

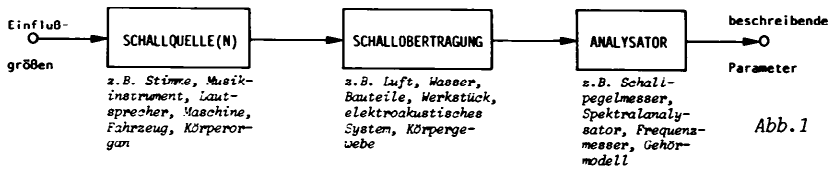
ASPEKTE UND MÖGLICHKEITEN DER GEHÖRBEZOGENEN SCHALLANALYSE UND -BEWERTUNG

E. Terhardt

Institut für Elektroakustik, Technische Universität München

1. Einleitung

Der Gesichtspunkt der Gehörbezogenheit stellt nur einen von drei wesentlichen Aspekten der Schallanalyse dar. Die beiden anderen sind die *quellenbezogene Analyse* und die *übertragungswegbezogene Analyse*. Abb.1 illustriert durch schemati-



sche Darstellung der üblichen Konfiguration von Schallquelle, Übertragungsweg und Analysator, was gemeint ist. Die quellenbezogene Analyse wird unter dem Gesichtspunkt durchgeführt, daß die Einflußgrößen, welche das Verhalten der Schallquelle bestimmen, ermittelt werden sollen (z.B. Maschinengeräuschanalyse, Herzschallanalyse). Bei der übertragungswegbezogenen Analyse stehen die physikalisch-technischen Eigenschaften der Übertragungsstrecke im Mittelpunkt (z.B. Werkstoffprüfung, medizinische Ultraschalldiagnose).

Im vorliegenden Beitrag wird ausschließlich die dritte Analyseart, die gehörbezogene Schallanalyse, betrachtet. Diese erstreckt sich auf die Extraktion von Parametern bzw. Meßwerten aus dem Schallsignal, welche bestimmten Gehörfunktionen bzw. Wahrnehmungen entsprechen sollen. Weil das Gehör an die Hörschalle der Umwelt durch Evolution hervorragend angepaßt ist, kann man davon ausgehen, daß die unter diesem Aspekt extrahierten Meßgrößen auch dort von Wert sind, wo es im Grunde um eine der beiden übrigen Analysearten geht.

In Bezug auf die Wahrnehmung von Schall durch das Gehör lassen sich zweckmäßigerweise die folgenden drei Aspekte unterscheiden.

(1) *Schall als Signal*, d.h. als Träger von Information. Beispiele: Sprache, Musik, Klingelzeichen, Alarmsirene. Im Bereich der Hörwahrnehmung sind als Informationsträger die Hörempfindungen, z.B. die Lautheit, die Tonhöhe und die Klangfarbe, anzusehen. Die unter diesem Aspekt durchgeführte Analyse erstreckt sich auf die Extraktion von Meßgrößen, welche den Hörempfindungen mehr oder weniger genau entsprechen.

(2) *Schall als Störung der Kommunikation*. Beispiele: Verstärkerrauschen, Umweltgeräusche. Die Analyse erstreckt sich auf die störende (z.B. verdeckende) Wirkung des Schalles.

(3) *Schall als psychologisch/physiologisch wirksamer Reiz*. Als Kriterien für Schallwirkungen dieser Art kommen die Störung von Tätigkeiten, emotionale Wirkungen und physiologisch meßbare Effekte in Frage. Derartige Wirkungen sind oft negativ, d.h. schädlich, und Schalle mit solchen Wirkungen bezeichnet man als Lärm. Jedoch gibt es auch positive (nützliche) Wirkungen dieser Kategorie, die beispielsweise in der Musiktherapie genützt werden.

Die unter vorstehend beschriebenen drei Gesichtspunkten zu analysierenden Parametergruppen, also die Verdeckungseffekte, die Hörempfindungen und die psychologisch/physiologischen Wirkungen, bilden eine Art Hierarchie, welche auf das in Abb.2 gezeigte Schema führt. Es stellt eine grobe Wiedergabe der Schallsignal-

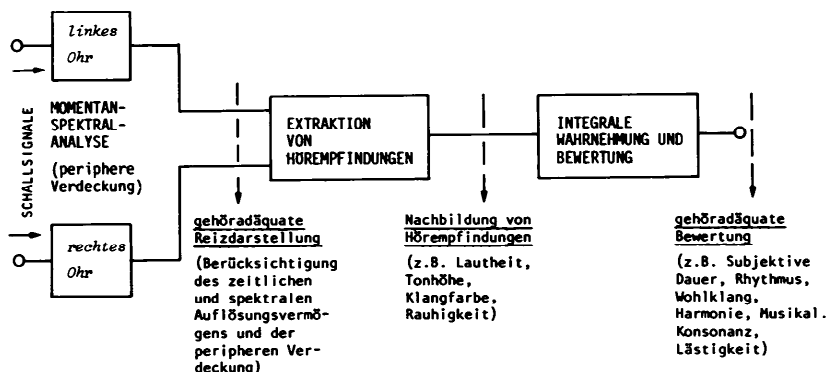


Abb.2. Schema der gehörbezogenen Schallanalyse und Bewertung.

verarbeitung durch das Hörsystem dar und impliziert insbesondere eine serielle Beziehung zwischen den genannten drei Parametergruppen.

Verdeckungseffekte sind einerseits zeitlich, andererseits spektral begrenzt. Das heißt, daß ein Schall auf einen anderen nur dann eine verdeckende Wirkung hat, wenn beide in nicht zu großem zeitlichen Abstand auftreten und ihre Spektralmuster ein Mindestmaß von Überlappung bzw. Nachbarschaft aufweisen. Die Verdeckungseffekte sind also ein ausgezeichnetes Mittel, um die tatsächlich relevante Frequenz- und Zeitselektivität des Gehörs quantitativ zu bestimmen. Dies ist in der Psychoakustik auch intensiv durchgeführt worden. Die Nachbildung der betreffenden Ergebnisse psychoakustischer Forschung ermöglicht eine gehöradäquate, d.h. insbesondere den Verdeckungseffekten Rechnung tragende Momentanspektralanalyse. In diesem Sinne liefert die erste Stufe des in Abb.2 dargestellten Schemas eine gehöradäquate Reizdarstellung durch Spektren. Sie ist in der Akustik seit langem üblich, wobei allerdings hinsichtlich der frequenz- und zeitselektiven Eigenschaften der Spektralanalyse die entsprechenden Gehördaten meist nur grob angenähert werden. Die zweite Stufe ermöglicht insbesondere die Nachbildung informationstragender Merkmale, das heißt von Hörempfindungen, und als darauf aufbauende dritte Stufe folgt die gehöradäquate Bewertung, das heißt die Nachbildung mehr oder weniger komplexer integraler Wahrnehmungs- und Bewertungsvorgänge.

