Objektbasierter Analyse/Synthese Audio Coder für sehr niedrige Datenraten

Heiko Purnhagen, Bernd Edler

Laboratorium für Informationstechnologie
Uni Hannover, Schneiderberg 32, D-30167 Hannover
Tel.: +49 511 762 5033, Fax: +49 511 762 5052
E-Mail: purnhage@tnt.uni-hannover.de, edler@tnt.uni-hannover.de

Übersicht

In diesem Beitrag wird ein Verfahren zur Codierung von Audiosignalen mit Datenraten zwischen 4 und 16 kbit/s vorgestellt. Es beruht auf einer objektbasierten Darstellung des Eingangssignals durch Überlagerung verschiedener Teilquellensignale. Hierfür werden geeignete Teilquellenmodelle eingeführt. Anschließend wird eine Analyse/Synthese-Struktur beschrieben, die zur Extraktion der Teilquellensignale, zur Auswahl geeigneter Teilquellenmodelle und zur modellbasierten Parameterschätzung geeignet ist. Besonders wird auch darauf eingegangen, wie sich hierbei und bei der Codierung der zu übertragenden Parameter Wahrnehmungseigenschaften des menschlichen Gehörs ausnutzen lassen.

Weiterhin wird die Effizienz dieses Verfahrens mit anderen Codierverfahren verglichen, welche für Anwendung im gleichen Datenratenbereich verfügbar sind.

1 Einführung

Für Anwendungen im Internet und im Multimediabereich ergeben sich neue Anforderungen an Verfahren zur Quellencodierung von Audiosignalen. Zum einen werden sehr effiziente Codierverfahren benötigt, um auch Musiksignale mit Datenraten übertragen zu können, bei denen bisher im wesentlichen Sprachsignale übertragen wurden. Zum anderen ist auch in zunehmendem Maß eine Interaktivität erwünscht, die flexible Zugriffs- und Manipulationsmöglichkeiten für die codierten Daten bietet.

Derzeit werden für die Codierung von Audiosignalen mit Datenraten unter 16 kbit/s im wesentlichen zwei Typen von Codierverfahren eingesetzt. Dies sind zum einen Sprachcodecs, die auf Verfahren der linearen Prädiktion basieren und durch Ausnutzung der spezifischen Eigenschaften von Sprachsignalen für diese eine sehr hohe Codiereffizienz erzielen [1]. Mit derartigen Codierverfahren läßt sich in einem Bereich von Datenraten zwischen 4 und 16 kbit/s eine sehr gute Qualität bei Sprachsignalen erzielen, während sie für Musiksignale nur bedingt geeignet erscheinen. Den zweiten Typ stellen Codierverfahren für beliebige Audiosignale dar, die auf einer Spektralzerlegung des Eingangssignals basieren und durch eine geschickte Ausnutzung von Wahrnehmungseigenschaften des menschlichen Gehörs besonders erfolgreich für die Codierung hochqualitativer Audiosignale eingesetzt werden [2]. Im Bereich von Datenraten unterhalb von 16 kbit/s, in dem in der Regel deutlich wahrnehmbare Codierfehler auftreten, weisen diese Verfahren Probleme auf, insbesondere bei Sprach- und Einzelinstrumentsignalen.

Mit dem hier vorgestellten Codierverfahren wird versucht, die Lücke zwischen den beiden erwähnten Verfahren zu schließen und eine effiziente Codierung von Einzelinstrumentsignalen und von musikalischen Szenen relativ geringer Komplexität bei Datenraten zwischen 4 und 16 kbit/s zu ermöglichen. Bei diesem Verfahren wird im Gegensatz zu den beiden oben beschriebenen Verfahren darauf verzichtet, die Wellenform des Eingangssignals bei der Decodierung wieder approximieren zu können. Jedoch soll der Klangeindruck des Eingangssignals so gut möglich nachgebildet werden. Hierzu wurde ein objektbasiertes Analyse/Syntheseverfahren entwickelt, bei welchem das Eingangssignal in Teilquellensignale zerlegt wird, die aus einzelnen Sinustönen, aus harmonischen Klängen, oder aus Rauschsignalen bestehen können. Im Gegensatz zu aus der Literatur bekannten Ansätzen zur Nachbildung von Audiosignalen durch Sinustöne und Rauschsignale [3] werden bei dem hier vorgestellten Verfahren bei der Bestimmung und besonders auch bei der Codierung der zu übertragenden Parameter die Wahrnehmungseigenschaften des menschlichen Gehörs in hohem Maße berücksichtigt. Dies ist besonders wichtig zur Erzielung einer ausreichenden Qualität bei einer starken Einschränkung der Anzahl zu übertragender Parameter bei Datenraten zwischen 4 und 8 kbit/s.

