

Berechnung von Erregungsverteilungen aus FTT-Spektren

Claus von Rücker

Fachgebiet Akustische Kommunikation, Technische Universität München

1. Erregung

Die Erregungsverteilung ist die über dem Ort der Basilarmembran aufgetragene Aktivität, die ein vom Gehör aufgenommener Schall erzeugt. Sie dient als Hilfsgröße zur Erklärung und Berechnung verschiedener psychoakustischer Empfindungsgrößen, wie beispielsweise der Lautheit oder der Tonhöhe, und kann als direkte Folge der Frequenzselektivität der Cochlea verstanden werden.

Im Bereich der Kernerregung ist die psychoakustische Erregung im wesentlichen gleich der physikalischen Anregung [9]. Außerhalb dieses Bereiches werden Flankenerregungen angenommen, die durch Verschiebung der Mithörschwellenmuster schmalbandiger Schalle entstehen.

Soll die Erregungsverteilung eines unbekanntes Schalls bestimmt werden, muß zunächst eine Spektralanalyse vorgenommen werden. Bei der Berechnung der Erregungsverteilung aus Spektren ist zu berücksichtigen, daß sich die Selektivität des gewählten Analyseverfahrens bereits im Spektrum niederschlägt.

2. Näherung für die Simultanhörschwelle von Sinustönen

Häufig wird für die Berechnung von Erregungsverteilungen eine von Terhardt [5] vorgeschlagene Näherung für die Mithörschwelle von Schmalbandrauschen verwendet, die aus zwei Geradenstücken besteht, wenn die Frequenzskala Bark-skaliert wird. Die Steigung der unteren Flanke wird dabei mit konstant 27 dB/Bark angenähert, die obere Flanke ist stark pegel- und leicht frequenzabhängig und hat für mittlere Pegel etwa die Steigung -16 dB/Bark. Ein Vergleich mit experimentellen Daten aus Mithörschwellenmessungen verschiedener Autoren zeigt, daß diese Näherung nur eine grobe Beschreibung der Daten erlaubt. Insbesondere wird der Verlauf der Mithörschwelle etwa 1 Bark über der Maskierfrequenz als zu hoch angegeben und die Verflachung der unteren Flanke bei kleinen Maskierpegeln wird nicht berücksichtigt. Deshalb wird im folgenden eine genauere Näherung vorgeschlagen.

Zwicker und Jaroszewski [8] haben die Mithörschwelle von Sinustönen für verschiedene Pegel auch an der unteren Flanke genau gemessen. Es zeigt sich, daß bei sinkendem Maskierpegel die untere Flanke flacher wird und geknickt verläuft. Deshalb wird hier eine Näherung durch zwei Geradenstücke mit den Steigungen S_{11} und S_{12} vorgeschlagen, in der auch die Meßdaten von Schöne [3] und Sonntag [4] berücksichtigt wurden:

$$S_{11} = \left[1,2 + 0,5 \cdot \frac{L_M}{\text{dB}} \right] \frac{\text{dB}}{\text{Bark}}; \quad -1 \leq \frac{z - z_M}{\text{Bark}} \leq 0, \quad (1)$$

$$S_{12} = \left[0,25 \cdot \frac{L_M}{\text{dB}} \right] \frac{\text{dB}}{\text{Bark}}; \quad \frac{z - z_M}{\text{Bark}} < -1, \quad (2)$$

wobei z_M die Tonheit und L_M den Pegel des maskierenden Sinustons bezeichnet.

Nach einem Vorschlag von Valenzuela [7] wird die obere Flanke des Mithörschwellen-Tonheitsmusters durch zwei Geradenstücke angenähert, deren frequenzabhängige Steigungen vor allem aus den Daten von Sonntag [4] ermittelt wurden. Damit wird sowohl die nichtlineare Auffächerung der oberen Flanke mit steigendem Pegel als auch deren Frequenzabhängigkeit genauer beschrieben. Für die Steigungen S_2 und S_3 ergibt sich:

$$S_2 = \left[-18,9 - \frac{6 \text{ kHz}}{f_M} + 0,19 \cdot \frac{L_M}{\text{dB}} \right] \frac{\text{dB}}{\text{Bark}}; \quad 0 < \frac{z - z_M}{\text{Bark}} \leq 1, \quad (3)$$

$$S_3 = \left[-14 - 0,33 \cdot \frac{f_M}{\text{kHz}} + 0,13 \cdot \frac{L_M}{\text{dB}} \right] \frac{\text{dB}}{\text{Bark}}; \quad \frac{z - z_M}{\text{Bark}} > 1. \quad (4)$$

