

Codierung von Audiosignalen mit überlappender Transformation und adaptiven Fensterfunktionen

Coding of Audio Signals with Overlapping Block Transform and Adaptive Window Functions

Von Bernd Edler

*Beitrag aus dem Institut für Theoretische Nachrichtentechnik und Informationsverarbeitung
der UNIVERSITÄT HANNOVER, den Professor Dr.-Ing. H. G. Musmann und seine Mitarbeiter
Herrn Prof. Dr. phil. nat. Dietrich Wolf zum 60. Geburtstag widmen*

Übersicht:

Das hier vorgestellte Verfahren der Bitratenreduktion für Audiosignale basiert auf überlappenden Transformationen mit „Time Domain Aliasing Cancellation“, deren Fensterfunktionen und Transformationslängen in Abhängigkeit vom Eingangssignal umgeschaltet werden. Die adaptive Fensterung verbessert das Verhalten der Transformationscodierung mit überlappenden Blöcken, die sich durch einen hohen Codierungsgewinn auszeichnet, beim Auftreten von Impulsen und Amplitudensprüngen im Eingangssignal.

Abstract:

The method of bit rate reduction for audio signals presented in this paper is based on overlapping transforms with „Time Domain Aliasing Cancellation“. By adapting the window functions and the transform lengths to the input signal, the performance of the transform coder with overlapping blocks in presence of impulses and rapid attacks in the input signal is improved.

Für die Dokumentation:

Transformationscodierung / Transformation mit überlappenden Blöcken / adaptive Fensterfunktionen / Vorecho-Unterdrückung / Maskierungswirkung im Gehör

1. Einleitung

Bei der Bitratenreduktion für Tonsignale erweist sich die Verwendung von Transformationen mit überlappenden Blöcken und nicht-rechteckigen Fensterfunktionen als vorteilhaft. Sie ermöglichen z.B. einen hohen Codierungsgewinn bei korrelierten Eingangssignalen und eine gute Anpassung des Quantisierungsfehler-Spektrums an die spektrale Maskierungswirkung des Nutzsignals im menschlichen Gehör [4, 5]. Hierbei zeichnet sich die

„Overlapping Block Transform“ (OBT) dadurch aus, daß durch die Überlappung keine zusätzlichen Koeffizienten entstehen, das Eingangssignal aber dennoch durch das Prinzip des „Time Domain Aliasing Cancellation“ (TDAC) [1, 2] exakt aus den (nicht quantisierten) Koeffizienten wiedergewonnen werden kann.

Neben den Vorteilen eines hohen Codierungsgewinns und geringer Blockgrenzeffekte führt die OBT jedoch den

Nachteil der Verlängerung der sogenannten Vorechos (Quantisierungsfehler, die sprunghaften Anstiegen der Signalamplitude vorausgehen) mit sich.

Durch Überlagerung und Addition (Overlap-Add) der Ausgangssignale aufeinanderfolgender, um N Abtastwerte versetzter Blöcke ergibt sich folgendes Ausgangssignal im Überlappungsbereich (hier: $0 \leq n \leq N-1$):

$$y(n) = f^2(N+n) \cdot x(n) + f(N+n) \cdot f(2N-1-n) \cdot x(N-1-n) + f^2(n) \cdot x(n) - f(n) \cdot f(N-1-n) \cdot x(N-1-n). \quad (4)$$

In diesem Beitrag wird eine Codierung von Tonsignalen mit überlappender Transformation und adaptiven Fensterfunktionen vorgestellt. Dieses Verfahren nutzt die Vorteile der Transformationen nach dem TDAC-Prinzip und bietet gleichzeitig ein günstiges Verhalten beim Auftreten von Impulsen und Amplitudensprüngen im Eingangssignal.

