

VERFAHREN ZUR GEHÖRBEZOGENEN FREQUENZANALYSE

E. Terhardt

Lehrstuhl für Elektroakustik der Technischen Universität München

1. Einleitung

Die Wahrnehmung von Klängen ist unter anderem durch das Auftreten von Spektraltonhöhen gekennzeichnet, welche den Frequenzen einzelner sinusförmiger Klangkomponenten entsprechen. In diesem Sinne ist ein Klangsinal $p(t)$, welches aus N Komponenten besteht, durch

$$p(t) = \sum_{n=1}^N \hat{p}_n(t) \cos[\omega_n(t) + \varphi_n(t)] \quad (1)$$

zu beschreiben. Die Aufgabe besteht darin, die im allgemeinen zeitvariablen Frequenzen $f_n = \omega_n / (2\pi)$ als Funktion der Zeit zu bestimmen. Dies kann unter anderem dadurch bewerkstelligt werden, daß man die Lage der Maxima eines zeitvariablen Betragspektrums auf der Frequenzachse, welche von den einzelnen Komponenten erzeugt werden, ermittelt. An das zeitvariable Spektrum sind bei Anwendung dieser Methode folgende beide Forderungen zu stellen.

(1) Die effektive Länge des Analyseintervalls T sollte möglichst gering sein, damit zeitliche Frequenzänderungen korrekt wiedergegeben werden können. Zugleich sollte die Analysebandbreite B , das heißt die Breite der Gipfel der Betrags-Spektralfunktion, ebenfalls gering sein, damit die Frequenzbestimmung möglichst genau wird. Das Produkt BT hängt von der Gewichtung des Signals innerhalb des Analyseintervalls ab ("Zeitfenster"). Zu fordern ist also eine Gewichtsfunktion mit möglichst kleinem BT -Produkt.

(2) Die zeitvariable Spektralfunktion $P(f, t)$ muß als quasi-kontinuierliche Funktion der Zeit zur Verfügung stehen. Das heißt, die Zeitabstände T_y zwischen aufeinanderfolgend berechneten Spektren sollen so klein sein, daß zeitliche Änderungen der Spektralparameter beobachtet werden können. Dadurch ergibt sich zwischen T_y und T ein Zusammenhang der Art $T_y = kT$, wobei k davon abhängt, wie groß die Parameteränderungen sein sollen, um als signifikant zu gelten; sinnvoll erscheint der Bereich $2 \leq k \leq 10$. Beispielsweise nimmt das Gehör Rauigkeit bis zu Schwankungsfrequenzen von etwa 300 Hz wahr. Um solche Schwankungen der Spektralparameter noch beobachten zu können, muß $T_y < 1,5$ ms gewählt werden, und dies ist andererseits nur sinnvoll, wenn das Analyseintervall T kürzer als etwa 15 ms ist.

Die gegenwärtig Überwiegend gebräuchliche Spektraltransformation ist durch ein Analyseintervall T endlicher Länge, Berechnung der Spektralwerte im Frequenzabstand $1/T$ und eine innerhalb des Analyseintervalls symmetrische Signalgewichtung gekennzeichnet; zur Berechnung wird die Methode der schnellen Fouriertransformation (FFT) benutzt. Das BT -Produkt hat je nach Gewichtsfunktion Werte zwischen etwa 0,8 und 2. Der Vorteil kurzer Rechenzeit des FFT-Verfahrens kommt nur zum tragen, wenn die zeitliche Überlappung aufeinanderfolgender Analyseintervalle gering, wenn also $T_y \approx T$ ist. Für die oben erläuterte quasi-kontinuierliche Spektralberechnung hat das FFT-Verfahren keine Vorteile; eine fortlaufende, rekursive Berechnung des Spektrums erscheint eher angemessen und erfolgversprechend. Weiterhin zeigt sich, daß das üblicherweise erreichte BT -Produkt zu groß ist, um sowohl dem Zeit- als auch dem Frequenzauflösungsvermögen des Gehörs in vollem Umfang gerecht zu werden.

