

Entwicklung und Implementierung eines einkanaligen,
adaptiven akustischen Rückkopplungsunterdrückers auf
Notchfilterbasis für den Musik-Liveeinsatz

Christian Göttlinger

21. Januar 2008

Zusammenfassung

Entwicklung und Implementierung eines einkanaligen, adaptiven akustischen Rückkopplungsunterdrückers für den Musik-Liveeinsatz.

Neben einer umfassenden Literaturrecherche im Vorfeld erfolgt eine Evaluierung und Problemverifizierung mit eigenen Simulationen und Messungen. Das Vorhandensein einer Rückkopplung wird kontinuierlich ohne hörbare Messsignale bestimmt und soll auch in Anwesenheit eines musikalischen Eingangssignals einsetzbar sein. Zielsetzung ist die zuverlässige Unterdrückung bei geringer objektiver Signaldegradierung. Zur Beurteilung wird die „Bark Spectral Distortion“ (BSD) herangezogen. Da das System auf die zeitvariante Übertragungsfunktion reagieren können muss, ist eine adaptive Steuerung der Systemeigenschaften erforderlich.

Die erste Algorithmenentwicklung erfolgt in MATLABTM. Die Evaluierung des zeitlichen Konvergenzverhaltens unter simulierten Realbedingungen erfolgt mithilfe von PureData (PD).

Development and implementation of an adaptive, single-channel acoustic feedback-suppressor for live-music applications.

An extensive literature research is followed by evaluations and a problem verification with simulations and measurements. The presence of feedback will be evaluated continuously without using additional measurement signals. A reliable suppression with low objective signal degradation is the aim of this project. The signal degradation will be evaluated by the "Bark Spectral Distortion" (BSD)-measure. An adaptive control of the system is required because of the time-variant characteristic of the transfer function. The initial development of the algorithms is carried out in MATLABTM. The convergence behavior under simulated real-world conditions will be evaluated in Pure Data (PD).

Inhaltsverzeichnis

1	Einleitung	1
2	Prinzipieller Systemaufbau	2
2.1	Das Feedback-Unterdrückungssystem	2
2.2	Das Evaluierungssystem	2
2.2.1	Evaluierung der Detektionseigenschaften(-features)	2
2.2.2	Evaluierung des Echtzeitverhaltens	2
3	Durchführung: Evaluierung der Detektionsmerkmale	5
3.1	Die Features im Überblick	5
3.1.1	Peakhöhe	6
3.1.2	Zeitliches Amplitudenderivat	6
3.1.3	Phasenderivat zweiter Ordnung $\frac{d^2\Phi}{dfdt}$	7
3.1.4	Phasenderivat zweiter Ordnung $\frac{d^2\Phi}{dt^2}$	8
3.2	Phasenderivate im Detail	8
3.2.1	Algorithmus der „Finiten Differenzapproximation“	8
3.2.2	Spekulativer Unwrap-Algorithmus	9
3.3	Statistische Ergebnisse	10
4	Optimierung des Gesamtsystems	11
4.1	Qualitätsmaße	11
4.1.1	Übersicht über mögliche Maße	11
A	Rekursive Fensterung	14
B	Oft verwendete Bezeichner	15

