

# Kanalvokoder mit Synthesefiltern zweiter Ordnung

Seminararbeit aus Algorithmen in Akustik und Computermusik 2

Klaus Dobbler  
Florian Pausch

Betreuung: Dr. Franz Zotter, DI Matthias Frank  
Graz, 15. April 2011



institut für elektronische musik und akustik



## Zusammenfassung

Der Kanalvokoder wurde in den 1930er Jahren von Homer Dudley (*AT&T Laboratories, Circuit Research Department*) entwickelt und wird seither in verschiedenen Gebieten eingesetzt. Während er in der Telekommunikation vor allem der Codierung und Übertragung von Sprache dient, kommt er im Bereich Sprachsignalverarbeitung zur Trennung von Quelle und Filter (Vokaltrakt und Glottis) zum Einsatz, um nur zwei Anwendungen zu nennen.

In dieser Arbeit wird der Kanalvokoder zur Bestimmung der spektralen Einhüllenden eines Fourierkurzzeitpektrums (Analysestufe) und dessen Nachbildung durch Kanalimpulsantworten mit aufgeprägten, optimalen Gewichtskoeffizienten (Synthesestufe) verwendet. Es wird zusätzlich eine Methode beschrieben, eine minimalphasige Impulsantwort aus dem Betragsfrequenzgang eines Systems mittels cepstraler Analyse zu bestimmen.

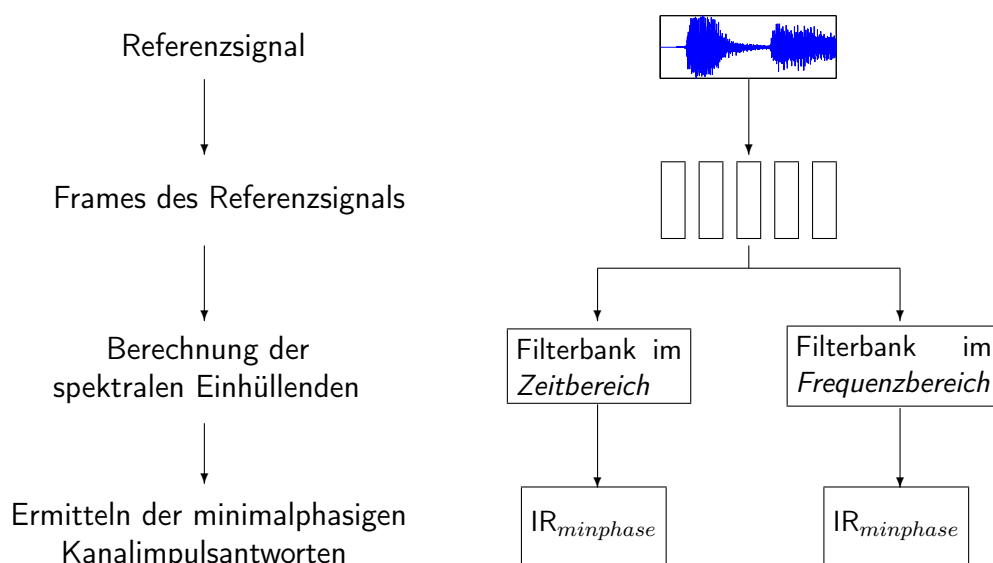
## Inhaltsverzeichnis

<b>1</b>	<b>Einleitung</b>	<b>4</b>
<b>2</b>	<b>Analysestufe des Kanalvokoders</b>	<b>4</b>
2.1	Ermittlung der spektralen Einhüllenden im Zeitbereich . . . . .	4
2.2	Ermittlung der spektralen Einhüllenden im Frequenzbereich . . . . .	6
2.3	Berechnung der minimalphasigen Impulsantworten . . . . .	7
2.3.1	Minimalphasige Systeme . . . . .	8
2.3.2	Verwendung des reellen Cepstrums . . . . .	9
<b>3</b>	<b>Synthesestufe des Kanalvokoders</b>	<b>11</b>
3.1	Design der Synthesefilterbank . . . . .	11
3.2	Berechnung der optimalen Gewichtungskoeffizienten . . . . .	11
<b>4</b>	<b>Ergebnisse</b>	<b>13</b>
<b>5</b>	<b>Echtzeit-Implementierung in Pure Data (PD)</b>	<b>16</b>
<b>6</b>	<b>Verbesserungen, Schlussfolgerung</b>	<b>18</b>

# 1 Einleitung

Ziel des Seminars war es, die zeitvariante Einhüllende des Spektrums eines Referenzsignals  $x(n)$  zunächst zu ermitteln (Abschnitt 2) und anschließend mithilfe einer Synthesefilterbank bestehend aus Bandpässen 2. Ordnung nachzubilden (Abschnitt 3). Diese verwendet weißes Rauschen als Eingangssignal  $\tilde{x}(n)$ . Die Filterbankimpulsantworten werden anschließend mittels berechneten optimalen Koeffizienten (Abschnitt 3.2) entsprechend gewichtet.

Um einen Überblick über die einzelnen Bearbeitungsschritte zu bekommen, sollen Abbildung 1 (Analysestufe) bzw. Abbildung 2 (Synthesestufe) dienen.



**Abb. 1:** Flowchart der Analysestufe mit Erklärung der einzelnen Verarbeitungsstufen

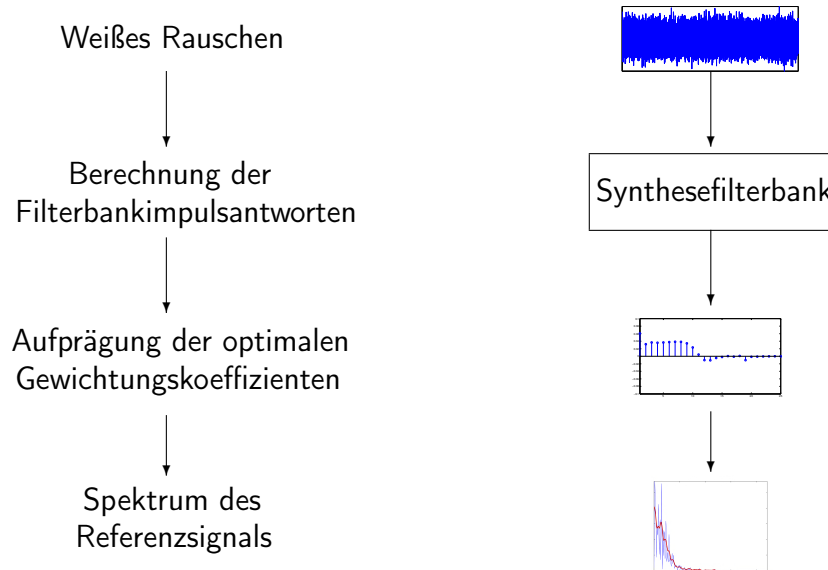
## 2 Analysestufe des Kanalvokoders

### 2.1 Ermittlung der spektralen Einhüllenden im Zeitbereich

Das zeitvariante Spektrum eines Eingangssignals  $x(n)$  kann auch als Filterbank mit  $B$  parallelen Kanälen (Bandpassfiltern) mit Impulsantwort  $h_b(n)$  ( $b = 1, 2 \dots B$ ) und dem entsprechendem Filterausgang, berechnet über die lineare, diskrete und kausale Faltungssumme