Nachdem auf diese Weise die wesentlichen Aspekte der gehörbezogenen Analyse hervorgehoben wurden, sollen im folgenden die Möglichkeiten diskutiert werden. Dabei wird nicht angestrebt, eine lückenlose Übersicht über die derzeit bekannten Forschungsergebnisse zu geben, sondern es wird vielmehr versucht, anhand von Beispielen aus der Arbeit des Instituts, in welchem der Verfasser arbeitet, tatsächlich gegebene wie auch potentielle Möglichkeiten der gehörbezogenen Schallanalyse und -Bewertung zu illustrieren. Dabei werden die in Abb.2 dargestellten drei Stufen der Verarbeitung der Reihe nach betrachtet: die Frequenz- und Zeitselektivität, die Extraktion von Hörempfindungen, und die integrale Wahrnehmung und Bewertung.

2. Frequenz- und Zeitselektivität

Bei der Wahrnehmung von Sprache oder beim Anhören eines Musikstückes wird eine außerordentlich große Informationsmenge aufgenommen. Dies wird dadurch ermöglicht, daß das Gehör sich wie ein Spektralanalysator verhält, der einerseits ein hohes Frequenzauflösungsvermögen hat, andererseits jedoch auch eine sehr feine Zeitauflösung. Erst durch die (bis an die physikalisch-systemtheoretisch möglichen Grenzen vorangetriebene) Kombination von hohem Frequenz- und Zeitauflösungsvermögen wird es möglich, Sprache zu verstehen und die einzelnen Musikinstrumente eines Orchesters im Zusammenspiel zu unterscheiden.

Frequenzselektivität. Ein Maß für die Frequenzselektivität des Gehörs ergibt sich aus der Fähigkeit, gleichzeitig und stationär dargebotene Sinustöne als individuelle reine Töne mit entsprechenden Tonhöhen wahrzunehmen. Haben die Teiltöne ungefähr gleiche Amplituden, so ist dazu im unteren Spektralbereich (unterhalb etwa 500 Hz) ein Frequenzabstand von etwa 20 bis 50 Hz, im höheren Spektralbereich von etwa 1/4 Frequenzgruppe erforderlich /38, 18, 24/. Die Frequenzgruppenbreite ist überhaupt ein brauchbares Maß zur Kennzeichnung der für das periphere Gehör charakteristischen "Filter-Durchlaßkurven" /44/. Sie entspricht ungefähr der effektiven Bandbreite der Durchlaßkurven. Sie beträgt im unteren Spektralbereich etwa 100 Hz, oberhalb 1 kHz etwa 20% der Mittenfrequenz /53/.

Die "Filter-Durchlaßkurven" des Gehörs werden üblicherweise durch Messung der Mithörschwellen von Sinustönen, verdeckt durch Schmalbandrauschen oder maskierende Sinustöne, bestimmt /53/. Ihre obere Flanke hat eine größte Steilheit von ungefähr 27 dB/Bark (1 Bark entspricht einer Frequenzgruppenbreite). Die untere Flanke ist vom Schallpegel abhängig und variiert von etwa 30 dB/Bark bei kleinen Pegeln bis etwa 6 dB/Bark bei sehr hohen Pegeln. Die Frequenzselektivität des Gehörs ist also vom Schallpegel abhängig, was ein ausgeprägt nichtlineares Verhalten darstellt. Dieses steht wahrscheinlich im Zusammenhang mit anderen nicht-linearen Effekten, beispielsweise den Kombinationstönen /49, 50/.

Die Frequenzselektivität eines realen Systems hängt im Effekt nicht nur von den Filterdurchlaßkurven bzw. deren Bandbreite ab, sondern wird zusätzlich entscheidend durch die Anzahl der wirksamen Filterkanäle beeinflusst. Eine größenordnungsmäßige Aussage über diese Anzahl läßt sich aus der Anzahl unterscheidbarer Sinustöne pro Frequenzintervall ableiten, das heißt, aus den Frequenzunterschiedsschwellen. Die Anzahl aneinandergereichter Frequenzunterschiedsschwellen pro Frequenzgruppe ist angenähert konstant und beträgt etwa 30 /53/. Im Spektralbereich höchster Frequenzauflösung (kleine Bandbreite), d.h. unterhalb etwa 1 kHz, sind demnach größenordnungsmäßig 250 Filterkanäle wirksam. Im restlichen Hörbereich sind es etwa doppelt so viele.

Zeitselektivität. Verwendet man bei der Bestimmung von Mithörschwellen anstelle des stationären Maskierers einen Maskierimpuls und anstelle des stationären Testtones einen kurzen Tonimpuls, so kann man (unter Beachtung der mit dem Zeitmuster verknüpften spektralen Wirkungen) die Selektivität des Gehörs hinsichtlich der Auflösung von Zeitstrukturen quantitativ erfassen. Man unterscheidet Vor-, Simultan- und Nachverdeckung, je nachdem, ob der Testimpuls kurz vor dem Maskierer, gleichzeitig mit demselben, oder kurz danach erscheint. Die Vorverdeckung ist nur gering /10, 48, 11/. Meßbare Effekte treten hier nur für Verzögerungszeiten zwischen Test- und Maskierschall von weniger als 20 ms auf. Die Nachverdeckung dauert wesentlich länger (bis einige 100 ms). Letztere hängt auch deutlich von der Maskierimpulsdauer ab /52/. Je kürzer der Maskierimpuls, um so rascher klingt die Nachverdeckung ab. Die Abhängigkeit der Verdeckungseffekte von Frequenz und Zeit wurde systematisch z.B. von Fastl /12, 15/ gemessen. Damit wurde eine Beschreibung der Kombination von Frequenz- und Zeitauflösung durch das sogenannte Mithörschwellen-Tonheits-Zeitmuster möglich. Durch Messung der Mithörschwellen kurzer Testtonimpulse bei Verdeckung durch Maskierer mit schwankender Amplitude

kann man feststellen, wie "träge" das Gehör ist. Amplitudenänderungen, die sich innerhalb weniger Millisekunden abspielen, werden gerade noch im Mithörschwellen-Zeitmuster abgebildet /54/. Periodische Amplitudenschwankungen (Amplitudenmodulation des Maskierers) bilden sich bis zu Schwankungsfrequenzen von etwa 500 Hz ab /48/. Wird als Maskierer ein tiefer Sinuston fester Amplitude angewendet, so findet man sogar dessen Schwingungsstruktur im Mithörschwellen-Zeitmuster kurzer Testtonimpulse wieder ("Mithörschwellen-Periodenmuster" /46/).

Frequenz-Zeitverhalten. Versucht man, auf dieser Grundlage die erste Stufe der Schallverarbeitung durch das Gehör (die Momentan-Spektralanalyse) zusammenfassend zu kennzeichnen, so ergibt sich folgendes Bild.

