LAUFZEITOPTIMIERTE GERÄUSCHREDUKTIONSVERFAHREN BASIERENDAUF OVERLAP-SAVE-STRUKTUREN MIT PROJEKTIONSFILTERNÄHERUNGEN

Arthur Wolf[1], *Bernd Iser*[1] *Gerhard Schmidt*[2]

SVOX Deutschland GmbH[1] *Christian-Albrechts-Universität zu Kiel*[2]
Magirus-Deutz-Str. 16, 89077 Ulm *Kaiserstr. 2, 24143 Kiel*
E-Mail: arthur.wolf/bernd.iser@svox.com *E-Mail: gus@tf.uni-kiel.de*

Kurzfassung: Digitale Filterbänke sind wesentliche Bestandteile von aktuellen Algorithmen der Sprach- und Audiosignalverarbeitung. Sie sind ein wichtiger Bestandteil von Systemen zur Echounterdrückung und Bandbreitenerweiterung. Ein weiteres Anwendungsgebiet ist die Geräuschreduktion zur Sprachsignalverbesserung. Die Wahl der verwendeten Filterbank hat einen signifikanten Einfluss auf die Leistungsfähigkeit eines solchen Systems in Bezug auf Signalqualität, Rechenkomplexität und Signalverzögerung. Bei der Signalverarbeitung eines Innenraumkommunikationssystems müssen alle diese Kriterien berücksichtigt werden. Eine *overlap-save*-basierte FFT-Filterbank kann für eine Geräuschreduktion mit geringer Verzögerung verwendet werden. Dabei müssen die Filterkoeffizienten der Geräuschreduktion durch ein Projektionsfilter modifiziert werden, um zyklische Faltungseffekte zu unterdrücken. Der Nachteil einer solchen Filterbankstruktur ist die erhöhte Rechenkomplexität. Durch die Wahl einer geeigneten Projektionsfilternäherung kann diese bei nahezu gleichbleibender Signalqualität erheblich reduziert werden.

1 Einleitung

Die Kommunikation im Kraftfahrzeug während der Fahrt gestaltet sich aufgrund des hohen Hintergrundgeräuschpegels oftmals schwierig. Dies gilt insbesondere, wenn sich der eine Gesprächspartner auf einem der Vordersitze und der andere auf einem der Rücksitze befindet. Abhilfe können hier sogenannte Innenraumkommunikationssysteme (ICC-Systeme) schaffen. Hierzu wird die Sprache des jeweiligen Kommunikationspartners mit einem oder mehreren Mikrofonen aufgenommen und nach entsprechender Signalverarbeitung über diejenigen Lautsprecher, welche sich in der Nähe der übrigen Passagiere befinden, wiedergegeben. Neben dem Sprachsignal nehmen die Mikrofone auch Störungen (z.B. Motor-, Wind-, und Reifenabrollgeräusche) auf. Wird das geräuscherfüllte Mikrofonsignal lediglich verstärkt über die zuhörernahen Lautsprecher wiedergegeben, so wirken sich die aufgenommenen Störungskomponenten negativ auf die Sprachqualität aus. Durch eine geeignete Signalverarbeitung (Geräuschreduktion) können diese Störungen vor der Signalwiedergabe reduziert werden.

Um die Rechenkomplexität der Signalverarbeitung zu reduzieren, werden die meisten Algorithmen für Sprachsignalverbesserung im Frequenzbereich realisiert. Infolge der Blockbildung und Filterbankverarbeitung wird dabei eine Verzögerung von mehreren Millisekunden eingefügt, was sich bei ICC-Systemen negativ auf die Sprachverständlichkeit und Systemqualität auswirkt [1].

In der Literatur sind verschiedene Ansätze für eine laufzeitoptimierte Signalverarbeitung bekannt, z.B. die Filterbanksummationsmethode, die oftmals in Hörgeräten Anwendung findet [2]. Daneben kann man auch eine *overlap-save*-basierte FFT-Filterbank für eine Geräuschreduktion mit geringer Verzögerung verwenden. Eine weitere Anwendung für *overlap-save*-basierte Verarbeitung ist die *overlap-save*-FDAF (*frequency domain adaptive filtering*) [3]. Um dabei die Anteile der zyklischen Faltung zu reduzieren, muss die Impulsantwort beschränkt werden. Üblicherweise werden hierzu die Filterkoeffizienten im Zeitbereich durch ein Projektionsfilter gefenstert, wodurch die Rechenkomplexität (aufgrund der zusätzlichen FFT bzw. IFFT) ansteigt. Dieser zusätzliche Rechenaufwand kann durch eine Faltung im Fre-

quenzbereich mit einer geeigneten zyklischen Projektionsfilternäherung reduziert werden. Gleichzeitig kann die Sprachqualität nahezu unverändert erhalten bleiben.

In dieser Veröffentlichung wird ein laufzeitoptimiertes, *overlap-save*-basiertes Verfahren zur Geräuschreduktion beschrieben. Zur Reduktion der daraus folgenden zyklischen Faltungseffekte wird ein Verfahren zum Entwurf von zyklischen Projektionsfiltern vorgestellt. Dabei wird auch die Auswirkung des Projektionsfilters auf die Filterkoeffizienten des Geräuschreduktionsfilters und somit auch auf das Ergebnis der Geräuschreduktion berücksichtigt. Zusätzlich wird das vorgestellte Verfahren mit einer *overlap-add*-basierten Verarbeitung hinsichtlich der Rechenkomplexität, Laufzeit und Sprachqualität verglichen.