Abb. 1 zeigt die mit Hilfe der oben angegebenen Steigungen ermittelten Erregungspegel-Tonheitsmuster für einen 1 kHz (8,5 Bark) Sinuston bei vier verschiedenen Pegeln. Die Geradenstücke werden dazu derart an den Wert der Kernerregung (physikalische Anregung) angesetzt, daß sie ohne Sprung anschließen. Das Übertragungsmaß a_0 von Außen- und Mittelohr wird hier aus Gründen der Übersichtlichkeit vernachlässigt, könnte aber an dieser Stelle berücksichtigt werden.

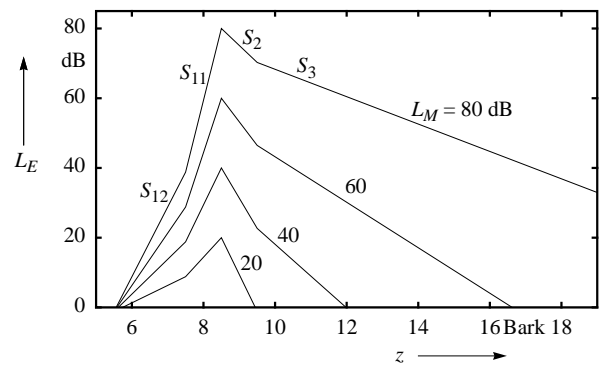


Abb. 1: Erregungspegel-Tonheitsmuster eines Sinustones der Frequenz 1 kHz (8,5 Bark) für vier Pegel, ermittelt durch Anlegen der Flankenerregungen in Form der angegebenen Näherungen in (1)-(4).

3. Spektralanalyse

Zur Berechnung der Erregung wird häufig *a priori*-Wissen über die im Signal enthaltenen Fourier-Komponenten vorausgesetzt. Ist das Schallsignal diesbezüglich unbekannt, muß zunächst eine Spektralanalyse vorgenommen werden.

Die Fourier-t-Transformation nach Terhardt [6] ist eine Kurzzeit-Spektraltransformation, deren Parameter (Bandbreite, Frequenzskalierung) gehöradäquat eingestellt werden können. Schlang und Mummert [2] konnten zeigen, daß durch die Wahl einer geeigneten Fensterfunktion die Selektivität erhöht werden kann, ohne dabei das Zeitverhalten stark negativ zu beeinflussen.

Der allgemeine Ansatz für das zeitvariante, komplexe FTT-Spektrum des kausalen Signals $s(t)$ lautet:

$$S(f, t) = \int_0^t s(\tau) h(t - \tau) e^{-j2\pi f\tau} d\tau \quad (5)$$

Dabei ist $h(t)$ eine gleitende Fensterfunktion, mit der das Zeitsignal $s(t)$ gewichtet wird. Wird diese in der Form $t^3 e^{-at}$ ($t \geq 0$) gewählt, ergibt sich die sog. FTT 4. Ordnung, da $h(t)$ der Laplace-Transformierten eines Tiefpaß 4. Ordnung entspricht. Die Funktion $a(f)$ wird proportional zur Bark-Skala gewählt, womit sich eine gehöradäquate, frequenzabhängige Analysebandbreite ergibt.