Nach einem kurzen Überblick über die Theorie des TDAC wird gezeigt, daß die Fensterfunktionen der OBT von Block zu Block variiert werden können. Diese Umschaltung der Überlappung ermöglicht auch die Verwendung unterschiedlicher Transformationslängen in aufeinanderfolgenden Blöcken. Für die Steuerung der Überlappung wird anschließend ein Verfahren mit einer Vorab-Analyse des Eingangssignals vorgestellt. Es schaltet bei Sprüngen der Signalamplitude von maximaler Überlappung auf geringe Überlappung und von großen auf kleine Transformationslängen um.

Abschließend werden Ergebnisse der Transformationscodierung mit adaptiver Fensterung bei einer Bitratenreduktion von 16 auf 2,5 bit pro Abtastwert vorgestellt und mit den Ergebnissen bei Verwendung fester Fensterlängen verglichen.

2. Überblick über die Theorie der Transformationen mit überlappenden Blöcken und „Time Domain Aliasing Cancellation“

In [1, 2] werden Transformationen vorgestellt, die überlappende, nicht-rechteckige Fensterfunktionen ermöglichen, ohne daß durch die Überlappung zusätzliche Transformationskoeffizienten entstehen. Bei einer derartigen Transformation der Länge N werden zunächst $2N$ Werte des Eingangssignals $x(n)$ mit einer Fensterfunktion $f(n)$ multipliziert und nach der Transformation im Frequenzbereich um den Faktor 2 unterabgetastet. Dies führt zu folgender Transformationsgleichung:

$$X(k) = \sqrt{2/N} \cdot \sum_{n=0}^{2N-1} f(n) \cdot x(n) \cdot \cos[\pi(k + \frac{1}{2})(n + N/2 + \frac{1}{2})/N]. \quad (1)$$

Die Rücktransformation bildet die N Transformationskoeffizienten auf $2N$ Abtastwerte ab, die wiederum mit $f(n)$ multipliziert werden:

$$x'(n) = \sqrt{2/N} \cdot f(n) \cdot \sum_{k=0}^{N-1} x(k) \cdot \cos[\pi(k + \frac{1}{2})(n + N/2 + \frac{1}{2})/N] \quad (2)$$

Das Ausgangssignal der Rücktransformation $x'(n)$ entspricht jedoch nicht mehr dem Eingangssignal $x(n)$:

$$x'(n) = \begin{cases} f^2(n) \cdot x(n) - f(n) \cdot f(N-1-n) \cdot x(N-1-n) & \text{für } 0 \leq n < N, \\ f^2(n) \cdot x(n) + f(n) \cdot f(3N-1-n) \cdot x(3N-1-n) & \text{für } N \leq n < 2N. \end{cases} \quad (3)$$

Dem mit $f^2(n)$ gewichteten Eingangssignal sind Störkomponenten überlagert, die als Aliasing im Zeitbereich bezeichnet werden können, da sie durch die Unterab-tastung im Frequenzbereich entstehen.

Fordert man für die exakte Rekonstruktion $y(n) = x(n)$, so müssen die beiden Aliaskomponenten eliminiert werden:

$$f(n) \cdot f(N-1-n) = f(N+n) \cdot f(2N-1-n). \quad (5)$$

Gleichzeitig müssen die beiden gewichteten Signalkomponenten in ihrer Summe das Eingangssignal wiedergeben:

$$f^2(n) + f^2(N+n) = 1. \quad (6)$$

Bedingung (5) läßt sich in Verbindung mit Bedingung (6) in eine einfache Symmetriebedingung überführen:

$$f(n) = f(2N-1-n). \quad (7)$$

Als Beispiele für Fensterfunktionen, die diese Bedingungen erfüllen, seien hier das Sinus-Fenster (**Bild 1a**), das Sinus-Rechteck-Fenster (**Bild 1b**) und als Sonderfall auch das Rechteckfenster (**Bild 1c**) aufgeführt.

