

Thilo Thiede

Eine rechenzeiteffektive gehörrichtige Filterbank mit signalabhängiger Filtercharakteristik

Ende 1998 wurde von der internationalen Telekommunikationsunion (ITU) eine Empfehlung für ein Verfahren zur objektiven Messung der wahrnehmbaren Qualität digital übertragener Audiosignale herausgegeben [2]. Die in diesem Standard enthaltene Modellvariante für die höchsten Genauigkeitsanforderungen ("advanced version") verwendet neben einem transformationsbasierten Gehörmodell auch eine gehörangepaßte Filterbank mit signalabhängigen Filtercharakteristiken. Der vorliegende Artikel beschreibt diese Filterbank sowie eine rechenzeiteffektive Filterstruktur zu ihrer Realisierung.

1 EINLEITUNG

Bei der digitalen Übertragung und Speicherung von Audiosignalen werden oft verlustbehaftete Datenreduktionsverfahren verwendet, in denen die spektrale Verteilung der entstehenden Quantisierungsfehler an die im menschlichen Gehör auftretenden Wahrnehmungsschwellen angepaßt wird. Die auf diese Weise in ihrer Lästigkeit reduzierten Störungen bleiben dabei in vollem Umfang physikalisch vorhanden, so daß konventionellen Meßverfahren, die lediglich die insgesamt vorhandenen Störungen erfassen, nicht zur Bestimmung der wahrgenommenen Qualität solcher Signale verwendet werden können. Daher wird die Qualität von gehörangepaßten Codierverfahren üblicherweise mittels subjektiver Hörtests bestimmt, die jedoch einen sehr hohen Zeit- und Kostenaufwand erfordern. Dieser Aufwand kann durch die Verwendung

objektiver Meßverfahren, die die zum subjektiven Qualitätseindruck führenden physiologischen und kognitiven Vorgänge simulieren, erheblich reduziert werden.

Um die Vielzahl der im Laufe der vergangenen Jahre hierzu vorgeschlagenen Meßmethoden durch ein einheitliches Verfahren zu ersetzen, wurde Ende 1998 in der ITU-R die Empfehlung BS.1387 „Methode zur objektiven Messung der wahrgenommenen Audioqualität“ verabschiedet [2]. Eine ausführliche Beschreibung der in dieser Empfehlung beschriebenen Meßmethode findet sich u. a. in [6]. Das Verfahren enthält sowohl ein transformationsbasiertes als auch ein filterbankbasiertes Gehörmodell. Letzteres besteht aus einer gehörangepaßten Filterbank mit signalabhängigen Filtercharakteristiken, die in ähnlicher Form bereits in dem Meßverfahren DIX [5] verwendet wurde. Diese Filter lassen sich als einfache FIR-Filter beschreiben, können aber auch durch einen rekursiven Algorithmus berechnet werden, wodurch der erforderliche Rechenaufwand erheblich verringert werden kann.

2 BESCHREIBUNG DER FILTERSTRUKTUR

Eine Sinusschwingung kann durch Berechnung der Ausdrücke $a_{n+1}=a_n \cdot \cos(\varphi) - b_n \cdot \sin(\varphi)$ und $b_{n+1}=a_n \cdot \sin(\varphi) + b_n \cdot \cos(\varphi)$ rekursiv fortgesetzt werden. Diese Berechnungsvorschrift kann als IIR-Filter erster Ordnung mit einem komplexen Koeffizienten $e^{j\varphi}$ gedeutet werden, das eine unendlich lange Impulsantwort in Form einer Sinusschwingung hat (d. h. das Filter ist nicht stabil). Indem am Filtereingang das um einen geeigneten Phasenwinkel gedrehte Eingangssignal mit einer Verzögerung von N Abtastwerten zum ursprünglichen Eingangssignal addiert wird, kann die Impulsantwort des Filters nach N Samples künstlich

abgebrochen werden. Bei Vernachlässigung der begrenzten Rechengenauigkeit kann das Filter dann als theoretisch stabil betrachtet werden. Das Filter hat keine Polstellen, aber N Nullstellen. Es verhält sich damit wie ein FIR-Bandpaß N -ter Ordnung, obwohl es als rekursives Filter aufgebaut ist. Aufgrund dieser Eigenschaft wird dieses (linearphasige) Filter im Folgenden als RFIR (rekursives FIR-Filter) bezeichnet. Wenn die komplexe Multiplikation in Real- und Imaginärteil aufgeteilt wird, ergibt sich die in Abb. 1 gezeigte Filterstruktur.