Kapitel 1

Einleitung

Kapitel 2

Prinzipieller Systemaufbau

2.1 Das Feedback-Unterdrückungssystem

2.2 Das Evaluierungssystem

2.2.1 Evaluierung der Detektionseigenschaften(-features)

2.2.2 Evaluierung des Echtzeitverhaltens

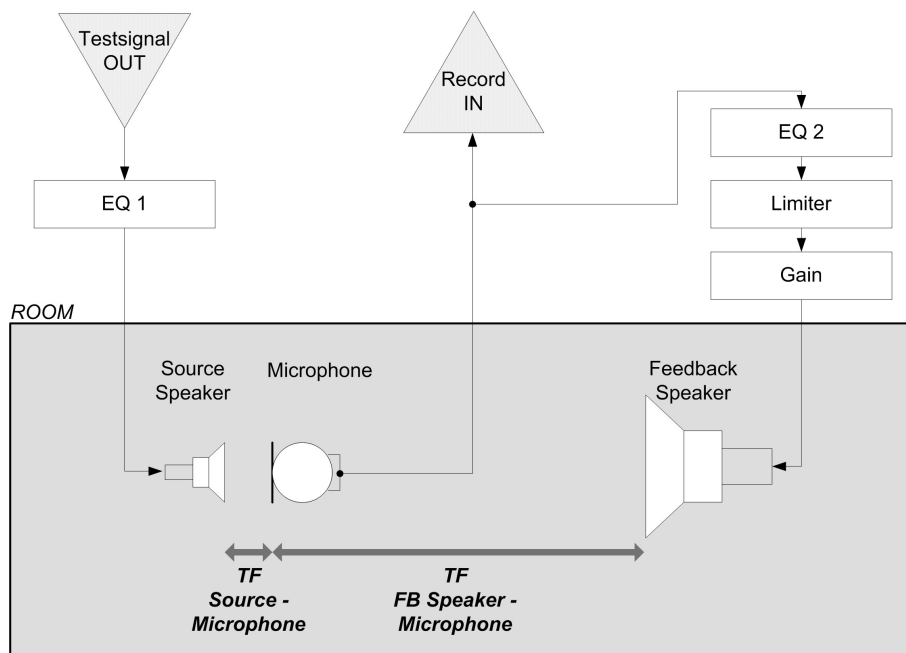


Abbildung 2.1: Das Messsystem insgesamt

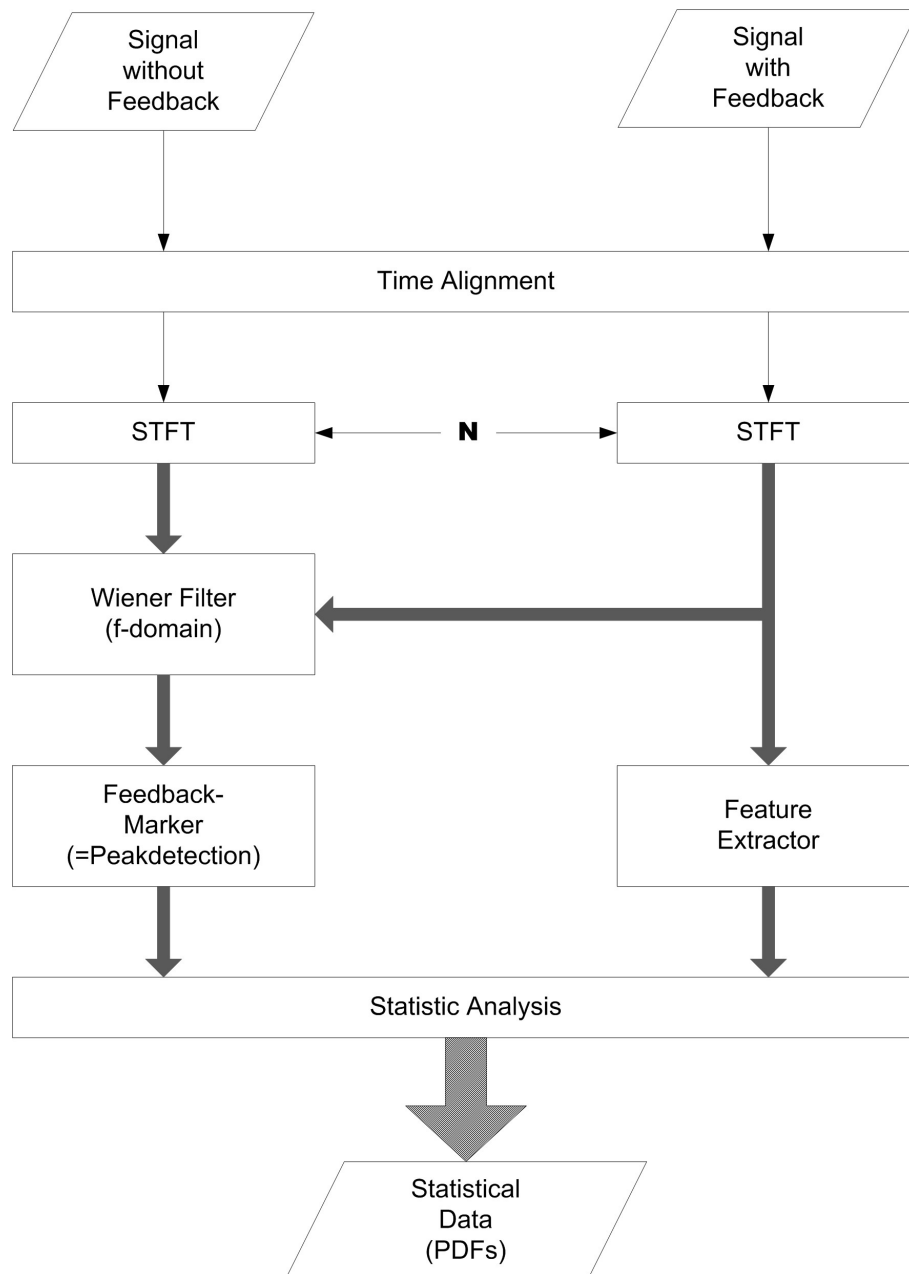


Abbildung 2.2: Das FeatureTester

Kapitel 3

Durchführung: Evaluierung der Detektionsmerkmale

3.1 Die Features im Überblick

Ein Feature ist ein hervorstechendes Merkmal oder Charakteristikum eines Vorgangs. Der Vorgang ist nun die Rückkopplung, beziehungsweise deren Vorboten, das längere Nachschwingen einzelner Frequenzen. Dieser kommt sowohl im Zeit- wie auch im Frequenzbereich nur lokal vor: Das Feedback kann auftreten oder auch nicht (je nach Positionsveränderung des Mikrofons) und tritt schmalbandig bei einigen Frequenzen auf. Um eine sichere Unterscheidung von vorhandenem Nutzsignal und auch eine möglichst wenig invasive Beseitigung zu gewährleisten empfiehlt sich also die Detektion im Zeit-/Frequenzbereich. Es wird also eine STFT durchgeführt. Mit der STFT in Rohform lässt sich zuerst einmal nicht viel anfangen. Sie muss geeignet weiter transformiert werden. Bei gelungener Verarbeitung sind die Wertebereiche Feedback-kein Feedback möglichst gut getrennt. Es lässt sich dann ziemlich sicher beurteilen, ob in diesem Bin eine Rückkopplung stattfindet oder nicht.

Diese Verarbeitungsschritte werden aus physikalischen Beobachtungen entwickelt. Ein einfaches Beispiel wäre die Feststellung, dass im Falle einer starken Rückkopplung eine einzige Frequenz eine starke Überhöhung aufweist. Mittels eines sinnvoll gesetzten Thresholds auf die absolute Amplitude ließe sich dies im einfachsten Fall detektieren.

