

# Spektraltonhöhenanalyse unter Berücksichtigung von Akzentuierung

Claus von Rücker

Fachgebiet Akustische Kommunikation, Technische Universität München

Spektraltonhöhen sind diejenigen Tonhöhen, die im Gegensatz zu virtuellen Tonhöhen unmittelbar mit den Diskontinuitäten des Fourier-Spektrums eines Schallsignals korrespondieren. In einem hierarchischen Modell der Hörwahrnehmung (vgl. dazu Terhardt [9, S.311]) kommt ihnen die Bedeutung auditiver Primärobjekte zu. Analog zu primären Konturen in der visuellen Wahrnehmung können Spektraltonhöhen als primäre auditive Konturen verstanden werden. Die Extraktion von Spektraltonhöhen aus dem zeitvarianten Fourier-Spektrum des Schallsignals kann daher im übertragenen Sinne als *Konturierung* bezeichnet werden. Bei der Konturierung durch das Gehör treten spektrale und zeitliche Kontrasteffekte auf, die zur Steigerung der wahrgenommenen Prominenz von Spektraltonhöhen führen können. Diese sog. *Akzentuierung* steht im Zusammenhang mit der Tatsache, daß das Gehör jeden nur erdenklichen Hinweis zum Aufspüren der informationstragenden Signalmerkmale nutzt.

Im folgenden wird ein System zur Spektraltonhöhenanalyse vorgestellt, das diesen Vorgängen Rechnung trägt. Abb. 1 zeigt schematisch den Aufbau des auf einem Digitalrechner implementierten Systems; die Funktion der einzelnen Blöcke wird in den folgenden Abschnitten erläutert.

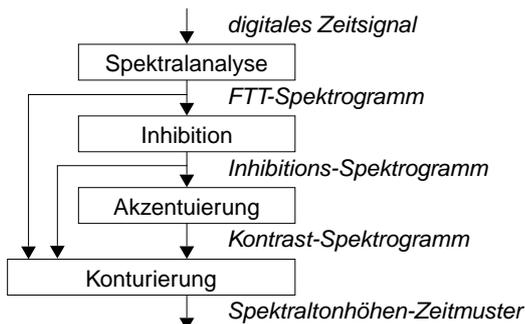


Abb. 1: Aufbau des Systems zur Spektraltonhöhenanalyse.

## 1. Spektralanalyse

Zur Nachbildung der Frequenzselektivität der Cochlea wird die digital implementierte Fourier-t-Transformation (FTT) [7] 4. Ordnung [4] verwendet. Bandbreite und Analysefrequenzabstand werden proportional zur Frequenzunterschiedsschwelle für Sinustöne (SPINC-Skalierung [8]) eingestellt. Der Parametersatz (siehe Tab. 1) erlaubt für den vorliegenden Anwen-

Tab. 1: Parameter der Spektralanalyse

Abtastfrequenz des Zeitsignals	16 kHz
Auswertintervall	1 ms
Abstand der Analysefrequenzen	8 SPINC
Anzahl der Analysefrequenzen	245
Frequenzbereich	20 Hz - 7,53 kHz
3 dB-Bandbreite der Analyse	25 SPINC

dungsfall eine optimale Anpassung an die Gehöreigenschaften unter besonderer Berücksichtigung der für die Tonhöhenwahrnehmung maßgeblichen Kriterien bei gleichzeitig effizienter Berechnung. Das Ausgangssignal der Einzelkanäle ist derart kalibriert, daß ein Frequenzgang entsteht, der dem gespiegelten Verlauf der Hörschwelle für Sinustöne [6] entspricht. Da das Zeitsignal der Spektralkanäle bandbegrenzt ist, muß das Spek-

trum nicht zu jedem Abtastzeitpunkt ausgegeben werden, sondern nur einmal je Auswertintervall. Nach dem Logarithmieren der Amplituden liegt am Ausgang des Funktionsblocks "Spektralanalyse" das diskrete, zeitvariante FTT-Leistungsspektrum vor, das sog. *FTT-Spektrum*.

## 2. Inhibition

Im Zusammenhang mit der Wahrnehmung von Spektraltonhöhen können gewisse Wirkungen lateraler Inhibition psychoakustisch nachgewiesen werden. So hat beispielsweise Fastl [1] die Wahrnehmung von Tonhöhen bei spektral begrenztem Rauschen untersucht. Bei ausreichend großer Bandbreite des Rauschens korrespondiert die Tonhöhenempfindung dabei mit der Frequenz der Bandgrenze. Dieses Ergebnis kann als Kantenüberhöhung interpretiert werden.

