

Carl von Ossietzky-Universität Oldenburg

Studiengang Diplom-Physik

Diplomarbeit

**Psychoakustische Modellierung
der virtuellen Tonhöhe**

vorgelegt von: Matthias Vormann

Betreuender Gutachter: Dr. Reinhard Weber

Zweiter Gutachter: Prof. Dr. Volker Mellert

Oldenburg, 25. September 1995

1 EINLEITUNG	1
1.1 Zielsetzung dieser Arbeit	2
1.2 Grundzüge des Tonhöhenberechnungsmodells	3
2 VIRTUELLE TONHÖHE.....	5
2.1 Tonhöhenempfindung	5
2.2 Parameter der Tonhöhe für reine Töne.....	7
2.3 Tonhöhe von Klängen	9
2.4 Virtuelle Tonhöhe.....	11
3 FOURIER-ZEIT-TRANSFORMATION.....	14
3.1 Definition der Fourier-t-Transformation (FTT)	14
3.2 FTT als gehörgerechte Audiotransformation	15
3.3 Bedeutung der Fensterfunktion für die FTT	16
3.4 Zeit- und Frequenzauflösung der FTT	17
3.5 Swing	19
3.6 Glättung des FTT-Spektrums	20
3.7 Diskrete FTT	24
3.8 Inverse FTT.....	25
3.9 Beispiel zur FTT	25
3.10 Bilanzierende Bewertung der FTT	28
4 TEILTONZEITMUSTER (TTZM).....	29
4.1 Bildung des TTZM.....	30
4.1.1 Berechnung der Kurzzeitspektren	30
4.1.2 Berechnung der Leistungsspektren.....	31
4.1.3 Glättung der Leistungsspektren	31
4.1.4 Berücksichtigung des zeitabhängigen Auswertintervalls	31
4.1.5 Erkennung der Teiltöne	32
4.2 Vernetzung des TTZM zum Teiltonlinienzeitmuster	35
4.2.1 Numerische Beschreibung des Vernetzungsalgorithmus.....	39
4.2.2 Beispiel für ein TTLZM	41

4.3 Überprüfung des Konzepts des TTLZM	42
4.3.1 Resynthese	42
4.3.2 Berücksichtigung der Fensterfunktionen der FFT bei der Re- synthese	44
4.3.3 Hörversuche.....	45
5 MODELL DER VIRTUELLEN TONHÖHE.....	46
5.1 Modell nach Terhardt et al. 1982	46
5.2 Modifiziertes Modell	49
5.2.1 Vorverarbeitung	49
5.2.2 Überschußpegel	49
5.2.3 Tonhöhenverschiebung	52
5.2.4 Gewichtung der spektralen Komponenten	53
5.2.5 Virtuelle Tonhöhen-Extraktor (VTE)	55
5.2.6 Vergleich von Spektral- und virtuelle Tonhöhenmuster	60
5.2.7 Implementierter Tonhöhenberechnungsalgorithmus	61
5.3 Zeitkontinuierliche Verarbeitung	62
6 PROGRAMMODULE, WÄHLBARE PARAMETER UND PROGRAMMTESTS	69
6.1 Die einzelnen Programmodule.....	70
6.2 Parameter der Modelle	71
6.3 Test der Modelle.....	73
6.3.1 Stationäre Testsignale.....	73
6.3.2 Nichtstationäre Testsignale	80
7 ANWENDUNG DES MODELLS AUF UMWELTGE- RÄUSCHE	82
8 ZUSAMMENFASSUNG UND AUSBLICK	97

Anhänge:

Anhang A	FOURIER-TRANSFORMATION.....	100
Anhang B	EIGENSCHAFTEN DER FOURIER-TRANSFORMATION	106
Anhang C	EFFEKTIVE FENSTERLÄNGE UND BANDBREITE DER FTT	107
Anhang D	PROGRAMME ZUR BERECHNUNG DER VIRTUELLEN TONHÖHE	109
Anhang E	HILFSPROGRAMME.....	116
Anhang F	LITERATURVERZEICHNIS	118
Anhang G	DANKSAGUNG UND ERKLÄRUNG.....	123

1 EINLEITUNG

Das Ziel der vorliegenden Arbeit ist die Entwicklung eines Modells zur Vorhersage der virtuellen und spektralen Tonhöhe von Umweltschallen.

Die Hörempfindung ‘virtuelle Tonhöhe’ tritt auf, wenn aus harmonischen Klängen die kleinsten Harmonischen samt Grundfrequenz entfernt werden. Es entstehen so Residualklänge, die die gleiche Hörempfindung hervorrufen wie die zugehörenden, vollständigen, harmonischen Klänge.

Die Wahrnehmung von virtueller Tonhöhe erfolgt in vielen Bereichen der alltäglichen Schallwahrnehmung, ohne daß sich der Wahrnehmende dieses immer bewußt ist. So werden z.B. beim Telefonieren aufgrund der Übertragungseigenschaften des Telefonsystems nur Frequenzen oberhalb von ca. 300 Hz übertragen [41], [56]. Für eine typische männliche Stimme mit einer Grundfrequenz von etwa 100 Hz bedeutet dies, daß die untersten drei Spektrallinien durch Hochpaßfilterung bei der Übertragung aus dem Spektrum entfernt werden. Trotzdem ruft das gefilterte Sprachsignal die gleiche Tonhöhe hervor wie ein Sinuston von etwa 100 Hz (Residuum, vgl. z.B. [58]).

Zur Erklärung nimmt man an, daß im menschlichen Gehör die Tonhöhe des Sprachsignals durch Auffinden der Grundfrequenz der im Spektrum vorhandenen Teillinien gebildet wird. Da Frequenzanteile, die eine spektrale Tonhöhe von etwa 100 Hz hervorrufen, nicht im Spektrum vorhanden sind, wird diese Tonhöhenwahrnehmung als ‘virtuelle Tonhöhe’ bezeichnet [24], [41], [42], [44], [49], [50], [56], [58]. Allgemein entstehen virtuelle Tonhöhen, wenn von einem breitbandigen Linienspektrum nur ein höherfrequenter Teil hörbar ist [56].

Messungen haben gezeigt, daß die virtuelle Tonhöhe eine komplexe Empfindung ist, die das Gehör aus den spektralen Tonhöhen ableitet [56]. Die spektralen Tonhöhen hingegen sind psychoakustische Empfindungsgrößen, die durch einzelne Spektrallinien des Spektrums erzeugt werden, vgl. Kap. 2 und 3.9.

Weitere Beispiele für die alltägliche Wahrnehmung virtueller Tonhöhen sind neben Sprache [41], [49], [50], [56] u.a. die Schalle von Musikinstrumenten wie kleine Streicher (Violine, Viola) [8], [24], Glockenklängen [34], [49], [50] und musikalischen Akkorden [49], [50].

Virtuelle Tonhöhen lassen sich nicht ohne weiteres aus dem Spektrum eines Schalles voraussagen [44]. Um virtuelle Tonhöhen zu prognostizieren, müssen eine Reihe psychoakustischer Zusammenhänge berücksichtigt werden, die Ein-

fluß auf die Tonhöhenempfindung haben. In den letzten Jahrzehnten wurden unterschiedliche Vorschläge gemacht, die virtuelle Tonhöhe zu berechnen. Diese Vorschläge beruhen auf verschiedenen psychoakustischen Annahmen, bei denen jedoch grundsätzlich nur zwei Ansätze zu unterscheiden sind: Die Ortstheorie, bei der die Verteilung der Anregungen auf der Basilarmembran betrachtet wird und die Zeittheorie, der die neuronalen Aktivitäten der Hörnerven als Grundlage dienen [22]. Zur ersten Gruppe gehören zum Beispiel die Arbeiten von Goldstein 1973, Wightman 1973 [54], Terhardt 1974 [44] und Terhardt et al. 1982 [50]. Diese Modelle gehen im wesentlichen von den Orten der Anregung auf der Basilarmembran aus, d.h. sie benutzen die Frequenz- Ortstransformation, um auf verschiedene Weise einen Mustererkennungsprozeß, der aus den physikalisch vorhandenen Komponenten eines Signals die virtuelle Tonhöhe berechnet, anzuwenden. Die zweite Gruppe von Modellen arbeitet im Zeitbereich. Das wesentliche Merkmal dieser Modelle ist die Auswertung der zeitlichen Informationen der Spikes, d.h. der neuronalen Aktivitäten der Hörnerven mit Autokorrelationsfunktionen. Nach diesem Prinzip arbeiten z.B. die Modelle von Licklider 1951, Broadbent 1975, Moore 1982, van Noorden 1982, Loeb et al. 1983, Lyon 1984, de Cheveigné 1986, Patterson 1987, Lazzaro und Mead 1989, Slaney und Lyon 1990, sowie Meddis und Hewitt 1991 [22], [23].

Ein Hybridmodell aus beiden Gruppen wurde von Goldstein und Sruлович 1977, bzw. Sruлович und Goldstein 1983 vorgestellt [22].

1.1 Zielsetzung dieser Arbeit

Ziel dieser Arbeit ist die Implementation eines Tonhöhenberechnungsmodells, das in der Lage sein soll, sowohl spektrale wie auch virtuelle Tonhöhen zu erkennen und nach ihrer Ausprägtheit zu bewerten. Diese Auswertung soll quasi-zeitkontinuierlich stattfinden, um die Tonhöhenverläufe über der Zeit wiederzugegeben.

Das Modell soll für eine möglichst große Gruppe von Umweltgeräuschen anwendbar sein und auf einer Workstation bzw. einem PC implementiert werden. Eine Spezialisierung auf bestimmte Geräuschklassen, wie z.B. auf Sprachsignale oder Motorengeräusche, soll in dieser Arbeit nicht erfolgen.

1.2 Grundzüge des Tonhöhenberechnungsmodells

Als Basis für den Tonhöhenberechnungsalgorithmus wird das Tonhöhenmodell zur Berechnung virtueller Tonhöhen von Terhardt, Stoll und Seewann [50] verwendet. In diesem Modell wird zur Berechnung des Spektrums der untersuchten Signale eine Fast-Fourier-Transformation benutzt. Terhardt et al. weisen jedoch darauf hin, daß diese Art der Transformation nicht als gehörgerecht bezeichnet werden kann [50]. Deshalb wird in der vorliegenden Arbeit die Fast-Fourier-Transformation durch die Fourier-Zeit-Transformation [46], [47] ersetzt, die wesentlich gehörgerechtere Eigenschaften besitzt. Zur Extraktion der tonalen Komponenten wird in dieser Arbeit das Teiltonzeitmuster von Heinbach [12], [13] verwendet. Durch den Einsatz des Teiltonzeitmusters kann eine allgemeinere Detektion der tonalen Komponenten erfolgen als bei Verwendung des von Terhardt et al. vorgeschlagenen Algorithmus, der als eine unter vielen anderen Möglichkeiten zur Bestimmung von tonalen Komponenten bezeichnet wird [50]. Eine Erweiterung des Konzepts der Teiltonzeitmuster zum Teiltonlinienzeitmuster nach Heldmann [14] wird in dem hier vorgestellten Tonhöhenberechnungsalgorithmus adaptiert, um die Verarbeitung an die Eigenschaften des menschlichen Gehöres anzugleichen.

Die eigentliche Berechnung der virtuellen und spektralen Tonhöhen wird in dieser Arbeit mit dem von Terhardt entwickelten ‘Virtuelle Tonhöhen-Extractor’ [44] durchgeführt.

Es wird überprüft, inwieweit der Algorithmus zur Berechnung virtueller Tonhöhen von Heldmann [14], der mit diesem Modell Elektromotorengeräusche untersucht, für beliebige Signale übernommen werden kann.

In Kapitel 2 werden die Grundlagen der Wahrnehmung virtueller und spektraler Tonhöhen erläutert. In Kapitel 3 wird die Fourier-Zeit-Transformation nach Terhardt [46], [47] vorgestellt. Es werden einige Modifikationen des Modells vorgenommen und die Implementation des modifizierten Fourier-Zeit-Transformationsalgorithmus beschrieben. In Kapitel 4 wird die Theorie und Programmierung des Teiltonzeitmusters [12], [13] und des Teiltonlinienzeitmusters [14] dargelegt. Kapitel 5 beschreibt die Tonhöhenwahrnehmung quantitativ und berücksichtigt ihre Umsetzung im vorliegenden Modell. Der ‘Virtuelle Tonhöhen-Extractor’ von Terhardt [44] wird diskutiert und die Programmierung dieses Algorithmus durchgeführt. In Kapitel 6 werden die im Rahmen dieser Arbeit erstellten Programme zur Berechnung virtueller und spektraler Tonhöhen vorgestellt, verifiziert und ihre Parameter erläutert. Kapitel 7 zeigt die Anwendung

der selbst erstellten Algorithmen auf Umweltgeräusche. In Kapitel 8 wird schließlich eine Zusammenfassung der vorliegenden Arbeit gegeben, sowie ein Ausblick für die weitere Forschung beschrieben.

2 VIRTUELLE TONHÖHE

Gegenstand dieser Arbeit ist die rechnergestützte Bestimmung der virtuellen Tonhöhe aus den Zeitsignalen beliebiger Umweltschalle. Daher soll kurz eingeführt werden, was unter ‘virtueller Tonhöhe’ zu verstehen ist, durch welche Schallgrößen sie bestimmt wird und wie sie experimentell ausgemessen werden kann.

Die Daten und Fakten dieses Kapitels stammen im wesentlichen aus [56], [57] und [58]. Weitere allgemeine Eigenschaften der virtuellen Tonhöhe fassen [43] und [45] zusammen. Die Ausprägtheit spektraler Tonhöhen untersuchen [9] und [18], die Ausprägtheit der virtuellen Tonhöhe wird in [10] analysiert.

2.1 Tonhöhenempfindung

Die psychoakustische Empfindungsgröße *Tonhöhe* beschreibt die durch akustische Reize beim Menschen hervorgerufenen Eindrücke in Bezug auf die spektrale Frequenzverteilung der Reize. Um die Tonhöhe als psychoakustische Empfindungsgröße gegenüber allem, was i.A. Sprachgebrauch mit ‘Tonhöhe’ beschrieben wird, zu unterscheiden, wird die Tonhöhe in der DIN 1320 [64] folgendermaßen definiert: ‘Zuordnung eines Tones oder Tongemisches zu einer bestimmten Skala zwischen tiefen und hohen Tönen’. In der DIN 1320 wird weiterhin angemerkt, daß der Begriff der Tonhöhe der Musik entstammt, jedoch allgemein ein psychologisches Merkmal des Hörereignisses charakterisiert. In der physikalischen Akustik wird die Tonhöhe durch die Frequenz gekennzeichnet.

Der Zusammenhang zwischen Tonhöhe und Frequenz eines reinen Tones wurde bereits in vielen Meßreihen mit einer großen Anzahl von Versuchspersonen ausgemessen. Bei diesen Meßreihen stellen Versuchspersonen eine veränderliche Reizgröße relativ zu einer anderen, fest vorgegebenen Reizgröße so ein, daß die damit verbundenen Empfindungsgrößen in einem bestimmten Verhältnis zueinander stehen. In diesem Fall bedeutet dies, daß die Versuchspersonen die Frequenz eines Testtones f_T so eingeregeln, daß sie die halbe, bzw. die doppelte Tonhöhe hervorruft wie ein Referenzton fester, vorgegebener Frequenz f_r . Die so gewonnene Tonhöhe wird als *Verhältnistonhöhe* H_v bezeichnet, ihre Einheit ist das *mel*.

Abb 2.1 zeigt die so bestimmten Versuchsergebnisse. Aufgetragen ist die Verhältnistönhöhe über der Referenzfrequenz f_r (doppeltlogarithmisch), gemittelt über viele Versuchspersonen und Einzelversuche. Unterhalb von 500 Hz ist die Verhältnistönhöhe (in mel) praktisch identisch der Frequenz eines reinen Tones (in Hz). Oberhalb von 500 Hz bleibt die Verhältnistönhöhe mit steigender Frequenz des Tones mehr und mehr hinter der Tonfrequenz zurück.

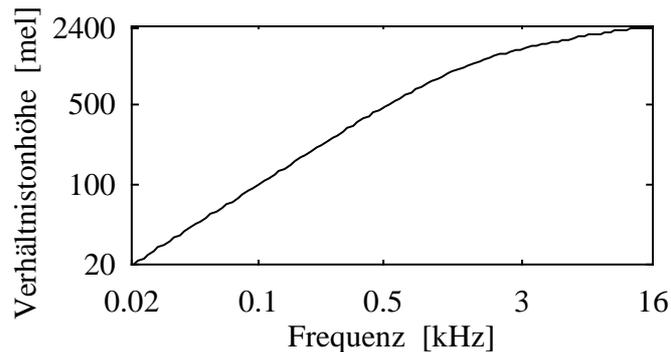


Abb. 2.1: Doppeltlogarithmische Darstellung der Verhältnistönhöhe über der Frequenz, Daten nach [57].

Die in Abb. 2.1 dargestellte Eigenschaft der Verhältnistönhöhe kann durch die Anatomie des menschlichen Ohres erklärt werden [56]. Dazu wird die natürliche Skala der Basilarmembran nach [56], Abb. 2.2, betrachtet. Die Frequenz- Orts-Transformation auf der Cochlea bedingt u.a. folgende Eigenschaft: die Verhältnistönhöhe in mel ist proportional zur Tonheit, d.h. proportional zur Frequenzgruppeneinteilung in Bark nach Zwicker (1 Bark = 100 mel). Aus Abb. 2.2 geht

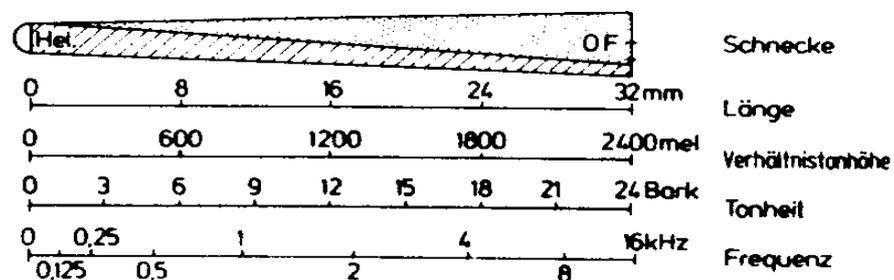


Abb. 2.2: Natürliche Skalen der Basilarmembran. Darstellung entnommen aus [56] (modifiziert), Erläuterungen siehe Text.

hervor, daß die Tonheit, und somit auch die Verhältnistönhöhe, linear zur Länge der Basilarmembran (gemessen entlang der Basilarmembran vom Helicotrema zum ovalen Fenster) zunimmt. Im Gegensatz dazu ist der Ort, an der ein Ton bestimmter Frequenz seine Haupterregung auf der Basilarmembran ausbildet, für Frequenzen unterhalb 500 Hz in etwa linear, für größere Frequenzen etwa logarithmisch. Da der Quotient aus Verhältnistönhöhe und dem Ort der Haupterregung einer Frequenz auf der Basilarmembran der gleiche sein muß wie der in Abb. 2.1 gemessene Quotient aus Verhältnistönhöhe und der dargebotenen Frequenz, kann die Kurvenform der Abb. 2.1 so veranschaulicht werden.