Eine wichtige Eigenschaft der OBT ist die Äquivalenz der Transformationskoeffizienten zu Abtastwerten der Teilbandsignale einer modulierten Filterbank [1, 2], d. h. die Transformationskoeffizienten geben die Struktur des Eingangsspektrums wieder.

Im Vergleich zu nicht-überlappenden Transformationen zeichnet sich die OBT durch einen höheren Codierungsgewinn [6] und reduzierte Blockgrenzeffekte aus.

3. OBT mit adaptiven Fensterfunktionen

3.1 Umschaltung der Fensterform

Die OBT weist gegenüber nicht-überlappenden Transformationen den Nachteil längerer Vorechos auf. Vorechos sind Quantisierungsfehler im Ausgangssignal, die sprunghaften Anstiegen der Signalamplitude vorausgehen. Bei einer Transformationslänge von N beträgt deren maximale Dauer wegen der überlappenden Blöcke $2N$

Abtastwerte im Gegensatz zur maximalen Dauer von N Abtastwerten bei nicht-überlappenden Transformationen. Dadurch können sie bei gleicher Amplitude leichter wahr-

genommen werden. Durch die im folgenden beschriebene Anpassung der Fensterfunktionen an das Eingangssignal

lassen sich die Vorteile der OBT, nämlich der hohe Codierungsgewinn und die geringen Blockgrenzeffekte, mit dem Vorteil der kürzeren Vorechos nicht-überlappenden Transformationen vereinigen.

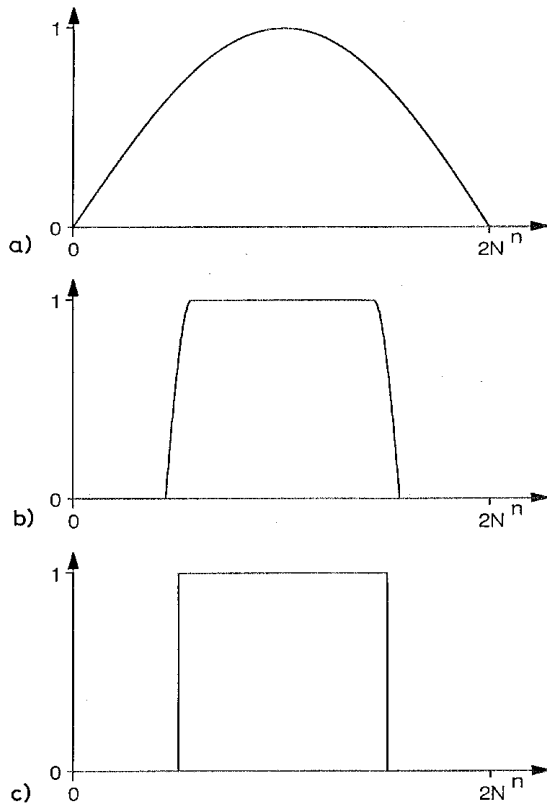


Bild 1: Fensterfunktionen für OBT

- a) Sinus-Fenster
b) Sinus-Rechteck-Fenster
c) Rechteck-Fenster

Verwendet man in zwei aufeinanderfolgenden Blöcken unterschiedliche Fensterfunktionen $f(n)$ und $g(n)$, so ergibt sich folgendes Ausgangssignal im Überlappungsbereich ($0 \leq n \leq N-1$):

$$y(n) = f^2(N+n) \cdot x(n) + f(N+n) \cdot f(2N-1-n) \cdot x(N-1-n) + g^2(n) \cdot x(n) - g(n) \cdot g(N-1-n) \cdot x(N-1-n). \quad (9)$$

