

# Rücktransformation des Kurzzeitspektrums der Fourier-t-Transformation und Ansatz für eine gehörgerechte Transformationskodierung

M. MUMMERT

Lehrstuhl für Elektroakustik der Technischen Universität München

## Einleitung

Die Fourier-t-Transformation nach Terhardt [1] ist eine gehörgerechte Spektraltransformation, bei der sich die Zeitfensterlänge umgekehrt proportional zur Frequenzgruppenbreite des Gehörs verhält. Die Frequenzabhängigkeit des Fensters scheint einer Rücktransformation im Wege zu stehen, ein iterativer Ansatz von Owens und Murphy [2] führt zu deutlichen Übertragungsfehlern. In diesem Beitrag wird eine Lösung des Problems beschrieben.

Fortgeschrittene gehörbezogene Datenreduktionsverfahren für digitale Audiosignale arbeiten im Spektralbereich (Übersicht in [3]). Es konkurrieren zwei Ansätze, die Transformations- und die Teilbandkodierung. Der erste basiert auf Blocktransformationen, deren Frequenz/Zeit-Auflösungsvermögen nicht dem Gehör angepaßt werden kann. Der zweite Ansatz erlaubt hier eine Annäherung, die jedoch an starre Bandraaster gebunden ist. Dagegen besitzt die FTT zusammen mit ihrer Rücktransformation die Grundlage einer Frequenz/Zeit-kontinuierlichen Spektraltransformation. Durch Diskretisierung in Zeit und Frequenz entsteht ein Transformationspaar, das als Rahmen einer *gehörgerechten* Transformationskodierung vorgestellt wird.

## Fourier-t-Transformation (FTT)

Das FTT-Spektrum eines kausalen Signals  $s(t)$  wird allgemein angegeben durch:

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

wobei  $h(\omega, t)$  eine frequenzabhängige Fensterfunktion ist. Eine systemtheoretische Interpretation ergibt sich formal durch Betrachtung des Zeitverlaufs an einer festen Analysefrequenz  $\omega_A$ :

$$S_{\omega_A}(t) = S(\omega, t)|_{\omega=\omega_A} = (s(t)e^{-j\omega_A t}) * h_{\omega_A}(t). \quad (2)$$

Hierin wird  $s(t)$  auf einen komplexen Modulator gegeben, der eine spektrale Verschiebung um  $-\omega_A$  durchführt. Die nachfolgende zeitliche Faltung entspricht einer Tiefpaßfilterung durch ein kausales System, dessen Impulsantwort mit der Fensterfunktion identisch ist. Gehörgerechte Fensterfunktionen sind solche, die nicht oder bestenfalls schwach oszillieren und die keine spektralen Nebenmaxima aufweisen [4]. Die folgenden Betrachtungen werden am Beispiel der nichtoszillierenden Fensterfunktionen  $n$ -ter Ordnung mit  $n \geq 2$  durchgeführt:

$$h_{\omega_A}(t) = 2a \frac{(at)^{n-1}}{(n-1)!} e^{-at}, \quad \text{mit } a = \pi \frac{B_{3dB}}{\sqrt{2^{\frac{1}{n}} - 1}}. \quad (3)$$

Das zugehörige System hat einen  $n$ -fachen Pol bei  $-a$ . Die Bandbreite  $B_{3dB}$  verhält sich prportional zur Frequenzgruppenbreite  $\Delta f_G$ , für die ein analytischer Ausdruck verwendet wird [5].

## Rücktransformation der FTT (RFTT)

Zur Durchführung der RFTT wird das FTT-Spektrum mit einer frequenzabhängigen Korrekturfunktion  $x_{\omega_A}(t)$  gefaltet. Durch Fourier-Rücktransformation erhält man das Synthesesignal:

$$\hat{s}(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} (S_{\omega_A}(t) * x_{\omega_A}(t)) e^{j\omega_A t} d\omega_A. \quad (4)$$

Aus systemtheoretischer Sicht wird der Zeitverlauf des FTT-Spektrums bei  $\omega_A$  über ein Korrektursystem gefiltert und durch einen komplexen Modulator an die ursprüngliche spektrale Lage zurückgeschoben. Insgesamt wird dann über ein Kontinuum solcher Signale in  $\omega_A$  aufsummiert. Durch Einsetzen von Gl.(2) in (4) und Umformung läßt sich das Übertragungsverhalten von FTT und RFTT als Gesamtsystem angeben. Die Impulsantwort lautet:

$$g(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} (h_{\omega_A}(t) * x_{\omega_A}(t)) e^{j\omega_A t} d\omega_A. \tag{5}$$

