

# 1 Einleitung

Die Bewertung der subjektiven Qualität von audiosignalverarbeitenden Systemen ist seit den Anfängen der Entwicklung solcher Systeme von großem Interesse. In den meisten Fällen wird ein System aber nicht über seine subjektiven Eigenschaften charakterisiert, sondern es werden technische Größen wie der Klirrfaktor und die Abweichung von einem idealen Frequenzgang angegeben, da diese Größen relativ einfach zu bestimmen sind. Erst mit der Entwicklung von verlustbehafteten Audiokomprimierungsverfahren entstand die Notwendigkeit, die subjektive Qualität mit Hilfe von Messverfahren bestimmen zu können, um aufwändige Hörtests in der Entwicklungsphase zu vermeiden. Bei den Komprimierungsverfahren lässt sich die Qualität nicht mit klassischen Methoden, wie der Bestimmung des Frequenzgangs und des Klirrfaktors, angeben, da es sich um in höchstem Maße nichtlineare und zeitvariante Systeme handelt. In den letzten Jahren wurden objektive Messverfahren auch für die Anwendung mit herkömmlichen Audiosystemen wie Verstärkern entwickelt, um die subjektive Störwirkung von nichtlinearen Verzerrungen zu beschreiben.

Dies hat seine Ursache darin, dass in der Audiotechnik häufig verwendete Größen zur Quantifizierung der nichtlinearen Verzerrungen von Audiosystemen nicht dazu geeignet sind, die tatsächliche Wahrnehmbarkeit und Störwirkung der Verzerrungsprodukte zu beschreiben. Die bestehenden Größen sind zudem auf die Messung der Eigenschaften analoger Systeme zugeschnitten, deren Nichtlinearitäten reguläre Verzerrungen erzeugen. Bei digitalen Systemen entstehen oft nichtreguläre Verzerrungen und somit ist die Angabe z.B. des Klirrfaktors für solche Systeme nicht angemessen.

Die überwiegende Anzahl der Arbeiten auf dem Gebiet der perceptuellen Audioqualitätsbeurteilung beschäftigt sich aus den genannten Gründen mit der Bewertung der Qualität von Komprimierungsverfahren. Daneben existieren einige Arbeiten, die Klirverzerrungen mit Hilfe von psychoakustischen Modellen bewerten. Das älteste Verfahren zur Bewertung der Hörbarkeit des Klirrfaktors von Audiosystemen ist *Auditory Spectral Distance* [39], neuere Verfahren sind *Distortion Score* [68] und  $R_{\text{nonlin}}$  [12]. Eine weitere Größe zur Beschreibung der Wahrnehmbarkeit von nichtlinearen Verzerrungen ist die *GedLee metric* [22]. Keines dieser Verfahren berücksichtigt jedoch das Vorhandensein linearer Verzerrungen.

Aus diesen Gründen wurde ein neues Messverfahren entwickelt, das die Hörbarkeit nichtlinearer Verzerrungen bestimmt und zusätzlich deren Auswirkungen auf psychoakustische Empfindungsgrößen berechnet. Im Rahmen dieser Dissertation werden schwach nichtlineare, zeitinvariante Audiosysteme wie Verstärker betrachtet. Bei diesen Systemen ist die Angabe eines Frequenzgangs und somit die Unterscheidung zwischen linearen und nichtlinearen Verzerrungen im Gegensatz zu Audiocodecs möglich. Das Messverfahren ist in zwei Teile gegliedert. Im ersten Schritt wird mit einer effizienten Messung der linearen

## 1 Einleitung

re Frequenzgang des untersuchten Systems bestimmt. Dazu wird ein iteratives Verfahren mit binären Pseudozufallsfolgen als Testsignalen verwendet. Der Vorteil dieses Verfahrens im Vergleich zu anderen Verfahren zur Frequenzgangmessung (z.B. Sweeps) liegt darin, dass der Frequenzgang auf einem DFT-Raster gewonnen wird und somit einfach ein lineares Ersatzsystem in Form eines FIR-Filters angegeben werden kann. Gegenüber einer Messung mit Einzeltönen zeichnet sich das Verfahren durch eine hohe Effizienz aus. Im zweiten Schritt des Messverfahrens können beliebige Signale, also auch Musiksignale, als Testsignal benutzt werden. Mit Hilfe des im ersten Schritt bestimmten linearen Ersatzsystems lassen sich die durch nichtlineare Verzerrungen entstandenen Störprodukte separieren und somit psychoakustisch bewerten.