Aufgabe ist es auch geeignete Verfahren zu evaluieren. Einige davon sind so oder in ähnlicher Form bereits im Einsatz, andere scheinen zum jetzigen Zeitpunkt noch nicht weitere Verbreitung gefunden zu haben.

Sämtliche Amplituden sind in dB , falls nicht anders angegeben. Der ungeheuer große Wertebereich wird so „zusammengestaucht“. Außerdem lassen sich die dynamischen Verläufe (Abklingen, Aufschwingen) gut mit e -Funktionen beschreiben. Beim Wechsel in die logarithmische Domäne werden diese Kurven dort linear.

3.1.1 Peakhöhe

Ein naheliegendes Maß ist die Höhe einer Erhebung im Spektrum. Da die absolute Verstärkung und damit die maximalen Amplitudenwerte des Signalanteils nicht bekannt sind, ist der oben erwähnte absolute Grenzwert wenig sinnvoll. Also besser relative Amplitudenwerte, die einen Bezug zur Umgebung liefern. So gibt es nun hierfür 3 Varianten:

- **Peakhöhe zu direkten Nachbarn:**

Hier wird einfach die Differenz der Amplitude in dB des aktuellen Bins von den beiden Nachbarn ermittelt:

$$\delta f A_{near}[\tilde{n}, k] = A[\tilde{n}, k] - 1/2 \cdot (A[\tilde{n}, k - 1] + A[\tilde{n}, k + 1])$$

- **Peakhöhe zu übernächsten Bins:**

Hier werden statt der unmittelbar benachbarten Bins erst die beiden übernächsten zum Vergleich herangezogen. Der Grundgedanke dahinter ist der, dass bei einer Überhöhung sehr viel Energie in die benachbarten Bins fließt. Die Amplitudendifferenz des Hamming-Fensters vom ersten zum zweiten Frequenzbin beträgt zum Beispiel gerade einmal $-7,4dB$. Als Nachteil schrumpft natürlich die Auflösung.

$$\delta f A_{wide}[\tilde{n}, k] = A[\tilde{n}, k] - 1/2 \cdot (A[\tilde{n}, k - 2] + A[\tilde{n}, k + 2])$$

