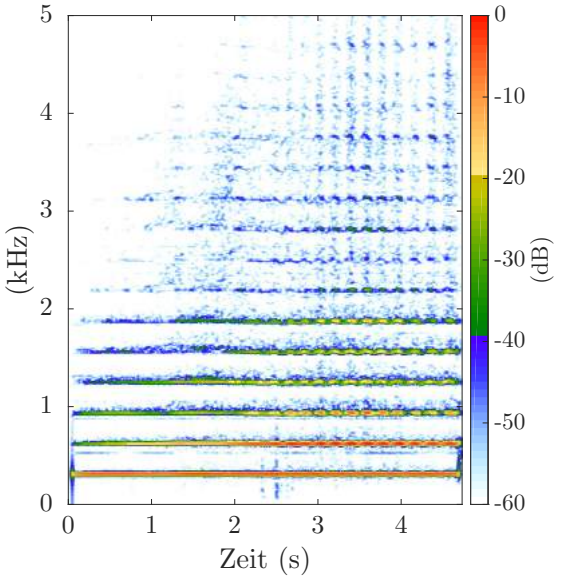
(c) Cepstrogramm

(d) Modulation Power Spectrum

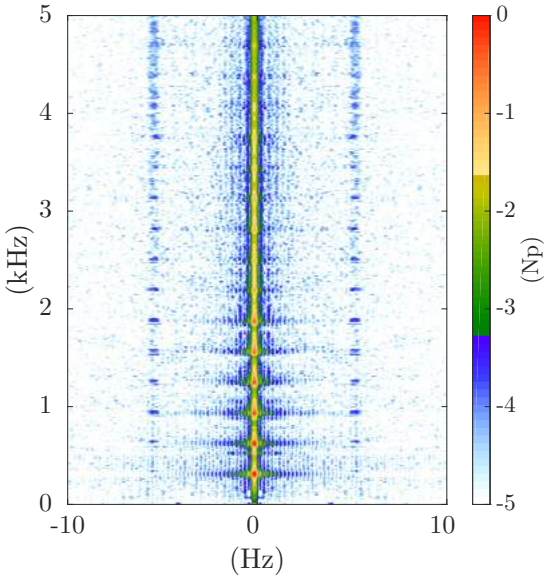
Abb. 3.25: Haydn Streichquartett. Fensterlänge: 30 ms, Hopsize: 330 Samples, Fenster: Hann, vierfache Überlappung. Um eine bessere Auflösung der Modulationen entlang der beiden Achsen zu erhalten, wurde ein Oversampling-Faktor von 10 verwendet.

3.5.2 Flöte mit Tremolo

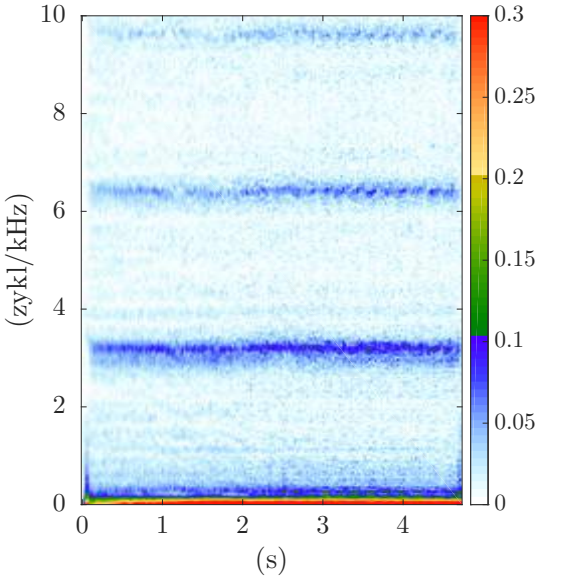
Im Gegensatz zum Vibrato, welches Frequenzmodulationen der Obertöne enthält, werden bei einem Tremolo die Obertöne amplitudenmoduliert. Als Beispiel wird ein Flötenton betrachtet, bei dem die Obertöne zu Beginn eine konstante Amplitude aufweisen. Ab ca. 2 Sekunden sind im Spektrogramm Amplitudenschwankungen zu erkennen, die sich durch periodische Auf- und Abwärtsbewegungen der Amplituden bemerkbar machen. Aufgrund der Periodizität dieser Schwankungen ist dies im zeitlichen Modulationsspektrum zu erkennen. Bei ca. ± 5 Hz wird im zeitlichen Modulationsspektrum diese Periodizität bemerkbar und im vertikalen Abstand der Teiltöne aufgetragen. Im Cepstrogramm sind nur die Periodizitäten der Obertonstruktur des Klanges zu erkennen. Eine Periodizität der Amplitudenschwankungen ist nur schwer zu finden. Dadurch ist in diesem Fall eine Interpretation des MPS durch Heranziehung des zeitlichen Modulationsspektrums anschaulicher. Die Bereiche entlang der zeitlichen Modulationslinie von ± 5Hz weisen in vertikaler Richtung eine Periodizität aufgrund der Obertonstruktur auf. Eine Fouriertransformation entlang der vertikalen Richtung liefert jene Bereiche, in denen sowohl zeitliche als auch spektrale Modulationen erkennbar sind und zwar bei allen Vielfachen des Grundtons von ca. 320 Hz (3.1 zykl/kHz) und bei den zeitlichen Modulationen von ± 5 Hz. Jedoch ist bei den positiven zeitlichen Modulationen mehr konzentrierte Energie im MPS zu erkennen. Dies ist dadurch zu erklären, dass im Spektrogramm nicht nur eine Modulation der Amplitude stattfindet, sondern auch gewissen Obertönen eine Frequenzmodulation aufgeprägt ist. Die Auf- und Abwärtsbewegungen sind in diesem Beispiel ungefähr mit gleicher Intensität im zeitlichen Modulationsspektrum zu sehen. Die geringere Periodizität der zeitlichen Modulationen entlang der spektralen Modulationsachse führt zu einer geringeren Konzentration der Energie im MPS bei den spektralen Modulationen.



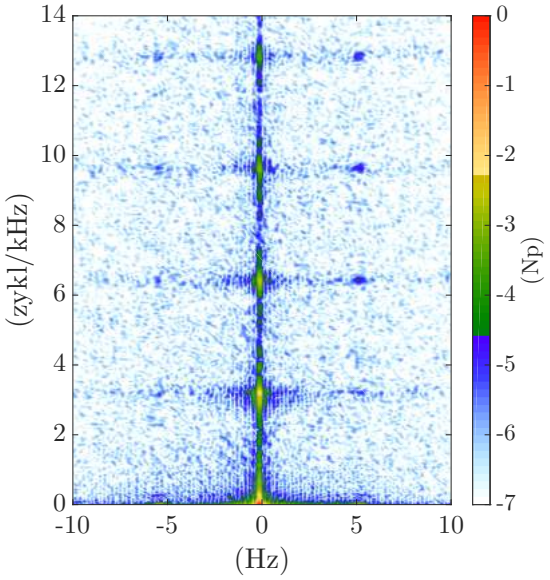
(a) Spektrogramm



(b) Zeitliches Modulationsspektrum



(c) Cepstrogramm

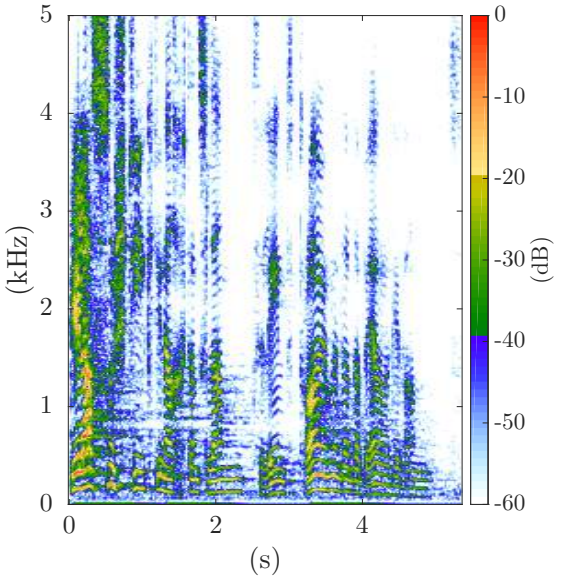


(d) Modulation Power Spectrum

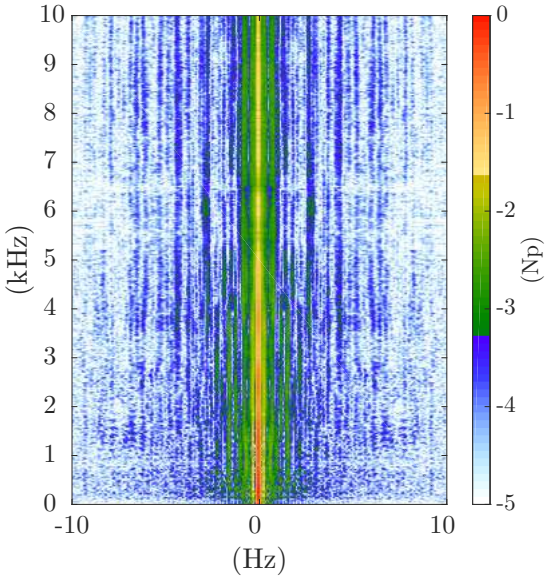
Abb. 3.26: Flöte mit Tremolo. Fensterlänge: 30 ms, Hopsize: 330 Samples, Fenster: Hann, vierfache Überlappung. Um eine bessere Auflösung der Modulationen entlang der beiden Achsen zu erhalten, wurde ein Oversampling-Faktor von 10 verwendet.

3.5.3 Sprache

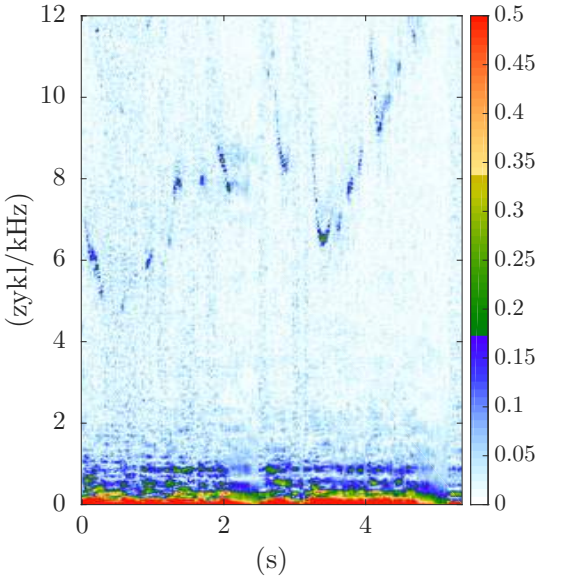
Bei Sprache sind generell drei Bereiche entlang der spektralen Achse zu finden. Bei tiefen spektralen Modulationen entsteht eine dreiecksförmige Energiefläche, welche die Formanten der Sprache beschreibt. Diese geringen spektralen Modulationen werden durch den Vokaltrakt erzeugt und beschreiben die spektrale Einhüllende der Sprache. Die höheren spektralen Modulationen ergeben sich aufgrund der harmonischen Struktur der Sprache, also dem Frequenzabstand der Teiltöne. Für einen männlichen Sprecher (Abb. 3.27) und eine weibliche Sprecherin (Abb. 3.28) ergeben sich verschiedene Bereiche entlang der spektralen Modulationsachse, die durch die harmonische Struktur beschrieben werden. Tiefe männliche Sprache erzeugt höhere spektrale Modulationen, da sich mehr Teiltöne innerhalb eines kHz befinden. Bei einer weiblichen Sprecherin befindet sich der Grundton grundsätzlich in einem höheren Bereich, was in niedrigeren spektralen Modulationen resultiert. Dies ist auch in den beiden Abbildungen (jeweils in (d) zu sehen. Die maximale spektrale Modulation bei einer weiblichen Sprecherin liegt in diesem Fall bei ca. 8 zykl/kHz. Bei einem männlichen Sprecher hingegen werden Laute erzeugt, die auch bei spektralen Modulationen über 8 zykl/kHz Energien erzeugen. Ein weiteres Merkmal von Sprache ist, dass positive und negative zeitliche Modulationen ungefähr die selbe Energie aufweisen. Das bedeutet, dass Bewegungen nach oben (linker Quadrant) und Bewegungen nach unten (rechter Quadrant) mit gleicher Energie im Signal vorkommen.



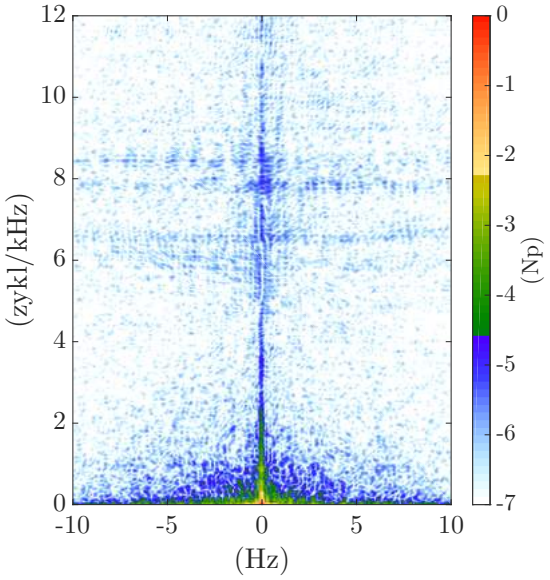
(a) Spektrogramm



(b) Zeitliches Modulationsspektrum

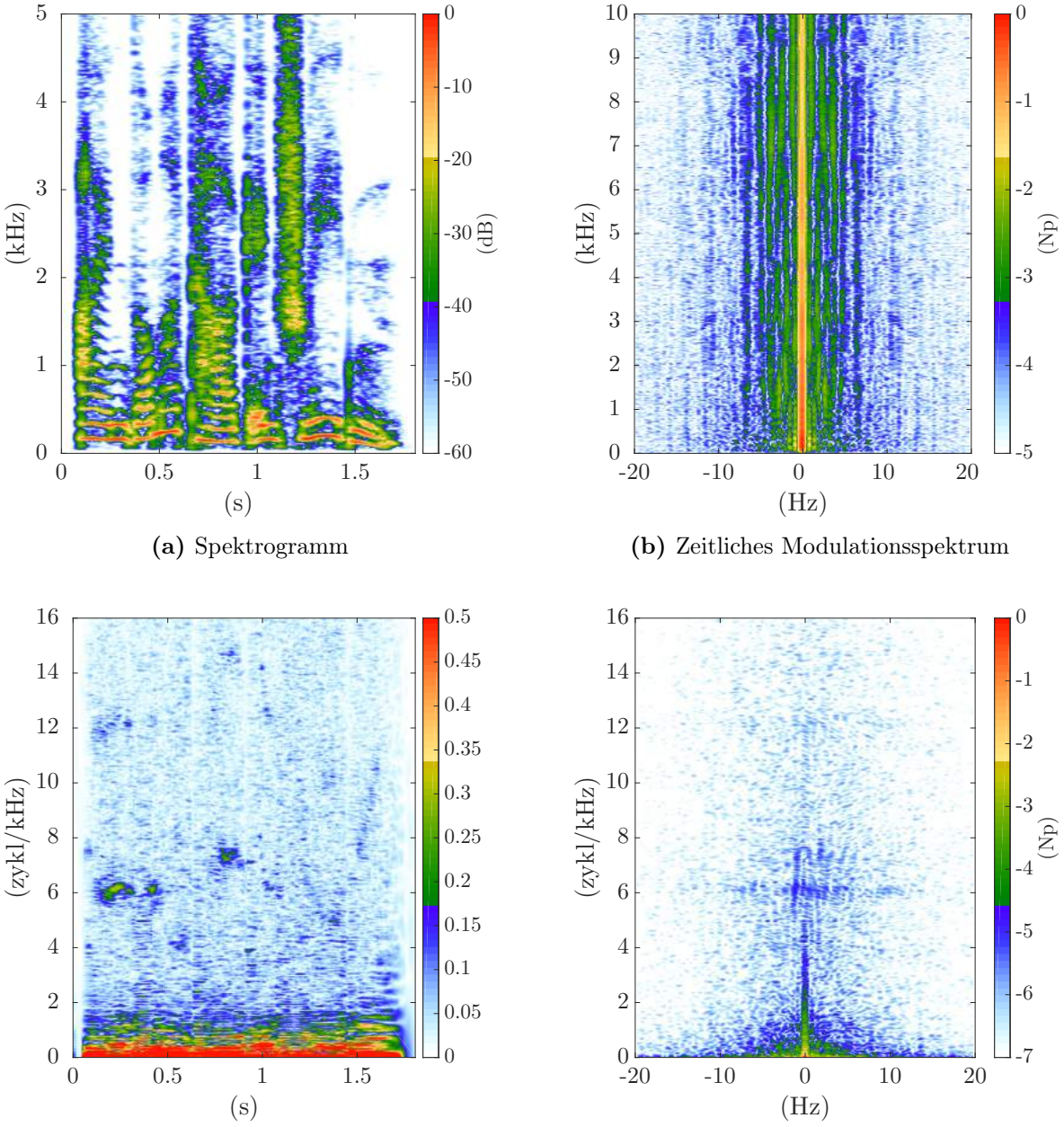


(c) Cepstrogramm



(d) Modulation Power Spectrum

Abb. 3.27: Männlicher Sprecher. Fensterlänge: 30 ms, Hopsize: 330 Samples, Fenster: Hann, vierfache Überlappung. Um eine bessere Auflösung der Modulationen entlang der beiden Achsen zu erhalten, wurde ein Oversampling-Faktor von 10 verwendet.



(a) Spektrogramm

(b) Zeitliches Modulationsspektrum

(c) Cepstrogramm

(d) Modulation Power Spectrum

Abb. 3.28: Weibliche Sprecherin. Fensterlänge: 100 ms, Hopsize: 1102 Samples, Fenster: Hann, vierfache Überlappung. Um eine bessere Auflösung der Modulationen entlang der beiden Achsen zu erhalten, wurde ein Oversampling-Faktor von 10 verwendet.