Das vorliegende Messverfahren unterscheidet sich grundsätzlich von anderen Verfahren darin, dass keine Metrik in Form eines Einzelmaßes zur Bewertung der Qualität bestimmt wird. Neben der Bestimmung einer Schwelle für die Hörbarkeit unterschiedlicher Arten nichtlinearer Verzerrungen werden die Auswirkungen der Verzerrungen auf die psychoakustischen Empfindungsgrößen Lautheit, Schärfe, Rauigkeit und Klanghaftigkeit berechnet. Dies liefert eine bessere Einsicht in die Auswirkungen der Nichtlinearitäten, als dies ein Einzelmaß zu tun vermag. Mit dieser Betrachtung lassen sich in gewissem Umfang die unterschiedlichen klanglichen Auswirkungen unterschiedlicher Verzerrungstypen beschreiben.

Die berechneten Empfindungsgrößen werden mit Hilfe eines filterbankbasierten psychoakustischen Modells bestimmt. Zusätzlich wird mit dem gleichen psychoakustischen Modell die Hörbarkeit der Verzerrungen untersucht. Die Verwendung eines einheitlichen Modells für beide Berechnungen ist wichtig, um Widersprüche im Messverfahren zu vermeiden. So dürfen z.B. Störungen, die nicht wahrnehmbar sind, sich nicht als hörbare Änderungen der Empfindungsgrößen auswirken.

In Kapitel 2 werden die Grundlagen der Psychoakustik vorgestellt, auf denen die perzeptuelle Bewertung der Verzerrungen basiert. Sowohl allgemeine Grundlagen, als auch eine Übersicht über die Empfindungsgrößen und die Arbeitsweise perzeptueller Messverfahren werden dargestellt. In Kapitel 3 wird die iterative, auf Pseudozufallsfolgen basierende Methode zur Messung des Frequenzgangs beschrieben, die die erste Stufe des Messverfahrens bildet. Anhand von Messungen an Verzerrerschaltungen mit unterschiedlichen Verzerrungscharakteristiken wird die Methode mit bekannten Verfahren verglichen und auch ihre Eignung zur Messung nichtlinearer Verzerrungen gezeigt.

In Kapitel 4 wird die zweite Stufe des Messverfahrens, die Berechnung der psychoakustischen Empfindungsgrößen, im Detail beschrieben. Es werden einige Beispiele gegeben, die die Eignung des Verfahrens für die Bewertung der Hörbarkeit von Verzerrungen demonstrieren. Im 5. Kapitel werden schließlich die Ergebnisse von Messungen an verschiedenen nichtlinearen Systemen dargestellt. Dabei werden die Ergebnisse von Berechnungen unter MATLAB und von Messungen an Verzerrerschaltungen betrachtet.

# 2 Psychoakustische Grundlagen

## 2.1 Grundlegende Höreigenschaften

Bevor auf die psychoakustischen Eigenschaften des Gehörs eingegangen wird, soll zunächst der Aufbau und die Funktionsweise des Ohrs dargestellt werden.

### 2.1.1 Aufbau des Ohrs

Der Aufbau des menschlichen Ohrs ist in Abbildung 2.1 dargestellt. Es lässt sich in drei Teilbereiche untergliedern. Das Außenohr besteht aus der Ohrmuschel (1) mit dem Ohr läppchen (3) und dem Gehörgang (2). Die Ohrmuschel dient der Bündelung des Schalls und spielt durch ihre unsymmetrische Form auch beim Richtungshören eine wichtige Rolle. Der Gehörgang hat einen ausgeprägten Frequenzgang. Physikalisch betrachtet handelt es sich um ein an einer Seite offenes Rohr mit einer Länge von etwa 2 cm. Ein solches Gebilde hat eine Resonanzfrequenz bei  $L = \lambda/4$ , was etwa 4 kHz entspricht. Dies führt zu der hohen Empfindlichkeit des Gehörs bei dieser Frequenz.

Das Mittelohr dient der Impedanzanpassung zwischen Außenohr mit Luft als schallübertragendem Medium und dem flüssigkeitsgefüllten Innenohr. Es besteht aus dem Trommelfell (9), einer dünnen Membran, die den Abschluss des Gehörgangs bildet. Am Trommelfell ist der Hammer (8), einer der drei Gehörknöchelchen, befestigt. Die Gehörknöchelchen Hammer, Amboss (10) und Steigbügel (11) übertragen den Schall vom Trommelfell zum ovalen Fenster, dem Eingang des Innenohres mit einem Übersetzungsverhältnis von 2:1. Diese Übersetzung ist zusammen mit dem Verhältnis der Flächen von Trommelfell und der Fläche des ovalen Fensters von 15:1 für die Impedanzanpassung verantwortlich [84].