Idealerweise sollte dies eine Dirac-Funktion sein. Die Korrekturfunktion  $x_{\omega_A}(t)$  soll nun so bestimmt werden, daß ein möglichst steiler Impuls entsteht. Dazu wird der Zeitpunkt  $t_{max,\omega_A}$  und der Wert  $h_{max,\omega_A}$  bei der maximalen Öffnung der Fensterfunktion definiert:

$$max\{|h_{\omega_A}(t)|\} = |h_{\omega_A}(t_{max,\omega_A})| \quad \text{und} \quad h_{max,\omega_A} = h_{\omega_A}(t_{max,\omega_A}). \tag{6}$$

Es ergibt sich  $t_{max,\omega_A} = (n-1)/a$  und  $h_{max,\omega_A} = 2a(n-1)^{n-1} e^{-(n-1)} / (n-1)!$  für die Fensterfunktionen nach Gl.(3) und  $n \geq 2$ . Der verwendete Ansatz der Korrekturfunktion enthält, als System betrachtet, ein reines Laufzeitglied und einen zeitkonstanten komplexen Übertragungsfaktor:

$$x_{\omega_A}(t) = \frac{e^{-j\omega_A t_{max,0}}}{h_{max,\omega_A}} \delta(t - (t_{max,0} - t_{max,\omega_A})). \tag{7}$$

Damit wird Gl.(5) zu:

$$g(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{h_{\omega_A}(t - (t_{max,0} - t_{max,\omega_A}))}{h_{max,\omega_A}} e^{-j\omega_A t_{max,0}} e^{j\omega_A t} d\omega_A. \tag{8}$$

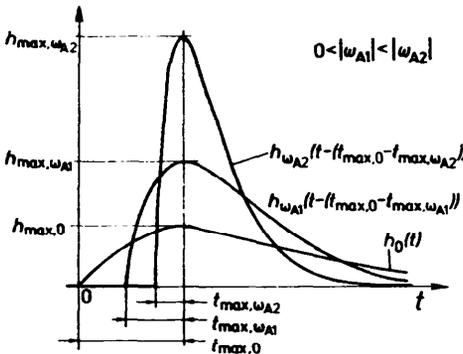


Fig.1: Auswirkung der Verzögerung im Korrektursystem  $x_{\omega_A}(t)$  auf die zeitliche Lage der Fensterfunktion  $h_{\omega_A}(t)$  für drei Analysefrequenzen  $\omega_A$ . Normalerweise beginnt die Fensterfunktion bei  $t = 0$ . Wegen des Verlaufs der Frequenzgruppenbreite würde das Maximum, ausgehend von einem Höchstwert  $t_{max,0}$  bei  $\omega_A = 0$ , mit steigendem  $\omega_A$  früher erreicht werden. Durch die Verzögerung  $t_{max,0} - t_{max,\omega_A}$  wird das Maximum unabhängig von  $\omega_A$  bei  $t_{max,0}$  erreicht.  $h_{\omega_A}(t)$  nach Gl.(3) für  $n = 2$ .

Die Auswirkung der zeitlichen Verzögerung der Fensterfunktion im Zähler von Gl.(8) veranschaulicht Fig.1. Die Verzögerung gewährleistet, daß die maximale Öffnung für alle  $\omega_A > 0$  über der bei  $\omega_A = 0$  zum Zeitpunkt  $t_{max,0}$  zu liegen kommt. Der Wert im Nenner normiert alle Fensterfunktionen auf Maximalwert eins. Der erste Exponentialfaktor bewirkt schließlich, daß formal insgesamt  $g(t) = \delta(t - t_{max,0})$  für den Zeitpunkt  $t = t_{max,0}$  gilt. Allerdings ist dies für andere Zeiten nicht mehr der Fall. Statt einer analytischen Lösung von Gl.(8) bietet sich eine numerische Auswertung im Frequenzbereich an. Durch Fouriertransformation und Umformung ergibt sich die Gesamtübertragungsfunktion:

$$G(\omega) = \frac{e^{-j\omega t_{max,0}}}{2\pi} \int_{-\infty}^{\infty} \frac{H_{\omega_A}(\omega - \omega_A)}{h_{max,\omega_A}} e^{j(\omega - \omega_A)t_{max,\omega_A}} d\omega_A. \tag{9}$$

