

# FFT-FILTERBANK FÜR HÖRGERÄTE-ANWENDUNGEN - REDUKTION VON ZEITBEREICHS-ALIASING

Uwe Raß<sup>1)</sup>, Bernd Beigel<sup>1)</sup>, Gerhard H. Steeger<sup>2)</sup>

<sup>1)</sup> Siemens Audiologische Technik GmbH, Erlangen, <sup>2)</sup> Georg-Simon-Ohm-Fachhochschule Nürnberg  
Uwe.Rass@med.siemens.de, Bernd.Beigel@med.siemens.de, Gerhard.Steeger@fh-nuernberg.de

## Kurzfassung

Filterbänke lassen sich mit relativ geringem Rechenaufwand mit Hilfe der FFT-Transformation und des Overlap-Add-Verfahrens realisieren. Die berechneten Frequenzkanäle besitzen konstante Bandbreite und können durch geeignete Zusammenfassung z.B. an die auditorischen Filtern angepaßt werden. Die mit dem Overlap-Add-Verfahren verbundene Unterabtastung der Kanalsignale ermöglicht eine Reduktion des Rechenaufwands für den Hörgeräte-Algorithmus. Dieser Algorithmus berechnet im Frequenzbereich eine dem Signal angepaßte Übertragungsfunktion. Die Anwendung des Overlap-Add-Verfahrens fordert nun, daß die zugehörige Impulsantwort in der Länge beschränkt ist. Dies kann aber i.a. beim Entwurf im Frequenzbereich nicht berücksichtigt werden. Bei Nichteinhaltung dieser Bedingung entstehen Verzerrungsprodukte (Aliasing) im Ausgangssignal, die im vorliegenden Beitrag untersucht werden. Es wird ein Verfahren vorgestellt, wie mit vertretbarem Aufwand die Verzerrungen reduziert werden können. In einem Hörtest mit 5 normalhörenden Versuchspersonen wird untersucht, in wie weit die Aliasing-Verzerrungen hörbar sind.

## 1 Einleitung

Die Bezeichnung "Overlap-Add" wurde von Stockham [5] für zeitinvariante FIR Filterung von zeitlich unbeschränkten Signalen eingeführt. Allen und Rabiner [1] erweiterten Stockhams Idee auf den Fall adaptiver bzw. zeitvarianter Filterung von nichtstationären Signalen. Wichtige Anwendungen wie z.B. Dynamikkompensation von Sprache oder Störgeräuschreduktion wurden bereits von Allen und Rabiner genannt. Jedoch birgt die Verallgemeinerung des Overlap-Add-Verfahrens auf adaptive Filterung auch Nachteile. In vielen Fällen kann die Adaptionsregel für diese Filter nur im Frequenzbereich formuliert werden, weil dazu z.B. das Kurzzeit-Spektrum benötigt wird. Die zur wirksamen adaptiven Übertragungsfunktion gehörende Länge der Impulsantwort läßt sich leider nicht mit einfachen Mitteln bestimmen. Eine Beschränkung der Länge dieser Impulsantwort ist allerdings notwendig, um eine verzerrungsfreie Signalverarbeitung auf Basis des Overlap-Add-Verfahrens zu realisieren.

## 2 Entstehung von Aliasing-Verzerrungen

Im Overlap-Add-Verfahren wird zunächst das Eingangssignal  $v(k)$  in Blöcke der Länge  $N$  geteilt, wobei sich aufeinander folgende Blöcke um  $N_0=N/K$  Abtastwerte überlappen [1]. Jeder Block wird mit einer Hann-Fensterfunktion [2] multipliziert und mit einer FFT der Länge  $M$  in den Frequenzbereich transformiert. Das so berechnete Kurzzeit-Spektrum  $V_p(\mu)$  des Eingangssignals wird mit der adaptiven Übertragungsfunktion  $H_p(\mu)$  (mit  $\mu=0, 1, \dots, M-1$ ) multipliziert. Der Index  $p$  bezeichnet die laufende Blocknummer. Nach Berechnung der inversen FFT werden die Zeitsignalblöcke mit der gleichen Überlappung  $N_0$  aufaddiert. Die verschobenen und aufaddierten Fensterfunktionen müssen sich zu 1 ergänzen, um das Signal nicht zu verzerrern. Diese Eigenschaft ist bei sog. Raised-Cosine-Fenstern wie dem Hann-Fenster und entsprechender Wahl von  $N_0$  gegeben.