2.2 Parameter der Tonhöhe für reine Töne

Der wesentliche Parameter der Tonhöhenwahrnehmung für reine Töne ist die Frequenz. Für eine vollständige Erfassung der Tonhöhe müssen jedoch noch weitere Größen wie der Schalldruckpegel des Tones [19], [52] oder Maskierungseffekte durch andere Töne oder Rauschen [36], [52] berücksichtigt werden [20], [53].

Die Abhängigkeit der Tonhöhe vom Schalldruckpegel wird vermessen, indem ein Vergleichston regelbarer Frequenz f_r und eines Schalldruckpegels von 40 dB mit einem Testton vorgegebenen Schalldruckpegels und vorgegebener Frequenz f_T so abgeglichen wird, daß beide die gleiche Tonhöhe hervorrufen. Auf diese Weise wird die *Tonhöhenverschiebung* v ermittelt:

$$(2.1) \quad v = \frac{f_r - f_T}{f_T}.$$

Sie ist ein Maß für die prozentuale Verschiebung zwischen Vergleichs- und Testtonfrequenz, die dafür erforderlich ist, daß von beiden Tönen der gleiche Tonhöhen Eindruck hervorgerufen wird.

Abb. 2.3 zeigt die Tonhöhenverschiebung v in Abhängigkeit vom Schalldruckpegel für verschiedene Testtonfrequenzen f_T als Parameter.

Mit der Tonhöhenverschiebung v kann die Tonhöhe H dargestellt werden als:

$$(2.2) \quad H = \frac{f}{[Hz]} \cdot (1 + v) [pu]$$

Die Tonhöhe in Gl. (2.2) wird in Einheiten der äquivalenten Frequenz pu (engl. 'pitch units') gemessen. Das bedeutet, daß die Empfindungsgröße Tonhöhe eines Testtones den gleichen Zahlenwert bekommt wie die Frequenz des Referenztones f_r , der die gleiche Tonhöhe hervorruft. Durch diese Vereinbarung wird

die Empfindungsgröße Tonhöhe abgebildet auf die Reizgröße 'Frequenz des Referenztones'.

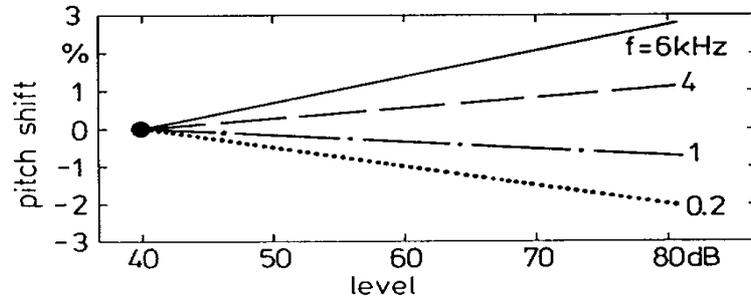


Abb. 2.3: Pegelabhängigkeit der Tonhöhe von reinen Tönen. Aufgetragen ist die Tonhöhenverschiebung v über dem Schalldruckpegel für verschiedene Testtonfrequenzen f_T als Parameter (hier mit f bezeichnet). Die Daten stammen aus [58]. Terhardt [44] schlägt für den Referenzton einen Schalldruckpegel von 60 dB vor. In beiden Fällen ist bemerkenswert, daß die Tonhöhe für hohe Frequenzen mit steigendem Schalldruckpegel ansteigt, während sie für tiefe Frequenzen sinkt.

Die Tonhöhe (reiner Töne) wird ebenfalls durch gegenseitige Maskierung, durch verdeckende Störgeräusche oder durch rauschhafte Anteile beeinflusst, siehe Abb 2.4:

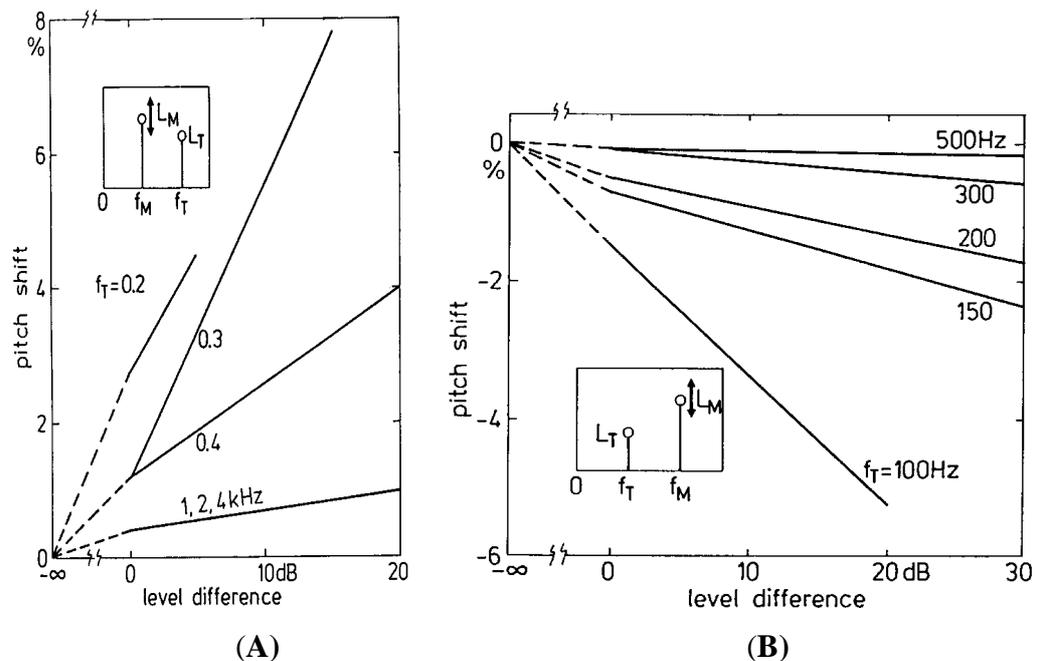


Abb. 2.4: Maskierungseinflüsse auf die Tonhöhe reiner Töne. Daten nach [58], Erläuterungen siehe Text.

Abb. 2.4 zeigt die Tonhöhenverschiebung ν eines Testtones mit Testtonfrequenz f_T als Funktion des Pegelunterschieds zwischen seinem (konstanten) Pegel L_T und einem maskierenden Ton der Frequenz f_M mit kontinuierlich veränderbarem Pegel L_M . Es sind verschiedene Messungen mit unterschiedlicher Testtonfrequenz f_T als Parameter dargestellt. In Abb. 2.4A wird die Frequenz des maskierenden Tones f_M jeweils eine Oktave tiefer gewählt als die Frequenz des Testtones.

Abb. 2.4B zeigt den gleichen Zusammenhang, jedoch wird f_M hier um eine Oktave höher dargeboten als f_T . Beide Diagramme zeigen, daß für kleine Frequenzen die Tonhöhenverschiebung einen größeren Betrag aufweist als für große Frequenzen. Dieses wird durch den Oktavabstand von Testton und Maskierer bedingt, der für kleinere Frequenzen absolut gesehen kleiner ist als für große Frequenzen.

Die Richtung der Tonhöhenverschiebung (also das Vorzeichen von ν) hängt im wesentlichen von der Orientierung des Maskierers zum Testton ab; im Spektralbereich betrachtet ‘schiebt’ der Maskierer den Testton von sich weg. Das bedeutet: für $f_T < f_M$ ergeben sich negative, für $f_T > f_M$ positive Tonhöhenverschiebungen.

Auch durch Schmalbandrauschen wird eine ausgeprägte Tonhöhe hervorgerufen. Nach [56] lassen sich die hier erläuterten Resultate grundsätzlich auf Schmalbandrauschen übertragen.

2.3 Tonhöhe von Klängen

Die Summe mehrerer reiner Töne wird als *Tonkomplex* oder *Klang* bezeichnet. Können die einzelnen Tonfrequenzen als ganzzahlige Vielfache einer gemeinsamen Grundschiwingung angegeben werden, so handelt es sich um einen *harmonischen Klang*.

Harmonische Klänge rufen in der Regel nur eine einzige Tonhöhe hervor, zumindest nur eine dominante Tonhöhe. Diese Tonhöhe liegt in der Größenordnung der Tonhöhe, die ein reiner Ton mit der Frequenz der Grundschiwingung des Klanges hervorruft. Eine qualitative Betrachtung zeigt Abb. 2.5:

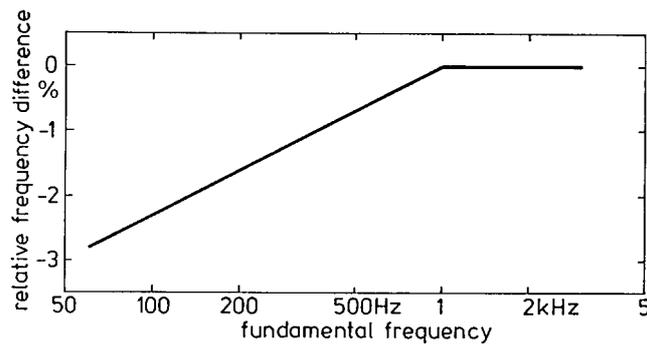


Abb. 2.5: *Tonhöhe von Klängen. Aufgetragen ist die Tonhöhenverschiebung v (hier als relative Frequenzdifferenz bezeichnet) gegen die Grundfrequenz des Klanges. Daten nach [58], weitere Erläuterungen siehe Text.*

In Abb. 2.5 ist die Tonhöhenverschiebung $v = (f_T - f_g)/f_g$ eines Klanges zwischen seiner Grundfrequenz f_g und einem Vergleichston der Frequenz f_T gegen die Grundfrequenz f_g des Klanges aufgetragen. Dabei wird die Vergleichstonfrequenz f_T so bestimmt, daß sie die gleiche Tonhöhe hervorruft wie der Klang. Die genaue Beschaffenheit des Klanges ist hier nicht weiter von Bedeutung (vgl. [58]), lediglich die Tendenz zu negativen Tonhöhenverschiebungen für $f_g < 1$ kHz ist interessant. Die höheren Harmonischen des Klanges können in diesem Fall als ‘Maskierer’ der Grundschwingung betrachtet werden. Wie im vorigen Abschnitt dargelegt, ‘schieben’ die Harmonischen die Grundfrequenz in Richtung kleinerer Frequenzen.

Auch bei Klängen ist die Tonhöhe pegelabhängig. Abb. 2.6 zeigt die Tonhöhenverschiebung eines harmonischen Klanges mit $f_g = 200$ Hz in Abhängigkeit vom dargebotenen Schalldruckpegel des Klanges. Der Pegel des Vergleichstones L_T wird dabei auf konstant 50 dB gehalten.

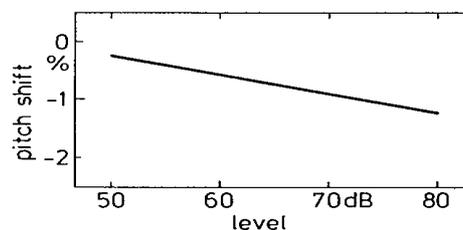


Abb. 2.6: *Pegelabhängigkeit der Tonhöhe von Klängen. Tonhöhenverschiebung v über dem Darbietungspegel für einen Klang mit der Grundfrequenz 200 Hz. Daten entnommen aus [58].*

Ein ähnliches Verhalten zeigt auch der reine Ton mit 200 Hz in Abb. 2.3. Daraus folgt, daß die Tonhöhe eines Klanges im wesentlichen von seinen niedrigsten Frequenzkomponenten, insbesondere also der Grundfrequenz, abhängt.

2.4 Virtuelle Tonhöhe

Werden die kleinsten Harmonischen samt Grundfrequenz aus einem harmonischen Klang entfernt, so erhält man einen sogenannten *Residualklang*. Dieser ruft im wesentlichen die gleiche Tonhöhe hervor wie der vollständige Klang. Die zugehörige Hörempfindung wird *virtuelle Tonhöhe* genannt.

Den Existenzbereich der virtuellen Tonhöhe zeigt Abb. 2.7. Aufgetragen ist die Grundfrequenz f_g (d.h. der Abstand der einzelnen Harmonischen) über der kleinsten, im Spektrum des Residualklanges physikalisch vorhandenen Frequenz f_k .

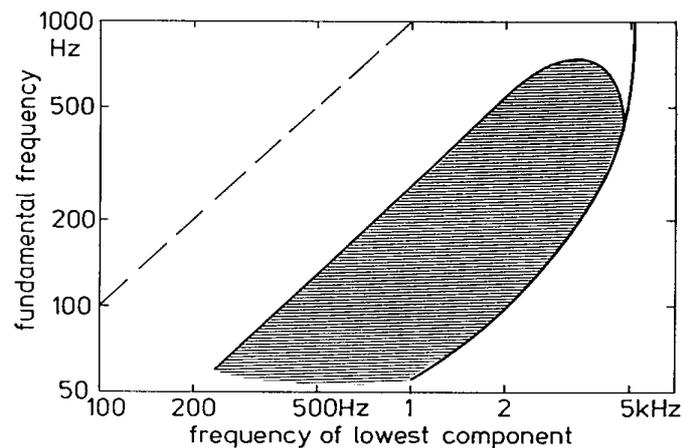


Abb. 2.7: Existenzbereich der virtuellen Tonhöhe. Aufgetragen ist die Grundfrequenz f_g gegen die kleinste im Spektrum vorhandene Frequenzkomponente f_k . Virtuelle Tonhöhen können nur durch Residualklänge hervorgerufen werden, deren Grundfrequenz f_g und kleinste im Spektrum vorhandene Frequenzkomponente f_k innerhalb des Bereiches liegen, der zur linken durch $f_g = f_k$ (gestrichelte Gerade) und zur rechten durch die fett ausgezogene Linie begrenzt wird. [58]

Dieser Bereich wird zur linken durch die Identität (gestrichelte Linie) von kleinster vorhandener Frequenzkomponente f_k und Grundfrequenz f_g des Residualklanges gegeben. Auf der rechten Seite wird der Existenzbereich durch die experimentell für Klänge mit einer großen Anzahl von Frequenzkomponenten ermittelte, ausgezogene Linie begrenzt. Der schraffierte Bereich markiert dasjenige Gebiet, in dem ein amplitudenmodulierter Ton mit einem Modulationsgrad

von eins eine virtuelle Tonhöhe hervorruft [58]. Aus Abb. 2.7 geht hervor, daß die kleinste vorhandene Frequenzkomponente f_k in jedem Fall kleiner als 5 kHz sein muß, damit der Residualklang noch eine virtuelle Tonhöhe hervorrufen kann. Je kleiner die Grundfrequenz des Residualklangs wird, desto kleiner muß die kleinste vorhandene Frequenz f_k sein, um den Existenzbereich der virtuellen Tonhöhe nicht zu verlassen.

Die maximale Anzahl der Spektrallinien eines Klanges, die zum virtuellen Tonhöhereindruck beiträgt, wird folgendermaßen abgeschätzt [58]: aus Abb. 2.7 ergibt sich, daß für eine Grundfrequenz von $f_g = 100$ Hz die kleinste Frequenzkomponente von $f_k = 2$ kHz nicht überschritten werden darf, um den Existenzbereich der virtuellen Tonhöhe nicht zu verlassen. Dann gilt: $f_k = 20 \cdot f_g$ und somit ist f_k die zwanzigste Harmonische des betrachteten Residualklanges. Der Wert von zwanzig wird für keine andere Grundfrequenz f_g überschritten. Mit zunehmender Grundfrequenz nimmt diese Zahl bis auf etwa fünf ab. Daraus folgt, daß die maximale Anzahl der Spektrallinien, die zum virtuellen Tonhöhereindruck beiträgt, bei ca. 20 liegt.

Die virtuelle Tonhöhe hängt, wie die Tonhöhe reiner Töne, vom Darbietungspegel ab. Der genaue Zusammenhang wird u.a. in [58] dargestellt.

Die Art der Erzeugung eines Residualklangs spielt keine Rolle für die virtuelle Tonhöhe. Es ist unerheblich, ob die niedrigsten Harmonischen eines Klanges durch Filterung aus dem Spektrum entfernt oder durch Rauschen vollständig maskiert werden; beide Varianten rufen die gleiche virtuelle Tonhöhe hervor [56].

Die virtuelle Tonhöhe *nichtharmonischer Klänge* wird in Kapitel 6 diskutiert. Sie wird dort dazu benutzt, das in dieser Arbeit implementierte Modell zur Bestimmung virtueller Tonhöhen zu verifizieren.

Ein visuelles Analogon zur virtuellen Tonhöhe [58] ist in Abb. 2.8 beschrieben. Auf der linken Seite wird das Wort 'PITCH' mit Buchstaben dargestellt, deren Ränder durch ausgezogene Linien symbolisiert sind. Dies ist das Analogon zu einem vollständigen, harmonischen Klang. Auf der rechten Seite von Abb. 2.8 sind nicht mehr alle Begrenzungen der Buchstaben eingezeichnet, ähnlich einem Residualklang, dem einige Harmonische und die Grundfrequenz fehlen. Trotzdem wird das Wort 'PITCH' nach wie vor erkannt, so wie einem Residualklang die virtuelle Tonhöhe zugeordnet wird, die in etwa der Grundfrequenz des zugeordneten harmonischen Klangs entspricht. Terhardt bezeichnet dieses Phänomen als Gestalterkennung [44], [50].



PITCH PITCH

Abb. 2.8: *Optisches Beispiel für die virtuelle Tonhöhe. Vollständige Buchstaben, d.h. Analogon für vollständige Klänge auf der linken Seite, demgegenüber nur teilweise ausgezogene Buchstaben, d.h. Analogon für Residualklänge auf der rechten Seite. Das Wort 'PITCH' kann in beiden Fällen abgelesen werden. [58].*

3 FOURIER-ZEIT-TRANSFORMATION

Die Fourier-Transformation ist ein grundlegendes Hilfsmittel der Signalverarbeitung. Um die Grundlage für die Darstellung beliebiger Signale sowohl im Zeit- als auch im Frequenzbereich zu schaffen, sind im Anhang A die Grundlagen der Fourier-Transformation zusammengestellt, im Anhang B sind ihre wichtigsten Eigenschaften aufgelistet.

In diesem Kapitel wird eine Erweiterung der Fourier-Transformation auf die sogenannte Fourier-t-Transformation (FTT, Fourier-Zeit-Transformation) nach Terhardt [46] vorgenommen, die einen integralen Bestandteil der vorliegenden Arbeit bildet. Diese Transformation wird weiterentwickelt und der „herkömmlichen“ Fouriertransformation in Bezug auf eine gehörgerechte Spektralanalyse gegenübergestellt.

Dabei wird besonderer Wert auf die Programmierung und Anpassung des FTT-Algorithmus für den Einsatz auf einer Workstation gelegt. Die für die Implementation des FTT- Algorithmus wichtigen Überlegungen sind in den folgenden Abschnitten zusammengefaßt und erläutert. Alle Abbildungen sind, soweit nicht ausdrücklich anders vermerkt, mit dem in dieser Arbeit entwickelten und programmierten FTT- Algorithmus berechnet worden.