Kapitel 4

Verarbeitung in der MPS-Domäne

In diesem Kapitel sollen Techniken präsentiert werden, mit denen es möglich ist, das MPS zu manipulieren. Dies geschieht im ersten Schritt anhand von Filtertechniken, die in der Bildverarbeitung üblich sind. Weiters soll es möglich sein, Bereiche in der MPS-Domäne an- oder abzusenken, um zum Beispiel Glissandi in einem Klang manipulieren zu können. Dazu wird eine zweidimensionale Gauß-Verteilung verwendet. Anhand von verschiedenen Klängen soll auch gezeigt werden, wie der Formantbereich und der Bereich von höheren zeitlichen und spektralen Modulationen in einem MPS ausgetauscht werden kann. Abschließend wird untersucht, welche Auswirkung eine Verzerrung der zeitlichen oder spektralen Modulationsachse mit sich zieht und was passiert, wenn eine der Achsen im MPS gespiegelt wird.

4.1 Filterung

Eine Filterung in der zweidimensionalen Domäne wird genau wie in der eindimensionalen Domäne als linearer Vorgang durchgeführt. Dabei wird das Eingangs/Ausgangsverhalten eines linearen, zeitinvarianten zweidimensionalen Filters im Bildbereich durch die Faltungsoperation und im Frequenzbereich durch Multiplikation des Eingangssignalspektrums mit der Übertragungsfunktion des Filters beschrieben. In der Bildverarbeitung wird die Verarbeitung im Bildbereich üblicherweise mit Filterkernen (Matrix) vollzogen [26]:

$$h \star f = \sum_{i=-a}^{a} \sum_{k=-b}^{b} h(i,k) \cdot f(x-i, y-k) \tag{4.1.1}$$

Ein Filterkern der Ordnung 3 für ein Tiefpassfilter[1] wird im Bildbereich folgendermaßen beschrieben:

$$h = \frac{1}{9} \cdot \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} \tag{4.1.2}$$

Dabei wird der Filterkern über einen Bereich des Eingangsbildes gelegt[2]. Danach wird jeder Bildpunkt mit dem darüberliegenden Wert des Filterkerns multipliziert und anschließend all diese Ergebnisse addiert. Aufgrund der Symmetrie des Filterkerns wird das Ergebnis dieses Filterungsprozesses in den mittleren Bildpunkt des Ergebnisbildes gesetzt. Diese Operation wird für alle neuen Bildpunkte ausgeführt, indem der Filterkern über das gesamte Eingangsbild geschoben wird. Dabei wird dieses Filter dazu verwendet, Rauschen zu eliminieren oder Kanten zu verwaschen.
In dieser Arbeit wird eine Filterung oder Manipulation ausschließlich im Frequenzbereich durchgeführt. Da es sich bei der Berechnung des Modulation Power Spectrums um eine Transformation eines Spektrogramms in den Frequenzbereich handelt, müssen neue Begriffe definiert werden. Der Bildbereich entspricht dem Bereich des Spektrogramms. Da in der Literatur üblicherweise Bilder im Bildbereich mit einem kleinen Buchstaben beschrieben werden und im Frequenzbereich mit einem Großbuchstaben, wird hier folgende Nomenklatur verwendet. Für den Bildbereich wird weiter der Großbuchstabe verwendet, da es sich bei einem Spektrogramm um eine Spektraldarstellung über die Zeit handelt. Für den Frequenzbereich, also die Transformation des Spektrogramms anhand der zweidimensionalen Fouriertransformation, wird ein Dach über die Variable gesetzt. Der Frequenzbereich wird forthin als MPS-Domäne bezeichnet. Dies wird in Abbildung 4.1 verdeutlicht.

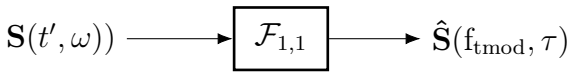


Abb. 4.1: Zur Nomenklatur eines Spektrogramms im Bild- und Frequenzbereich.

Eine Filterung in der MPS-Domäne bedeutet eine Manipulation der Fouriertransformation des Spektrogramms und anschließender Rücktransformation. Ein Spektrogramm der Größe M × N, wo M die Anzahl an Frames und N die Frequenzpunkte im Spektrogramm bezeichnet, wird über folgende Gleichung gefiltert:

[1] Dieser Filterkern wird auch als Mittelwertfilter oder Boxfilter bezeichnet.
[2] Dieser Filterkern wird vor der Faltungsoperation um 180° gedreht. Dies ergibt sich aus der Definition der Faltungsoperation.

$$\mathbf{S_m}(t', \omega) = \mathcal{F}_{\text{-1,-1}}\left[\hat{\mathbf{S}}(\mathrm{f_{tmod}}, \tau) \circ \hat{\mathbf{F}}(\mathrm{f_{tmod}}, \tau)\right]. \tag{4.1.3}$$

Dabei wird $\hat{\mathbf{F}}$ als das Filter bezeichnet, mit welchem $\hat{\mathbf{S}}$ manipuliert wird. Es ist zu beachten, dass es sich beim Produkt in 4.1.3 um eine punktweise Multiplikation[3] handelt. Das Ergebnis einer zweidimensionalen Fouriertransformation liefert eine Matrix mit der selben Größe wie das Spektrogramm. Von diesem Ergebnis können nun einige Begriffe für die MPS-Domäne abgeleitet werden:

$$\text{Realteil} \qquad \hat{\mathbf{R}} = \mathrm{Re}\,(\hat{\mathbf{S}}) \tag{4.1.4}$$

$$\text{Imaginärteil} \qquad \hat{\mathbf{I}} = \mathrm{Im}\,(\hat{\mathbf{S}}) \tag{4.1.5}$$

$$\text{MPS (Betragsspektrum)} \qquad |\hat{\mathbf{S}}(\mathrm{f_{tmod}}, \tau)| = \left[\hat{\mathbf{R}}^2(\mathrm{f_{tmod}}, \tau) + \hat{\mathbf{I}}^2(\mathrm{f_{tmod}}, \tau)\right]^{1/2} \tag{4.1.6}$$

$$\text{Phasenspektrum} \qquad \hat{\phi}(\mathrm{f_{tmod}}, \tau) = \tan^{-1}\left[\frac{\hat{\mathbf{I}}(\mathrm{f_{tmod}}, \tau)}{\hat{\mathbf{R}}(\mathrm{f_{tmod}}, \tau)}\right] \tag{4.1.7}$$

$$\text{Polar Darstellung} \qquad \hat{\mathbf{S}}(\mathrm{f_{tmod}}, \tau) = |\hat{\mathbf{S}}(\mathrm{f_{tmod}}, \tau)|\, e^{\mathrm{j}\phi(\mathrm{f_{tmod}}, \tau)} \tag{4.1.8}$$

$$\text{Powerspektrum} \qquad \hat{\mathbf{P}}(\mathrm{f_{tmod}}, \tau) = |\hat{\mathbf{S}}(\mathrm{f_{tmod}}, \tau)|^2 \tag{4.1.9}$$

Dabei ist anzumerken, dass der Filterungsprozess in erster Linie auf das Betragsspektrum (MPS) angewendet wird. Bei anderen Manipulationen, die im späteren Verlauf noch erläutert werden, können auch komplexe Werte von $\hat{\mathbf{S}}$ bearbeitet werden.

Signalverarbeitungskette

Im Zuge der Signalverarbeitung soll es möglich sein, ein MPS nach der Manipulation wieder in ein Zeitsignal überführen zu können. Dazu wird das manipulierte MPS über die Phase wieder in eine komplexe Form gebracht. Nach anschließender Fourierrücktransformation erfolgt eine Umkehrung der Logarithmierung über die Exponentialfunktion. Das vorliegende neue Spektrogramm wird über die Spektrogramminversion in ein Zeitsignal rücktransformiert, um den Effekt hörbar zu machen (Abb. 4.2).

[3] Dies wird auch als Hadamard Produkt bezeichnet.

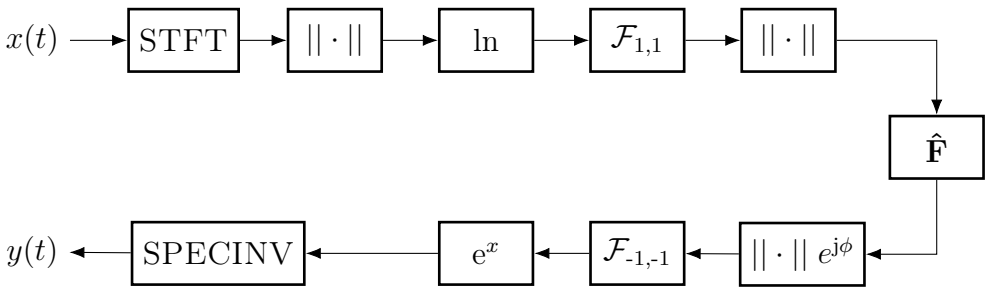


Abb. 4.2: Signalverarbeitungskette

4.1.1 Tiefpass-Filterung

Eine Tiefpass-Filterung in der MPS-Domäne bedeutet, dass höhere zeitliche oder spektrale Modulationen ausgeblendet werden. Dabei wird ein bestimmter Bereich an niedrigeren Modulationen durchgelassen. Diese Gewichtung kann anhand einer Filterfunktion für die Bildverarbeitung so beschrieben werden [27]:

$$\hat{\mathbf{F}}^{c}_{\mathrm{lp}}(\mathrm{f}_{\mathrm{tmod}}, \tau) = \begin{cases} 1 & \sqrt{\mathrm{f}^2_{\mathrm{tmod}} + \tau^2} \leq c \\ 0 & \text{sonst} \end{cases}$$

Diese Filterfunktion wird als idealer Tiefpass bezeichnet und beschreibt einen Zylinder in der MPS-Domäne mit dem Radius c, welcher als Grenzfrequenz bezeichnet wird. Dieses Filter weist den Nachteil auf, dass es sehr steilflankig verläuft und damit im Bildbereich Artefakte erzeugt, welche in der Bildverarbeitung als *blurring* bezeichnet werden. Dieses Artefakt erzeugt ein verschwommenes Bild, da der Übergang bei der Grenzfrequenz hart verläuft. Ein solches Filter kann im Zeitbereich durch eine rundsymmetrische Sincfunktion beschrieben werden. Je kleiner die Grenzfrequenz des Filters in der MPS-Domäne, desto breiter wird die Sincfunktion im Bildbereich. Da im Bildbereich die Sinc-Funktion mit dem Spektrogramm gefaltet wird, wirken die Nebenkeulen auch in größeren Bereichen des Bildes, welche dadurch ein verschwommenes Bild erzeugen. Es ist nun erforderlich, einen weicheren Übergang zwischen Durchlassbereich und Sperrbereich zu erzeugen. Dazu bieten sich mehrere Filter an:

2D-Butterworth-Filter

$$\hat{\mathbf{F}}^{c}_{\mathrm{lp}}(\mathrm{f}_{\mathrm{tmod}}, \tau) = \frac{1}{1 + \left(\frac{\mathrm{f}^2_{\mathrm{tmod}} + \tau^2}{c^2}\right)^{2n}} \tag{4.1.10}$$

2D-Gauß-Filter

$$\hat{\mathbf{F}}_{\mathrm{lp}}^{\sigma}(\mathrm{f_{tmod}}, \tau) = \mathrm{e}^{-\frac{\mathrm{f_{tmod}}^2+\tau^2}{2\sigma^2}} \tag{4.1.11}$$

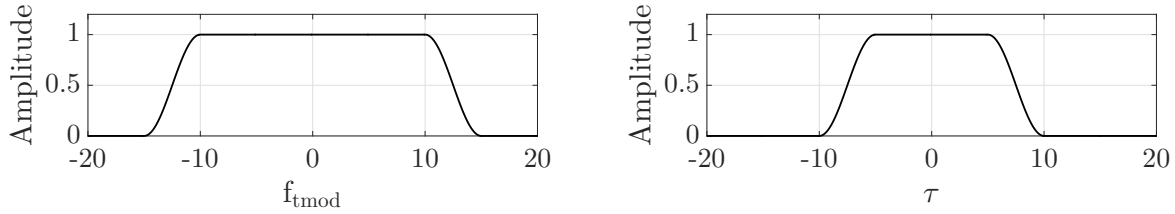
Filterung entlang einer Achse

Da es aber auch möglich sein soll, eine Filterung durchzuführen, bei der nur spektrale Modulationen gefiltert werden sollen und keine zeitlichen Modulationen, sind Filter mit einer kreisrunden Charakteristik unbrauchbar. Deswegen wird ein rechteckiger Bereich definiert, in dem Modulationen bestehen bleiben. Außerhalb dieses Rechtecks werden Modulationen unterdrückt. Der Übergang wird über ein Cosinus-Roll-Off gesteuert, dessen Breite einstellbar ist.

2D-Cosinus-Rolloff Tiefpass-Filter

$$\hat{\mathbf{F}}_{\mathrm{lp}}^{f_g,\tau_g}(\mathrm{f_{tmod}}, \tau) = \begin{cases} 1 & |\mathrm{f_{tmod}}| < f_g \wedge |\tau| < \tau_g \\ \cos\left(\frac{|\mathrm{f_{tmod}}|-f_g}{d_f}\frac{\pi}{2}\right)^2 & |\mathrm{f_{tmod}}| > f_g \wedge |\mathrm{f_{tmod}}| < f_g + d_f \\ \cos\left(\frac{|\tau|-\tau_g}{d_\tau}\frac{\pi}{2}\right)^2 & |\tau| > \tau_g \wedge |\tau| < \tau_g + d_\tau \\ 0 & \text{sonst} \end{cases}$$

Die Parameter f_g und τ_g geben die obere Grenzfrequenz in der jeweiligen Domäne an. Die Parameter d_f und d_τ geben die Breite des Cosinus-Rolloff an. Im weiteren Verlauf wird nur mehr dieses Filter zur Tiefpass-Filterung verwendet.

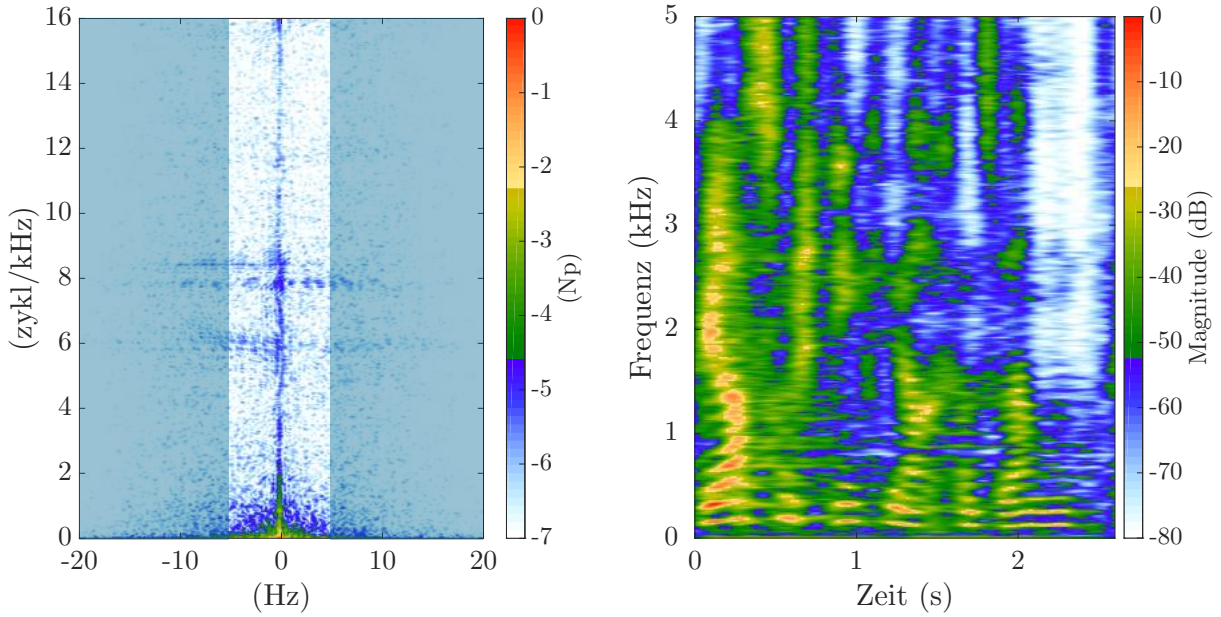


(a) Filter $\mathrm{f_{tmod}}$ - Achse **(b)** Filter τ - Achse

Abb. 4.3: Darstellung eines Tiefpass-Filters jeweils entlang einer Achse mit den Grenzfrequenzen $f_g = 10$ Hz (a) und $\tau_g = 5$ zykl/kHz (b) und den Breite-Faktoren $d_f = 5$ Hz (a) und $d_\tau = 5$ zykl/kHz (b).

Tiefpass-Filterung von zeitlichen Modulationen

Als Beispiel werden bei einem männlichen Sprecher temporale Modulationen größer als 5 Hz weggefiltert (Abb. 4.4). Dabei ist im Spektrogramm klar zu erkennen, dass die spektrale Information im Klangsignal bestehen bleibt, wohingegen zeitliche Informationen verwaschen werden und die Artikulation der Sprache verloren geht. Wird diese Grenzfrequenz noch weiter verringert, sinkt auch die Sprachverständlichkeit, da Sprache aufgrund ihrer zeitlichen Modulationen erkannt wird.



(a) MPS **(b)** Spektrogramm

Abb. 4.4: Darstellung des tiefpassgefilterten MPS bei $f_g = 5$ Hz und $\tau_g = 50$ zykl/kHz und dem daraus resultierenden Spektrogramm.

