

UNTERSUCHUNG EINER GEHÖRBEZOGENEN SPEKTRALANALYSE MITTELS RESYNTHESE
 W. Heinbach
 Lehrstuhl für Elektroakustik der Technischen Universität München

1. Einleitung

Bei den meisten Modellen der Hörempfindungen (z.B. Tonhöhe, Lautheit) ist eine Spektralanalyse der Zeitfunktion notwendig. In der Regel wird dabei eine Datenreduktion vorgenommen, in dem z.B. nur das Betragsspektrum verwendet wird. Dabei sollte jedoch sichergestellt sein, daß die für das Gehör wesentliche Information nicht verloren geht. Dies kann mit Hilfe der Resynthese überprüft werden. Dazu wird ein akustisches Zeitsignal $p(t)$ analysiert und mit den Analysedaten ein neues Zeitsignal $q(t)$ erzeugt. Die beiden Signale $p(t)$ und $q(t)$ werden durch Anhören verglichen. Der Einfluß von Analyseparametern zeigt sich in einer mehr oder weniger guten Übereinstimmung der Hörwahrnehmungen.

In diesem Beitrag wird über Untersuchungen des zeitlich geglätteten und abgetasteten Betragsspektrums der sogenannten Fourier-t-Transformation (FTT) /1,2/ berichtet.

2. Spektralanalyse

Die Fourier-t-Transformation

$$\underline{p}(\omega, t) = \int_0^t p(x) e^{-a(t-x)} e^{-j\omega x} dx \quad ; t > 0 \quad (1)$$

erscheint zur Durchführung einer gehörbezogenen Spektralanalyse besonders geeignet /2,3/. Mit der Transformationskonstanten a wird die effektive Länge des Analysefensters bestimmt und damit die Bandbreite $B=a/\pi$. Die wesentlichen Vorteile dieser Transformation gegenüber anderen Verfahren (z.B. FFT) sind:

- Quasi-kontinuierliche Funktion der Zeit;
- BT-Produkt (Bandbreite mal eff. Analysefensterlänge) ≈ 0.3 ;
- keine Beschränkung der Stützwerte auf ganzzahlige Verhältnisse;
- Analysefrequenzen und -bandbreiten frei wählbar.

Wird a frequenzabhängig gemacht, so empfiehlt sich die Normierung auf $1/a$:

$$\underline{p}(\omega, t) = 2a \int_0^t p(x) e^{-a(t-x)} e^{-j\omega x} dx \quad ; t > 0 \quad (2)$$

Die Parameter der Transformation (2) werden entsprechend der Frequenzgruppenbreite des Gehörs nach /4/ eingestellt:

$$1 \text{ Bark} \approx \Delta f_G = 25 + 75(1 + 1.4(f/\text{kHz})^2)^{0.69} \quad (3)$$

Die Abstände $\Delta\omega$ zwischen den Analysefrequenzen ω sind dann abhängig von der Frequenz; ebenso die Bandbreiten B .

Das Betragsquadratspektrum erhält man aus (2) mit:

$$F(\omega, t) = |\underline{p}(\omega, t)|^2 \quad (4)$$

Führt man eine Resynthese mit dem gesamten Betragsspektrum durch, so erhält man:

$$q(t) = \sum_{i=0}^{\infty} \sqrt{F(\omega_i, t)} \cdot \cos(\omega_i t) \quad (5)$$

Das entstehende Zeitsignal $q(t)$ weist eine sehr hohe Rauigkeit auf.

Den Grund für diese Rauigkeit zeigt folgende Betrachtung:
Analysiert man eine Zeitfunktion

$$p(t) = A_s \cos(\omega_s t + \varphi_s) \quad , \quad (6)$$

so ergibt sich mit (2) und (4) das Betragsquadratspektrum:

$$F(\omega, t) = \frac{A_s^2}{1 + \left[\frac{\omega - \omega_s}{a} \right]^2} \quad (7)$$

Lässt man zur Vereinfachung die Frequenz ω_s in (6) mit einer beliebigen Analyse- bzw. Synthesefrequenz ω_1 in (5) zusammenfallen, so erhält man mit (5):

$$q(t) = A_s \cos(\omega_s t) + \sum_{j=0}^{i-1} \frac{A_s \cos(\omega_j t)}{\sqrt{1 + \left[\frac{\omega_j - \omega_s}{a} \right]^2}} + \sum_{j=i+1}^{\infty} \frac{A_s \cos(\omega_j t)}{\sqrt{1 + \left[\frac{\omega_j - \omega_s}{a} \right]^2}} \quad . \quad (8)$$

Bei Verwendung des gesamten Betragspektrums erhält man außer einem Anteil bei ω_s auch Anteile benachbarter Frequenzen, deren Amplitude mit wachsendem Abstand zu ω_s abnimmt. Deren Interferenzen erzeugen den Rauigkeitseindruck.