Zur Nachbildung lateraler Inhibition wird im vorliegenden System jedes FTT-Pegelspektrum mit einem diskreten, SPINC-skalierten Kern in Frequenzrichtung gefaltet. Form und Größe dieses Faltungskerns (siehe Abb. 2) orientieren sich an Daten

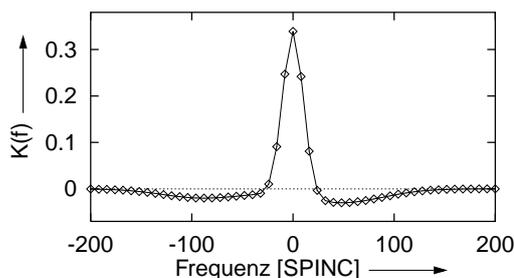


Abb. 2: Faltungskern zur Nachbildung lateraler Inhibition.

psychoakustisch gemessener Suppression von Sinustönen [3]. Mathematisch kann er im Kontinuierlichen beschrieben werden durch die Summe dreier unterschiedlich gewichteter, normierter Gauß-Funktionen  $G(f)$  der Form

$$G(f) = \frac{2\sqrt{\ln 2}}{B} \cdot \exp\left[-\left(\frac{2\sqrt{\ln 2}}{B} \cdot f\right)^2\right]. \quad (1)$$

Dabei ist  $B$  die Halbwertsbreite von  $G(f)$ . Der Faltungskern  $K(f)$  ergibt sich mit dem Gewicht  $w_{EX}$  der exzitatorischen Region und den Gewichten  $w_{IU}$  und  $w_{IO}$  von unterer und oberer hemmender Region zu

$$K(f) = w_{EX}G_{EX}(f) - w_{IU}G_{IU}(f - f_{IU}) - w_{IO}G_{IO}(f - f_{IO}). \quad (2)$$

Die hemmenden Bereiche sind dabei um  $f_{IU}$  bzw.  $f_{IO}$  gegen den exzitatorischen Bereich verschoben; die Halbwertsbreiten  $B_{IU}$  bzw.  $B_{IO}$  betragen etwa das Vierfache von  $B_{EX}$ . Die Gewichte  $w$  werden derart eingestellt, daß  $\int K(f)df = 0,5$  (Frequenz SPINC-skaliert). Damit wird erreicht, daß die Hälfte des Gleichanteils im Spektrum erhalten bleibt.

Das durch diskrete Faltung berechnete *Inhibitions-Spektrum* unterscheidet sich vom FTT-Spektrum durch folgende Eigenschaften: Spektrale Kontraste werden verstärkt, gering ausgeprägte Maxima der spektralen Hüllkurve werden unterdrückt, Kanten in spektraler Richtung erscheinen überhöht. Es macht eine Aussage über die relative Ausprägtheit der spektralen Maxima gegenüber ihrer näheren Umgebung.

### 3. Akzentuierung

Diejenigen Signalkomponenten eines Schalls, die transiente Anteile im zeitvarianten Spektrum verursachen, erfahren eine Steigerung ihrer wahrgenommenen Prominenz, sie werden *akzentuiert*. Der Funktionsblock "Akzentuierung" dient der Detektion dieser Signalkomponenten. Zu diesem Zweck wird das Inhibitions-Spektrogramm zunächst mit einem nichtlinearen Tiefpaß nach einem Ansatz von Zwicker [10] in Zeitrichtung geglättet. Dabei bleiben steigende Flanken unverändert, fallende Flanken werden entsprechend dem zeitlichen Verlauf der Nachverdeckung des Gehörs geglättet. Ein anschließender Hochpaß detektiert die transienten Anteile. Spektralbereiche, in denen große zeitliche Kontraste auftreten, verursachen im *Kontrast-Spektrogramm* große Amplituden. Dabei wird durch die nichtlineare Glättung eine Zunahme der Leistung in einem Spektralbereich höher bewertet, als ein Rückgang. Das Kontrast-Spektrogramm macht eine Aussage über diejenigen Akzentuierungseffekte, die einer zeitlichen Kontrastverstärkung entsprechen.

### 4. Konturierung

Die Aufgabe des Funktionsblocks "Konturierung" ist die Extraktion der Spektraltonhöhen unter Verwendung der drei vorliegenden Spektrogramme. Entsprechend ihrer wahrgenommenen Prominenz soll ihnen ein zeitvariantes Tonhöhengewicht zugeordnet werden. Dazu werden zunächst für jedes Auswertintervall nach dem Verfahren von Heinbach [2] lokale spektrale Maxima im Inhibitions-Spektrogramm detektiert. Diesen wird ein Teilton mit der Frequenz  $f$  und dem Pegel  $L_f$  des Maximums zugeordnet. Gleichzeitig wird der Pegel des FTT-Spektrogramms  $L$  und der Pegel im Kontrast-Spektrogramm  $L_A$  bei der gleichen Frequenz mit dem Teilton abgespeichert. Verfolgt man den Verlauf der Maxima über der Zeit, können durch Verwendung geeigneter Kriterien (geringer Frequenz- und Pegelunterschied benachbarter Teiltöne) sog. Teiltonlinien gebildet werden [4]. Tonale Komponenten werden durch längere Linien repräsentiert; dementsprechend werden nur Linien mit einer Dauer von über 80 ms extrahiert. Sie sind Kandidaten für Spektraltonhöhen; ihre zeitvariante Prominenz wird durch das folgende Verfahren bestimmt.