3.1 Definition der Fourier-t-Transformation (FTT)

Die in dieser Arbeit verwendete Fourier-t-Transformation nach Terhardt [46], [47] basiert auf einem Ansatz von Flanagan und ist folgendermaßen definiert:

$$(3.1) \quad \bar{F}(\omega, t) = \int_0^t f(t') \cdot h(\omega, t - t') e^{-i\omega t'} dt' \quad \text{für } t > 0, \quad t \geq t'$$

Dies ist ein Fourierintegral, in dem eine frequenzabhängige Fensterfunktion $h(\omega, t - t')$, mit der das Signal $f(t)$ gewichtet wird, enthalten ist. Die untere Integralgrenze beträgt Null, die obere gibt den Analysezeitpunkt t vor.

Flanagan [nach Terhardt, 46] schlägt eine Exponentialfunktion erster Ordnung als Fensterfunktion vor, Schlang und Mummert [33] dehnen die Betrachtungen auf Exponentialfunktionen zweiter Ordnung aus:

$$(3.2) \quad \text{1. Ordnung: } h_1(\omega, t - t') = \eta_1 \cdot e^{-a(\omega)(t-t')}$$

$$(3.3) \quad \text{2. Ordnung: } h_2(\omega, t - t') = \eta_2 \cdot (t - t') \cdot e^{-a(\omega)(t-t')}$$

Diese Fensterfunktionen sind phänomenologisch begründet. Sie bilden das „natürliche Vergessen“ [46] des menschlichen Gehöres nach und haben sich in der Anwendung als zweckmäßig erwiesen [14]. Die Bedeutung der *Transformationskonstanten* $a(\omega)$ und der *Normierungskonstanten* η_1, η_2 wird in den folgenden Abschnitten erörtert. Der in dieser Arbeit entwickelte FTT- Algorithmus verwendet wahlweise Fensterfunktionen erster oder zweiter Ordnung.

3.2 FTT als gehörgerechte Audiotransformation

Bei geeigneter Wahl der frequenzabhängigen Transformationskonstanten $a(\omega)$ kann eine Anpassung an das Frequenz- und Zeitaufhebungsvermögen des menschlichen Gehörs erreicht werden [14]. Für den FTT- Algorithmus wird $a(\omega)$ nach einem Vorschlag von Flanagan [vgl. Terhardt, 46] proportional zur *Frequenzgruppenbreite* $\Delta f_G(\omega)$ nach Zwicker [55] angesetzt:

$$(3.4) \quad a(\omega) = \pi \cdot \Delta f_G(\omega) \frac{b}{[\text{Bark}]}$$

Dabei stellt b die *Analysebandbreite des Fensters*, gemessen in Bark, dar. Die Frequenzgruppenbreite $\Delta f_G(\omega)$ wird in dem hier realisierten FTT- Algorithmus approximiert durch [59]:

$$(3.5) \quad \Delta f_G(\omega) = 25 + 75 \left(1 + 1,4 \left(\frac{\omega}{2\pi[\text{kHz}]} \right)^2 \right)^{0,69}$$

Die kontinuierlichen Analysefrequenzen ω werden durch Einführung des Parameters *Frequenzabstand der Analysefrequenzen* Δz , gemessen in Bark, für die Ausführung des FTT- Algorithmus auf einer Workstation diskretisiert [14]:

$$(3.6) \quad \omega_j = 2\pi(f_u + j \cdot \Delta z) \quad j = 0..N$$

Durch geeignete Wahl von Δz ist eine Anpassung an die Frequenzauflösung des menschlichen Gehöres¹ gewährleistet [14]. Die Parameter *unterste Analysefrequenz*² f_u und *Gesamtzahl der Analysefrequenzen* N können dabei in Abhängigkeit von dem interessierenden Analysebereich gewählt werden.

¹ Der eben wahrnehmbare Frequenzunterschied von (frequenzmodulierten) Sinustönen beträgt etwa $1/25$ bis $1/50$ der Frequenzgruppenbreite, also 0,04 bis 0,02 Bark [59]. Heldmann [14] benutzt für seine Berechnungen einen Wert von $\Delta z = 0,03$ Bark.

² Die unterste Analysefrequenz f_u sollte nicht zu klein gewählt werden. Wird f_u zu tief abgesenkt, so approximiert Gl (3.5) die Frequenzgruppenbreite nur sehr schlecht. Als Anhaltswert für f_u kann ein Wert von 100 Hz gelten [14].

3.3 Bedeutung der Fensterfunktion für die FTT

Die Normierungskonstanten η_1 und η_2 der Fensterfunktionen Gl. (3.2) und (3.3) werden in dem hier implementierten FTT- Algorithmus so bestimmt, daß die Fensterfläche für alle Frequenzen konstant ist. Dadurch wird der Einfluß der frequenzabhängigen Fensterlänge und -form auf die Amplituden der Spektrallinien kompensiert [33].

$$(3.7) \quad \eta_1 = 2 \cdot a(\omega)$$

$$(3.8) \quad \eta_2 = 2 \cdot a^2(\omega)$$

Einige normierte Fensterfunktionen für exemplarisch ausgewählte Frequenzen sind in Abb. 3.1 und Abb. 3.2 dargestellt. Die Werte für die beiden Abbildungen werden mit Gl. (3.2) und (3.3) berechnet, die auch im FTT- Algorithmus enthalten sind.

Schlang und Mummert [33] zeigen, daß das resynthetisierte Sprachsignal einer FTT unter Verwendung einer Fensterfunktion 1. Ordnung hallig klingt. Die Qualität der resynthetisierten Sprachschalle kann erheblich verbessert werden, wenn Fensterordnungen größer 1 benutzt werden. Als „geeigneten Kompromiß zwischen Berechnungsaufwand und Nutzen“ der FTT bezeichnen Schlang und Mummert die zweite Fensterordnung.

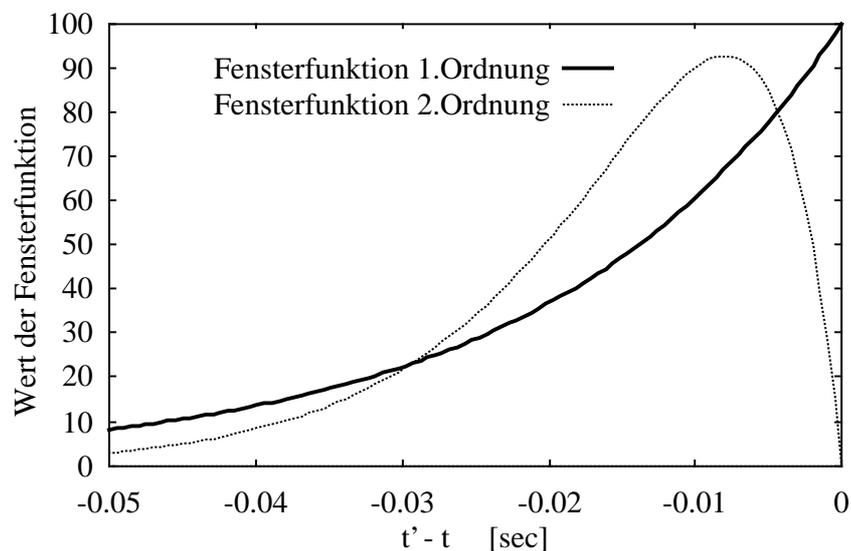


Abb. 3.1: Normierte Fensterfunktionen für Analysefrequenzen von 1 kHz:

1. Ordnung: $b = 0.1$ Bark, entsprechend $a(\omega) = 50$ Hz,
2. Ordnung: $b = 0.25$ Bark, entsprechend $a(\omega) = 126$ Hz

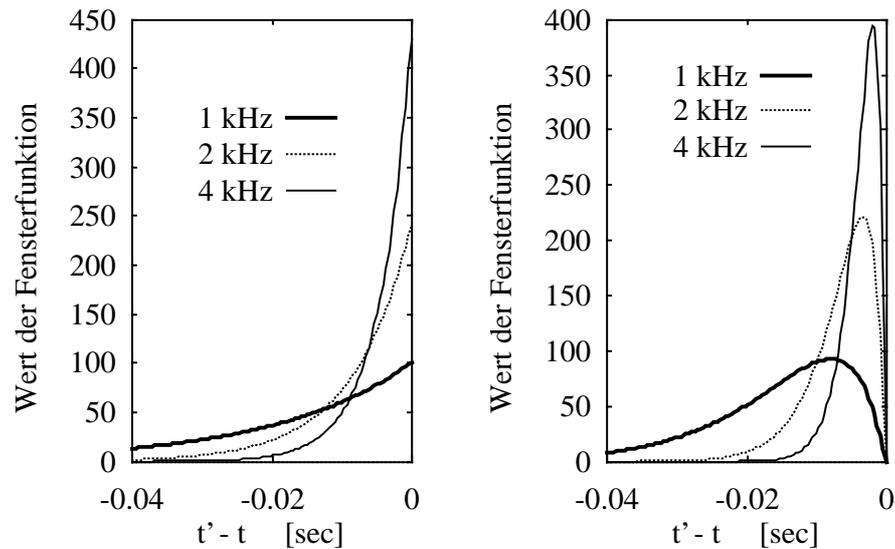


Abb. 3.2: Normierte Fensterfunktionen für ausgewählte Analysefrequenzen:
 links: 1. Ordnung, 1, 2 und 4 kHz mit $a(\omega) = 50, 120$ und 215 Hz,
 rechts: 2. Ordnung, 1, 2 und 4 kHz mit $a(\omega) = 126, 301$ und 538 Hz.

Die Qualitätssteigerung wird vor allem dadurch bedingt, daß die Analysebandbreite b einer FTT mit Fenstern 2. Ordnung aufgrund größerer Flankensteilheit gegenüber FTTs mit Fenstern 1. Ordnung von 0,1 auf 0,25 Bark erhöht werden kann. Hierdurch wird die Zeitauflösung verbessert, siehe Abb. 3.2 und Abb. 3.3. Für Fensterfunktionen höherer Ordnung als 2 kann die Analysebandbreite b maximal auf 0,3 Bark erhöht werden, der Berechnungsaufwand steigt jedoch beträchtlich [33].

3.4 Zeit- und Frequenzauflösung der FTT

Unter Voraussetzung eines eingeschwungenen FTT-Spektrums eines Sinustones der Frequenz ω_0 , das heißt $t \gg \{a(\omega)\}^{-1}$ und $a(\omega) \ll \omega, \omega_0$, ergibt sich die Frequenz- und Zeitauflösung der FTT wie in Tab. 3.1 dargestellt. Die Berechnung dieser Werte ist im Anhang C ausgeführt.

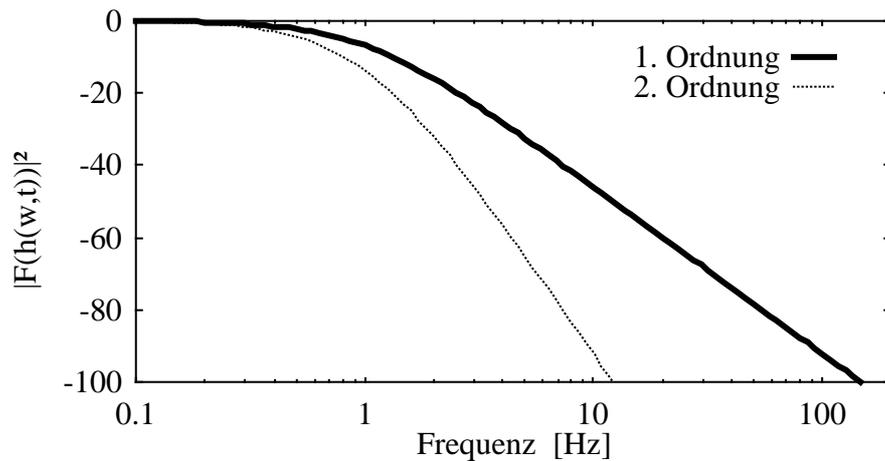


Abb. 3.3: *Fouriertransformierte der verwendeten Fensterfunktionen. Das Spektrum der Fensterfunktionen Gl. (3.2) und (3.3) ist doppelt logarithmisch gegen die Frequenz aufgetragen. Für beide Fensterfunktionen wird die gleiche Transformationskonstante verwendet. Das Analyseintervall erstreckt sich von $-\infty$ bis zum Analysezeitpunkt t . Die Transformierten ergeben sich, indem Gl. (3.2) und (3.3) mittels [5, Tab. 4.4.2.2.] umgeformt werden, vgl. Anhang A und B.*

	1. Ordnung	2. Ordnung	menschliches Gehör ³
Zeitauflösung, bzw. Fensterbreite T	$\frac{1}{a(\omega)}$	$\approx \frac{2,987}{a(\omega)}$ ⁽⁴⁾	≈ 10 msec für 0,1 kHz ≈ 6 msec für 1 kHz ≈ 2 msec für ≥ 3 kHz
Frequenzauflösung, bzw. Bandbreite B	$\frac{a(\omega)}{\pi}$	$\frac{\sqrt{\sqrt{2}-1}}{\pi} a(\omega)$	0,2 - 0,5 Bark
$B \cdot T$ -Produkt ⁵	$\cong 0,318$	$\cong 0,612$	-

Tabelle 3.1: *Zeit- und Frequenzauflösung der FTT*

³ Die angegebenen Werte für das menschliche Gehör sind lediglich als Anhaltsgrößen zu verstehen. Sowohl die Zeit- als auch die Frequenzauflösung sind abhängig von Frequenz, Pegel, Art und Darbietungsdauer des untersuchten Geräusches. Die Zeitauflösung wird als Grenzwert der Wahrnehmung der Rauigkeit [6], [58], [38] aufgefaßt. Die Frequenzauflösung wird als Grenze der Ununterscheidbarkeit zweier Tonhöhen gegenüber einer einzigen angesehen. Der hier angegebene Wert entstammt Terhardt [38].

⁴ Heldmann gibt für die effektive Fensterlänge $\sqrt{5}/a(\omega)$ an [14].

⁵ Es handelt sich um die *effektiven* Werte für Bandbreite und Fensterlänge, vgl. Anhang A4.

Aus Tab. 3.1 wird 0,318 als *BT*- Produkt für die FTT erster Ordnung, 0,612 als *BT*- Produkt für die FTT zweiter Ordnung entnommen. Sie ist in beiden Fällen kleiner als das *BT*- Produkt der FFT, welches 1 beträgt, vgl. Anhang A4. Eine FTT erster oder zweiter Ordnung hat daher eine bessere Zeitauflösung bei gleicher Frequenzauflösung, bzw. eine bessere Frequenzauflösung bei gleicher Zeitauflösung als eine FFT. Da zudem die Frequenz- und Zeitauflösung der FTT frequenzabhängig wählbar ist, werden diese Eigenschaften des menschlichen Gehöres durch eine FTT besser nachgebildet als dies durch eine FFT erfolgen kann.

3.5 Swing

Aufgrund der Unstetigkeit im Analysezeitpunkt ‘schwingt’ das Spektrum einer FTT 1. Ordnung für Sinustöne. Dieser Effekt entsteht durch die Gewichtung mit der Fensterfunktion, die eine Sprungstelle im Analysezeitpunkt besitzt (vgl. Abb. 3.1): Für aufeinanderfolgende Analysezeitpunkte werden jeweils andere Ausschnitte aus dem Zeitsignal (d.h. der Sinusfunktion) analysiert. Innerhalb dieser aufeinanderfolgenden Ausschnitte verschiebt sich die spektrale Zusammensetzung des analysierten Signals jeweils um einen kleinen Anteil. Aufgrund der Periodizität der untersuchten Sinusfunktion wird der gleiche Ausschnitt eine Periodendauer der Sinusfunktion später erneut analysiert. Schon nach der halben Periodendauer wird ein Ausschnitt der Sinusfunktion analysiert, der das umgekehrte Vorzeichen besitzt. Im Spektrum unterscheiden sich Ausschnitte, die jeweils um die halbe Periodendauer der Sinusfunktion verschoben sind, daher nur durch umgekehrtes Vorzeichen. Im Rahmen dieser Arbeit werden jedoch ausschließlich Leistungsspektren ausgewertet, somit ist dieser Unterschied bedeutungslos. Die ‘Frequenz der Schwingung des Leistungsspektrums’ ist daher doppelt so groß wie die Frequenz des untersuchten Sinustones. Terhardt bezeichnet dieses Phänomen als ‘Swing’ [46].

Abb. 3.4 zeigt die mit dem in dieser Arbeit implementierten FTT- Algorithmus berechneten FTT- Leistungsspektren eines Sinustones für eine FTT erster und zweiter Ordnung zu verschiedenen Analysezeitpunkten. Die Periodendauer des ‘Swings’ beträgt 0,5 msec. Abb. 3.4 zeigt, daß das Phänomen des ‘Swinging’ für FTTs 2. Ordnung nicht meßbar auftritt.

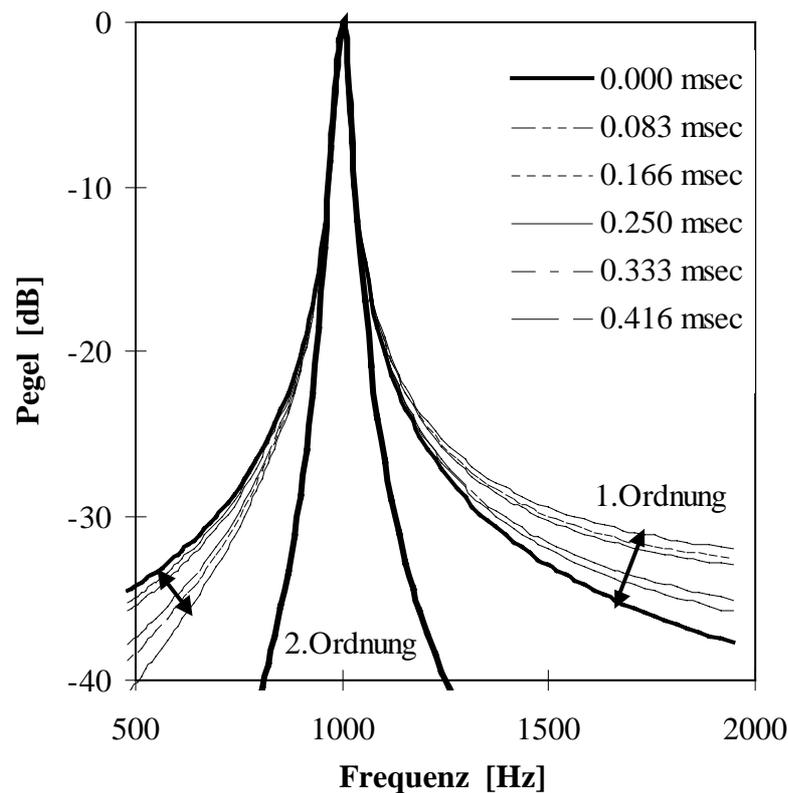


Abb. 3.4: *Swing eines 1 kHz- Sinustones. Dargestellt sind 6 eingeschlungene FTT-Leistungsspektren⁶ 1. Ordnung eines 1 kHz- Sinustones zu verschiedenen, äquidistanten Zeitpunkten. Die Zeitangaben sind relativ zur „Swing“-Periode bezeichnet. Das „Swinging“ mit einer Frequenz von 2 kHz ist zu erkennen. Zum Vergleich ist das FTT-Leistungsspektrum zweiter Ordnung (dicke Linie) für die gleichen Zeitpunkte aufgetragen: für ein Analysefenster zweiter Ordnung tritt kein „Swinging“ auf.*

3.6 Glättung des FTT-Spektrums

Bisher wurden die Fensterfunktionen nur im Analysezeitraum von $-\infty$ bis zum Analysezeitpunkt t betrachtet, bzw. ein eingeschlungenes FTT-Spektrum vorausgesetzt. Dies ist jedoch nicht immer der Fall. Wird zum Beispiel ein Signal rechteckförmig ein- und ausgeschaltet, so ergeben sich im FTT-Leistungsspektrum Nebenmaxima.

