

LATTICE-QUANTISIERUNG VON LP-FILTERKOEFFIZIENTEN AUF DER BASIS DER LINE-SPECTRAL-FREQUENCIES (LSF)

Jürgen Schnitzler

Institut für Nachrichtengeräte und Datenverarbeitung (IND)

RWTH Aachen, D-52056 Aachen

Telefon: +49.241.806982, Fax: +49.241.8888186, E-mail: Juergen.Schnitzler@ind.rwth-aachen.de

KURZFASSUNG

Bei der prädiktiven Codierung von Sprachsignalen mit Bandbreiten von 6 ... 8 kHz ergeben sich im Vergleich mit der Codierung von Telefonsprache erhöhte Anforderungen an die Quantisierung der Prädiktorkoeffizienten. In diesem Beitrag wird für die Koeffizientenquantisierung eine Hybridlösung aus mehreren Lattice-Quantisierern und einem trainierten Vektorquantisierer vorgestellt, mit dem eine transparente Quantisierung mit etwa 3 Bit pro LPC-Koeffizient bei deutlich reduziertem Speicher- und Rechenaufwand erreicht wird. Ein von Xie und Adoul vorgestellter Ansatz [1] wurde dazu modifiziert und erweitert.

1. EINFÜHRUNG

In den Algorithmen zur Quellencodierung von Sprachsignalen kommt heute vorwiegend das Prinzip der linear prädiktiven Codierung (LPC) zum Einsatz. Die N_p Koeffizienten der linearen Prädiktoren (LP-Koeffizienten) sind dabei geeignet zu quantisieren und zu übertragen. Als zur Quantisierung bzw. Codierung geeignete Darstellung der LP-Koeffizienten $a_i, i = 1, \dots, N_p$ haben sich dabei die sog. Line Spectral Frequencies (LSF) [2] durchgesetzt. In zahlreichen Veröffentlichungen, die sich mit effizienten Quantisierungsalgorithmen auf der Grundlage der LSF-Parameter beschäftigen, werden neben skalaren hauptsächlich vektorielle Quantisierer (VQ) vorgeschlagen (z.B. [2]). Man findet Split-VQ- und Multi-Stage-VQ-Verfahren, außerdem differentielle oder prädiktive Ansätze ([3], [4]). Im 8 kbit/s Standard ITU-T G.729 wird z.B. eine prädiktive, zweistufige Split-Vektor-Quantisierung eingesetzt [5].

Im Falle einer breitbandigen Sprachcodierung mit einer entsprechend höheren LP-Filterordnung sind neben einer hohen Sprachqualität auch vermehrt der erhöhte Speicherbedarf und Suchaufwand bei einer (Vektor-) Quantisierung mit trainierten Codebüchern zu berücksichtigen. Daher wird ein Entwurf einer hybriden Struktur bestehend aus mehreren sog. Lattice-Vektorquantisierern (LVQ) und einem kleinen trainierten Vektorquantisierer vorgestellt. Der zugrundeliegende Ansatz beruht auf einer Untersuchung von Xie und Adoul [1].

Das hier vorgestellte Entwurfsbeispiel ist angepaßt auf eine Signalbandbreite von 6 kHz und eine LP-Filterordnung von $N_p = 14$, entsprechend einer Anwendung in einem Teilband-Sprachcodec mit ungleichmäßiger Bandaufteilung [6]. Die LP-Koeffizienten werden einmal pro Analyserah-

men (10 ms) berechnet und quantisiert.

2. LATTICE-VEKTORKOMPANDERSYSTEM

Als Alternative zu den bekannteren trainierten LBG-Vektorquantisierern (LBG-VQ) [7], bieten sich bei bestimmten Anwendungen die sog. *Lattice-Vektorquantisierer* (LVQ) an. Bei diesen sind die Codevektoren der Dimension N auf einem regelmäßigen Gitter im N -dimensionalen Raum angeordnet (z.B. [8]).

Für solche „algebraischen“ Vektorquantisierer lassen sich schnelle Quantisierungsalgorithmen angeben, die meist *nicht* von der Quantisierungsrate und nur linear von der Dimension abhängen z.B. [9]. Da die Lagen der Codevektoren analytisch bekannt sind, lassen sich die Codebuchindizes durch eine Codierungsvorschrift bestimmen; eine Speicherung des Codebuchs entfällt.