Betrachtet man nun die Bedingungen für die exakte Rekonstruktion

$$f(N+n) \cdot f(2N-1-n) = g(n) \cdot g(N-1-n) \quad (10)$$

$$f^2(N+n) + g^2(n) = 1, \quad (11)$$

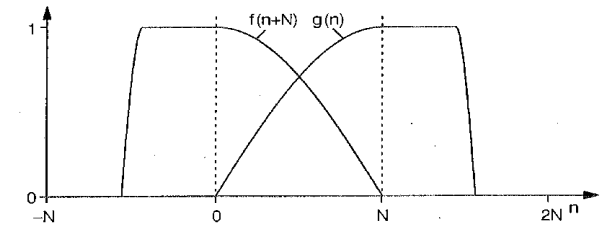


Bild 2: Unterschiedliche Fensterfunktionen in aufeinanderfolgenden Blöcken

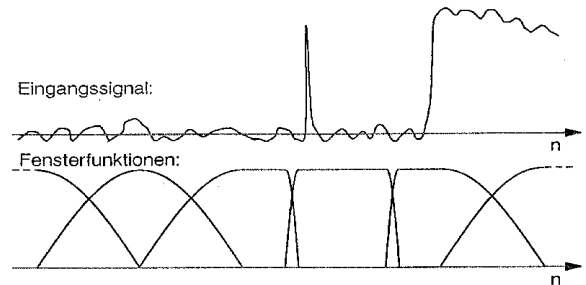


Bild 3: An das Eingangssignal angepaßte Fensterfunktionen

so fällt insbesondere auf, daß dabei nur jeweils die zweite Hälfte der Fensterfunktion des ersten Blocks und die erste Hälfte der Fensterfunktion des zweiten Blocks eingehen. Dies wiederum bedeutet, daß in jedem Block die beiden Hälften der Fensterfunktion unabhängig voneinander gewählt werden können, jedoch zusammen mit den zugehörigen Fensterhälften der Nachbarblöcke die Bedingungen (10) und (11) einhalten müssen. Somit kann die Fensterfunktion einer Transformation in einer Hälfte eine maximale und in der anderen Hälfte eine geringere Überlappung aufweisen (Bild 2).

Beim Auftreten von Impulsen oder Amplitudensprüngen im Eingangssignal können daher die Fensterfunktionen so umgeschaltet werden, daß der Impuls bzw. Sprung

immer in einen Block mit kurzen Überlappungsbereichen fällt (Bild 3).

3.2 Umschaltung der Transformationslängen

Verwendet man bei einer OBT der Länge N_1 eine Fensterfunktion mit einer Überlappungslänge von L Abtastwerten in der zweiten Fensterhälfte

$$f(n) = \begin{cases} 1 & N_1 \leq n \leq 3N_1/2 - L/2 - 1 \\ f_L(n - 3N_1/2 + L/2) & 3N_1/2 - L/2 \leq n \leq 3N_1/2 + L/2 - 1 \\ 0 & 3N_1/2 + L/2 \leq n \leq 2N_1 - 1, \end{cases} \quad (12)$$

so kann $g(n)$ auch die Fensterfunktion einer OBT mit anderer Länge N_2 sein

$$g(n) = \begin{cases} 0 & 0 \leq n \leq N_2/2 - L/2 - 1 \\ g_L(n - N_2/2 + L/2) & N_2/2 - L/2 \leq n \leq N_2/2 + L/2 - 1 \\ 1 & N_2/2 + L/2 \leq n \leq N_2 - 1, \end{cases} \quad (13)$$

und gewährleistet dennoch die exakte Rekonstruktion, wenn keine der beiden Transformationslängen kleiner als die Überlappungslänge ist, und wenn die Fensterfunktionen im gemeinsamen Überlappungsbereich die Bedingungen

$$f_L(n) \cdot f_L(L-1-n) = g_L(n) \cdot g_L(L-1-n) \quad (14)$$

$$f_L^2(n) + g_L^2(n) = 1 \quad (15)$$

erfüllen.

Hiermit wurde gezeigt, daß zusätzlich zur Länge der Überlappungsbereiche auch die Transformationslängen signaladaptiv umgeschaltet werden können. Die Verringerung der Frequenzauflösung (Anzahl der Koeffizienten) durch die Umschaltung auf kürzere Transformationslängen bei impuls- und sprungartigen Eingangssignalen bringt kaum Nachteile mit sich, da derartige Signale relativ breite und flache Spektren aufweisen.