Das Innenohr setzt sich aus *Cochlea* (6) und dem Gleichgewichtsorgan (Vestibulärorgan (4)) zusammen, die weitgehend unabhängig voneinander funktionieren.

Die schneckenförmige *Cochlea* ist das Hörorgan des Innenohrs. Sie ist mit Lymphe gefüllt und in den sehr harten Schädelknochen eingebettet. Der Steigbügel überträgt durch das ovale Fenster die Schallschwingungen auf die inkompressible Lymphflüssigkeit. Ein Druckausgleich ist aufgrund der harten Wände nur über das runde Fenster, einer Membran ähnlich dem ovalen Fenster, möglich. Zwischen dem runden und dem ovalen Fenster sitzt die Basilarmembran mit dem Cortischen Organ in der Lymphflüssigkeit, auf dem sich die für das Hören zuständigen Sinneszellen befinden. Die Basilarmembran teilt den Schnecken gang mit zweieinhalb Windungen auf ihrer gesamten Länge von etwa 32 mm in zwei Bereiche, die *Scala Tympani* und die *Scala Vestibuli*. An der Spitze der Schnecke sind die beiden Bereiche über das *Helicotrema* direkt miteinander verbunden. Diese Ver-

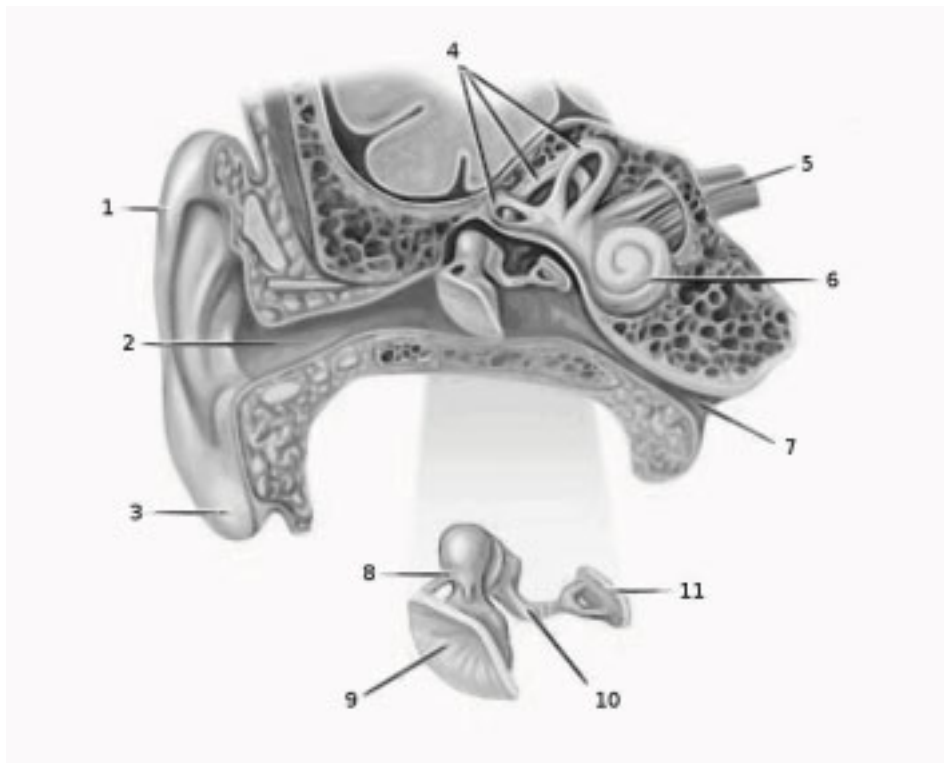


Abbildung 2.1: Aufbau des menschlichen Gehörs aus [1].

bindung dient dem Druckausgleich bei sehr niedrigen Frequenzen, die z.B. bei vertikalen Bewegungen des Kopfes durch Druckunterschiede auftreten.