- **Peakhöhe zu gemittelter Umgebung:**

Prinzipiell wird hierbei der Spitzenwert mit einem Mittel der Umgebung verglichen. Dies kann mit Moving-Average-Filtern oder anderen Arten der Fensterung geschehen. In unserem Fall erscheint mir ein rekursiver Filter mit Vergessensfaktor sehr günstig, da er sehr effizient ist und auch durch die immanente Gewichtung das direkte Umfeld betont. Ein paar Details zu dieser Gewichtung befinden sich im Anhang A. Die Lücke von einem Bin jeweils wie beim vorhergehenden Verfahren bleibt ebenso bestehen.

$$\begin{aligned} \delta f A_{av}[\tilde{n}, k] &= A[\tilde{n}, k] - 1/2 \cdot A_{AvDown}[\tilde{n}, k - 2] + A_{AvDown}[\tilde{n}, k + 2]) \\ A_{AvDown}[\tilde{n}, k] &= \alpha A_{AvDown}[\tilde{n}, k - 1] + (\alpha - 1)A[\tilde{n}, k] \\ A_{AvUp}[\tilde{n}, k] &= \alpha A_{AvUp}[\tilde{n}, k + 1] + (\alpha - 1)A[\tilde{n}, k] \end{aligned}$$

3.1.2 Zeitliches Amplitudenderivat

Als weiteres markantes Kennzeichen einer Rückkopplung ist der zeitliche Verlauf: wird wegen einer Änderung der geometrischen Situation die Feedback-Grenze überschritten, so schwingt das pfeifende Störsignal sehr schnell auf, bis es von einem Limiter begrenzt wird. Auch davor schwingt das ganze System auf einzelnen Frequenzen nach Anregung deutlich länger nach. Das Ausgabesignal klingt „hallig“.

Auch hier gibt es zumindest 2 Varianten:

- **Absolutes Amplitudenderivat:**

Die Amplitude der FFT-Bins des letzten Frames wird von den gleichen Bins des aktuellen Frames abgezogen und genormt. Durch diese Norm erhält man die Amplitudensteigung zum Zeitpunkt der Messung in dB/s . Alle Amplitudenveränderungen jenseits einer Obergrenze können als Rückkopplung eingestuft werden.

$$\delta t A[\tilde{n}, k] = (A[\tilde{n}, k] - A[\tilde{n} - 1, k]) \cdot \frac{f_s}{M_{step}}$$

f_s Samplefrequenz
 M_{step} Schrittweite der STFT

- **Relatives Amplitudenderivat:**

Die Steigung eines Frequenzbins über die Zeit allein ist bei vorhandenem Signal eventuell nicht aussagekräftig genug. Sie wird zusätzlich mit der der Gesamtenergie ins Verhältnis gesetzt. Die Gesamtenergie ist die Summe der Amplitudenquadrate eines STFT-Frames dividiert durch deren Anzahl im Quadrat. Die Logarithmierung erfolgt erst zum Schluss. Durch die Summierung der Quadrate erhält ein Frequenzbin um so mehr Anteil an der Gesamtleistung, je mehr er im Spektrum herausragt. Gleiches gilt natürlich für die Amplituden/Energieänderung. Die Differenz zwischen dem absoluten Amplitudenderivat eines Bins und dem nun logarithmischen Gesamtanstieg der Amplitude sollte also bei einem alles überschattenden Aufschwingvorgang gegen Null tendieren.

$$\delta t A_{rel}[\tilde{n}, k] = (A[\tilde{n}, k] - A[\tilde{n} - 1, k] - \delta t A_{global}[\tilde{n}]) \cdot \frac{f_s}{M_{step}}$$

$$\delta t A_{global}[\tilde{n}] = 10 \cdot \log_{10} \frac{\sum_{k=0}^N \tilde{A}^2[\tilde{n}, k] - \tilde{A}^2[\tilde{n} - 1, k]}{N^2}$$

$\tilde{A}$ nicht logarithmische Amplitude
 f_s Samplefrequenz
 S_{step} Schrittweite der STFT