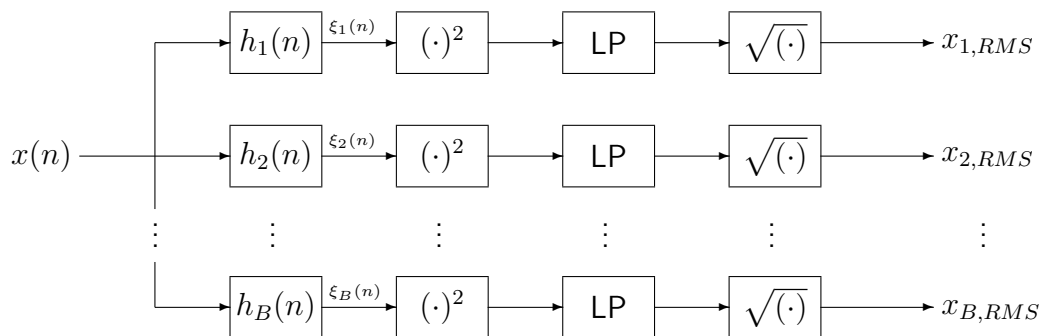
$$\xi_b[n] = \sum_{k=0}^{\infty} x[k]h_b[n-k], \quad (1)$$

betrachtet werden (Abbildung 3). Um eine gehörangepasste Schätzung der spektralen Einhüllenden zu erhalten, kann beispielsweise eine Terz-/Oktav- oder Bark-Filterbank



**Abb. 2:** Flowchart der Synthesestufe mit Erklärung der einzelnen Verarbeitungsstufen

verwendet werden. Der jeweilige Filterausgang wird zunächst quadriert, tiefpassgefiltert und radiziert, was zu einem Effektivwert  $x_{b,RMS}$  ( $b = 1, 2 \dots B$ ) führt.



**Abb. 3:** Analyse-Filterbank im Zeitbereich zur Ermittlung der spektralen Einhüllenden

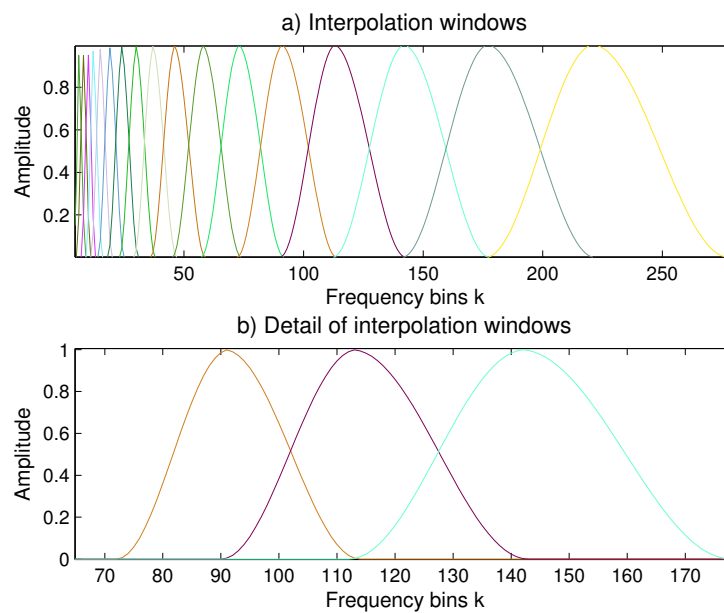
Trägt man diese Effektivwerte nun über der Frequenz auf, erhält man eine Schätzung der spektralen Einhüllenden des Eingangssignals  $x(n)$ .

Um eine lineare Frequenzachse zu erhalten (wird für die Berechnung einer minimalphasigen Impulsantwort benötigt, Abschnitt 2.3), bedarf es eines Mappings mittels Interpolationsfenstern. Dabei werden die fehlenden FFT-Werte durch Hanningfenster interpoliert. Es ist zu beachten, dass die Fenster mit steigender Frequenz breiter werden (vgl. logarithmische Anordnung der Filterbank).

Zur Vermeidung von Einbrüchen in der geschätzten Einhüllenden (dies führt zu Artefakten bei der Berechnung der minimalphasigen Impulsantworten), überlappen die Fenster

bis zu den jeweiligen benachbarten Mittenfrequenzen. Durch die mit der Frequenz ansteigende Breite, kommt es dazu, dass diese nicht symmetrisch sind, sondern die rechte Hälfte mehr Frequenzbins als die linke Hälfte abdeckt. Abbildung 4 zeigt die Interpolationsfenster aller Kanäle.

Summiert man die Interpolationsfenster (Abbildung 5), sind die auftretenden Probleme im tieffrequenten Bereich bei unzureichender FFT-Auflösung (Überhöhungen) anzumerken, weshalb wir uns für den Algorithmus im Frequenzbereich (Abschnitt 2.2) entschieden haben.

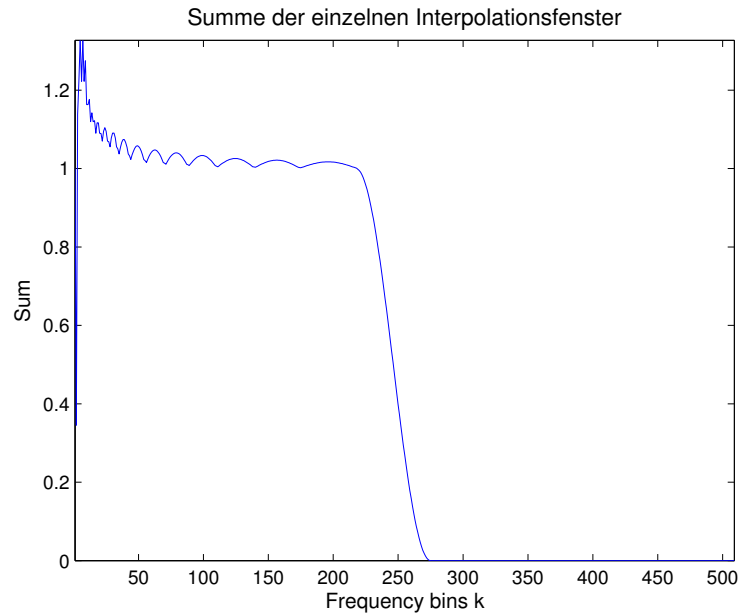


**Abb. 4:** a) Darstellung der sich überlappenden Interpolationsfenster  
b) Ausschnitt für drei Kanäle im mittleren Frequenzbereich

## 2.2 Ermittlung der spektralen Einhüllenden im Frequenzbereich

Ein elegantere (und auch genauere) Lösung zur Schätzung der Einhüllenden eines zeitvarianten DFT-Spektrums des Eingangssignals  $x(n)$  kann direkt im Frequenzbereich realisiert werden. Das Design der einzelnen Filterkanäle erfolgt im Frequenzbereich, wobei jeder Kanal die Energie des entsprechenden Frequenzbereichs enthält (dies entspricht genau den aufsummierten, durch den Fenstertyp festgelegten, gewichteten Einzelenergien der betrachteten Bins). Um Amplitudenwerte zu erhalten, muss man aus den Energiewerten die Wurzel ziehen.