Lattice-Quantisierer verhalten sich allerdings nur dann optimal, wenn die zu quantisierenden Vektoren möglichst gleichmäßig in einem endlichen, geometrisch einfach zu beschreibenden Gebiet des Vektorraums verteilt sind. Im allgemeinen müssen die Vektoren vor Anwendung einer LVQ geeignet transformiert werden, um diese Bedingungen herzustellen (*Kompandersystem*, vgl. Bild 1).

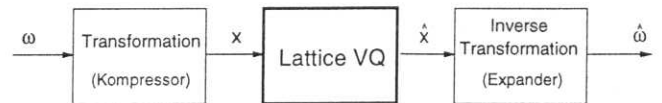


Bild 1. Lattice-Vektorkompandersystem

Im vorliegenden Fall werden die LSF-Parameter $\omega_i, i = 1 \dots 14$ entsprechend Tabelle 1 in 4 Blöcke eingeteilt. Die LSF $\omega_3, \omega_7, \omega_{11}$ werden hier als *Stützstellen* bezeichnet. Diese werden zunächst mit einem LBG-VQ gemeinsam

Block 1	ω_1, ω_2	↪ Lattice VQ
Block 2	$\omega_4, \omega_5, \omega_6$	
Block 3	$\omega_8, \omega_9, \omega_{10}$	
Block 4	$\omega_{12}, \omega_{13}, \omega_{14}$	
Stützstellen	$\omega_3, \omega_7, \omega_{11}$	↪ LBG-VQ

Tabelle 1. Einteilung der LSF-Parameter

quantisiert. Ein Codebuch mit 512 Vektoren (9 Bit) erweist sich hierzu als ausreichend. Die quantisierten Stützstellen-LSF $\hat{\omega}_3, \hat{\omega}_7, \hat{\omega}_{11}$ werden für die Transformation der in den Blöcken $k = 1, 2, 3, 4$ liegenden restlichen LSF benötigt. Wie in [1] ergeben die Transformationen innerhalb der

einzelnen Blöcke Parameter x_{kj} , die den auf die jeweilige Blocklänge normierten Differenzen benachbarter LSF entsprechen. Beispiel: Intervall $k = 2$, $\omega_4 \leq \omega_i \leq \omega_6$:

$$x_{21} = \frac{\omega_4 - \hat{\omega}_3}{\hat{\omega}_7 - \hat{\omega}_3}, \quad x_{22} = \frac{\omega_5 - \omega_4}{\hat{\omega}_7 - \hat{\omega}_3}, \quad x_{23} = \frac{\hat{\omega}_7 - \omega_6}{\hat{\omega}_7 - \hat{\omega}_3} \quad (1)$$

Der Zusammenhang zwischen den LSF ω_i und den x_{kj} -Parametern ist in Bild 2 veranschaulicht. In den einzel-

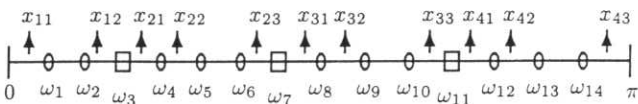


Bild 2. Zusammenhang zwischen LSF-Parametern und ω_i transformierten Parametern x_{kj}

nen Blöcken lassen sich die Parameter x_{kj} als Koordinaten von Vektoren \mathbf{x}_k in Vektorräumen der Dimension $N = 2$ ($k = 1$) bzw. $N = 3$ ($k = 2, 3, 4$) interpretieren. Aufgrund der monoton ansteigenden Sortierung der ω_i ergeben sich die Eigenschaften $0 < x_{kj} < 1$ und $0 < \sum_j x_{kj} < 1$, $k = 1 \dots 4$, d.h., die so gegebenen Vektoren liegen innerhalb eines Dreiecks bzw. Simplex. In diesen – je nach gewünschter Quantisierungsaufösung geeignet zu skalierenden – Gebieten werden jeweils Lattices des Grundtyps D_2 bzw. D_3 [8] angeordnet. Dieser Typ ist durch die Menge aller Punkte im zwei- bzw. dreidimensionalen Vektorraum gegeben, die ganzzahlige Koordinaten und nur geradzahlige (oder nur ungeradzahlige) Koordinatensummen besitzen.