2 Schnelle Faltung

Eine mögliche Realisierung für eine recheneffiziente und laufzeitoptimierte Filterbank ist die Verwendung einer *overlap-save*-basierten Verarbeitung. Die *overlap-save*-Struktur reduziert, im Vergleich zur häufig verwendeten *overlap-add*-Struktur, die Signalverzögerung, da für den Ausgabeblock nach der Synthesefilterbank nicht erst noch die verzögerten und überlappenden Anteile addiert werden müssen. Die theoretischen Grundlagen der *overlap-save*-basierten Verarbeitung beruhen auf der schnellen Faltung. Hierfür wird die Faltung im Zeitbereich (lineare Faltung) durch eine Multiplikation im Frequenzbereich (zirkulare Faltung) durchgeführt. Die Signalanteile im Frequenzbereich werden blockweise recheneffizient mit Hilfe des FFT-Algorithmus bestimmt. Dabei müssen bestimmte Randbedingungen berücksichtigt werden, um die durch die zirkulare Faltung entstehenden Artefakte zu unterdrücken. Für die Herleitung der blockweisen Verarbeitung im Frequenzbereich wird zunächst die blockweise Faltung des Filtervektors $\mathbf{h}_\mathrm{z} = [h_0, ..., h_{N-1}]^\mathrm{T}$ der Länge N mit dem Mikrofonsignal $x(n)$ betrachtet. Der Vektor mit den L aktuellen Werten des Filterausgangs $\mathbf{y}_\mathrm{z}(k) = [y(kL-L+1), ..., y(kL-1), y(kL)]^\mathrm{T}$ kann berechnet werden als

$$\mathbf{y}_\mathrm{z}(k) = \begin{bmatrix} x(kL-L+1) & \dots & x(kL-N-L+2) \\ \vdots & \dots & \vdots \\ x(kL-1) & \dots & x(kL-N) \\ x(kL) & \dots & x(kL-N+1) \end{bmatrix} \begin{bmatrix} h_0 \\ h_1 \\ \vdots \\ h_{N-1} \end{bmatrix} = \mathbf{X}_\mathrm{z}(k)\,\mathbf{h}_\mathrm{z}. \tag{1}$$

k ist hier der Zeitindex des um den Faktor L unterabgetasteten Signals, z ist der Index für Matrizen im Zeitbereich und f für Matrizen im Frequenzbereich. Um die schnelle Faltung durchführen zu können, muss die Gleichung erweitert werden, bis man eine zirkulare $M \times M$ große Matrix $\overline{\mathbf{X}}_\mathrm{z}(k)$ erhält, mit $M = N + L - 1$. Aus diesem Grund wird der Filtervektor um $L-1$ Nullen erweitert

$$\overline{\mathbf{h}}_\mathrm{z} = [\mathbf{h}_\mathrm{z}^\mathrm{T}, 0, \dots, 0]^\mathrm{T}. \tag{2}$$

Wird auch der Ausgangssignalvektor um $N-1$ Werte erweitert, kann das Gleichungssystem mit einer zirkularen Anregungssignalmatrix angegeben werden [4]

$$\overline{\mathbf{y}}_\mathrm{z}(k) = \overline{\mathbf{X}}_\mathrm{z}(k)\,\overline{\mathbf{h}}_\mathrm{z}. \tag{3}$$

Dabei gilt für die zirkulare Anregungsmatrix

$$\overline{\mathbf{X}}_\mathrm{z}(k) = \begin{bmatrix} x(kL-N-L+1) & \dots & x(kL-N-L+3) & x(kL-N-L+2) \\ \vdots & \ddots & \vdots & \vdots \\ x(kL-2) & \dots & x(kL) & x(kL-1) \\ x(kL-1) & \dots & x(kL-N-L+1) & x(kL) \\ x(kL) & \dots & x(kL-N-L+2) & x(kL-N-L+1) \end{bmatrix}, \tag{4}$$

welche schon durch eine einzelne Zeile vollständig beschrieben ist, z.B. durch die letzte Zeile $\overline{\mathbf{x}}_\mathrm{z}(k) = [x(kL), ..., x(kL-N-L+2), x(kL-N-L+1)]^\mathrm{T}$. Die Faltung in Gl. (3) kann als elementweise Multiplikation im Frequenzbereich realisiert werden. Nach der Rücktransformation des Ergebnisses entsprechen

die letzten L Werte dem gesuchten Ergebnis. In Matrixnotation kann das Ausgangssignal wie folgt angegeben werden:[1]

$$\mathbf{y}_z(k) = \mathbf{C}\,\mathbf{F}^{-1}\mathrm{diag}\{\mathbf{F}\,\overline{\mathbf{x}}_z(k)\}\,\mathbf{F}\,\overline{\mathbf{h}}_z. \tag{5}$$

3 Teilbandgeräuschreduktion mit Projektionsfilter

Die hier angewendete Geräuschreduktion basiert auf dem Optimierungskriterium nach Wiener und wird recheneffizient im Frequenzbereich durchgeführt. Die hier vorgestellte Filterbank beruht auf der *overlap-save*-basierten schnellen Faltung. In diesem Abschnitt wird die Teilbandgeräuschreduktion und das für die *overlap-save*-basierte Verarbeitung notwendige Projektionsfilter vorgestellt.