⁶ Streng genommen handelt es sich nicht um das Leistungsspektrum, sondern um das Betragsquadratspektrum. Die Proportionalitätskonstante, die insbesondere für die Einheit der Leistung wichtig ist, wird hier nicht berücksichtigt [24], [21].

Das FTT-Fenster ist zwar prinzipiell unendlich lang, die Signallänge sei jedoch beschränkt und von der Dauer T . Die Entstehung der Nebenmaxima kann somit durch eine Faltung der Spektren des FTT-Fensters (Abb. 3.3), eines virtuellen Rechteckfensters⁷ der Länge T und der periodisch fortgesetzten Signalfunktion erklärt werden.

Abb. 3.5A zeigt die Faltung des Spektrums eines Rechteckfensters der Dauer T mit dem Spektrum des FTT-Fensters 1. Ordnung für einen Testton von 1 kHz, Abb. 3.5C zeigt den gleichen Zusammenhang für das Spektrum eines FTT-Fensters 2. Ordnung. Alle vier Einzeldiagramme der Abb. 3.5 wurden mit dem in dieser Arbeit erstellten FTT- Algorithmus berechnet.

In den Einzeldiagrammen der Abb. 3.5 ist jeweils das Leistungsspektrum der FTTs über der Frequenz aufgetragen. Die Beschränkung auf das Leistungsspektrum erfolgt, weil im Rahmen dieser Arbeit die Phaseninformation der Spektren vernachlässigt werden kann. Die Motivation hierfür wird in Abschnitt 4.1.2 gegeben. Das Betragsquadrat- oder Leistungsspektrum $P(\omega, t)$ stellt sich dar als:

$$(3.9) \quad P(\omega, t) = |\bar{F}(\omega, t)|^2$$

In Abb. 3.5A und Abb. 3.5C ist zu erkennen, daß nach etwa 128 msec die Spektren „eingeschwungen“ sind, d.h. die Nebenmaxima sind nicht mehr durch das Gehör wahrnehmbar (Detektionsschwelle etwa 1 dB nach [58]).

Während der ersten 128 msec sind die durch die Analyse bedingten Nebenmaxima jedoch so ausgeprägt, daß sie durch das Gehör wahrgenommen würden, wenn sie im Testsignal (1 kHz Sinus) vorhanden wären. Da dies nicht der Fall ist, müssen die Nebenmaxima unterdrückt werden, um eine gehörgerechte Spektraldarstellung des Signals zu erhalten.

Dazu wird in dem FTT- Algorithmus dieser Arbeit eine zeitliche Glättung der Leistungsspektren $P(\omega, t)$ mittels eines Tiefpasses erster Ordnung (Integrator) durchgeführt [47]:

$$(3.10) \quad S(\omega, t) = \frac{1}{T_g(\omega)} \int_0^t P(\omega, t') \cdot e^{-(t-t')/T_g(\omega)} dt'$$

$S(\omega, t)$ stellt das geglättete Leistungsspektrum der FTT dar, $T_g(\omega)$ ist die *Glättungszeitkonstante*. Sie muß so gewählt werden, daß einerseits das günstige BT-Produkt der FTT nicht beeinträchtigt wird, andererseits die Nebenmaxima so gut wie möglich unterdrückt werden. Durch eine analytische Betrachtung des tran-

⁷ Das Spektrum eines Rechteckfensters ist die sinc-Funktion, vgl. z.B. [1].

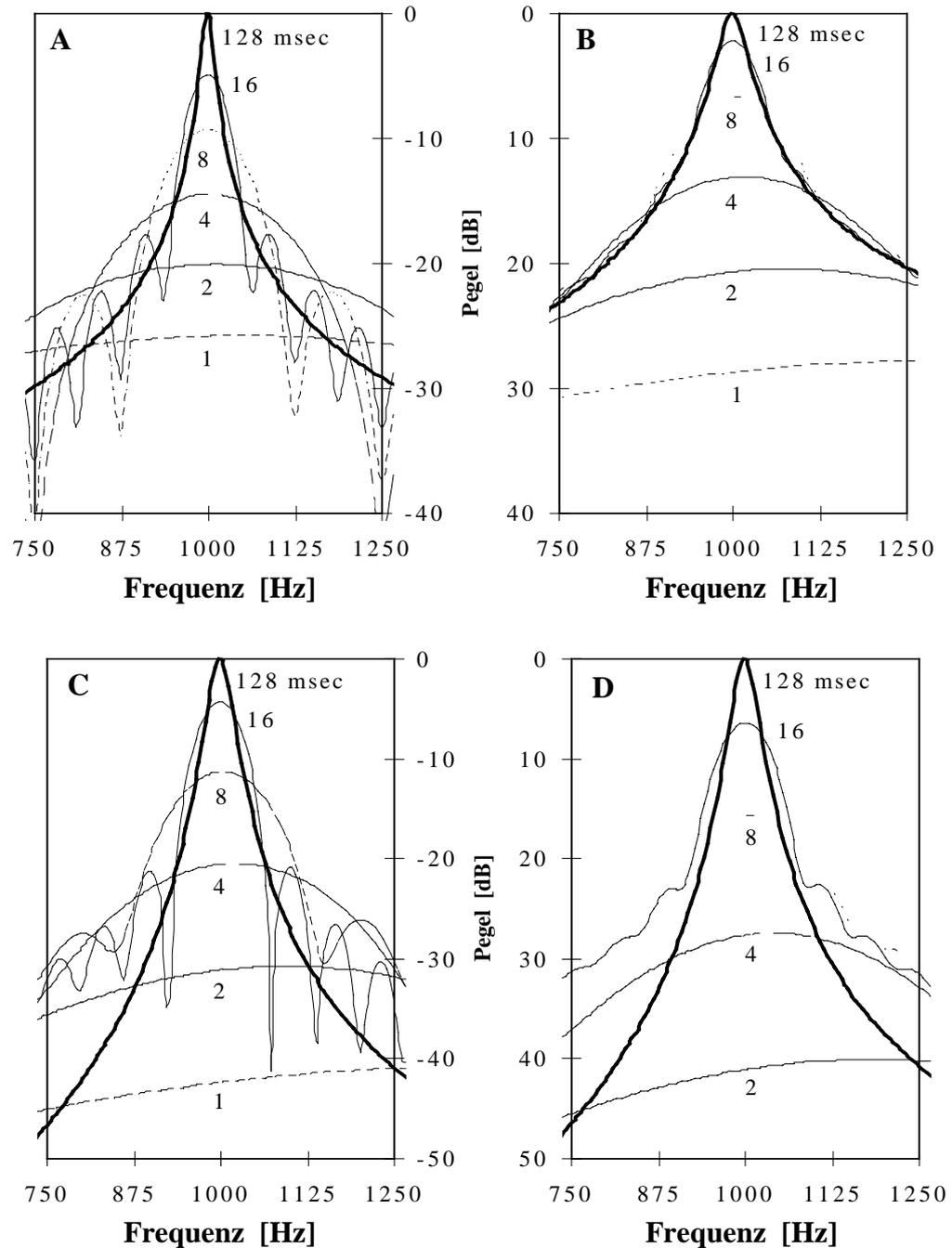


Abb. 3.5: FTT-Spektren eines zum Zeitpunkt $\Theta = 0$ rechteckförmig eingeschalteten Sinustones der Frequenz 1 kHz. Die Zahlen in den Diagrammen geben die verstrichene Zeit relativ zu Θ an, Erläuterung und Interpretation siehe Text.

oben: FTT-Spek. 1. Ordnung: **A:**, 'ungeglättet'. **B:** 'geglättet'.

unten: FTT-Spek. 2. Ordnung: **C:**, 'ungeglättet'. **D:** 'geglättet'.

sienten Anteils der FTT, Gl. (3.1) mit der Fensterfunktion erster Ordnung für feste Frequenzen ergeben sich sinnvolle Werte von etwa $0,12/a(\omega)$ bis $0,54/a(\omega)$ für $T_g(\omega)$ [13].

Für den vorliegenden FTT- Algorithmus wird die frequenzabhängige Glättungszeitkonstante $T_g(\omega)$ an der Zeitauflösung des Gehörs orientiert. Nach [38] ist die Zeitauflösung unterhalb von 3 kHz etwa proportional dem Kehrwert der Frequenzgruppenbreite, oberhalb von 3 kHz bleibt sie konstant bei etwa 2 msec, siehe auch Tab. 3.1. Heinbach [13] zeigt, daß mit der Dimensionierung

$$(3.11) \quad T_g(\omega) = \begin{cases} 0,2 / a(\omega) & \text{für } \omega < 2\pi \cdot 3\text{kHz} \\ 0,2 / a(2\pi \cdot 3\text{kHz}) & \text{für } \omega \geq 2\pi \cdot 3\text{kHz} \end{cases}$$

nicht nur für reine Sinustöne, sondern auch für komplexe Zeitsignale die Nebenmaxima ausreichend gedämpft werden (für FTT-Spektren 1. Ordnung). Nach Vereinfachung ergibt sich:

$$(3.12) \quad T_g(\omega) = \begin{cases} \approx 0,64 / \Delta f_g(\omega) & \text{für } \omega < 2\pi \cdot 3\text{kHz} \\ \approx 1,25\text{msec} & \text{für } \omega \geq 2\pi \cdot 3\text{kHz}. \end{cases}$$

Abb. 3.5B, D zeigen das gleiche FTT-Spektrum wie Abb. 3.5A, C, jedoch diesmal ‘geglättet’ mittels eines Tiefpasses mit der Glättungszeitkonstanten nach Gl. (3.11). Die FTT zweiter Ordnung ist mit den gleichen Parametern $T_g(\omega)$ und $a(\omega)$ ‘geglättet’ wie für die FTT erster Ordnung angegeben⁸.

Aus Abb. 3.5C ist abzulesen, daß die Glättung für die FTT mit einer Fensterfunktion erster Ordnung zu jedem Zeitpunkt der Analyse die Nebenmaxima unterdrückt. Diese Unterdrückung wird durch die Anpassung der Glättungszeitkonstanten, die unter Verwendung einer FTT erster Ordnung ermittelt wurde, erreicht.

Aus Abb. 3.5D geht hervor, daß etwa 8 msec nachdem das Signal rechteckförmig eingeschaltet wurde, kleine Nebenmaxima auftreten. Die Nebenmaxima haben jedoch nur einen Pegelunterschied von 0,5 dB gegenüber den unmittelbar angrenzenden Minima. Das Ziel, den Pegelunterschied unter den Wert von 1 dB, also unter die Wahrnehmbarkeitsschwelle des menschlichen Gehörs zu drücken, ist auch hier gut erfüllt.

⁸ Dies sind die Parameter, wie Heldmann sie in seiner Dissertation benutzt [14].

3.7 Diskrete FTT

Durch analoge Diskretisierung der FTT, wie im Anhang A4 für die FFT durchgeführt, zeigt sich eine für die numerische Berechnung der FTT wichtige Eigenschaft: Aus der Definition der FTT, Gl. (3.1), wird mit $t = m \cdot T_s$, $t' = j \cdot T_s$ und $f(j \cdot T_s) \Rightarrow f_j$:

- 1. Ordnung:

$$(3.13) \quad \bar{F}_1(\omega, m) = \eta_1 \cdot T_s \cdot \sum_{j=0}^m f_j \cdot e^{-a(\omega)(m-j)T_s} \cdot e^{-i\omega j T_s}$$

Daraus kann durch geschicktes Zerlegen der Summe ein rekursiver Algorithmus gewonnen werden [46]:

$$(3.14) \quad \bar{F}_1(\omega, m+1) = \bar{F}_1(\omega, m) \cdot e^{-a(\omega)T_s} + T_s \cdot \eta_1 \cdot f_{m+1} \cdot e^{-i\omega(m+1)T_s}$$

Aufgrund der Rekursivität ist diese Gleichung hervorragend für den Einsatz auf Digitalrechnern geeignet. Deshalb wird die Rekursion (3.14) in dem hier vorgestellten FTT- Algorithmus zur Berechnung von FTTs erster Ordnung verwendet.

- 2. Ordnung:

$$(3.15) \quad \bar{F}_2(\omega, m) = T_s^2 \cdot \eta_2 \cdot \sum_{j=0}^m f_j \cdot e^{-a(\omega)(m-j)T_s} \cdot e^{-i\omega j T_s} \cdot (m-j)$$

Um eine Rekursion zu erhalten, wird die Summe der Gl. (3.14) zerlegt:

$$(3.16) \quad \bar{F}_2(\omega, m+1) = e^{-a(\omega)T_s} \cdot \left(\bar{F}_2(\omega, m) + \frac{\eta_2}{\eta_1} \cdot T_s^2 \cdot \bar{F}_1(\omega, m) \right)$$

Gl. (3.16) wird in dem hier vorgestellten FTT- Algorithmus zur Berechnung der FTT 2. Ordnung implementiert. Die Rekursion (3.16) erfordert etwa den doppelten Rechenaufwand gegenüber derjenigen für die FTT 1. Ordnung.

- Glättung:

Der Glättungstiefpaß nach Gl. (3.10) wird ebenfalls diskretisiert:

$$(3.17) \quad S(\omega, m) = \frac{T_s}{T_g(\omega)} \sum_{j=0}^m P(\omega, j) \cdot e^{-T_s(m-j)/T_g(\omega)}$$

Die Rekursion lautet:

$$(3.18) \quad S(\omega, m+1) = S(\omega, m) \cdot e^{-T_s/T_g(\omega)} + \frac{T_s}{T_g(\omega)} P(\omega, m+1)$$

Gl. (3.18) wird zur Glättung der FTT- Leistungsspektren in dem in dieser Arbeit vorgestellten FTT- Algorithmus verwendet.

3.8 Inverse FTT

Die inverse FTT für ein Fenster erster Ordnung kann relativ einfach hergeleitet werden. Sie lautet [46]:

$$(3.19) \quad f(t') = \frac{e^{a(t-t')}}{2\pi} \int_{-\infty}^{+\infty} \overline{F}(\omega, t) \cdot e^{-i\omega t'} d\omega \quad \text{für } t \geq t'$$

Wesentlich komplizierter ist es, die inverse FTT für höhere Fensterordnungen zu berechnen. Durch die frequenzabhängige Transformationskonstante $a(\omega)$ entstehen frequenzabhängige Laufzeiten der spektralen Komponenten eines Testsignals. Diese Laufzeitunterschiede müssen durch eine Korrekturfunktion ausgeglichen werden, um ein Signal aus seinem FTT-Spektrum rekonstruieren zu können. Hierzu hat Mummert [27] ausführliche systemtheoretische Betrachtungen angestellt. Eine Näherung zur Invertierung der FTT höherer Ordnung (≥ 2) wird im nächsten Kapitel dargestellt.

3.9 Beispiel zur FTT

Abb. 3.6 stellt beispielhaft ein mit dem in dieser Arbeit vorgestellten FTT-Algorithmus berechnetes FTT-Spektrum dar. Auf der Abszisse ist die Zeit relativ zum Beginn der FTT-Analyse aufgetragen, die Ordinate repräsentiert die Tonheit in Bark, d.h. die Frequenzgruppeneinteilung nach Zwicker [55]. Der Pegel des FTT-Leistungsspektrums ist farblich skaliert in der Darstellung enthalten und wird in dB abgetragen. Die Skalierung des Pegels wird relativ zum Pegelmaximum des FTT-Spektrums vorgenommen, welches auf 0 dB festgelegt ist. Daher treten nur negative Pegelwerte auf, vgl. Pegellegende auf der rechten Seite von Abb. 3.6.

Dargestellt ist das FTT-Spektrum des Satzes 'Käse riecht manchmal sehr', gesprochen von einem männlichen Sprecher (Autor). Der Satz liegt auf DAT-Kassette vor, gesampelt mit 48 kHz. Zur Berechnung des Spektrums werden folgende Parameter verwendet:

Fensterfunktion:	2. Ordnung
Glättung des Leistungsspektrums:	wie in Kapitel 3.6
Frequenzabstand der Analysefrequenzen Δz :	0,03 Bark
Analysebandbreite b :	0,25 Bark

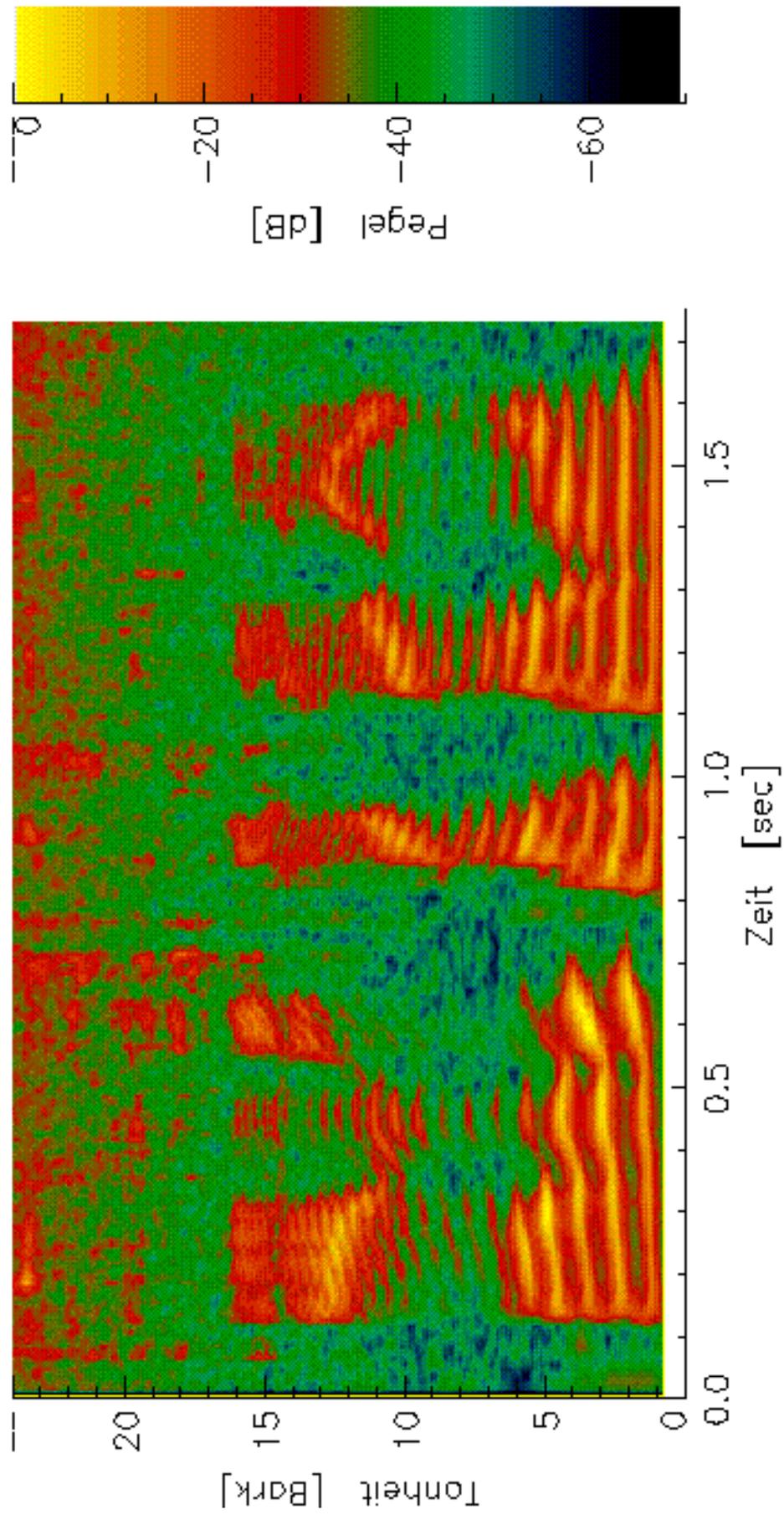


Abb 3.6: FTT-Spektrum für das Beispielsignal aus Abschnitt 3.9.