In Abschnitt 2 wird zunächst das Konzept der objektbasierten Audiocodierung vorgestellt. Anschließend wird in Abschnitt 3 die Vorgehensweise zur Extraktion der Teilquellensignale und zur zugehörigen Parameterschätzung beschrieben. In Abschnitt 4 wird auf die Codierung der Parameter und in Abschnitt 5 auf die Decodierung und Synthese eingegangen. Einige experimentelle Ergebnisse und Ergebnisse subjektiver Bewertungen werden in Abschnitt 6 dargestellt. Abschnitt 7 beschließt diesen Beitrag mit einer Zusammenfassung und einem Ausblick auf mögliche Weiterentwicklungen.

2 Objektbasierte Audiocodierung

Das hier vorgestellte Codierverfahren basiert auf dem Ansatz, das Eingangssignal in Objekte zu zerlegen, die im folgenden als Teilquellensignale bezeichnet werden. Die Zerlegung ist dabei abhängig von den sog. Teilquellenmodellen mit deren Hilfe die Teilquellensignale parametrisch beschrieben werden können. Da es im allgemeinen sehr schwierig oder sogar unmöglich ist, ein Musiksignal exakt in die Signale einzelner Musikinstrumente zu zerlegen, werden hier Teilquellenmodelle nach ihrer Eignung für die effiziente Codierung von Musiksignalen ausgewählt. Während in einer ersten Implementierung des Analyse/Synthese Audio Coders [4] nur ein Teilquellenmodell für einzelne Sinustöne eingesetzt wurde, werden gegenwärtig folgende Teilquellenmodelle verwendet:

- Einzelne Sinustöne
- Harmonische Klänge
- Rauschsignale

Abbildung 1 zeigt die prinzipielle Funktionsweise des auf diesen Modellen basierenden Audiocoders. Das Audiosignal wird zunächst durch eine Separierungseinheit in geeignete Teilquellensignale zerlegt. Anschließend werden die zugehörigen Modellparameter geschätzt, quantisiert, codiert und übertragen. Zur Erzielung der höchstmöglichen Klangqualität wird die Quantisierung und Codierung der ermittelten Parameter durch ein geeignetes Wahrnehmungsmodell gesteuert. Falls die Datenrate nicht für die Codierung sämtlicher Teilquellensignale ausreicht, beinhaltet diese Steuerung auch die Auswahl von Teilquellensignalen gemäß ihrer Wichtigkeit für den subjektiven Klangeindruck.

Um die Codierverzögerung auf sinnvolle Werte für Anwendungen im Kommunikationsbereich zu beschränken, erfolgt die Verarbeitung des Audiosignals blockweise.

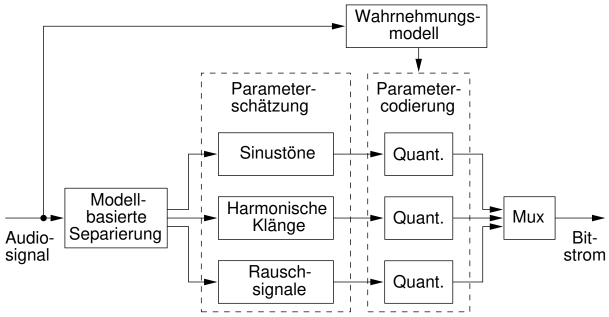

Abbildung 1: Blockdiagramm eines Audiocoders mit mehreren Teilquellenmodellen.

Bei der Decodierung werden zunächst die übertragenen Parameter rekonstruiert, aus welchen anschließend Teilquellensignale synthetisiert werden. Das Ausgangssignal ergibt sich durch Überlagerung sämtlicher Teilquellensignale (Abbildung 2).

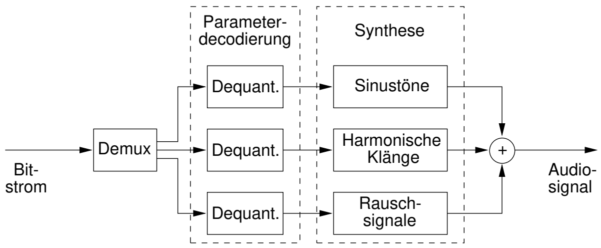

Abbildung 2: Blockdiagramm eines Audiodecoders mit mehreren Teilquellenmodellen.