Liegt das Zeitsignal in abgetasteter Form vor, kann das Betragsquadratspektrum $P(f, t) = |S(f, t)|^2$ bei den Analysefrequenzen zu jedem Abtastzeitpunkt effizient digital berechnet werden. Danach wird das Spektrum mit einem Tiefpaß 1. Ordnung in Zeitrichtung geglättet und einmal pro Auswertintervall ausgegeben. Tab. 1 gibt Auskunft über die weiteren Parameter der Analyse, die an dieser Stelle nicht weiter diskutiert werden sollen. Das Spektrum eines Sinustones ist im eingeschwungenen Zustand nicht zeitabhängig und zeigt einen glockenförmigen Verlauf, der der Fourier-Transformierten der Fensterfunktion entspricht. Soll aus einem solchen

Tab. 1: Parameter der Spektralanalyse

Abtastfrequenz des Zeitsignals	16 kHz
Auswerteintervall	4 ms
Abstand der Analysefrequenzen	0,05 Bark
Anzahl der Analysefrequenzen	380
Frequenzbereich	20 Hz - 5,5 kHz
3 dB-Bandbreite der Analyse B	0,2 Bark

Spektrum eine Erregungsverteilung berechnet werden, muß das berücksichtigt werden. Bei Voruntersuchungen zeigte sich, daß die FFT-Analyse die große Steigung der unteren Flanke der Erregungsverteilung eines schmalbandigen Signals erst erreicht, wenn mindestens die FFT 4. Ordnung verwendet wird.

4. Berechnung der Erregungsverteilung

In jedem Auswerteintervall wird aus dem FFT-Kurzzeitspektrum der Erregungsverlauf ermittelt, indem die Beiträge aller Frequenzstützstellen linear überlagert werden (Intensitätsaddition). Durch Normierung der errechneten Summenintensitäten auf die Anzahl der in eine Analysebandbreite fallenden Frequenzstützstellen kann das richtige Pegelniveau wiederhergestellt werden. Der Beitrag einer Stützstelle wird durch das Erregungspegelmuster eines Sinustones mit den zugehörigen Werten von Frequenz und Pegel angenommen. Damit durch die Überlagerung der Spektralverteilung mit den Erregungsmustern die richtige Verteilung entsteht, müssen die Näherungen der unteren Flanke Gl.(1,2) für die oben angegebenen Analyseparameter wie folgt angepaßt werden:

$$S'_{11} = \left[2 + 0,6 \cdot \frac{L_M}{\text{dB}} \right] \frac{\text{dB}}{\text{Bark}}; \quad -1 \leq \frac{z - z_M}{\text{Bark}} \leq 0, \quad (6)$$

$$S'_{12} = \left[0,275 \cdot \frac{L_M}{\text{dB}} \right] \frac{\text{dB}}{\text{Bark}}; \quad \frac{z - z_M}{\text{Bark}} < -1. \quad (7)$$

Die Näherung der oberen Flanke kann unverändert übernommen werden, da sich die Form der Spektralverteilung bei den hier gewählten Analyseparametern nur innerhalb der unmittelbaren Umgebung des Maximums auf die obere Flanke stark auswirkt. Abb. 2 zeigt die auf diese Weise ermittelte Erregungsverteilung eines 1 kHz Sinustons. Im Vergleich zu den vorgegebenen Mustern in Abb. 1 ist das Maximum verrundet, der Verlauf sonst über weite Bereiche identisch.

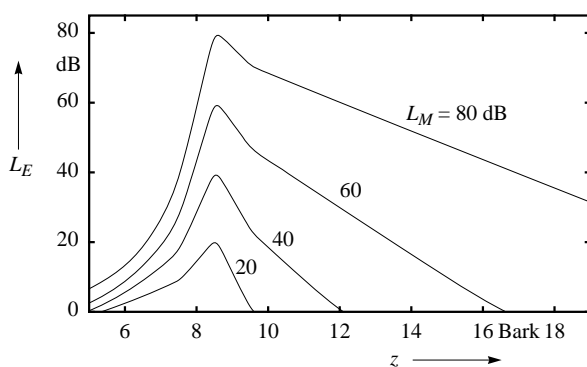


Abb. 2: Erregungspegel-Tonheitsmuster eines Sinustons der Frequenz 1 kHz (8,5 Bark) für vier Pegel, ermittelt durch Berechnung aus dem FFT-Spektrum.

5. Erregungsverteilung eines harmonischen komplexen Tons

Als Anwendungsbeispiel wurde die Erregungsverteilung eines harmonischen komplexen Tons mit Hilfe des beschriebenen Verfahrens ermittelt. Dabei sollte geklärt werden, ob die lineare Über-

lagerung der Einzelerrungen auch bei einem komplexen Signal zum richtigen Ergebnis führt.

