Breitbandsprachcodierung auf Basis von Filterbänken und Verdeckungseffekten

Ralf Th. Pietsch Arild Lacroix

Institut für Angewandte Physik, Johann Wolfgang Goethe-Universität, Frankfurt am Main

1 Einführung

Die Qualität schmalbandiger Telefonsprache mit Bandbreiten von ca. 3,4 kHz ist für fortschrittliche Anwendungen in der Telekommunikation nicht befriedigend. Breitbandsprache, mit einer Bandbreite von 7 kHz (Datenrate 16 kHz), ermöglicht höchste Sprachqualität im Vergleich zu Schmalbandsprache. Bei der Übertragung von Breitbandsprache ist eine Datenreduktion unerläßlich, weil die Datenrate von PCM Breitbandsprache mehr als doppelt so groß ist, wie die für Schmalbandsprache.

In diesem Beitrag wird ein Sprachcodierungssystem vorgestellt, welches sich aufgrund der niedrigen Laufzeit gut für die Kommunikation mit Breitbandsprache eignet.

Die Frequenzanalyse und -synthese wird durch Filterbänke auf der Basis von rekursiven Filtern realisiert. Rekursive Filter erlauben niedrige Laufzeiten in der Filterbank. Die Codierung findet im wesentlichen unter Ausnutzung von Irrelevanzen statt. Hierbei spielen, neben optimierten nicht uniformen Quantisierern, insbesonderen psychoakustische Verdeckungseffekte eine wichtige Rolle.

Bild 1 zeigt den Aufbau des gesamten Systems, bestehend aus dem Coderteil mit vier Modulen und dem Decoderteil mit zwei Modulen. Das erste Modul „Vorverarbeitung“ im Coder, realisiert eine Pegelkontrolle und eine Pausendetektion. Die folgende „Analysefilterbank“ wird durch Filterbankbäume auf der Basis rekursiver Filter realisiert, hierbei wurden sowohl unterschíedliche Filterbankbäume, als auch unterschiedliche Filterprototypen untersucht. Das Modul „Verdeckungsprozessor“ entscheidet auf der Basis psychoakustischer Verdeckungseffekte, welche Teilbänder übertragen werden müssen und welche entfernt werden können. Für die Übertragung über einen Kanal werden die Teilbänder im Modul „Bitzuweisung und Quantisierung“ vorbereitet. Verschiedene Techniken für die Codierung der Teilbänder wurden untersucht. Das erste Modul im Decoder, die „Bitdecodierung“ rekonstruiert aus dem übertragenen Signal die einzelnen Teilbandsignale, welche durch die „Synthesefilterbank“ zu einem Ausgangssignal rekonstruiert werden.

2 Frequenzanalyse und -synthese

Zur Frequenzanalyse und -synthese der Sprachsignale werden in Bäumen angeordnete Zwei-Kanal-Filterbänke auf der Basis rekursiver Filter eingesetzt. Da die Kommunikation zwischen Gesprächspartner bei einer hohen Signallaufzeit stark beeinträchtigt wird, spielt in einem Dialogsystem die Laufzeit eine wesentliche Rolle. Die Verwendung rekursiver Filter erlaubt hohe Flankensteilheiten und eine gute

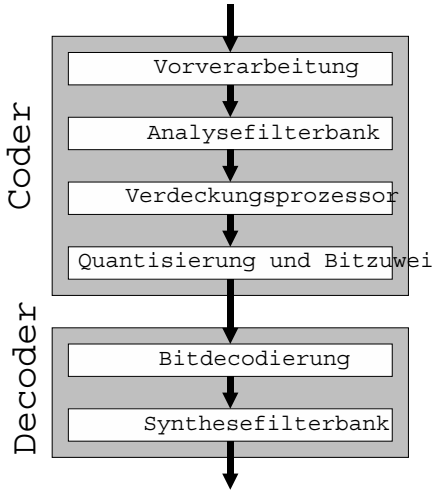

Bild 1: Modularer Aufbau des Systems bestehend aus Encoder und Decoder.