Im Frequenzbereich unterhalb etwa 1 kHz ist die effektive Filterbandbreite (entsprechend der Frequenzgruppenbreite) 100 bis 160 Hz, und die Anzahl der Filterkanäle beträgt 2 bis 3 pro 10 Hz Frequenzintervall. Im Frequenzbereich oberhalb 1 kHz steigt die effektive Filterbandbreite bis zur Größenordnung 1 kHz an, und die Anzahl der Filterkanäle pro 10 Hz Frequenzintervall beträgt im Mittel nur noch 0,3. Aus diesen Zahlen ergibt sich, daß im Spektralbereich unterhalb 1 kHz das Frequenzauflösungsvermögen optimal und angenähert gleichförmig ist, wobei die Zeitauflösung eingeschränkt ist. Oberhalb 1 kHz sind die Verhältnisse umgekehrt: Die Zeitauflösung erreicht ihr Optimum, auf Kosten der Frequenzauflösung. Diese Charakteristika sind durch übliche Verfahren der Fourieranalyse, beispiels-

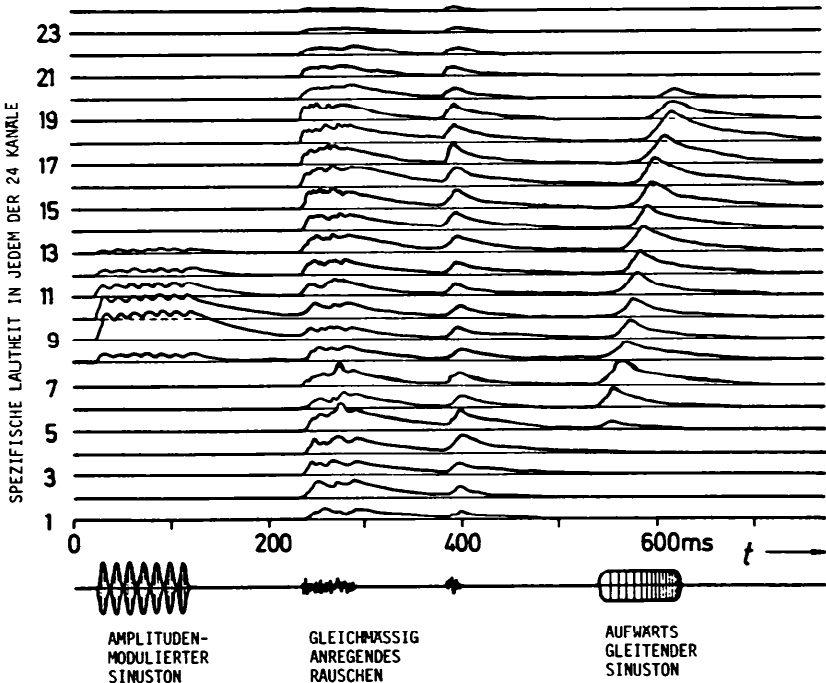


Abb.3. Beispiele für gehöradäquate Reizdarstellung durch Momentan-Spektralanalyse. Verteilung der Spezifischen Lautheit auf 24 Spektralkanäle für einen amplitudenmodulierten Sinuston (1 kHz, Mod.Freq. 70 Hz), gleichmäßig anregendes Rauschen (langer bzw. kurzer Impuls), und einen aufwärts gleitenden Ton. Nach /55/.

weise die FFT, nicht gut nachzubilden. Bessere Nachbildungen sind die von Zwicker /47/ in einem Lautheitsmeßgerät, bzw. von Zwicker, Terhardt und Paulus /55/ im Zusammenhang mit der Vorverarbeitung von Sprache beschriebenen Filterbänke. Abb.3 zeigt als Beispiel die von einem derartigen System in 24 Kanälen gelieferten Signale, welche weitgehend einem gehöradäquaten Momentanspektrum entsprechen (nach /55/).

Die gehöradäquate Reizdarstellung durch Momentanspektren ist demnach ein prinzipiell weitgehend gelöstes Problem. Jedoch hat diese Lösung bisher nicht allzu-große Verbreitung bei der Schallanalyse gefunden, vermutlich deshalb, weil sie verhältnismäßig aufwendig ist.

3. Extraktion von Hörempfindungen

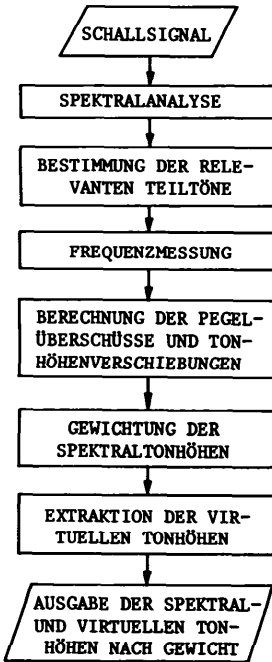
Hörempfindungen, die als eine Art von informationstragenden Merkmalen im Wahrnehmungsbereich angesehen werden können, sind die Lautstärke (Lautheit), die Tonhöhe, die Klangfarbe (ihrerseits aus mehreren Komponenten zusammengesetzt) und die akustische Rauigkeit (d.h. die Empfindung für rasche Schallschwankungen). All diese Hörempfindungen bedürfen zu ihrer Extraktion aus dem Schallsignal der Spektralanalyse. Ihre Extraktion setzt also die Reizdarstellung in Form eines gehöradäquaten Momentanspektrums voraus.

Lautheit. Seit den grundlegenden Arbeiten von Fletcher und Munson /16/ und Zwicker /41, 42, 47/ ist die Extraktion dieser Empfindungsgröße weitestgehend geklärt, und zwar sowohl für stationäre als auch für instationäre Schalle. Das Bildungsprinzip der Lautheit besteht darin, daß aus der spektralen Schallenergieverteilung durch eine nichtlineare Bewertung zunächst eine Verteilung der sogenannten Spezifischen Lautheit gebildet und diese sodann längs der Frequenzkoordinate und über ein gewisses Zeitintervall integriert wird. Als Frequenzkoordinate wird üblicherweise die sogenannte Tonheit z (Maßeinheit 1 Bark) benutzt /43/. Sie ist derart definiert, daß dem Tonheitsintervall 1 Bark die Frequenzbandbreite einer Frequenzgruppe entspricht. Die "Zeitkonstante der Lautheit" beträgt etwa 100 ms. Dies bedeutet, daß ein Schall fester Intensität unabhängig von seiner Dauer stets gleich laut erscheint, sofern die Dauer größer als 100 ms ist; ist sie kürzer, so nimmt die Lautheit mit der Dauer ab /20/. Die Gesetzmäßigkeiten der Lautheitsempfindung sind beispielsweise in /41, 42, 47/ beschrieben.

Beschränkt sich eine gehörbezogene Schallanalyse auf den Aspekt der Schallstärke, so wird vielfach zur Reduzierung des Aufwands auf ein stark vereinfachtes Verfahren zurückgegriffen. Insbesondere wird häufig einfach der A-bewertete Schallpegel herangezogen. Die dadurch bedingten Abweichungen von einer tatsächlich gehöradäquaten Schallstärkebewertung können allerdings bedenkliche Dimensionen annehmen, derart, daß in Einzelfällen gar eine Verminderung des L_A -Wertes mit einer Erhöhung der Lautheit einhergeht /51/.