2. Das Fourier-t-Spektrum

Die schon seit längerem bekannte (vgl. /1/) Spektraltransformation

$$P(f, t) = \int_{-\infty}^t p(x) e^{-a(t-x)} e^{-j\omega x} dx \quad (\text{mit } \omega = 2\pi f) \quad (2)$$

erscheint zur Erfüllung der genannten Forderungen eher geeignet; sie wird wegen der zeitvariablen Obergrenze des Fourierintegrals als Fourier-t-Transformation (FTT) bezeichnet /2/. Die Signalgewichtung mit $e^{-a(t-x)}$ ist einseitig und das Analyseintervall endet abrupt im variablen Zeitpunkt t . Das Analyseintervall ist im Prinzip unendlich lang; seine effektive Länge ist jedoch endlich und kann durch die Zeitkonstante $1/a$ gekennzeichnet werden. Die Analysebandbreite erweist sich als $B = a/\pi$, so daß $BT = 1/\pi$ bzw. rund 0,3 beträgt /2/. Aus den aufeinanderfolgenden Abtastwerten des Signals $p(mT_x)$ ($T_x =$ Abtastintervall) kann die komplexe Spektralfunktion in Zeitabständen des Abtastintervalls berechnet werden nach der Formel

$$P(f, m+1) = P(f, m) e^{-aT_x} + T_x p[T_x(m+1)] e^{-j\omega(m+1)T_x} \quad (3)$$

Die Analysebandbreite kann auf einfache Weise der Frequenzselektivität des Gehörs angepaßt werden, indem a frequenzabhängig gemacht wird /2/. Der optimale Frequenzabstand der mit Gl.(3) berechneten Spektralwerte liegt in der Größenordnung der eben wahrnehmbaren Frequenzunterschiede, das heißt, bei etwa 1/25 bis 1/50 der Frequenzgruppenbreite. Da der gesamte hörbare Frequenzbereich 24 Frequenzgruppen ausmacht, genügen zu seiner vollständigen gehöradäquaten Spektralzerlegung 600 bis 1200 Frequenzstützwerte.

Als Beispiel zeigt Fig.1 das FTT-Betragspektrum eines Zweiklangs, berechnet mit einer Analysebandbreite von 1/10 Frequenzgruppe, und zwar in den beiden Zeitpunkten, wo die Lücke zwischen den beiden Maxima am stärksten (0 ms) bzw. am wenigsten (10 ms) ausgeprägt ist. Man kann daraus entnehmen, daß in der Tat einerseits die beiden Klangkomponenten hinreichend ausgeprägte Maxima hervorrufen, andererseits die Zeitstruktur der Schwebung adäquat wiedergegeben wird. Bei dem in Fig.1 gewählten Frequenzabstand von 50 Hz liegt die Grenze der Klangzerlegung durch das Gehör. Das heißt, daß bei größeren Frequenzabständen zwei Spektraltonhöhen statt einer einzigen gehört werden. Man kann abschätzen, daß diesem Verhalten durch das FTT-Betragspektrum mit einer Analysebandbreite von etwa der doppelten Frequenzunterschiedsschwelle Rechnung getragen wird, also einer Bandbreite, die etwa halb so groß ist wie die in Fig.1 gewählte.

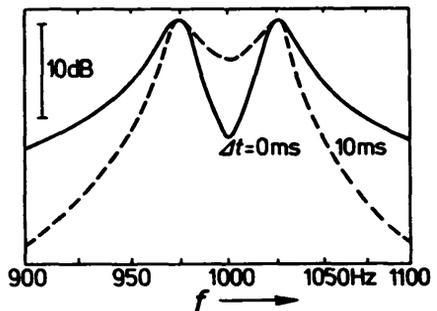


Fig.1. FTT-Betragspektren (logarithmisch) eines stationären Zweiklangs aus Sinustönen (975+1025 Hz) in den Zeitpunkten 0 und 10 ms relativ zur Schwebungsperiode.