Statistische Analysen zeigen, daß die transformierten Vektoren zumindest in den zentralen Bereichen dieser Gebiete angenähert gleichverteilt auftreten, während die Randbereiche nur sehr dünn besetzt sind.

Die Quantisierung eines N-dimensionalen Vektors im D_N -Lattice erfordert im wesentlichen einfache Rundungsoperationen auf ganzzahlige Koordinaten, z.B. [8], die zu einem optimalen Codevektor im Sinne eines minimalen euklidischen Fehlers führen.

Mit den gegebenen Dreiecks- bzw. Simplexbegrenzungen lassen sich für diesen Latticetyp zudem bis zur Dimension 3 einfache geschlossene Indizierungsvorschriften angeben [1].

3. OPTIMIERTE LATTICE-STRUKTUR

Die auf die Dreiecks- bzw. Simplex-Gebiete begrenzten Lattices werden hier als S_N -Lattice bezeichnet. Bild 3 zeigt als Beispiel ein S_3 -Lattice der Ordnung $L = 8$, das 95 gültige Codevektoren enthält. Als Ordnung des Lattice wird hier die maximale Summe der Lattice-Koordinaten bezeichnet. Um der nur eingeschränkt gleichmäßigen Verteilung Rechnung zu tragen, werden zunächst die Punkte in den Randbereichen ausgeschlossen (*Reduced Simplex* (RS_N)-Lattice). Derartige Beschränkungen sollten allerdings an die Grundform des S_N -Lattice angepaßt sein, um die Indizierung nicht zu erschweren. Wie Bild 4 zeigt, werden daher ein minimaler Koordinatenwert γ_{\min} sowie zwei diagonalen Schranken μ_{\min} und μ_{\max} eingeführt. Sodann wird eine Überlagerung von zwei um einen halben Rasterabstand versetzten RS_N -Lattices betrachtet. Die genannten Schranken sollen für das normale RS_N -Lattice und das sogenannte *Co-Lattice* (RS_N^*) unabhängig gewählt werden, vgl. Bild 4.

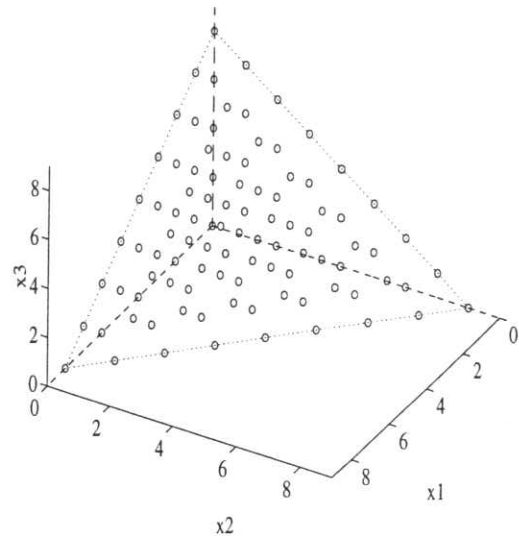


Bild 3. Beispiel eines S_3 -Lattice: Ordnung $L = 8$, 95 Lattice-Punkte

Die resultierende, hier als *Combined Reduced Simplex* (CRS_N)-Lattice bezeichnete Struktur weist in einem Zentralbereich eine hohe Dichte, in den Randbereichen eine geringere Dichte zulässiger Codevektoren auf. Bei einem sorgfältigen Entwurf liegen nur noch etwa 1 % der Vektoren außerhalb der zusätzlich eingeführten Grenzen. Für diese seltenen Fälle sind allerdings zusätzliche Maßnahmen erforderlich, die ggf. den nächsten zulässigen Codevektor am Rand der CRS_N -Struktur ermitteln. Außerdem ist bei der Parameterwahl zu beachten, daß die Anzahl der Codevektoren in beiden Teil-Lattices etwa gleich ist und nahe unterhalb einer Zweierpotenz liegen sollte. Ein zusätzliches Bit wird zur Unterscheidung von normalem und Co-Lattice