Tonhöhe. Die Bedeutung der Tonhöhe als informationstragendes Merkmal ist weitgehend evident. In der Musik stellt sie den wesentlichen Informationsträger dar. Aber auch in der Sprache und bei der Identifikation und Beurteilung vieler Umweltschalle ist sie von großer Wichtigkeit. Neben der Lautheit ist die Tonhöhe die zweite grundlegende Hörempfindung. Sie hat sich im Laufe der psychoakustischen Bemühungen um ihr Verständnis als mindestens ebenso kompliziert herausgestellt wie die Lautheit. Die Komplikationen sind im wesentlichen durch ein scheinbar widersprüchliches Verhalten des Gehörs bedingt: Einerseits verhält es sich wie ein Spektralanalysator; es zerlegt Klänge in ihre Teiltöne, so daß deren Tonhöhen als Resultat der Spektralanalyse angesehen werden können. Andererseits zeigt die Erfahrung, daß es zahlreiche Klänge gibt, die zwar aus vielen Teiltönen bestehen, jedoch im wesentlichen nur eine einzige Tonhöhe hervorrufen. Es sind dies die harmonischen Klänge (sog. harmonische Komplexe Töne, also z.B. die Töne

von Musikinstrumenten), deren wahrgenommene Tonhöhe in der Regel der Grundfrequenz entspricht. In diesen Fällen scheint das Gehör sich nicht wie ein Spektralanalysator, sondern eher wie ein solcher der zeitlichen Periodizität zu verhalten (vgl. hierzu /19/). Der scheinbare Widerspruch wird durch die Erkenntnis gelöst, daß es zwei grundsätzlich verschiedene Arten von Tonhöhe gibt, und zwar die sogenannte Spektraltonhöhe einerseits und die Virtuelle Tonhöhe andererseits /26/. Die Spektraltonhöhe ist die Tonhöhe einzelner Sinustöne und als solche im vorstehend diskutierten Sinne durch Spektralanalyse erklärbar. Sie ist insofern eine Art Primärempfindung, das heißt, eine Hörempfindung, welche in vergleichsweise peripheren Bereichen des Wahrnehmungsprozesses zustandekommt. Demgegenüber wird die Virtuelle Tonhöhe als eine Hörempfindung aufgefaßt, die in höheren Verarbeitungszentren entsteht. Sie wird aus den tonalen Bestandteilen des Spektrums durch eine Art *pattern recognition* gebildet. Sie hängt also ebenso wie die Spektraltonhöhe von Spektralmerkmalen ab, wenn auch auf kompliziertere Weise. Das Gehör verhält sich demnach durchaus einheitlich wie ein Spektralanalysator und es gibt keinen Widerspruch. Dieses Konzept der Tonhöhenwahrnehmung wird insbesondere auch den zu beobachtenden Mehrdeutigkeiten der Tonhöhe von Klängen gerecht: Je nach Darbietungsbedingungen und individueller Aufmerksamkeit kann man eine oder mehrere Spektraltonhöhen bzw. Virtuelle Tonhöhen wahrnehmen /26, 27, 31/.



Ein Tonhöhenanalyseverfahren, welches auf den geschilderten Prinzipien beruht, hat die in Abb.4 gezeigte Struktur. Nach der Spektralanalyse werden zunächst die "relevanten Teiltöne" bestimmt, das heißt, diejenigen Teiltöne, welche nicht infolge wechselseitiger Verdeckung unhörbar bleiben. Die Frequenzmessung und die Bestimmung der eventuell durch Teilmaskierung hervorgerufenen Tonhöhenverschiebungen liefert danach die genauen Tonhöhen der relevanten Teiltöne. Durch Berechnung ihrer sogenannten Pegelüberschüsse /31, 32/ und eine spektrale Gewichtung /37/ wird sodann ein Maß für die Deutlichkeit der einzelnen Spektraltonhöhen gewonnen. Das Ensemble von Spektraltonhöhen mit den zugehörigen Gewichten ("Deutlichkeiten") wird als *Spektraltonhöhenmuster* bezeichnet. Dieses beschreibt einerseits die primären Tonhöhenwahrnehmungen, andererseits bildet es die Grundlage für die Extraktion des zweiten wesentlichen Musters, nämlich des Musters der Virtuellen Tonhöhen.

Abb.4. Blockschema der Tonhöhenanalyse. Relevante Teiltöne sind solche, die nicht verdeckt sind. Die Extraktion der Virtuellen Tonhöhen erfolgt nach dem Prinzip der subharmonischen Koizidenzdetektion (Abb.5). Nach /37/.

Die Extraktion der Virtuellen Tonhöhen (und ihrer Gewichte) erfolgt nach einem neuerdings beschriebenen Verfahren /37/, welches eine Weiterentwicklung des schon früher entwickelten Verfahrens der "subharmonischen Koizidenzdetektion" /31, 33/ darstellt.

Das letztgenannte Prinzip ist in vereinfachter Weise durch ein Beispiel in Abb.5 veranschaulicht. Die Darstellung ist von oben nach unten zu lesen. Während an der oberen Abszisse die Frequenz (im Sinne der Spektralvariablen) angegeben ist, zeigt die untere Abszisse die Virtuelle Tonhöhe im "pitch-units Maß" pu an. Dieses Maß hat den Vorteil, daß sein Zahlenwert gleich dem der Frequenz des äquivalenten Sinustones ist. Im dargestellten Beispiel (harmonischer Residualton mit

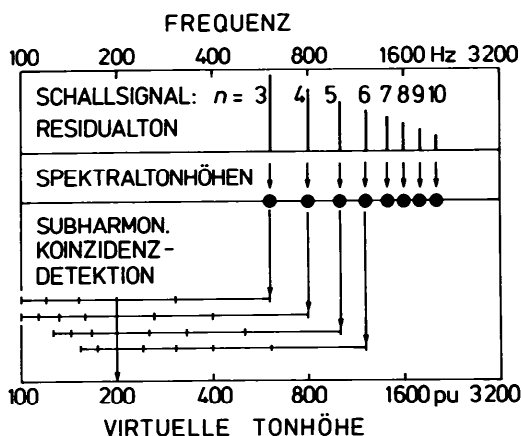


Abb.5. Veranschaulichung des Prinzips der subharmonischen Koinzidenzdetektion zur Extraktion der Virtuellen Tonhöhe aus dem Schallspektrum. pu bedeutet "pitch units".

ergeben sich zahlreiche potentielle Virtuelle Tonhöhen, die im allgemeinen nicht miteinander übereinstimmen. Im Falle eines harmonischen Reizspektrums wird sich jedoch bei derjenigen Tonhöhe, welche der Grundfrequenz entspricht, eine Koinzidenz ergeben, wie im Bild bei 200 pu angezeigt. Von allen potentiellen Virtuellen Tonhöhen besitzt demnach die derart ausgezeichnete das höchste Gewicht und wird als die vorherrschende angesehen. Dieses Verfahren ist beispielsweise geeignet, um die Grundfrequenzextraktion von Sprache zu bewerkstelligen /33/.