Die Impulsantwort  $h_p(k)$  entsteht durch inverse FFT-Transformation der adaptiven Übertragungsfunktion  $H_p(\mu)$ . Aliasing-Verzerrungsprodukte werden im Ausgangssignal  $y(k)$  auftreten, falls die Länge  $L$  dieser Impulsantwort  $h_p(k)$  die folgende Bedingung verletzt:

$$L \leq M - N + 1 \quad (1)$$

Wird nun die Übertragungsfunktion  $H_p(\mu)$  im Frequenzbereich durch Anwendung einer Adaptionsregel (z.B. Mehrkanal-Dynamikkompensation) modifiziert, läßt sich die Bedingung (1) nicht ohne weiteres nachprüfen. Im allgemeinen wird die Länge  $L$  der Impulsantwort nicht der Gleichung (1) genügen. Die Aliasing-Verzerrungen können klein gehalten werden, wenn die FFT-Länge  $M$  sehr viel größer gewählt

wird als die Fensterlänge  $N$ . Dies erhöht allerdings den Rechenaufwand für den Hörgeräte-Algorithmus erheblich.

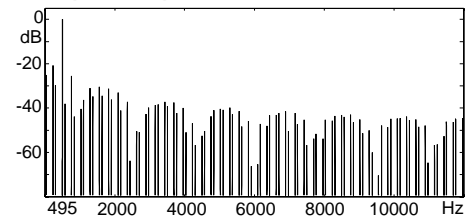


Bild 2: Ausgangsspektrum mit Aliasingverzerrungen für ein sinusförmiges Eingangssignal.

Durch Zeitbereichs-Aliasing werden nicht-harmonische Verzerrungsprodukte im gesamten Frequenzbereich erzeugt. Bild 1 zeigt z.B. das Ausgangsspektrum bei sinusförmiger Anregung mit der Frequenz 495 Hz. Dabei wurde eine konstante Übertragungsfunktion eingestellt, die das Frequenzband von 500 Hz bis 1 kHz um 40 dB absenkt. Man sieht, daß Verzerrungsprodukte bis zur Nyquist-Frequenz von 12 kHz vorkommen. Bei Signalen mit einem breiteren Spektrum, wie z.B. bei Sprache, könnten die Verzerrungsprodukte maskiert sein. Die Frage der Hörbarkeit wird in Abschnitt 4 weiter untersucht.

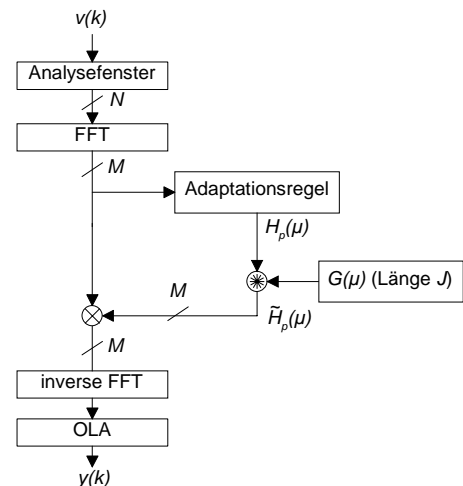


Bild 2: Blockschaltbild des modifizierten Overlap-Add-Verfahrens.

## 3 Verfahren zur Reduktion von Aliasing-Verzerrungen

Um Probleme mit Zeitbereichs-Aliasing zu vermeiden, müßte man nach jeder Änderung der adaptiven Übertragungsfunktion einen Filterentwurf durchführen, der die Erfüllung der Bedingung (1) sicherstellt. Dazu könnte man  $H_p(\mu)$  in den Zeitbereich transformieren und mit einer passenden Fensterfunktion auf die Länge  $L$  begrenzen [4]. Jedoch verlangt dieses Vorgehen einen sehr großen Rechenaufwand. Deshalb schlagen wir vor, die Approximation im Frequenzbereich durchzuführen (Bild 2). Die Fensterung von  $h_p(k)$  im Zeitbereich entspricht dabei der zyklischen Faltung von  $H_p(\mu)$  mit dem Spektrum  $G(\mu)$  der Fensterfunktion. Da  $G(\mu)$  i.a.  $M$  von Null verschiedene Werte hat und damit eine zyklische Faltung mit  $H_p(\mu)$   $M^2$  Multiplikationen erfordern würde, bringt das zunächst noch keinen Vorteil. Entwirft man aber die Funktion  $G(\mu)$  direkt im Frequenzbereich mit den Formeln, die eigentlich für den Zeitbereich angegeben werden (z.B. [2]), so kann man die Anzahl  $J$  von Null verschiedener Koeffizienten vorgeben. Dieses Vorgehen ist aufgrund der Dualität von Zeit- und Frequenzbereich erlaubt. Die so entworfenen Fensterfunktionen haben eine beschränkte Länge  $J$  im Frequenzbereich, folglich ist die Zeitbereichsdarstellung nicht zeitlich beschränkt, sondern zeigt Haupt- und Nebenkeulen. Um eine hohe Reduktion der Aliasing-Komponenten zu erreichen, sollten die

