Praktische Erfahrungen mit der fraktalen Sprachkompression

Stephan Schneider
BTU Cottbus, Lehrstuhl Interkulturalität
schneide@tu-cottbus.de

Abstract: Mit fraktalen Verfahren läßt sich eine Signalformkodierung für Sprachsignale realisieren, die der Vektorquantisierung ähnelt. Ausgehend von theoretischen Kennwerten wie dem SNR werden die praktischen Eigenschaften einer fraktalen Sprachkodierung im Zeit- und im Frequenzgang beschrieben. Der Beitrag gibt Hinweise für die konkrete Implementierung und schlägt ein Dateiformat für frakal kodierte Sprachsignale vor.

1 Einleitung

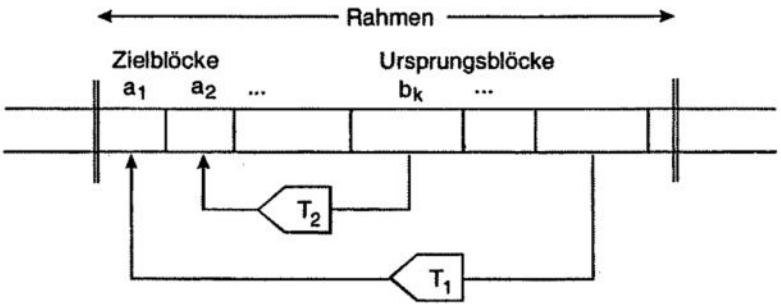
Ob ein Verfahren zur Sprachkompression brauchbar ist, entscheidet sich in der praktischen Erprobung und in Hörtests. Das in [10] vorgestellte fraktale Verfahren wurde deshalb implementiert und praktischen Tests unterzogen.

2 Das Kompressionsverfahren

2.1 Kodierung und Rekonstruktion

Das Verfahren zur fraktalen Kompression von Sprachsignalen wurde bereits an anderer Stelle ausführlich beschrieben (z.B. [6], [10], [11])so daß hier eine kurze Wiederholung genügt.

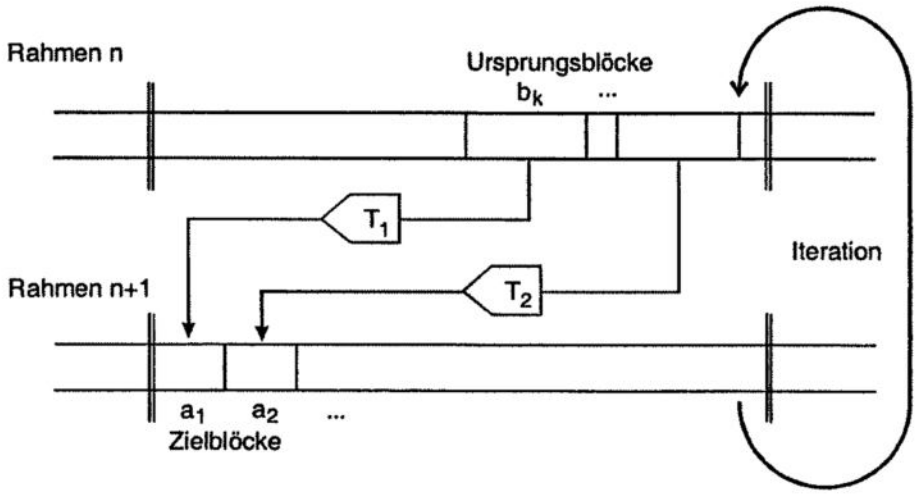
Das digitalisierte Sprachsignal wird zunächst in Rahmen mit einer konstanten Länge von M Abtastwerten eingeteilt, wobei M im Bereich von 1.000 bis 50.000 Abtastwerten liegt. Das Problem unvollständiger Rahmen wird weiter unten erörtert. Jeder Rahmen wird lückenlos in sog. "Zielblöcke" a_k mit der Länge N unterteilt, wobei N beim Grundverfahren konstant bleibt und zwischen 8 und 64 Abtastwerten beträgt.