Zunächst erfolgt die Berechnung eines Kurzzeitfourierspektrums des Eingangssignals  $x(n)$  durch Hanning-Fensterung (Framesize = 1024 Samples, dies entspricht einer zeitlichen Auflösung von rund 23 ms @  $f_s = 44.1$  kHz, was für Sprach- oder Musiksignale ausreichend ist) und anschließender Fast-Fouriertransformation (FFT). Auch für das



**Abb. 5:** Aufsummierte Hanning Interpolationsfenster

Glättungsfenster wird ein Hanning-Fenster verwendet. Der Grad der Glättung ist über die Fensterbreite einstellbar (je breiter, desto höher die anschließende Glättung). Die Schätzung der Einhüllenden erfolgt nun mithilfe der zirkulären Faltung im Frequenzbereich

$$\hat{X}(k) = \sqrt{|X(k)|^2 \circledast w(k)},$$

$|X(k)|^2$  ... Kurzzeitleistungsdichtespektrum des Eingangssignals

$\circledast$  ... zirkuläre Faltungsoperation

$w(k)$  ... Hanning-Glättungsfenster

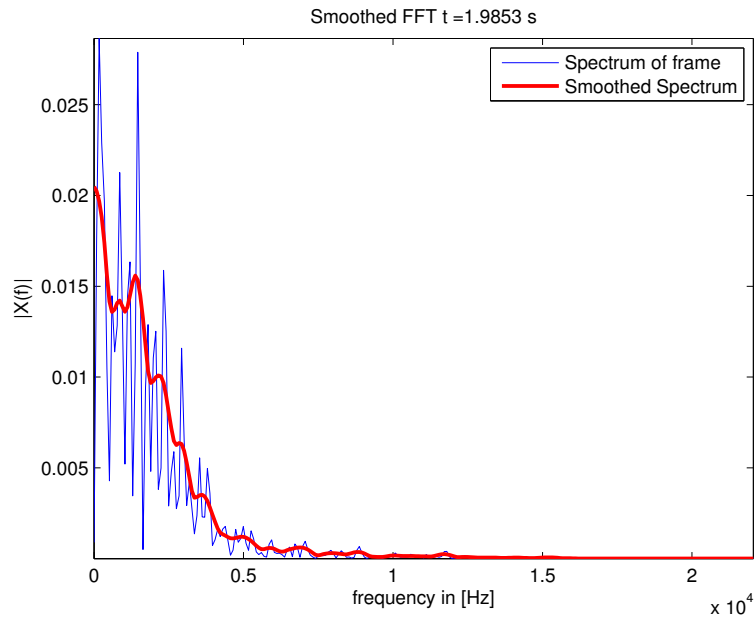
die über einen FFT-IFFT-Algorithmus [ZÖ2] effizient berechnet werden kann:

$$\hat{X}(k) = \text{Re}\left\{\sqrt{\text{IFFT}[\text{FFT}\{|X|^2\} \cdot \text{FFT}\{w(k)\}]}\right\}.$$

In Abbildung (6) kann man als Ergebnis exemplarisch ein geglättetes Kurzzeitfourierspektrum sehen.

### 2.3 Berechnung der minimalphasigen Impulsantworten

Im letzten Schritt der Analysestufe erfolgt die Berechnung der minimalphasigen Impulsantworten aus den spektralen Einhüllenden mit Hilfe einer cepstralen Methode (Abschnitt 2.3.2).



**Abb. 6:** Kurzzeitfourierspektrum und geglättetes Spektrum (Einhüllende) eines Musiksignals, Framesize:  $N_{FFT} = 1024$  Samples @  $f_s = 44.1$  kHz, Glättungsfenster: Hanning (10 Bins)

### 2.3.1 Minimalphasige Systeme

Minimalphasige Systeme besitzen spezielle Eigenschaften, die im Folgenden kurz erklärt werden sollen.

Ein minimalphasiges System  $H_{min}(z)$  . . .

- hat alle Pol- und Nullstellen innerhalb des Einheitskreises,
- ist ein stabiles, kausales System und
- besitzt eine stabile Inverse  $H_{min}^{-1}(z)$ .

Jedes stabile und kausale System  $H(z)$  kann in ein Allpasssystem und ein minimalphasiges System zerlegt werden:

$$H(z) = H_{min}(z) \cdot H_{ap}(z). \quad (2)$$

Ausgehend von dieser Beobachtung, kann folgende Eigenschaft abgeleitet werden.

**Minimale Gruppenlaufzeit** Die Gruppenlaufzeit eines stabilen und kausalen Systems ist über die negative Ableitung der Phase nach der Frequenz definiert:

$$\tau(\omega) = -\frac{\partial}{\partial \omega} \{ \arg[H(z)] \}. \quad (3)$$

Sie beschreibt die Verzögerung der einzelnen Frequenzen  $\omega$  in einem digitalen Filter und gibt daher Auskunft über den Grad der Nichtlinearität eines Systems hinsichtlich



Phasenverzerrungen. Das System aus Gl. (2) führt zu einer Gruppenlaufzeit von

$$\text{grd}\{H(z)\} = \text{grd}\{H_{\min}(z)\} + \text{grd}\{H_{\text{ap}}(z)\}. \quad (4)$$

Da bekannt ist, dass Allpasssysteme für alle Frequenzen  $\omega$  eine positive Gruppenlaufzeit aufweisen, muss das Minimalphasensystem ( $\tau(\omega)$  ist abschnittsweise negativ) die minimale Gruppenlaufzeit aufweisen [OSB99].

**Minimales Energiedelay** Vergleicht man die Impulsantworten von Systemen mit gleichem Betragsfrequenzgang (Abbildung 7), ist die Partialenergie des minimalphasigen Systems am wenigsten verzögert (d.h. die Energieverzögerung ist kausal (rechtsseitig) um  $n = 0$  konzentriert) [OSB99].