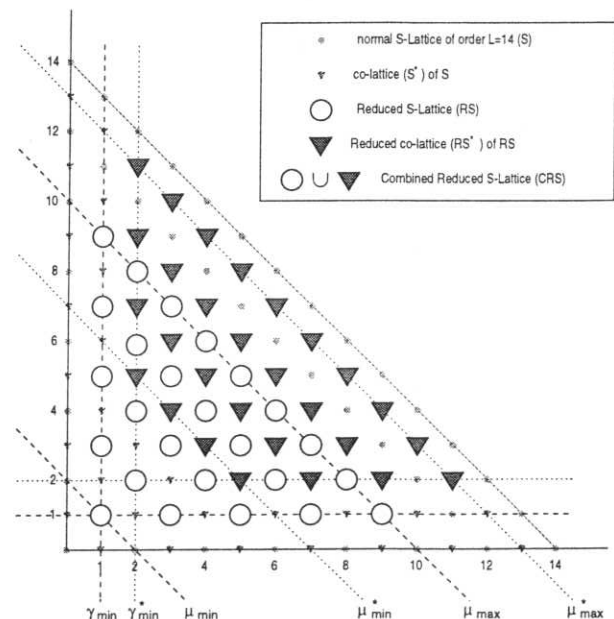


Bild 4. Beispiel eines CRS_2 -Lattice

benötigt.

Der hier vorgestellte erste Entwurf eines solchen, mit VQ-LVQ bezeichneten hybriden Schemas für die 14 LSF-Werte verwendet neben dem 9 Bit Vektorcodebuch für die drei Stützstellen noch eine 6 Bit LVQ (*RS*₂-Lattice) für ω_1, ω_2 , sowie 10 Bit, 9 Bit, bzw. 10 Bit LVQ-Konfigurationen (*CRS*₃ Lattices) für die restlichen LSF der anderen drei Blöcke, d.h. insgesamt 44 Bits.

Zur Beurteilung der LSF-Quantisierung wird ein spektrales Abstandsmaß *SD* (*Spectral Distortion*) [2] benutzt, das definiert ist als

$$SD = \left\{ \frac{100}{\pi} \int_0^\pi [\log_{10} |A_n(e^{j\omega})|^2 - \log_{10} |\hat{A}_n(e^{j\omega})|^2]^2 d\omega \right\}^{1/2}. \quad (2)$$

Hier stehen $A_n(e^{j\omega})$ und $\hat{A}_n(e^{j\omega})$ für den unquantisierten bzw. quantisierten Frequenzgang des LPC-Analysefilters $A(z) = 1 - \sum_{i=1}^{N_p} a_i z^{-i}$ im Rahmen n .

In guter Übereinstimmung mit der subjektiven Beurteilung wird eine Quantisierung als spektral transparent betrachtet [2], wenn die mittlere spektrale Verzerrung SD_{avg} etwa 1 dB beträgt, *SD* außerdem in weniger als $\Delta_{T_{ypI}} = 2\%$ aller Rahmen Werte zwischen 2 und 4 dB, und in nahezu keinem ($\Delta_{T_{ypII}} \approx 0$) Rahmen mehr als 4 dB annimmt.

Mit dieser 44 Bit-Allokation werden im Mittel eine Verzerrung von $SD_{avg} = 1.02$ dB erzielt. $\Delta_{T_{ypI}} = 0.7\%$ der *SD*-Werte liegen zwischen 2 und 4 dB und keine darüber ($\Delta_{T_{ypII}} = 0$). Diese Messungen wurden mit Sprachproben unterschiedlicher Sprecher (Gesamtlänge etwa 100 s) und einer Sprachaktivität oberhalb 95 % durchgeführt. Zum Training des Stützstellencodebuchs wurde deutsches, englisches und französisches Sprachmaterial mit einer Gesamtdauer von etwa 20 min verwendet.

Vergleicht man die erzielte Bitrate von 44 Bit/Rahmen (also 3,1 Bit/LSF) z.B. mit Ansätzen aus dem Bereich der Schmalbandcodierung (bei 2 ... 3 Bit/LSF je nach gewünschter Qualität, s.o.), erscheint eine weitere Optimierung sinnvoll.