Ein Anwendungsbeispiel des in Abb.4 gezeigten, umfassenden Tonhöhenextraktionsverfahrens auf Klänge, die anderweitig kaum oder nur schwer gehöradäquat zu beurteilen sind, nämlich auf Glockenklänge, wird in Abb.6 illustriert. Es handelt

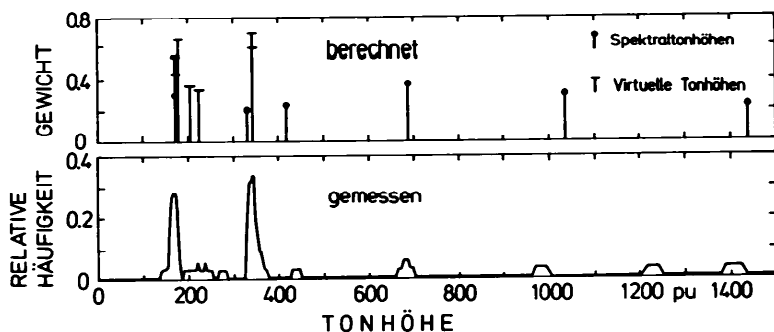


Abb.6. Oben: Mit dem Tonhöhenextraktionsverfahren berechnete Spektral- bzw. Virtuelle Tonhöhenmuster für den Klang einer Kirchenglocke. Unten: Histogramm der von 18 Versuchspersonen angegebenen Tonhöhen derselben Glocke. Nach /37/.

sich um eine ältere Kirchenglocke, welche durch Anschlag mit dem Originalklöppel von Hand zum Klingen gebracht und aus etwa 2 m Entfernung auf Tonband aufgenommen wurde. Das Frequenzspektrum zeigt eine große Zahl von Teiltönen, wobei die wahr-

den Harmonischen Nr. 3 bis 10) deuten die vertikalen Pfeile zunächst die erste Stufe der Verarbeitung an: Aus den Spektralkomponenten werden Spektraltonhöhen gebildet. Das Spektraltonhöhenmuster (ohne Gewichte) ist durch Punkte dargestellt. Die subharmonische Koinzidenzdetektion wird durch Anlegen eines Maßstabes, der die Tonhöhenintervalle der harmonischen Teiltonreihe trägt, an jede einzelne Spektraltonhöhe veranschaulicht. Die Markierungen auf dem Maßstab geben sodann in Bezug auf die untere Abszisse die Reihe der möglichen Virtuellen Tonhöhen an, welche von einer einzelnen Spektraltonhöhe abgeleitet werden können. Wird derselbe Maßstab der Reihe nach an die verschiedenen

Spektraltonhöhen angelegt, so

genommene Tonhöhe nicht ohne weiteres ersichtlich ist. Das Extraktionsverfahren liefert das in Abb.6 oben dargestellte Tonhöhenmuster (eine Kombination von Spektraltonhöhen- und Virtuelle Tonhöhenmuster /37/). Dieses Muster zeigt einerseits die Mehrdeutigkeit der Tonhöhenwahrnehmung an, andererseits weist es zwei ausgeprägte Gewichtsmaxima bei etwa 340 und 170 pu (tonhöhenäquivalente Frequenz 340 bzw. 170 Hz) auf. Da diese beiden Tonhöhen im Oktavabstand stehen, ist es vom Standpunkt der musikalischen Tonbezeichnung gleichgültig, welchem der beiden Maxima man letztlich die dominante musikalische Tonhöhe (in diesem Falle angenähert die Note F) zuordnet.

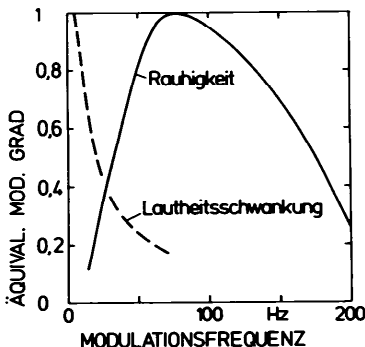
Als Kontrollexperiment wurden mit dem Glockenklang Hörversuche gemacht, in welchen 18 Versuchspersonen die dem Klang spontan zugeordneten Tonhöhen durch Abgleich eines Sinustones wiederholt anzugeben hatten /23/. Das Histogramm der Einstellungen ist im unteren Teil von Abb.6 dargestellt. Man erkennt ohne weiteres die hohe Korrelation mit den vom Extraktionsverfahren vorausgesagten Werten.

Klangfarbe. Die Klangfarbe ist als informationstragendes Merkmal von größter Bedeutung. Die Erkennung von Sprachlauten, die Identifikation der Quellen von Umweltgeräuschen, sowie von Musikinstrumenten beruhen zu einem erheblichen Teil auf der Klangfarbenwahrnehmung. Die psychoakustische Erforschung dieser Hörempfindung hat sich als außerordentlich schwierig erwiesen und es liegen bisher lediglich Teilergebnisse vor. Der gegenwärtige Stand der Forschung erlaubt deshalb eine gehörrädequate Beurteilung der Klangfarbenaspekte nur in begrenztem Maße.

Die einzige Klangfarbenkomponente, die bisher als brauchbare psychoakustische Größe etabliert wurde, ist die sogenannte Schärfe. Diese Komponente wurde durch von Bismarck /5/ als eine von drei wesentlichen Klangfarbenkomponenten identifiziert und näher untersucht. Sie hängt insbesondere mit der Schallintensität in höheren Spektralbereichen zusammen. Ihre Extraktion aus dem Schallreiz kann mit Hilfe des Bismarckschen Funktionsschemas durchgeführt werden.

Ein anderer Aspekt der Klangfarbe, nämlich derjenige der "Tonalität", wurde neuerdings aufgegriffen /4/. Dabei geht es um die Beschreibung derjenigen Klangfarbenunterschiede, die durch das Vorhandensein einer spektralen Feinstruktur, das heißt, tonalen Anteilen, neben der Grobstruktur (der spektralen Hüllkurve, z.B. dem Terzpegelspektrum) hervorgerufen werden. Weil das zuvor beschriebene Spektraltonhöhenmuster eine gehörrädequate Beschreibung der tonalen Schallanteile liefert, ist dieses Muster geeignet, um dem tonalen Aspekt der Klangfarbe Rechnung zu tragen.