3 Extraktion der Teilquellensignale

Die Zerlegung des Eingangssignals in einzelne Teilquellensignale und die Schätzung der Teilquellenparameter stellen den Kern des Analyse/Synthese Audio Coders dar. Dabei wird das Eingangssignal zunächst in Zeitabschnitte (Blöcke) fester Dauer (32 ms) unterteilt. Für jeden Block werden anschließend gemäß des verwendeten Quellenmodells die Parameter der Teilquellen ermittelt.

Für das Quellenmodell *einzelner Sinustöne* werden in einer Analyse/Synthese-Schleife (Abbildung 3) einzelne Sinustöne in der Reihenfolge ihrer Wichtigkeit aus dem Eingangssignal $x(n)$ extrahiert [4]. In jedem Schleifendurchlauf wird unter Berücksichtigung eines Psychoakustischen Modells zunächst ein grober Schätzwert $f_{m,i}$ der Frequenz des wichtigsten, noch nicht extrahierten Sinustons i aus dem Spektrum des Restsignals $e(n) = x(n) - s(n)$ ermittelt. Ausgehend von $f_{m,i}$ werden dann die Parameter Frequenz f_i, Amplitude a_i und Phase φ_i des Sinustons genau geschätzt [5]. Abschließend wird der Sinuston gemäß der geschätzten Parameter wieder synthetisiert und zum bisher synthetisierten Signal $s(n)$ addiert. Außerdem wird mit dem Psychoakustischen Modell aus den Parametern f_i und a_i der extrahierten Sinustöne die durch $s(n)$ hervorgerufene Maskierungsschwelle $M(k)$ berechnet [6].

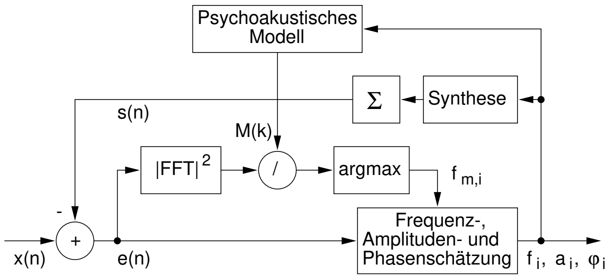

Abbildung 3: Analyse/Synthese-Schleife.

Um die Nachbildung transienter Signalanteile zu verbessern, wird aus dem Eingangssignal eine mit den Parametern r_{atk}, r_{dec} und t_{max} beschriebene Amplitudenhüllkurve ermittelt (Abbildung 4). Diese Hüllkurve wird dann für ausgewählte Sinustöne bei der Synthese verwendet.

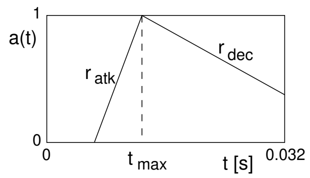

Abbildung 4: Parameter einer Hüllkurve.

Die Amplituden und Frequenzen der Sinustöne in aufeinanderfolgenden Blöcken werden miteinander verglichen, um so länger andauernde Sinustöne zu erkennen, die aus dem vorangegangenen Block fortgesetzt werden können.

Bei der Erweiterung des Quellenmodells um *harmonische Klänge* und *Rauschsignale* wird die Zerlegung des Eingangssignals und Schätzung der Teilquellenparameter in insgesamt drei Schritten durchgeführt:

Im ersten Schritt wird basierend auf einer Cepstralanalyse ein grober Schätzwert für die Grundfrequenz des ausgeprägtesten harmonischen Klangs im Eingangssignal bestimmt. Für den Grundton und die Obertöne dieses Klangs werden nun jeweils die genaue Frequenz und Amplitude geschätzt und aus den Frequenzen dieser Teiltöne wird die genaue Grundfrequenz des harmonischen Klangs berechnet. Da z.B. bei Saiteninstrumenten die Obertonfrequenzen nicht immer exakt ganzzahlige Vielfache der Grundtonfrequenz sind, wird zusätzlich ein Parameter ermittelt, der dieses sog. Stretching beschreibt.