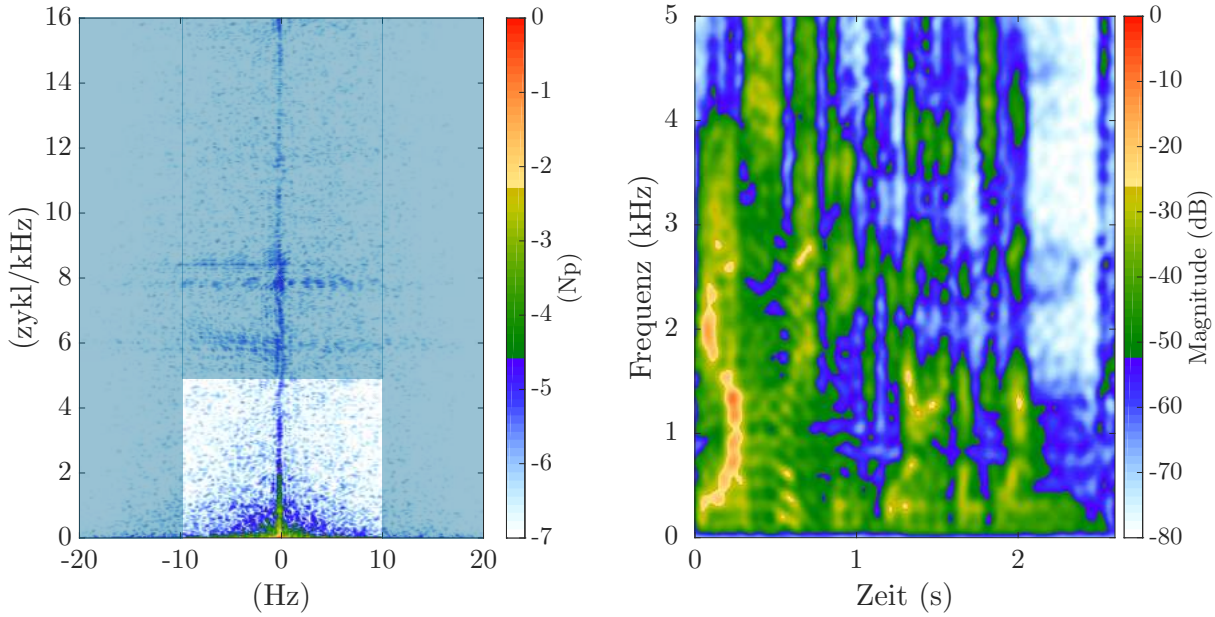
Filtervorgang:

$$\mathbf{S_m}(t', \omega) = \mathcal{F}_{\text{-1,-1}}\left[\hat{\mathbf{S}}(\mathrm{f_{tmod}}, \tau) \circ \hat{\mathbf{F}}_{\mathrm{lp}}^{5,50}(\mathrm{f_{tmod}}, \tau)\right]. \tag{4.1.12}$$

Klangbeispiel: `male_p_lp_s50_t5.wav`

Tiefpass-Filterung von zeitlichen und spektralen Modulationen

Werden hingegen auch spektrale Modulationen im Klangsignal unterdrückt, hat dies Auswirkungen auf die Tonhöhenbestimmung des Sprechers (Abb. 4.5). Im MPS ist zu sehen, dass spektrale Modulationen im Bereich von $6 - 8.5$ zykl/kHz vorkommen. Dies entpricht einer Grundtonhöhe von ca. 120 Hz bis 160 Hz. Da die spektralen Modulationen nach der Filterung fehlen, fehlt auch die Periodizität im Spektrum und somit die Obertongestalt. Eine Tonhöhenbestimmung wird somit erschwert. Die Sprachverständlichkeit bleibt hingegen erhalten, da der Formantbereich bestehen bleibt. Zusätzlich wurde auch die zeitliche Grenzfrequenz auf 10 Hz erhöht, was den Artikulationsverlust des vorherigen Beispiels wieder ausgleicht.



(a) MPS **(b)** Spektrogramm

Abb. 4.5: Darstellung des gefilterten MPS und dem daraus resultierenden Spektrogramm.

Filtervorgang:

$$\mathbf{S_m}(t', \omega) = \mathcal{F}_{\text{-1,-1}}\left[\hat{\mathbf{S}}(\mathrm{f_{tmod}}, \tau) \circ \hat{\mathbf{F}}_{\mathrm{lp}}^{10,5}(\mathrm{f_{tmod}}, \tau)\right]. \tag{4.1.13}$$

Klangbeispiel: `male_p_lp_s5_t10.wav`

4.1.2 Hochpass-Filterung

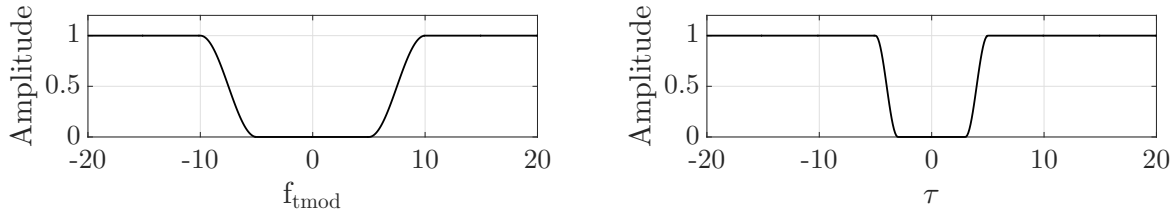
Eine Filterung, bei der höhere Modulationen durchgelassen und tiefe unterdrückt werden, wird als Hochpass-Filterung bezeichnet. Dadurch ist es möglich, höhere Modulationen von niedrigeren zu trennen. Im MPS befinden sich Formantbereiche eines Klanges bei niedrigen Modulationen entlang der τ-Achse, wohingegen Periodizitäten der Obertonstruktur bei höheren Modulationen zu finden sind. Bei einer eindimensionalen Darstellung werden cepstrale Lifter dazu eingesetzt, diese beiden Bereiche zu trennen. Ein cepstraler Hochpasslifter bei einer Grenzfrequenz von 1 zykl/kHz unterdrückt die Formanten, die breiter als der Bereich von einem kHz sind. Die Formanten können als Schwingung gesehen werden, bei denen die lokalen Maxima überlicherweise eine Breite von bis zu einem kHz haben. Diese Formanten, die dem Spektrum aufgeprägt sind, können mithilfe eines Hochpass-Filters unterdrückt werden.

2D-Cosinus-Rolloff Hochpass-Filter

Als Inversion des Tiefpass-Filters wird der zweidimensionale Hochpass-Filter mit einem Cosinus-Rolloff betrachtet:

$$\hat{\mathbf{F}}_{\mathrm{hp}}^{f_g,\tau_g}(\mathrm{f_{tmod}},\tau) = \begin{cases} 1 & |\mathrm{f_{tmod}}| > f_g \wedge |\tau| > \tau_g \\ \cos\left(\frac{|\mathrm{f_{tmod}}|-f_g}{d_f}\frac{\pi}{2}\right)^2 & |\mathrm{f_{tmod}}| < f_g \wedge |\mathrm{f_{tmod}}| > f_g - d_f \\ \cos\left(\frac{|\tau|-\tau_g}{d_\tau}\frac{\pi}{2}\right)^2 & |\tau| < \tau_g \wedge |\tau| > \tau_g + d_\tau \\ 0 & \text{sonst} \end{cases}$$

Die Parameter f_g und τ_g geben die Grenzfrequenz in der jeweiligen Domäne an. Die Parameter d_f und d_τ geben die Breite des Cosinus-Rolloffs an (Abb. 4.6).



(a) Filter $\mathrm{f_{tmod}}$ - Achse **(b)** Filter τ - Achse

Abb. 4.6: Darstellung eines Hochpassfilters jeweils entlang einer Achse mit den Grenzfrequenzen $f_g = 10$ Hz (a) und $\tau_g = 5$ zykl/kHz (b) und den Breite-Faktoren $d_f = 5$ Hz (a) und $d_\tau = 2$ zykl/kHz (b).

Hochpass-Filterung von spektralen Modulationen

Als konkretes Beispiel wird eine Sprecherin betrachtet, deren Modulationen unter 1 zykl/kHz ausgeblendet werden (Abb. 4.7).

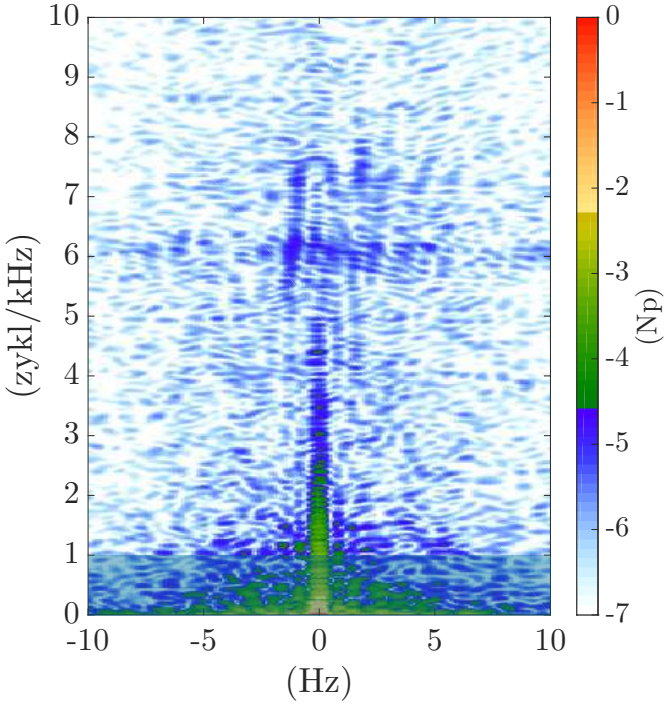


Abb. 4.7: Hochpassfilterung der spektralen Modulationen.

In Abb. 4.8 ist ein Frame bei Sekunde 0.24 zu sehen. Die blaue Linie gibt den spektralen Verlauf zu diesem Zeitpunkt ohne Filterung an. Die Grenzfrequenz der roten Linie beträgt 0.1 zykl/kHz, womit spektrale Modulationen, die sich im Spektrum über einen Bereich von mehr als 10000 Hz erstrecken, weggefiltert werden. Es ist zu sehen, dass die Formantstruktur noch erhalten bleibt, jedoch die Amplituden der einzelnen Frequenzen reduziert wurden. Dies liegt daran, dass auch der Gleichanteil und somit die Verschiebung des Spektrums in vertikaler Richtung entfernt wurde. Bei einer Erhöhung der Grenzfrequenz werden lokale Maxima im Spektrum entfernt, die sich über einen kleinen Bereich erstrecken. Ab 1 zykl/kHz werden lokale Maxima weggefiltert, die sich über einen Bereich von 1 kHz erstrecken. Befindet sich die Grenzfrequenz noch unterhalb der spektralen Modulation der Grundtonhöhe, ist die Sprachstruktur noch zu erkennen. Da aber die Formantstruktur durch eine Erhöhung völlig verloren geht, sinkt auch die Sprachverständlichkeit. Ab ca. 2 zykl/kHz ist

zwar der Obertongehalt zu hören, die Formantstruktur versinkt aber sukzessive in einem Rauschteppich.

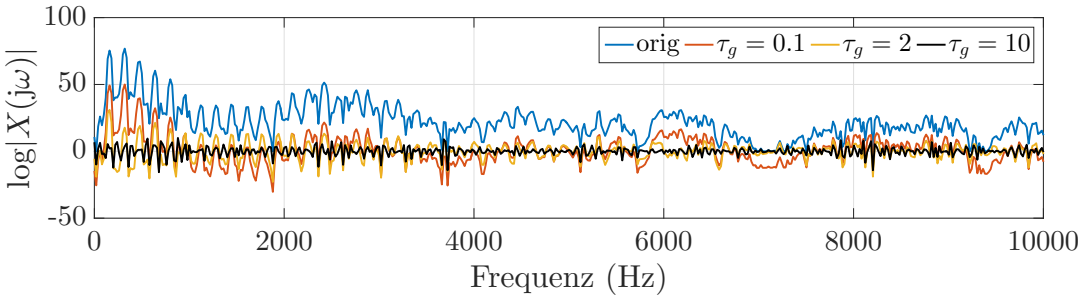
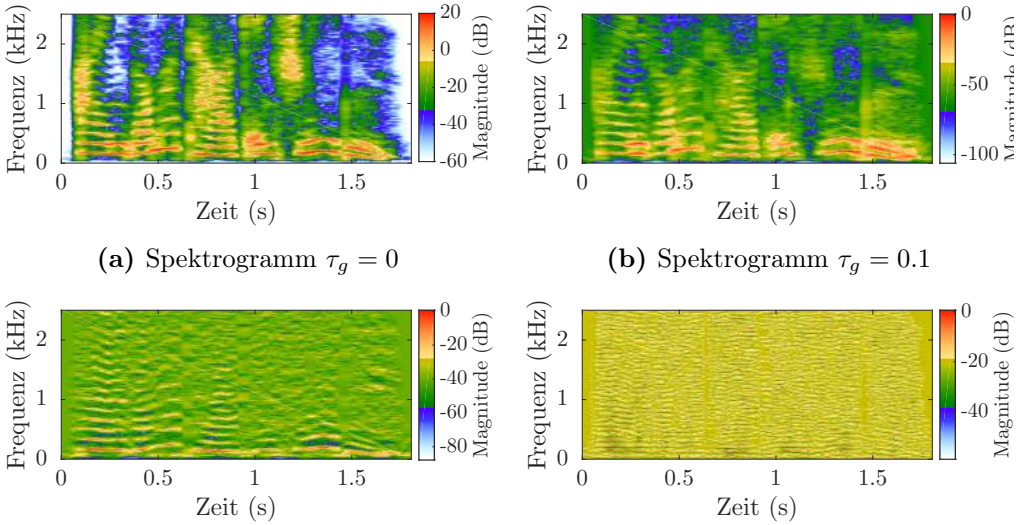


Abb. 4.8: Darstellung des Spektrums eines Frames bei Sekunde 0.24 nach einer Hochpassfilterung der spektralen Modulationen.

Übersteigt die Grenzfrequenz die Anzahl an Modulationen, bei der sich der Grundton und seine Vielfachen befinden, ist das ursprüngliche Klangsignal nicht mehr zu erkennen. Dieser Sachverhalt ist auch in den Spektrogrammen in Abb. 4.9 (b)-(d) ersichtlich. Die Formantstruktur ist in (c) noch zu erkennen, verschwindet jedoch völlig in (d).



(a) Spektrogramm $\tau_g = 0$ **(b)** Spektrogramm $\tau_g = 0.1$

(c) Spektrogramm $\tau_g = 2$ **(d)** Spektrogramm $\tau_g = 10$

Abb. 4.9: Hochpassfilterung von spektralen Modulationen.

Filtervorgang:

$$\mathbf{S_m}(t', \omega) = \mathcal{F}_{\text{-1,-1}}\left[\hat{\mathbf{S}}(\mathrm{f_{tmod}}, \tau) \circ \hat{\mathbf{F}}_{\mathrm{hp}}^{0,1}(\mathrm{f_{tmod}}, \tau)\right]. \tag{4.1.14}$$

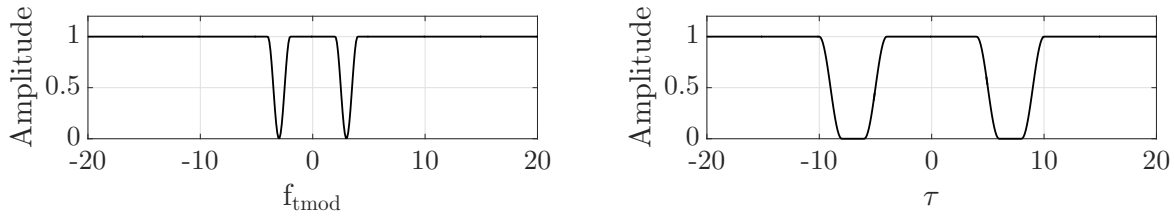
Klangbeispiel: `male_p_hp_s1_t0.wav`

4.1.3 Notch-Filterung

Ein Notch-Filter[4] wird dazu verwendet, um einzelne Bereiche eines MPS zu unterdrücken und zwar Bereiche, die sich zwischen zwei Grenzfrequenzen befinden. Es werden also für jede Achse zwei Grenzfrequenzen benötigt. Die Grenzfrequenz mit dem Index 1 gibt jeweils die untere Grenzfrequenz an. Dabei bezeichnen f_{g1} und f_{g2} die Grenzfrequenzen der zeitlichen Modulationsachse und τ_{g1} und τ_{g2} die Grenzfrequenzen der spektralen Modulationsachse. Die Bereiche eines Notch-Filters werden demnach definiert mit:

$$\hat{\mathbf{F}}_{\text{notch}}^{f_{g_1}\text{-}f_{g_2},\tau_{g_1}\text{-}\tau_{g_2}}(\mathrm{f}_{\text{tmod}},\tau) = \begin{cases} 1 & (|\mathrm{f}_{\text{tmod}}| > f_{g_1} \wedge |\tau| > \tau_{g_1}) \vee (|\mathrm{f}_{\text{tmod}}| < f_{g_2} \wedge |\tau| < \tau_{g_2}) \\ \cos\left(\frac{|\mathrm{f}_{\text{tmod}}|-f_g}{d_f}\frac{\pi}{2}\right)^2 & (|\mathrm{f}_{\text{tmod}}| > f_{g_1} \wedge |\mathrm{f}_{\text{tmod}}| < f_{g_1}+d_f) \vee (|\mathrm{f}_{\text{tmod}}| < f_{g_2} \wedge |\mathrm{f}_{\text{tmod}}| > f_{g_2}-d_f) \\ \cos\left(\frac{|\tau|-\tau_g}{d_\tau}\frac{\pi}{2}\right)^2 & (|\tau| > \tau_{g_1} \wedge |\tau| < \tau_{g_1}+d_\tau) \vee (|\tau| < \tau_{g_2} \wedge |\tau| > \tau_{g_2}-d_\tau) \\ 0 & \text{sonst} \end{cases}$$

Es ist damit möglich, gezielt Modulationen, die sich abseits der beiden Achsen befinden, zu manipulieren. Befinden sich zum Beispiel in einem Klang zeitliche periodische Bewegungen von Obertönen aufwärts, die in einem ganzzahligen Verhältnis zum Grundton stehen, wird dies im MPS im linken oberen und rechten unteren Quadranten dargestellt. Eine Bewegung nach unten wird demnach im rechten oberen und linken unteren Quadranten abgebildet. Die genaue Position ergibt sich aus den zeitlichen bzw. spektralen Modualtionen.