Rauhigkeit. Die letzte der hier betrachteten grundlegenden Hörempfindungen ist die akustische Rauhigkeit. Dabei handelt es sich um diejenige Empfindung, welche durch rasche Amplitudenschwankungen hervorgerufen wird. Schallfluktuationen mit



Schwankungsfrequenzen unterhalb etwa 20 Hz werden nicht als Rauhigkeit, sondern als Schwankung der Lautheit (Tremolo) empfunden. Diese Empfindung nimmt mit anwachsender Schwankungsfrequenz rasch ab, wie in Abb.7 dargestellt. Statt dessen entsteht die Rauhigkeitsempfindung. Abb.7 zeigt einen typischen Verlauf dieser Empfindung als Funktion der Schwankungsfrequenz. Er gilt für den Fall, daß die Rauhigkeit durch Schallanteile im Spektralgebiet oberhalb etwa 1 kHz hervorgerufen wird. Wenn tiefere

Abb.7. Wahrgenommene Stärke von Amplitudenschwankungen (schematisch), ausgedrückt durch äquival. Modulationsgrad, als Funktion der Schwankungsfrequenz.

Spektralanteile maßgebend sind, erreicht die Rauhigkeit ihren Maximalwert schon bei kleineren Schwingungsfrequenzen, weil die Spektralzerlegung durch das periphere Gehör keine höheren Schwingungsfrequenzen wirksam werden läßt.

Die Rauhigkeit ist als weiteres informationstragendes Merkmal anzusehen. Sie repräsentiert diejenigen Schallmerkmale, die beispielsweise eine tiefe oder heisere Stimme rau, ein Summersignal schnarrend, ein Motorrad knatternd erscheinen lassen. Sie trägt wesentlich zur Verminderung des Wohlklangs (Sensorische Konsonanz, vgl. Abschn. 4) beliebiger Schalle, sowie zur Dissonanz musikalischer Klänge bei. Ihre Bildung aus dem Momentanspektrum ist im Prinzip durch zahlreiche psychoakustische Untersuchungen geklärt /3, 24, 25, 28, 39, 40, 13, 1/, jedoch steht ein genaues, allgemeingültiges Funktionsschema noch nicht zur Verfügung. Neben dem in Abb.7 sichtbaren bandpaßartigen Frequenzgang (hinsichtlich der Schwingungsfrequenz) ist für die Rauhigkeit vor allem die Eigenschaft charakteristisch, nahezu ausschließlich von der *relativen* Größe der Amplitudenschwankungen (beispielsweise dem Modulationsgrad eines amplitudenmodulierten Signals) abzuhängen. Der Schallpegel, und damit letztlich die absolute Größe der Amplitudenschwankungen, spielt nur eine geringe Rolle /28/. Dies ist der Grund dafür, daß schwache, rau klingende Schallanteile, die zugleich mit starken, nicht rau klingenden Anteilen auftreten, den Wohlklang des Gesamtschalles stark herabmindern können (es sei denn, die schwachen Spektralanteile würden durch die starken völlig verdeckt).

4. Integrale Wahrnehmung und Bewertung

Im Schema der gehörbezogenen Analyse (Abb.2) wurde eine Vielfalt von Wahrnehmungs- und Bewertungsaspekten in der letzten Verarbeitungsstufe ("integrale Wahrnehmung und Bewertung") zusammengefaßt. Durch die Bezeichnung "integral" soll angedeutet werden, daß hiermit Wahrnehmungen und Schalleigenschaften gemeint sind, die durch das Zusammenwirken von mehreren, möglicherweise zahlreichen, zeitlich bzw. spektral getrennten Anteilen zustandekommen. Dazu werden beispielsweise die Wahrnehmung der Subjektiven Dauer eines Schalles, der Rhythmus, der Wohlklang und die Lästigkeit gerechnet. Wenn auch nicht sicher ist, daß die gesamte Schallverarbeitung im Prinzip die in Abb.2 angenommene serielle Struktur hat, so besteht doch eine gewisse Wahrscheinlichkeit dafür und man kann annehmen, daß die gehörbezogene Extraktion der Hörempfindungen (also die ersten beiden Stufen der Verarbeitung nach Abb.2) eine sehr gute, möglicherweise sogar notwendige Voraussetzung für die adäquate Beurteilung der genannten integralen Aspekte ist. Dies soll an einigen Beispielen erläutert und diskutiert werden.

Subjektive Dauer. Schallereignisse, die kürzer als etwa 800 ms dauern, weisen eine Diskrepanz zwischen ihrer tatsächlichen (objektiven) und subjektiv geschätzten Dauer auf, die umso größer wird, je kürzer die Dauer ist /45, 6/. Wird beispielsweise die objektive Dauer eines Schallimpulses halbiert, so erscheint der Impuls dem Gehör deutlich *mehr* als halb so lang /45/. Vergleicht man die wahrgenommene Dauer eines Schallimpulses mit derjenigen einer Schallpause, die zwischen zwei längeren Impulsen liegt, so ist die Subjektive Dauer der Pause systematisch geringer als diejenige des Impulses, obwohl die entsprechenden objektiven Dauern gleich sind /7/. Burghardt /8/ hat mit Hilfe eines Funktionsschemas gezeigt, daß dieses Verhalten durch die Annahme erklärt werden kann, eine für die Wahrnehmung maßgebende Zwischengröße habe ein Ein- und Ausschwingverhalten, das durch eine Zeitkonstante von 200 ms beschrieben wird, und für die subjektive Dauer eines Schallereignisses sei das Zeitintervall zwischen den beiden Zeitpunkten maßgebend, in denen diese Größe eine bestimmte Schwelle über- bzw. unterschreitet. Fastl /14/ hat ein ähnliches Modell angegeben, welches sich auf die Mithörschwellen-Zeitmuster stützt und eine andere Definition der Schwelle benützt.

Es scheint, daß man in der Tat die gehörmäßige Schätzung der Dauer eines Schalles als eine "integrale Bewertung" im vorstehend erläuterten Sinne auffassen kann. Dann ist konsequenterweise anzunehmen, daß diese Bewertung auf der Basis der Hörempfindungen erfolgt. Von diesen dürfte insbesondere die Lautheit maßgebend sein. In der Tat läßt sich zumindest das Burghardsche Modell recht gut auf die Lautheit als Zwischengröße beziehen, denn deren Ein- und Ausschwingzeitkonstante ist von derselben Größenordnung (100 ms) wie die nicht näher bezeichnete Zwischengröße des Funktionsschemas.

Rhythmus. Rhythmus entsteht dadurch, daß die einzelnen Schallereignisse einer längeren Folge von Schallvorgängen durch "integrale Bewertung" miteinander in Zusammenhang gebracht bzw. zeitlich aufeinander bezogen werden. Daraus ergibt sich die grundsätzliche Frage, wodurch Schallvorgänge zu "Ereignissen" werden und welches die effektiven Zeitpunkte der Ereignisse sind. Durch Untersuchungen der einfachsten Art von Rhythmus, nämlich des gleichmäßigen Rhythmus, wurde versucht, Antworten auf diese Fragen zu finden /34, 21/. Die Ergebnisse dieser Untersuchung hat Schütte /22/ in einem Funktionsschema zusammengefaßt, welches mit den erläuterten Modellen der Subjektiven Dauer interessante Gemeinsamkeiten aufweist. Zur Beschreibung der Ereigniszeitpunkte von Schallvorgängen verschiedenster Hüllkurven erweist sich wiederum die Annahme einer Zwischengröße mit einer Zeitkonstante der Größenordnung 100 ms als notwendig und adäquat. Die bisherigen Ergebnisse und Modellansätze stehen also mit dem geschilderten Konzept, die Rhythmuswahrnehmung als einen Vorgang der integralen Bewertung von Lautheitsverläufen aufzufassen, zumindest nicht in Widerspruch.