Zur Analyse der Beiträge zur *SD* Daher wurden in mehreren Simulationen jeweils entweder nur die LSF-Stützstellen oder nur ein Block von LSF (vgl. Tabelle 1) quantisiert. Es zeigt sich, daß die Stützstellenquantisierung mit dem LBG-Vektorquantisierer den größten Beitrag zu *SD* liefert. Ein Vorschlag zur weiteren Verbesserung der Stützstellenquantisierung wird nachfolgend behandelt.

4. VERBESSERTE QUANTISIERUNG DER STÜTZSTELLEN

Codebuchsuche mit gewichtetem Fehlermaß

Bislang wurde für die Codebuchsuche ein einfaches euklidisches Fehlermaß (MSE) eingesetzt. Eine Verbesserung ist hier durch ein subjektives, d.h. gehörbezogenes gewichtetes Fehlermaß (WMSE) zu erwarten (vgl. [2]), in dem Formanten des Spektrums des LP-Synthesefilters stärker als Täler sowie Komponenten bei niedrigeren Frequenzen etwas stärker als solche bei höheren Frequenzen gewichtet werden.

Als modifizierte Fehlerfunktion wird daher zur Quantisie-

rung der Stützstellen eingeführt:

$$d_{edge} = \sum_{j \in \{3,7,11\}} [c_j v_j (\omega_j - \hat{\omega}_j)]^2. \quad (3)$$

Die Wichtung setzt sich zusammen aus einem festen Anteil (hier wurde $c_3 = 1.0$, $c_7 = 0.95$ und $c_{11} = 0.9$ gewählt) und einem variablen Anteil, in den die Charakteristik des aktuellen Rahmens eingeht. Vereinfachend im Vergleich zu [2] oder [3] wird die invertierte Differenz

$$v_j = (\omega_{j+1} - \omega_{j-1})^{-1} \quad (4)$$

der die Stützstelle umgebenden LSF-Parameter gewählt, da dicht zusammenliegende Line Spectral Frequencies die Formanten bestimmen. Mit diesen Wichtungen ergeben sich leichte Verbesserungen bezüglich SD_{avg} und $\Delta_{T_{ypI}}$.

Prädiktive Quantisierung der Stützstellen

Bei dem bisher vorgestellten Ansatz der VQ-LVQ wurde noch nicht die zeitliche Korrelation der Prädiktionsparameter über mehrere Rahmen hinweg ausgenutzt. Mithilfe der Transformation und der LVQ wurden bereits die Intraframe-Korrelation der LSF ausgenutzt. Desweiteren besteht eine Korrelation der ω_i über mehrere Rahmen (Länge 10 ms) hinweg (Interframe-Korrelation). Diese kann, wie etwa auch im 8 kbit/s-Standardcodec ITU-T G.729, durch eine prädiktive Codierung verwertet werden (z.B. [5]).

In diesem Sinne wird für die Stützstellen $\omega_3, \omega_7, \omega_{11}$ jeweils einzeln ein fester Prädiktor der Länge $N_p^{(Q)}$ entworfen. Dabei werden die zeitlichen Abhängigkeiten durch einen *Moving Average* (MA) Prozeß der Ordnung $N_p^{(Q)}$ modelliert, um im Empfänger eine nichtrekursive Synthesestruktur zu ermöglichen und so eine Fehlerfortpflanzung über mehr als $N_p^{(Q)}$ Rahmen zu verhindern. Für die Analyseseite erhält man somit, wie Bild 5 zeigt, eine IIR-Struktur (AR-Prädiktor), wobei eine 'closed-loop'-Quantisierung bevorzugt wird.