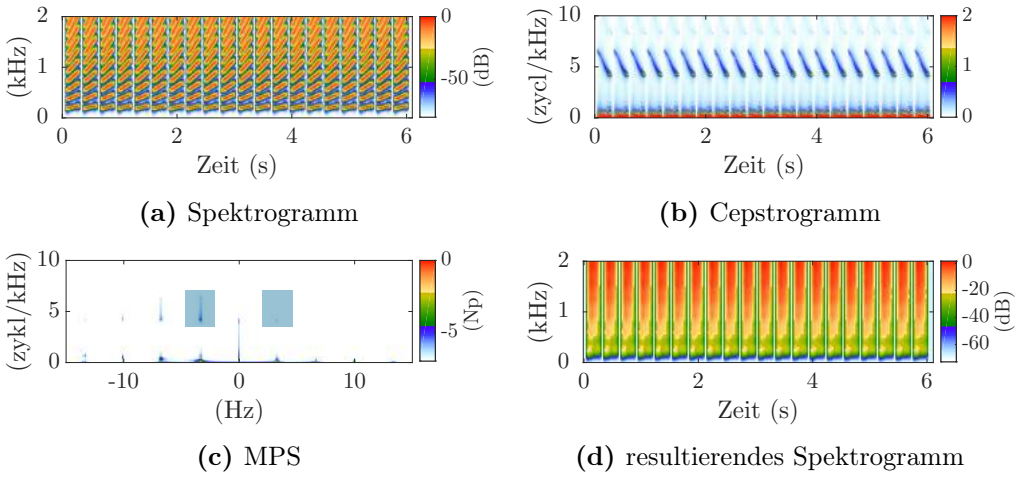
(a) Filter f_{tmod} - Achse **(b)** Filter τ - Achse

Abb. 4.10: Darstellung eines Notchfilters jeweils entlang einer Achse mit den Grenzfrequenzen $f_{g1} = 10$ Hz und $f_{g2} = 10$ Hz (a) und $\tau_{g1} = 5$ zykl/kHz und $\tau_{g2} = 5$ zykl/kHz (b) und den Breite-Faktoren $d_f = 1$ Hz (a) und $d_\tau = 2$ zykl/kHz (b).

Als Beispiel wird ein kurzes Chirp-Signal betrachtet, welches Obertöne mit ganzzahligen Vielfachen enthält (Abb 4.11 (a)). Diese Chirp-Signale bewegen sich in kurzen Abständen mit der Frequenz nach oben. Der Zeitbereich, in dem ein Chirp erklingt, ist 200 ms lang. Dabei verläuft der Chirp mit der niedrigsten Frequenz von $150-250$ Hz. Der erste Oberton

[4] In der englischsprachigen Literatur wird ein Kerb-Filter auch als Notch-Filter bezeichnet.

bewegt sich von $300 - 500$ Hz. Dazwischen befindet sich Stille mit einer Dauer von 100 ms. Das ergibt eine Periodendauer von 300 ms. Im Cepstrogramm (b) ist zu sehen, wie sich die Cepstren der Chirps über die Zeit verhalten. Die Rahmonische mit der geringsten Frequenz macht eine Bewegung nach unten und die periodische Fortsetzung bleibt erhalten, da die Fouriertransformation nur entlang der Frequenzachse verläuft. Im Cepstrogramm ist zu sehen, dass die Abstände der Linien in der horizontalen Achse immer gleich sind. Daraus resultiert die Abbildung im MPS bei genau dieser Periodizität und zwar bei 3 Hz, was genau der Periodendauer aus dem Spektrogamm entspricht. Der Bereich der spektralen Modulationen ergibt sich über die Periodizität der Obertöne im Spektrogramm. Da sich der Frequenzabstand der Obertöne vom Beginn eines Chirps bis zum Ende verändert, verändern sich auch die spektralen Modulationen. Der Frequenzabstand beträgt zu Beginn 150 Hz (6,66 zykl/kHz) und zum Ende 250 Hz (4 zykl/kHz). Dies wird im MPS durch eine vertikale Linie bei den Vielfachen von 5 zeitlichen Modulationen und im Bereich von $4 - 6,66$ zykl/kHz wiedergegeben. Diese Abbildung in der linken Hälfte rührt daher, dass jeder Chirp eine positive Sweeprate besitzt und somit anhand der Phase nur in dieser abgebildet wird. Würden sich die Chirps von oben nach unten im Spektrogramm bewegen, würde dies in der rechten Hälfte des MPS abgebildet werden.



(a) Spektrogramm **(b)** Cepstrogramm

(c) MPS **(d)** resultierendes Spektrogramm

Abb. 4.11: Notch-Filterung eines periodischen Gauß-Chirps mit Obertönen. Bei 3-4 Hz zeitlichen Modulationen und 4-6 zykl/kHz spektralen Modulationen wird der Bereich im MPS abgesenkt.

Ein Notchfilter kann in diesem Beispiel dazu verwendet werden, die Periodizität in den Obertönen zu zerstören. Dazu müssen die Energien im Bereich von $4 - 6,66$ zykl/kHz und bei 3 Hz (zeitlich) gleich Null gesetzt werden. Das Spektrogramm, welches aus dieser Filterung resultiert, wird über folgende Berechnung erreicht und in (d) dargestellt:

$$\mathbf{S_m}(t', \omega) = \mathcal{F}_{\text{-1,-1}}\left[\hat{\mathbf{S}} \circ \hat{\mathbf{F}}_{\text{notch}}^{\text{3-5,4-6}}\right]. \tag{4.1.15}$$

Es ist zu sehen, dass die Periodizität entlang der Frequenzachse, also der Obertongehalt, nicht weiter vorhanden ist. Die Periodizität in der Zeit wurde jedoch nicht zerstört, da im MPS auch Vielfache der zeitlichen Modulationen vorhanden sind. Diese Vielfachen tragen dazu bei, dass die Periodizität auch durch eine Notchfilterung nicht zerstört wird. Würden sich diese Chirps im Spektrogramm von oben nach unten bewegen, würden die Linien im Cepstrogramm nach oben verlaufen und im MPS ergibt sich eine Bereich im rechten oberen Quadranten.

4.2 Weitere Manipulationen

In diesem Abschnitt werden unter anderem Manipulationen betrachtet, in denen das MPS mit einer Maske multipliziert wird, die Werte größer als 1 enthält. Weiters werden Verzerrungen der Achsen und Spiegelungen um eine Achse erläutert. Den Abschluss bildet ein Morphingprozess, in dem Bereiche im MPS von zwei verschiedenen Klängen zusammengeführt werden.

4.2.1 Anheben oder Absenken von Bereichen

Das oben in Abschnitt 4.1.3 beschriebene Notchfilter ist symmetrisch für eine Achse, was bedeutet, dass sowohl positive als auch negative zeitliche und spektrale Modultionen gefiltert werden. Es soll nun möglich sein, nur einen Quadranten[5] des MPS zu manipulieren. Zu diesem Zweck werden zweidimensionale Gauß-Filter eingesetzt, mit denen ein Bereich angehoben oder abgesenkt werden kann. Diese weisen ihr Maximum bei der zu filternden zeitlichen und spektralen Modulation auf und sind um den Nullpunkt MPS(0,0) punktsymmetrisch.

$$\hat{\mathbf{F}}_{\text{gauß}}^{\mu_f,\mu_\tau,\beta} = 1 + (\beta - 1) \cdot \exp\Big(-\frac{1}{2}(\mathbf{x} - \mu_f)^T \Sigma^{-1} (\mathbf{x} - \mu_\tau)\Big) \tag{4.2.1}$$

[5] Aufgrund der Symmetrie der zweidimensionalen Fouriertransformation wird eine Filterung jeweils bezüglich der Punktsymmetrie ausgeführt. Eine Filterung im ersten Quadranten entspricht gleichzeitig auch einer Filterung im dritten Quadranten. Genauso entspricht einer Filterung im zweiten Quadranten einer Filterung im vierten Quadranten.

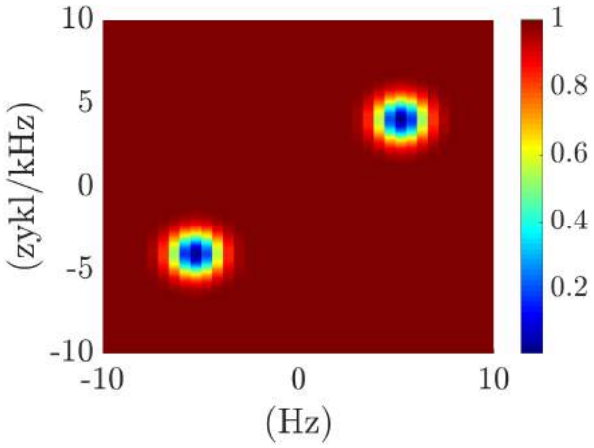
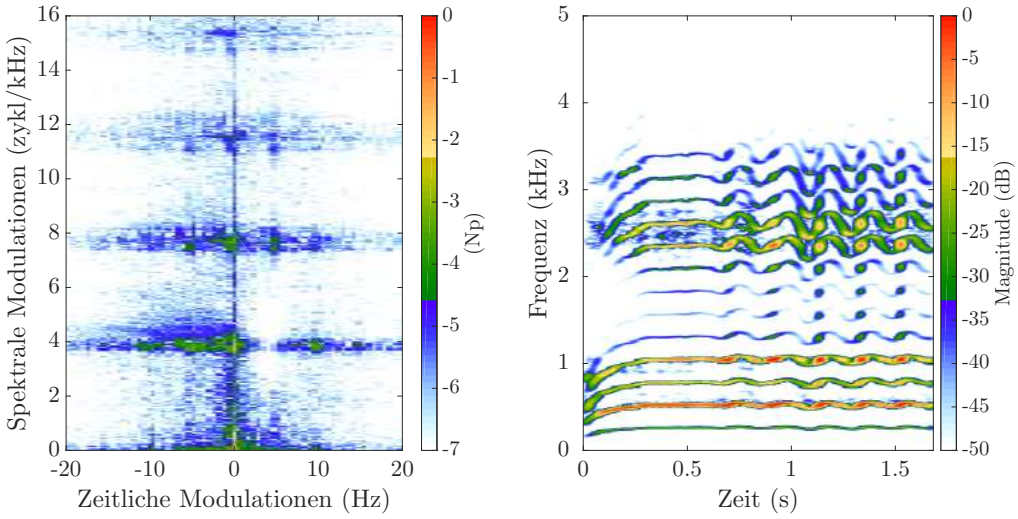


Abb. 4.12: Gauß-Filter, mit der das MPS manipuliert wird. Parameter: $\mu_f = 5$ Hz, $\mu_\tau = 4$ zykl/kHz, , $\beta = 0.01$, $\sigma_f = \sigma_\tau = 0.8$.

Dabei wird das MPS des Baritons verwendet. Im Spektrogramm des Baritons ist zu sehen, dass es seitens der Obertöne sowohl Bewegungen nach oben, als auch nach unten gibt. Die Bereiche, in denen die Momentanfrequenz eine positive Ableitung besitzt, werden im MPS im zweiten Quadranten abgebildet und Bewegungen, bei denen der Sinus eine negative Ableitung besitzt, werden im ersten Quadranten aufgetragen. Wird nun ein Filter dazu verwendet, um eine dieser Bewegungen zu unterdrücken, wird der dementsprechende Bereich im MPS ausgeblendet (Abb. 4.14 (a)).



(a) MPS **(b)** Spektrogramm

Abb. 4.13: Darstellung MPS und dem daraus resultierenden Spektrogramm nach Absenken bei $f_{tmod} = 5$ Hz und $\tau = 4$ zykl/kHz.

Das resultierende Spektrogramm (b) ergibt sich mit

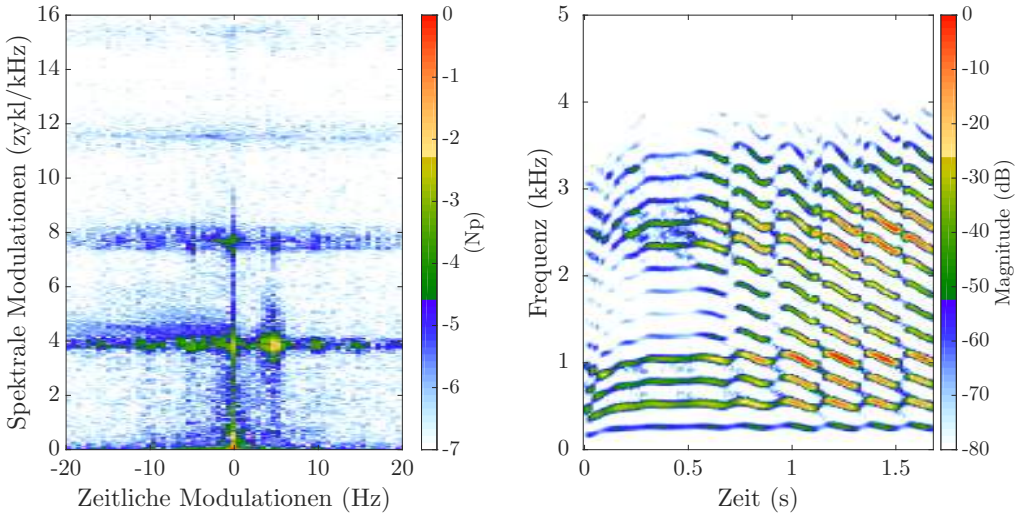
$$\mathbf{S_m}(t', \omega) = \mathcal{F}_{\text{-1,-1}}\left[\hat{\mathbf{S}} \circ \hat{\mathbf{F}}^{5,4,0.01}_{\text{gauß}}\right]. \tag{4.2.2}$$

Klangbeispiel: `bariton_p_red0.01_s4_t5.wav`

In Abb. 4.13 ist zu sehen, dass im Spektrogramm Bewegungen nach unten weniger Energie aufweisen, als Bewegungen nach oben. Eine Multiplikation mit einer Maske, in der Werte vorhanden sind, die größer als 1 sind, wird als Expansion bezeichnet. Werte zwischen 0 und 1 demnach als Kompression. Dies ist anhand des Logarithmus zu erklären. Da von der STFT der Betrag und anschließend der Logarithmus berechnet wurde, ergibt eine Multiplikation in der MPS-Domäne eine Potenzierung in der zeitlichen Domäne nach der Resynthese.

$$\mathbf{S} = \hat{\mathbf{F}} \cdot \log(\hat{\mathbf{S}}) = \log(\hat{\mathbf{S}}^{\hat{\mathbf{F}}}) \tag{4.2.3}$$

Um dies zu verdeutlichen, wird das MPS mit dem selben Gauß-Filter wie im vorigen Beispiel multipliziert, jetzt jedoch mit $\beta = 5$. Bei der Resynthese werden Bewegungen nach unten mit dem Faktor 5 potenziert, was auch im resultierenden Spektrogramm (Abb. 4.14) zu sehen ist.



(a) MPS **(b)** Spektrogramm

Abb. 4.14: Darstellung MPS und dem daraus resultierenden Spektrogramm nach einer Potenzierung bei $f_{tmod} = 5$ Hz und $\tau = 4$ zykl/kHz mit $\beta = 5$.

Klangbeispiel: `bariton_p_add_s4_t5_beta5.wav`

Anders als bei der Multiplikation mit einem Filter in der MPS-Domäne erzeugt eine Addition oder Subtraktion mit einer Maske keine Expansion oder Kompression in diesem Bereich. Es wird dadurch der Bereich in einem linearen Maßstab angehoben oder abgesenkt und nicht wie bei der Multiplikation über die Potenz verstärkt.