Besonders die letzte Eigenschaft ist im Rahmen dieser Arbeit wichtig. Betrachtet man nämlich die Kanalimpulsantworten der Synthesefilterbank (Abbildung 10), kann man erkennen, dass auch diese ihre Energie um  $n = 0$  konzentriert haben. Die Nachbildung der Impulsantworten der Analysefilterbank durch Linearkombination und optimale Gewichtung ist daher nur möglich, wenn auch jene ein minimales Energiedelay haben, also minimalphasig sind.

### 2.3.2 Verwendung des reellen Cepstrums

Im Allgemeinen kann ein und derselbe Betragsfrequenzgang  $|H(z)|$  durch unterschiedliche Pol-/Nullstellenkonstellationen repräsentiert werden. Als Veranschaulichung dieser Mehrdeutigkeit sollen die Systeme in Abbildung (7) dienen.

Unter Ausnutzung der Relation, dass die Fouriertransformation des geraden Anteils einer reellen Folge  $h_e(n)$  gleich dem Realteil ihres Spektrums entspricht ( $\text{FFT}\{h_e(n)\} = H_R(z)$ ), kann man aus dem Betragsfrequenzgang die zugehörige minimalphasige Impulsantwort berechnen:

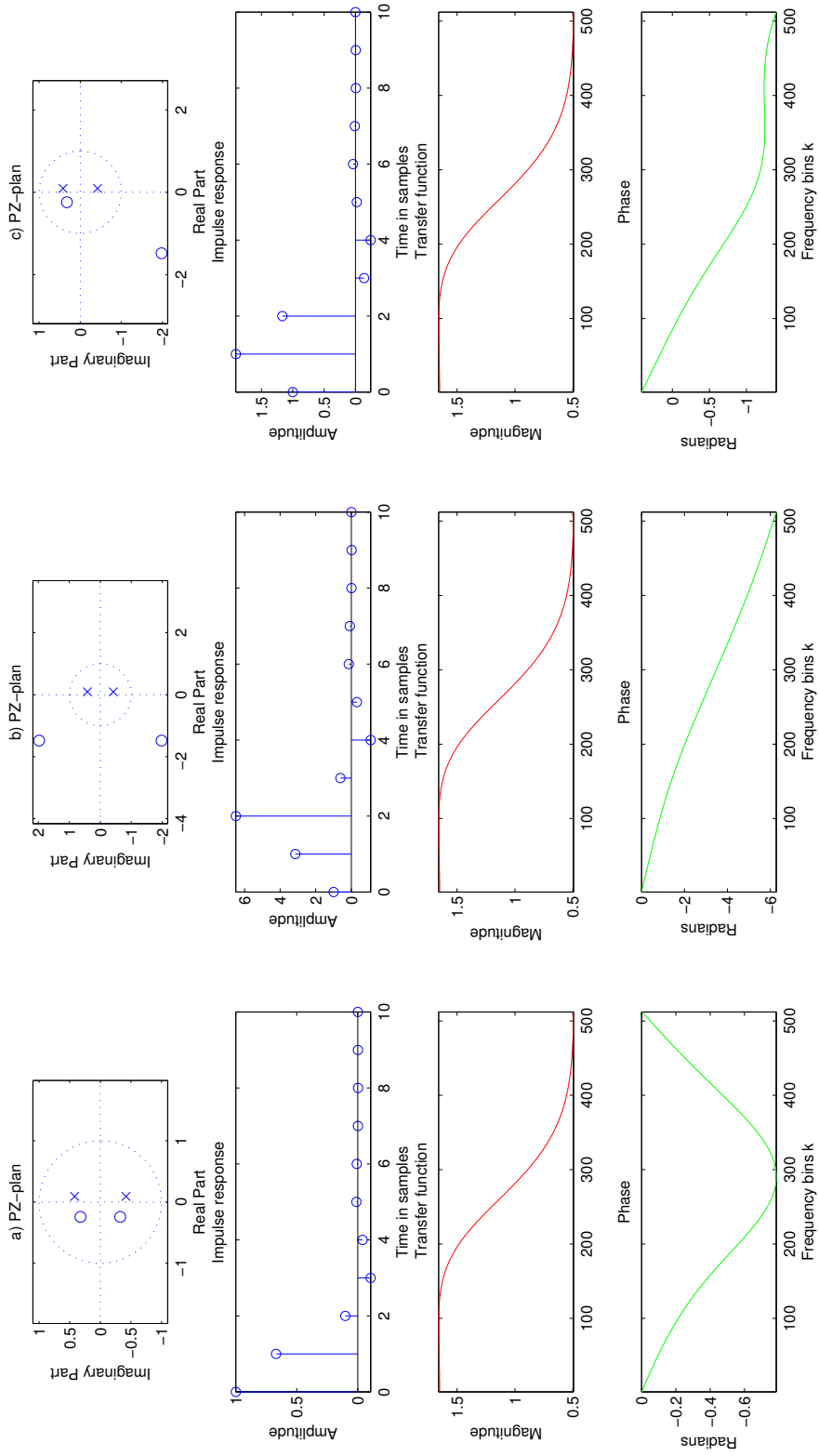
$$\begin{aligned} \hat{H}(z) &= \log\{H(z)\} = \log|H(z)| + j \arg\{H(z)\} \\ \hat{C}(z) &= \log|H(z)| \\ \hat{h}(n) &= \text{IDFT}\{\hat{H}(z)\} \\ \hat{c}(n) &= \text{IDFT}\{\text{Re}\{\hat{H}(z)\}\} = \dots \\ &= \text{IDFT}\{\log|H(e^{j\omega})|\} = \text{IDFT}\{\hat{C}(z)\}. \end{aligned}$$

Man sieht, dass das reelle Cepstrum  $\hat{c}(n)$  mit der geraden Komponente des komplexen Cepstrums  $\hat{h}(n)$  gleichzusetzen ist.

Da jede minimalphasige Folge ein kausales Cepstrum besitzt und jede kausale Folge  $h(n)$  durch ihren geraden Anteil mittels folgender Beziehung

$$h(n) = h_e(n)u_+(n) = \hat{c}(n)u_+(n),$$

mit  $u_+(n) = \begin{cases} 0 & n < 0 \\ 1 & n = 0, N/2 \\ 2 & n > 0, \end{cases} \quad N \dots \text{Anzahl der DFT-Bins,}$



**Abb. 7:** Systeme mit gleichem Betragsamplitudengang  $|H(z)|$ : (a) minimalphasiges System, (b) maximalphasiges System, (c) gemischtphasiges System; jeweils mit Impulsantwort, Betragsgang und Phasengang

in Verbindung steht, kann das zur minimalphasigen Folge zugehörige Cepstrum berechnet werden.  $h(n)$  entspricht also dem Cepstrum einer minimalphasigen Folge  $h_{min}(n)$  [PL06, T.N01].

Dieser Berechnungsvorgang ist in Abbildung 8 dargestellt.