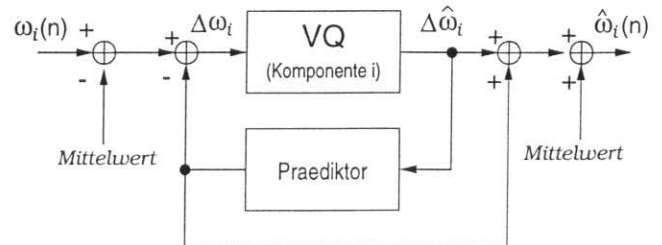


Bild 5. Prädiktive Vektorquantisierung mit AR-Prädiktoren, dargestellt für die i -te Komponente (n : Zeitindex)

Dieses Verfahren soll hier als PVQ-LVQ (Prädiktive Vektorquantisierung - Lattice-Vektorquantisierung) bezeichnet werden. Eine Prädiktorordnung von $N_p^{(Q)} = 4$, erweist sich als ausreichend und führt zu Prädiktionsgewinnen von 6.1, 5.9 und 5.5 dB für ω_3, ω_7 bzw. ω_{11} . Diese lassen sich derart nutzen, daß zunächst das Vektorcodebuch für die Prädiktionsresiduen $\Delta\omega = (\Delta\omega_3, \Delta\omega_7, \Delta\omega_{11})$ auf 8 Bit reduziert wird. Eine weitere Verkleinerung der LVQ-Struktur

Block 1	ω_1, ω_2	LVQ (RS_2)	6 Bit
Block 2	$\omega_4, \omega_5, \omega_6$	LVQ (CRS_3)	10 Bit
Block 3	$\omega_8, \omega_9, \omega_{10}$	LVQ (CRS_3)	9 Bit
Block 4	$\omega_{12}, \omega_{13}, \omega_{14}$	LVQ (CRS_3)	9 Bit
Stützstellen	$\omega_3, \omega_7, \omega_{11}$	PVQ (LBG)	8 Bit
Summe			42 Bit

Tabelle 2. Bitaufteilung für PVQ-LVQ-Konfiguration

im Block $k = 4$ (LSF ω_{12}, ω_{13} und ω_{14}) auf 9 Bit führt schließlich insgesamt auf eine 42 Bit-Konfiguration dieses PVQ-LVQ-Schemas (vgl. Tabelle 2), das vergleichbare SD-Werte ($SD_{avg} = 0.98$ dB, $\Delta_{Typ1} = 1.3$ % und $\Delta_{Typ2} = 0$) wie die zuvor beschriebene 44 Bit-VQ-LVQ-Anordnung liefert.

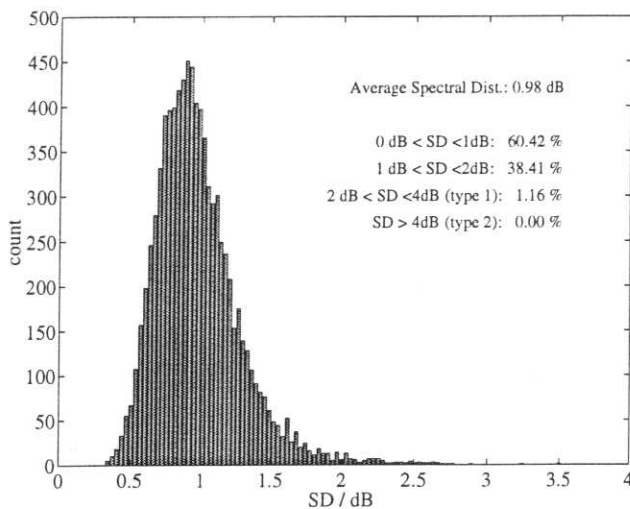


Bild 6. Histogramm der spektralen Verzerrung SD für das PVQ-LVQ-System (42 Bit pro Rahmen)

In Bild 6 ist ein Histogramm der SD-Werte für diese Konfiguration dargestellt.

Zur vergleichenden Bewertung der PVQ-LVQ wurde ein PMAVQ-System (Predictive Moving Average Split Vector Quantization) nach [3] für eine Prädiktorordnung $N_p = 14$ untersucht, bei dem ebenfalls AR-Prädiktoren der Ordnung 4 für alle 14 LSF eingesetzt und die Residuen blockweise vektoriell mit 4 jeweils 1024 Vektoren umfassenden Codebüchern der Dimensionen 3-3-4-4 quantisiert werden. Die Quantisierungsgüte ($SD_{avg} = 1.01$ dB, $\Delta_{Typ1} = 1.91$ % und $\Delta_{Typ2} = 0.01$) ist zwar vergleichbar, jedoch liegen die gemessenen Rechenzeiten bei der hier vorgeschlagenen PVQ-LVQ-Konfiguration um Faktoren 15 ... 20 niedriger, und der Speicheraufwand ist mit einem dreidimensionalen 8 Bit-Vektorcodebuch um den Faktor 19 geringer gegenüber dem PMAVQ-System.