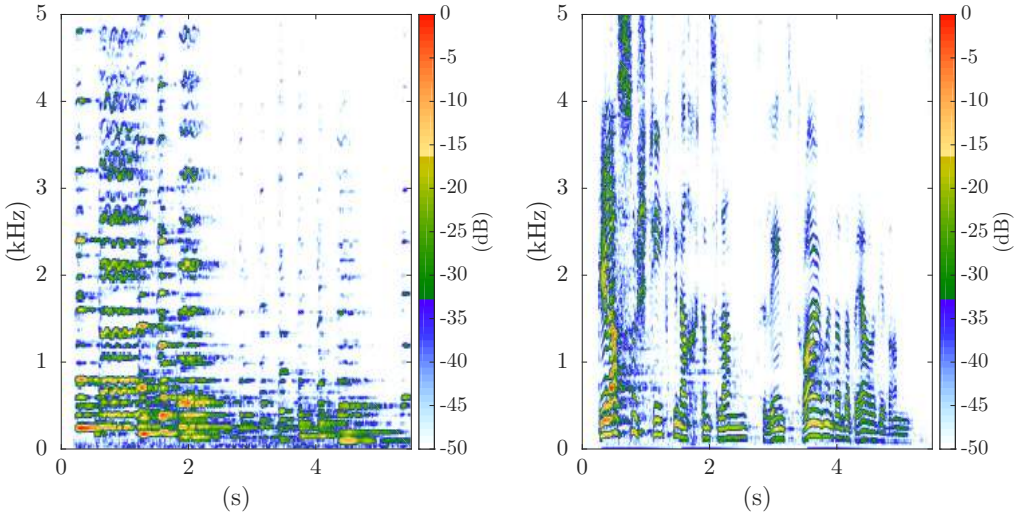
4.2.2 Morphing

Eine andere Art der Manipulation ergibt sich, wenn der Formantbereich eines Klanges mit dem Bereich der Teiltonstruktur eines anderen Klanges kombiniert wird. Dies geschieht anhand einer Hoch- bzw. Tiefpassfilterung. Beim ersten Klang wird der Kern des MPS durch ein Hochpassfilter gefiltert und gespeichert. Dies entspricht dem Bereich eines MPS, welches höhere zeitliche und spektrale Modulationen enthält. Der Bereich, in dem der Hochpassfilter einsetzt, kann beliebig variiert werden. Wird nun der Formantbereich eines zweiten Klanges durch eine Tiefpassfilterung bei den selben Grenzfrequenzen durchgeführt und mit dem MPS des hochpassgefilterten ersten Klanges zusammengeführt, erfolgt eine Fusionierung der beiden Klänge. Je nach Positionierung der Grenzen werden andere Ergebnisse erzielt. Der Manipulationsvorgang kann wie folgt notiert werden:

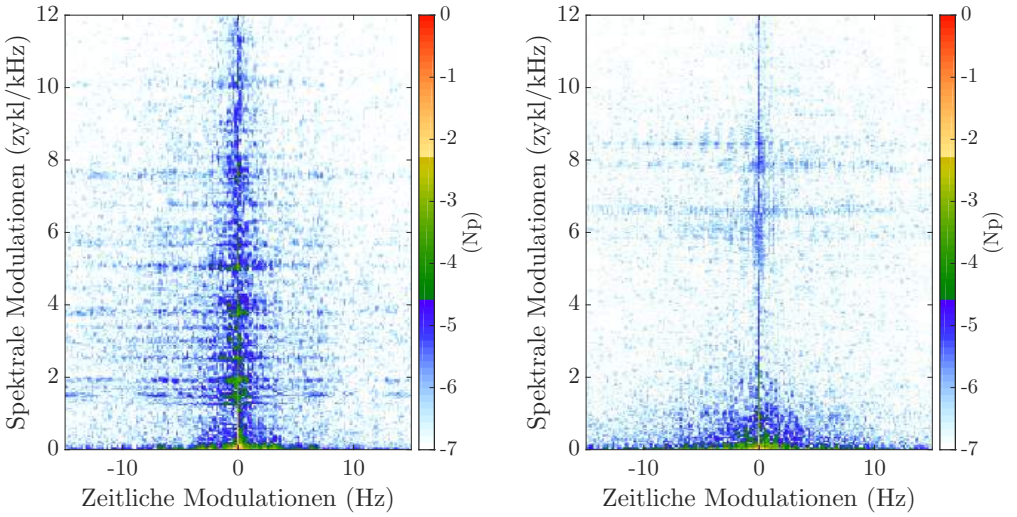
$$\mathbf{S_m}(t', \omega) = \mathcal{F}_{\text{-1,-1}}\big[\hat{\mathbf{S}}_1 \circ \hat{\mathbf{F}}^{5,5}_{\text{hp}} + \hat{\mathbf{S}}_2 \circ \hat{\mathbf{F}}^{5,5}_{\text{lp}}\big]. \tag{4.2.4}$$

Der erste Klang wird mit einer Hochpass-Maske gefiltert und zum tiefpassgefilterten zweiten Klang addiert. Nach anschließender Fourierrücktransformation ergibt sich das neue Spektrogramm. Als Beispiel wird der Hochpassbereich des Haydn-Quartetts mit dem Tiefpassbereich des männlichen Sprechers zusammengeführt (Abb. 4.15). Als Grenzfrequenzen werden $\tau_g = 5$ zykl/kHz und $f_g = 5$ Hz gesetzt. Im resultierenden MPS (e) ist zu sehen, dass der Tiefpassbereich der Sprache mit dem Hochpassbereich des Haydn-Quartetts fusioniert wurde. Der gesamte Formantbereich und die tiefen spektralen Modulationen (hohe Grundtöne im Klang) wurden durch den Formantbereich des Sprechers ersetzt. Im resultierenden Klang ist die Sprache gut verständlich. Dies liegt daran, dass der Formantbereich der Sprache im resultierenden MPS nach wie vor enthalten ist. Die Tonhöhen der Sprache sind aber nicht wahrnehmbar. Vielmehr wurde der Teiltonbereich der Sprache mit jenem des Haydnquartetts zusammengeführt. Dies ist daran zu hören, dass, wenn der Sprecher spricht, in seinem Klang das Streichquartett erklingt.

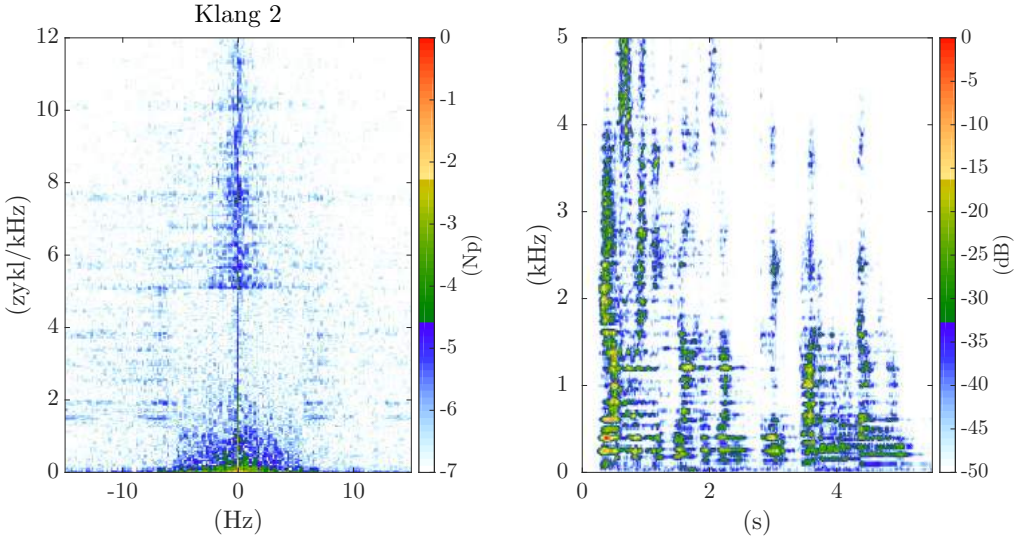
Klangbeispiel: `morph_haydn_male_s5_t5.wav`



(a) Spektrogramm 1 **(b)** Spektrogramm 2



(c) MPS 1 **(d)** MPS 2



(e) MPS Morph **(f)** Spek Morph

Abb. 4.15: Morphingprozess

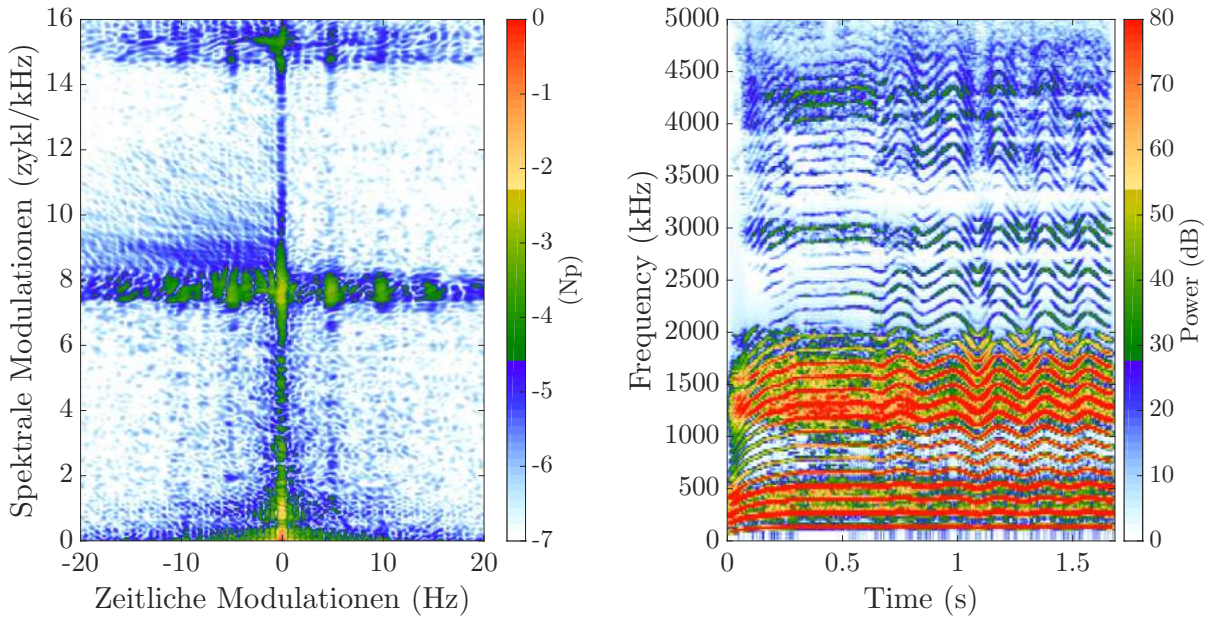
4.2.3 Verzerrung der Achsen

Es wurde untersucht, welche Auswirkung eine Verzerrung der zeitlichen bzw. der spektralen Modulationsachse mit sich zieht. Eine Umtastung der Quefrenzen entlang der spektralen Achse erfolgt über folgende Transformation $\tau_0 \rightarrow \tau_1$:

$$\tau_1 = \frac{\tau_0}{2^{d/12}} \tag{4.2.5}$$

wobei τ_0 die spektralen Quefrenzen bezeichnet und τ_1 die neuen Quefrenzen. Mit d wird der Verzerrungsfaktor eingeführt, der den Grad der Verzerrung angibt. Genau wie im Frequenzbereich ein Faktor $d = 12$ eine Verschiebung der Frequenzen um eine Oktave nach oben bedeutet, bewirkt er entlang der spektralen Achse eine Verschiebung der spektralen Modulationen um das Doppelte nach oben.
Im folgenden Beispiel (Abb. 4.16) ist die Auswirkung des Verzerrungsgrades 12 zu sehen. Im MPS auf der linken Seite ist zu sehen, dass sich bei einem Faktor $d = 12$ die Bereiche im MPS von 4 auf 8 zykl/kHz bewegen. Dies ergibt im Spektrogramm auf der rechten Seite eine „Verdichtung“ der Obertöne innerhalb eines kHz. Da die Anzahl der zykl/kHz der Anzahl an Obertönen im Spektrogramm innerhalb eines kHz entspricht, wird ersichtlich, dass sich ca. 7 Obertöne und der Grundton innerhalb dieses Bereichs befinden. Der Klang wurde somit nach unten transponiert und der Grundton befindet sich jetzt bei ca. 125 Hz.

Verzerrung der spektralen Modulationsachse mit $d = 12$



(a) MPS **(b)** Spektrogramm

Abb. 4.16: Darstellung MPS und dem daraus resultierenden Spektrogramm nach Verzerrung der spektralen Modulationsachse.

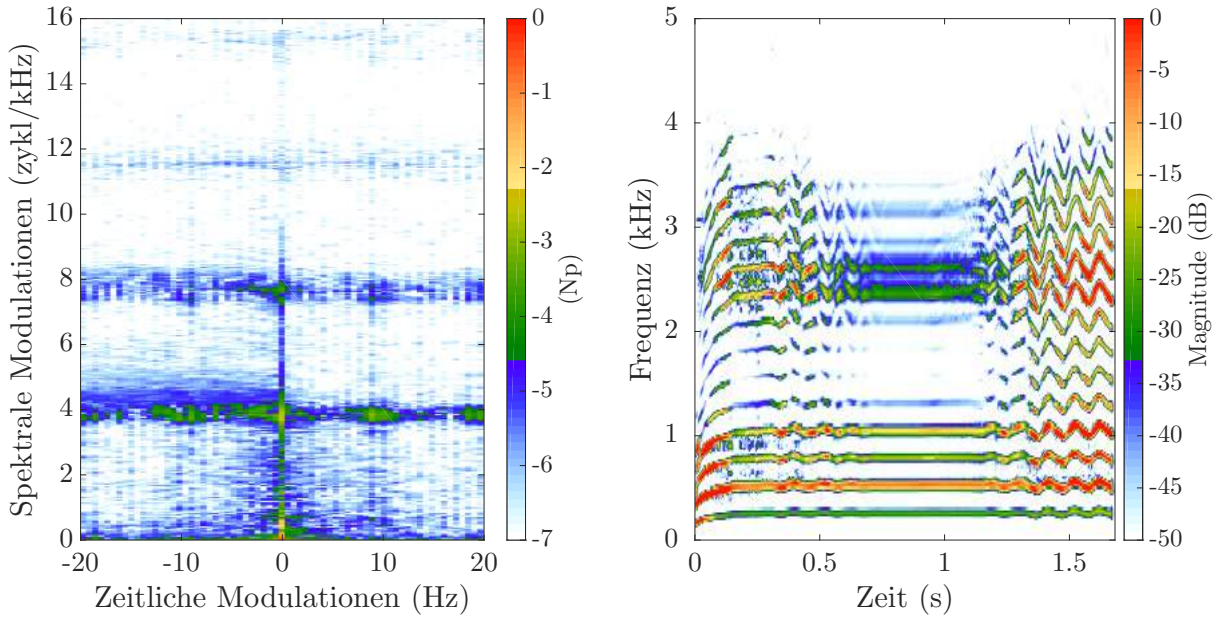
Klangbeispiel: `male_p_dist_s_d12_s.wav`

Verzerrung der zeitlichen Modulationsachse

Bei einer Verzerrung der zeitlichen Modulationsachse wird die selbe Transformation verwendet wie für die spektrale Achse, jedoch erfolgt nun eine Umtastung der zeitlichen Modulationen:

$$f_1 = \frac{f_0}{2^{d/12}} \tag{4.2.6}$$

Im MPS in Abb. 4.17 ist zu sehen, dass sich die Bereiche mit großer Modulationsenergie nach außen bewegen, was eine Veränderung der Modulationsfrequenz bedeutet. Im Spektrogramm wird ersichtlich, dass sich bei frühen und späten Zeiten des Klanges die zeitlichen Modulationsfrequenzen von 5 Hz zu 10 Hz verändert haben. Ein Verzerrungsfaktor von $d = 12$ entspricht auch hier einer Verdoppelung der Frequenz.



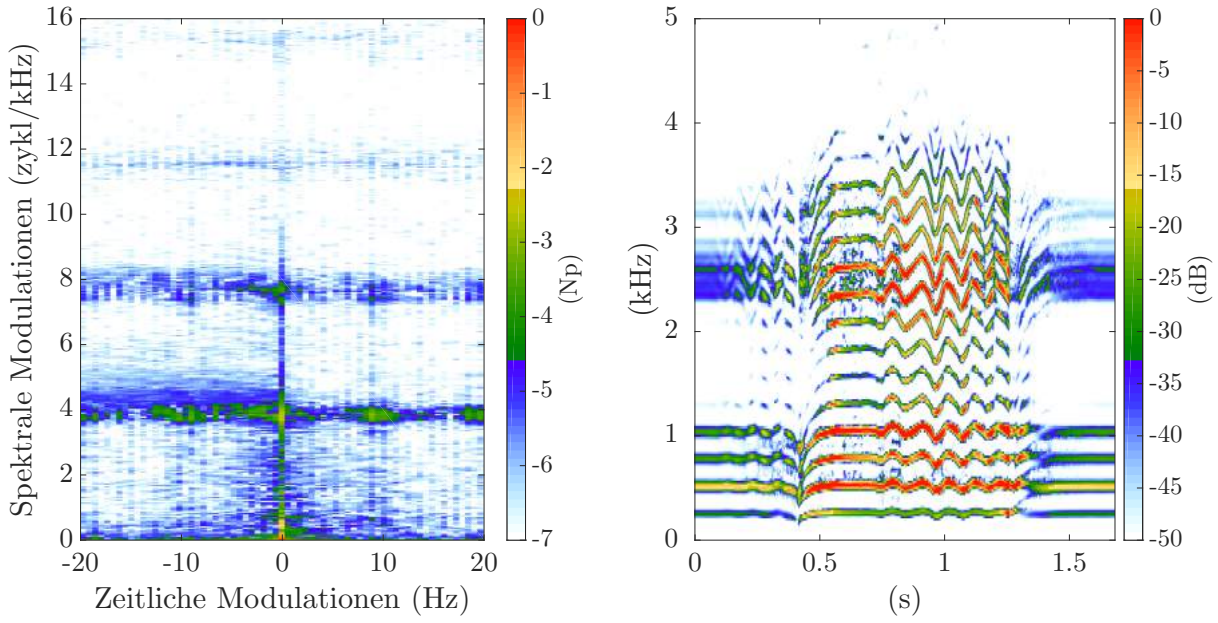
(a) MPS **(b)** Spektrogramm

Abb. 4.17: Darstellung MPS und dem daraus resultierenden Spektrogramm nach Verzerrung der zeitlichen Modulationsachse.

Klangbeispiel: `bariton_p_dist_z_d12_außen.wav`

Es besteht nun die Frage, warum die Signalenergie an die Randbereiche verschoben wird und wie es möglich ist, Bereiche zu zentrieren. Aufgrund der Linearität und damit der Kommutativität spielt es keine Rolle, ob zuerst eine Fouriertransformation entlang der vertikalen Achse (Cepstrogramm), oder eine Fouriertransformation entlang der zeitlichen Achse (zeitliches Modulationsspektrum) durchgeführt wird. Wird der Weg über das Cepstrogramm betrachtet, finden sich entlang jeder Zeile komplexe Werte für die entsprechenden spektralen Modulationen τ_i. Eine Fouriertransformation entlang jeder Zeile resultiert in der entsprechenden τ_i-Zeile im MPS. Wird das MPS nun entlang der f_{tmod}-Achse gedehnt, bedeutet das, dass die Phasenfortschreitung vom Ursprung bei $f_{tmod} = 0$ Hz hin zu höheren Frequenzen verlangsamt wird. Eine langsamere Phasenfortschreitung bedeutet in der Urdomäne (Cepstrogramm), dass die Energieanteile mehr Richtung Ursprung konzentriert werden. Da man sich das Spektrogramm als periodische Fortsetzung vorstellen muss, wird die Signalenergie nun an die Randbereiche, also an den Beginn und das Ende verschoben. Aufgrund der Konzentration der Signalenergie an den Rändern entsteht ein Loch in der Mitte. Um die Signalenergie in der Mitte zu konzentrieren, müssen die Randbereiche des

Cepstrogramms in die Mitte verschoben werden. Dies wird erreicht, wenn über die Funktion `fftshift` die Quadranten des Cepstrogramms getauscht werden und erst danach die Fouriertransformation entlang der Zeitachse berechnet wird. Dieser Prozess wird bei der Rücktransformation über die Funktion `ifftshift` wieder rückgängig gemacht. Dadurch wird die Signalenergie in der Mitte konzentriert (Abb. 4.18), da durch die Vertauschung der Quadranten der Signalbeginn in die Mitte verschoben wird.