Aus Abb. 3.6 ist die Grundfrequenz der männlichen Stimme ablesbar. Sie beträgt etwa 105 Hz. Die Oberschwingungen der Stimme sind gut zu erkennen, insbesondere ist die unterschiedliche Intensitätsverteilung zwischen den einzelnen Silben und Lauten des Satzes nachzuvollziehen. Die höchsten Oberschwingungen der Stimme reichen bis etwa 16 Bark (ca. 3150 Hz). Dies stimmt gut mit den Daten von [25] überein.

Im Bereich oberhalb von 20 Bark treten Frequenzkomponenten auf, die durch Hintergrundrauschen bei der Aufnahme des Satzes hervorgerufen werden. Diese rauschhaften Anteile machen sich in der gewählten Darstellung besonders stark bei großen Frequenzen bemerkbar, weil die Frequenzgruppenbreite zu großen Frequenzen hin zunimmt. Aufgrund der anwachsenden Breite der Frequenzgruppen fällt daher für höhere Frequenzen mehr Energie in die entsprechende Frequenzgruppe und somit steigt der Pegel des Rauschens an.

Eine zeitliche Beschreibung des Spektrums des Testsatzes wird im folgenden durchgeführt. Dabei wird im wesentlichen nur auf [25, Kapitel 7.2] zurückgegriffen und keine weiteren Ergebnisse der Sprachverarbeitung berücksichtigt.

Der Satz beginnt bei etwa 80 msec mit dem Laut 'k'. Ebenso wie das 't' bei 780 msec ist das 'k' ein Plosivlaut, welcher durch ein nahezu gleichmäßig angeregtes Spektrum charakterisiert wird. Bei dem 't' kann man dies gut erkennen, wenn man die eingangs über die Energieverteilung aufgrund der Frequenzgruppenbreite gemachten Aussagen auf das Spektrum des 't' (und analog auch auf das des 'k') überträgt.

An das 'k' schließt sich der kontinuierliche Übergang von 'ä' auf 's' und schließlich 'e' an. Dieser Übergang findet zwischen 120 und 310 msec statt. Das Spektrum, d.h. die Formantverteilung verändert sich in dieser Zeit für die Vokale 'ä' und 'e' sowie den stimmhaften Reibelaut 's' wie es nach [25, Abb. 7.16 und 7.25] zu erwarten ist.

Der Anfang des nächsten Wortes ('riecht') ist stark mit dem Ende des Wortes 'Käse' verschliffen, das 'r' (310 bis 390 msec) besteht hauptsächlich aus Frequenzanteilen unterhalb von 5 Bark. Beim nachfolgenden 'ie', welches etwa von 400 bis 500 msec andauert, tritt die Formantstruktur besonders deutlich hervor. Daran schließt sich der Frikativ 'ch' an (530 bis 680 msec), der durch den hochfrequenten Bereich oberhalb von etwa 12 Bark hervorgerufen wird. Noch deutlicher tritt der Frikativ 'ch' beim Wort 'manchmal' zwischen 950 und 1050 msec auf. Im Wort 'manchmal' kommen drei Nasale vor, zweimal 'm' und einmal 'n'. Nasale werden rein periodisch durch die Stimmritze angeregt [25, Seite 226] und zeichnen sich daher durch ein streng periodisches Spektrum aus:

‘man’ von 810 bis 950 msec und ‘mal’ von 1120 bis 1330 msec. Das ‘l’ aus ‘manchmal’ ist ein Liquidlaut, der genau wie die Nasale, periodisch durch die Stimmritze angeregt wird [25] und daher eine ähnliche spektrale Struktur wie die Nasale aufweist.

Im Zeitbereich von 1350 bis 1620 msec wird das Wort ‘sehr’ analysiert. Es ist wiederum eine Verschleifung des Wortes ‘sehr’ mit dem vorangehenden Wort ‘manchmal’ zu erkennen. Wie für die zuvor diskutierten Worte können auch für das Wort ‘sehr’ die Formanten der einzelnen Phoneme aus dem Diagramm abgelesen werden.

3.10 Bilanzierende Bewertung der FTT

Die in dem hier implementierten FTT- Algorithmus enthaltenen Rekursionen erlauben die effiziente Berechnung von FTT- Spektren. Besonders wenn zeitlich quasikontinuierliche Spektralverläufe (d.h. z.B. für jeden Abtastpunkt) berechnet werden sollen, zeigen sich die deutlich günstigeren Rechenzeiten einer FTT gegenüber einer FFT, vgl. [47].

Zudem wird die FTT, im Gegensatz zur FFT, mit einer dem Gehör nachempfundenen Frequenzauflösung ausgeführt. Daher ist die FTT besser als die FFT geeignet, als Vorverarbeitung für komplexe Musterbeschreibungen im Gehör zu dienen, mittels derer akustische Merkmale extrahiert werden können [51]. Besonders im Hinblick auf die nächsten Kapitel ist dies von entscheidender Bedeutung.

4 TEILTONZEITMUSTER (TTZM)

Zur Bestimmung der tonalen Komponenten aus dem FTT- Spektrum wird in dieser Arbeit das Teiltonzeitmuster von Heinbach [12], [13] benutzt. Das Teiltonzeitmuster (TTZM) wurde von Heinbach als gehörgerechte Signalverarbeitung eingeführt, um eine an das menschliche Gehör angepasste Übertragung, Speicherung, Bearbeitung und graphische Darstellung von Audiosignalen zu ermöglichen. Insbesondere können mit Hilfe des TTZM informationsgewinnende Mustererkennungsprozesse des Gehörs nachgebildet werden. Heinbach stützt sich hierbei auf Untersuchungen von Terhardt [41], [42], der folgende Ergebnisse gefunden hat:

- Spektraltonhöhen sind akustische Gegenstücke zu visuellen Konturen.
- Nach dem Konzept der Spektraltonhöhen läßt sich jede Spektraltonhöhe unmittelbar auf eine physikalische Ursache zurückführen, aber nicht jede spektrale Komponente eines Schalles erzeugt eine Spektraltonhöhe.
- Spektraltonhöhen sind die Träger der Informationen von Audiosignalen für die Gesamttonhöhenwahrnehmung, die virtuelle Tonhöhe, die Klangfarbenwahrnehmung und die tonalen Eigenschaften.

Heinbach erweitert diese Untersuchungen und findet, daß die Spektraltonhöhen das Ergebnis einer der ersten informationsgewinnenden Entscheidungsprozesse des menschlichen Gehöres sind [13]. Dieses Ergebnis wird durch ein Analyse- durch Syntheseverfahren bestätigt.

Heldmann kann ebenfalls durch ein Analyse- durch Syntheseverfahren zeigen, daß eine Vernetzung im TTZM die Signalverarbeitung des menschlichen Gehörs nachbildet [14].

Das in dieser Arbeit verwendete Konzept des TTZM ist in Abb. 4.1 dargestellt: Von einem Eingangssignal $s(t)$ wird für jeden Abtastwert mittels der zuvor diskutierten FTT ein neues Kurzzeitspektrum gebildet. Anschließend wird daraus das Leistungsspektrum berechnet und dieses zeitlich geglättet. Die weitere Auswertung findet in zeitlichen Abständen eines frequenzabhängigen, gehörgerechten Auswertintervalls $T_{fA}(f)$ statt. Aus dem so erhaltenen Spektrum werden die Teiltöne (TT), d.h. die grundsätzlich durch einen menschlichen Hörer wahrnehmbaren tonalen Komponenten, extrahiert. Diese fünf Schritte definieren das Teiltonmuster (TTM). Durch zeitlich fortlaufende Aneinanderreihung der TTM erhält man das Teiltonzeitmuster (TTZM), welches durch 'Vernetzung' ins Teiltonlinienzeitmuster (TTLZM) überführt werden kann. Die in der vorliegen-

den Arbeit vorgenommene Umsetzung der fünf oben definierten Einzelschritte für die Implementation des TTZM wird für jeden Schritt separat in den nächsten Abschnitten erläutert.

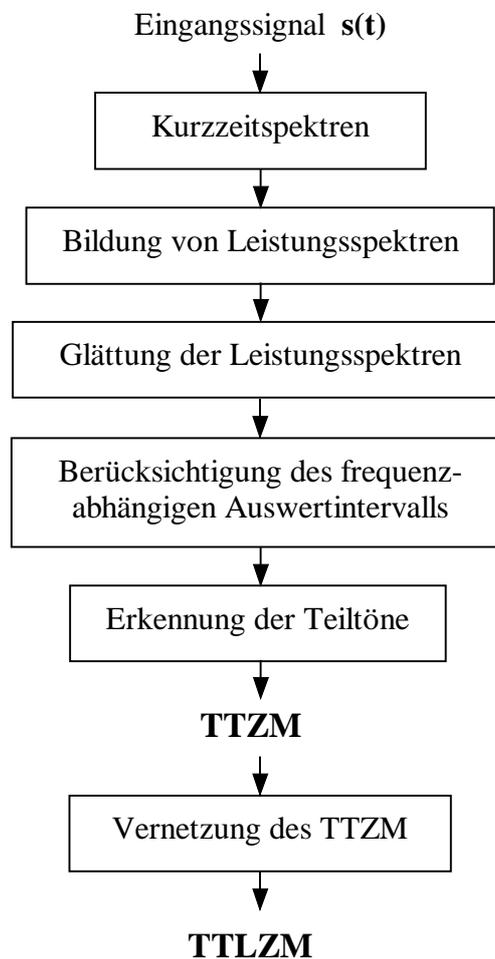


Abb. 4.1: Einzelschritte zur Bildung des TTZM nach Heinbach [12], [13] bzw. des TTLZM nach Heldmann [14].

4.1 Bildung des TTZM

4.1.1 Berechnung der Kurzzeitspektren

Die Spektralanalyse wird in der vorliegenden Arbeit mittels einer FTT, wie im Kapitel 3 definiert, ausgeführt. Die FTT- Parameter können dabei je nach Bedarf gewählt werden. Eine sinnvolle Wahl dieser Parameter wird in den nächsten Abschnitten diskutiert.

4.1.2 Berechnung der Leistungsspektren

Nach Untersuchungen von Patterson [28] und Wightman [54] ist die Empfindungsgröße Tonhöhenwahrnehmung von virtuellen und spektralen Tönen im wesentlichen nicht von der Phasenlage der Komponenten abhängig. Analog dazu finden Plomp und Steneken [29], daß auch die Klangfarbenwahrnehmung nicht wesentlich von der Phaseninformation des Signals abhängt.

Aufgrund dieser Ergebnisse wird die Phaseninformation der Spektren in der vorliegenden Arbeit vernachlässigt. Es werden nur Betragsquadratspektren (hier auch synonym als Leistungsspektren bezeichnet, vgl. Fußnote auf Seite 20) betrachtet. Die Datenmenge reduziert sich dabei um 50%.

Dies ist auch die Vorgehensweise von Heinbach [13]. Er zeigt, daß das Leistungsspektrum ausreichend ist, um alle für den Tonhöhen Eindruck und die Klangfarbe wichtigen Eigenschaften eines Signals im Teiltonzeitmuster zu repräsentieren.

Die Bildung der FTT- Leistungsspektren wird in dieser Arbeit unter Verwendung von Gl. (3.9) durchgeführt.

4.1.3 Glättung der Leistungsspektren

Die Glättung der Leistungsspektren wird für das hier vorgestellte Modell mittels eines Tiefpasses 1. Ordnung nach Gl (3.10) bzw. (3.18) durchgeführt , vgl. Abschnitt 3.6.

Ein Beispiel für ein FTT- Spektrum, welches mit dem hier dargestellten Modell berechnet wurde, zeigt Abb. 3.6 des vorigen Kapitels.

4.1.4 Berücksichtigung des zeitabhängigen Auswertintervalls

Die FTT berücksichtigt das frequenzabhängige Frequenzauflösungsvermögen des Gehörs. Um der Zeitauflösung des Gehörs Rechnung zu tragen, wird das geglättete FTT- Leistungsspektrum in dem hier erstellten Algorithmus nur zu äquidistanten, aber frequenzabhängigen Analysezeitpunkten neu ausgewertet. Innerhalb dieses Auswertintervalls werden die spektralen Komponenten des letzten Zeitpunkts für die weitere Auswertung beibehalten. Erst nach Ablauf des Analyseintervalls werden die Komponenten aktualisiert. Durch diese Maßnahme wird laut Heldmann [14] gewährleistet, daß nur spektrale Anteile berücksichtigt werden, die nach Frequenz und Dauer prinzipiell wahrnehmbar sind.

Aus diesem Grund wird in dem vorliegenden Modell die Länge des Auswertintervalles $T_{fA}(f)$ der Wahrnehmungsgrenze nichtstationärer Schalle angepaßt. Heldmann gibt hierfür folgende Näherung an [14]:

$$(4.1) \quad T_{fA}(f) = \begin{cases} T_A \cdot 10 \cdot \frac{\Delta f_G(100\text{Hz})}{\Delta f_G(f)} & f \leq 5600\text{Hz} \\ T_A & f > 5600\text{Hz} \end{cases}$$

Dabei ist die Länge des Auswertintervalls T_A für große Frequenzen der entscheidende Parameter bei der Anpassung an das menschliche Gehör. Heldmann benutzt $T_A = 1$ msec [14].

Heldmann erläutert nicht näher, wie diese Näherung zustande kommt. Ein Vergleich von Gl (4.1) mit der Dimensionierung $T_A = 1$ msec und der Wahrnehmungsgrenze der Rauigkeit als ein Maß für die maximale Zeitauflösung des Gehörs [38] ergibt für Frequenzen $f \leq 3$ kHz eine gute Übereinstimmung. Für Frequenzen $f > 3$ kHz ist $T_{fA}(f)$ etwa um den Faktor zwei kleiner als die Wahrnehmungsgrenze der Rauigkeit.

4.1.5 Erkennung der Teiltöne

Die Lage der spektralen Maxima im eingeschwungenen FTT-Spektrum hängt nicht von der Transformationskonstanten $a(\omega)$ der FTT ab, vgl. Gl. (3.4). $a(\omega)$ beeinflusst lediglich die Steigung der Flanken [12], [13]. Das Spektrum eines Sinustones ist daher vollständig durch die Angabe von Frequenz und Pegel des Maximums bestimmt [47], [50]. Die Bestimmung von Teiltönen kann somit durch eine Maximumdetektion im Spektrum erfolgen. Die Art der Maximumdetektion hängt dabei entscheidend von der Absicht der Untersuchung ab [44].

In der vorliegenden Arbeit wird eine Maximumdetektion mit Schwellenkriterium nach Heinbach [12] verwendet, um Teiltöne in natürlichen Schallen aufzufinden. Bei dieser Art der Maximumbestimmung wird wie folgt vorgegangen:

- Bestimmung der lokalen Maxima
- Überprüfung des Schwellenkriteriums
- Untersuchung auf globale Maxima

Die Maximumdetektion mit Schwellenkriterium nach Heinbach wird anhand von Abb. 4.2 schematisch erläutert. In Abb. 4.2 ist das FTT-Spektrum eines Signals zu einem festen Zeitpunkt t_k dargestellt. Aufgetragen ist der Pegel über der Frequenz. Die Pfeile geben die Lage der lokalen Maxima wieder, die gestrichelten Linie geben den Bereich des Schwellenkriteriums an. Als Schwellenkriterium wird die Ausprägtheit eines Maximums herangezogen, welche

als die kleinere Pegeldifferenz ΔL zwischen einem lokalen Maximum und den unmittelbar angrenzenden lokalen Minima definiert ist [13].

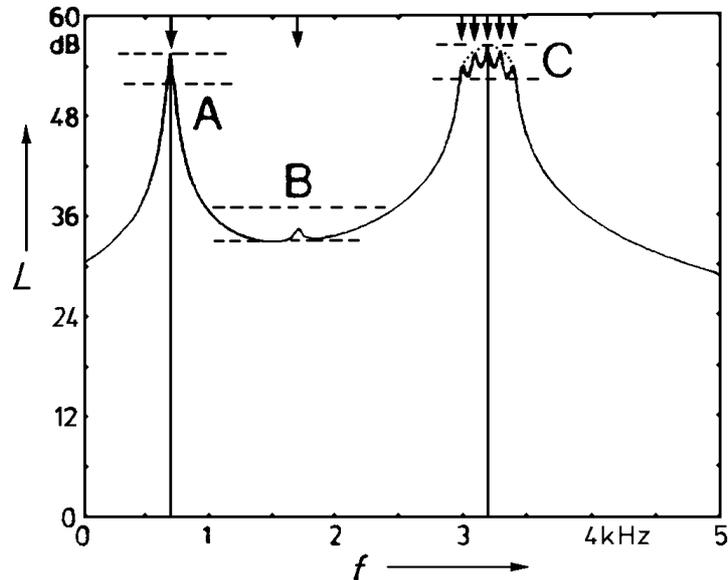


Abb. 4.2: Maximumdetektion mit Schwellenkriterium nach Heinbach [12], Erläuterungen siehe Text. Diagramm entnommen aus [12].

Region A in Abb. 4.2 stellt ein lokales Maximum dar. Das Schwellenkriterium ist erfüllt, d.h. die Voraussetzungen für einen TT sind gegeben. Region B hat ebenfalls ein lokales Maximum, jedoch wird hier kein TT ausgebildet, da das Schwellenkriterium nicht erfüllt ist. Bei C wird das Schwellenkriterium zwar nicht unmittelbar erfüllt, jedoch liegen die an das untersuchte lokale Maximum angrenzenden lokalen Maxima und Minima innerhalb des Bereiches, der durch das Schwellenkriterium vorgegeben ist. In diesem Fall spricht man für das höchste lokale Maximum von einem globalen Maximum. Globalen Maxima werden entsprechend den Erfahrungen über Spektraltonhöhen [56], [57], [58] ebenfalls TT zugeordnet.