Der Schneckengang verjüngt sich zur Schneckenspitze hin von etwa 3 mm Durchmesser an der Basis auf 1,5 mm Durchmesser an der Spitze. Die Basilarmembran verjüngt sich ebenfalls, allerdings in umgekehrter Richtung. Sie hat eine Breite von etwa 0,2 mm an der Basis und 0,6 mm an der Spitze der Schnecke.

Die Basilarmembran führt eine Frequenz/Ortstransformation durch, die es dem Gehör ermöglicht, verschiedene Frequenzen zu unterscheiden. Die Anregung der Lympfflüssigkeit durch das ovale Fenster hat eine Wanderwelle über die Basilarmembran zur Folge. Bei verschiedenen Schallfrequenzen werden somit unterschiedliche Bereiche der Membran zu maximalen Auslenkungen angeregt. Dadurch werden im Cortischen Organ je nach Frequenz Haarzellen an unterschiedlichen Orten erregt. Dabei liegt die zeitliche Auflösung der Haarzellen nur in der Größenordnung von etwa einer Millisekunde. Das bedeutet, dass die grobe zeitliche Struktur des Schallsignals im Zeitbereich, die Feinstruktur dagegen im Frequenz- bzw. Ortsbereich vom Gehör analysiert wird [72].

### 2.1.2 Frequenzskalen

Die Aufteilung der Frequenzbereiche auf der Cochlea folgt weder einem linearen noch einem logarithmischen Verlauf. Verschiedene Skalen für die Frequenzachse wurden in der Literatur vorgeschlagen, etwa ERB (*Equivalent Rectangular Bandwidth*), Spinc (*Spectral*

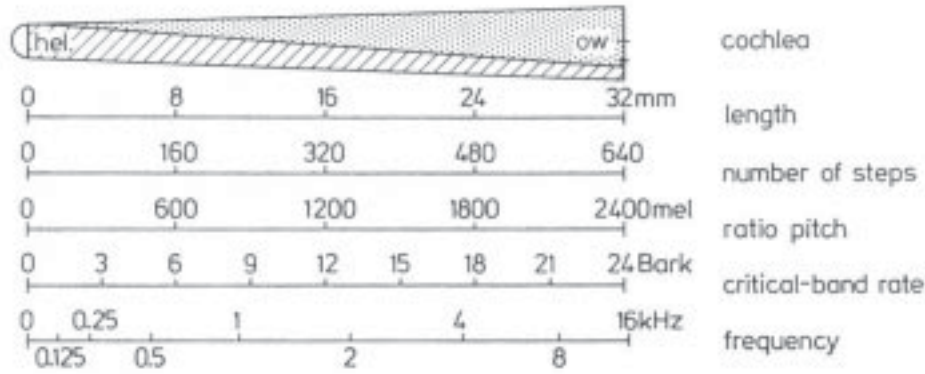


Abbildung 2.2: Cochlea mit Frequenzskalaren. Entnommen aus [84].

*Increment*) und Bark (benannt nach Heinrich Barkhausen), von denen sich die Barkskala (Tonheitsskala, critical band rate) in der perzeptuellen Messtechnik am weitesten verbreitet hat [75]. Abbildung 2.2 zeigt die Lage der Barkskala auf der Cochlea [84].

Die Definition Barkskala beruht auf der Tatsache, dass das Gehör bestimmte Frequenzbereiche als Frequenzbänder (Barkbänder, critical bands) zusammenfasst. Sie teilt den Frequenzbereich von 0-16 kHz in 24 Bänder ein. Unterhalb von 500 Hz beträgt die Bandbreite konstant 100 Hz. Bei höheren Frequenzen ist die Bandbreite eines Barkbands etwa 20 % der Mittenfrequenz.

Die Bandbreiten der kritischen Bänder lassen sich zum Beispiel mit Lautheitsmessungen von Bandpassrauschen mit variabler Bandbreite aber konstantem Schalldruck bestimmen. Beim Erhöhen der Bandbreite des Rauschsignals bleibt die empfundene Lautheit zunächst konstant. Überschreitet die Bandbreite des Rauschens einen bestimmten Wert, so nimmt die Lautheit des Bandpassrauschens trotz gleichbleibenden Schalldruckpegels zu. Die Bandbreite, bei der dieser Effekt auftritt, entspricht genau einer kritischen Bandbreite.

Bei der Aufteilung des Hörfrequenzbereichs in Barkbänder ist zu beachten, dass die Bänder nicht an feste Grenzfrequenzen gebunden sind, sondern beliebig verschoben werden können. Das oben beschriebene Experiment mit Bandpassrauschen zur Bestimmung der kritischen Bandbreiten kann also bei jeder beliebigen Mittenfrequenz des Rauschens durchgeführt werden.