Nebenkeulen möglichst klein sein. Für diese Anwendung haben sich die sog. Discrete Prolate Spheroidal Sequences (DPSS; [6]) als optimal erwiesen, da sie die höchst mögliche Energiekonzentration in der Hauptkeule besitzen. Damit wird das Energieverhältnis von Nutz- zu Störsignal maximiert. Diese Fenster können leicht mit den Entwurfsparametern Anzahl der Koeffizienten im Frequenzbereich und Breite der Hauptkeule im Zeitbereich entworfen werden.

#### 4 Messung der Performance

Die Wirksamkeit des Verfahrens wurde mit einem Dynamikkompresionsalgorithmus mit 3 Frequenzkanälen und kurzen Zeitkonstanten untersucht. Das Fensterspektrum zur Aliasing-Reduktion hatte  $J=9$  von Null verschiedene Koeffizienten. Für das Overlap-Add-Verfahren wurde eine FFT-Länge von 256 und ein Hann-Fenster der Länge  $N=180$  verwendet. Aufeinanderfolgende Blöcke des Eingangssignal überlappten sich um  $N/2$ . Ein Sprachsignal wurde mit und ohne Zusatzalgorithmus zur Aliasing-Reduktion verarbeitet. Das jeweilige aliasing-freie Referenzsignal wurde durch Faltung der Eingangssignale mit den jeweiligen Impulsantworten  $h_p(k)$  im Zeitbereich berechnet. Durch Subtraktion des Referenzsignals von den Testsignalen erhält man den zeitlichen Verlauf der Aliasing-Komponenten, wie sie in Bild 3 dargestellt sind. Das linke Bild zeigt die Aliasing-Verzerrung im Ausgangssignal ohne, und das rechte Bild mit dem beschriebenen Zusatzalgorithmus. Man sieht eine deutliche Reduktion der Aliasing-Komponenten (hier: 27 dB).

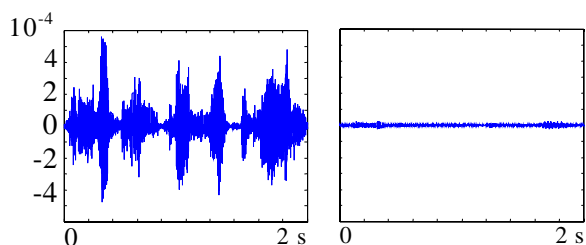


Bild 3: Aliasing-Verzerrungen eines Sprachsignals: links ohne, rechts mit Aliasing-Reduktionsverfahren nach Abschnitt 3.

#### 5 Subjektive Untersuchung

Die Aliasing-Signalenergie ist der Energie des Eingangssignals direkt proportional. Insofern unterscheidet sich diese Art von Verzerrungen von anderen additiven Störungen wie z.B. dem Mikrofonrauschen. Der Signal-Rauschabstand ist aus diesem Grund kein geeignetes Instrument, um die Qualität des Ausgangssignals zu beurteilen. Daher wurde das Verfahren in einem Hörversuch mit 5 normalhörenden Versuchspersonen untersucht.

Da die Signalverarbeitung erheblichen Einfluß auf die entstehenden Verzerrungsprodukte hat, wurde für alle Versuchspersonen die gleiche Einstellung des Dynamikkompresionsalgorithmus gewählt, der auch schon für die Messungen in Bild 3 verwendet wurde. Als Eingangssignale wurde männliche und weibliche Sprache in Ruhe, Sprechgesang ("Tom's Diner" von Suzanne Vega), Trompete und Flöte gewählt. Vor allem die Musiksignale werden als kritisch angesehen, da bei tonalen Anteilen nicht-harmonische Verzerrungen besonders störend wahrgenommen werden. Jedes Signal wurde mit 3 verschiedenen Einstellungen verarbeitet:

N	ohne Aliasing-Reduktionsverfahren
Z	mit Fensterung der verarbeiteten Zeitsignale nach der inversen FFT [7]
F	zyklische Faltung im Frequenzbereich nach Abschnitt 3

Das Verfahren Z reduziert nur Verzerrungen an den Rändern des Zeitsignals, benötigt aber im allgemeinen einen geringeren Rechenaufwand als das Verfahren F.