Wohlklang (Sensorische Konsonanz). Es ist eine allgemeine Erfahrung, daß Schalle mehr oder weniger wohlklingend empfunden werden, je nach ihrer zeitlichen Struktur und spektralen Zusammensetzung. Dieser Aspekt der integralen Bewertung ist weitgehend unabhängig von individuellen Gemütszuständen oder anderen Einflüssen. Er wird deshalb auch als Sensorische Konsonanz bezeichnet /29/. Um die allgemeinen Erfahrungen hierüber durch systematische Daten zu ergänzen, wurden Wohlklangsbewertungen mit 17 Umweltschallen durchgeführt /35, 36/. Dabei zeigte sich, daß in der Tat sehr verschiedenartige Schalle mit Erfolg im Wohlklang beurteilt werden und man eine Reihenfolge der Schalle hinsichtlich ihres Wohlklangs erhält, die mit der allgemeinen Hörerfahrung gut übereinstimmt. Tabelle I zeigt die in den Tests ermittelte Rangfolge.

RANG	SCHALLBEZEICHNUNG
1	Musikal. Akkord
2	Frauenstimme /a/
3	Glockenlärten
4	Männerstimme /a/
5	Reiner Ton 1 kHz
6	Wasserhahn
7	Staubsauger
8	Auto
9	Motorrad
10	Telefon
11	Schreibmaschine
12	Flugzeug
13	Weißes Rauschen
14	AM Sinuston
15	El. Kaffeemühle
16	Schlagbohrmasch.
17	Kreissäge

Tabelle I. Rangfolge von 17 Umweltschallen, ermittelt durch Paarvergleich und Wohlklangsschätzung. Nach /35/.

Entsprechend dem in Abb.2 veranschaulichten Konzept kann man annehmen, daß auch der Wohlklang sich durch integrale Bewertung zuvor ermittelter Hörempfindungen beschreiben lassen sollte (serielle Struktur des Schemas Abb.2). In der Tat wurde eine hohe Korrelation zwischen den geschilderten Wohlklangsbewertungen und den Hörempfindungen bzw. Empfindungsaspekten *Rauhigkeit*, *Schärfe* und *Tonalität* gefunden /36/. *Rauhigkeit* und *Schärfe* korrelieren mit dem Wohlklang negativ, d.h. ihr Auftreten setzt ihn herab, während die *Tonalität* (im Sinne des Auftretens tonaler Anteile im Schall) ihn erhöht. Über die Extraktion der genannten Empfindungskomponenten *Rauhigkeit*, *Schärfe* und *Tonalität* eröffnet sich somit ein Weg zur gehörbezogenen Schallbewertung hinsichtlich des Wohlklangs. Eine Anwendung dieses Konzepts auf Glockenklänge findet sich im Tagungsbeitrag von Aures /2/.

Lästigkeit. Ein Beispiel für einen aktuellen Parameter, der in einem gewissen Zusammenhang mit dem Wohlklang steht, ist die Lästigkeit. Führt man sich die eingangs erwähnten Aspekte der Schallanalyse vor Augen, so wird deutlich, daß Lästigkeit zumindest dann vorliegt, wenn

- (1) ein Kommunikationsprozeß gestört wird, und/oder
- (2) eine Tätigkeit gestört wird, und/oder
- (3) ungünstige psychologisch/physiologische Wirkungen (Stress, Ermüdung, Widerwillen) hervorgerufen werden.

Es scheint, daß der Wohlklang zumindest mit der letztgenannten Komponente der Lästigkeit in Zusammenhang steht, insofern, als ein sehr mißklingender Schall eher Stress, Ermüdung bzw. Widerwillen erzeugt und damit lästig wird, als ein wohlklingender (vgl. hierzu /9/). Damit ist jedoch nicht gesagt, daß das Problem der Lästigkeit durch Untersuchung des Wohlklangs im wesentlichen zu lösen sei. Vielmehr bedarf es dazu umfangreicher Untersuchungen der übrigen obengenannten Kriterien.

Musikalische Konsonanz. Die gehöradäquate Bewertung des Phänomens der Musikalischen Konsonanz scheiterte bisher zum erheblichen Teil daran, daß recht unterschiedliche Meinungen darüber bestehen, was man darunter verstehen soll, das heißt, an konzeptuellen Schwierigkeiten. Deshalb wurde versucht, ein Konzept der Musikalischen Konsonanz zu entwickeln, welches dem alltäglichen Begriffsgebrauch möglichst nahekommt und doch einigermaßen präzise ist /29/. Danach werden unter dem Begriff Musikalische Konsonanz zum einen das Phänomen des Wohlklangs (Sensorische Konsonanz; ein nicht musikspezifisches Phänomen), zum anderen die mit dem Sammelbegriff *Harmonie* gemeinten musikspezifischen Phänomene zusammengefaßt (letztere sind insbesondere die Tonverwandtschaft und die Grundtonbezogenheit). Dieses Konzept stimmt im wesentlichen mit der Helmholtzschen Auffassung von den musikalisch relevanten Tonempfindungen /17/ überein, jedoch wurde von Helmholtz eine andere Terminologie benützt und insbesondere der Konsonanzbegriff weit enger gefaßt als üblich.

Mit Hilfe dieses Konzepts der Musikalischen Konsonanz konnte diese weitgehend auf bekannte Gesetzmäßigkeiten der Hörempfindungen (im Sinne einer "integralen Bewertung" derselben) zurückgeführt werden. Die erste Komponente, die Sensorische Konsonanz, erweist sich insbesondere als von der Rauigkeit abhängig (was schon Helmholtz festgestellt hatte). Die zweite Komponente, die Harmonie, konnte weitgehend durch die Gestalttheorie der Tonhöhenwahrnehmung erklärt werden. Insbesondere erklärt die Theorie der Virtuellen Tonhöhe ohne weiteres die Tatsache, daß musikalischen Akkorden, wenn sie "harmonisch" sind, bestimmte Grundtöne zugeordnet werden. Die Akkordgrundtöne sind nichts anderes als die Virtuellen Tonhöhen dieser Klänge. Damit ergibt sich die Möglichkeit der Analyse musikalischer Klänge hinsichtlich ihrer tonal/harmonischen Eigenschaften im Sinne der Musiktheorie durch Anwendung des beschriebenen Tonhöhenextraktionsverfahrens (Abb.4). Gegenüber der konventionellen Musiktheorie bedeutet dies insofern einen Fortschritt, als das Verfahren die realen akustischen Eigenschaften (Instrumentation) der Klänge berücksichtigt, im Gegensatz zur bloßen Notation.