3.1.3 Phasenderivat zweiter Ordnung $\frac{d^2\Phi}{dfdt}$

In neueren Veröffentlichungen () werden Phasenderivate zweiter Ordnung zur Verfeinerung von Spektrogrammen (FulopFitz 2007, Fulop 2007) oder Messergebnissen (Zotter) herangezogen. Die erste Ableitung $\frac{d\Phi}{dt}$ ergibt die Phasenfortschreitung pro Sekunde. Anders betrachtet ist dies auch nichts anderes als die Momentanfrequenz des jeweiligen Kanals/Bins (Channelized Instantaneous Frequency - $CIF[Hz]$). Bei einem einzigen enthaltenen Sinus ist diese über die Zeit konstant. Auf eine STFT angewendet beschreibt die zweite Ableitung $\frac{d^2\Phi}{dtdf}$ die Änderung der Momentanfrequenz von einem Bin zum nächsten. Im Falle eines Signals mit nur einem Sinus in einem Frequenzband fließt die Phaseninformation durch den

Leakage-Effekt mit abnehmender Intensität in alle anderen Bins. Kommen nun noch andere Rauschelemente oder Signale hinzu, so überlagern sich deren Effekte. Je bedeutender ein Signal ist, desto mehr fließt es in andere Bins. Die Phasenfortschreitung dieser Bins nähert sich an. Bei dominanten sinusförmigen Signalen oder scharfen spektralen Spitzen tendiert dieses Phasenderivat vor allem im direkten Umfeld gegen Null - je geringer der Unterschied von einem Bin zum andern, desto prominenter ist die spektrale Spitze. Physikalisch setzt sie also ähnlich der oben beschriebenen Peakhöhe an, jedoch mit dem Vorteil, dass sie sinusförmige Signalanteile von kurzen Rauschpeaks unterscheiden kann.

3.1.4 Phasenderivat zweiter Ordnung $\frac{d^2\Phi}{dt^2}$

Hiermit kann die Änderung der Phasenfortschreitung über die Zeit betrachtet werden. Bei einem Sinussignal mit konstanter Frequenz beträgt diese 0. Es lässt sich somit eine Aussage über die Frequenzänderungen oder -modulationen von Signalspitzen über die Zeit treffen. Außerdem lassen sich laute geräuschhafte Signalanteile und Impulse aussortieren, da diese keine zeitlich konstanten Phasenfortschreitungen produzieren. Diese Methode erscheint sehr vielversprechend. Ein massgebliches Kriterium für Rückkopplungen ist eine stabile additive Beziehung der verschiedenen Signalwege in der Akustik. Dies induziert eine stabile Phasenlage der einzelnen Wege im Frequenzbereich. Durch einfaches Schütteln oder Bewegen des Mikrofons um ein paar Zentimeter oder auch durch eine Modulation des Ausgabesignals lässt sich ein Feedback oft unterbinden. Werden also im Mikrofonsignal einige Signalanteile frequenziell zu statisch, so liegt der Verdacht einer positiv additiven Phase im Raum nahe und eine Grundbedingung für eine Rückkopplung ist erfüllt.

3.2 Phasenderivate im Detail

3.2.1 Algorithmus der „Finiten Differenzapproximation“

Dieses Verfahren ist das effizienteste aller für die Offline-Analyse verwendeten Verfahren. Es benötigt für die Erstellung von Derivaten zweiter Ordnung nur zwei FFTs pro STFT-Frame im Gegensatz zu 3 wie bei dem Verfahren nach Auger-Flandrin (FulopFitz 2006). Die Ergebnisse beider Methoden unterschieden sich nicht allzu deutlich voneinander - im Echtzeitfall geht Effizienz vor Genauigkeit.

Die erste FFT ($STFT_{del}$) wird zeitlich ein Sample vor der anderen ($STFT$) ausgeführt.

Die Momentanfrequenz CIF [Hz] der Bins lauten nun:

(Ganz exakt sind Momentanfrequenz eines um einen halben Sample verzögerten STFT-Frames)

$$CIF[\tilde{n}, k] = \frac{-f_s}{2\pi} \text{mod}(\angle(STFT_{del}[\tilde{n}, k]) - \angle(STFT[\tilde{n}, k]), 2\pi)$$

Für eine exakte Berechnung der Ableitungen zweiten Grades zu den FFT-Zeitpunkten würde nun sehr eine zusätzliche FFT und nachfolgende Berechnungen erfordern. Einfacher, wenn auch physikalisch nicht ganz korrekt, ist es die Übergänge der Momentanfrequenz zwischen in Frequenz und Zeit benachbarten Frequenzbins als linear zu betrachten. Derivate zweiten Grades reduzieren sich somit zu einfachen Subtraktionen.