Abbildung~1: Fraktale Kodierung: Bestimmung einer optimalen Transformation T_k für jeden Zielblock a_k eines Rahmens.

Die Kodierung besteht darin, dass jeder Zielblock durch eine affine Transformation T_k angenähert wird, wobei die Daten der Transformation anstelle der Abtastwerte gespeichert werden. Die Transformation T_k bildet einen bestimmten Bereich des Rahmens, der als Ursprungsblock b_j bezeichnet wird, so ab, daß der Zielblock im Sinne eines Abstandsmaßes optimal angenähert wird. Als Abstandsmaß wurde hier das mittlere Fehlerquadrat (MSE) gewählt. Die Gesamtheit aller Transformationen T_k bildet die fraktale Kodierung eines Rahmens. Bei der Rekonstruktion eines Rahmens werden alle Transformationen T_k des Rahmens hintereinander durchgeführt, bis ein neuer Rahmen berechnet ist. Diese Prozedur wird solange wiederholt, bis der ursprüngliche Rahmen wiederhergestellt ist. Der Rahmen, mit dem die Iteration begonnen wird, kann dabei beliebige Abtastwerte enthalten.

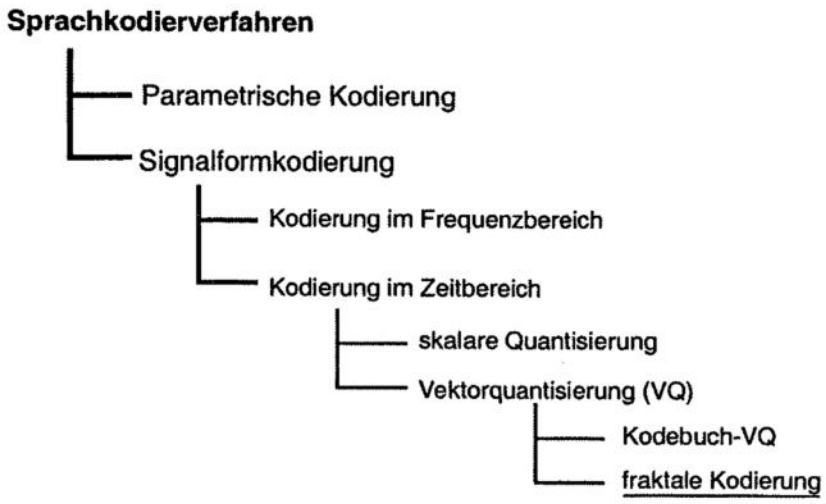
Damit die wiederholt ausgeführten Transformationen konvergieren, müssen sie bestimmte Eigenschaften aufweisen. Dazu gehört, dass der Ursprungsblock durch die Transformation zeitlich verkürzt, d.h. "kontrahiert" werden muß. In der Praxis geschieht die Kontraktion entweder durch Unterabtastung oder durch Mittelwertbildung ([7]).



Abbildung~2: Rekonstruktion eines fraktal kodierten Signals. Die Transformationen T_k werden hintereinander ausgeführt und mehrfach iteriert.

2.2 Einordnung des Verfahrens

Das vorliegende Verfahren gehört, da es auf nahezu beliebige Signale anwendbar ist und auf keinem Signalmodell beruht, zur Klasse der Signalformkodierung. Da das Verfahren im Zeitbereich arbeitet und blockweise vorgeht, kann es der Vektorquantisierung (VQ) zugeordnet werden. Bild 3 zeigt die Einordnung des vorliegenden Verfahrens in den "-Stammbaum" der Kodierverfahren (vgl. [2], [12]).



Abbildung~3: Stammbaum der Sprachkodierverfahren.