3.1 *Overlap-save*-FFT-Filterbank mit Geräuschreduktion

Wird das Sprachsignal $s(n)$ von einem additiven Störgeräusch $b(n)$ überlagert, ist es das Ziel einer Geräuschreduktion mit dem Filter $\mathbf{h}_{\mathrm{z,opt}}$ die Anteile des Hintergrundgeräuschs $b(n)$ zu unterdrücken ohne das Sprachsignal $s(n)$ anzugreifen. Dabei ist das Mikrofonsignal $x(n)$ definiert als $x(n) = s(n) + b(n)$.

Weiterhin wird angenommen, dass die Signale $b(n)$ und $s(n)$ orthogonal sind, somit gilt $\mathrm{E}\{s(n)\,b(n+i)\} = 0,\ \forall\, i$. Das verarbeitete geräuschreduzierte Signal wird mit $\hat{s}(n)$ bezeichnet. Abb. 1 zeigt das Blockschaltbild der zugrundeliegenden Signalverarbeitung.

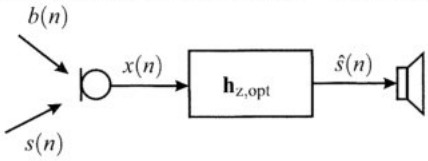

Abbildung 1 - Filterung eines geräuschbehafteten Signals zur Störungsreduktion.

Das Geräuschreduktionsfilter $\mathbf{h}_{\mathrm{z,opt}}$ wird häufig als Wiener-Filter im Frequenzbereich $H(e^{j\Omega})$ mit Hilfe der spektralen Leistungsdichten des geräuscherfüllten Signals $S_{xx}(e^{j\Omega})$ und des Störsignals $S_{bb}(e^{j\Omega})$ bestimmt. Es kann als Funktion dieser Größen angegeben werden

$$H(e^{j\Omega}) = f\left(\frac{S_{bb}(e^{j\Omega})}{S_{xx}(e^{j\Omega})}\right). \tag{6}$$

In der Praxis werden die spektralen Leistungsdichten durch spektrale Kurzzeitleistungsdichten angenähert. Details zur Herleitung des Wiener-Filters, die Realisierung im Frequenzbereich und die Schätzung der spektralen Kurzzeitleistungsdichten sind z.B. in [5] zu finden.

Das geräuschreduzierte Signal $\hat{s}(n)$ wird durch eine Faltung des geräuscherfüllten Signals $x(n)$ mit dem Geräuschreduktionsfilter $\mathbf{h}_z$ berechnet. Die Faltung kann auch als Multiplikation im Frequenzbereich durchgeführt werden. Eine laufzeit- und recheneffiziente Geräuschreduktion kann im Teilbandbereich mit der *overlap-save*-basierten FFT-Filterbank umgesetzt werden. Die hierfür notwendige Signalverarbeitung wird ähnlich dem *overlap-save*-FDAF- (*frequency-domain adaptive filter*) Algorithmus[2] implementiert. Aufgrund der Eigenschaften der FFT wird eine zyklische Faltung durchgeführt, deshalb müssen die Filterkoeffizienten im Zeitbereich mit einem Projektionsfilter (*constraint*) beschränkt werden, d.h. Teile des Filtervektors werden zu Null gesetzt, um die zyklischen Faltungseffekte zu unterdrücken.

Zur Beschreibung der blockweisen Signalverarbeitung ist die Matrixnotation zweckmäßig. Die FFT-Blockgröße ist M, der Rahmenversatz beträgt $L < M$ Abttasttakte. Der Block des Eingangssignals besteht aus L neuen und aus $N = M - L$ alten Abtastwerten. Zur Beschreibung des Verfahrens in Matrixnotation werden folgende Matrizen benötigt:

$$\begin{aligned}
\mathbf{W} &= \begin{bmatrix} \mathbf{I}_{N\times N} & \mathbf{0}_{N\times L} \\ \mathbf{0}_{L\times N} & \mathbf{0}_{L\times L} \end{bmatrix}, \quad \mathbf{C} = \begin{bmatrix} \mathbf{0}_{N\times L} \\ \mathbf{I}_{L\times L} \end{bmatrix}, \quad \mathbf{R} = \begin{bmatrix} \mathbf{0}_{M-P\times P} & \mathbf{I}_{M-P\times M-P} \\ \mathbf{I}_{P\times P} & \mathbf{0}_{P\times M-P} \end{bmatrix}, \\
\mathbf{F} &= M\times M\ \mathrm{FFT-Matrix}, \quad \mathbf{I} = \text{Einheitsmatrix}.
\end{aligned} \tag{7}$$

[1] Der Operator diag{...} generiert – sofern er auf einen Vektor angewendet wird – aus einem Vektor eine Diagonalmatrix. Für die Definition der Matrizen **C** und **F** sei an dieser Stelle auf Gl. 7 verwiesen.