Auf diese Weise erhält man das TTM für den festen Zeitpunkt t_k . Bei Durchführung des Verfahrens zu jedem Analysezeitpunkt gewinnt man das TTZM.

Das TTZM für das Beispielsignal aus Kapitel 3.9 ist in Abb. 4.3 dargestellt. Es wurde mit dem hier vorgestellten und implementierten TTZM-Algorithmus erstellt. Die Art der Darstellung ist identisch zu der des FFT-Spektrums, vgl. Abb. 3.6. Die Tonheit wird auf der Ordinate, die Zeit auf der Abszisse und der Pegel farblich codiert über der durch Zeit und Tonheit aufgespannten Ebene aufgetragen.

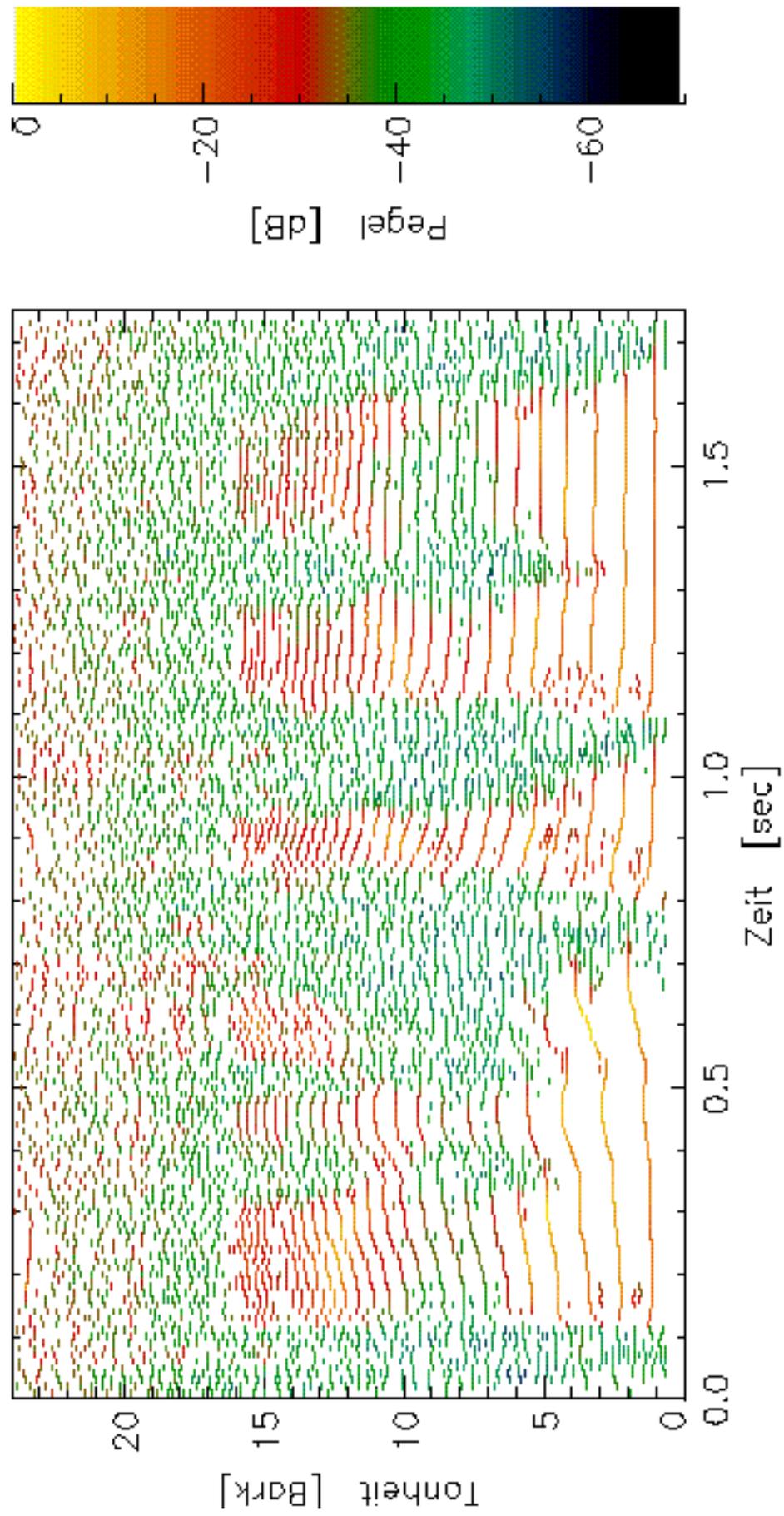


Abb 4.3: TTZM des Beispielsignals aus Kapitel 3.9.

Die Grundfrequenz und die Oberschwingungen des Sprachsignals können aus dem Diagramm abgelesen werden. Das ist ebenfalls im FTT- Spektrum (Abb. 3.6) möglich, jedoch sind die Pegelmaxima aus dem TTZM wesentlich genauer zu entnehmen. Zusätzlich treten einige Teiltonlinienverläufe, die einen kleinen Pegel besitzen, im TTZM (Abb. 4.3) deutlicher hervor als im FTT- Spektrum (Abb. 3.6). Die rauschhaften Anteile des FTT- Spektrums zeichnen sich im TTZM durch ungeordnete, statistisch verteilte Teiltöne aus. Diese rauschhaften Anteile sind in Abb. 4.3 und Abb. 3.6 oberhalb von ca. 16 Bark gut zu erkennen.

4.2 Vernetzung des TTZM zum Teiltonlinienzeitmuster

Beim Übergang vom TTZM zum Teiltonlinienzeitmuster (TTLZM) werden Teiltöne ähnlicher Frequenz für aufeinanderfolgende TTM prinzipiell so miteinander verbunden, daß 'Linien' entstehen, die sich durch das TTZM ziehen. Diese 'Linien' geben den zeitlichen Verlauf der einzelnen Teiltöne wieder. Lassen sich einzelne TT nicht in geeigneter Weise mit anderen TT aus dem vorangegangenen oder nachfolgenden TTM verbinden, so werden diese TT als rauschhaft betrachtet und vernachlässigt. Ebenso werden alle TT behandelt, die einer 'Linie' angehören, die eine gewisse Zeitdauer unterschreitet.

Dieser Mustererkennungsalgorithmus, der aus dem TTZM das TTLZM berechnet, bewirkt zwei Dinge:

- 1.) Es wird eine Rauschbefreiung durchgeführt. Dieses Ziel verfolgt u.a. Mumert [26].
- 2.) Nur die vom menschlichen Gehör grundsätzlich wahrnehmbaren 'Linien' werden im weiteren betrachtet. Das heißt die 'Linien', die eine frequenzabhängige Mindestlänge unterschreiten und deshalb nicht als Ton spezifischer Frequenz identifizierbar sind, werden nachfolgend vernachlässigt.

Der in diesem Modell verwendete Musterkennungsalgorithmus geht auf Heldmann [14] zurück und wird durch fünf Kriterien definiert⁹:

- Abstandskriterium
- erweitertes Abstandskriterium
- Maximalabstandskriterium
- Pegelkriterium
- Mindestlängenkriterium

Die einzelnen Kriterien werden im folgenden beschrieben. Dabei gelten folgende Vereinbarungen:

- f_t^i Diskrete Darstellung der Frequenz (in Bark) eines Teiltones zum Zeitpunkt t . Die Diskretisierung erfolgt mittels Gl. (3.6):
 $f_t^i = f_u + i \cdot \Delta z$. Diese Vereinbarung bedeutet implizit, daß die Frequenzen nach aufsteigender Größe geordnet sind.
- S_t^i Pegel des geglätteten FTT- Leistungsspektrums des Teiltones f_t^i zur Zeit t . Berechnet wird S_t^i nach Gl. (3.10) bzw. (3.18).
- (f_t^i, S_t^i) Teilton zum Zeitpunkt t mit Frequenz f_t^i und Pegel G_t^i .
- $\{(f_t^i, S_t^i)\}$ Satz aller Teiltöne, die das TTM zum Zeitpunkt t definieren.
- Δi Parameter um den maximal erlaubten Frequenzabstand Δf_{\max} (in Bark) zweier zeitlich aufeinanderfolgender Teiltöne zu charakterisieren: $\Delta f_{\max} = \Delta i \cdot \Delta z$.
- $l_{\min}(f)$ Frequenzabhängige Mindestlänge der Teiltonlinie.

Abstandskriterium

- „Ein Teilton des TTM zum Zeitpunkt t ist nur mit demjenigen Teilton zum Zeitpunkt $t+1$ zu verknüpfen, der den geringsten Frequenzabstand aufweist.“

Abb. 4.4 zeigt die Vernetzung des Teiltons $f_t^{i_0}$ zum Zeitpunkt t mit dem Teilton $f_{t+1}^{j_0}$ zum Zeitpunkt $t+1$. Die Zuordnung erfolgt auf diese Weise, weil der

⁹ Die Darstellung der Definition erfolgt anders als bei Heldmann. Das bei Heldmann implizit enthaltene Maximalabstandskriterium wird hier als eigenständiges Kriterium aufgefaßt. Dieser Umstand ändert nichts an der Aussage der Definition.

Frequenzabstand von $f_t^{i_0}$ zu $f_{t+1}^{j_0}$ geringer ist als der Frequenzabstand von $f_t^{i_0}$ zu $f_{t+1}^{j_1}$.

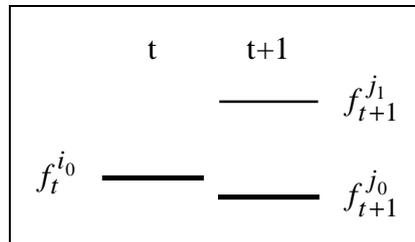


Abb. 4.4: Abstandskriterium.

Erweitertes Abstandskriterium

- „Ein Teilton zum Zeitpunkt $t+1$ ist nur mit demjenigen Teilton des Zeitpunktes t zu verbinden, der den geringsten Frequenzabstand besitzt.“

Abb. 4.5 verdeutlicht den Sinn des erweiterten Abstandskriteriums: Für $f_t^{i_0}$ ist $f_{t+1}^{j_1}$ die Linie mit dem geringsten Frequenzabstand, für $f_t^{i_1}$ ist es $f_{t+1}^{j_2}$. Während für $f_t^{i_1}$ auch das erweiterte Abstandskriterium zutrifft, d.h. für $f_{t+1}^{j_2}$ ist auch $f_t^{i_1}$ der nächste zeitliche Nachbar (die Linie mit dem geringsten Frequenzabstand), gilt dies für $f_t^{i_0}$ nicht. In diesem Fall ist $f_t^{i_1}$ der nächste zeitliche Nachbar zu $f_{t+1}^{j_0}$. Eine Vernetzung der Teiltöne findet daher nur im ersten Fall statt.

Das erweiterte Abstandskriterium sorgt so für eine Eineindeutigkeit der Zuordnung.

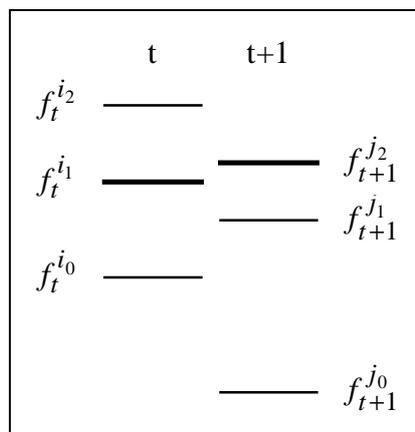


Abb. 4.5: Erweitertes Abstandskriterium.

Maximalabstandskriterium

- „Der maximale Frequenzabstand zweier Teiltöne zum Zeitpunkt t und $t+1$ darf eine bestimmte Maximalgröße $\Delta i \cdot \Delta z$ nicht überschreiten.“

Abb. 4.6 zeigt die Erfüllung des Maximalabstandskriteriums: $|f_{t+1}^{j_0} - f_t^{i_0}| \leq \Delta i \cdot \Delta z$. Daher werden die beiden Teiltonlinien verknüpft. In Abb. 4.7 dürfen die Teiltonlinien nicht verknüpft werden, weil gilt: $|f_{t+1}^{j_0} - f_t^{i_0}| > \Delta i \cdot \Delta z$. Das Maximalabstandskriterium ist somit nicht erfüllt.

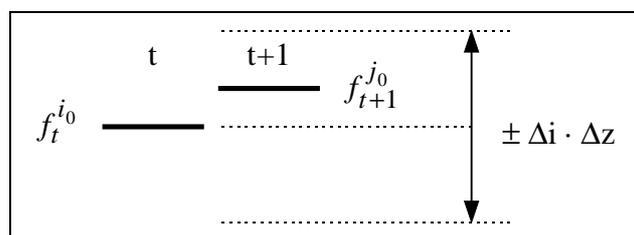


Abb. 4.6: Maximalabstandskriterium erfüllt.

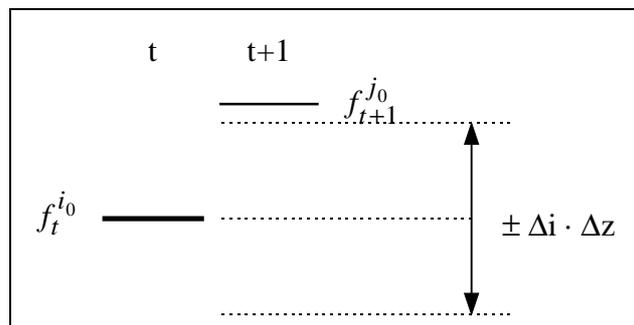


Abb. 4.7: Maximalabstandskriterium nicht erfüllt.

Pegelkriterium

- „Bei gleichem Frequenzabstand dominiert der Teilton mit dem größeren Pegel.“

Abb. 4.8 zeigt die Anwendung des Pegelkriteriums. Der Frequenzabstand von $f_t^{i_0}$ zu $f_{t+1}^{j_0}$ bzw. $f_{t+1}^{j_1}$ ist gleich groß. $f_t^{i_0}$ wird mit demjenigen Teilton zum Zeitpunkt $t+1$ verbunden, der den größeren Pegel, $\max(S_{t+1}^{j_0}, S_{t+1}^{j_1})$, aufweist.

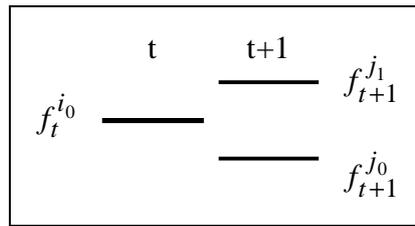


Abb. 4.8: Pegelkriterium.

Mindestlängenkriterium

- „Eine Teiltonlinie hat eine frequenzabhängige Mindestlänge $l_{min}(f)$ aufzuweisen.“

Wie bereits angedeutet, gewährleistet dieses Kriterium, daß nur solche ‘Linien’ detektiert werden, die aufgrund ihrer Länge prinzipiell vom Gehör wahrnehmbar sind. In dem vorliegenden Modell wird $l_{min}(f)$ nach einem Vorschlag von Heldmann [14] approximiert. Gl. (4.2) ist nach Daten von [58] gefittet und wird auf ähnlich Weise wie Gl. (4.1) ausgedrückt:

$$(4.2) \quad l_{min}(f) = \begin{cases} l_m \cdot \frac{\Delta f_G(100\text{Hz})}{\Delta f_G(f)} & f \leq 5600\text{Hz} \\ l_m / 10 & f > 5600\text{Hz} \end{cases}$$

Der Parameter l_m in Gl. (4.2) bedeutet die zeitliche Mindestlänge der Teiltonlinien für $l_{min}(100\text{ Hz})$. Dieser Parameter kann in der Implementation des TTLZM vom Benutzer frei gewählt werden. Gl. (4.2) berücksichtigt, daß für große Frequenzen die frequenzabhängige Mindestlänge $l_{min}(f)$ abnimmt. Heldmann [14] wählt als Faktor zwischen der frequenzabhängigen Mindestlänge für kleine und große Frequenzen den Faktor 10. Die Daten für diese Approximation entstammen [58]. Dabei wird das frequenzabhängige Zeitauflösungsvermögen des Gehöres durch die reziproke Frequenzgruppenbreite berücksichtigt.

4.2.1 Numerische Beschreibung des Vernetzungsalgorithmus

Die numerische Umsetzung der zuvor dargestellten Kriterien erfolgt hier etwas anders als bei Heldmann [14]. Änderungen werden vorgenommen, um den Algorithmus strukturiert und geschlossen darstellen zu können. Dadurch fällt die implizite Beschreibung des erweiterten Abstandskriteriums bei Heldmann zugunsten einer einfacheren Darstellung fort. Der Algorithmus läuft in folgenden Schritten ab:

1. Schritt

Für jeden TT (f_t^i, S_t^i) eines TTM $\{(f_t^i, S_t^i)\}$ wird der Teilton (f_{t+1}^j, S_{t+1}^j) des folgenden TTM $\{(f_{t+1}^j, S_{t+1}^j)\}$ gesucht, der den geringsten Frequenzabstand gfa aufweist. Diese Umsetzung des Abstandskriteriums geschieht durch Vergleichen der Werte von f_t^i mit f_{t+1}^j . Im Falle einer *Entartung* besitzen zwei TT (f_{t+1}^j, S_{t+1}^j) , bzw. $(f_{t+1}^{j+1}, S_{t+1}^{j+1})$ den gleichen, geringsten Frequenzabstand $|-gfa|$, bzw. gfa zu (f_t^i, S_t^i) .

2. Schritt

Es wird überprüft, ob das Maximalabstandskriterium erfüllt ist, d.h. ob $|gfa| \leq d_{max}$ eine wahre Aussage ist. Wenn die Aussage wahr ist, wird mit Schritt drei fortgefahren, andernfalls zum Unterprogramm ‘Linie beenden’ verzweigt.

3. Schritt

Das erweiterte Abstandskriterium wird durch folgende Bedingungen überprüft:

$$(4.3) \quad \text{für } f_{t+1}^j < f_t^i \quad \frac{|f_t^i - f_t^{i-1}|}{2} < |gfa|$$

$$(4.4) \quad \text{für } f_{t+1}^j > f_t^i \quad \frac{|f_t^{i+1} - f_t^i|}{2} < |gfa|$$

Wird eine dieser Bedingungen erfüllt, so wird mit dem nächsten Schritt fortgefahren, ansonsten zum Unterprogramm ‘Teiltonlinie beenden’ verzweigt.

Im Falle der Entartung muß für den TT mit der kleineren Frequenz (f_{t+1}^j, S_{t+1}^j) Bedingung (4.3), für den TT mit der größeren Frequenz $(f_{t+1}^{j+1}, S_{t+1}^{j+1})$ Bedingung (4.4) angewandt werden. Ist nur eine der beiden Bedingungen erfüllt, so wird der TT, für den die Aussage falsch ist, nicht weiter berücksichtigt und nur für denjenigen TT, der die wahre Aussage liefert, wird das Verfahren mit Schritt 5 weitergeführt. Dafür muß der ‘gültige’ TT unter (f_{t+1}^j, S_{t+1}^j) abgespeichert werden (falls er bisher unter $(f_{t+1}^{j+1}, S_{t+1}^{j+1})$ abgespeichert war).