### 3 Synthesestufe des Kanalvokoders

Laut Vorgabe war das Spektrum eines Referenzsignals  $x(n)$  mithilfe einer Synthesefilterbank mit  $M = 25$  Kanälen unter Verwendung von Filtern zweiter Ordnung bei weißem Rauschen als Eingangssignal  $\tilde{x}(n)$  zu rekonstruieren (Abbildung 9).

#### 3.1 Design der Synthesefilterbank

Die Synthesefilterbank besteht aus  $M$  parallelen Kanälen (Bandpassfilter zweiter Ordnung, Typ Butterworth), die jeweils mit weißem Rauschen angeregt werden<sup>1</sup>. Die Impulsantworten  $\eta_m$  ( $m = 1 \dots M$ ) haben wie bereits erwähnt eine sehr geringe Energieverzögerung (Abbildung 10).

#### 3.2 Berechnung der optimalen Gewichtungskoeffizienten

Für jeden Zeitpunkt  $n^2$  wird nun ein Satz optimaler Gewichtungskoeffizienten  $\mathbf{w}$  (Abbildung 11) berechnet und auf die einzelnen Filterbankimpulsantworten aufgeprägt.

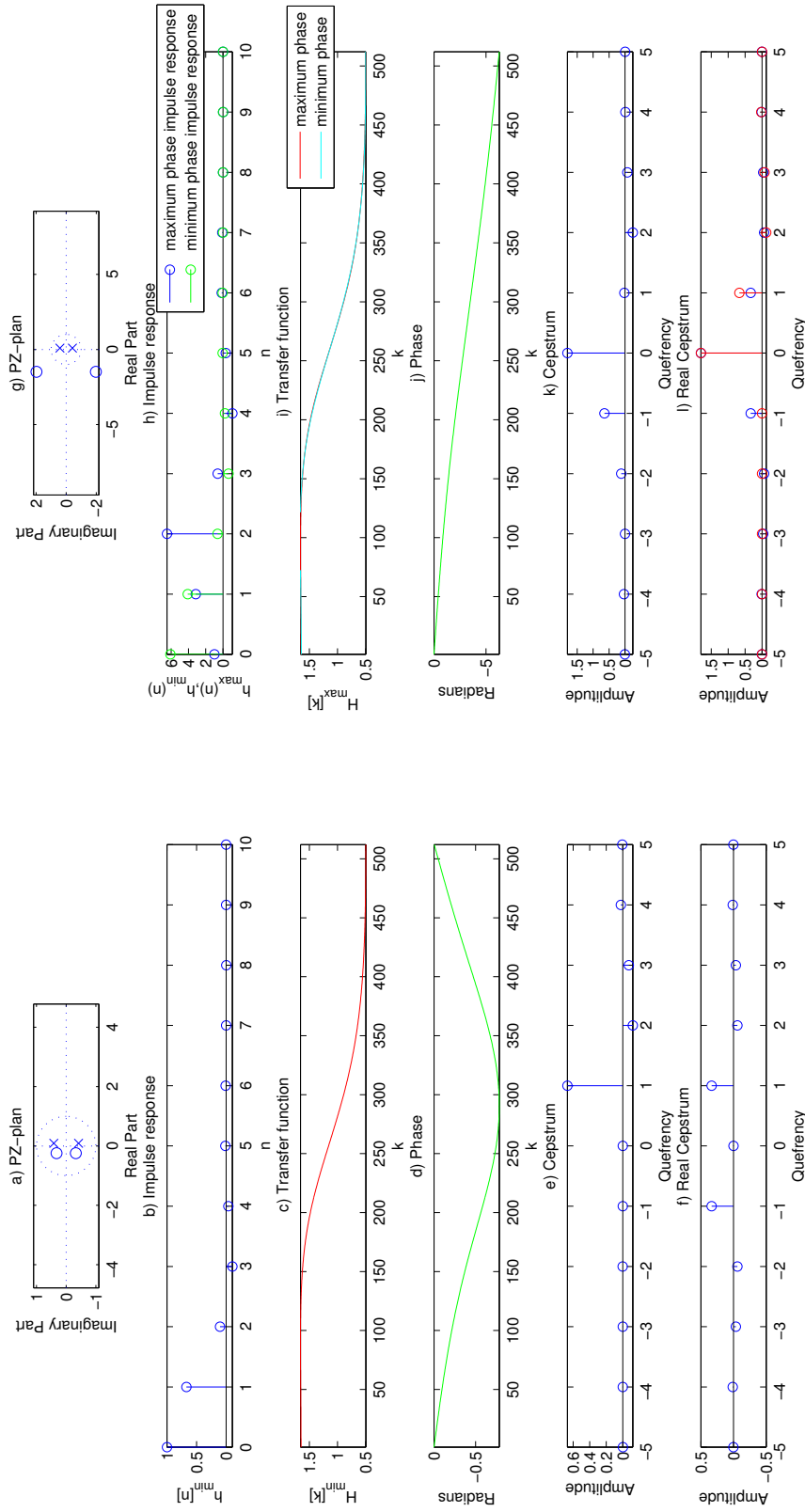
Mit dem Ansatz für eine Fehlerfunktion  $\|\mathbf{e}\| = \|(\mathbf{m} - \mathbf{H}\mathbf{w})\|$  ( $\mathbf{m}$  entspricht hierbei den minimalphasigen Impulsantworten der Analysestufe,  $\mathbf{H}$  jenen der Synthesefilterbank) ergibt sich für den quadratischen Fehler die Kostenfunktion

$$\begin{aligned} J(\mathbf{w}) &= \mathbf{e}^T \mathbf{e} = (\mathbf{m} - \mathbf{H}\mathbf{w})^T (\mathbf{m} - \mathbf{H}\mathbf{w}) = \\ &= \mathbf{m}^T \mathbf{m} - (\mathbf{H}\mathbf{w})^T \mathbf{m} - (\mathbf{H}\mathbf{w})^T \mathbf{m} + \mathbf{w}^T \mathbf{H}^T \mathbf{H} \mathbf{w} = \\ &= \mathbf{m}^T \mathbf{m} - 2\mathbf{m}^T \mathbf{H}\mathbf{w} + \mathbf{w}^T \mathbf{H}^T \mathbf{H} \mathbf{w}. \end{aligned}$$

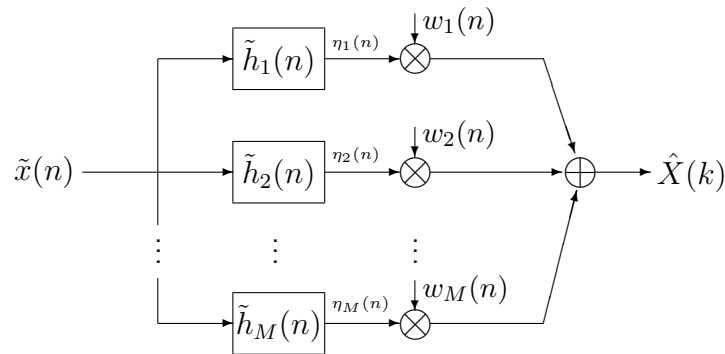
---

1. Anzumerken ist hier, dass die Mittenfrequenz des höchsten Bandes knapp über  $\frac{f_s}{4}$  liegt, da über dieser Frequenz keine Butterworth-Bandpassfilter 2. Ordnung erzeugt werden können.