[2] Recheneffizientes Verfahren zur Echokompensation im Teilbandbereich [3].

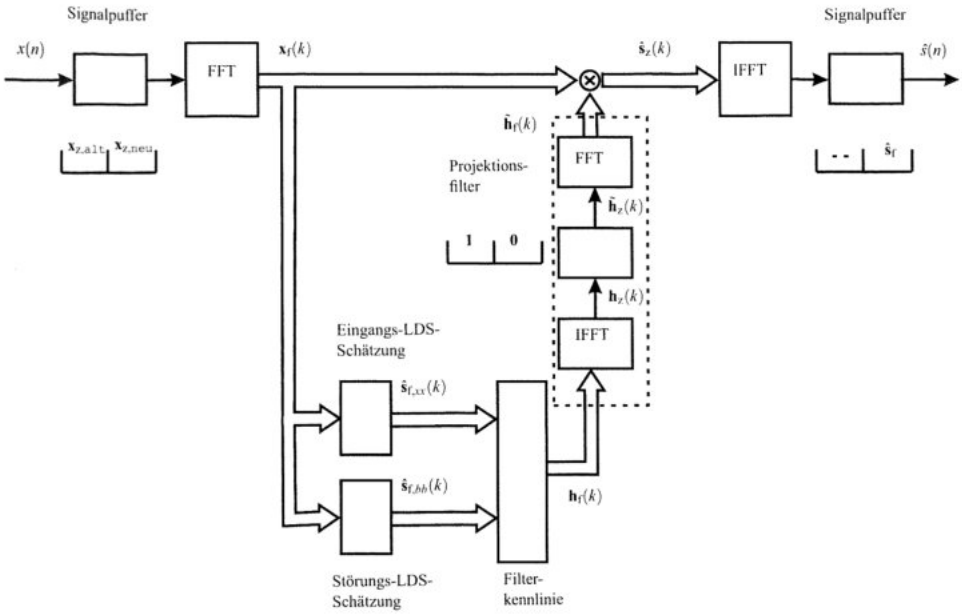

Abbildung 2 - Teilbandgeräuschreduktion mit einer *overlap-save*-basierten FFT-Filterbank (LDS steht für Leistungsdichtespektrum). Das Verfahren beruht auf der blockweisen schnellen Faltung.

Der Vektor des Eingangssignals im Frequenzbereich $\mathbf{x}_{\mathrm{f}}(k)$ wird aus dem blockweise verarbeiteten Mikrofonsignal $x(n)$ bestimmt

$$\mathbf{x}_{\mathrm{f}}(k) = \mathbf{F}\,\mathbf{x}_{\mathrm{z}}(k) = \mathbf{F}\,[x(kL-L), \ldots, x(kL+M-1)]^{\mathrm{T}}. \tag{8}$$

Das Geräuschreduktionsfilter aus Gl. (6) wird in Vektornotation wie folgt angegeben

$$\mathbf{h}_{\mathrm{f}}(k) = f\left(\left[\mathrm{diag}\{\hat{\mathbf{s}}_{\mathrm{f},xx}(k)\}\right]^{-1}\hat{\mathbf{s}}_{\mathrm{f},bb}(k)\right). \tag{9}$$

Zur Reduktion der zyklischen Faltungseffekte muss ein Teil der Filterkoeffizienten im Zeitbereich auf Null gesetzt werden (siehe auch Gl. (2)). Hierzu werden zunächst die Filterkoeffizienten in den Zeitbereich zurück transformiert

$$\mathbf{h}_{\mathrm{z}}(k) = \mathbf{F}^{-1}\,\mathbf{h}_{\mathrm{f}}(k). \tag{10}$$

Der Frequenzgang $\mathbf{h}_{\mathrm{f}}(k)$ ist aufgrund der Symmetrie der spektralen Leistungsdichten reell und symmetrisch, also nullphasig. Resultierend aus den zyklischen Eigenschaften der IFFT ist die Folge periodisch mit der Periodenlänge M. Wegen der Frequenzgangeigenschaften ist sie außerdem reell und gerade

$$h_{\mathrm{z},M-m}(k) = h_{\mathrm{z},m}(k) \quad \text{bzw.} \quad h_{\mathrm{z},-m}(k) = h_{\mathrm{z},m}(k). \tag{11}$$

Durch Rechtsverschiebung um $P \leq \frac{M}{2}$ und Fensterung der dann ersten $2P$ Werte erhält man eine zum Punkt $m = P$ symmetrische Impulsantwort:

$$\begin{aligned} \tilde{h}_{\mathrm{z},m}(k) &= h_{\mathrm{z},m-P}(k),\ m \in \{0,1,\ldots,2P\}, \\ \tilde{h}_{\mathrm{z},P+p}(k) &= \tilde{h}_{\mathrm{z},P-p}(k) = h_{\mathrm{z},p}(k),\ p \in \{0,1,\ldots,P\}, \end{aligned} \tag{12}$$

also ein linearphasiges, kausales FIR-Filter vom Grad $2P$ [6]. Aufgrund dieser Eigenschaften müssen die Filterkoeffizienten um P zyklisch mit der Rotationsmatrix $\mathbf{R}$ verschoben werden, um ein kausales