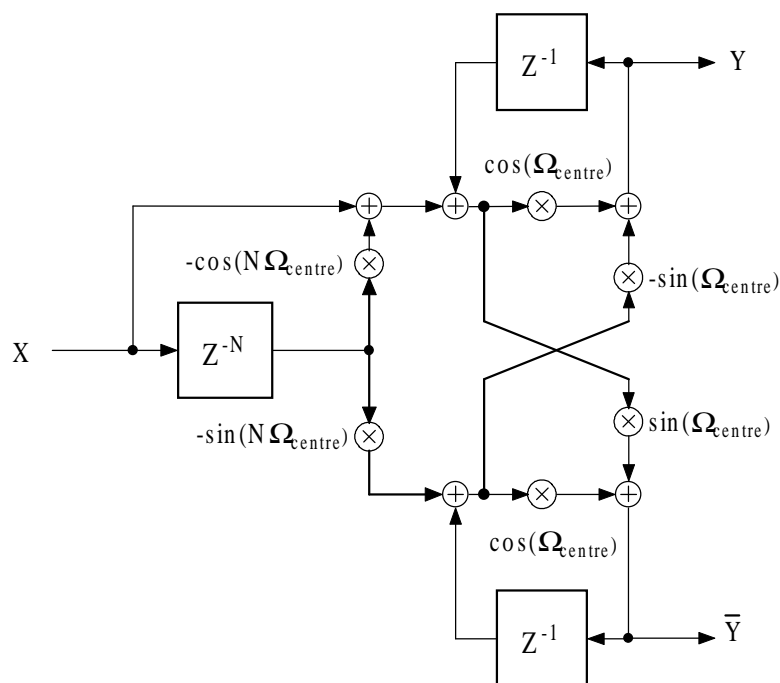


Abb. 1:

Rekursive Filterstruktur zur Realisierung eines linearphasigen Bandpasses mit Real- und Imaginärteilausgang.

Das in Abb. 1 gezeigte Filterelement weist noch keine sehr selektive Bandpaßcharakteristik auf. Die Amplitude der Nebenmaxima fällt nur umgekehrt proportional zum Abstand von der Mittenfrequenz ab. Durch Parallelschalten

von $K+1$ Filtern mit leicht versetzten Mittenfrequenzen kann die Selektivität des Filters jedoch erhöht werden. Bei einer geeigneten Gewichtung der parallelgeschalteten Teilfilter läßt sich ein Abfall der Nebenmaxima des so definierten RFIR-Filters K -ter Ordnung mit der $(K+1)$ -ten Potenz des Abstands zur Filtermittenfrequenz erreichen. Die Impulsantwort des Filters hat dann die Form

$$a_K(n) = \sin^K\left(\frac{\pi}{N}n\right) \cdot \cos\left(\frac{\Omega_{centre}}{f_{samp}} \cdot n\right) \quad \left| \quad 0 \leq n < N \right. \quad (1)$$

für den Realteil und

$$a_K(n) = \sin^K\left(\frac{\pi}{N}n\right) \cdot \sin\left(\frac{\Omega_{centre}}{f_{samp}} \cdot n\right) \quad \left| \quad 0 \leq n < N \right. \quad (2)$$

für den Imaginärteil. Mit $K=2$ entspricht die Filtercharakteristik damit der Frequenzselektivität einer Spektrallinie einer DFT mit Kosinusquadratfenster. Bei höherer Filterordnung nähert sich die Hüllkurve der Impulsantwort einer Gausskurve. Allerdings wird der beschriebene Algorithmus ineffizient, wenn K sehr groß wird. Für das in der ITU-Empfehlung BS.1387 [2] verwendete Gehörmodell hat sich eine Filterordnung von $K=2$ als ausreichend erwiesen.