Im zweiten Schritt werden mit der oben beschriebenen Analyse/Synthese-Schleife einzelne Sinustöne aus dem Eingangssignal extrahiert. Anhand der Frequenz wird für jeden extrahierten Sinuston entschieden, ob sich um einen Teilton des harmonischen Klangs oder um einen einzelnen Sinuston handelt. Nach Abschluß der Analyse/Synthese-Schleife werden auch die Teiltöne des harmonischen Klangs, die noch nicht extrahiert wurden, vom Restsignal $e(n)$ abgezogen.

Im dritten Schritt wird angenommen, daß mit den Teiltönen des harmonischen Klangs sowie den einzelnen Sinustönen alle ausgeprägten tonalen Signalanteile aus dem Eingangssignal extrahiert wurden, so daß im verbleibenden Restsignal $e(n)$ nur noch rauschartige Signalanteile enthalten sind. Das Restsignal wird daher als Rauschsignal mit einer spektralen Einhüllenden modelliert, die dem Betragsspektrum des Restsignals entspricht. Zur Beschreibung der spektralen Einhüllenden werden als Parameter die ersten Koeffizienten einer Cosinusreihenentwicklung des Betragsspektrums verwendet, die mit Hilfe einer Diskreten Cosinustransformation (DCT) ermittelt werden.

4 Codierung

Für die Übertragung der Parameter der Teilquellensignale zum Decoder ist eine geeignete Quantisierung und Codierung erforderlich. Dabei werden sowohl bei der Auswahl der zu übertragenen Parameter als auch bei der Quantisierung der Parameter die Wahrnehmungseigenschaften des menschlichen Gehörs berücksichtigt.

Im Fall des Quellenmodells *einzelner Sinustöne* müssen nur die Frequenzen und Amplituden der Sinustöne übertragen werden, da die Phasen der

Sinustöne keinen signifikanten Einfluß auf den wahrgenommen Klangeindruck haben. Bei einem im aktuellen Block neu beginnenden Sinuston werden die Frequenz und Amplitude des Sinustons jeweils mit logarithmischer Kennlinie quantisiert, um die Wahrnehmung von Tonhöhe und Lautstärke zu berücksichtigen. Bei einem fortgesetzten Sinuston werden dagegen nur die Änderung der Frequenz und Amplitude übertragen, da hierfür eine geringere Bitanzahl ausreicht. Falls erforderlich werden auch die Parameter der Amplitudenhüllkurve quantisiert und übertragen.

Bei einer Datenrate von 6 kbit/s können für jeden Block die Parameter von etwa 10 bis 20 Sinustönen übertragen werden. Die Anzahl der übertragenen Sinustöne variiert von Block zu Block und hängt unter anderem davon ab, wieviele Sinustöne aus dem vorangegangenen Block fortgesetzt werden können. Bei der Auswahl der zu übertragenen Sinustöne wird berücksichtigt, in welcher Reihenfolge die Sinustöne in der Analyse/Synthese-Schleife aus dem Eingangssignal extrahiert wurden. Auf diese Weise ist sichergestellt, daß jeweils die für den Klangeindruck wichtigsten Sinustöne übertragen werden.

Bei der Erweiterung des Quellenmodells um *harmonische Klänge* und *Rauschsignale* müssen auch die Parameter dieser Teilquellensignale quantisiert und übertragen werden. Die Grundfrequenz des harmonischen Klangs und die Amplituden der Teiltöne werden mit logarithmischer Kennlinie quantisiert. Da speziell bei einer niedrigen Grundfrequenz viele (bis zu 100) Obertonamplituden zu codieren sind, werden ab dem 10. Oberton die Obertöne zu Gruppen zusammengefaßt und es wird nur die mittlere Obertonamplitude einer Gruppe übertragen. Auf diese Weise wird die mit zunehmender Frequenz abnehmende Frequenzselektivität des menschlichen Gehörs ausgenutzt.

Die Parameter, die die spektrale Einhüllende des Rauschsignals beschreiben, werden zunächst normiert und der zur Rekonstruktion notwendige Skalierungsfaktor wird mit logarithmischer Kennlinie quantisiert. Anschließend werden die normierten Parameter gleichförmig quantisiert und ebenfalls übertragen.

Die Aufteilung der zur Verfügung stehenden Datenrate auf die Parameter der einzelnen Sinustöne, des harmonischen Klangs und des Rauschsignals wird in Abhängigkeit von der Wichtigkeit der einzelnen Teilquellensignale für den Klangeindruck automatisch gesteuert.

