

## Die Bedeutung der Fensterfunktion für die Fourier-t-Transformation als gehörgerechte Spektralanalyse

M. SCHLANG, M. MUMMERT\*

(TU München und Siemens AG, München; \* Lehrstuhl für Elektroakustik der TU-München)

### 1. Einleitung

Bei der Berechnung von Kurzzeitspektren ist aufgrund der Unschärferelation das Produkt aus Zeitauflösung und Bandbreite begrenzt. Die Fourier-Time-Transformation (FTT) nach Terhardt [1] weist hier einen günstigen Wert auf. Sie ist zudem für eine Audiotransformation besonders geeignet, da für jede Analysefrequenz unabhängig das Zeitfenster bzw. die Bandbreite frei wählbar sind. Die Berechnung der FTT für zeitdiskrete Signale ist durch Rekursionsformeln besonders einfach.

Wird die FTT jedoch mit der Fensterfunktion von der Form  $e^{-at}$  für eine Audiotransformation verwendet, so klingt ein anhand des FTT-Spektrums synthetisiertes Sprachsignal leicht hallig. Dies läßt sich durch das relativ langsame Ausklingen dieser Fensterfunktion erklären. Im vorliegenden Beitrag wird gezeigt, wie die gehörgerechte Spektralanalyse durch Modifikation der Fensterfunktion deutlich verbessert werden kann.

### 2. Die Fourier-t-Transformation als gehörgerechte Spektraltransformation

Die gehörgerechte Spektralanalyse eines Signals  $s(t)$  mit der FTT nach Terhardt [1] basiert auf einem Ansatz von Flanagan [2], der ein allgemeines, zeitvariantes komplexes Spektrum  $S(\omega, t)$  angibt:

$$S(\omega, t) = \int_0^t s(x) h(t-x) e^{-j\omega x} dx. \quad (1)$$

Dies ist ein Fourierintegral, in dem eine zeitlich verschiebbare Fensterfunktion  $h(t)$ , mit der das Signal  $s(t)$  gewichtet wird, enthalten ist. Die untere Integralgrenze beträgt Null da für  $s(t)$  Kausalität vorausgesetzt wird, die obere gibt den Analysezeitpunkt vor. Für eine feste Analysefrequenz  $\omega = \omega_A$  kann (1) auch als Faltung der Fensterfunktion mit dem mit  $\omega_A$  komplex modulierten Signal  $s(t)$  interpretiert werden.

$$q_{\omega_A}(t) = (s(t) e^{-j\omega_A t}) * h(t) \Big|_{\omega = \omega_A} \quad (2)$$

Man erhält somit den komplexen Zeitverlauf  $q_{\omega_A}(t)$  des FTT-Spektrums an der Frequenz  $\omega_A$  durch spektrale Verschiebung des Eingangssignals  $s(t)$  und anschließende Filterung mit einem Tiefpass mit der Impulsantwort  $h(t)$ . Seine verdoppelte 3dB-Grenzfrequenz  $f_g$  ergibt die Analysebandbreite  $B$ .

Um der spektralen Auflösung des Gehörs Rechnung zu tragen, werden für die gehörgerechte Signalanalyse die Analysefrequenzen  $\omega_A$  entlang der Frequenzachse im diskreten Abstand von 0,05 Bark festgelegt und die Analysebandbreite wird proportional zur Frequenzgruppenbreite gewählt.

### 3. Teiltonanalyse und Synthese

Die Beurteilung der Qualität des Analysesystems kann durch subjektive Hörtests erfolgen. Deshalb wurde ein Testsystem implementiert, das eine Synthese des analysierten Signals mit Hilfe des *Teiltonzeitmusters* (TTZM) [3] ermöglicht.

Zunächst wird, um unerwünschte Nebenmaxima zu unterdrücken, der quadrierte Betrag des Spektrums  $S(\omega, t)$  für jede Analysefrequenz mit einem Tiefpass erster Ordnung geglättet. Seine Grenzfrequenz liegt um den Faktor 2,5 über der jeweiligen Analysebandbreite. Aus dem geglätteten Spektrum werden dann in Auswertintervallen von 1 ms die Maxima zum Teiltonzeitmuster ausgewählt, die eine Ausgeprägtheit von mindestens 0,5 dB aufweisen. Die Amplituden aller anderen Spektralfrequenzen werden zu Null gesetzt, die Phasenbeziehungen nicht ausgewertet.