Die oben beschriebene gewichtete Summation von $K+1$ parallelen Bandpaßfiltern kann als Faltung des Amplitudenfrequenzganges des Filters mit der Fouriertransformierten eines Zeitfensters aufgefaßt werden. Daher wird ein solches Filter im Folgenden als *FDC (frequency domain convolution)-Filter* bezeichnet.

3 AUFBAU DER FILTERBANK

Die in der ITU-Empfehlung BS.1387 [2] verwendete Filterbank läßt sich durch 40 FDC-Filter zweiter Ordnung realisieren, zwischen deren Ausgängen eine nach Real- und Imaginärteil getrennte gewichtete Summation vorgenommen wird, um die für das Gehörmodell geforderten exponentiellen Filterflanken zu erzeugen. Die Flankensteilheiten werden dabei entsprechend der in [4] vorgeschlagenen Näherung in Abhängigkeit vom Signalpegel gewählt. Dieses Vorgehen entspricht formal der üblicherweise in transformationsbasierten Gehörmodellen vorgenommenen spektralen Verschmierung, jedoch mit dem entscheidenden Unterschied, daß diese Verschmierung vor der Gleichrichtung vorgenommen wird. Es handelt sich daher um eine quasi-lineare Operation, und die aus der Filtertheorie bekannte Relation zwischen Frequenzgang und Impulsantwort bleibt erhalten. Die durch die Verschmierung verringerte spektrale Auflösung führt daher zu einer Verkürzung der Impulsantwort und somit zu einer erhöhten zeitlichen Auflösung. Dies gilt näherungsweise auch noch bei der Modellierung pegelabhängiger Filterflanken, wobei die flacheren Filterflanken bei hohen Signalpegeln gleichzeitig zu einer erhöhten zeitlichen Auflösung führen.

4 ABTASTUNG IM ZEIT- UND FREQUENZBEREICH

Ausgehend von einer mit der Filtercharakteristik gefalteten kontinuierlichen Zeit-Frequenz-Repräsentation der Signale kann die erforderliche Anzahl von Filterkanälen sowie die mögliche Unterabtastung der Filterausgänge mittels des Abtasttheorems abgeschätzt werden. Daraus ergibt sich eine Anzahl von 80 Filterbändern und eine Abtastung nach jedem achten Eingangswert. Bei der

experimentellen Optimierung des Modells zeigte sich jedoch, daß diese Werte auf 40 Filterbänder und eine Abtastung nach jedem 32. Eingangswert reduziert werden konnten, ohne die Zuverlässigkeit des Modells negativ zu beeinflussen.

Durch Ausnutzung dieser Unterabtastung sowie einiger Symmetrieeigenschaften der Filterkoeffizienten läßt sich der Rechenaufwand gegenüber der in Abb. 1 gezeigten Filterstruktur nochmals nahezu halbieren [7].

5 GLEICHRICHTUNG

Für ein physiologisch korrektes Modell der Signalverarbeitung im menschlichen Gehör wäre eigentlich eine Halbwellengleichrichtung die adäquate Gleichrichtungsmethode. Dies würde jedoch zu einer Reihe von Problemen bei der Weiterverarbeitung führen. Es wurde daher statt dessen die auch bei der Berechnung von DFT-Spektren verwendete Methode der Betragsbildung über die Hilbert-Transformierte gewählt. Dabei werden die Betragsquadrate des Signals und seiner Hilbert-Transformierten (die auch als Imaginärteil eines komplexen Signals oder als um 90° phasenverschobenes Signal interpretiert werden kann) addiert, woraus sich, wie sich mit Hilfe der Additionstheoreme für trigonometrische Funktionen leicht zeigen läßt, für den Fall einer Sinusschwingung das Amplitudenquadrat des Signals ergibt. Als besonders günstig erweist es sich hier, daß die in Abb. 1 gezeigte Filterstruktur bereits die Hilbert-Transformierte des Signals liefert. Der Hauptvorteil dieser Gleichrichtungsmethode liegt in der möglichen Unterabtastung der Filterausgänge und darin, daß für stationäre Eingangssignale auch ohne nachfolgende Glättung stabile Ausgangsamplituden vorliegen.