5 Decodierung und Synthese

Bei der Decodierung eines Blocks werden zunächst, wenn vorhanden, die Parameter für die Amplitudenhüllkurve rekonstruiert, aus welchen anschließend die Hüllkurvenfunktion generiert wird.

Für Sinustöne werden entsprechend der in Abschnitt 4 beschriebenen Codierung die Parameter Frequenz und Amplitude rekonstruiert. Wird für einen Sinuston signalisiert, daß er vom vorangegangenen Block fortgesetzt werden soll, so wird bei der anschließenden Synthese die Phase stetig fortgesetzt. Auch bei Frequenz- und Amplitudenänderungen werden mittels Interpolation Unstetigkeiten vermieden. Bei Tönen, die im aktuellen Block beginnen, wird durch Einblenden dafür gesorgt, daß beim Einschaltvorgang keine unangenehmen Störgeräusche entstehen. Ebenso werden Töne, die nicht mehr fortgesetzt werden, ausgeblendet. Wird für einen Ton die Verwendung der Amplitudenhüllkurve signalisiert, so wird zusätzlich eine Multiplikation mit der zuvor generierten Hüllkurvenfunktion durchgeführt.

Für harmonische Klänge werden nach der Rekonstruktion der Grundfrequenz die Teiltonamplituden rekonstruiert und den entsprechenden Teiltonfrequenzen zugeordnet. Die anschließende Synthese der Teiltöne entspricht der Synthese einzelner Sinustöne.

Rauschsignale werden synthetisiert, indem zunächst die spektrale Einhüllende mittels einer inversen DCT rekonstruiert wird. Hieraus wird ein komplexwertiges Kurzzeitspektrum gebildet, indem die Werte der spektralen Einhüllenden als Beträge mit Zufallswerten für die Phasen kombiniert werden. Anschließend erhält man mittels einer inversen DFT das synthetisierte Teilquellensignal.

Eine interessante Eigenschaft des vorgestellten Codierverfahrens ergibt sich aus der Möglichkeit, bei der Synthese die Blocklänge zu variieren. Hierdurch läßt sich die Wiedergabegeschwindigkeit bei der Decodierung bei unveränderter Tonhöhe beliebig manipulieren.

Ebenso ist es möglich, sämtliche Frequenzparameter vor der Synthese mit einem Faktor zu multiplizieren und so die Tonhöhe des decodierten Signals auf sehr einfache Weise unabhängig von der Wiedergabegeschwindigkeit zu beeinflussen.

Diese Möglichkeiten zur unabhängig voneinander einstellbaren Tonhöhe und Geschwindigkeit eröffnen besonders im Multimediabereich neue Anwendungen, wenn z.B. unabhängig voneinander codierte Audiosignale zur Generierung einer neuen Szene synchronisiert werden müssen.

6 Ergebnisse

Zur Beurteilung des objektbasierten Analyse/Synthese Audio Coders wird zunächst das Verhalten der Analyse/Synthese-Schleife betrachtet. In Abbildung 5 sind die Kurzzeitspektren des Originalsignals und des aus den extrahierten einzelnen Sinustönen synthetisierten Signals für einen Block aus dem Signal "male speech" dargestellt. Außerdem ist die für das synthetisierte Signal ermittelte Maskierungsschwelle eingezeichnet. Es ist zu erkennen, daß in diesem Fall 13 Sinustöne extrahiert wurden, mit denen die wesentlichen tonalen Spektralkomponenten des Originalsignals gut nachgebildet werden.

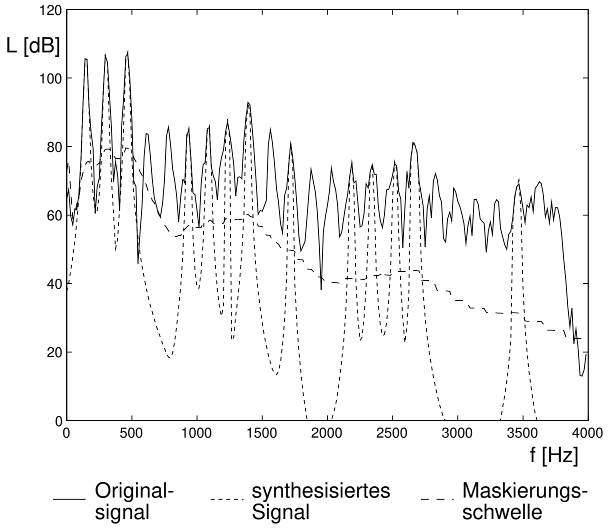

Abbildung 5: Signalspektren für "male speech".