(a) MPS **(b)** Spektrogramm

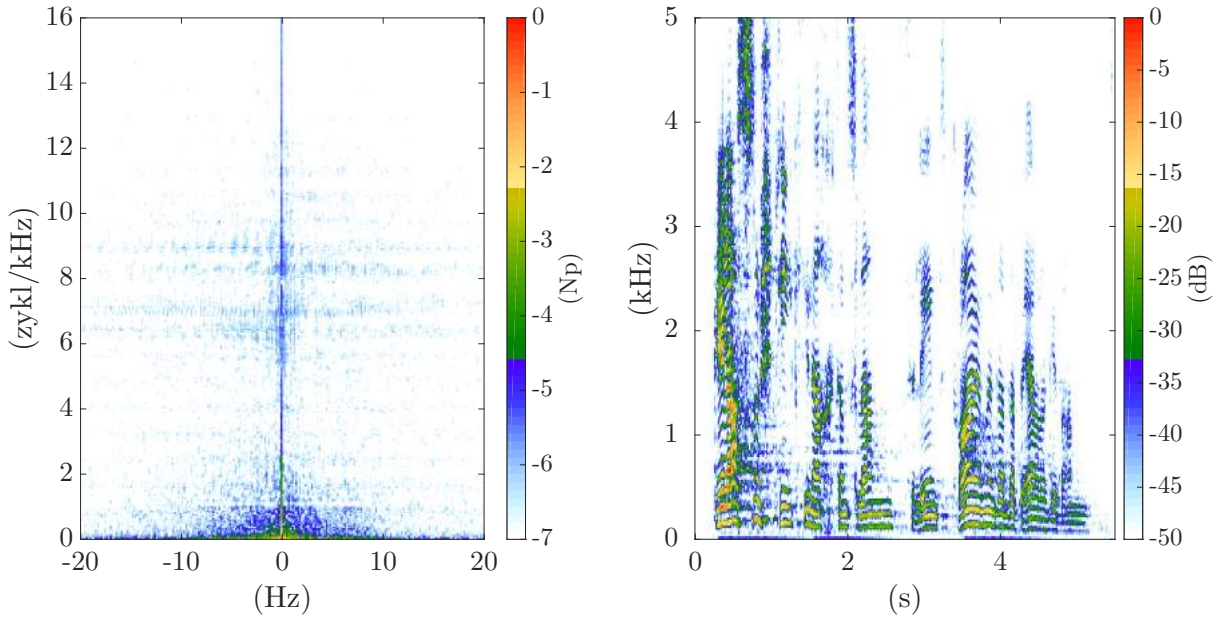
Abb. 4.18: Darstellung MPS und dem daraus resultierenden Spektrogramm nach Verzerrung der zeitlichen Modulationsachse. Im Gegensatz zu Abb. 4.17 ist zu sehen, dass sich der Bereich mit erhöhter zeitlicher Modulationsenergie nun in der Mitte konzentriert.

Klangbeispiel: `bariton_p_dist_z_d12_mitte.wav`

4.2.4 Verzerrung des Hochpassanteils

Eine weitere Möglichkeit der Klangmanipulation besteht darin, einen Klang in einen Hochpass- und in einen Tiefpassanteil zu teilen und nur den Hochpassanteil zum Beispiel entlang der spektralen Achse nach oben zu verzerren. Der Formantbereich wird hingegen nicht umgetastet, was bedeutet, dass die Formantbereiche erhalten bleiben. Ein Formant bei 0.5 zykl/kHz würde nach einer Umtastung um $d = 12$ nach oben einem Formanten entsprechen, der eine

Breite von einem zykl/kHz entspricht. Die Breite im Spektrum würde demnach von 2000 Hz zu 1000 Hz verkleinert. Diese Umtastung im Formantbereich wird unterlassen und es werden nur die höheren spektralen Modulationen umgetastet. Die Grundtöne des Klanges verschieben sich nach unten. Die Breiten der Formanten wurden jedoch nicht verändert. Dies ist in Abbildung 4.19 und 4.20 zu erkennen.



(a) MPS **(b)** Spektrogramm

Abb. 4.19: Darstellung MPS und dem daraus resultierenden Spektrogramm nach Verzerrung der spektralen Modulationsachse des Hochpassbereichs. Es ist zu erkennen, dass im Übergangsbereich bei $\tau_g = 1$ zykl/kHz ein weißer Streifen entstanden ist. Dies resultiert aufgrund der Umtastung des Hochpassbereichs. Im Spektrogramm (b) wurden die Teiltöne um einen Halbton ($d = 1$) nach unten verschoben. Die Formanten wurden jedoch nicht verändert.

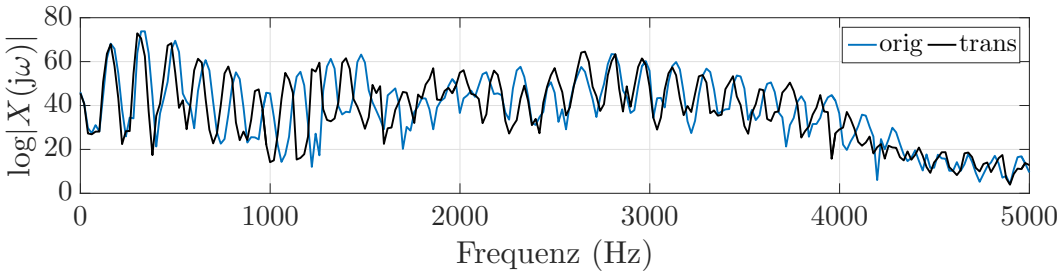


Abb. 4.20: Vergleich der Spektren bei Sekunde 0.36. Es wurden die spektralen Modulationen größer als 1 zykl/kHz um den Faktor $d = 1$ umgetastet. Der Faktor $d = 1$ bewirkt, dass die Grundtöne des Sprechers um einen Halbton nach unten transponiert werden. Dadurch, dass der Formantbereich nicht umgetastet wurde, haben sich die Formantbreiten nicht verändert. Im Gegensatz dazu würden bei einem Pitchshifter auch die Formantbreiten dementsprechend verändert.

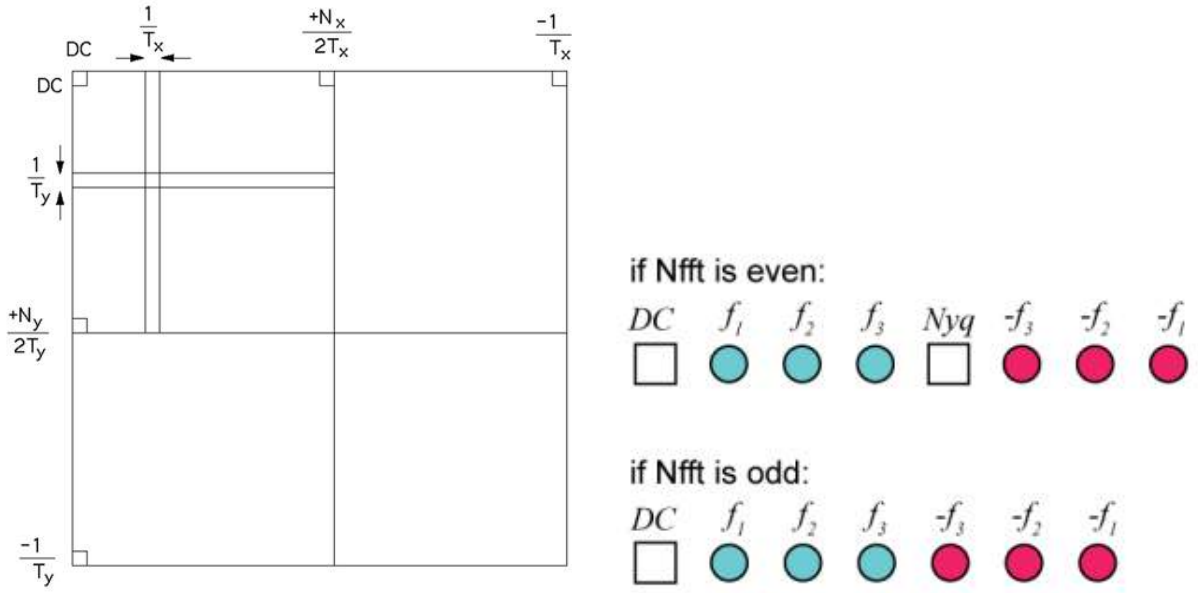
4.2.5 Spiegelung

Spiegelung des MPS um die τ-Achse

Zum Abschluss soll gezeigt werden, welche klanglichen Auswirkungen auftreten, wenn eine Spiegelung um eine der beiden Achsen mit sich zieht. Dies bedeutet, dass Bereiche, die sich in der rechten Hälfte des MPS befinden, also positive zeitliche Modulationen aufweisen, in den Bereich negativer zeitlicher Modulationen gespiegelt werden und vice versa. Dabei muss die Matrix, die aus der zweidimensionalen Fouriertransformation resultiert, betrachtet werden. Führt man in der Fouriertransformation die Variablentransformationen $\tau \rightarrow -\tau$ und $\omega \rightarrow -\omega$ durch, ergibt sich:

$$x(-\tau) \circ\!\!-\!\!\bullet\, X(-\mathrm{j}\omega) \tag{4.2.7}$$

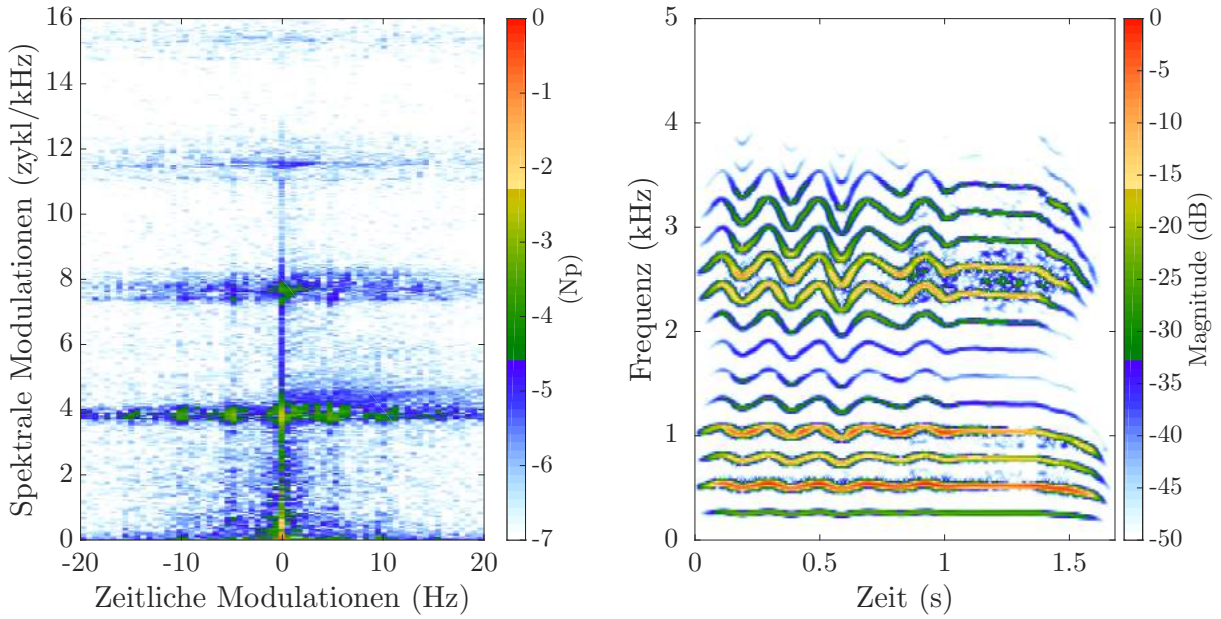
Dabei wird klar, dass eine Zeitspiegelung einer Frequenzspiegelung entspricht. In Abb. 4.21 ist zu sehen, dass sich der Gleichanteil an der Position (0,0) befindet. Das bedeutet, dass bei einer Spiegelung der Matrix um die τ-Achse die erste Spalte nicht mitgespiegelt werden darf, da sonst die kleinste negative Frequenz mit dem Gleichanteil getauscht wird und dies zu Unregelmäßigkeiten führt[6]. Nimmt man aber die erste Spalte nicht mit in die Spiegelung der Matrix, ergibt dies die Zeitumkehr des Signals.



(a) 2DFT **(b)** FFT

Abb. 4.21: Ergebnis einer zweidimensionalen Fouriertransformation (a). Der Gleichanteil befindet sich bei Position (0,0). Aufteilung der positven und negativen Frequenzen bei geraden und ungeraden FFT-Längen.

[6] Dieser Fehler wurde zuerst gemacht und resultiert in einer Auslöschung bei den mittleren Zeiten im Spektrogramm.



(a) MPS **(b)** Spektrogramm

Abb. 4.22: Darstellung MPS und dem daraus resultierenden Spektrogramm nach Spiegelung um die spektrale Modulationsachse.

Klangbeispiel: `bariton_p_mir_d2.wav`

Spiegelung des MPS um die f_{tmod}-Achse

Bei dieser Spiegelung ergibt sich noch kein sinnvolles Ergebnis.

Kapitel 5

Erweiterte Signalverarbeitung

Die MPS-Berechnung und die entsprechende Signalresynthese können als Spezialfall einer mehrstufigen Kette von Signaltransformationen angesehen werden. Dabei ist es üblich, zuerst die STFT und den Betrag des Signals zu berechnen, anschließend den Logarithmus zu bilden und bei der Resynthese die Inversion des Logarithmus, die Exponentialfunktion, anzuwenden. Ein verallgemeinertes Modell, welches optional einerseits auf die Betragsbildung vor der Logarithmierung verzichtet und andererseits anstatt der Logarithmierung funktionale Wurzeln von Logarithmus und Exponentiation verwendet, soll hier vorgestellt werden. In Kapitel 2 wurde beschrieben, wie anhand der Tetration mehrmals hintereinander ausgeführte Funktionen erzeugt werden können. Dabei ergibt sich die Logarithmierung für den Parameter $c = -1$ und die Exponentiation für den Parameter $c = 1$. Anhand des Parameters c ist es nun möglich, beliebig zwischen den Bereichen der Logarithmierung und der Exponentiation überzublenden. Zum Beispiel stellt sich für $c = 0$ der Fall $y = x$ ein, wobei die (komplexe[1]) Amplitude nicht bearbeitet wird.

[1] Falls nach der STFT kein Betrag gebildet wird.

5.1 Komplexe Signalverarbeitungskette

Wird nach der STFT auf die Betragsbildung verzichtet, liegen die Werte der STFT in komplexer Form vor. Dies ist in Abbildung 5.1 zu sehen.

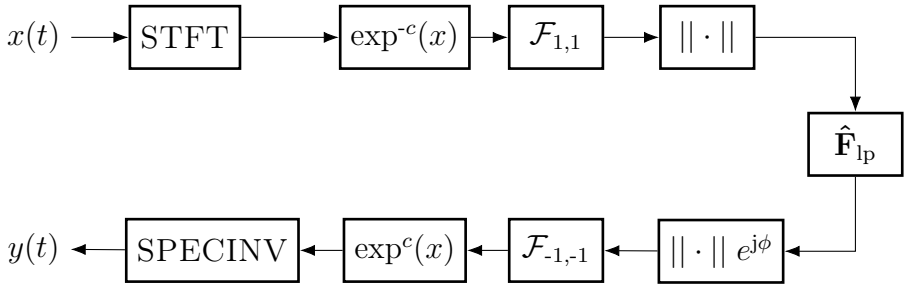


Abb. 5.1: Signalverarbeitungskette ohne Betragsbildung nach der STFT und natürliche Tetration mit einem beliebigen Faktor.

Bei $c = 1$ führt die Berechnung von $\exp^{-c}(x)$ zum komplexen Logarithmus:

$$\log z = \log|z| + \mathrm{j}\arg(z) \tag{5.1.1}$$

Ein Unterschied zur Berechnung mit Betragsbildung stellt der imaginäre Anteil mit der Phase der STFT zusätzlich zum Betrag der STFT dar. In der MPS-Domäne wird anschließend eine spektrale Tiefpassfilterung durchgeführt. Nach der Fourier-Rücktransformation und einer Umkehrung des Logarithmus anhand von $\exp^{1}(x)$ wird das resultierende Spektrogramm über die Spektrogrammінversion wieder in ein Zeitsignal übergeführt. Der imaginäre Anteil der Phase der STFT bringt in dieser Berechnung keinen klanglichen Mehrwert im Gegensatz zur Berechnung des Logarithmus ausschließlich vom Betrag der STFT. Ein Spezialfall stellt sich ein, wenn eine Hopsize von $R = 1$ und $c = 0$ verwendet wird. Eine Hopsize von 1 resultiert in einer STFT, bei der in jeder Spalte die Spektren der um ein Sample verschobenen Fenster des Signals aufgetragen sind. Wird zusätzlich auf die Betragsbildung verzichtet, liegen die Spektren in komplexer Form vor. Aufgrund von $c = 0$ werden die komplexen Amplituden nicht verändert. Nach anschließender zweidimensionaler Fourier-Transformation werden im MPS die Spektren der um 1 Sample verschobenen Fenster des Zeitsignals in jeder Zeile aufgetragen. Eine zeitliche Tiefpassfilterung des MPS ergibt eine Tiefpassfilterung jedes Spektrums des Eingangssignals und entspricht somit einer normalen Tiefpassfilterung und nicht der Tiefpassfilterung der Modulationen.

5.2 Signalverarbeitung anhand von funktionalen Wurzeln

Anhand des Faktors c ist es nun möglich, auch nicht ganzzahlige Werte für die Exponentialfunktion zu verwenden und damit anhand von funktionalen Wurzeln die Amplitude der STFT nach der Betragsbildung zu manipulieren. Eine genaue Untersuchung der Effekte bei Verwendung eines Faktors $c \neq \pm 1$ ist im Zuge dieser Arbeit aus Zeitgründen nicht mehr vorgesehen. Es wurden jedoch Tests durchgeführt, um die Effekte der Filterung in einer solchen Domäne zu untersuchen. Die Funktionen für den Wert $c = 0.5$ sind in Abb. 5.2 abgebildet.