6 EIGENSCHAFTEN DER FILTERBANK

6.1 Spektrale Charakteristik

Die spektrale Charakteristik der Filterbank ergibt sich aus der Faltung der in (Gl. 1) und (Gl. 2) mit einer exponentiellen Verschmierungsfunktion nach [4]. Über der Bark-Skala nach [8] sind die Charakteristiken aller Teilfilter bis auf ihre Mittenfrequenz gleich und bestehen aus einer exponentiell ansteigenden Flanke, deren Steilheit bei niedrigen Signalpegeln bei ca. 24 dB/Bark liegt und bei hohen Signalpegeln bis auf 4 dB/Bark abflacht. Es folgt ein etwa 0.6 Bark breiter Durchlaßbereich und eine mit ca. 31 dB/Bark steil abfallende obere Flanke.

6.2 Zeitliche Auflösung

Die folgenden Abbildungen veranschaulichen das signalabhängige Verhalten der Filterbank anhand der Impulsantworten ihrer Teilfilter. Abbildung 2 zeigt die Impulsantwort der Filterbank ohne zeitliche Verschmierung. Man erkennt, daß die zeitliche Auflösung im oberen Frequenzbereich um ein Vielfaches höher ist als bei niedrigen Frequenzen (was einer der wichtigsten Vorzüge von Filterbänken gegenüber einer DFT ist).

Abb. 3 zeigt die Impulsantworten für ein einzelnes Filterband für zwei Impulse verschiedener Intensität. Die Abbildung zeigt, daß sich bei hoher Signalintensität die Impulsantwort des Filters tatsächlich entsprechend der durch die Pegelabhängigkeit der Filter verringerten spektralen Auflösung deutlich verkürzt.

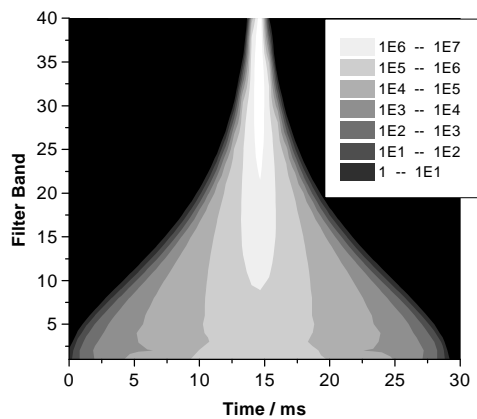


Abb. 2: Hüllkurve der Impulsantworten der FDC-Filterbank.

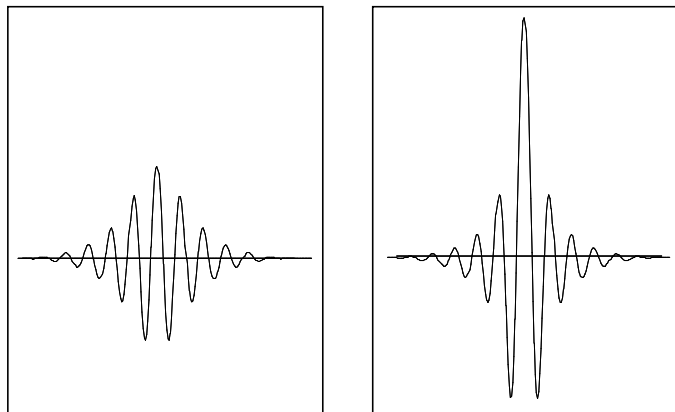


Abb. 3: Impulsantworten für ein einzelnes Filterband für zwei Impulse verschiedener Intensität (links: niedrige Energie, rechts: hohe Energie). Aus [5].