Der erste Ausdruck in (8) stellt das Amplitudenmaximum über der Frequenz dar. Eine Resynthese mit diesem Maximum ergibt eine Zeitfunktion, deren Amplitude und Periode mit (6) übereinstimmt, nicht jedoch mit deren Phase. Es wird daher weiter untersucht, welche Informationen die relativen Maxima des Betragspektrums über der Frequenz enthalten. Ein relatives Maximum liegt dann vor, wenn das Vorzeichen von $dF(\omega, t)/d\omega$ nach negativen Werten wechselt und eine Schwelle ΔL_{\max} überschritten wird. Die Schwelle ΔL_{\max} wird dabei von den umliegenden Minima aus gerechnet. Wendet man dieses Verfahren auf das Betragspektrum einer Zeitfunktion wie (6) an, so zeigt dieses bei Einschaltvorgängen (z.B. $t \approx 0$) Nebenmaxima, die mit zunehmender Zeit verschwinden. Um eine falsche Frequenzzuordnung durch solche Nebenmaxima zu verhindern, wird eine zeitliche Glättung des Leistungsspektrums durchgeführt:

$$G(\omega, t) = \frac{1}{T_G} \int_0^t F(\omega, x) e^{-(t-x)/T_G} dx \quad ; t > 0 \quad . \quad (9)$$

Der Wert der Glättungszeitkonstanten T_G hängt von der Bandbreite und damit von der Transformationskonstanten a ab; zur Unterdrückung von Nebenmaxima muß ein Mindestwert von $T_G = 0.3/a$ verwendet werden /3/.

Das Spektrum nach (9) kann sich wegen der Glättung und der endlichen Bandbreiten nicht beliebig schnell ändern. Es bietet sich also an, dieses nur zu bestimmten, äquidistanten Zeitpunkten $t = kT_A$ auszuwerten. Die Auswertung kann z.B. in der Bestimmung der Maxima bestehen.

Das Abtastintervall T_A hängt von der Glättungszeitkonstanten T_G ab. Betrachtet man die Glättung nach (9) als Tiefpass erster Ordnung, so hat dieser eine Grenzfrequenz von $f_{T_G} = 1/(2\pi T_G)$; als obere Grenze für T_A ergibt sich dann $T_A = \pi T_G$. Tastet man das gesamte Spektrum zu gleichen Zeitpunkten ab, so richtet sich T_A nach der Glättungszeitkonstanten der höchsten Analysefrequenz. Die Analyse liefert somit in Zeitabständen T_A eine Anzahl von m Maximumparametern:

$$M(kT_A) = \{\omega_1, A_1; \omega_2, A_2; \dots; \omega_m, A_m\} \quad . \quad (10)$$

3. Synthese

Die Synthese erfolgt in Zeitabschnitten der Länge T_A , dem Abtastintervall des geglätteten Betragsspektrums. Jedem Maximum in (10) wird eine Sinusschwingung mit dessen Frequenz und Amplitude zugeordnet. Die Startphase wird aus der vorhergehenden Synthese entnommen, wenn dort eine Frequenz innerhalb eines bestimmten Bereiches v der aktuellen Frequenz vorhanden war. Ansonsten wird sie zu Null gesetzt. Dadurch werden Phasensprünge der Zeitfunktion von einem Abschnitt zum nächsten weitgehend vermieden.

$$q_1(t) = A_1 \sin(\omega_1 t + \varphi_1(kT_A)) \quad ; t > kT_A \quad (11)$$

$$\text{mit } \varphi_1(kT_A) = \varphi_1((k-1)T_A) + \omega_1 T_A \quad \text{wenn } \omega_1 \pm v \in M((k-1)T_A) \quad (12)$$

$$\text{sonst } \varphi_1(kT_A) = 0 \quad .$$

Die einzelnen Sinusschwingungen werden addiert und ergeben die Zeitfunktion im Syntheseabschnitt. Bis auf die Übernahme der Phasen erfolgt keine Interpolation zwischen den einzelnen Abschnitten. Frequenz und Amplitude einer Einzelschwingung bleiben innerhalb eines Abschnittes konstant.

$$q(t) = \sum_{i=1}^m q_i(t) \quad ; t > kT_A \quad (13)$$

4. Ergebnisse

Die Resynthese mit den Parametern ω_m , A_m der Maxima des zeitlich geglätteten Betragsspektrums der FTT zeigt, daß diese die für das Gehör wesentliche Information noch enthalten.

Dies bezieht sich auf alle Schalle, die zur Untersuchung verwendet worden sind, nämlich **Sprache**, **Glockenklänge**, **Simustöne** mit **AM** und **FM** sowie **Rauschen**. Die Qualität des resynthetisierten Schalles hängt von den verwendeten Analyseparametern ab. Die besten Ergebnisse erzielt man mit:

- einem Abstand der Analysefrequenzen entsprechend 0.05 Bark
- Analysebandbreiten zu 0.1 Bark
- einer Glättungszeitkonstanten zu 0.35/a.

Durch eine Frequenzinterpolation wird der Abstand der Analysefrequenzen mit 0.05 Bark sowohl der Frequenzunterschiedsschwelle als auch der Klangzerlegung durch das Gehör gerecht. Ab 0.1 Bark Analysefrequenzabstand wird die Klangzerlegung zunehmend schlechter.