Dieses Teiltonzeitmuster kann mit einer *Resynthese* hörbar gemacht werden [3]. An den im TTZM vorgegebenen Frequenzen werden für jedes Auswertintervall Sinusgeneratoren mit den entsprechenden Amplituden angeregt, die zwischen den Intervallgrenzen linear interpoliert werden. Enthält das TTZM für ein Folgeintervall eine benachbarte Frequenz, so wird für einen stetigen Phasenübergang an der Intervallgrenze gesorgt. Die Summe der Signale der Generatoren ergibt das Synthesesignal.

#### 4. Kriterien zur Auswahl der Fensterfunktion

Zur Bewertung der Eignung einer Fensterfunktion  $h(t)$  bzw. ihrer Laplace-Transformierten  $H(p)$  für die FTT lassen sich vier Kriterien angeben. Hierfür werde zunächst ein kausales Testsignal  $s(t) = A \sin(\omega_T t)$  für  $t \geq 0$  in (2) eingesetzt und umgeformt:

$$q_{\omega_A}(t) = \frac{A}{2} e^{j(\omega_T - \omega_A)t} * h(t) + \frac{A}{2} e^{j(\omega_T + \omega_A)t} * h(t) \quad \text{für } t \geq 0. \quad (3)$$

Für die Betrachtung des wesentlichen Verhaltens des FTT-Spektrums an der Stelle  $\omega_A$  kann unter der Voraussetzung  $|\omega_T - \omega_A| \ll |\omega_T + \omega_A|$  und  $2\pi f_g \ll |\omega_T + \omega_A|$  der zweite Term in (3) vernachlässigt werden. Wird  $\omega_T' = \omega_T - \omega_A$  substituiert, so ergibt sich nach einer Laplacetransformation aus (3) eine Betrachtung, die äquivalent ist zur Antwort des Tiefpasses  $H(p)$  bei einem kausalen Eingangssignal  $s'(t) = A/2 e^{j\omega_T' t}$ :

$$Q_{\omega_A}(p) = \frac{A}{2} \frac{1}{p - j\omega_T'} H(p). \quad (4)$$

Die stationäre Selektionskurve  $|H(p)|_{p=j\omega}$  muß als *erstes* Kriterium monoton über  $|\omega|$  abfallen. Wären nicht monotone Bereiche vorhanden, so würde das stationäre Testsignal im Spektrum Nebenmaxima aufweisen.

Als *zweites* Kriterium darf  $H(p)$  als konjugiert komplexe Pole nur solche mit geringer Polgüte haben. Die Sprungantwort von  $H(p)$  wird damit nur geringe Überschwinger zeigen. Andernfalls gäbe es Nebenmaxima bei nichtstationären Signalen. Sei hierzu  $H(p)$  durch ein konjugiert komplexes Polpaar  $p_1, p_1^*$  charakterisiert. Nach Abklingen der Einschalteteffekte kann das Abschalten von  $s'(t)$  zu  $t = t_1 \gg 0$  so beschrieben werden:

$$|q_{\omega_A}(t)| = \text{const} \left| \frac{e^{p_1(t-t_1)}}{(p_1 - j\omega_T')(p_1 - p_1^*)} + \frac{e^{p_1^*(t-t_1)}}{(p_1^* - j\omega_T')(p_1^* - p_1)} \right| \quad (5)$$

Die komplexen Zeiger der beiden Summanden drehen gegenläufig und verkürzen sich mit der Zeit. Ihre Beträge sind für  $\omega_T' = 0$  und für  $|\omega_T'| \gg |p_1|$  gleich, sodaß  $|q_{\omega_A}(t)|$  nachfolgend Nulldurchgänge mit der Frequenz  $2\text{Im}\{p_1\}$  aufweist. Für Werte  $|\omega_T'| \approx |\text{Im}\{p_1\}|$  sind die Zeiger ungleich lang, umso mehr, je größer die Polgüte  $\eta = |p_1|/|2\text{Re}\{p_1\}|$  ist. Folglich kann  $|q_{\omega_A}(t)|$  hier keine Nulldurchgänge haben und es stellen sich Nebenmaxima ein. Kleine Polgüten verlangsamen die Zeigerdrehung. Deshalb ist ihre Länge und