2. Für sehr gute Rekonstruktionsergebnisse reicht es sowohl in der Analysestufe als auch in Synthesestufe nur jeden 512-ten Zeitpunkt zu berechnen.



**Abb. 8:** Darstellung der Berechnung einer minimalphasigen Impulsantwort  
 (a), (g) Pol-Nullstellendiagramme; (b), (h) Impulsantworten (grün: generierte minimalphasige Impulsantwort)  
 (c), (i) Betragsfrequenzgänge; (d), (j) Phasengänge; (e), (k) Cepstre  
 (f), (l) reelle Cepstre (rot: kausales Cepstrum zur Erzeugung der minimalphasigen Impulsantwort))



**Abb. 9:** Synthesefilterbank im Zeitbereich zur Rekonstruktion der spektralen Einhüllenden eines Referenzsignals

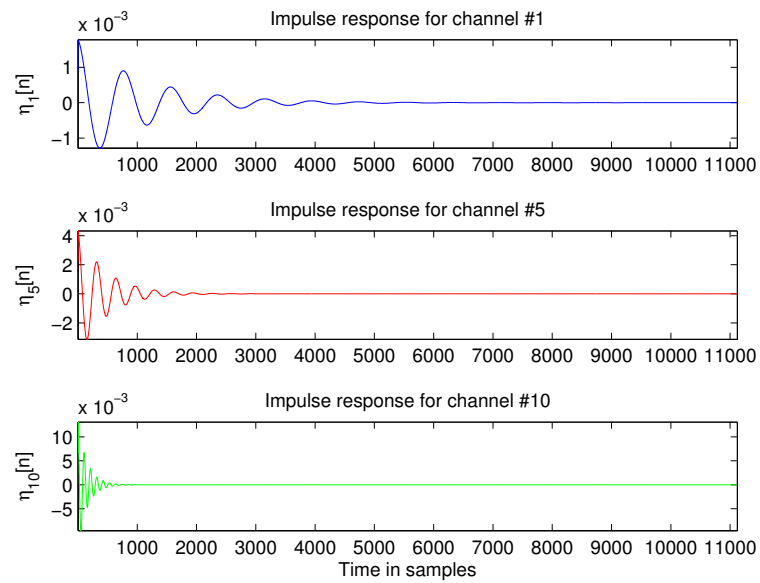
Um die optimalen Gewichtungskoeffizienten (optimal im Sinne eines *Minimum Squared Errors*) zu ermitteln, bildet man den Gradienten bezüglich  $\mathbf{w}$  und setzt die Kostenfunktion anschließend zu Null:

$$\begin{aligned}\nabla_{\mathbf{w}}\{J(\mathbf{w})\} &= \frac{\partial}{\partial \mathbf{w}}\{J(\mathbf{w})\} = \\ &= -2\mathbf{m}^T\mathbf{H} + 2\mathbf{H}^T\mathbf{H}\mathbf{w} = \mathbf{0} \\ \mathbf{H}^T\mathbf{H}\mathbf{w} &= \mathbf{m}^T\mathbf{H} \\ \mathbf{w}_{MSE} &= (\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\mathbf{m}\end{aligned}$$

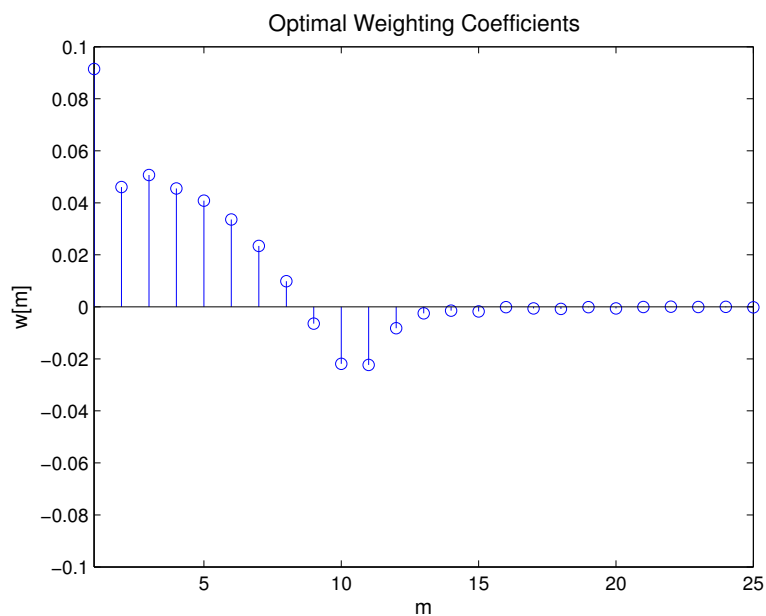
## 4 Ergebnisse

Als Ergebnis der Synthesestufe erhält man ein genähertes Spektrum des Referenzsignals. Nach Rücktransformation in den Zeitbereich und Kontrolle mittels Gehör, ist festzustellen, dass das Rekonstruktionssignal methodenbedingt keine Tonhöhen beinhaltet. Nichtsdestotrotz kann man das Referenzsignal sehr gut erahnen (Rhythmus, Dynamik, spektraler Inhalt). Diese Parameter reichen aus, um zwischen Sprache und Musik zu unterscheiden. Auch die Sprachverständlichkeit genügt, um den synthetisierten Inhalt zu verstehen. Bei Musik ist auffallend, dass der Nachhall sehr gut zu hören ist.