3 Der Signal-Rauschabstand

Für ein Verfahren, das zu der Gruppe der Signalformkodierer gehört, ist der Signal-Rauschabstand (SNR) eine wichtige Größe. In Bezug auf Sprachkodierverfahren muß aber betont werden, dass der SNR nur wenig über die subjektiv empfundene Sprachqualität aussagt.

Für die Vektorquantisierung ([3], [8]) mit einer Blocklänge von N unkorrelierten, gaußverteilten Abtastwerten gilt der folgende Zusammenhang ([1]):

$$SNR_{\mathrm{dB}} = 20\frac{\log_2 M}{N}\log_{10} 2 \approx 6.03\frac{\log_2 M}{N} \tag{1}$$

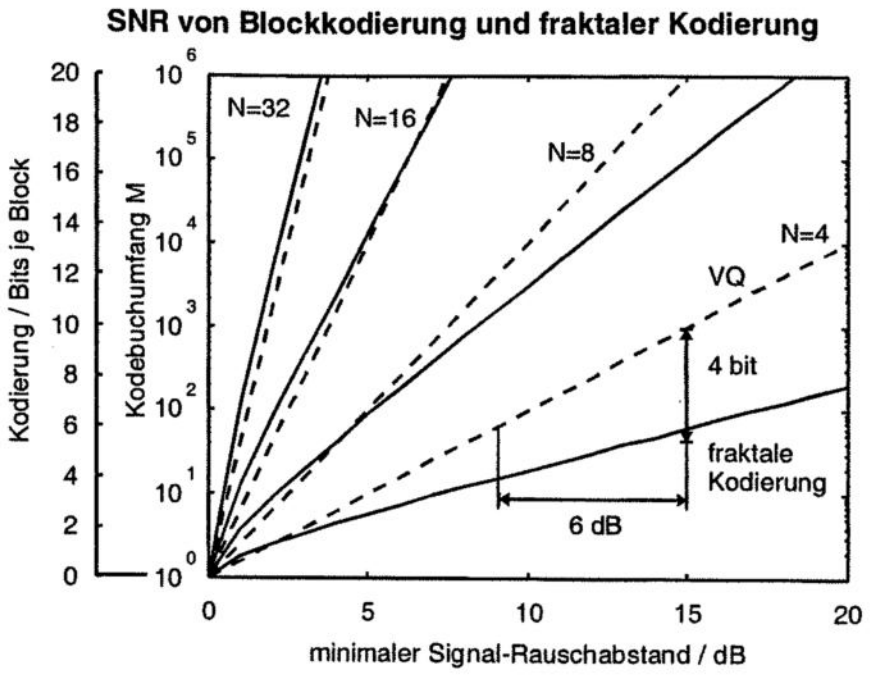
M bezeichnet die Zahl der Kodebuchvektoren, die $\log_2 M$ Bits zur Kodierung benötigen. Bei Korrelationen verbessert sich der SNR erheblich, was durch einen "Gewinn" G_N ausgedrückt, der von der Korrelationsmatrix und damit auch von der Blocklänge N abhängt ([5]):

$$SNR_{\mathrm{dB}} \approx 6.03\frac{\log_2 M}{N} + G_N \tag{2}$$

Mit Hilfe eines statistischen Modells, in dem die Amplitudenverteilung des Sprachsignals als Gaußverteilung angenähert wurde, konnte die untere Grenze des SNR für das oben beschriebene fraktale Kodierverfahren analytisch berechnet werden. Diese untere Grenze gilt für den Fall statistisch unabhängiger, gaußverteilter Abtastwerte und bestimmt den SNR, der für eine bestimmte Rahmenlänge M und Blocklänge N mindestens erreicht wird. Der Rechengang sowie die Gleichungen sind in [11] zu finden.

Zum direkten Vergleich sind die Ergebnisse für die fraktale Kodierung wie für die Vektorquantisierung in der untenstehenden Grafik (Bild 4) gemeinsam dargestellt.