Die spektrale Maskierung wird durch Bestimmung des Pegelüberschusses für jeden Teilton einer Linie nach [6] unter Berücksichtigung aller, d.h. auch der nicht in Linien integrierten Teiltöne (geräuschhafte Anteile) abgeschätzt. Aus dem Pegelüberschuß und durch Verwendung einer bandpaßartigen Gewichtsfunktion, die den spektralen Dominanzbereich nachbildet, wird gemäß [6] ein vorläufiges Tonhöhengewicht  $0 < \tilde{W}_S < 1$  errechnet. Das Tonhöhengewicht  $W_S$  ergibt sich daraus durch Addition des durch zeitliche Akzentuierung verursachten Zuwachses gemäß  $W_S = \tilde{W}_S + \tilde{W}_S \cdot |L_A|$ . Da bei der vorliegenden Parametrisierung im interessierenden Frequenzbereich etwa  $|L_A| < 3$  gilt, bewegen sich die Werte für das Tonhöhengewicht  $W_S$  auch etwa im Bereich zwischen 0 und 3. Um diejenigen Spektraltonhöhen zu extrahieren, die für die Wahrnehmung von Bedeutung sind, kann durch Anwendung eines einfachen Schwellenkriteriums eine Auswahl getroffen werden. Es lautet: Für mindestens einen Zeitpunkt  $t$  muß  $W_S(t) > 0,2$  gelten. Um allen bekannten Akzentuierungseffekten Rechnung zu tragen, werden die Methoden zur Berechnung der Tonhöhengewichte noch verfeinert und weitere Auswahlkriterien entwickelt.

### 5. Anwendung

Summerfield *et al.* [5] konnten die Akzentuierung von Spektraltonhöhen durch ein Experiment nachweisen, bei dem ein harmonischer komplexer Ton mit glatter spektraler Hüllkurve eine Vokalwahrnehmung hervorruft. Durch die Wirkung eines vorausgehenden harmonischen komplexen Tons mit Lücken bei den Formantfrequenzen des Vokals wird die wahrgenommene Prominenz derjenigen Spektralkomponenten akzentuiert, die auf die spektralen Lücken folgen. Berechnet man das Spektraltonhöhen-Zeitmuster des Versuchsschalls mit dem oben beschriebenen Verfahren, wird die akzentuierende Wirkung des vorausgehenden Schalls durch vorübergehend erhöhte Tonhöhengewichte im Bereich der Formantfrequenzen (300 Hz und 900 Hz) des wahrgenommenen Vokals /u/ erkennbar (Abb. 3).

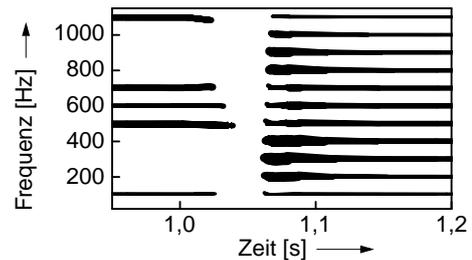


Abb. 3: Ausschnitt aus dem Spektraltonhöhen-Zeitmuster des Versuchsschalls von Summerfield *et al.* [5]. Die Strichstärke der Linien ist ein Maß für das Tonhöhengewicht.

Die Spektraltonhöhen-Zeitmuster von bandbegrenztem Rauschen entsprechend der Experimente von Fastl [1] zeigen übereinstimmend mit dem Versuchsergebnis, daß vor allem an den Bandkanten Tonhöhen extrahiert werden, die naturgemäß einer starken Fluktuation unterworfen sind (siehe Abb. 4).

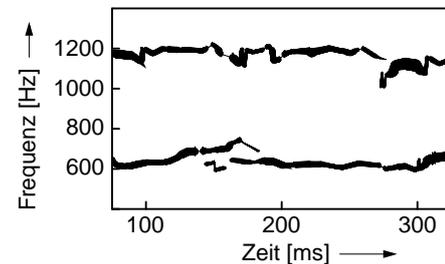


Abb. 4: Ausschnitt aus dem Spektraltonhöhen-Zeitmuster von bandbegrenztem Rauschen (Bandgrenzen 600 Hz und 1,2 kHz). Die Strichstärke der Linien ist ein Maß für das Tonhöhengewicht.

### Literatur

- [1] Fastl, H. (1971). *Acustica*, 25: 350-354.
- [2] Heinbach, W. (1988). *Acustica*, 67: 113-121.
- [3] Houtgast, T. (1974). *Lateral Suppression in Hearing*. TNO Report, Institute for Perception, The Netherlands.
- [4] Mummert, M. (1998). *Sprachcodierung mit Spektrogrammkonturen*. Fortschr.-Ber. 10/522, VDI Verlag.
- [5] Summerfield, Q., Haggard, M., Foster, J., Gray, S. (1984). *Perception & Psychophysics*, 35: 203-213.
- [6] Terhardt, E. (1982). *J. Acoust. Soc. Am.*, 71: 679-688.
- [7] Terhardt, E. (1985). *Acustica*, 57: 242-256.
- [8] Terhardt, E. (1992). *Acustica*, 77: 40-42.
- [9] Terhardt, E. (1998). *Akustische Kommunikation*. Springer.
- [10] Zwicker, E. (1984). *J. Acoust. Soc. Am.*, 75: 219-223.