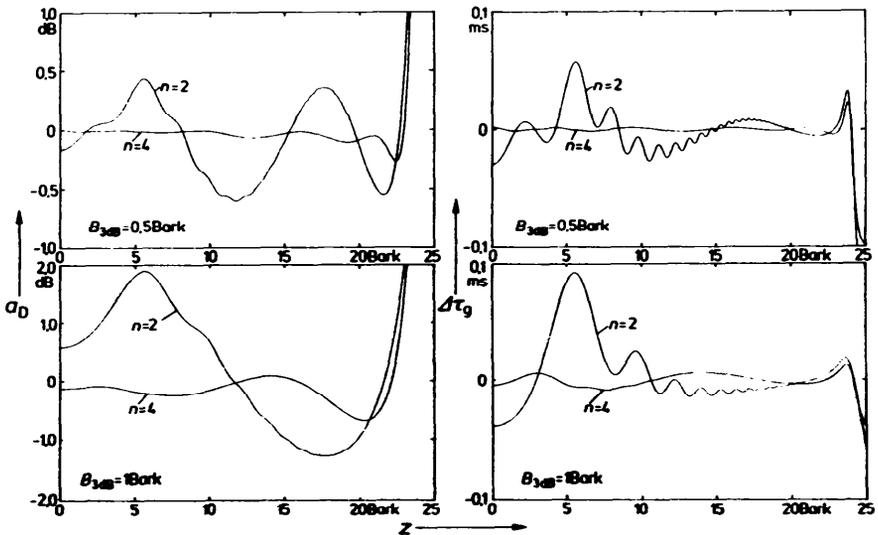


Fig.2: Frequenzverlauf von Dämpfung  $\alpha_D$  und Laufzeitschwankung  $\Delta\tau_g$  der Gesamtübertragungsfunktion  $G(\omega)$  aus FFT und RFFT, dargestellt über der Tonheit  $z$ . Fensterfunktionen nach Gl.(3),  $n = 2$  und 4. Fensterbandbreite oben, 0,5 Bark und 1 Bark unten. Betrag der Analysefrequenzen bis 24 Bark ( $\approx 15500$  Hz).

In der praktischen Anwendung ist der Bereich des Betrags der Analysefrequenzen auf den Hörbereich zu beschränken. Die Ergebnisse einer numerische Auswertung sind in Fig.2 als Dämpfung  $\alpha_D$  und Gruppenlaufzeitschwankung  $\Delta\tau_g = \tau_g - t_{max,\beta}$  von  $G(\omega)$  dargestellt. Allgemein läßt sich schließen, daß sowohl bei Vergrößern von  $n$ , wie auch bei Verkleinern von  $B_{3dB}$  der Schwankungsbereich von  $\alpha_D$  und  $\Delta\tau_g$  kleiner wird. Für  $B_{3dB} \leq 1$  Bark reicht  $n \geq 4$  und für  $B_{3dB} \leq 0,5$  Bark reicht  $n \geq 2$  aus, um die Abweichungen von einer idealen Übertragungsfunktion innerhalb des Analysefrequenzbereichs nicht wahrnehmbar werden zu lassen.

### Gehörgerechte Transformationskodierung

Wird das Integral in Gl.(4) durch eine Summe über endlich viele diskrete Analysefrequenzen im Abstand  $\Delta\omega_A \sim \Delta f_G$  ersetzt, so entsteht eine Filterbankstruktur. Die zeitdiskretisierte Ausführung dieser Struktur mit der Abtastrate  $1/T$  läßt sich zu einer gehörgerechten Transformationskodierung ausbauen. Fig.3 stellt einen Kanal hieraus dar. Die Impulsantwort von Analyse- und Synthesefilter in Reihe bestimmt die effektive Analysefensterfunktion  $h_{\omega_A}(n)$ . Die effektive Korrekturfunktion  $x_{\omega_A}(n)$  muß im zuvor behandelten Sinne hierauf abgestimmt sein. Sie ist die Impulsantwort der Reihenschaltung von Analyse- und Synthesekorrekturen, dividiert durch  $\Delta\omega_A$ . Die Berücksichtigung von  $\Delta\omega_A$  folgt aus der Diskretisierung von Gl.(4). Das Kanalsignal nach dem Analysetiefpaß belegt ein viel kleineres Frequenzband als das Eingangssignal. Deshalb kann die Abtastrate vor der Kodierung um den kanalabhängigen, ganzzahligen Faktor  $M_{\omega_A} < 1/(B_{3dB}^A T)$  erniedrigt werden. Nach der Kodierung muß die ursprüngliche Rate durch Einschalten von Nullwerten wieder erreicht werden. Der Synthesetiefpaß interpoliert die Zwischenwerte. Bei geeigneter Wahl der Tiefpaßarten und von  $M_{\omega_A}$  wird das durch Abtastratenveränderung und Kodierung hinzugekommene Kanalfehlensignal unter die Mithörschwelle des Kanalnutzsignals gedrückt. In diesem Punkt erscheint das Verfahren durch die Anpassung an die Frequenzgruppe als besonders wirkungsvoll.