Zwicker [84] gibt als Approximation für die Tonheit in Abhängigkeit von der Frequenz

$$\frac{z}{\text{Bark}} = 13 \arctan \left( 0,76 \frac{f}{\text{kHz}} \right) + 3,5 \arctan \left( \frac{f}{7,5 \text{ kHz}} \right)^2 \quad (2.1)$$

an. Die Breite eines Barkbandes lässt sich näherungsweise mit

$$\frac{\Delta f_G}{\text{Hz}} = 25 + 75 \left[ 1 + 1,4 \left( \frac{f}{\text{kHz}} \right)^2 \right]^{0,69} \quad (2.2)$$

berechnen. Eine einfacher zu invertierende Formel zur Berechnung der Tonheit stammt von Traunmüller [76]

$$\frac{z}{\text{Bark}} = \frac{26,81f/\text{Hz}}{1960 + f/\text{Hz}} - 0,53. \quad (2.3)$$

Die inverse Funktion dazu lautet:

$$\frac{f}{\text{Hz}} = 1960 \frac{z/\text{Bark} + 0,53}{26,28 - z/\text{Bark}}. \quad (2.4)$$

Eine weitere, bei gehörangepassten Qualitätsbewertungsverfahren oft anzutreffende Formel stammt von Schröder [61]:

$$\frac{z}{\text{Bark}} = 7 \operatorname{asinh} \left( \frac{f}{650 \text{ Hz}} \right). \quad (2.5)$$

Diese Annäherung lässt sich ebenfalls leicht invertieren :

$$\frac{f}{\text{Hz}} = 650 \sinh \left( \frac{z}{7 \text{ Bark}} \right). \quad (2.6)$$

Die Approximation von Schröder liefert nur für den Sprachfrequenzbereich gute Werte für die Tonheit, oberhalb von 4 kHz ist die Abweichung zur Barkskala relativ groß und erreicht ca. 2 Bark bei 10 kHz.

Die Pegel in den Barkbändern über der Tonheit bezeichnet man als Erregungspegel. Sie bildet die Grundlage zur Berechnung fast aller gehörbezogenen Empfindungsgrößen. Teilweise werden auch die Zeitsignale in einem Band zur Berechnung weiterer Größen benötigt.

### 2.1.3 Hörfläche

Die beschriebene Anatomie des Gehörs bedingt einen frequenzabhängigen Verlauf der Ohrempfindlichkeit, die sich in der Ruhehörschwelle ausdrückt. Die Ruhehörschwelle gibt den minimalen Schalldruck an, der notwendig ist, um einzelne Dauertöne mit einer Dauer von  $t > 200$  ms einer bestimmten Frequenz zu hören. Der Verlauf der Ruhehörschwelle ist bei normal hörenden Menschen individuell unterschiedlich [85]. Die Ruhehörschwelle bei 1 kHz wird auf einen Pegel von 0 dB festgelegt.

Zu tiefen Frequenzen hin nimmt die Empfindlichkeit auf Grund der Kopplung von *Scala Tympani* und *Scala Vestibuli* über das *Helicotrema* stark ab. Bei einer Frequenz von 50 Hz liegt die Ruhehörschwelle bei etwa 40 dB. Außerdem erzeugen der Blutkreislauf und Muskelbewegungen ein tieffrequentes Ohrinnengeräusch, das leise Signale verdeckt.

Bei etwa 4 kHz ist unser Gehör am empfindlichsten, was auf die Gehörgangresonanz zurückzuführen ist. Die Ruhehörschwelle liegt dort bei etwa -4 dB.

Oberhalb von 12 kHz nimmt die Empfindlichkeit wegen der Trägheit der Gehörknöchelchen wieder ab. Für Frequenzen größer 18 kHz sind unabhängig vom Schalldruck keine Töne mehr hörbar. Diese obere Grenzfrequenz des Gehörs gilt für Menschen im Alter von etwa 20-25 Jahren und nimmt mit zunehmendem Alter stark ab.