In Bild 4 hat es den Anschein, als würde die fraktale Kodierung die VQ in der Bitrate unterbieten. Es ist jedoch zu beachten, daß nur die Bits erfaßt wurden, die zur Kodierung der Ursprungsblöcke im Rahmen notwendig sind. Da noch Bits hinzukommen, mit



Abbildung~4: Vergleich des SNR von Blockkodierung (strichliert) und fraktaler Kodierung (durchgezogene Linie).

denen die übrigen Parameter der Transformationen T_k kodiert werden müssen, geht der Vorsprung der fraktalen Kodierung wieder verloren.

Für korrelierte Abtastwerte läßt sich mit der fraktalen Kodierung genau wie bei der VQ (vgl. [4]) ein erheblich höherer SNR erreichen. Ausschlaggebend dafür ist, wie auch die "rate-distortion-theory" ([1]) vorhersagt, die Korrelationsmatrix der Abtastwerte eines Blockes.

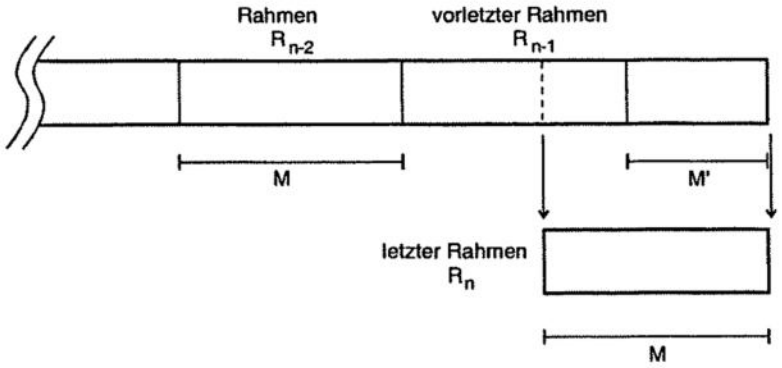
4 Praktische Implementierung

4.1 Problem der unvollständigen Rahmen

Rahmenorientierte Verfahren stehen immer vor dem Problem, unvollständige Rahmen behandeln zu müssen. Dieses Problem ist bei der fraktalen Kodierung dadurch verschärft, dass zum einen die Rahmen besonders groß gewählt werden müssen, um einen akzeptablen SNR zu erreichen, und dass zum anderen der Rahmen selbst das Kodebuch der fraktalen Kodierung ist. Zwei bekannte Möglichkeiten, unvollständige Rahmen zu kodieren sind damit ausgeschlossen:

- Der unvollständige Rahmen kann nicht unverändert kodiert werden, da er weniger Abtastwerte enthält und der erreichbare SNR somit niedriger ist.
- Der unvollständige Rahmen kann auch nicht mit beliebigen Werten, z.B. mit Nullwerten aufgefüllt werden, da er das Kodebuch für die fraktale Kodierung ist.

Das Problem wurde dadurch gelöst, dass der unvollständige Rahmen mit Abtastwerten aus dem vorangegangenen Rahmen aufgefüllt wird (Bild 8).



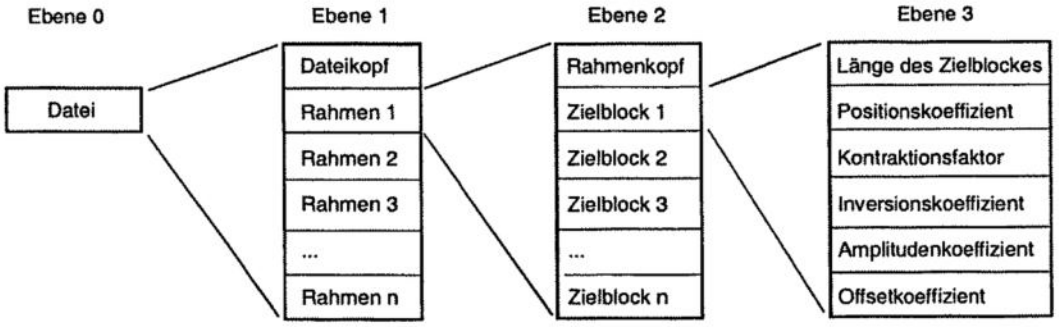
Abbildung~5: Auffüllen eines unvollständigen Rahmens mit Abtastwerten aus dem vorangegangenen Rahmen.