Für $f_{t+1}^j = f_t^i$, d.h. $gfa = 0$, wird Schritt 3 und 4 übergangen und gleich Schritt 5 ausgeführt werden.

Schritt 4:

Dieser Schritt ist ausschließlich für entartete TT relevant. Es wird das Pegelkriterium angewandt und der TT mit dem größeren Pegelwert, $\max(S_{t+1}^j, S_{t+1}^{j+1})$,

berücksichtigt und als (f_{t+1}^j, S_{t+1}^j) abgespeichert. Der TT mit dem kleineren Pegelwert wird vernachlässigt.

Schritt 5:

Alle Bedingungen sind erfüllt, d.h. der TT (f_{t+1}^j, S_{t+1}^j) ist die Fortsetzung zum TT (f_t^i, S_t^i) . War der TT (f_t^i, S_t^i) bereits Bestandteil einer Teiltonlinie, so wird der TT (f_{t+1}^j, S_{t+1}^j) einfach hinten an die schon bestehende Teiltonlinie angehängt. Ansonsten wird eine neue Teiltonlinie mit den TT (f_t^i, S_t^i) und (f_{t+1}^j, S_{t+1}^j) begonnen. Anschließend wird der Laufindex i der TT (f_t^i, S_t^i) des TTM $\{(f_t^i, S_t^i)\}$ um eins erhöht und für den nächsten TT wieder mit Schritt 1 begonnen.

Sind alle TT des TTM $\{(f_t^i, S_t^i)\}$ sukzessive abgearbeitet, wird zum nächsten TTM $\{(f_{t+1}^i, S_{t+1}^i)\}$ übergegangen und erneut mit Schritt 1 fortgefahren. Diese Schleife wird solange durchgeführt, bis das letzte TTM bearbeitet worden ist. Am Schluß wird noch einmal abschließend zum Unterprogramm 'Teiltonlinie beenden' verzweigt, um alle noch nicht beendeten Teiltonlinien zu beenden und im Falle der Erfüllung des Mindestlängenkriteriums abzuspeichern.

Teiltonlinie beenden

Wenn das Ende einer Teiltonlinie aufgetreten ist, wird die Länge der Teiltonlinie mit der frequenzabhängigen Mindestlänge $l_{min}(f)$ verglichen. Dafür wird die mittlere Frequenz (arithmetisches Mittel) \bar{f} der TT dieser Teiltonlinie berechnet und als repräsentativ für die gesamte Teiltonlinie angesehen. Wenn die Länge der Teiltonlinie kleiner als $l_{min}(\bar{f})$ nach Gl. (4.2) ist, dann wird die Teiltonlinie im weiteren nicht mehr betrachtet, ansonsten werden die einzelnen TT der Teiltonlinie abgespeichert. Der Algorithmus beginnt wieder bei Schritt 1, nachdem der Laufindex i des TTM $\{(f_t^i, S_t^i)\}$ um eins erhöht worden ist.

Wird ein einzelner, rauschhafter TT detektiert, dann wird der TT nicht weiter betrachtet, der Laufindex i der TTM $\{(f_t^i, S_t^i)\}$ wird um eins erhöht und die Berechnung bei Schritt 1 fortgeführt.

4.2.2 Beispiel für ein TTLZM

Ein Beispiel für ein TTLZM, berechnet wie im Abschnitt 4.2.1 erläutert, zeigt Abb. 4.9. Dort ist das TTLM (Abb. 4.3) des Beispielsignals aus Kapitel 3.9 mit

folgenden Parametern zum TTLZM vernetzt worden: $l_m = l_{min}(100 \text{ Hz}) = 40 \text{ msec}$, maximaler Frequenzabstand $\Delta f \cdot \Delta z = 0,24 \text{ Bark}$.

In Abb. 4.9 wird die Darstellungsweise, die bereits aus Abb. 4.3 und 3.6 bekannt ist, beibehalten. Der Pegel ist farblich kodiert über der Zeit und der Tonheit aufgetragen. Dabei werden in allen drei Abbildungen exakt die gleichen Farben für gleiche Pegelwerte verwendet.

Beim Vergleich des TTLZM (Abb. 4.9) mit dem TTLM (Abb. 4.3) fällt auf, daß die rauschhaften TT durch die Vernetzung entfernt werden, und so eine Rauschbefreiung durchgeführt wird. Diesen Effekt benutzt u.a. Mummert [26]. Die Rauschbefreiung ist im weiteren für die Bestimmung der spektralen und virtuellen Tonhöhen wichtig, da rauschhafte spektrale Komponenten keinen Einfluß auf die Tonhöhenwahrnehmung oder die Klangfarbe des untersuchten Signals haben. Im TTLZM Abb. 4.9 sind ausschließlich die tonalen Komponenten des untersuchten Signals vorhanden. Deshalb eignet es sich gut als Ausgangspunkt für die Bestimmung der spektralen und virtuellen Tonhöhen mit dem Algorithmus von Terhardt et al. [50], vgl. auch Kapitel 5.

4.3 Überprüfung des Konzepts des TTLZM

Um das Konzept des TTLZM zu überprüfen, wird eine Resynthese der analysierten Signale durchgeführt. Mittels eines Analyse durch Syntheseverfahrens (vgl. z.B. [32]) werden die Parameter des Modells so bestimmt, daß sie die Eigenschaften des Gehörs so gut wie möglich nachbilden. Die Gültigkeit des Modells ergibt sich dann aus der Reproduzierbarkeit der analysierten Signale im Hinblick auf Klangfarbe und Tonhöhenwahrnehmung.

4.3.1 Resynthese

Die Resynthese von Signalen aus dem TTLZM oder dem TTZM geht auf [13], [14] und [27] zurück. Die grundsätzliche Idee dabei ist, diskrete ‘Sinusgeneratoren’ jeweils mit den Frequenzen und Amplituden der Teiltöne anzuregen, die im TTLZM bzw. im TTLM vorhanden sind. Die Summe aller dieser Sinussignale definiert das resynthetisierte Signal zu dem betrachteten Zeitpunkt. Bei sukzessiver Durchführung dieses Verfahrens für alle aufeinanderfolgenden, diskreten Zeitpunkte wird das vollständig resynthetisierte Signal gewonnen. Dies ist auch das Prinzip der Fourier-Synthese (vgl. Anhang A und z.B. [5], [35], [37]).

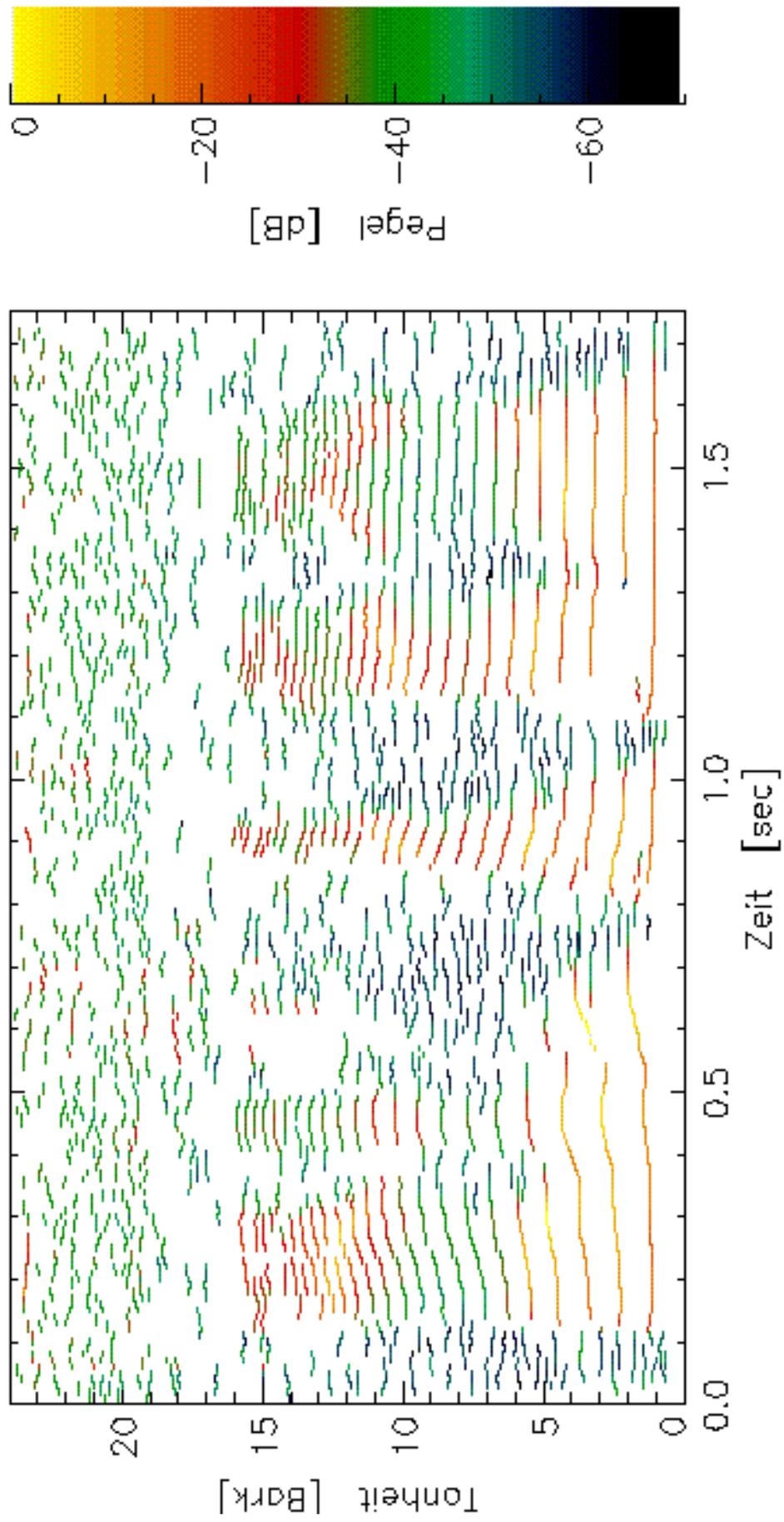


Abb 4.9: TTLZM des Beispielsignals aus Kapitel 3.9.

Im signaltheoretischen Sinn ist für dieses Verfahren die Phaseninformation der einzelnen Sinuskomponenten unerlässlich. Aus diesem Grunde muß die Phaseninformation, die durch die Bildung des Leistungsspektrums (vgl. Kapitel 2.6) verloren gegangen ist, ergänzt werden. Dazu wird zu Beginn einer Teiltonlinie die Phase zu Null gesetzt. Für jeden weiteren Teilton der gleichen Teiltonlinie wird die Phase anschließend solange fortgeführt, bis die Teiltonlinie endet. Ändert die Teiltonlinie von Teilton (f_t^i, S_t^i) zu Teilton (f_{t+1}^i, S_{t+1}^i) ihre Frequenz, so muß die Phase entsprechend korrigiert werden, damit kein Sprung im resynthetisierten Zeitsignal auftritt. Um Diskretisierungseffekte zu vermeiden, wird die Summe aller Sinusgeneratoren, d.h. die Amplitude des resynthetisierten Zeitsignals zu den 'Abtastzeitpunkten', an denen die Phase von Teilton zu Teilton übergeben wird, mittels kubischer Splines geglättet. Der in dieser Arbeit verwendete Algorithmus zur Glättung des Zeitsignals mittels kubischer Splines entstammt [30].

4.3.2 Berücksichtigung der Fensterfunktionen der FFT bei der Resynthese

Die FFT berücksichtigt die frequenzabhängigen Laufzeiten spektraler Komponenten auf der Basilarmembran durch die frequenzabhängige Analysefensterlänge und -form (vgl. Kap. 3). Durch Laufzeitunterschiede verschiedener Frequenzen ist somit eine 'Gleichzeitigkeit' der Analyse nicht mehr gegeben, d.h. im Zeitsignal gleichzeitig beginnende, unterschiedliche Frequenzen werden im Spektrum nicht gleichzeitig detektiert [27], [14]. Die tiefere Frequenz wird später wahrgenommen als die höhere, und zwar umso mehr, je größer die Differenz zwischen den beiden Frequenzen ist. Dies entspricht den Eigenschaften des Gehörs [57].

Bei einer Resynthese des Originalsignals aus dem TTLZM muß dieser Effekt berücksichtigt werden, um das Originalsignal exakt wieder herstellen zu können. Die tiefen Frequenzen müssen um eine frequenzabhängige Zeitspanne gegenüber den hohen Frequenzen in der Zeit zurückgeschoben werden. Diese frequenzabhängige Zeitspanne ist nach [27] durch die Fensterfunktion der FFT definiert. Sie ergibt sich als die Zeitdifferenz zwischen dem Maximum der verwendeten Fensterfunktion für die untersuchte Frequenz gegenüber dem Maximum der gleichen Fensterfunktion für eine Referenzfrequenz von z.B. 20 Hz. Mit diesem Zusammenhang wird die Verschiebung der spektralen Komponenten vorgenommen.

4.3.3 Hörversuche

Im Rahmen dieser Arbeit wird die Resynthese wie oben beschrieben ausgeführt, um die Funktion der zuvor beschriebenen Algorithmen zu überprüfen. Insbesondere soll die Programmierung der Algorithmen getestet werden. Dafür werden verschiedene Testsignale analysiert und resynthetisiert. Der quantitative Vergleich von reinen und amplituden- bzw. frequenzmodulierten Sinussignalen liefert gute Ergebnisse. Desweiteren wird ein qualitativer Hörvergleich mit Sprachsignalen vorgenommen, in dem die Ähnlichkeit des resynthetisierten mit dem Originalsignal verglichen wird. Dieser Hörtest soll nicht als eigene Messung, Prüfung oder Anpassung der verwendeten Literaturdaten verstanden werden, sondern lediglich als Test für die korrekte Umsetzung der hier verwendeten Algorithmen auf einem PC oder einer Workstation. Es werden vier Versuchspersonen danach befragt, ob sie einen Unterschied zwischen dem Original und der Resynthese erkennen und wenn ja, ob die tonalen Eigenschaften wie Tonhöhe und Klangfarbe unterschiedlich sind. Im Gegensatz zu [14] und [27] können alle Versuchspersonen einen Unterschied feststellen, der sich in zusätzlichem Rauschen und hochfrequenten Störgeräuschen des resynthetisierten Signals bemerkbar macht. Die Tonhöhe des Sprachsignals ist jedoch in beiden Signalen so gut wie identisch. Die zusätzlich im resynthetisierten Signal auftretenden, hochfrequenten Geräusche und das Rauschen sind von der Art der Resynthese abhängig. Je nach verwendeter Resynthesemethode ändert sich der subjektive Eindruck des Rauschens und der hochfrequenten Anteile. Die tonalen Komponenten bleiben jedoch identisch. Da Heldmann [14] den von ihm benutzten Resynthese-Algorithmus nicht näher beschreibt, können keine weiteren Vergleiche zwischen Heldmanns und meinen Ergebnissen angestellt werden. Es ist zu vermuten, daß Heldmann eine andere Anpassung der Amplitudenwerte bei dem Übergang von einem zum nächsten Teilton benutzt. Da die FTT- und TTZM- Analyse speziell für die Bestimmung der Tonhöhe optimiert ist und die Tonhöhe identisch resynthetisiert werden kann, wird die Annahme gemacht, daß die Algorithmen korrekt auf den PC bzw. der Workstation umgesetzt wurden. Weitere Tests des hier vorgestellten Modells und seiner Implementation werden in den nächsten Kapiteln vorgestellt.

5 MODELL DER VIRTUELLEN TONHÖHE

Wie in der Einleitung erwähnt, existieren zahlreiche Modelle zur Berechnung der virtuellen Tonhöhe. Im Rahmen dieser Arbeit wird das Modell nach Terhardt [44] von 1979 und die Erweiterung von Terhardt, Stoll und Seewann [50] aus dem Jahre 1982 benutzt. Dieser Ansatz wird ausgewählt, weil er neben der Berechnung der virtuellen und spektralen Tonhöhen auch eine Bewertung der Ausgeprägtheit von Tonhöhen ermöglicht. Auf diese Weise ist eine Abschätzung der Tonhöhenwahrnehmung, virtuell oder spektral, verfügbar. Das Modell eignet sich daher zur Tonhöhenbeurteilung beliebiger Geräusche.

5.1 Modell nach Terhardt et al. 1982

Das Modell nach Terhardt et al. besteht aus fünf Schritten, siehe Abb. 5.1:

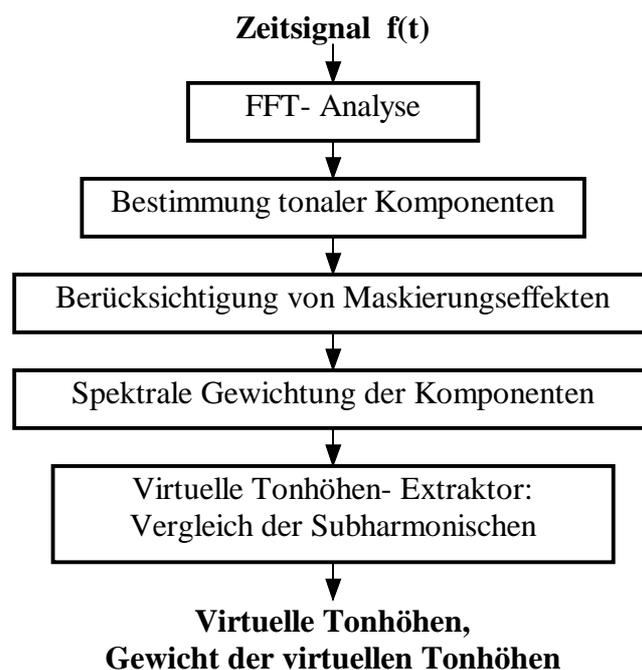


Abb. 5.1: Modell nach Terhardt et al. [50], Erläuterungen siehe Text.

Im ersten Schritt führt Terhardt eine FFT mit 400 Samples für das Zeitsignal $f(t)$ durch, um das Leistungsspektrum des Signals zu berechnen. Der untersuchte Spektralbereich erstreckt sich von 20-5000 Hz, d.h. die Frequenzauflösung beträgt 12,5 Hz. Bei einer verwendeten Abtastrate von 10 kHz muß das Zeitfenster hierfür zu 80 msec Länge gewählt werden. Als Fensterfunktion wird das

Hanningfenster benutzt. Aus den Ausführungen der Kapitel 3 und 4 ergibt sich, daß dieser Ansatz nicht sehr gehörgerecht ist.