FIR-Filter zu erhalten. Durch die Multiplikation mit der Fenstermatrix $\mathbf{W}$ werden die hinteren Filterkoeffizienten auf Null gesetzt

$$\tilde{\mathbf{h}}_{\mathrm{z}}(k) = \mathbf{W}\,\mathbf{R}\,\mathbf{h}_{\mathrm{z}}(k). \tag{13}$$

Anschließend werden die Filterkoeffizienten $\tilde{\mathbf{h}}_{\mathrm{z}}(k)$ zurück in den Frequenzbereich transformiert

$$\tilde{\mathbf{h}}_{\mathrm{f}}(k) = \mathbf{F}\,\tilde{\mathbf{h}}_{\mathrm{z}}(k) \tag{14}$$

und mit dem Spektrum des Mikrofonsignals multipliziert

$$\hat{\mathbf{s}}_{\mathrm{f}}(k) = \mathrm{diag}\{\tilde{\mathbf{h}}_{\mathrm{f}}(k)\}\,\mathbf{x}_{\mathrm{f}}(k), \tag{15}$$

um den geräuschreduzierten Spektralvektor $\hat{\mathbf{s}}_{\mathrm{f}}(k)$ zu erhalten. Nach der Rücktransformation des Vektors in den Zeitbereich entsprechen die letzten L Koeffizienten dem verarbeiteten Block im Zeitbereich

$$[\hat{s}(kL-L), \ldots, \hat{s}(kL-1)] = \mathbf{C}\,\mathbf{F}^{-1}\,\hat{\mathbf{s}}_{\mathrm{f}}(k). \tag{16}$$

In Abb. 2 ist die Teilbandgeräuschreduktion, beschrieben durch die Gl. (8) – (16), abgebildet.

3.2 Näherung des Projektionsfilters

Im vorangegangenen Abschnitt wurde eine *overlap-save*-basierte Signalverarbeitung vorgestellt. Die Rücktransformation in den Zeitbereich und die Fensterung mit einem Rechteckfenster ist notwendig, um die zyklischen Faltungseffekte zu unterdrücken. Zusätzlich müssen die Filterkoeffizienten verschoben werden, um eine kausale Impulsantwort zu erhalten. Die Gewichtung mit einem Rechteckfenster (Multiplikation im Zeitbereich) kann alternativ recheneffizient im Frequenzbereich durch eine zyklische Faltung mit einem geeigneten Projektionsfilter durchgeführt werden.

Die schnelle Faltung kann so abgeändert werden, dass das Filter im Zeitbereich die Nullen in der Mitte haben darf, um darauf ein zu $\frac{M}{2}$ symmetrisches Fenster anwenden zu können. Hierfür muss anstatt der Filterkoeffizienten $\mathbf{h}_{\mathrm{z}}(k)$ der Block des Eingangssignals $\mathbf{x}_{\mathrm{z}}(k)$ durch Multiplikation mit der Matrix $\mathbf{R}$ zyklisch verschoben werden

$$\mathbf{x}'_{\mathrm{f}}(k) = \mathbf{F}\,\mathbf{R}\,\mathbf{x}_{\mathrm{z}}(k). \tag{17}$$

Das Ausgangssignal bleibt somit unverändert. Daraus ergibt sich die Entwurfsvorschrift für das neue Rechteckfenster

$$\mathbf{w}_{\mathrm{z}} = [1_0, \ldots, 1_{\frac{N}{2}-1}, 0_{\frac{N}{2}}, \ldots, 0_{\frac{N}{2}+L-1}, 1_{\frac{N}{2}+L}, , \ldots, 1_{M-1}]. \tag{18}$$

Dieses Projektionsfilter $\mathbf{w}_{\mathrm{z}}$ sollte so abgeändert werden, dass die zyklische Faltung im Frequenzbereich mit diesem Fenster möglichst recheneffizient durchgeführt werden kann. Hierfür wird das Projektionsfilter $\mathbf{w}_{\mathrm{z}}$ durch ein cosinusartiges symmetrisches Projektionsfilter $\mathbf{w}_{\mathrm{z,soft}}$, bzw. ein angenähertes Projektionsfilter $\hat{\mathbf{w}}_{\mathrm{z,soft}}$ (besitzt eine geringe Anzahl von Null verschiedener Koeffizienten im Frequenzbereich) ersetzt.

Es können je nach Systemanforderungen (Sprachqualität, Laufzeit, Rechenkomplexität) verschiedene symmetrische cosinusartige Fenster entworfen werden. Folgende Parameter können für den Entwurf modifiziert werden:

- q_1 - Breite des Rechteckanteils vorne,
- q_2 - Breite der Cosinusflanke,
- q_3 - Breite des Nullanteils bis zur Mitte im Zeitbereich,
- q_{4-6} - die dazugehörigen symmetrischen Anteile,
- Q - die Anzahl der verwendeten Koeffizienten für die Näherung im Frequenzbereich.