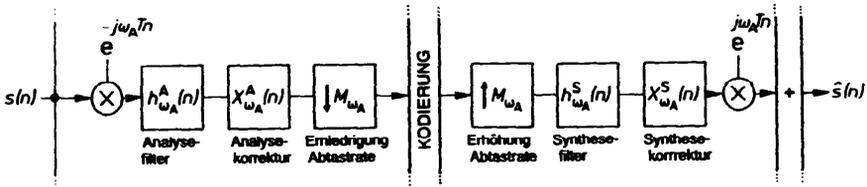


Fig. 3: Blockschaltbild eines Kanals der gehörgerechten Transformationscodierung für Signale mit der Abtastrate  $1/T$ .  $\omega_A$  ist die Kanalmittenfrequenz. Das komplette Schema erhält man durch Parallelschaltung solcher Kanäle. Kanalabstand und Kanalbandbreite sind proportional zur Frequenzgruppenbreite des Gehörs.

Das Gesamtsystem ohne einen Kodierungsteil wurde mit kanalweise identischen Analyse- und Synthesefiltern bei einer Abtastrate von  $1/T = 12.8$  kHz realisiert. Als Filtertyp kommt jeweils ein IIR-Besseltiefpass 8. Grades mit  $B_{3dB}^A = B_{3dB}^S = 0.7$  Bark zum Einsatz. Der Kanalabstand beträgt  $\Delta\omega_A = 2\pi B_{3dB}^A$ , die Kanäle decken annähernd den Frequenzbereich  $\pm 1/2T$  ab. Die Kanalabtastrate vor der Kodierung liegt ungefähr bei  $2B_{3dB}^A$ . Die Summenabtastrate der vor der Kodierung anliegenden Spektralwerte liegt somit um den Faktor zwei über der Abtastrate des Eingangssignals. Hierin sind Real- und Imaginärteile separat gezählt und spektrale Symmetrien berücksichtigt. Erste Hörversuche mit Sprache zeigten keine wahrnehmbaren Veränderungen im Synthesesignal.

Es ist wichtig, daß die Laufzeit aus der zeitkontinuierlichen Korrekturfunktion genauer als ein Abtastwert in der diskreten Realisierung aufgelöst wird, sonst kann es Frequenzgangverfälschungen geben. Deshalb werden diskrete Allpassglieder vom Grad 1 zur Feinjustierung eingesetzt.

### Zusammenfassung

Zur Fourier-t-Transformation (FTT), einer gehörgerechten Kurzzeitspektralanalyse mit Frequenzgruppenanpassung des Analysefensters, kann eine Rücktransformation (RFTT) angegeben werden. Darin wird das Spektrum mit einer Korrekturfunktion modifiziert und das Fourier-Rücktransformationsintegral angewandt. Durch die Korrekturfunktion erreichen die Analysefenster an allen Frequenzen zum selben Zeitpunkt eine gleichhohe Maximalöffnung und es rückt der Syntheszeitpunkt des Integrals auf diesen Zeitpunkt zurück. Ein vorhandener Restfehler wird bei geeigneter Wahl der Fensterfunktion unhörbar. Aus der Diskretisierung von FTT und RFTT in Zeit und Frequenz ergibt sich ein Rahmen für eine gehörgerechte Transformationskodierung. Die durch Unterabtastung praktisch erreichte Summenabtastrate des Spektrums vor der Kodierung lag um den Faktor 2 höher als die Abtastrate des Zeitsignals. Die Frequenzgruppenanpassung des Gesamtsystems verspricht eine effiziente Absenkung von Kodierungsfehlern unter die Mithörschwelle.

### Literatur

- [1] Terhardt, E.: "Fourier transformation of time signals, conceptual revision", *Acoustica* 57 (1985) S.242 - 256.
- [2] Owens, F.J.; Murphy, M.S.: "Non-uniform RFT filterbank design for speech processing", *EURO-SPEECH '89*, Paris (1989), Bd. 1, S.605-608.
- [3] Krahe, D.: "Was genügt dem Gehör? - Datenreduktion bei digitalen Audiosignalen", *Fortschritte der Akustik, DAGA '89*, Bad Honnef (1989), S.141-155.
- [4] Schlang, M.; Mummert, M.: "Die Bedeutung der Fensterfunktion für die Fourier-t-Transformation als gehörgerechte Spektralanalyse", *Fortschr. d. Ak., DAGA '90*, Bad Honnef (1990), S.1043-1046.
- [5] Zwicker, E.; Terhardt, E.: "Analytical expressions for critical bandwidth and critical bandwidth as a function of frequency", *J. Acoust. Soc. Amer.* 68 (1980) S.1523-1525.