Zur experimentellen Bestimmung der Erregungsverteilung eines harmonischen komplexen Tons ist die Messung der Simultanhörschwelle praktisch unmöglich, da bei der Messung Kombinationstöne und Schwebungen auftreten, die das Ergebnis verfälschen würden. Deshalb wird hier die durch Messung der Pulsationsschwelle bestimmte Erregung als Vergleich herangezogen. Diese kann als weitgehend identisch mit der aus Simultanhörschwellen bestimmten Erregung angenommen werden. Abb. 3 zeigt das Pulsationsschwellenmuster eines harmonischen komplexen Tons der Grundfrequenz 250 Hz bestehend aus 10 Harmonischen gleichen Pegels (70 dB SPL, gemessen von Houtgast [1]). Die ersten sieben

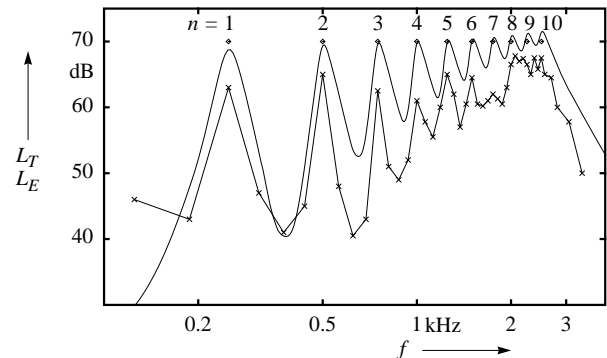


Abb. 3: Pulsationsschwellenmuster (X) eines harmonischen komplexen Tons nach Houtgast [1]. Rauten: Pegel der Harmonischen. Durchgezogene Linie: aus dem FFT-Spektrum berechnete Erregungsverteilung.

oder acht Harmonischen verursachen ausgeprägte Maxima im Pulsationsschwellenmuster. Diese Harmonischen können als Spektraltonhöhen einzeln herausgehört werden. Die Maxima des Pulsationsschwellenmusters liegen methodisch bedingt deutlich unter dem Pegel der Harmonischen.

Da bei den hier gewählten Analyseparametern der Frequenzabstand der Harmonischen immer deutlich über der Analysebandbreite liegt, ist das FFT-Spektrum und damit auch die daraus ermittelte Erregungsverteilung (siehe Abb. 3) des Tons nicht zeitabhängig. Wie das Pulsationsschwellenmuster zeigt die Erregungsverteilung eine mit wachsender Frequenz abnehmende Ausprägung der Maxima. Die drei letzten Maxima liegen weniger als 3 dB über den umgebenden Minima, was damit als sinnvolle Ausprägungsschwelle für die Zuordnung einer Spektraltonhöhe definiert werden könnte. Die Minima liegen wegen der Asymmetrie der Flanken nicht beim arithmetischen Mittelwert der Frequenzen zweier benachbarter Maxima, sondern deutlich darüber. Der im Vergleich zur Pulsationsschwelle steile Abfall unter 200 Hz ist darauf zurückzuführen, daß der Einfluß der Ruhehörschwelle bei der Berechnung noch nicht berücksichtigt wurde. Insgesamt zeigt sich auch für diesen aus mehreren Komponenten bestehenden Schall ein sinnvolles Ergebnis.

Literatur

- [1] Houtgast, T. (1974). TNO Report, Institute for Perception, The Netherlands.
- [2] Schlang, M., Mummert, M. (1990). DAGA'90, S. 1043-1046.
- [3] Schöne, P. (1979). Acustica, 43: 197-204.
- [4] Sonntag, B. (1982). Acustica, 52: 95-97.
- [5] Terhardt, E. (1979). Hearing Research, 1: 155-182.
- [6] Terhardt, E. (1985). Acustica, 57: 242-256.
- [7] Valenzuela, M.N. (1997). DAGA'97, in diesem Tagungsband.
- [8] Zwicker, E., Jaroszewski, A. (1982). J. Acoust. Soc. Am., 71: 1508-1512.
- [9] Zwicker, E., Fastl, H. (1990). Psychoacoustics, Springer.