Dabei gilt für die Parameter des Projektionsfilters: $q_1 + q_2 = P$, $q_3 = M/2 - P$ und $q_3 + q_4 \geq L$. Von der Wahl der Parameter P und L hängt die Verzögerung D der Teilbandgeräuschreduktion ab: $D = L + P$. Das mit den Parametern q_{1-6} entworfene Projektionsfilter $\mathbf{w}_{\mathrm{soft}}$ wird nach dem Entwurf im Zeitbereich in den Frequenzbereich transformiert, um anschließend die symmetrischen reellwertigen Teilbandkoeffizienten

$\mathbf{w}_{\mathrm{f,soft},\frac{Q+1}{2},\ldots,M-\frac{Q-1}{2}-1} = 0$ zu setzen. Dadurch erhält man das angenäherte Projektionsfilter $\hat{\mathbf{w}}_{\mathrm{f,soft}}$ mit wenigen Koeffizienten ungleich Null für eine recheneffiziente zyklische Faltung. Das Projektionsfilter $\hat{\mathbf{w}}_{\mathrm{f,soft}}$ sollte trotz der wenigen Koeffizienten im Frequenzbereich sowohl im Durchlass, wie auch im Sperrbereich möglichst gut dem urspünglichen Filter (insbesondere jenem Bereich, der zuvor auf Null gesetzt wurde) $\mathbf{w}_{\mathrm{z}}$ angenähert sein, um die zyklischen Faltungseffekte zu unterdrücken. Abb. 3 zeigt einige mögliche Projektionsfilter mit verschiedenen Parametern.

Durch den Gebrauch des Projektionsfilters $\hat{\mathbf{w}}_{\mathrm{f,soft}}$ werden die Gl. (10) – (14) ersetzt durch eine zyklische Faltung

$$\tilde{\mathbf{h}}_{\mathrm{f}}(k) = \hat{\mathbf{w}}_{\mathrm{f,soft}} \circledast \mathbf{h}_{\mathrm{f}}(k). \tag{19}$$

Vergleicht man die verarbeiteten Signale der *overlap-save*-basierten Filterbank mit denen der *overlap-add* basierten Filterbank fällt auf, dass die Erstgenannten kratzende Artefakte beinhalten. Sie entstehen durch die aneinander gesetzten Signalblöcke. Dieses wird durch das überlappende Aufaddieren bei einer *overlap-add*-basierten Verarbeitung unterdrückt. Die Artefakte können bei einer *overlap-save*-basierten Verarbeitung durch eine zeitliche Glättung der Teilbandfilterkoeffizienten ausreichend gut unterdrückt werden

$$\tilde{\mathbf{h}}'_{\mathrm{f}}(k) = \beta\, \tilde{\mathbf{h}}'_{\mathrm{f}}(k-1) + (1-\beta)\, \tilde{\mathbf{h}}_{\mathrm{f}}(k). \tag{20}$$

Die Wahl der Glättungskonstanten β aus dem Intervall $0 \leq \beta \leq 1$ ist wichtig für die Signalqualität des verarbeiteten Signals. Bei einem zu großen Wert sind die Artefakte deutlich hörbar, bei einem zu kleinen Wert klingt das Signal durch die Glättung der Filterkoeffizienten hallig. Aus subjektiven Tests haben sich Werte im Bereich von $\beta \approx 0.3$ als geeignet herausgestellt. Somit lässt sich der Ausgangsblock des verarbeiteten Signals im Frequenzbereich wie folgt bestimmen

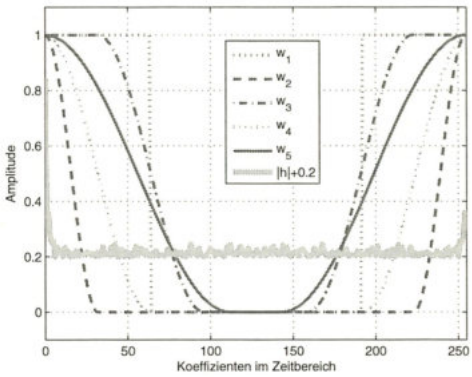

Abbildung 3 - Verschiedene Projektionsfilter im Zeitbereich mit folgenden Parametern ($L = 32$): w_1 mit $q_1 = 64$, $q_2 = 0$, $q_3 = 64$, $D = 96$; w_2 mit $q_1 = 0$, $q_2 = 32$, $q_3 = 96$, $D = 64$; w_3 mit $q_1 = 32$, $q_2 = 64$, $q_3 = 32$, $D = 128$; w_4 mit $q_1 = 0$, $q_2 = 64$, $q_3 = 64$, $D = 96$; w_5 mit $q_1 = 0$, $q_2 = 112$, $q_3 = 16$, $D = 144$. Zusätzlich ist noch ein Beispiel für den Betrags der Filterkoeffizienten im Zeitbereich $|h|$, zur besseren Darstellung um 0.2 nach oben versetzt, dargestellt.

$$\hat{\mathbf{s}}'_{\mathrm{f}}(k) = \mathrm{diag}\{\tilde{\mathbf{h}}'_{\mathrm{f}}(k)\}\, \mathbf{x}'_{\mathrm{f}}(k). \tag{21}$$