Damit die Abtastwerte, mit denen der Rahmen aufgefüllt wurde, nicht ein zweites Mal kodiert werden, muß nur der Beginn des ersten Zielblockes a_1 des unvollständigen Rahmens nach hinten verschoben werden. Es werden nur die Diese Verschiebung muß in der kodierten Datei gespeichert werden.

Mit dieser Lösung kann eine gleichbleibende Kodierqualität sichergestellt werden, ohne den Rechenaufwand oder den Speicherbedarf zu erhöhen.

4.2 Dateiformat

Für die Speicherung fraktal kodierter Sprachsignale wurde ein einfaches, rahmenorientiertes Dateiformat entworfen. Aufgrund seines experimentellen Charakters wurde auf eine Kanalkodierung sowie auf Maßnahmen der Kodesicherung und -korrektur verzichtet. Bild 6 zeigt den Aufbau des Formates.



Abbildung~6: Experimentelles Dateiformat für fraktal kodierte Signale.

Der Dateikopf enthält Angaben über die allgemeinen Parameter der fraktalen Kodierung, die verwendeten Rahmen- und Blocklängen und, sofern konstante Kontraktionsfaktoren und Blocklängen verwendet wurden, auch diese beiden Angaben.

5 Eigenschaften des rekonstruierten Sprachsignals

5.1 Charakterisierung im Zeitbereich

Da das hier behandelte Verfahren im Zeitbereich arbeitet, bietet sich zunächst eine Betrachtung im Zeitbereich an. Bild 7(a) zeigt das Zeitsignal des Wortes "Jawohl", digitalisiert mit einer Abtastfrequenz von 11,025 kHz bei einer gleichförmigen Auflösung von 8 bit. Darunter ist das Zeitsignal nach der Kodierung mit einer konstanten Blocklänge von $N = 16$ und der darauffolgenden Rekonstruktion dargestellt und in Bild 7(c) die Differenz beider Zeitsignale.