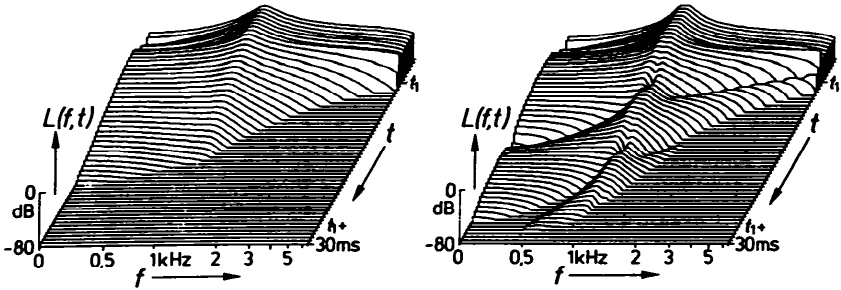


Fig. 1: FTT Spektrum eines zu  $t_1$  abgeschalteten Tons für Analysetiefpässe mit einem Poolpaar, Polgüte 0,5 (links) bzw. 0,7 (rechts), Bandbreite  $B = 1$  Bark zur besseren Sichtbarkeit der Nebenmaxima.

somit  $|q_{\omega_A}(t)|$  bereits stärker zurückgegangen, wenn durch ihre Drehung Nulldurchgänge in  $|q_{\omega_A}(t)|$  hervorgerufen werden.

Das *dritte* Kriterium fordert eine kleine Laufzeit von  $H(p)$ . Da in der FTT verschiedene Analysebandbreiten gleichzeitig verwendet werden, müssen die sich ergebenden Laufzeitunterschiede klein gehalten oder andernfalls kompensiert werden.

Als *viertes* Kriterium gilt die Normierung  $|H(0)| = 2$ , denn für die Frequenz  $\omega_T' = 0$  soll sich der stationäre Zustand  $|q_{\omega_A}(t)| = A$  (Amplitude des Signals nach (3)) einstellen.

Die Familie der unendlich langen, asymmetrischen Fensterfunktionen  $h_n(t)$  nach [4], deren Laplace-Transformierte  $H_n(p)$  einen reellen Pol  $n$ -ter Ordnung an der Stelle  $a$  aufweist, kann für kleine  $n$  allen diesen Kriterien genügen,

$$h_n(t) = \frac{2a}{(n-1)!} (at)^{n-1} e^{-at}; \quad H_n(p) = 2 \left( \frac{a}{p+a} \right)^n; \quad \tau_g(0) = \frac{n}{a}; \quad B = \frac{a}{\pi} \sqrt{\frac{1}{2^n - 1}}; \quad (6)$$

wobei  $\tau_g$  die Gruppenlaufzeit bei  $\omega = 0$  und  $B$  die Bandbreite ist. Die Exponentialfunktion modelliert in Näherung das vom Menschen bekannte natürliche Vergessen. Der Fall  $n = 1$  entspricht der in [1] vorgeschlagenen Fensterfunktion. Der Fall  $n = 2$  wird in [5] beschrieben und  $n = 3$  wird von [2] vorgeschlagen. Die Folge der Fenster strebt für  $n \rightarrow \infty$  gegen einen Gaußtieffpass mit einer unendlichen Laufzeit.

### 5. Die Fensterfunktion zweiter Ordnung

Unter dem Gesichtspunkt der Qualität des synthetisierten Sprachsignals ergibt sich für  $n = 2$  ein geeigneter Kompromiß zwischen Aufwand und Nutzen, insbesondere wird gegenüber  $n = 1$  eine deutliche Verbesserung erreicht. Dies liegt daran, daß die Analysebandbreite  $B$  in diesem Fall von 0,1 auf 0,25 Bark erhöht werden kann. Eine Erklärung hierfür liefert die verbesserte Weitabselektion des zugehörigen Tiefpasses aufgrund seiner erhöhten Flankensteilheit. Dadurch kann die Bandbreite vergrößert werden, ohne daß benachbarte Maxima beeinträchtigt werden. Die Vergrößerung erhöht das Zeitauflösungsvermögen. Obwohl sich nach (6) die Gruppenlaufzeit für Fenster höherer Ordnung vergrößert, wird dieser Effekt durch die Erhöhung der Bandbreite übertroffen, womit eine genauere Gleichzeitigkeit der Analyse gegeben ist.