5. Schlußbemerkungen

Mit Hilfe der diskutierten Aspekte und Möglichkeiten der gehörbezogenen Schallanalyse sollte vor allem diesem Begriff ein Sinn gegeben werden, welcher einerseits psychoakustischen Erkenntnissen, andererseits potentiellen Anwendungen gerecht wird. Durch die hierarchische Gliederung der Verarbeitungsstufen (gehöradäquate Reizdarstellung; Hörempfindungen; integrale Bewertung) sollte hervorgehoben werden, daß einerseits Wege gangbar erscheinen, auf denen gehörbezogene Schallanalyse in wissenschaftlich begründeter Weise bewerkstelligt werden kann, und andererseits innerhalb dieses Schemas noch viele Fragen offen sind, so daß hier zugleich eine Art Forschungskonzept vorliegt.

Der Verfasser dankt den Herren Prof.Dr.-Ing. E. Zwicker und Dr.-Ing. H. Fastl für wertvolle Anregungen. Dieser Beitrag ist im Rahmen des Sonderforschungsbereichs 50 "Kybernetik", gefördert durch die Deutsche Forschungsgemeinschaft, entstanden.

Schrifttum

- /1/ Aures, W., in "Fortschritte der Akustik", VDE-Verlag, Berlin, pp.623-626 (1980). /2/ Aures, W., Wohlklangsbeurteilung von Kirchenglocken, DAGA 81 (dieser Berichtsband). /3/ Békésy, G., Experiments in Hearing, McGraw-Hill, New York (1960). /4/ Benedini, K., Biol. Cybernetics 34, 111-117 (1979). /5/ Bismarck, G., Acustica 28, 146-172 (1974). /6/ Burghardt, H., Acustica 28, 278-284 (1973). /7/ Burghardt, H., Acustica 28, 284-290 (1973). /8/ Burghardt, H., Kybernetik 12, 21-29 (1972). /9/ Cardozo, B.L. und van Lieshout, R.A., in IPO Annual Progr. Rep. 14 (1979), pp. 27-32. /10/ Elliott, L.L., Audiology 10, 65-76 (1971). /11/ Fastl, H., Acustica 35, 287-302 (1976). /12/ Fastl, H., Acustica 36, 317-331 (1977). /13/ Fastl, H., in "Psychophysics and Physiology of Hearing" (E.F.Evans&J.P.Wilson eds.), Academic Press, London, pp. 403-414 (1977). /14/ Fastl, H., J.Acoust.Soc. Am. 61, 162-168 (1977). /15/ Fastl, H., Acustica 43, 282-294 (1979). /16/ Fletcher, H. und Munson, W.A., J.Acoust.Soc.Am. 9, 1-10 (1937). /17/ Helmholtz, H., Die Lehre von den Tonempfindungen, Vieweg, Braunschweig (1863). /18/ Plomp, R., J.Acoust.Soc.Am. 36, 1628-1636 (1964). /19/ Plomp, R. & Smoorenburg, G.F. (Hrsg.), "Frequency Analysis and Periodicity Detection in Hearing", Sijthoff, Leiden (1970) /20/ Port, E., Frequenz 13, 242-245 (1959). /21/ Schütte, H., Acustica 41, 197-206 (1978). /22/ Schütte, H., Biol. Cybernetics 29, 49-55 (1978). /23/ Seewann, M. und Terhardt, E., in "Fortschritte der Akustik", VDE-Verlag, Berlin, pp.635-638 (1980). /24/ Terhardt, E., Acustica 20, 210-214 (1968). /25/ Terhardt, E., Acustica 20, 215-224 (1968). /26/ Terhardt, E., Acustica 26, 173-199 (1972). /27/ Terhardt, E., J.Acoust.Soc.Am.55, 1061-1069 (1974). /28/ Terhardt, E., Acustica 30, 201-213 (1974). /29/ Terhardt, E., Acustica 36, 121-137 (1976). /30/ Terhardt, E., in "Psychophysics and Physiology of Hearing" (E.F.Evans&J.P. Wilson, eds.) Academic Press, London, pp. 381-390 (1977). /31/ Terhardt, E., Hearing Research 1, 155-182 (1979). /32/ Terhardt, E., in "Hearing Mechanisms and Speech" (O.Creutzfeld, H.Scheich, C.Schreiner, eds.), Springer, Berlin, pp. 281-291 (1979). /33/ Terhardt, E., in "Fortschritte der Akustik", VDE-Verlag, Berlin, pp. 667-670 (1980). /34/ Terhardt, E. und Schütte, H., Acustica 35, 122-126 (1976) /35/ Terhardt, E. und Stoll, G., in "Fortschritte der Akustik", VDE-Verlag, Berlin pp. 583-586 (1978). /36/ Terhardt, E. und Stoll, G., Acustica (1981), im Druck. /37/ Terhardt, E., Stoll, G. und Seewann, M., "Quantitative evaluation of pitch and pitch salience for complex tonal signals", zur Veröfö. eingereicht. /38/ Thurlow, W.R. und Bernstein, S., J.Acoust.Soc.Am. 29, 515 (1957). /39/ Vogel, A., Acustica 32, 300-306 (1975). /40/ Vogel, A., Biol. Cybernetics 18, 31-40 (1975). /41/ Zwicker, E., Acustica 8, 237-258 (1958). /42/ Zwicker, E., Acustica 10, 304-308 (1960). /43/ Zwicker, E., Acustica 10, 185 (1960). /44/ Zwicker, E., in "Frequency Analysis and Periodicity Detection in Hearing" (R.Plomp&G.F.Smoorenburg, eds.), Sijthoff, Leiden (1970), pp. 376-394. /45/ Zwicker, E., Acustica 22, 214-218 (1970). /46/ Zwicker, E., Acustica 34, 189-199 (1976). /47/ Zwicker, E., J. Acoust.Soc.Am. 62, 675-682 (1977). /48/ Zwicker, E., in "Kybernetik 1977" (G. Hauske&E. Butenandt, Hrsg.), Oldenbourg, München (1978), pp. 248-262. /49/ Zwicker, E., Hearing Research 1, 283-292 (1979). /50/ Zwicker, E., Biol. Cybernetics 35, 243-250 (1979). /51/ Zwicker, E., in "Fortschritte der Akustik", VDE-Verlag, Berlin (1980), pp. 159-162. /52/ Zwicker, E. und Fastl, H., Acustica 26, 78-82 (1972). /53/ Zwicker, E. und Feldtkeller, R., "Das Ohr als Nachrichteneempfänger", Hirzel, Stuttgart (1967). /54/ Zwicker, E. und Schütte, H., Acustica 29, 343-347 (1973). /55/ Zwicker, E., Terhardt, E. und Paulus, E., J.Acoust.Soc.Am. 65, 487-498 (1979).