3. Frequenzanalyse

Die Genauigkeit der Frequenzbestimmung durch das Gehör wird sehr einfach durch die Frequenzunterschiedsschwelle einzelner Sinustöne als Funktion von deren Dauer gekennzeichnet. Fig.2 gibt neuere Daten von Fastl & Hesse /3/ wieder, und zwar in normierter Darstellung (Abszisse: Testtondauer T_T mal Testtonfrequenz f_T , d.h. Anzahl der Tonperioden; Ordinate: Eben wahrnehmbarer Frequenzunterschied Δf bezogen auf Testtonfrequenz). Die bei den Testtonfrequenzen 125 Hz, 1000 Hz und 4000 Hz gefundenen Meßpunkte sind durch Kurven (gestrichelt bzw. punktiert) angenähert, um ihre Charakteristika deutlich zu machen. Bei Tondauern über etwa 200ms ist Δf im wesentlichen konstant. Bei kurzen Tondauern, gemessen in der Anzahl von Perioden, konvergieren alle Meßwerte offenbar gegen ein- und dieselbe Gerade mit der Steigung -1 . Darin zeigt sich das Verhalten eines Frequenzanalysators, welcher nach dem "Maximumdetectionsprinzip" arbeitet, unabhängig davon, welche Signalgewichtung im Analyseintervall stattfindet.

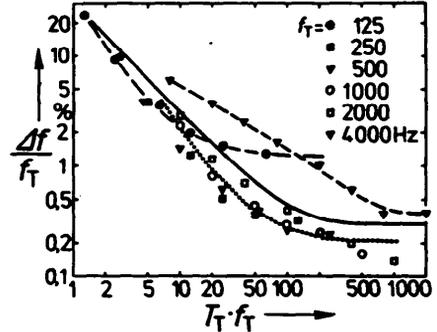
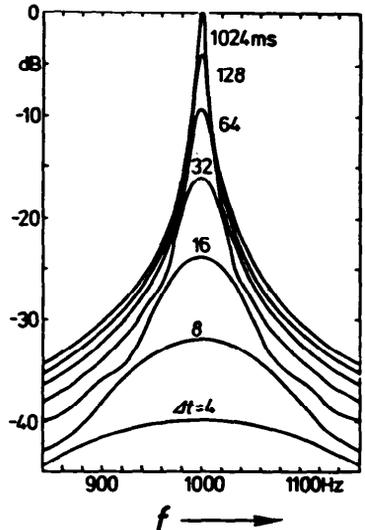


Fig.2. Eben wahrnehmbare Frequenzunterschiede Δf von Sinustönen, bezogen auf die Tonfrequenz f_T , als Funktion der Tondauer T_T , welche auf die Tonperiode normiert ist, d.h., die Anzahl dargebotener Tonperioden (Abszisse); nach /3/. Gestrichelt: Interpolierte Darstellung der Daten bei 125 und 4000 Hz. Punktiert: ebenso für 1000 Hz. Durchgezogene Kurve: Resultat aus FFT-Analyse mit 0,6% Analysebandbreite.

Das FFT-Betragspektrum eines Sinussignals hat im stationären Fall (d.h., lange nach dem Einschalten) keine Nebenmaxima, so daß die Maximumdetektion eindeutig ist. Unmittelbar nach dem Einschalten des Signals treten jedoch rasch veränderliche Nebenmaxima auf /2/. Um auch sehr kurz dauernden Klangkomponenten Frequenzen zuordnen zu können, müssen die Nebenmaxima vom Hauptmaximum unterschieden und ignoriert werden. Dies kann am einfachsten durch zeitliche Glättung des Betragspektrums erreicht werden. Fig.3 zeigt das mit einer Zeitkonstante von 100ms geglättete FFT-Betragspektrum eines 1000Hz-Tones in verschiedenen Zeitpunkten nach