In Abbildung 6 sind die Parameter Frequenz und Amplitude der extrahierten *einzelnen Sinustöne* und der Teiltöne des *harmonischen Klangs* für einen Ausschnitt aus dem Signal "female vocal" dargestellt. Für lange Zeitabschnitte liegt dabei ein stimmhafter Laut vor, dessen Grundton und Obertöne gut zu erkennen sind. Im Zeitbereich um 2,5 s ist ein ausgeprägter Formant bei ca. 2,6 kHz zu erkennen und im Zeitbereich um 2,7 s ist eine Änderung der Grundfrequenz zu beobachten. In kurzen Zeitabschnitten um 2,25 s und 3,0 s liegt ein stimmloser Zischlaut vor, der durch ein *Rauschsignal* nachgebildet wird. In diesen Zeitabschnitten sind daher nur Sinustöne mit sehr geringen Amplituden zu erkennen.

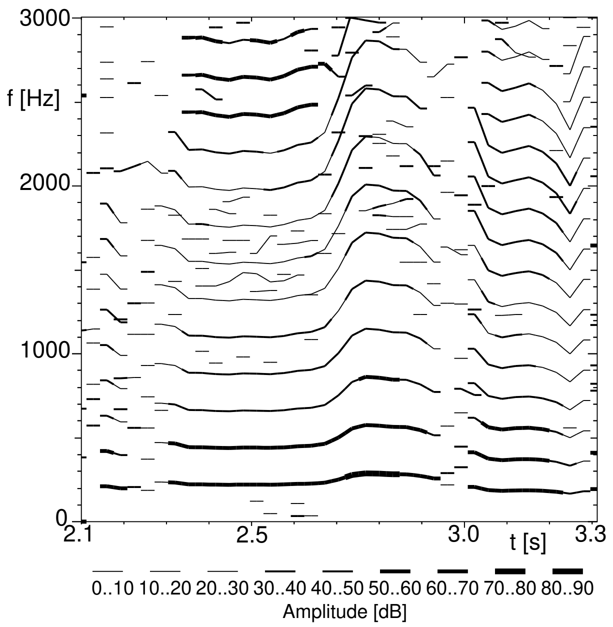

Abbildung 6: Parameter der Sinustöne für "female vocal".

Eine erste Implementierung des vorgestellten objektbasierten Analyse/Synthese Audio Coders, die nur das Teilquellenmodell für *einzelne Sinustöne* unterstützte [4], wurde bei den laufenden ISO/IEC MPEG-4 Standardisierungsaktivitäten als Vorschlag eingebracht [7] und in Hörtests mit anderen Vorschlägen verglichen. Hierbei ergab sich, daß dieses Verfahren typischerweise bei Datenraten um 6 kbit/s für Einzelinstrumentsignale und musikalische Szenen relativ geringer Komplexität eine deutlich bessere Tonqualität ermöglicht als Sprachcoder und Coder, die auf einer Spektralzerlegung des Eingangssignals basieren. Bei Sprachsignalen weisen jedoch erwartungsgemäß die spezialisierten Sprachcoder eine deutlich höhere Codiereffizienz auf. Bei sehr komplexen Musiksignalen erzielen andererseits die Coder, die auf einer Spektralzerlegung des Eingangssignals basieren, mit Abstand die beste Tonqualität. Daher wurde entschieden, in diesem Datenratenbereich alle drei erwähnten Verfahren zu unterstützen [8, 9].

Mittlerweile wurde auch der hier vorgestellte Vorschlag zur Erweiterung um die Teilquellenmodelle für *harmonische Klänge* und *Rauschsignale* bei MPEG-4 eingebracht [10] und im einem sog. Core-Experimentes untersucht. Dabei hat sich eine deutliche Verbesserung der subjektiven Tonqualität bei komplexem Musiksignalen gezeigt, so daß nun alle drei Teilquellenmodelle bei MPEG-4 berücksichtigt werden.