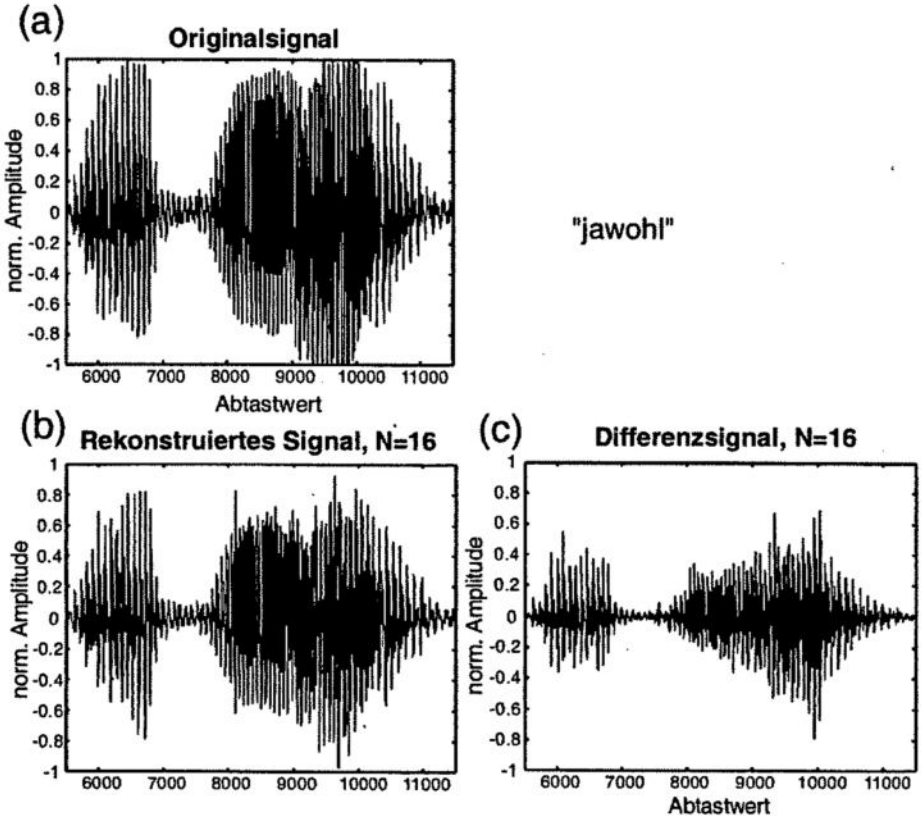
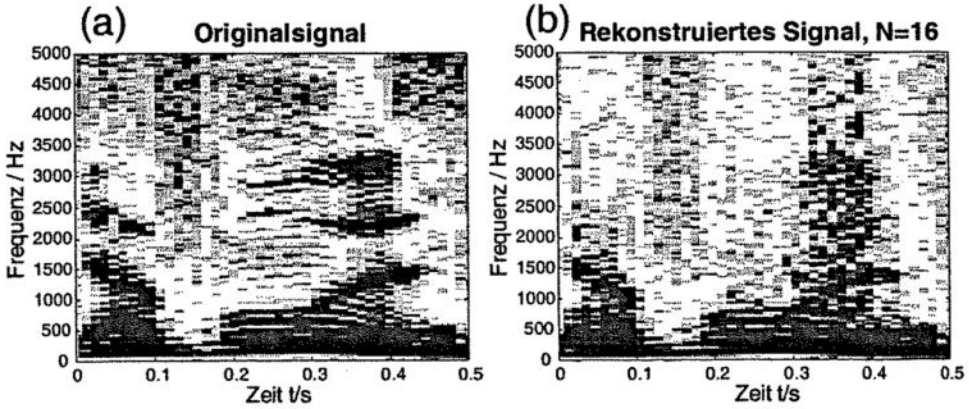


Abbildung 7: Zeitsignale des Originals (a) und nach der Kodierung und Rekonstruktion (b). Differenzsignal (c) zwischen (a) und (b).

Wie die Bilder zeigen, ist das Differenzsignal proportional zur Signalamplitude. Weitere Untersuchungen ergaben, daß das Differenzsignal mit der Blocklänge N zunimmt, da die Näherung der Zielblöcke mit größerem N immer ungenauer wird, wie sich auch im Verhalten des SNR zeigt. Bei Hörproben konnten zum Teil die gesprochenen Wörter und Sätze aus dem Differenzsignal herausgehört werden.

5.2 Charakterisierung im Frequenzbereich

Um die Sprachqualität und der Sprachverständlichkeit zu beurteilen, sind spektrale Darstellungen meist aussagefähiger als Zeitsignale. Aus diesem Grund wurden Spektrogramme der obigen Zeitsignale vor und nach dem Kodierungs- und Rekonstruktionsprozess aufgenommen (Bild 8).



Abbildung~8: Spektrogramme des Wortes "jawohl" vor (a) und nach einer fraktalen Kodierung (b) mit konstanter Blocklänge $N = 16$.