Teilbandtrennung bei niedriger Laufzeit im Vergleich zu linearphasigen nichtrekursiven Filtern. Tief- und Hochpaßfilter der Filterbänke werden auf der Basis der Allpaßzerlegung – welche eine effektive Implementierung in Bezug auf die Prozessorlast darstellt – realisiert. Verschiedene Filterbankbäume mit unterschiedlicher Kanalaufteilung wurden untersucht. Hierbei haben sich Filterbänke mit 15 und 19 Kanälen als vorteilhaft herausgestellt.

Bei den eingesetzten Filterbänken auf Basis rekursiver Filter tritt Phasenverzerrung auf, damit besitzen sie insbesondere keine konstante Gruppenlaufzeit. Phasenverzerrungen werden mit unterschiedlichen Methoden verringert: Unterschiedliche Prototypfilter in unterschiedlichen Ebenen des Filterbankbaums, Minimierung der Laufzeitspitzen durch Filterung und Laufzeitkompensationen innerhalb des Filterbankbaums [1]. Zur Abschätzung der Laufzeit einer solchen Filterbank kann eine *effektive Laufzeit* definiert werden, nach der 50% der Signalleistung des Eingangssignals am Ausgang vorgelegen hat. Liegt ein leistungserhaltenes System[1] vor, also ein System dessen Impulsantwort die Gesamtleistung Eins besitzt, so entspricht die effektive Laufzeit dem Zeitpunkt, an dem die akkumulierte Leistung der Impulsantwort den Wert 0,5 erreicht, bzw. überschreitet. Daß die Definition der effektiven Laufzeit sinnvoll ist, sieht man im Vergleich zur Laufzeit von linearphasigen Filtern. Linearphasige Filter besitzen symmetrische bzw. antisymmetrische Impulsantworten. Die Gruppenlaufzeit dieser Filter ist konstant, und

[1] Dieses System muß natürlich auch zeitinvariant und linear sein.

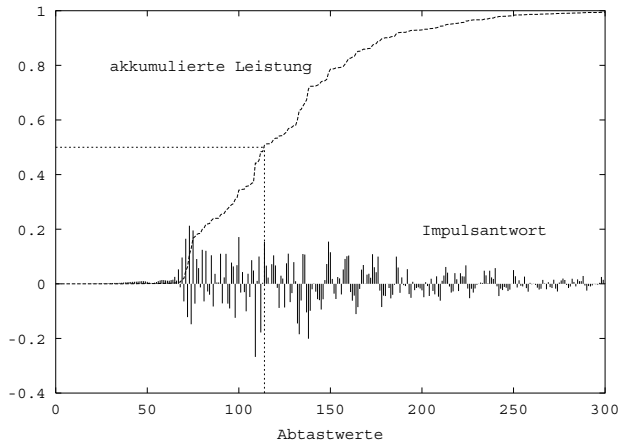

Bild 2: Impulsantwort einer der verwendeten 15-Kanal-Filterbank und der Verlauf der akkumulierten Leistung. Markiert ist die effektive Laufzeit, also der Zeitpunkt, an welchem die akkumulierte Leistung der Impulsantwort 0,5 überschreitet.

entspricht der halben Ordnung des Filters. Die effektive Laufzeit ist aufgrund der Symmetrie der Impulsantworten halb so lang wie die Impulsantwort. Gruppenlaufzeit und effektive Laufzeit sind also für linearphasige Filter gleich, was eine gewisse Berechtigung für die Definition der effektiven Laufzeit darstellt.

Bild 2 zeigt die Impulsantwort und den Verlauf der akkumulierten Leistung einer der verwendeten 15-Kanal-Filterbänke. Markiert ist der Zeitpunkt, an dem die akkumulierte Leistung der Impulsantwort 0,5 überschreitet, was der effektiven Laufzeit entspricht, da die Filterbank leistungserhaltend ist. Als Wert für die effektive Laufzeit ergibt sich im dargestellten Fall 114 Abtastwerte, bzw. 7,125 ms. In Bild 2 ist ferner zu sehen, daß die Impulsantwort sehr kompakt ist, der wesentliche Teil hat eine Ausdehnung von weniger als 200 Abtastwerten (entspr. 12,5 ms) und beginnt nach ca. 60 Abtastwerten (3,75 ms).