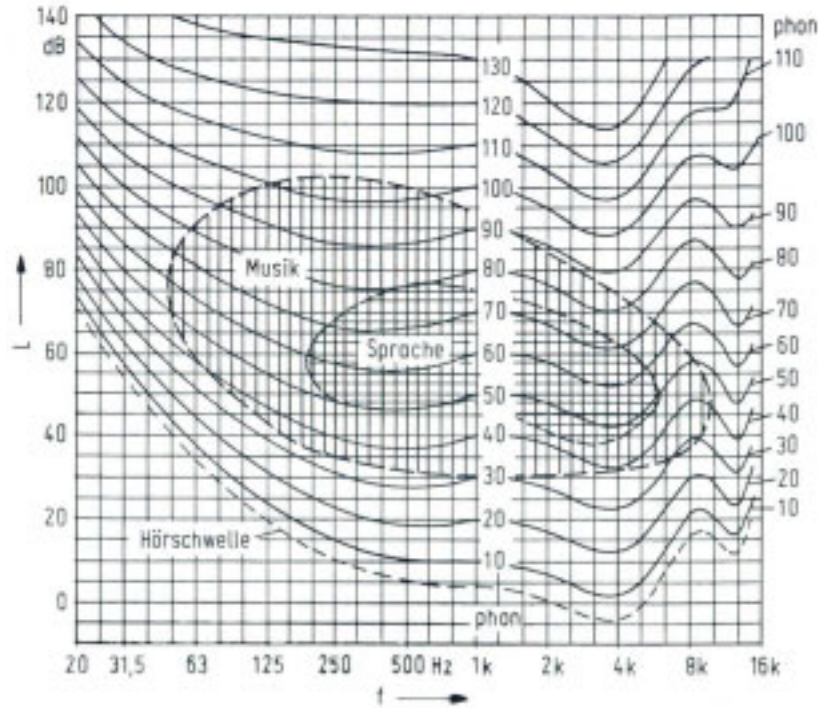


Abbildung 2.3: Hörfläche mit Kurven gleicher Lautheit. Entnommen aus [21].

Die mittlere Ruhehörschwelle normal hörender Erwachsener kann durch

$$\frac{L_A}{\text{dB}} = 3,64 \left( \frac{f}{\text{kHz}} \right)^{-0,8} - 6,5e^{-0,6\left(\frac{f}{\text{kHz}} - 3,3\right)^2} + 10^{-3} \left( \frac{f}{\text{kHz}} \right)^4 \quad (2.7)$$

angenähert werden [71].

Für Tonimpulse mit einer Dauer kleiner 200 ms steigt die Ruhehörschwelle an. Für Impulszeiten kleiner 100 ms wächst der benötigte Schalldruckpegel um etwa 10 dB, wenn die Impulsdauer um den Faktor 10 verringert wird. Daraus kann man auf die Integrationsdauer des Gehörs von etwa 100 ms schließen [84].

Die Hörfläche erlaubt die Darstellung aller hörbaren Schallereignisse in einem Pegel-/Frequenzdiagramm, das in Abbildung 2.3 dargestellt ist. Die Ruhehörschwelle bildet die untere Grenze der Hörfläche, nach oben wird sie durch die Schmerzgrenze bei einem Schalldruckpegel von ca. 130 dB beschränkt.

Innerhalb der Hörfläche sind die Bereiche eingezeichnet, in denen sich Sprache bzw. Musik abspielen. Musikalische Darbietungen finden im Frequenzbereich zwischen 40 Hz und 10 kHz statt und weisen Pegel zwischen etwa 20 dB und 95 dB auf. Sprache nimmt noch einen kleineren Teil der Hörfläche ein. In der Hörfläche befinden sich Sprachsignale im Bereich zwischen 100 Hz und 7 kHz mit Pegeln zwischen etwa 35 dB und 70 dB.

### 2.1.4 Kurven gleicher Lautheit

Die wahrgenommene Lautheit eines Tones ist abhängig vom Pegel des Tons und von seiner Frequenz. Der Frequenzabhängigkeit wird man durch Einführung der Phon-Skala gerecht, die sich auf den Pegel eines Tones bei 1 kHz im Freifeld bezieht. Die Lautheit eines Tones beliebiger Frequenz, ausgedrückt in Phon, entspricht der Lautheit eines 1 kHz Tones mit entsprechendem Schalldruckpegel. Zeichnet man die Kurven für konstante Phon-Werte über der Frequenz, erhält man die Kurven gleicher Lautheit. Die Kurve für 3 Phon entspricht dabei der Ruhehörschwelle.

Für kleine Pegel verlaufen die Kurven fast parallel zur Ruhehörschwelle, bei zunehmendem Pegel werden sie unterhalb von 200 Hz immer flacher.