Wie in Bild 8(b) deutlich zu erkennen ist, werden die Formanten über 2 kHz nicht mehr oder nur stark gestört wiedergegeben. Dieses Verhalten ist auf die Kontraktion zurückzuführen, die, wie in Abschnitt 2.1 beschrieben, durch Unterabtastung oder durch Mittelwertbildung das Sprachsignal blockweise tiefpaßfiltert. Da der Kontraktionsfaktor in den obigen Experimenten auf $a = 2$ eingestellt wurde, liegt die Grenzfrequenz bei der halben Grenzfrequenz des Originalsignals, welche durch die Nyquistfrequenz f_N (= halbe Abtastfrequenz) gegeben ist. Damit liegt die Grenzfrequenz der fraktalen Kodierung bei $f_g = f_N/2 = f_s/4$ d.h. bei 11,025 KHz / 4 $\approx$ 2,7 kHz. Ein verringerter Kontraktionsfaktor könnte zwar die Grenzfrequenz nach oben verschieben, würde aber das Konvergenzverhalten verschlechtern oder sogar zur Divergenz führen.

Die fraktale Kodierung bringt zusätzlich noch ein breitbandiges Störsignal in das Spektrum ein, das in Hörproben als Prasseln wahrzunehmen ist. Dieses Störsignal kann durch eine geeignete Filterung ([10]) aus dem rekonstruierten Signal entfernt werden.

6 Ausblick

Ein Verfahren der Signalformkodierung von Sprache, wie die hier vorgestellte fraktale Kodierung, kann in Bezug auf die Signalqualität und den Kompressionsfaktor mit den parametrischen Sprachkodierverfahren kaum konkurrieren. In der weiteren Arbeit soll daher versucht werden, fraktale Parameter aus dem Sprachsignal zu extrahieren wie z.B. eine fraktale Dimension ([9]), um daraus ein Signalmodell und ein parametrisches Kodierverfahren zu entwerfen.

Literatur

[1] T. Berger. *Rate distortion theory. A mathematical basis for data compression.* Prentice-Hall, Englewood Cliffs, 1971.

[2] K. Fellbaum. *Sprachverarbeitung und Sprachübertragung.* Nachrichtentechnik. Springer, Berlin, 1984.

[3] A. Gersho and V. Cuperman. Vector quantization: A pattern matching technique for speech coding. *IEEE Commun. Mag.*, 21:15 ff, Dec. 1983.

[4] J.J.Y. Huang and P.M. Schultheiss. Block quantization of correlated gaussian random variables. *IEEE Trans. Commun. Syst.*, CS-11:289 – 296, Sept. 1963.

[5] N.S. Jayant and P. Noll. *Digital coding of waveforms.* Prentice-Hall, London, 1984.

[6] L. Lundheim. *Fractal Signal Modelling for Source Coding.* PhD thesis, Department of Electrical and Computer Engineering, Trondheim, September 1992.

[7] L. Lundheim. On the use of subsampling in fractal image compression. In Y. Fisher, editor, *Fractal Image Compression and Analysis*, pages 49 – 64, Berlin, 1998. NATO ASI Series F: Computer and Systems Sciences, Vol. 159, Springer.

[8] J. Makhoul, S. Roucos, and H. Gish. Vector quantization in speech coding. *Proceedings of the IEEE*, 73(11):1551–1588, November 1985.

[9] B.B. Mandelbrot. *Die fraktale Geometrie der Natur.* Birkhäuser, Basel, 1991.

[10] S. Schneider. Zum Einsatz fraktaler Methoden in der Sprachkompression. In D. Mehnert (Hrsg.), *Elektronische Sprachsignalverarbeitung, 10. Konferenz, Tagungsband*, Bd. 16 der *Studientexte zur Sprachkommunikation*, 104–111, 1999.

[11] S. Schneider. *Entwicklung und Analyse eines fraktalen Kodierverfahrens für Sprachsignale.* PhD thesis, BTU Cottbus, Fakultät für Maschinenbau und Elektrotechnik (eingereicht), 2000.

[12] P. Vary, U. Heute, and W. Hess. *Digitale Sprachsignalverarbeitung.* Informationstechnik. Teubner, Stuttgart, 1998.