Die beschriebene 42 Bit-Konfiguration der PVQ-LVQ ist derzeit für die Quantisierung der LP-Filterkoeffizienten eines Breitband-Codieralgorithmus (50-7000 Hz) vorgesehen, bei dem für ein Teilbandsignal (Abtastrate 12 kHz) eine LP-Analyse der Ordnung 14 durchgeführt wird ([6], [10]).

5. ZUSAMMENFASSUNG

In diesem Beitrag wurde ein hybrides Schema zur Quantisierung von LP-Filterkoeffizienten von Sprachsignalen vorgestellt. Dabei werden die in LSF-Darstellung vorliegenden Koeffizienten blockweise geeignet transformiert und mit optimierten Lattice-Strukturen vektoriell quantisiert. Die zur Transformation benötigten „Stützstellen“-LSF werden vorab mit einer prädiktiven Vektorquantisierung codiert. Eine Konfiguration mit 42 Bit pro Koeffizientensatz bei einer LP-Filterordnung von $N_p = 14$ führte zu einer transparenten Quantisierung mit einem im Vergleich zum Stand der Technik (z.B. [3]) erheblich reduzierten Rechen- und Speicheraufwand.

DANKSAGUNG

Der Autor dankt dem Technologiezentrum der Deutschen Telekom AG, Darmstadt, hier besonders Herrn G. Schröder für die Unterstützung dieser Untersuchung. Spezieller Dank gilt auch Herrn Prof. P. Vary und Herrn J.W. Paulus für Anregungen und Diskussion.

LITERATUR

- [1] M. Xie und J. Adoul, Fast and Low Complexity LSF Quantization Using Algebraic Vector Quantizer in *Proceedings ICASSP*, pp. 716–719, IEEE, 1995.
- [2] K. Paliwal und B. Atal, Efficient Vector Quantization of LPC Parameters at 24 Bits/Frame *IEEE Transactions on Speech and Audio Processing*, vol. 1, pp. 3–13, January 1993.
- [3] R. Salami, C. Laflamme, J.-P. Adoul, und D. Massaloux, “A Toll Quality 8kb/s Speech Codec for the Personal Communications System (PCS),” *IEEE Trans. Vehicular Technology*, vol. 43, pp. 808–816, August 1994.
- [4] H. Ohmuro, M. Takehiro, K. Mano, und S. Miki, Coding of LSP Parameters Using Interframe Moving Average Prediction and Multi-Stage Vector Quantization in *Proceedings Workshop on Speech Coding*, pp. 63–64, IEEE, 1993.
- [5] ITU-T Recommendation G.729, “Coding of Speech at 8kbps using conjugate-structure algebraic-code-excited linear-prediction (CS-ACELP).”
- [6] J. W. Paulus und J. Schnitzler, 16 kbit/s Wideband Speech Coding Based on Unequal Subbands in *Proc. Int. Conf. Acoust., Speech, Signal Processing, ICASSP*, (Atlanta, Georgia), pp. 255–258, IEEE, 1996.
- [7] Y. Linde, A. Buzo, und R. Gray, An Algorithm for Vector Quantizer Design *IEEE Transactions on Communications*, vol. COM-28, pp. 84–94, January 1980.
- [8] J. Conway und N. Sloane, *Sphere Packings, Lattices and Groups*. Springer, 2 ed., 1993.
- [9] M. Herbert, “Lattice-Quantisierung von Sprach- und Sprachmodellsignalen.” Dissertation, Universität Erlangen-Nürnberg, 1991.
- [10] J. W. Paulus und J. Schnitzler, Wideband Speech-coding for the GSM Fullrate Channel? in *Proc. ITG Fachtagung Sprachkommunikation*, (Frankfurt am Main), 1996.