4 Evaluierung

Für die Evaluierung wurde eine Filterbank mit folgenden Parametern verwendet: Abtastrate $f_{\mathrm{s}} = 22050$ Hz, FFT-Blockgröße $M = 256$ Abtastwerte, Rahmenversatz bzw. Blockgröße des verarbeiteten Signals $L = 32$ Abtastwerte. Untersucht wurden die Rechenkomplexität, die durch die Signalverarbeitung eingefügte Laufzeit und die Sprachqualität. Miteinander verglichen wurden eine *overlap-add* basierte FFT-Filterbank mit Hannfenster und zwei verschiedene *overlap-save*-basierten FFT-Filterbänke mit Projektionsfiltern. Um die tonalen Reststörungen zu unterdrücken wurde bei der Verarbeitung das Störgeräusch um den Faktor 3 überschätzt, die maximale Sperrdämpfung beträgt 10 dB [5]. Die *overlap-save*-FFT-Filterbänke wurden für zwei verschiedene Projektionsfilter untersucht, ein Rechteckfenster $\mathbf{w}_{\mathrm{z}}$ ($q_1 = 112$, $q_2 = 0$, $q_3 = 16$, $D = 144$) und ein cosinusartiges Fenster $\hat{\mathbf{w}}_{\mathrm{z,soft}}$ ($q_1 = 0$, $q_2 = 64$, $q_3 = 64$, $D = 96$, $Q = 7$). Abb. 4 zeigt die untersuchten Projektionsfilter. Um eine recheneffiziente zyklische Faltung im Frequenzbereich durchzuführen wurde das Projektionsfilter $\mathbf{w}_{\mathrm{f,soft}}$ durch eine Näherung $\hat{\mathbf{w}}_{\mathrm{f,soft}}$ ersetzt

$$\hat{\mathbf{w}}_{\mathrm{f,soft}} = [65, 54.98, 31.98, 10.45, 0, ..., 0, 10.45, 31.98, 54.98]. \tag{22}$$

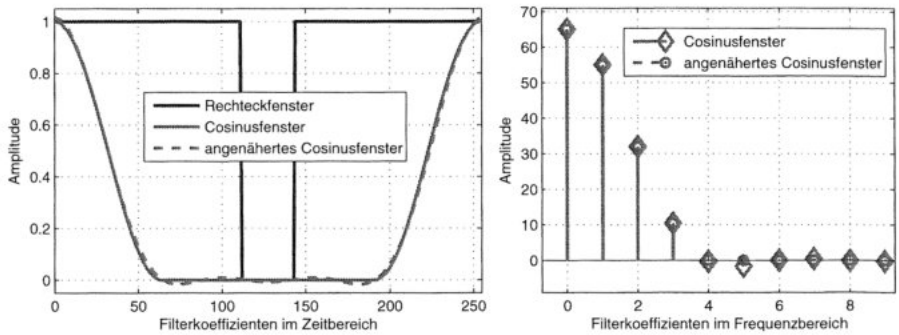

Abbildung 4 - Projektionsfilter $\mathbf{w}_z, \mathbf{w}_{z,\text{soft}}$ und $\hat{\mathbf{w}}_{z,\text{soft}}$ im Zeitbereich und Filterkoeffizienten $\mathbf{w}_{f,\text{soft}}$ und $\hat{\mathbf{w}}_{f,\text{soft}}$ im Frequenzbereich.

Tab. 1 enthält algorithmische Komplexität und die Laufzeit für die untersuchten Filterbankstrukturen. Für die Berechnung der Komplexität wurden bereits verschiedene Optimierungen (Radix-2-FFT, reellwertige Eingangssignale, symmetrisches Spektrum, usw.) berücksichtigt. Die Anzahl der Multiplikationen und Additionen wurde getrennt betrachtet und setzt sich zusammen aus den Rechenoperationen der FFT und den zusätzlich notwendigen Operationen [2]. Die Berechnung der Filterkoeffizienten wurde nicht berücksichtigt. Die Laufzeit setzt sich zusammen aus der Verzögerung der Verarbeitung und der Blockbildung. Die Evaluierung zeigt, dass sich für die Methode (C) im Vergleich zu (A) eine deutlich geringere Laufzeit, bei einer nicht wesentlich erhöhten Rechenkomplexität ergibt. Für die Methode (B) wird zwar auch die Laufzeit reduziert, jedoch ist die Rechenkomplexität im Vergleich zu (A) gestiegen. Für die Untersuchung der Sprachqualität wurden die verschiedenen Geräuschreduktionsstrukturen mit

Tabelle 1 - Algorithmische Komplexität pro Abtastwert für verschiedene Realisierungen einer Teilbandgeräuschreduktion (Index: ana, syn - Multiplikation mit dem Analyse- und Synthesefenster, weight - Multiplikation mit dem Geräuschreduktionsfilter, add - Aufaddieren überlappender Anteile, pro - Multiplikation mit dem Projektionsfilter, conv - Faltung mit dem Projektionsfilter).