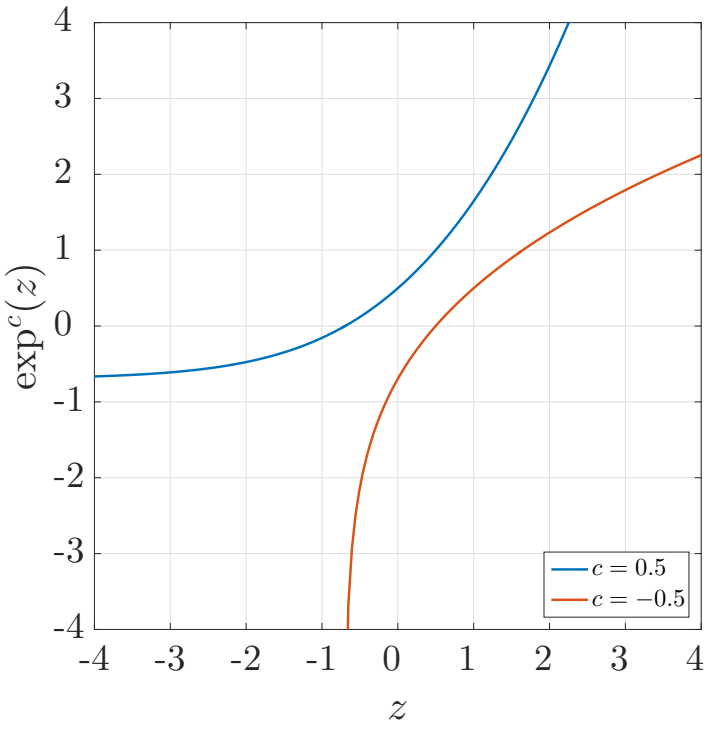


Abb. 5.2: Darstellung der natürlichen Exponentiation für $c = \pm 0.5$.

Dabei ist zu sehen, dass für größere Werte von z die Amplitude nicht in dem Ausmaß komprimiert wird wie bei der Logarithmierung. Die Signalverarbeitungskette für dieses Szenario ist in Abb. 5.3 abgebildet.

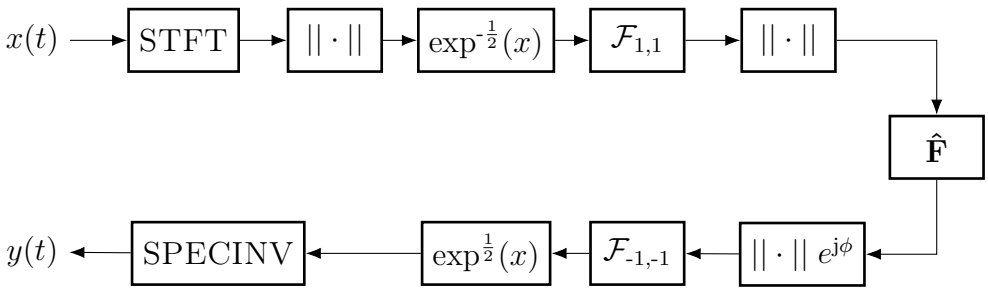


Abb. 5.3: Signalverarbeitungskette anhand der Tetration.

Ein Faktor $c = 0$ hat zur Folge, dass die Amplituden nach der Betragsbildung nicht komprimiert werden.

Kapitel 6

Zusammenfassung und Ausblick

In dieser Arbeit wurden unterschiedliche Darstellungsmöglichkeiten von Modulationen verschiedener Klänge diskutiert und präsentiert. Zeitliche Modulationen kennzeichnen sich hierbei dadurch aus, dass die zeitliche Hüllkurve eines Klanges moduliert wird. Im zeitlichen Modulationsspektrum ist ersichtlich, welchen Amplitudenschwankungen einzelne Frequenzen unterliegen, wohingegen die Periodizität der Teiltonstruktur anhand der Cepstralanalyse beschrieben werden kann. Eine Darstellung dieser Cepstralanalyse über die Zeit wird als Cepstrogramm bezeichnet und wurde in einzelnen Exempeln gezeigt. Zum Beispiel sind die Grundtonverläufe von einzelnen Klängen und Zusammenklängen erst im Cepstrogramm sichtbar.

Im Zuge der Masterarbeit wurde außerdem eine gemeinsame Darstellung der zeitlichen bzw. der spektralen Modulationen im Modulation Power Spectrum präsentiert. Dies resultiert aus einer zweidimensionalen Fouriertransformation eines Spektrogramms. Im dabei entstehenden MPS wird ersichtlich, dass der Formantbereich eines Klanges im Ursprung und der Teiltonbereich bei den entsprechenden spektralen Modulationen konzentriert wird. Anhand von Filterungen und anderen Manipulationen wurde gezeigt, wie zeitliche bzw. spektrale Modulationen beeinflusst werden können. Bei einer Filterung wird ein Klang beispielsweise in der MPS-Domäne, also dem Frequenzbereich, mit einer Filtermaske multipliziert. Einzelne Modulationsbereiche können so unterdrückt werden. Eine weitere Untersuchung ergab sich am Beispiel von Glissandi. Im MPS werden Glissandi nach oben oder nach unten getrennt voneinander dargestellt. Dadurch ist es zum Beispiel möglich, ein Vibrato zu manipulieren. Eine Möglichkeit, den Formantbereich eines Klanges mit der Teiltonstruktur eines anderen zu verbinden, wurde in der Arbeit anhand des Morphingprozesses gezeigt. Ei-

ne Umtastung der Modulationsachsen führt dabei entlang der zeitlichen Modulationsachse hin zu höheren zeitlichen Modulationen zu einer Verlangsamung der Phasen im MPS. Im Spektrogramm bedeutet das eine Konzentration der Energie in der Nähe des Ursprungs, also den Randbereichen des Spektrogramms. Eine Umtastung entlang der spektralen Modulationsachse hingegen bedeutet eine Transposition des Klanges nach oben oder nach unten. Wird ein Klang in einen Hochpass- und in einen Tiefpassanteil geteilt und anschließend der Hochpassanteil, also die Teiltonstruktur, umgetastet, hat dies zur Folge, dass nur die Teiltonstruktur nach oben oder unten bewegt wird. Der Formantbereich wird hingegen nicht verändert. Ein verallgemeinertes Modell einer solchen Transformation wurde in Kapitel 5 vorgestellt. Anhand dieses Modells wurde die Möglichkeit der Klangmanipulation mittels funktionalen Wurzeln erweitert und untersucht.

Weitere Ideen zur Klangmanipulation wurden bereits gedanklich skizziert und diskutiert: Bei Sprache beispielsweise ergibt sich als Formantbereich eine dreieckige Fläche im Bereich des Ursprungs. Es stellt sich daher die Frage, ob eine Filterung der Dreiecksfläche in Verbindung mit der Teiltonstruktur eines anderen Klanges eine Verbesserung im Gegensatz zur rechteckigen Filterung darstellt. Eine mögliche Filterung, die andere geometrische Figuren als Filter verwendet, könnte für zukünftige Manipulationen wichtig sein. Es soll weiters möglich sein, die Teiltonstruktur eines Klanges hin zu einem anderen Zeitpunkt des Klanges zu verschieben. Dadurch werden im MPS nur Bereiche der Teiltonstruktur umgetastet. Beim Morphingprozess wurde der Hochpassanteil eines Klanges mit dem Tiefpassanteil eines anderen Klanges kombiniert. Es könnte untersucht werden, welche Auswirkungen es mit sich bringt, wenn der Sperrbereich eines Filters nicht völlig abgesenkt wird. Diese Ideen könnten in zukünftigen Arbeiten noch untersucht werden, um weitere klangliche Auswirkungen mittels MPS erforschen zu können.

Anhang A

Fourier-Transformationen

A.1 STFT eines Gauß-Chirp

Es wird das folgende Integral gelöst

$$\mathbf{STFT}\{g_h(t)\}(t', \mathrm{j}\omega) = G_h(t', \mathrm{j}\omega) = \int_{-\infty}^{+\infty} g_h(t) w(t - t') e^{-\mathrm{j}\omega t} \mathrm{d}t$$

Dabei ist

$$g_h(t) = \sum_{n=1}^{N} g_n(t) = \sum_{n=1}^{N} e^{\mathrm{j}2\pi n(f_m + \frac{S_0}{2}t)t}.$$

$$
\begin{aligned}
G_h(t', \mathrm{j}\omega) &= \int_{-\infty}^{+\infty} \sum_{n=1}^{N} g_n(t) w(t-t') e^{-\mathrm{j}\omega t} \mathrm{d}t \\
&= \int_{-\infty}^{+\infty} \sum_{n=1}^{N} e^{\mathrm{j}2\pi n f_m t}\ e^{\mathrm{j}\pi n S_0 t^2} w(t-t') e^{-\mathrm{j}\omega t} \qquad \text{mit} \qquad t'' = t - t' \mathrm{d}t \\
&= \int_{-\infty}^{+\infty} \sum_{n=1}^{N} e^{\mathrm{j}2\pi n f_m (t''+t')}\ e^{\mathrm{j}\pi n S_0 t''^2}\ e^{\mathrm{j}\pi n S_0 2 t'' t'}\ e^{\mathrm{j}\pi n S_0 t'^2}\ e^{-\mathrm{j}\omega(t''+t')} w(t'') \mathrm{d}t'' \\
&= \sum_{n=1}^{N} e^{\mathrm{j}2\pi n f_m t'}\ e^{\mathrm{j}\pi n S_0 t'^2}\ e^{-\mathrm{j}\omega t'} \int_{-\infty}^{+\infty} e^{\mathrm{j}2\pi n f_m t''}\ e^{\mathrm{j}\pi n S_0 t''^2} e^{\mathrm{j}2\pi n S_0 t'' t'} e^{-\mathrm{j}\omega t''} w(t'') \mathrm{d}t'' \\
&= \sum_{n=1}^{N} e^{\mathrm{j}\pi n (2 f_m + S_0 t') t' - \mathrm{j}\omega t'} \int_{-\infty}^{+\infty} e^{-\mathrm{j}t''(-2\pi n f_m - 2\pi n S_0 t')} e^{\mathrm{j}S_0 \pi n t''^2} w(t'') e^{-\mathrm{j}\omega t''} \mathrm{d}t'' \\
&= \sum_{n=1}^{N} e^{\mathrm{j}\pi n (2 f_m + S_0 t') t' - \mathrm{j}\omega t'} \int_{-\infty}^{+\infty} e^{-\mathrm{j}t''(-2\pi n f_m - 2\pi n S_0 t')} \underbrace{e^{\mathrm{j}S_0 \pi n t''^2} w(t'')}_{Eq.3.4.4} e^{-\mathrm{j}\omega t''} \mathrm{d}t'' \\
&= \sum_{n=1}^{N} e^{\mathrm{j}\pi n (2 f_m + S_0 t') t' - \mathrm{j}\omega t'}\ G_{\mathrm{comp}}(\mathrm{j}\omega - 2\pi(f_m n + S_0 n t'))
\end{aligned}
\tag{A.1.1}
$$

A.2 Fourier-Transformation einer Parabel

Die logarithmierte Magnitude des Spektrums eines Gauß-Chirp zentriert um 0 Hz ist gegeben mit:

$$
G_{\mathrm{spec}}(\omega) = \log\left(\sqrt{\frac{\pi}{\sqrt{a^2+b^2}}}\right) \cdot \left[-\omega^2 \cdot \frac{a}{a^2+b^2} \right] \tag{A.2.1}
$$

Um das reelle Cepstrum zu erhalten, wird eine Fouriertransformation berechnet.

$$
G_p(\tau) = \mathcal{F}^{-1}[G_{\mathrm{spec}}(\omega)] = \frac{1}{2\pi} \int_{-\infty}^{+\infty} G_{\mathrm{spec}}(\omega)\, e^{\mathrm{j}\omega\tau}\ \mathrm{d}\omega \Big| = -c \frac{1}{2\pi} \underbrace{\int_{-\frac{\omega_b}{2}}^{+\frac{\omega_b}{2}} \omega^2 e^{\mathrm{j}\omega\tau}\ \mathrm{d}\omega}_{X_p} \tag{A.2.2}
$$

Mit der Konstante c:

$$
c = \log\left(\sqrt{\frac{\pi}{\sqrt{a^2+b^2}}}\right) \cdot \frac{a}{a^2+b^2}.
$$

Das folgende bestimmte Integral wird gelöst:

$$
\begin{aligned}
X_p &= \int_{-\frac{\omega_b}{2}}^{\frac{\omega_b}{2}} \underbrace{\omega^2}_{u} \cdot \underbrace{e^{-\mathrm{j}\omega t}}_{v'} \mathrm{d}\omega \qquad \text{mit partieller Integration} \qquad u \cdot v - \int u' \cdot v \\
&= \omega^2 \cdot \frac{e^{-\mathrm{j}\omega t}}{-\mathrm{j}t} \Big|_{-\frac{\omega_b}{2}}^{\frac{\omega_b}{2}} - \int_{-\frac{\omega_b}{2}}^{\frac{\omega_b}{2}} 2\omega \cdot \frac{e^{-\mathrm{j}\omega t}}{-\mathrm{j}t} \mathrm{d}\omega = -\omega^2 \cdot \frac{e^{\mathrm{j}\omega t}}{\mathrm{j}t} \Big|_{-\frac{\omega_b}{2}}^{\frac{\omega_b}{2}} - 2\omega \cdot \frac{e^{-\mathrm{j}\omega t}}{-t^2} \Big|_{-\frac{\omega_b}{2}}^{\frac{\omega_b}{2}} + \int_{-\frac{\omega_b}{2}}^{\frac{\omega_b}{2}} 2 \cdot \frac{e^{-\mathrm{j}\omega t}}{-t^2} \mathrm{d}\omega \\
&= -\omega^2 \cdot \frac{e^{-\mathrm{j}\omega t}}{\mathrm{j}t} \Big|_{-\frac{\omega_b}{2}}^{\frac{\omega_b}{2}} + 2\omega \cdot \frac{e^{-\mathrm{j}\omega t}}{t^2} \Big|_{-\frac{\omega_b}{2}}^{\frac{\omega_b}{2}} + 2 \cdot \frac{e^{-\mathrm{j}\omega t}}{\mathrm{j}t^3} \Big|_{-\frac{\omega_b}{2}}^{\frac{\omega_b}{2}} \\
&= -\left(\frac{\omega_b}{2}\right)^2 \frac{e^{-\mathrm{j}\frac{\omega_b}{2}t}}{\mathrm{j}t} + \left(\frac{\omega_b}{2}\right)^2 \frac{e^{+\mathrm{j}\frac{\omega_b}{2}t}}{\mathrm{j}t} + 2\left(\frac{\omega_b}{2}\right) \cdot \frac{e^{-\mathrm{j}\frac{\omega_b}{2}t}}{t^2} + 2\left(\frac{\omega_b}{2}\right) \cdot \frac{e^{+\mathrm{j}\frac{\omega_b}{2}t}}{t^2} + 2 \cdot \frac{e^{-\mathrm{j}\frac{\omega_b}{2}t}}{\mathrm{j}t^3} - 2 \cdot \frac{e^{+\mathrm{j}\frac{\omega_b}{2}t}}{\mathrm{j}t^3} \\
&= +{\omega_b}^2 \frac{\sin\left(\frac{\omega_b}{2}t\right)}{2t} + 2\omega_b \frac{\cos\left(\frac{\omega_b}{2}t\right)}{t^2} - 4\frac{\sin\left(\frac{\omega_b}{2}t\right)}{t^3}
\end{aligned}
$$

Eingesetzt in A.2.2 ergibt:

$$
G_p(\tau) = -c\frac{1}{2\pi}\left(-{\omega_b}^2 \frac{\sin\left(\frac{\omega_b}{2}\tau\right)}{2\tau} - 2\omega_b \frac{\cos\left(\frac{\omega_b}{2}\tau\right)}{\tau^2} + 4\frac{\sin\left(\frac{\omega_b}{2}\tau\right)}{\tau^3} \right) \tag{A.2.3}
$$

was hier abgebildet wird:

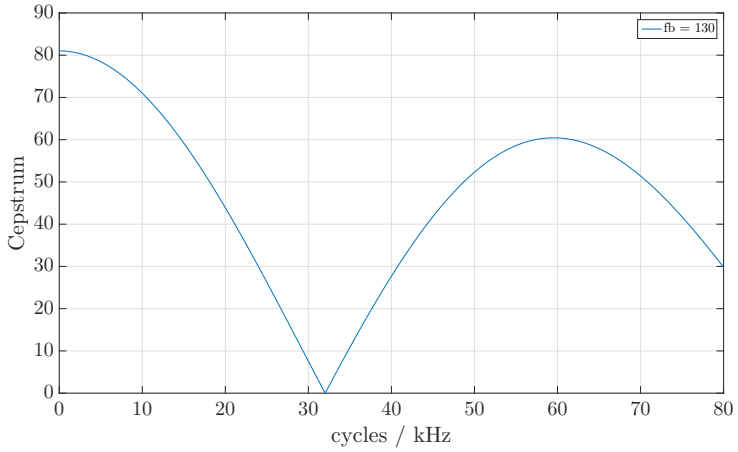


Abb. A.1: Betrag des Cepstrums der Funktion $c \cdot \omega^2$.