Fig.3. Geglättete FFT-Betragspektren (logarithmisch) eines 1000 Hz-Tones, welcher bei $\Delta t=0$ eingeschaltet wurde. Die Glättung erfolgte mit einem Tiefpaß 1. Ordnung, angewendet auf das Leistungsspektrum. $B=6$ Hz.



dessen Einschalten. Die Analysebandbreite ist $B=6$ Hz, woraus sich als Länge des Analyseintervalls $1/(6\pi) s = 50$ ms ergibt. Man erkennt, daß bereits unmittelbar nach dem Einschalten eine eindeutige, wenn auch zunächst ungenaue, Maximumbestimmung möglich ist. Legt man eine Amplitudenauflösung des Detektors von 1dB zugrunde (Dies ist ein Wert, der für zahlreiche Detektionsleistungen des Gehörs typisch ist), dann kann man die zur Erkennung nötige Frequenzverschiebung eines 1000 Hz-Tones als Funktion von dessen Dauer aus Fig.3 ablesen. Der Verlauf der solcherart definierten Frequenzunterschiedsschwelle ist in Fig.2 als durchgezogene Kurve eingetragen. Wegen der normierten Darstellung gilt sie für alle Tonfrequenzen, sofern die relative Analysebandbreite von 0,6% konstant gehalten wird. Ihre Übereinstimmung mit den Meßwerten des Frequenzbereichs 250 Hz bis 2000 Hz kann im Hinblick auf die großen individuellen Unterschiede und die ausgeprägte Abhängigkeit von der psychoakustischen Meßmethode (vgl. /3/) bereits als recht befriedigend bezeichnet werden und ist im Einklang mit der bekannten Vorstellung, daß in diesem Bereich und im stationären Fall im wesentlichen die relative Frequenzunterschiedsschwelle frequenzunabhängig ist. Brauchbare Übereinstimmung mit den Meßwerten bei 125 Hz kann einfach durch Vergrößerung der relativen Analysebandbreite um etwa einen Faktor 3 bis 5 erzielt werden, denn dadurch wird bei konstantem Detektionskriterium (1 dB Pegeldifferenz) die Kurve um denselben Faktor angehoben und nach links verschoben. Eine wesentliche Verbesserung der Übereinstimmung der theoretischen Kurve mit den Meßwerten bei 4kHz ist auf diese Weise nicht möglich, weil deutliche Diskrepanzen nur im Bereich mittlerer Tondauern (10 bis 500 Perioden) vorhanden sind. Dieses Ergebnis deutet darauf hin, daß möglicherweise bei hohen Tonfrequenzen bisher unbekannte sekundäre Effekte Einfluß auf die Frequenzunterschiedsschwelle haben.

4. Zusammenfassung

Von einem gehörbezogenen Spektralanalysator ist zu fordern, daß er in relativ kleinen Zeitabständen Spektralfunktionen liefert, welche sowohl hinsichtlich Frequenz- als auch Zeitauflösung als Grundlage entsprechender Hörempfindungen dienen können. Die gegenwärtig verbreiteten, mit der Schnellen Fouriertransformation arbeitenden Verfahren erfüllen diese Forderung nur in stark eingeschränktem Maße. Demgegenüber scheint es, daß die geschilderte Fourier-t-Transformation geeignet ist, die Forderung zu erfüllen. Dies wird insbesondere am Beispiel der eben wahrnehmbaren Frequenzunterschiede von Sinustönen gezeigt. Die psychoakustisch gefundenen Frequenzunterschiedsschwellen erweisen sich als weitgehend in Einklang mit dem Prinzip der Frequenzanalyse nach der Maximumdetektionsmethode.

Literaturangaben

- /1/ Flanagan, J.L.: Speech analysis, synthesis, and perception. Springer, Heidelberg (2.Aufl., 1972), S. 141-150.
- /2/ Terhardt, E.: Fourier transformation of time signals, conceptual revision. *Acustica* (1985), im Druck.
- /3/ Fastl, H. und Hesse, A.: Frequency discrimination for pure tones at short durations. *Acustica* 56, 41-47 (1984).

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