6.3 Rechenaufwand

Der Rechenaufwand für die in Abb. 1 gezeigten Filterstruktur beträgt 5 Additionen und 6 Multiplikationen pro Abtastwert. Durch Ausnutzung der Unterabtastung und der Symmetrieeigenschaften der Filterkoeffizienten lässt sich der Rechenaufwand auf ca. 4 Additionen und 2 Multiplikationen pro Abtastwert verringern. Da für die verwendeten FDC-Filter 2. Ordnung je drei Teilfilter parallelgeschaltet werden und die Filterbank aus 40 solchen Filtern besteht, beträgt der Rechenaufwand insgesamt also ca. 720 Rechenoperationen pro Abtastwert oder, bei einer Abtastrate von 48 kHz, ca. 34 Millionen Rechenoperationen pro Sekunde des Eingangssignals. Der Rechenaufwand ist damit zwar immer noch deutlich größer als in transformationsbasierten Verfahren, aber weit geringer als bei den meisten anderen in Gehörmodellen verwendeten Filterbänken.

6.4 Stabilität

Wie anfangs erwähnt, ist die beschriebene Filterstruktur nur unter der Voraussetzung unbegrenzter Rechengenauigkeit vollkommen stabil. Eine begrenzte Fließkommagenauigkeit führt zu einer mit der Signaldauer allmählich ansteigenden rauschartigen Störung, die solange toleriert werden kann, solange sie deutlich unterhalb der modellierten Ruhehörschwelle liegt. Bei einfacher (32 Bit) Fließkommagenauigkeit ist dies für Signaldauern bis ca. 60 Sekunden der Fall. Diese ist somit für die bei Codec-Vergleichen üblichen Testsignale im Allgemeinen ausreichend. Bei doppelter (64 Bit) Fließkommagenauigkeit erhöht sich diese Dauer theoretisch auf das 10^{12} fache. Somit lassen sich die Filter in diesem Fall als absolut stabil betrachten.

7 ZUSAMMENFASSUNG

Es wurde ein neues rechenzeiteffektives Filterbankkonzept vorgestellt, das sich zur Realisierung der in der ITU-R Empfehlung BS.1387 „Methode zur objektiven Messung der wahrgenommenen Audioqualität“ [2] festgelegten Filtercharakteristiken eignet. Die Filterbank ermöglicht eine gehörangepaßte Signalzerlegung mit signalabhängiger spektraler und zeitlicher Charakteristik bei vergleichsweise geringem Rechenaufwand.

8 LITERATUR

- [1] Glasberg, B. R.; Moore, B. J.: „DERIVATION OF AUDITORY FILTER SHAPES FROM NOTCHED NOISE DATA". *Hearing Research, Vol. 47*, 1990, S. 103-138.
- [2] Recommendation ITU-R BS.1387, *Method for objective measurements of subjectively perceived audio quality*, 1998.
- [3] Schroeder, M. R.; Atal, B. S.; Hall, J. L.: „OPTIMIZING DIGITAL SPEECH CODERS BY EXPLOITING MASKING PROPERTIES OF THE HUMAN EAR". *J. Acoust. Soc. Am.*, Vol. 66, S. 1647-1652, 1979.
- [4] Terhardt, E.: CALCULATING VIRTUAL PITCH. *Hearing Research, Vol. 1*, 1979, S. 155-182.
- [5] Thiede, T.; Kabot, E.: „A NEW PERCEPTUAL QUALITY MEASURE FOR BIT RATE REDUCED AUDIO". *Proceedings of the 100th AES Convention, Copenhagen 1996*, Preprint 4280.
- [6] Thiede, T.; Treurniet, W. C.; Bitto, R.; Schmidmer, C.; Sporer, T.; Beerends, J. G.; Keyhl, M.; Colomes, C.; Lever, M.; Stoll, G.; Brandenburg, K.; Feiten, B.: PEAQ - DER KÜNFTIGE ITU-STANDARD ZUR OBJEKTIVEN MESSUNG DER WAHRGENOMMENEN AUDIOQUALITÄT. 20. Tonmeistertagung, Karlsruhe, 22.11.1998.
- [7] Thiede, T.: PERCEPTUAL AUDIO QUALITY ASSESSMENT USING A NON-LINEAR FILTER BANK. Dissertation am Institut für Nachrichtentechnik und Theoretische Elektrotechnik der Technischen Universität Berlin. Eingereicht am 27.10.1998, Veröffentlichung ca. Mitte 1999.
- [8] Zwicker, E.; Feldtkeller, R.: „DAS OHR ALS NACHRICHTENEMPFÄNGER". Stuttgart: Hirzel Verlag, 1967.