	FFT	Zusätzliche Operationen	Summe
(A) *overlap-add*-FFT-FB			
Multiplikationen	$2\frac{M}{L}\log_2 M$	$\frac{1}{L}(M_{\text{ana}} + M_{\text{syn}} + M_{\text{weight}})$	152
Summationen	$3\frac{M}{L}\log_2 M$	$\frac{1}{L}(M_{\text{add}})$	200
Laufzeit in Abtastwerten			288
(B) *overlap-save*-FFT-FB mit $\mathbf{w}_z$			
Multiplikationen	$4\frac{M}{L}\log_2 M$	$\frac{1}{L}(M_{\text{pro}} + M_{\text{weight}})$	272
Summationen	$6\frac{M}{L}\log_2 M$		384
Laufzeit in Abtastwerten			144
(C) *overlap-save*-FFT-FB mit $\hat{\mathbf{w}}_{f,\text{soft}}$			
Multiplikationen	$2\frac{M}{L}\log_2 M$	$\frac{1}{L}(M_{\text{weight}} + \frac{M}{2}Q_{\text{conv}})$	164
Summationen	$3\frac{M}{L}\log_2 M$	$\frac{1}{L}(\frac{M}{2}Q_{\text{conv}})$	220
Laufzeit in Abtastwerten			96

einem CMOS-Test (*complementary mean opinion score*) evaluiert. Hierfür wurde von einer Gruppe von 8 Experten jeweils 15 verarbeitete Sprachäußerungen in einem direkten Vergleich bewertet. Die Tabelle 2 zeigt die Ergebnisse dieser Evaluierung. Die Signale wurden dabei über Kopfhörer angehört und nicht über ein ICC-System im Fahrzeug wiedergegeben, so dass in diesem Fall die Laufzeit unberücksichtigt blieb. Der CMOS-Test belegt, dass die Sprachqualität bei allen 3 Methoden vergleichbar ist, es gab keine Präferenz. Fasst man die Ergebnisse zusammen, dann ist die Methode (C) für Anwendungen mit gerin-

Tabelle 2 - Ergebnisse des CMOS-Tests für die Sprachqualität des verarbeiteten Signals in Prozent.

	X=(A) vers. Y=(B) in %	X=(A) vers. Y=(C) in %	X=(B) vers. Y=(C) in %
X viel schlechter als Y	0	0	0
X schlechter als Y	5	2,5	0
X ist wenig schlechter als Y	22,5	17,5	15
X und Y sind gleich gut	60	50	57,5
X wenig besser als Y	5	17,5	20
X besser als Y	7,5	12,5	7,5
X viel besser als Y	0	0	0

ger Verzögerung gut geeignet. Es konnte gezeigt werden, dass diese bei geringerer Laufzeit und gleicher Rechenkomplexität dieselbe Sprachqualität wie die Methode (A) erreicht.

5 Zusammenfassung und Ausblick

Die Störungen im Mikrofonsignal eines ICC-Systems müssen durch eine geeignete Geräuschreduktion unterdrückt werden. Dabei müssen die Signalqualität, die Rechenkomplexität und die Signalverzögerung berücksichtigt werden. Es konnte gezeigt werden, dass sich eine *overlap-save*-basierte FFT-Filterbank gut dazu eignet. Durch die Verwendete *overlap-save*-Struktur konnte die Laufzeit und durch die Wahl eines geeigneten Projektionsfilters die Rechenkomplexität reduziert werden. Das zur Unterdrückung der zyklischen Faltungseffekte notwendige Projektionsfilter wurde hergeleitet und ein einfaches Entwurfsverfahren für ein symmetrisches, cosinusförmiges Filter beschrieben. Die Untersuchung der verschiedenen Strukturen hat gezeigt, dass durch eine *overlap-save*-basierte Verarbeitung die Laufzeit erheblich reduziert werden kann. Durch die zyklische Faltung mit dem angenäherten Projektionsfilter im Frequenzbereich konnte zusätzlich die Rechenkomplexität reduziert werden. Die Untersuchung der Sprachqualität zeigte, dass die Reduktion der Laufzeit und der Rechenkomplexität sich nicht negativ auf die Sprachqualität des verarbeiteten Signals auswirken.

Bei dem hier vorgestellten Entwurfsverfahren für das angenäherte Projektionsfilter wird zunächst die Laufzeit durch die Wahl der Parameter q_{1-3} festgelegt und dann die Rechenkomplexität durch den Parameter Q bestimmt. Es können aber noch weitere geeignete Projektionsfilter nach zusätzlichen Optimierungskriterien, z.B. Verzerrung der Filterkoeffizienten im Durchlass- und Sperrbereich, untersucht werden. Alternativ können auch Strukturen, welche eine Kombination der *overlap-save*- und *overlap-add*-basierten Verarbeitung vereinen, betrachtet werden, um die Vorteile der beiden Methoden zu kombinieren.

Literatur

[1] T. Haulick, G. Schmidt: *Signal processing for in-car communication systems*, Signal Processing, Seiten 1307 – 1326, Juni 2006.

[2] H. W. Löllmann P. Vary: *Efficient non-uniform filter-bank equalizer*, Proc. EUSIPCO '05, Antalya, Turkey, September 2005.

[3] J. J. Shynk: *Frequency-domain and multirate adaptive filtering*, IEEE Signal Processing Mag., Seiten 14 – 35, Januar 1992.

[4] E. Hänsler, G. Schmidt: *Acoustic Echo and Noise Control: A Practical Approach*, John Wiley & Sons, 2004.

[5] P. Vary, R. Martin: *Digital Speech Transmission: Enhancement, Coding and Error Concealment*, John Wiley & Sons, 2006.

[6] P. Vary, U. Heute, W. Hess: *Digitale Sprachsignalverarbeitung*, Teubner, Stuttgart, 1998.