4. Steuerung der Fensteradaption

Will man, wie oben beschrieben, die Fensterfunktionen dem Eingangssignal anpassen, so wird ein Verfahren zur Erkennung von Impulsen und Amplitudensprüngen im Eingangssignal benötigt. Im folgenden wird ein einfaches Verfahren vorgestellt, das vor der weiteren Verarbeitung direkt aus dem Eingangssignal die Information zur Steuerung der Fensterfunktionen ermittelt. Es detektiert nur das Auftreten eines Impulses oder Sprungs in einem Block und nicht die genaue Lage, die für die Steuerung nicht von Bedeutung ist.

Betrachtet man das Verhältnis der Signalenergien in zwei aufeinanderfolgenden Blöcken, so sieht man, daß dieses große Werte annimmt, wenn die Amplitude des Eingangssignals sprunghaft ansteigt. Bleibt die Amplitude des Eingangssignals näherungsweise konstant, so nimmt dieses Verhältnis Werte um 1 an. Der Einfachheit halber können die Energien durch die Betragssummen der Abtastwerte innerhalb der Blöcke ersetzt werden. Die Erkennbarkeit der Amplitudensprünge wird noch erhöht, wenn nicht die Beträge der Abtastwerte, sondern die Beträge der Ausgangswerte eines einfachen Hochpaßfilters summiert werden. Zur Hochpaßfilterung werden jeweils die Differenzen zweier aufeinanderfolgender Abtastwerte gebildet:

$$c_1(v) = \sum_{n=0}^{N-1} |d(vN+n)| \bigg/ \sum_{n=0}^{N-1} |d(vN-N+n)| \quad (16)$$

mit $d(n) = x(n) - x(n-1)$.

Der Verlauf dieser Funktion ist in Bild 4 für ein Triangel-Signal mit vier Anschlägen grafisch dargestellt.

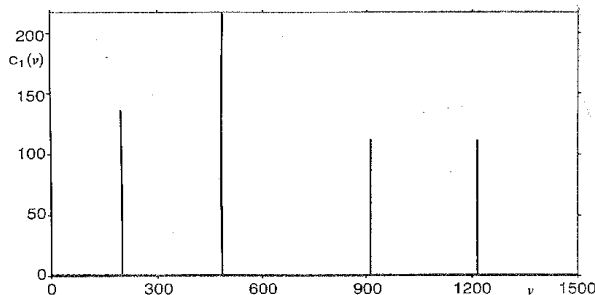


Bild 4: Kriterium zur Erkennung von Sprüngen der Signalamplitude

Sehr kurze Impulse und Amplitudensprünge am Blockende tragen jedoch nur wenig zur Gesamtenergie im Block bei und werden mit Hilfe von $c_1(v)$ nur schwer detektiert. Deshalb wird ein zweites Kriterium eingeführt, das den maximalen Betrag innerhalb eines Blocks zum Mittelwert der Beträge ins Verhältnis setzt:

$$c_2(v) = \max_{n=0}^{N-1} \{|x(vN+n)|\} \cdot N \bigg/ \sum_{n=0}^{N-1} |x(vN+n)|. \quad (17)$$

Die beiden vorgestellten Kriterien ergänzen sich und können die Fensterung so steuern, daß auf kurze Fensterlängen umgeschaltet wird, wenn einer der beiden Werte einen entsprechenden Schwellwert überschreitet.