Kleinere Werte der Bandbreite als 0.1 Bark wirken sich auf das Zeitauflösungsvermögen durch den Anstieg der effektiven Zeitfensterlänge aus, größere Werte als 0.15 Bark vermindern das Frequenzauflösungsvermögen.

Die Glättung sollte der Analysebandbreite angepaßt sein, da sonst die Vorteile der frequenzabhängigen Bandbreite, wie schnelles Einschwingen bei hohen Frequenzen verlorengehen. Bei Glättungszeitkonstanten unter 0.3/a treten störende Nebenmaxima auf, bei Werten über 1/a verschlechtert sich das Zeitauflösungsvermögen; das Signal weist dann eine zunehmende Halligkeit auf.

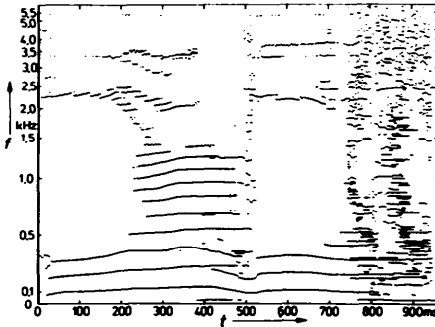


Fig. 1: Verlauf der Maximalfrequenzen über der Zeit. Abtastung: $T_A=1.25\text{ms}$; Glättungszeit: $0.35/a$

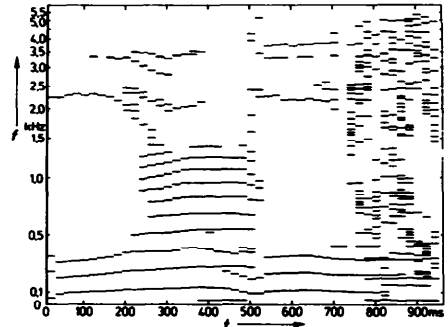


Fig. 2: Verlauf der Maximalfrequenzen über der Zeit. Abtastung: $T_A=20\text{ms}$; Glättungszeit: $0.5/a$

Gemeinsame Daten: Analysefrequenzabstand: $\Delta\omega\approx 0.05$ Bark; Bandbreite: 0.1 Bark. Analysierter Schall: "Er geht".

Den Verlauf der Frequenzen der Maxima (10) des abgetasteten, geglätteten Betragsspektrums über der Zeit zeigen Fig. 1 und Fig. 2. Die Frequenzachse ist entsprechend der Tonheit aufgetragen. Die Darstellung ist ähnlich einem Sonagramm, allerdings fehlt hier die Information über den Pegel.

Analysiert wurde das Sprachsignal "Er geht". Die Grundfrequenz bei etwa 100Hz und die Formanten des ersten "e" (bis 150ms) sind deutlich zu erkennen. Der Einfluß der frequenzabhängigen Glättung ist am Beginn des "t" bei 750ms zu erkennen; die Maxima bei tiefen Frequenzen treten in der Regel erst später auf.

Eine Datenreduktion erzielt man durch Abtasten des Spektrums nach (9) in größeren Zeitabständen T_A , wie in Fig. 2 für $T_A=20\text{ms}$ dargestellt. Hier machen sich die zum Teil großen Änderungen von einem zum nächsten Abschnitt durch Störungen bemerkbar. Trotzdem bleibt die Sprache gut verständlich, auch die Klangfarbe bleibt weitgehend erhalten.

5. Zusammenfassung

Mit Hilfe der Resynthese kann aus den relativen Maxima des geglätteten, abgetasteten Betragsspektrums der Fourier-t-Transformation eine Zeitfunktion erzeugt werden, deren Höreindruck weitgehend dem Originalsignal entspricht. Einen wesentlichen Anteil daran hat die Auswahl der Transformationsparameter Analysefrequenz, Bandbreite und Glättungszeit entsprechend der Frequenzgruppenbreite des Gehörs. Eine Datenreduktion lässt sich durch zeitliches Abtasten des Betragsspektrums erzielen. Die Ergebnisse deuten darauf hin, daß das Verfahren für alle Arten von Schallen geeignet ist.

6. Literatur

- /1/ Terhardt, E.: Fourier transformation of time signals, conceptual revision. *Acustica*, 57, 242-256 (1985)
- /2/ Terhardt, E.: Verfahren zur gehörbezogenen Frequenzanalyse. In: Fortschritte der Akustik - DAGA '85, S. 811, Bad Honnef: DPG-Verlag '85
- /3/ Feldtkeller, M.: Fourier-t-Transformation als gehörbezogene Spektralanalyse. Diplomarbeit am Lehrstuhl für Elektroakustik, TU München 1985
- /4/ Zwicker, E.: Psychoakustik. Springer, Heidelberg (1982), S. 67

Die Untersuchungen wurden im Sonderforschungsbereich 204 "Gehör", München, gefördert durch die Deutsche Forschungsgemeinschaft, durchgeführt.