Der Hörversuch wurde gemäß der ITU-Empfehlung BS-1116-1 [3] durchgeführt. Diese Methode ist besonders gut geeignet, um kleine, kaum hörbare Signalverzerrungen zu evaluieren. Der Versuchsperson werden 3 Signale A, B und C in frei wählbarer Reihenfolge und so oft wie gewünscht dargeboten. Signal A stellt immer das aliasing-freie Referenzsignal dar. Dies ist der Versuchsperson bekannt. Das Testsignal und nochmals das Referenzsignal werden zufällig den Signalen B und C zugeordnet, d.h. 2 der Signale sind identisch. Die

Versuchsperson hat die Aufgabe, die Unterschiede zwischen Signal A und B sowie zwischen Signal A und C auf einer Skala von 1 bis 5 zu bewerten. Die Ankerpunkte der kontinuierlichen Skala haben folgende Bedeutung:

5	keine hörbaren Unterschiede
4	hörbar, aber nicht störend
3	leicht störend
2	störend
1	sehr störend

Da der Versuchsperson bekannt ist, daß entweder Signal B oder C identisch mit A ist, wird sie einmal die Bewertung 5 vergeben. Wenn keine Unterschiede zwischen den 3 Signalen gehört werden, dürfen trotzdem B und C nicht beide mit 5 bewertet werden. Die Versuchspersonen sind dadurch angehalten, Referenz- und Testsignal zu raten. Ausgewertet wird die Differenz der Bewertungen zwischen dem Testsignal und der versteckten Referenz. Dieser Score ist negativ, falls das Testsignal schlechter bewertet wurde als das Referenzsignal. Tritt eine Verwechslung auf, d.h. wird das Testsignal fälschlicherweise als versteckte Referenz erkannt, so ergeben sich positive Scores. Sind keine Unterschiede zwischen Referenz- und Testsignal hörbar, ergibt sich im Mittel der Score 0.

Bild 4 zeigt die Ergebnisse des Hörversuchs für die beiden Aliasing-Reduktionsverfahren und für das Overlap-Add-Verfahren ohne Zusatzalgorithmus. Die paarweisen Unterschiede sind auf einem Signifikanzniveau von 1% jeweils signifikant (paarweiser Wilcoxon-Rangsummentest mit Korrektur für Mehrfachvergleiche [8]). Wird kein Aliasing-Reduktionsverfahren verwendet, ergibt sich ein mittlerer Score von -2, d.h. die Verzerrungen werden als leicht störend bewertet. Die Qualität der beiden Ausgangssignale bei Anwendung der Aliasing-Reduktionsverfahren Z und F wird deutlich besser bewertet. Verfahren Z mit einem mittleren Score von -0,2 zeigt einen deutlich größeren Interquartilabstand als das Verfahren F (Median-Score +0,1). D.h. es gibt Eingangssignale, für die hörbare Aliasing-Verzerrungen auftreten, wenn das Verfahren Z benutzt wird. Wird das hier vorgestellte Verfahren F angewandt, so sind die Aliasing-Komponenten praktisch nicht hörbar.

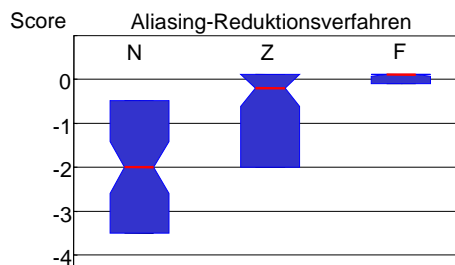


Bild 4: Ergebnis des Hörversuchs für die 3 untersuchten Einstellungen zur Aliasing-Reduktion.

#### Literatur

- [1] Allen, J.B., Rabiner, L.R. (1977). A unified approach to short-time Fourier analysis and synthesis. Proc. of the IEEE, Vol. 65, S. 1558-1564.
- [2] Harris, F.J. (1978). On the use of windows for harmonic analysis with the Discrete Fourier Transform. Proc. of the IEEE, Vol. 66, Nr. 1.
- [3] ITU (1997): Methods for the subjective assessment of small impairments in audio systems including multichannel sound systems. Rec. ITU-R-BS-1116-1.
- [4] Oppenheim, A.V., Schaffer, R.W. (1989). Discrete-time signal processing. Prentice Hall, Englewood Cliffs, NJ, USA.
- [5] Stockham, T.G. jr. (1966). High speed convolution and correlation. Spring Joint Computer Conf., AFIPS Proc. 28, S. 229-233.
- [6] Verma, T., Bilbao, S., Meng, T.H.Y. (1996). The Digital Prolate Spheroidal Window. Proc. Intl. Conf. on Acoustics, Speech and Signal Processing, ICASSP'96, Atlanta, GE, USA.
- [7] Wittkop, T. (1999). persönliche Mitteilung.
- [8] Wright, S. P. (1992). Adjusted p-values for simultaneous inference. Biometrics Nr. 48, S. 1005-1013.