3 Verdeckungsprozessor

Im Modul *Verdeckungsprozessor* wird auf der Basis der Teilbandsignalleistungen entschieden, welche Teilbänder übertragen werden müssen, und welche verworfen werden können. Diese Entscheidung beruht auf simultanen und zeitlichen Verdeckungseffekten. Der verwendete Algorithmus nähert die simultanen Mithörschwellen durch Dreiecksfunktionen an. Durch die ungleiche Bandaufteilung, wird der Verlauf der Dreiecksfunktionen für die simultanen Mithörschwellen in Ihrem Verlauf verzerrt und den Mithörschwellen für Schmalbandrauschen und Sinustöne aus [2] angenähert. Die zeitlichen Mithörschwellen werden durch exponentiell abfallende Kurven dargestellt. Die zeitliche Verdeckung wird für jedes Teilband separat berechnet.

Die psychoakustische Bewertung der Teilbandsignale basiert somit auf vier Parametern, welche in weiten Bereichen variiert werden können und angepaßt werden müssen. Für einen bestimmten Filterbankbaum und abhängig von den verwendeten Tiefpaß-Prototypen, muß dieser Parametersatz in der Art angepaßt werden, daß im Hinblick auf die Datenratenreduktion und auf die Sprachqualität die bestmögliche Wirkung des Verdeckungsprozessors erreicht wird.

4 Codierung der Teilbänder

Die Signale der zu übertragenden Teilbänder werden im Modul *Bitzuweisung und Quantisierung* mit dem Ziel die Datenrate weiter zu reduzieren codiert. Es kommen sowohl Irrelevanz- also auch Redundanzreduktionen zum Einsatz. Unterschiedliche Codierungen – insbesondere Quantisierungen – wurden untersucht. Als effektiv hat sich eine nicht uniforme Quantisierung mit optimalen Kennlinien herausgestellt, welche den kleinstmöglichen SNR realisiert. Die Kennlinien für diese Quantisierer wurden nach dem Lloyd- bzw. LBG-Verfahren [3] berechnet. Die codierten Teilbandsignale werden abschließend unter Verwendung angepasster Huffman-Codes nocheinmal codiert, was die Datenrate weiter reduziert.

5 Realisierung

Der Coder arbeitet mit einer Abtastrate von 16 kHz. Die Sprachproben werden durch externe Filter mit Bandgrenzen von 7 kHz gefiltert. Das System wurde komplett in der Programmiersprache C realisiert, und erlaubt somit eine schnelle Portierung auf andere Platformen. Entwickelt wurde es auf einer DSP32C-Entwicklungsumgebung unter DOS, inzwischen wurde es auf Linux portiert. Zur Erzeugung des Maschinencodes wird der GNU-Compiler egcs verwendet, wobei keine prozessorspezifischen Optimierungen benutzt werden. Das System – Coder und Decoder – läuft auf einem PC mit AMD K6-II 400 MHz Prozessor unter Linux, je nach eingesetztem Teilbandcodierverfahren, in ca. einem Viertel der Echtzeit und schneller.

Mit informellen Hörtests wurde die Sprachqualität auf der Basis einer fünfstufigen Skala geschätzt. Für die Bestimmung der Datenraten wurden verschiedene Testsätze und zusätzlich fließend gesprochene Sprache benutzt. Parameterkonfigurationen, welche zu Datenraten zwischen 32 und 16 kbit/s führen, wurden genauer untersucht und wiederholt bewertet. Die Sprachqualität für diese Konfigurationen erreicht die zwei höchsten Kategorien: Gut und sehr gut.

Literatur

[1] Pietsch, R. Th., A. Lacroix: *Psychoakustisch motivierte Echtzeitcodierung von Breitband-Sprachsignalen.* Tagungsband der Deutschen Arbeitsgemeinschaft für Akustik DAGA '98, Zürich, S. 344–345, 1998.

[2] Zwicker, E., H. Fastl: *Psychoacoustics.* Springer-Verlag, Berlin, 1990.

[3] Linde, Y., A. Buzo, R. M. Gray: *An Algorithm for Vector Quantizer Design.* IEEE Trans. on Comm., Vol. 28(No. 1):84–95, Jan. 1980.