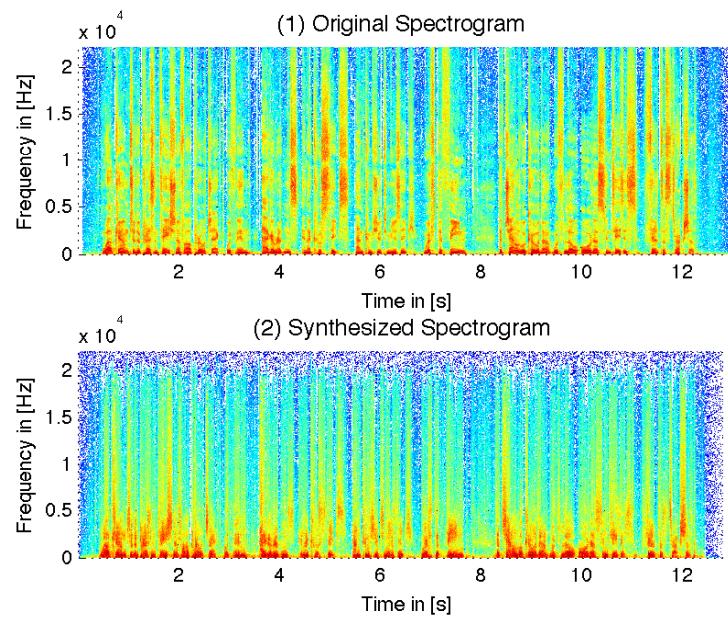
In den Abbildungen 12 und 13 kann man die Gegenüberstellung der Spektrogramme des originalen Sprachsignals und des synthetisierten Signals betrachten. Anzumerken ist die gute Übereinstimmung im tief- bis mittelfrequenten Bereich (bis ca. 5000 Hz), was für Sprache ausreichend ist. Die unvollständige beziehungsweise ungenaue Darstellung im höheren Frequenzbereich kommt einerseits durch die obere Grenzfrequenz der Synthesefilterbank (siehe Abschnitt 3.1) und andererseits durch die größer werdende Breite der Filterbankbänder zu hohen Frequenzen hin zustande. Das erklärt auch die vorhandenen Frequenzkomponenten oberhalb der Filterbankgrenzfrequenz, was besonders beim Spektrogramm des Musiksinal (Abbildung 13, unten) ersichtlich ist.



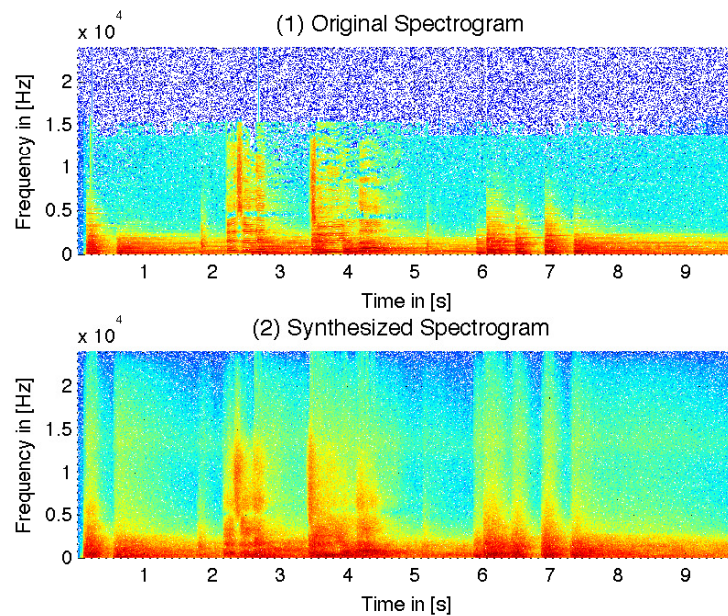
**Abb. 10:** Impulsantworten der Synthesefilterbank für den ersten, fünften und zehnten Filterbankkanal



**Abb. 11:** Satz optimaler Gewichtungskoeffizienten  $w$  für einen Zeitpunkt  $n$ ,  $M = 25$  Kanäle



**Abb. 12:** Syntheseresultate des Sprachsignals, (1) Spektrogramm des Originalsignals, (2) Spektrogramm des Synthesesignals,  $f_s = 44100$  Hz,  $N = 1024$  Bins



**Abb. 13:** Syntheseresultate des Musiksignals, (1) Spektrogramm des Originalsignals, (2) Spektrogramm des Synthesesignals,  $f_s = 44100$  Hz,  $N = 1024$  Bins

Der Algorithmus ist aufgrund der Verwendung der FFT und Ausnutzung von Cepstrum-Eigenschaften (Abschnitt 2.3.2) sowohl bei der Berechnung der spektralen Einhüllenden als auch bei der Ermittlung der minimalphasigen Impulsantworten recheneffizient implementiert. Auch der Rechenaufwand der Synthesestufe wird durch die niedrigen Filterordnungen gering gehalten.

## 5 Echtzeit-Implementierung in Pure Data (PD)

Zusätzlich zur Implementierung in MATLAB erfolgte eine Echtzeitimplementierung der Synthesestufe in PureData.

Beim Laden des Patches (Abbildung 14) werden die offline in MATLAB berechneten Koeffizienten geladen. Das *Metro*-Objekt dient dazu, diese Koeffizienten in Abständen von jeweils 11.6 ms (entspricht 512 Samples @  $f_s = 44100$  Hz) sequentiell auszulesen. Der jeweils aktuelle Koeffizientenvektor wird dem Subpatch *matlabCreatedFilterbank* übergeben.

Dieser Subpatch (Abbildung 15) besteht aus der Synthesefilterstruktur (25 Biquadfilter), welche dynamisch mit einem MATLAB-Skript erzeugt wurden. Der Vorteil dieser Herangehensweise liegt darin, dass sowohl die Koeffizienten der Filter nicht manuell eingetippt werden müssen als auch die Verbindung der einzelnen Objekte automatisch erfolgt.

Listing 1 beschreibt den gezeigten PD-Patch in der Abbildung 15 als ASCII-Zeichen.