7 Zusammenfassung

In diesem Beitrag wurde ein Verfahren zur objektbasierten Analyse/Synthese-Codierung von Audiosignalen vorgestellt. Es basiert darauf, daß Parameter zur Darstellung von Sinustönen, harmonischen Klängen und Rauschsignalen codiert und übertragen werden. Die zu übertragenden Parameter für die Teilquellensignale werden hierbei iterativ unter Einbeziehung eines Psychoakustischen Modells bestimmt. Zusammen mit einer geeigneten Codierung der Parameter wird es somit möglich, Wahrnehmungseigenschaften des menschlichen Gehörs zur Erzielung einer möglichst hohen Tonqualität bei Datenraten hinunter bis zu 4 kbit/s auszunutzen. Mit den genannten Teilquellenmodellen wird eine sehr effiziente Codierung für Musiksignale relativ geringer Komplexität ermöglicht, so daß dieser Ansatz bei der Entwicklung des zukünftigen MPEG-4 Standards zur Codierung von Audiosignalen berücksichtigt wird [11]. Ein weiterer Vorteil dieses Verfahrens für Anwendungen im Multimediabereich ist die Möglichkeit, bei der Decodierung die Geschwindigkeit und die Tonhöhe des generierten Audiosignals unabhängig voneinander zu variieren. Der hier vorgestellte Coder wird auch in einer Applikation als dynamisch skalierter Audioplayer für das Internet vorbereitet [12].

Um diesem Verfahren ein breiteres Anwendungsfeld zu eröffnen, wird derzeit an der Kombination mit einem parameterbasierten Sprachcoder gearbeitet, welcher für Sprachsignale bei Datenraten zwischen 2 und 4 kbit/s eine sehr hohe Codiereffizienz bietet und daher im MPEG-4 Standard berücksichtigt wird [13].

Die Forschungsarbeiten, die diesem Beitrag zugrunde liegen, werden im Auftrag und in Zusammenarbeit mit der Deutschen Telekom Berkom durchgeführt.

Literatur

[1] P. Noll, "Speech Coding for Communications," *Proc. Eurospeech*, 1993.

[2] K. Brandenburg und M. Bosi, "Overview of MPEG Audio: Current and Future Standards for Low Bit Rate Audio Coding," *J. Audio Eng. Soc.*, Vol. 45, No. 1/2, Jan./Feb. 1997.

[3] R. J. McAulay und T. F. Quatieri, "Speech Analysis/Synthesis Based on a Sinusoidal Representation," *IEEE Trans. on ASSP*, Vol. 34, No. 4, Aug. 1986.

[4] B. Edler, H. Purnhagen und C. Ferekidis, "ASAC - Analysis/Synthesis Audio Codec for Very Low Bit Rates," *100th AES Convention*, Mai 1996.

[5] S. Kay, "A Fast and Accurate Single Frequency Estimator," *IEEE Trans. on ASSP,* Vol. 37, No. 12, Dez. 1989.

[6] C. Ferekidis, "Kriterien für die Auswahl von Spektralkomponenten in einem Analyse/Synthese-Audiocoder," Diplomarbeit, Universität Hannover, Institut für Theoretische Nachrichtentechnik und Informationsverarbeitung, März 1997.

[7] B. Edler, H. Purnhagen und C. Ferekidis, "Detailed Technical Description of the MPEG-4 Audio Coding Proposal from University of Hannover and Deutsche Telekom AG." ISO/IEC JTC1/SC29/WG11 MPEG96/0843, März 1996.

[8] B. Edler, "Current Status of the MPEG-4 Audio Verification Model Developement," *101th AES Convention*, Nov. 1996.

[9] B. Edler, "Very Low Bit Rate Audio Coding Developement," *AES 14th International Conference*, Juni 1997.

[10] H. Purnhagen, B. Edler und C. Ferekidis, "Proposal of a Core Experiment for extended 'Harmonic and Individual Lines plus Noise' Tools for the Parametric Audio Coder Core." ISO/IEC JTC1/SC29/WG11 MPEG96/2480, Juli 1997.

[11] ISO/IEC JTC1/SC29/WG11, "MPEG-4 Audio Committee Draft." ISO/IEC JTC1/SC29/WG11 N1903, Okt. 1997.

[12] B. Feiten, R. Schwalbe und F. Feige, "Scalable Internet Audio Transmission." Deutsche Telekom Berkom, Vortrag eingereicht zur ECMAST'98, 1998.

[13] M. Nishiguchi, J. Matsumoto, S. Omori und K. Iijima, "Technical description of Sony IPC's proposal for MPEG-4 Audio and Speech Coding." ISO/IEC JTC1/SC29/WG11 MPEG95/0321, Nov. 1995.