Anhang B

Matlab Dateien

Es sollen hier kurz die MATLAB-Dateien die für die Manipulation des MPS verwendet wurden, erläutert werden. Weiters können mit diesen MATLAB-Dateien auch alle Plots erzeugt werden, die für die Darstellung von Interesse sind. Die Dateien für die Manipulationen können einfach mit dem Dateinamen aufgerufen werden und sind keine Funktionen, damit auch alle Variablen eingesehen werden können. Die Variablen die bei einer Manipulation entstehen sind erstens in allen Dateien gleich benannt und werden zusätzlich für die Darstellungen verwendet. Die Variablen für die Darstellungen sind in Tabelle B.1 gelistet:

Tab. B.1: Variablenbezeichnungen und deren Bedeutung

Variblenname	Bezeichnung
`sabs`	Spektrogramm vor der Filterung
`amp_fabso`	MPS vor der Filterung
`ceps`	Cepstrogramm vor der Filterung
`modspeco`	Zeitl. Modulationsspektrum vor der Filterung
`newamp_fabsn`	MPS nach der Filterung
`new_sabsDisp`	Spektrogramm vor der Filterung
`gainmap`	Filtermaske

In der Datei `MPS_plots.m` können folgende Darstellungen aus den oben angeführten Variablen erstellt werden: Spektrogramm vor der Filterung, MPS vor der Filtertung, Cepstrogramm vor der Filtertung, zeitliches Modulationsspektrum vor der Filtertung, MPS nach der Filtertung, Spektrogramm nach der Filtertung. Weiters kann auch die Maske mit der das MPS im Frequenzbereich multipliziert wird, dargestellt werden. Im mitgelieferten Ordner befinden sich noch weitere wichtige Unterordner. Einerseits der Ordner `modfilter` in

dem alle wichtigen Funktionen für die Berechnung enthalten sind. Manche Funktionen entstammen dem `Theunissen Lab`. Weiters werden in dem Ordner `Audio_Output` alle generierten Audio-Dateien automatisch abgelegt. Im Ordner `Audio` befinden sich einige Audio-Dateien mit denen die Skripts getestet wurden.

Im Folgenden sollen kurz die MATLAB-Skripts und deren Eingangsparameter erläutert werden.

Tiefpass-Filterung: `MPS_tiefpass.m`

```
input = audioread('soundfile.wav');
samprate = 44100;                                % Samplerate
tit = 'title';                                   % Name fuer Audioausgabe
a = 1;                                           % Oversamplingfaktor
d = 100;                                         % Stille zu Beginn und am Ende in ms
winlen = 30;                                     % Fensterlaenge in ms
winmode = 3;                                     % Fennstertyp (1->gauss, 2->hamm, 3->hann)
overlap = 4;                                     % Overlapfaktor
increment = winlen/overlap/1000*samprate;        % Vierfache Ueberlappung
wf_high = 5;                                     % spektrale Grenzfrequenz in zykl/kHz
wt_high = 5;                                     % zeitliche Grenzfrequenz in Hz
phase_comp = 1;                                  % Spektrogramminversion mit Phase (1), ohne (0)
path = 'Audio_Output/tiefpass';                  % Pfad fuer Audioausgabe
```

Hochpass-Filterung: `MPS_hochpass.m`

```
input = audioread('soundfile.wav');
samprate = 44100;                                % Samplerate
tit = 'title';                                   % Name fuer Audioausgabe
a = 1;                                           % Oversamplingfaktor
d = 100;                                         % Stille zu Beginn und am Ende in ms
winlen = 30;                                     % Fensterlaenge in ms
winmode = 3;                                     % Fennstertyp (1->gauss, 2->hamm, 3->hann)
overlap = 4;                                     % Overlapfaktor
increment = winlen/overlap/1000*samprate;        % Vierfache Ueberlappung
wf_high = 5;                                     % spektrale Grenzfrequenz in zykl/kHz
wt_high = 5;                                     % zeitliche Grenzfrequenz in Hz
phase_comp = 1;                                  % Spektrogramminversion mit Phase (1), ohne (0)
path = 'Audio_Output/hochpass';                  % Pfad fuer Audioausgabe
```

Notch-Filterung: MPS_notch.m

```
input = audioread('soundfile.wav');
samprate = 44100;                                   % Samplerate
tit = 'title';                                      % Name fuer Audioausgabe
a = 1;                                              % Oversamplingfaktor
d = 100;                                            % Stille zu Beginn und am Ende in ms
winlen = 30;                                        % Fensterlaenge in ms
winmode = 3;                                        % Fennstertyp (1->gauss, 2->hamm, 3->hann)
overlap = 4;                                        % Overlapfaktor
increment = winlen/overlap/1000*samprate;           % Vierfache Ueberlappung
wf_high = 0.006;                                    % untere spektrale Grenzfrequenz
wt_high = 1;                                        % untere zeitliche Grenzfrequenz
wf_it = 0.004;                                      % obere s. Grenzfrequenz = wf_high + wf_it
wt_it = 10;                                         % obere z. Grenzfrequenz = wt_high + wt_it
phase_comp = 1;                                     % Spektrogramminversion mit Phase (1), ohne (0)
path = 'Audio_Output/notch';                        % Pfad fuer Audioausgabe
```

Anheben und Absenken: MPS_anab.m

```
input = audioread('soundfile.wav');
samprate = 44100;                                   % Samplerate
tit = 'title';                                      % Name fuer Audioausgabe
a = 1;                                              % Oversamplingfaktor
d = 100;                                            % Stille zu Beginn und am Ende in ms
winlen = 30;                                        % Fensterlaenge in ms
winmode = 3;                                        % Fennstertyp (1->gauss, 2->hamm, 3->hann)
overlap = 4;                                        % Overlapfaktor
increment = winlen/overlap/1000*samprate;           % Vierfache Ueberlappung
mirror = 1;                                         % Spiegelung ein
sx2 = 2;                                            % Standardabweichung zeitlich
sy2 = 2;                                            % Standardabweichung zeitlich
Cxy = 0;                                            % Kovarianz
mx = 5;                                             % Mittelwert zeitlich
my = 4;                                             % Mittelwert spektral
beta = 5;                                           % Anhebungs- oder Absenkungsfaktor
phase_comp = 1;                                     % Spektrogramminversion mit Phase (1), ohne (0)
path = 'Audio_Output/anab';                         % Pfad fuer Audioausgabe
```

Morphing: MPS_morph.m

```
input1 = audioread('soundfile1.wav');
input2 = audioread('soundfile2.wav');
samprate = 44100;                                   % Samplerate
tit = 'title';                                      % Name fuer Audioausgabe
a = 1;                                              % Oversamplingfaktor
d = 100;                                            % Stille zu Beginn und am Ende in ms
winlen = 30;                                        % Fensterlaenge in ms
winmode = 3;                                        % Fennstertyp (1->gauss, 2->hamm, 3->hann)
overlap = 4;                                        % Overlapfaktor
increment = winlen/overlap/1000*samprate;           % Vierfache Ueberlappung
spek_grenze = 5;                                    % spektrale Grenzfrequenz in zykl/kHz
zeit_grenze = 5;                                    % zeitliche Grenzfrequenz in Hz
phase_comp = 1;                                     % Spektrogramminversion mit Phase (1), ohne (0)
path = 'Audio_Output/morph';                        % Pfad fuer Audioausgabe
```

Umtastung: MPS_umtastung.m

```
input = audioread('soundfile.wav');
samprate = 44100;                                   % Samplerate
tit = 'title';                                      % Name fuer Audioausgabe
a = 1;                                              % Oversamplingfaktor
d = 100;                                            % Stille zu Beginn und am Ende in ms
winlen = 30;                                        % Fensterlaenge in ms
winmode = 3;                                        % Fennstertyp (1->gauss, 2->hamm, 3->hann)
overlap = 4;                                        % Overlapfaktor
increment = winlen/overlap/1000*samprate;           % Vierfache Ueberlappung
w = 12;                                             % Umtastungsfaktor
w1 = 0;                                             % Umtastung der Frequenzachse
w2 = 1;                                             % Umtastung der Zeitachse
phase_comp = 1;                                     % Spektrogramminversion mit Phase (1), ohne (0)
path = 'Audio_Output/umtastung';                    % Pfad fuer Audioausgabe
```

Umtastung Hochpass: MPS_umhoch.m

```
input1 = audioread('soundfile1.wav');
input2 = audioread('soundfile1.wav');
samprate = 44100;                                   % Samplerate
tit = 'title';                                      % Name fuer Audioausgabe
a = 1;                                              % Oversamplingfaktor
d = 100;                                            % Stille zu Beginn und am Ende in ms
winlen = 30;                                        % Fensterlaenge in ms
winmode = 3;                                        % Fennstertyp (1->gauss, 2->hamm, 3->hann)
overlap = 4;                                        % Overlapfaktor
increment = winlen/overlap/1000*samprate;           % Vierfache Ueberlappung
spek_grenze = 5;                                    % spektrale Grenzfrequenz in zykl/kHz
zeit_grenze = 5;                                    % zeitliche Grenzfrequenz in Hz
w = 1;                                              % Umtastungsfaktor
phase_comp = 1;                                     % Spektrogramminversion mit Phase (1), ohne (0)
path = 'Audio_Output/umhoch';                       % Pfad fuer Audioausgabe
```

Spiegelung: MPS_spiegel.m

```
input1 = audioread('soundfile1.wav');
input2 = audioread('soundfile2.wav');
samprate = 44100;                                   % Samplerate
tit = 'title';                                      % Name fuer Audioausgabe
a = 1;                                              % Oversamplingfaktor
d = 100;                                            % Stille zu Beginn und am Ende in ms
winlen = 30;                                        % Fensterlaenge in ms
winmode = 3;                                        % Fennstertyp (1->gauss, 2->hamm, 3->hann)
overlap = 4;                                        % Overlapfaktor
increment = winlen/overlap/1000*samprate;           % Vierfache Ueberlappung
w1 = 1;                                             % Spiegelung um die tau-Achse
phase_comp = 1;                                     % Spektrogramminversion mit Phase (1), ohne (0)
path = 'Audio_Output/spiegelung';                   % Pfad fuer Audioausgabe
```

Tetration: MPS_tetration.m

Bei der Signalverarbeitung anhand der Tetration ergeben sich mehrere Möglichkeiten. Einerseits kann der Faktor c beliebig gewählt werden. Andererseits ist es möglich, nach der STFT-Berechnung nicht den Betrag zu berechnen. Es ergibt sich somit eine komplexe STFT. Anhand der zweidimensionalen Fouriertransformation wird das komplexe Cepstrum

entlang der spektralen Modulationsachse berechnet. Bei der Berechnung anhand des komplexen Logarithmus nach Gl. 5.1.1 muss aber die Phase beim imaginären Anteil abgewickelt werden. Dies wurde ebenfalls implementiert und das Programm berechnet automatisch das reelle oder das komplexe Cepstrum. Zusätzlich sind in dieser Datei mehrere Filter in der MPS-Domäne anwendbar. Anders als bei den anderen Dateien, wo der Filter schon vorgegeben war, kann die Filtercharakteristik hier beliebig eingestellt werden.

```
input = audioread('soundfile.wav');
samprate = 44100;                                % Samplerate
tit = 'title';                                   % Name fuer Audioausgabe
a = 1;                                           % Oversamplingfaktor
d = 100;                                         % Stille zu Beginn und am Ende in ms
winlen = 30;                                     % Fensterlaenge in ms
winmode = 3;                                     % Fennstertyp (1->gauss, 2->hamm, 3->hann)
overlap = 4;                                     % Overlapfaktor
increment = winlen/overlap/1000*samprate;        % Vierfache Ueberlappung
c = 1;                                           % Faktor fuer Exponentialfunktion
betr = 0;                                        % Betrag berechnen ja(1) nein(0)
f = @(c,z) tet(c + ate(z));                      % Funktion fuer iterated Exponential
method = 5;                                      % Methodenauswahl
phase_comp = 1;                                  % Spektrogramminversion mit Phase (1), ohne (0)
path = 'Audio_Output/tetration';                 % Pfad fuer Audioausgabe
```

Literaturverzeichnis

[1] Alois Sontacchi. *Entwicklung eines Modulkonzeptes für die psychoakustische Geräuschanalyse unter MatLab* . Diplomarbeit - IEM Graz, 1998.

[2] E. Terhardt. *On the Preception of Periodic Sound Fluctuation (Roughness)*. Acoustica 30, pp. 201-213, 1974.

[3] Juan G. Roederer. *Physikalische und psychoakustische Grundlagen der Musik*. Springer-Verlag, London, 1977.

[4] E. Zwicker and H. Fastl. *Psychoacoustics - Facts and Models*. Springer-Verlag, London, 2007.

[5] Klaus Genuit. *Sound Engineering im Automobilbereich* . Springer-Verlag, London, 2010.

[6] Taffeta M. Elliott and Frédéric E. Theunissen. *The Modulation Transfer Function for Speech Intelligibility*. PLoS Comput Biol 5(3): e1000302. doi:10.1371 journal.pcbi.1000302, 2009.

[7] Douglas C. Giancoli. *Physik – Lehr- und Übungsbuch*. Pearson Education Deutschland GmbH, München, 3. Auflage edition, 2010.

[8] Stefan Weinzierl. *Handbuch der Audiotechnik*. Springer, 2008.

[9] G.S. Moschytz and M. Hofbauer. *Adaptive Filter:*. Springer Berlin Heidelberg, 2000.

[10] D. Gabor. *Theory of communication*. Institution of Electrical Engineering, 1946.

[11] H.W. Schüßler. *Netzwerke, Signale und Systeme: Band 2 Theorie kontinuierlicher und diskreter Signale und Systeme*. Springer Berlin Heidelberg, 1991.

[12] J. E. Wilhjelm. Bandwidth expressions of gaussian weighted chirp. *Electronics Letters*, 25(5), 1993.

[13] Julius O. Smith. *Spectral Audio Signal Processing*. 2011. W3K Verlag.

[14] A. V. Oppenheim and R.W. Schafer. From frequency to quefrency: A history of the cepstrum. *IEEE Signal Processing Magazine*, 21(5), 2004.

[15] B.P. Bogert, Healy M.J.R., and J.W. Tukey. *The Quefrency Alanysis of Time series for Echoes: Cepstrum, Pseudo-Autocovariance, Cross-Cepstrum, and Saphe Cracking*. Proc. of the Symp. on Time Series Analysis, by M. Rosenblatt (Ed.), Wiley, NY, pp. 209-243, 1963.

[16] Alan V. Oppenheim, Ronald W. Schafer, and John R. Buck. *Zeitdiskrete Signalverarbeitung. Mit 112 Beispielen und 403 Aufgaben*. Pearson Studium, 2004.

[17] Bernd Jaehne. *Digitale Bildverarbeitung*. Springer, 1991.

[18] Daniel W. Griffin, Jae, S. Lim, and Senior Member. Signal estimation from modified short-time fourier transform. *IEEE Trans. Acoustics, Speech and Sig. Proc*, pages 236–243, 1984.

[19] Ackermann W. *Zum Hilbertschen Aufbau der reellen Zahlen*. Mathematische Annalen 99, 118-133. MR1512441, 1928.

[20] D. Kouznetsov. *Tetration as special Function*. Institute for Laser Science, University of Electro Communications, researcher 1-5-1 Chofugaoka, Chofushi, Tokyo, 182-8585, Japan, 2010.

[21] D. Kouznetsov. *Solution of F(z+1) = exp(F(z)) in Complex z-Plane*. MATHEMATICS OF COMPUTATION S 0025-5718(09)02188-7 Article electronically published on January 6, 2009, 2009.

[22] fsexp.cin is routine for the fast evaluation of natural tetration. URL `http://mizugadro.mydns.jp/t/index.php/Fsexp.cin`. Abgerufen am 11.12.2016.

[23] fslog.cin is routine for the fast evaluation of natural arctetration. URL `http://mizugadro.mydns.jp/t/index.php/Fslog.cin`. Abgerufen am 11.12.2016.

[24] Masashi Unoki. *Study on important role of temporal amplitude-modulation feature of speech for auditory perception.* School of Information Science, Japan Advanced Institute of Science and Technology, 2016.

[25] Nandini C. Singh and Frédéric E. Theunissen. *Modulation spectra of natural sounds and ethological theories of auditory processing.* Department of Psychology and Neuroscience Institute, University of California, Berkeley, 3210 Tolman Hall, Berkeley, California 94720-1650, 2003.

[26] Rafael C. Gonzalez and Richard E. Woods. *Digital Image Processing (3rd Edition).* Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 2006.

[27] Jae S. Lim. *Two-dimensional Signal and Image Processing.* Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1990.

[28] Brian C. J. Moore. *An Introduction to the Psychology of Hearing.* Academic Press, 1997.

[29] Stanley A. Gelfand. Hearing - An Introduction to Psychological and Physiological Acoustics. 1981.

[30] Taffeta M. Elliott, Liberty S. Hamilton, and Frédéric E. Theunissen. *Acoustic structure of the five perseptual dimensions of timbre in orchestral instrument tones.* Helen Wills Neuroscience Institute, University of California, Berkeley, California 94720, 2012.

[31] W. A. Sethares. *Tuning, Timbre, Spectrum, Scale.* Springer-Verlag, London, 2005.

[32] E. Zwicker. *Die Grenzen der Hörbarkeit der Amplitudenmodulation und der Frequenzmodulation eines Tones.* Mitteilung aus dem Institut fur Nachrichtentechnik der Technischen Hochschule Stuttgart, 1952.

[33] Leon Cohen. *Time-frequency Analysis: Theory and Applications.* Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1995.

[34] Yale E. Cohen, Frédéric Theunissen, Brian E. Russ, and Patrick Gill. *Acoustic Features of Rhesus Vocalizations and Their Representation in the Ventrolateral Prefrontal Cortex.* Journal of Neurophysiology Published 1 February 2007 Vol. 97 no. 2, 1470-1484 DOI: 10.1152/jn.00769.2006, 2007.

[35] R. Decorsiere and P. L. Søndergaard and E. N. MacDonald and T. Dau. Inversion of Auditory Spectrograms, Traditional Spectrograms, and Other Envelope Representations. 23(1):46–56, January 2015.

[36] Izrail Solomonovich Gradshteyn, Iosif Moiseevich Ryzhik, Alan Jeffrey, Daniel Zwillinger, and Inc. Scripta Technica. *Table of integrals, series, and products*. Academic Press, New York, USA, 1965.

[37] Adalbert Prechtl. Signale und Systeme 1. Vorlesungsskript - TU Wien, 2010.

[38] Milton Abramowitz and Irene A. Stegun. *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. Dover, New York, 1964.