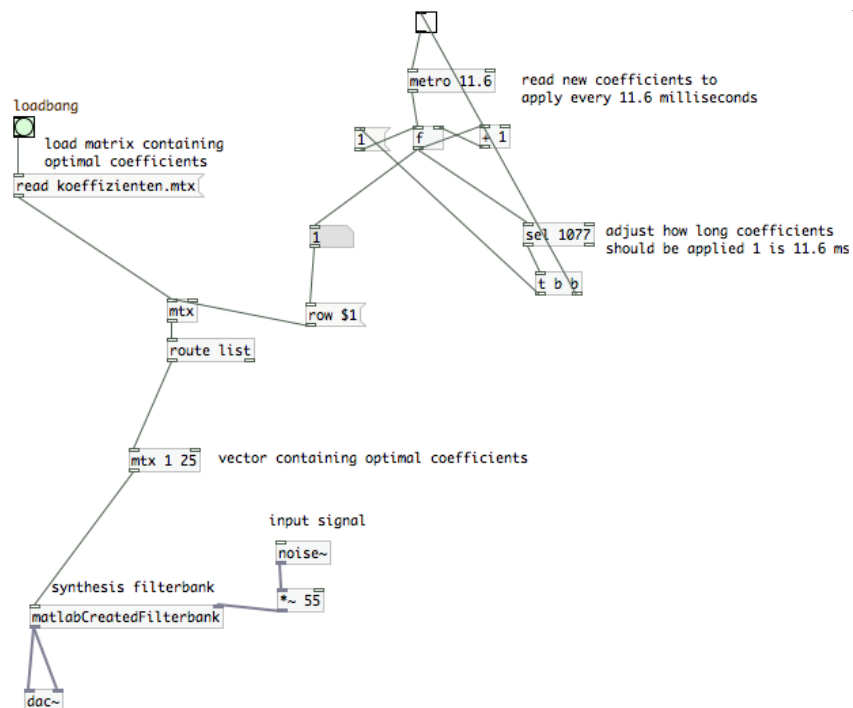


Abb. 14: Pure-Data-Patch für die Echtzeitsynthese



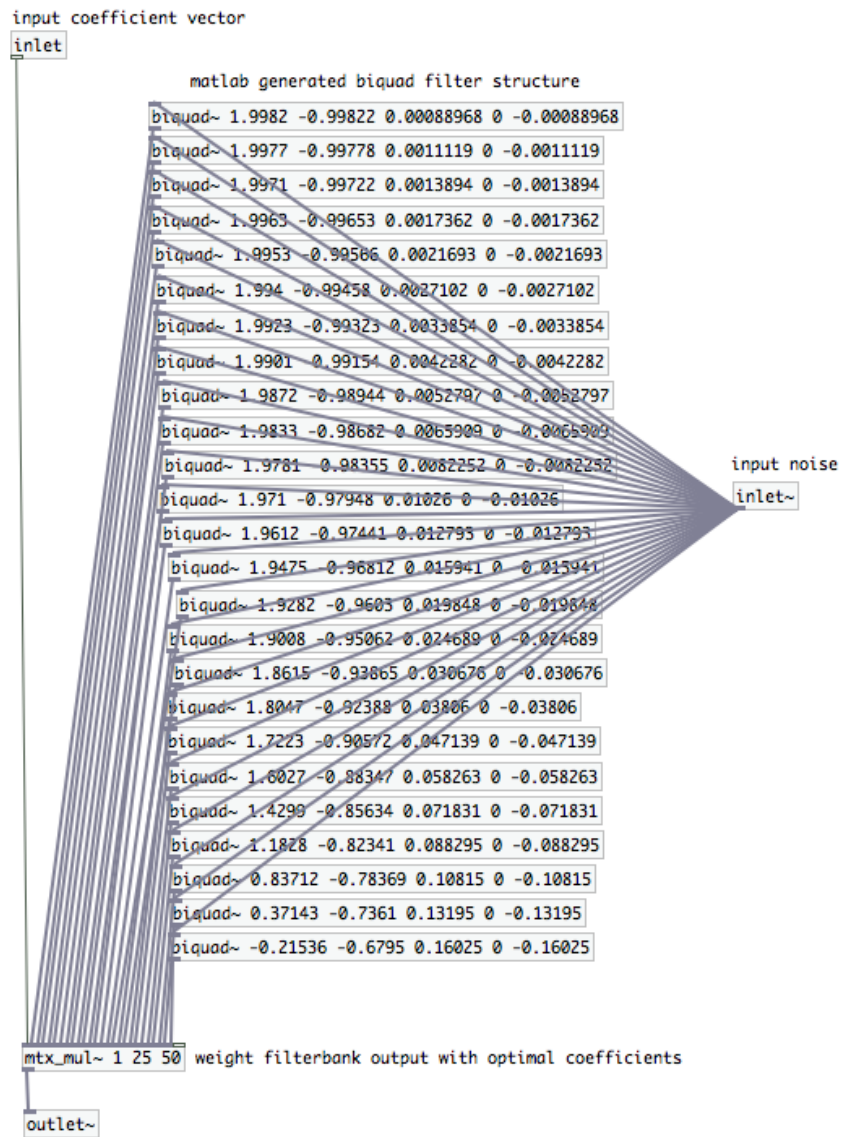


Abb. 15: Matlab erzeugte Synthese-Filterbank

**Listing 1:** Ausschnitt aus dem in MATLAB kreierten ASCII-Code

```

#N canvas 320 66 549 736 10;
#X obj 466 64 inlet~;
#X obj 106 -170 biquad~ 1.9982 -0.99822 0.00088968 0 -0.00088968;
.
.
.
#X obj 118 342 biquad~ -0.21536 -0.6795 0.16025 0 -0.16025;
#X obj 29 451 outlet~;
#X obj 21 -214 inlet;
#X obj 28 410 mtx_mul~ 1 25 50;
#X text 133 410 weight filterbank output with optimal coefficients;
#X text 20 -231 input coefficient vector;
#X text 463 44 input noise;
#X text 130 -192 matlab generated biquad filter structure;
#X connect 0 0 1 0;
.
.
.
#X connect 28 0 26 0;

```

## 6 Verbesserungen, Schlussfolgerung

Verbesserungen des Algorithmus könnten durch gehörgerechte Anpassungen erreicht werden – beispielsweise durch variable Fensterbreite (Verbreiterung gemäß Frequenzgruppen) beim Glättungsvorgang zur Ermittlung der spektralen Einhüllenden.

Will man auch tonale Signalanteile synthetisieren, braucht man einen Tonhöhen-Tracking-Algorithmus bzw. einen Pulsgenerator [B.G67]. Zur Schätzung der Glottisfrequenz bei Sprachsignalen kann das Signal invers mit seiner spektralen Einhüllenden (*Pre-Whitening*) gefiltert und damit das Quellsignal (Glottis) erhalten werden, aus welchem die Grundtonfrequenz ermittelt werden kann.

Durch eine Erweiterung der Echtzeitimplementierung in PD um die Analysestufe könnte man den Algorithmus auch im künstlerischen Bereich intuitiv und experimentell als Instrument bzw. als Effekt einsetzen.

Im Gegensatz zum traditionellen Kanalvokoder, bei welchem sowohl in der Analyse- als auch in der Synthesestufe durch die Überlappung der Filterbankkanäle Fehler entstehen, können im beschriebenen Algorithmus Fehler in der Synthesestufe durch die Berechnung der MSE-optimierten Gewichtungskoeffizienten minimiert werden. Die Abweichungen der Pegel in der Analysestufe können durch eine hohe FFT-Auflösung und geeignete Fensterung gering gehalten werden.

## Referenzen

- [B.G67] B.Gold, Ch.Rader, "The channel vocoder," *IEEE Transactions on audio and electroacoustics*, vol. -4C-15, KO. 4, December 1967.
- [OSB99] A. V. Oppenheim, R. W. Schafer, and J. R. Buck, *Discrete-Time Signal Processing (2nd Edition) (Prentice-Hall Signal Processing Series)*, 2nd ed. Prentice Hall, January 1999.
- [PL06] S.-C. Pei and H.-S. Lin, "Minimum-phase fir filter design using real cepstrum," *Circuits and Systems II: Express Briefs, IEEE Transactions on*, vol. 53, no. 10, pp. 1113 –1117, 2006.
- [T.N01] T.Nagarajan, V. Kamakshi Prasad and Hema A. Murthy, "The minimum phase signal derived from the magnitude spectrum and its application to speech segmentation," *Sixth Biennial Conference on Signal Processing and Communications, Bangalore, India*, 2001.
- [Zö2] U. Zölzer, *DAFX: Digital Audio Effects*. John Wiley & Sons, Mai 2002.