$$\begin{aligned}
\frac{d^2\Phi}{dt df} &= \frac{N_{frame}}{f_s} (CIF[\tilde{n}, k] - CIF[\tilde{n}, k-1]) + (CIF[\tilde{n}, k+1] - CIF[\tilde{n}, k])/2 \\
&= \frac{N_{frame}}{2 \cdot f_s} (CIF[\tilde{n}, k+1] - CIF[\tilde{n}, k-1]) \\
\\
\frac{d^2\Phi}{dt^2} &= \frac{f_s}{M_{step}} (CIF[\tilde{n}+1, k] - CIF[\tilde{n}, k]) + (CIF[\tilde{n}, k] - CIF[\tilde{n}-1, k])/2 \\
&= \frac{f_s}{2 \cdot M_{step}} (CIF[\tilde{n}+1, k] - CIF[\tilde{n}-1, k])
\end{aligned}$$

An erster Formel erkennt man auch gut die Auflösungsfähigkeit der STFT. Nur, wenn sich in den beiden benachbarten Bändern keine beziehungsweise nur sehr schwache Sinussignale befinden, kann im jeweiligen Band eine harmonische Komponente zweifelsfrei identifiziert werden. $\frac{d^2\Phi}{dt df}$ ist ohne Einheit. Es gibt die Änderung der Momentanfrequenz pro Hertz im Spektogramm an. Bei einem Wert von 0 sind starke harmonische Komponente im Band enthalten.

Die zweifache Ableitung nach der Zeit gibt die Veränderung der Momentanfrequenz pro Sekunde an. Der Wertebereich geht erneut von 0 (absolut statischer harmonischer Signalanteil, keine Modulation oder Veränderung) bis hin zu extrem großen Werten bei starken Impulsen. Zu ergänzen wäre noch, dass ein Kompromiss nötig ist. Wegen der Kausalität kennen wir den $CIF[\tilde{n}+1, k]$ zum Zeitpunkt $\tilde{n}$ noch nicht. Eine Verzögerung um eine Schrittlänge M_{step} ist jedoch nicht akzeptabel, da diese dann auch in der Ausgabe eingebaut werden müsste. Folglich stellen wir nur die Veränderung zwischen dem aktuellen und dem letzten Frame fest, die Tangente ist also eine halbe Schrittweite hinterher:

$$\frac{d^2\Phi}{dt^2} = (CIF[\tilde{n}, k] - CIF[\tilde{n}-1, k])$$

Zu erwähnen wäre noch, dass von beiden Ergebnissen natürlich nur der Betrag betrachtet wird.

3.2.2 Spekulativer Unwrap-Algorithmus

Das große Hindernis bei der Benutzung von Phasenderivaten aus der STFT besteht darin, dass bei einer Schrittlänge M_{step} von mehr als einem Sample die Phasenbeziehungen und -fortschreitungen vieldeutig werden. Die Phasenverläufe zwischen zwei Frames sind

bei steigenden zeitlichen Abständen derselben immer unvorhersagbarer, da die Anzahl der „Überschläge“ einer fortschreitenden, unbekannten Momentanfrequenz natürlich ebenso unbekannt ist. Die Momentanfrequenz eines Bins befindet sich eigentlich nur im Falle einer deutlichen tonalen Inhaltes innerhalb des Frequenzbereiches des Bins. Eine effiziente Nutzung der Phaseninformation zur Feedbackdetektion mit vergleichbarer Komplexität wie die amplitudenbasierten Algorithmen somit also in weite Ferne.

Dreht man die Betrachtungsweise ein wenig, so ergibt sich doch eine Möglichkeit: Wenn ein Sinusignal in einem Frequenzband enthalten ist, so befindet sich die Momentanfrequenz innerhalb des Bins. In diesem Fall lässt sie sich mittels einer Abschätzung mit nachfolgender Korrektur ermitteln. Auch die beiden Phasenderivate zweiter Ordnung gehen somit gegen Null. Ist kein Signal enthalten, so ist die Schätzung falsch, und das Ergebnis für die Momentanfrequenz und die Derivate sind durch die falsch berechnete Phasenüberschläge mehr oder weniger zufällig. Mittelt man nun nur ein wenig über die Zeit, sei es mit einem einfachen Zähler, so lässt sich dieses wohl gleich verteilte Rauschen von dem gegen Null tendierenden funktionierenden Fall trennen. Die Trennschärfe wird wohl schlechter sein als bei obigen exakteren Verfahren, aber dafür benötigt es einen Bruchteil an Rechenaufwand und Speicher.