5. Vergleich der Transformationscodierung mit und ohne Fensterung

Ein wichtiger Parameter in der Transformationscodierung ist die Transformationslänge. Bei der Codierung von Audiosignalen [7, 8, 9] bestimmt die Dauer der Vorverdeckung im menschlichen Gehör die obere Grenze für die Transformationslänge, wenn keine zusätzlichen Maßnahmen zur Vorecho-Unterdrückung durchgeführt werden [10]. Durch den Einsatz der adaptiven Fensterung

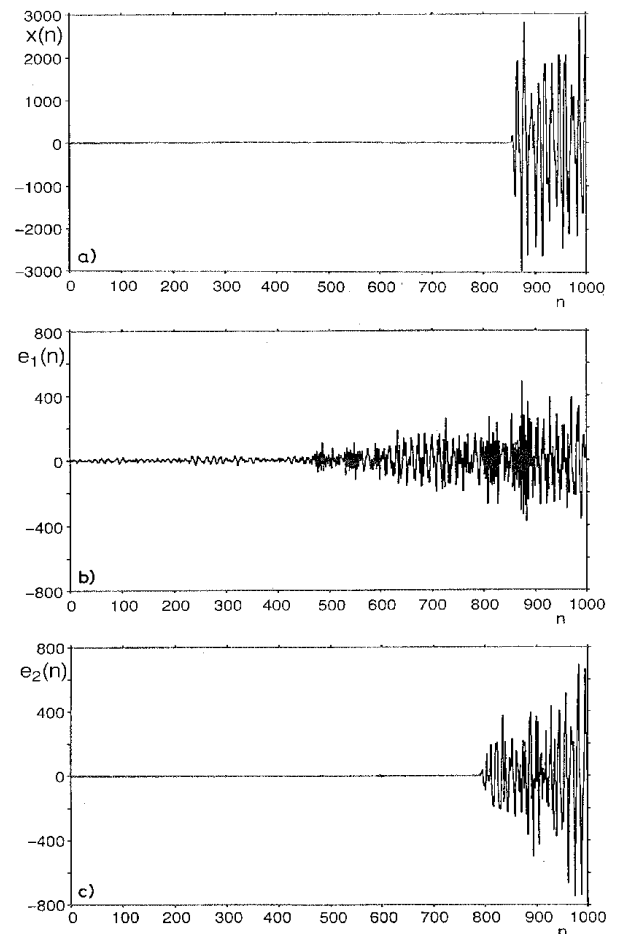


Bild 5: Anschlag einer Triangel

- a) Originalsignal $x(n)$
- b) Fehlersignal $e_1(n)$ ohne Fensteradaption
- c) Fehlersignal $e_2(n)$ mit Fensteradaption

ist die Transformationslänge hauptsächlich durch die maximal zulässige Verzögerung begrenzt.

Den folgenden Untersuchungen liegt eine Transformation der Länge $N=512$ mit einem Sinus-Fenster der Länge $2N=1024$ zugrunde. Bei adaptiver Fensterung wird nach der Detektion eines Impulses oder Amplitudensprungs anstelle einer Transformation der Länge $N_1=512$ vielmals eine Transformation der Länge $N_2=128$ mit Sinus-Rechteck-Fenster und Überlappung von jeweils $L=32$ Abtastwerten angewendet. Bei dieser Vorgehensweise muß als Steuerinformation jeweils nur ein zusätzliches Bit für jeden Block von 512 Abtastwerten übertragen werden.

Die Transformationskoeffizienten werden mittels adaptiver Quantisierung und adaptiver Bitzuweisung codiert, wobei spektrale Maskierungseigenschaften des menschlichen Gehörs ausgenutzt werden.

Die Ergebnisse einer Computersimulation mit einer Reduktion der Datenrate von 16 bit pro Abtastwert auf 2,5 bit pro Abtastwert wurden durch Hörversuche ausgewertet. Diese Untersuchungen ergaben ohne Fensteradaptation deutlich wahrnehmbare Vorechos bei kritischen Signalen. Wurde dagegen die hier beschriebene adaptive Fensterung eingesetzt, so traten bei gleicher Datenrate keine hörbaren Vorechos mehr auf.