Bei Vergrößerung von  $n$  auf Werte über 2 läßt sich die Bandbreite bestenfalls auf 0,3 Bark erhöhen, die daraus resultierende geringe Verbesserung der Zeitauflösung wird

durch den erhöhten Berechnungsaufwand und die Notwendigkeit der Laufzeitkompensation in Frage gestellt.

Bemerkenswert ist die kurzzeitige spektrale Verbreiterung beim Ein- und Ausschalten von stationären Sinustönen, die nur bei Fensterordnungen  $n > 1$  auftritt. Unmittelbar nach dem Schaltzeitpunkt hebt sich das FTT-Spektrum zu größeren  $\omega T$  hin an. Deshalb kann das Phänomen des Schaltknacks nur bei Spektralanalyse mit Fenstern der Ordnung  $n > 1$  in Erscheinung treten [6].

## 6. Realisierung der FTT als zeitdiskretes System

Zur Berechnung der FTT muß zunächst mit einem gängigen Filterentwurfverfahren die Z-Übertragungsfunktion  $H(z)$  eines rekursiven digitalen Tiefpasses zur zeitdiskreten Realisierung des gewünschten  $H(p)$  bestimmt werden. Für die Durchführung der Transformation gibt es zwei Möglichkeiten.

Das mit der Periode  $T$  abgetastete Signal  $s(nT)$  wird in einem *komplexen Modulator* mit der Funktion  $e^{-j\omega_A nT}$  multipliziert und anschließend der Real- und Imaginärteil jeweils getrennt mit  $H(z)$  gefiltert. Die Rekursionsformel für die Filterung ergibt sich aus der Differenzgleichung von  $H(z)$ . Die *sin-* und *cos-*Funktionen des Modulators können in Tabellen angelegt werden, auf die modulo, in durch die spezifische Analysefrequenz festgelegte Schrittweiten, zugegriffen werden kann [4].

Für die Realisierung als *komplexwertiges Digitalfilter* [5] wird (1) beidseitig mit  $e^{j\omega_A t}$  erweitert und auf die Betragsgleichung übergegangen. Die Modulation entfällt dadurch, es kommt das Digitalfilter

$$H'(z) = H(z) e^{-j\omega_A T} \quad (7)$$

zum Einsatz. Dabei handelt es sich um zwei verkoppelte reelle Filter, in denen die *sin-* und *cos-*Terme nicht mehr zeitabhängig sind. Der Berechnungsaufwand wird ab der Filterordnung größer eins von  $H(z)$  höher, der Modulator ist aber dafür entbehrlich.

## 7. Zusammenfassung

Der Einsatz einer Fensterfunktion zweiter Ordnung von der Form  $te^{-at}$  verbessert deutlich die Zeitauflösung der Fourier-t-Transformation als gehörgerechte Spektralanalyse gegenüber der Fensterfunktion erster Ordnung im Rahmen des verwendeten Testsystems. Grund hierfür ist die bessere Weitabselektion, die eine Vergrößerung der Analysebandbreite von 0,1 auf 0,25 Bark ermöglicht. Eine Steigerung auf Werte über 0,3 Bark ist aber hier auch bei Fenstern höherer Ordnung durch die Einschränkungen bei der Auswahl des Fensters nicht möglich.

## 8. Literatur

- [1] Terhardt, E.: "Fourier Transformation of Time Signals: Conceptual Revision", *Acustica* Vol. 57, 1985, S. 242-256.
- [2] Flanagan, J. L.: "Speech Analysis Synthesis and Perception", Springer-Verlag, Berlin, 1972.
- [3] Heinbach, W.: "Gehörgerechte Repräsentation von Audiosignalen durch das Teiltonzeitmuster", Dissertation, Techn. Univ. München, 1987.
- [4] Owens, F. J.; Murphy, M. S.: "A Short-Time Fourier Transform", *Signal Processing* Vol. 14, 1988, S. 3-10.
- [5] Schlang, M. F.: "An Auditory Based Approach for Echo Compensation with Modulation Filtering", *Europ. Conf. on Speech Communication and Technology, Proc. Eurospeech, Paris, 1989, S. 661-664.*
- [6] Mummert, M.: "Trennung von tonalen und geräuschhaften Anteilen im Sprachsignal", *Fortschritte der Akustik, DAGA'90, Bad Honnef, 1990, in diesem Tagungsband.*