Nun zum Algorithmus selbst:

Ein Hinweis voran: im folgenden wird als Wertebereich der Phase immer $[0, 2\pi[$ statt $]-\pi, \pi]$ verwendet, alle Vektorindizes beginnen bei 0. Wir kennen also die Startphase und setzen voraus, dass sich die Momentanfrequenz durch die Anwesenheit eines Sinussignales oder einer deutlichen spektralen Spitze zwischen der Ober- und Untergrenze des Bins befindet. Deren Phasenfortschreitung von einem Frame zum nächsten ist jeweils bekannt:

$$\begin{aligned}\delta\Phi_{UpperBound}[\tilde{n}, k] &= \\ \delta\Phi_{LowerBound}[\tilde{n}, k] &= \end{aligned}$$

$\Rightarrow$: kommt noch, habe im letzten Moment noch Fehler entdeckt, den ich nicht so stehen lassen möchte...

3.3 Statistische Ergebnisse

Wegen des deutlichen Leckeffekts eines Signals in einem Frequenzbin auf die beiden benachbarten bei der FFT werden nur lokale Maxima bei der Evaluierung herangezogen. Die beiden benachbarten Bins müssen also kleiner sein.

Kapitel 4

Optimierung des Gesamtsystems

4.1 Qualitätsmaße

4.1.1 Übersicht über mögliche Maße

Für die Beurteilung der Leistung des Gesamtsystems kann man auf bewährte Evaluierungsverfahren aus der Sprachverarbeitung beziehungsweise der verlustbehafteten Audiokodierung zurückgreifen. Hier bieten sich mehrere Ansätze an (jeweils pro STFT-Frame):

Varianz der Logarithmischen Übertragungsfunktion [1]

$$\sigma_H^2[\tilde{n}] = \frac{1}{N} \sum_{k=0}^{N-1} (20 \cdot \log_{10} |H[\tilde{n}, k]| - \bar{H}_{dB}[\tilde{n}, k])^2$$

$\bar{H}_{dB}[\tilde{n}, k]$: Mittelwert aus dem Betrag aller Frequenzen der Ü-fkt[dB]
 $\tilde{n}$: Framenummer
 k : Frequenzbin
 N : STFT-Framelänge

Diese Formel ist einfach anwendbar, da wir ja die Übertragungsfunktion von der verzerrten zur unverzerrten Wiedergabe ohne weiteres erstellen können. Sie berücksichtigt jedoch nicht die unterschiedliche Wahrnehmung bei hohen und tiefen Frequenzen sowie die tatsächliche dynamische Signaldegradierung.

Spectral Flatness Measure (*SFM*)[1][6]

$$SFM[\tilde{n}] = \frac{\bar{H}_{\text{geometrisch}}^2}{\bar{H}_{\text{arithmetisch}}^2} = \frac{\sqrt[N]{\prod_{k=0}^{N-1} |H[\tilde{n}, k]|^2}}{\frac{1}{N} \sum_{k=0}^{N-1} |H[\tilde{n}, k]|^2}$$

$\bar{H}_{\text{geom}}^2$: geometrisches Mittel
 $\bar{H}_{\text{arith}}^2$: arithmetisches Mittel

Der Wert bewegt sich zwischen 0 und 1. Je „welliger“ die Übertragungsfunktion ist, desto geringer wird der SFM . Auch hier können die gleichen Kritikpunkte wie oben angemerkt werden.

Bark Spectral Distortion (BSD)

$$BSD[\tilde{n}] = \frac{\sum_{i=0}^{M-1} (L_X[\tilde{n}, l] - L_Y[\tilde{n}, l])^2}{\sum_{i=0}^{M-1} (L_X[\tilde{n}, l])^2}$$

$L_X[\tilde{n}, l]$: Band l des Bark Spektrums des Eingangssignals
 $L_Y[\tilde{n}, l]$: Band l des Bark Spektrums des Ausgabesignals
 M : Anzahl der Barkbänder

Aus der „Leistungsdichtefunktion“ (PSD) werden die einzelnen Frequenzbänder in Barkbänder gruppiert und innerhalb dieser addiert. Sind definierte tonale Komponenten in einem Band vorhanden, so dominieren sie durch die Addition der Leistung ⁽²⁾ das jeweilige Band. Anschließend werden die Bänder zuerst in dB -Ebene und anschließend in Phon umgerechnet. Damit entsteht zusätzlich eine gehörsrichtige Frequenzgewichtung.