Zur Verdeutlichung der Verkürzung der Vorechos durch die adaptive Fensterung zeigt **Bild 5** den Anschlag $x(n)$ einer Triangel sowie die Differenzsignale zwischen Originalsignal und codiertem Signal ohne adaptive Fensterung $e_2(n)$. Die im Verhältnis zu $e_1(n)$ etwas größere Amplitude von $e_2(n)$ nach dem Anschlag macht sich nicht störend bemerkbar, da zu diesem Zeitpunkt die Momentanmaskierung durch das Nutzsignal [4] schon aktiv ist.

6. Zusammenfassung

Die in diesem Beitrag vorgestellte Transformationscodierung mit adaptiver Fensterung erlaubt den Einsatz

überlappender Transformationen, ohne dadurch den Nachteil längerer Vorechos aufzuweisen. Zur Einführung der adaptiven Fensterung in Verbindung mit den Transformationen nach dem Prinzip des TDAC wurde gezeigt, wie die Fensterfunktionen aufeinanderfolgender Blöcke variiert werden können und zusätzlich die Transformationslänge von Block zu Block umgeschaltet werden kann. Anschließend wurde ein Verfahren zur Steuerung der Fensterfunktionen vorgestellt. Die mit Hilfe einer adaptiven Codierung der Koeffizienten erzielten Ergebnisse zeigen die Leistungsfähigkeit des Verfahrens der adaptiven Fensterung. Mit einem zusätzlichen Übertragungsaufwand von nur einem Bit je Block von 512 Abtastwerten läßt sich die Dauer der Vorechos bis unter die Hörbarkeitsschwelle verkürzen.

Literatur:

- [1] Princen, J. P.; Bradley, A. B.: Analysis/Synthesis Filter Bank Design Based on Time Domain Aliasing Cancellation. IEEE Trans. on ASSP, ASSP-34 (Okt. 1986) S. 1153—1161.
- [2] Princen, J. P.; Johnson, A. W.; Bradley, A. B.: Subband/Transform Coding Using Filter Bank Designs Based on Time Domain Aliasing Cancellation. Proc. ICASSP 1987, S. 50.1.1—50.1.4.
- [3] Jayant, N. S.; Noll, P.: Digital Coding of Waveforms. Englewood Cliffs, N.J.: Prentice Hall, Inc., 1984.
- [4] Zwicker, E.: Psychoakustik. Berlin/Heidelberg/New York: Springer, 1982.
- [5] Schroeder, M. R.; Atal, B. S.; Hall, J. L.: Optimizing digital speech coders by exploiting masking properties of the human ear. JASA 66 (6) (Dec. 1979) S. 1647—1652.
- [6] Edler, B.: A fast overlapping block transform. Veröffentlichung in Vorbereitung.
- [7] Brandenburg, K.: High quality sound coding at 2.5 bit/sample. Preprint 2582 (D-2), 84, AES-Convention, Paris 1988.
- [8] Krahé, D.: Ein Verfahren zur Datenreduktion bei digitalen Audiosignalen unter Ausnutzung psychoakustischer Phänomene. Rundfunktech. Mitt. (1986) 3, S. 117—123.
- [9] Johnston, J.: Transform Coding of Audio Signals Using Perceptual Noise Criteria. IEEE Journal on sel. areas in communications 6 (Feb. 1988), S. 314—323.
- [10] Vaupel, Th.: Ein Kompanier zur Unterdrückung von hörbaren Störungen bei dynamischen Signalpassagen für ein Transformationscodierungsverfahren für qualitativ hochwertige Audiosignale (MSC). ITG-Fachbericht 106: Hörrundfunk (Nov. 1988) S. 209—215.

Dipl.-Ing. B. Edler, Institut für Theoretische Nachrichtentechnik und Informationsverarbeitung, Universität Hannover, Appelstraße 9a, 3000 Hannover
(Eingegangen am 7. 7. 1989)