Dieser Ansatz erscheint am sinnvollsten, da hier die vom Menschen wahrgenommene Beeinträchtigung ermittelt wird.

Fortgeschrittenere Ansätze wie „Modified Bark Spectral Distortion“ ($MBSD$)[5], „Enhanced Modified Bark Spectral Distortion“ ($EMBSD$)[4] oder „Perceptual Evaluation of Speech Quality“ ($PESQ$)[2] für die Evaluierung von Sprachkodierungssystemen berücksichtigen Maskierungseffekte. Diese nehmen jedoch zumeist Rauschen z.B. von der Quantisierung als Störquelle an und nicht eher statische, schmale Überhöhungen des Spektrums wie in diesem Fall. Um solch elaborierte Verfahren anwenden zu können, müsste erst ein anderes Maskierungsmodell in die BSD integriert werden. Letztere beiden sind noch dazu viel zu sehr auf Sprache als zugrunde liegendes Signal fixiert, um auch für musikalische Quellen aussagekräftig zu sein. Eine genaue Beschreibung des BSD -Algorithmuses befindet sich in [3].

Zwei kleine Anpassungen werden eingeführt:

Zum Vergleich mit dem feedbackbehafteten Signal wird nicht das ursprüngliche Audio-testmaterial herangezogen, sondern eine rückkopplungsfreie Aufnahme mit ausgeschaltetem Feedback- Lautsprecher. Damit fließen die Signalveränderungen durch das Mikrofon und den Messlautsprecher nicht in die Beurteilung mit ein. Diese Übertragungsfunktion soll mit dem Gerät ja nicht „verbessert“ werden. Im Idealfall fällt nur der Einfluß der Rückkopplungsschleife heraus.

Weiterhin werden in einem Zwischenschritt zuerst zwei BSD -Maße errechnet. BSD_{notch} ist die BSD der Barkbänder, die im Ausgangssignal kleiner sind als im Referenzsignal ohne Feedback. Bei BSD_{peak} ist es umgekehrt. Dadurch lässt sich beurteilen, ob das System eher zu stark (zu starke Filterung) oder zu schwach reagiert (zu viele Spitzen durch die Rückkopplung im Spektrum).

$$\begin{aligned}
 BSD[\check{n}] &= BSD_{notch}[\check{n}] + BSD_{peak}[\check{n}] \\
 BSD_{notch}[\check{n}] &= \sum (L_X[\check{n}, l] - L_Y[\check{n}, l])^2 \quad \text{falls } L_X > L_Y \\
 BSD_{peak}[\check{n}] &= \sum (L_X[\check{n}, l] - L_Y[\check{n}, l])^2 \quad \text{falls } L_X < L_Y
 \end{aligned}$$

Anhang A

Rekursive Fensterung

Anhang B

Oft verwendete Bezeichner

Φ	...	Phase
$\tilde{n}$	...	Framenummer der STFT
k	...	Frequenzbandnummer der STFT
N_{frame}	...	Framelänge der STFT
M_{step}	...	Schrittweite der STFT

Literaturverzeichnis

- [1] S. Goetze, M. Kallinger, K.-D. Kammeyer, and A. Mertins. Spatial sensitivity for listening room compensation. In *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA'07), Poster*, 2007.
- [2] ITU. Perceptual evaluation of speech quality (pesq): An objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs (itu-t p.862). Technical report, International Telecommunication Union (ITU), February 2001.
- [3] Shihua Wang, Andrew Sekey, and Allen Gersho. An objective measure for predicting subjective quality of speech coders. *IEEE J. Select. Areas Commun.*, 10:819–829, July 1992.
- [4] Wonho Yang. *Enhanced Modified Bark Spectral Distortion (EMBSD): an Objective Speech Quality Measure Based on Audible Distortion and Cognition Model*. PhD thesis, Temple University, 1999.
- [5] Wonho Yang, Majid Benbouchta, and Robert Yantorno. Performance of the modified bark spectral distortion as an objective speech quality measure. In *Proceedings of the 23rd International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 1998.
- [6] Zölzer. *Digitale Audiosignalverarbeitung*. Springer, 2005.