Anschließend erfolgt eine Bestimmung der tonalen Komponenten. Hierfür werden die lokalen Maxima im FFT-Spektrum ermittelt und mit den angrenzenden Spektralkomponenten verglichen [50], das genaue Verfahren soll hier nicht weiter betrachtet werden. Durch eine zusätzliche Frequenzapproximation gelingt es Terhardt, die tonalen Komponenten auf eine Genauigkeit von ± 1 Hz zu bestimmen. Terhardt [50] bezeichnet diese Vorgehensweise als „eine unter vielen anderen Möglichkeiten, die tonalen Komponenten zu erhalten. Je nach den Zielen einer Untersuchung können andere Methoden besser geeignet sein, tonale Komponenten zu bestimmen.“

Die gegenseitigen Maskierungseffekte der tonalen Komponenten berücksichtigt Terhardt im dritten Schritt. Zum einen werden die Überschußpegel LX_μ (vgl. Kapitel 5.3.2) berechnet: Komponenten mit einem Überschußpegel kleiner als 0 dB sind nicht hörbar und werden im weiteren vernachlässigt. Zum anderen wird die Tonhöhenverschiebung der spektralen Tonhöhen berücksichtigt, die von gegenseitiger Beeinflussung der tonalen Komponenten herrührt.

Auf die so gewonnenen Überschußpegel der spektralen Tonhöhen wird eine Gewichtsfunktion angewendet, die den Beitrag der spektralen Tonhöhen auf das Zustandekommen von virtuellen Tonhöhen berücksichtigt. Diese Funktion hängt sowohl vom Überschußpegel LX_μ als auch von der spektralen Lage der untersuchten Komponenten ab. Auf diese Weise wird der Überschußpegel LX_μ in das Gewicht der spektralen Tonhöhe WS_μ (vgl. Kapitel 5.3.4) transformiert.

Im letzten Schritt berechnet Terhardt die virtuelle Tonhöhe aus den spektralen Tonhöhen und ihren Gewichten. Dabei werden zu jeder Spektraltonhöhe die Subharmonischen gebildet und miteinander verglichen. Diesen Algorithmus bezeichnet Terhardt als ‘*Virtuelle Tonhöhen-Extraktor*’, VTE [44]. Die Funktionsweise des VTE soll Abb. 5.2 schematisch verdeutlichen.

Im oberen Teil der Abb. 5.2 ist ein Residualklang dargestellt, aus dem die ersten beiden Harmonischen (1. Harmonische = Grundfrequenz) entfernt wurden. Die obere Skala gibt die Frequenz der spektralen Komponenten des Residualklanges in Hz an, die Zahlen darunter zählen die Harmonischen ab. In der Strichlänge ist der Pegel der Komponenten kodiert. Darunter sind die spektralen Tonhöhen der Komponenten in äquivalenter Frequenz (pu) dargestellt. Die Strichlänge symbolisiert das Gewicht der spektralen Tonhöhen WS_μ . In der unteren Hälfte des Diagrammes ist beispielhaft für die untersten vier spektralen Komponenten des Residualklanges eine horizontale Gerade dargestellt, auf der die Positionen der

ersten acht Subharmonischen durch einen Punkt markiert sind. Zur virtuellen Tonhöhe (Pfeile auf der Skala ganz unten) gelangt man, indem die Übereinstimmungen der Subharmonischen aufgesucht und markiert werden. In diesem Beispiel wird die virtuelle Tonhöhe von 200 pu durch alle 4 Subharmonischen hervorgerufen. Diese Tonhöhe steht also für die dominanteste virtuelle Tonhöhe. Bei 100 und 400 pu liegen jeweils 2 Subharmonische. Das heißt auch hier wird eine virtuelle Tonhöhe detektiert. Die Ausgeprägtheit dieser virtuellen Tonhöhe ist jedoch wesentlich kleiner als die Ausgeprägtheit der virtuellen Tonhöhe bei 200 pu, angedeutet durch die Länge der Pfeile.

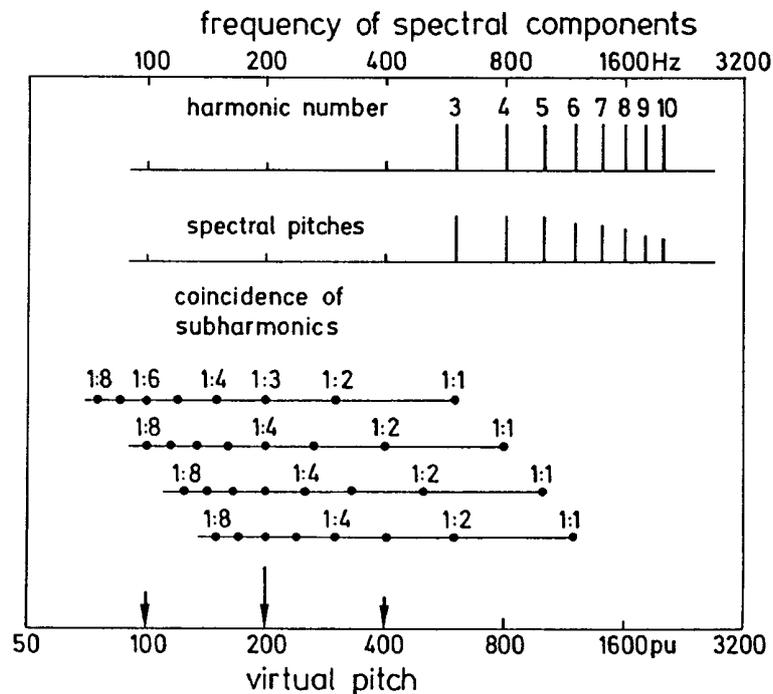


Abb. 5.2: Funktionsweise des virtuelle Tonhöhen- Extraktors. Darstellung nach [58], Erläuterung siehe Text.

Terhardt bemängelt in dem Modell nach Abb. 5.1 die schlechte Übereinstimmung der FFT- Analyse (Schritt 1) und der Extraktion tonaler Komponenten (Schritt 2) mit den Eigenschaften des menschlichen Gehörs. Insbesondere wird im ersten Schritt in Abb. 5.1 mit einer konstanten Frequenzauflösung Δf von 12,5 Hz gearbeitet. Dem Gehör wird jedoch eine konstante Güte q , d.h. eine konstante relative Frequenzauflösung $q = \Delta f/f$, eher gerecht [50].

5.2 Modifiziertes Modell

Kapitel 5.2 gibt eine Übersicht des Modells zur Berechnung virtueller Tonhöhen nach Terhardt et al. In diesem Kapitel wird ein neuer Tonhöhenextraktionsalgorithmus entwickelt, der die Erkenntnisse der FTT (Kapitel 3), des TTZM (Kapitel 4) und des VTE nach Terhardt [44], [50] (Kapitel 5.1) berücksichtigt.

5.2.1 Vorverarbeitung

An die Stelle der FFT tritt eine FTT und der Algorithmus zur Bestimmung der tonalen Komponenten wird durch das TTZM, bzw. das TTLZM ersetzt. Dadurch ist eine gehörangepaßte Vorverarbeitung gewährleistet, die gute Eingangsvoraussetzungen für die weitere Verarbeitung bietet.

Die zeitabhängigen Ausgangsgrößen des TTZM bzw. des TTLZM sind die Teiltonfrequenzen $f(t)$ und ihre Pegel $L(t)$. Die folgenden Ausführungen sind zunächst auf einen beliebigen, aber festen Zeitpunkt $t = t_k$ beschränkt: $f := f(t_k)$ und $L = L(t_k)$. Die tonalen Komponenten (bzw. Teiltöne des TTZM) seien nach aufsteigender Frequenz geordnet und durch den Laufindex μ abgezählt: $f_\mu < f_{\mu+1}$; L_μ ; $\mu = 1 \dots N$. Dabei ist N die Anzahl der vorhandenen Teiltöne.

5.2.2 Überschußpegel

Der *Überschußpegel* LX_μ ist ein Maß für die gegenseitige Maskierung der tonalen Komponenten. Er ist die Differenz des Schalldruckpegels einer einzelnen Komponente L_μ und dem Schalldruckpegel, welcher durch die Summe aller anderen Komponenten am Ort der untersuchten Frequenz f_μ durch Maskierung entsteht [44].

Anhand von Abb. 5.3 sollen die Konsequenzen der Definition von LX_μ für $N = 2$ verdeutlicht werden. Dargestellt ist die teilweise Maskierung eines reinen Tones mit der Frequenz f_μ , bzw. z_μ (z_μ ist die Transformation von f_μ in die Barkskala) durch einen weiteren Ton der Frequenz f_ν . Aufgetragen ist das Erregungsmuster (vgl. z.B. [44], [57], [58]) für den reinen Ton der Frequenz f_ν über der Barkskala in zwei Fällen: a.) $f_\mu > f_\nu$ und b.) $f_\mu \leq f_\nu$.

Der *Erregungspegel* $L_{E\nu}(f_\mu)$, d.h. die Erregung aufgrund der Frequenzkomponente f_ν am Ort der Frequenz f_μ , in Abb. 5.3 als $L_{\nu\mu}$ bezeichnet, ergibt sich demnach zu

$$(5.1) \quad L_{E\nu}(f_\mu) = L_\nu - s \cdot (z_\nu - z_\mu) = L_{\nu\mu}.$$

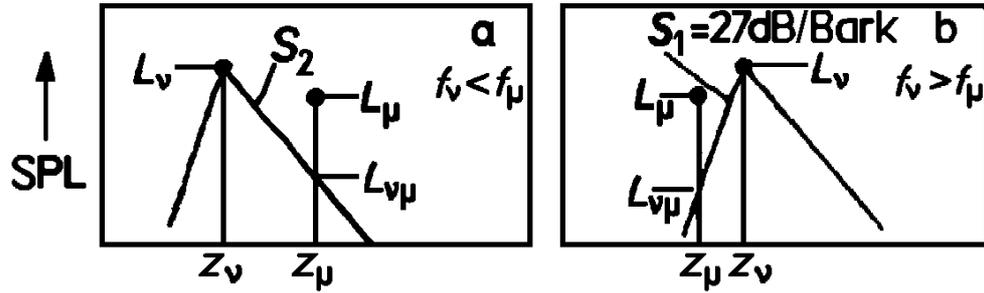


Abb. 5.3: Erregungspegelmuster. Aufgetragen ist das Erregungspegelmuster der Frequenz z_ν über der Frequenz in Bark nach [44]. Durch die Komponente z_ν mit dem Pegel L_ν wird am Ort der Komponente z_μ ein Erregungspegel von $L_{\nu\mu}$ hervorgerufen. Durch diese Maskierung wird der Pegel L_ν der Komponente z_ν um den Wert $L_{\nu\mu}$ verringert, d.h. der Überschußpegel LX_μ ergibt sich zu: $LX_\mu = L_\mu - L_{\nu\mu}$.

In Gl. (5.1) bedeutet s die Steigung, die in Abb. 5.3 als S_1 (für $f_\mu \leq f_\nu$) bzw. S_2 (für $f_\mu > f_\nu$) bezeichnet wird. Sie kann als geschlossener Ausdruck folgendermaßen angegeben werden:

$$(5.2) \quad s = \begin{cases} 27 \left[\frac{dB}{Bark} \right] & f_\mu \leq f_\nu \\ \left\{ -24 - \left(\frac{0.23[kHz]}{f_\nu} \right) + \left(\frac{0.2 \cdot L_\nu}{[dB]} \right) \right\} \left[\frac{dB}{Bark} \right] & f_\mu > f_\nu \end{cases}$$

Diese Approximation beruht auf Messungen des Erregungspegelmusters [44]. Zu beachten ist dabei die Frequenz- und Pegelabhängigkeit der Steigung für $f_\mu > f_\nu$. Im Fall einer größeren Anzahl von Komponenten als $N = 2$ muß die Summe über die Erregungspegel aller Komponenten f_ν am Ort der Frequenz f_μ erfolgen. Dabei wird $\nu = \mu$ bei der Aufsummierung übergangen.

Um Erregungspegel aufzuaddieren, müssen die entsprechenden Intensitäten summiert werden. Der Überschußpegel LX_μ ergibt sich zu:

$$(5.3) \quad LX_\mu = L_\mu - 20 \cdot \log \left(\sum_{v=1, v \neq \mu}^N 10^{L_{E_v}(f_\mu)/(20[dB])} \right) [dB].$$

Der so definierte Überschußpegel berücksichtigt die Ruheshchwelle des menschlichen Gehörs nicht. Daher können die Überschußpegel, die für verschiedene Frequenzen f_μ berechnet werden, nicht absolut miteinander verglichen werden. Um Abhilfe zu schaffen berücksichtigt Terhardt [50] die Ruheshchwelle $L_{Th}(f_\mu)$ bei der Berechnung des Überschußpegels LX_μ :

$$(5.4) \quad LX_{\mu} = L_{\mu} - 10 \cdot \log \left[\left(\sum_{v=1, v \neq \mu}^N 10^{L_{E_v}(f_{\mu})/(20[dB])} \right)^2 + I_{N_{\mu}} + 10^{L_{Th}(f_{\mu})/(10[dB])} \right] [dB]$$

Dafür muß die Ruheschwelle $L_{Th}(f_{\mu})$ für die Frequenz f_{μ} quantifiziert werden. Das geschieht in dem hier vorliegenden Modell mittels einer Näherung von [50] folgendermaßen:

$$(5.5) \quad L_{Th}(f_{\mu}) = \left[3,64 \cdot \left(\frac{f_{\mu}}{[kHz]} \right)^{-0,8} - 6,5 \cdot \exp \left(-0,6 \cdot \left(\frac{f_{\mu}}{[kHz]} - 3,3 \right)^2 \right) + 0,001 \cdot \left(\frac{f_{\mu}}{[kHz]} \right)^4 \right] [dB]$$

In Gl (5.4) ist die Rauschintensität $I_{N_{\mu}}$ enthalten. Dieser Summand berücksichtigt die Maskierung von f_{μ} durch die rauschhaften Anteile des Spektrums in der Frequenzgruppe um die untersuchte Frequenz: $[f_{\mu}-0,5 \text{ Bark}; f_{\mu}+0,5 \text{ Bark}]$. Hierauf wird im Abschnitt 5.2.2.1 näher eingegangen.

Wird der Überschußpegel LX_{μ} für alle Komponenten, d.h. $\mu = 1 \dots N$ berechnet, so erhält man das *Überschußpegelmuster* [44]. Dieses enthält nur die R Komponenten ($R \leq N$) mit $LX_{\mu} > 0$. Alle anderen Komponenten besitzen keine Bedeutung für das menschliche Gehör und werden deshalb im folgenden nicht weiter berücksichtigt. Dadurch verringert sich die Anzahl der für die folgenden Berechnungen erforderlichen Komponenten in der Regel erheblich.

Ein Beispiel für ein Überschußpegelmuster zeigt Abb. 5.9b. Abb. 5.9 soll die grundsätzliche Vorgehensweise bei der Bestimmung von virtuellen Tonhöhen verdeutlichen. Sie ist aus [50] entnommen, aber abgewandelt dargestellt. Bei [50] stellt Abb. 5.9a die tonalen Komponenten eines FFT-Leistungsspektrums dar. Grundsätzlich besteht aber kein Unterschied zum hier verwendeten TTLZM als Eingangsgröße für die Berechnung des Überschußpegelmuster.

In Abb. 5.9 wird ein Sprachsignal des (stationären) Vokals /a/ zu einem festen Zeitpunkt dargestellt. Dieses Beispiel steht in keiner Beziehung zum Beispiel aus Kapitel 3.9 und soll lediglich quantitativ die einzelnen Schritte der Analyse beschreiben.

5.2.2.1 Berechnung der Rauschintensität $I_{N_{\mu}}$

Die Rauschintensität $I_{N_{\mu}}$ repräsentiert diejenige Energie des Spektrums innerhalb der Frequenzgruppe um z_{μ} , die keiner tonalen Komponente zugeordnet werden kann [50].

Da der hier verwendete Extraktionsalgorithmus zur Bestimmung der tonalen Komponenten ein anderer ist als der von Terhardt verwendete, muß die Berechnung der Rauschintensität dem hier benutzten Extraktionsalgorithmus an-

gepaßt werden: In dem vorliegenden Modell werden die tonalen Komponenten durch einen Maximumdetektionsalgorithmus mit Schwellenkriterium und anschließender Vernetzung der Teiltöne bestimmt (TTLZM). Es bietet sich daher an, für die detektierten tonalen Komponenten f_μ ($\mu = 1..N$) des TTLZM jeweils die Frequenzgruppe um f_μ im FTT-Spektrum zu betrachten. Für diejenigen Frequenzen, die das Schwellenkriterium für eine beliebige tonale Komponente aus dieser Frequenzgruppe unterschreiten, werden die Intensitäten aufaddiert. Diese Summe stellt die Rauschintensität $I_{N\mu}$ dar.

Nach Terhardt kann $I_{N\mu}$ für stationäre Signale vernachlässigt werden [44]. Das liegt daran, daß in diesem Fall die Detektion der tonalen Komponenten mit seinem Modell gut funktioniert, d.h. die tonalen Komponenten auch als solche erkannt werden [50]. Die Zuordnung der Energie im Spektralbereich zu rauschhaften, bzw. tonalen Komponenten erfolgt eindeutig.

Durch die hier verwendete Vorverarbeitung, FTT und TTLZM, wird ebenfalls eine eindeutige Zuordnung rauschaft - tonal erreicht. Dies wird durch die Berücksichtigung globaler Maxima (vgl. Kapitel 4.1.5) bei der Bestimmung des TTZM erreicht. Es wird daher für dieses Modell angenommen, daß die Vernachlässigung der Rauschintensität keine gravierenden Fehler bei der Berechnung der virtuellen und spektralen Tonhöhen bewirkt.

Für eine möglichst exakte Analyse der Tonhöhengewichte, vgl. Kap 5.2.4- 5.2.6 muß auch in diesem Modell die Rauschintensität aufsummiert und bei der Berechnung des Überschußpegels berücksichtigt werden.

5.2.3 Tonhöhenverschiebung

Der Effekt der Tonhöhenverschiebung wurde bereits in Kapitel 2 erläutert. An dieser Stelle erfolgt nun eine Quantifizierung der Tonhöhenverschiebung v .

Um in dem vorliegenden Modell die spektrale Tonhöhe H_μ der μ ten Komponente (f_μ, L_μ) durch ihre Frequenz f_μ und ihre Tonhöhenverschiebung v_μ zu beschreiben, wird folgender Ausdruck verwendet, vgl Gl. (2.2):

$$(5.6) \quad H_\mu = \left(\frac{f_\mu}{[Hz]} \right) \cdot (1 + v_\mu) [pu]$$

Die Tonhöhenverschiebung wird in der vorliegenden Arbeit durch folgende Formel berücksichtigt, die von Terhardt stammt [50]:

$$\begin{aligned}
(5.7) \quad v_\mu &= 0,0002 \cdot \left(\frac{L_\mu}{[dB]} - 60 \right) \left(\frac{f_\mu}{[kHz]} - 2 \right) \\
&+ 0,015 \cdot \exp\left(\frac{-LX'_\mu}{20[dB]} \right) \cdot \left(3 - \ln\left(\frac{f_\mu}{[kHz]} \right) \right) \\
&+ 0,03 \cdot \exp\left(\frac{-LX''_\mu}{20[dB]} \right) \cdot \left(0,36 + \ln\left(\frac{f_\mu}{[kHz]} \right) \right)
\end{aligned}$$

Die Größen LX'_μ , LX''_μ werden in Gl. (5.8) bzw. (5.9) definiert. Der erste Summand in Gl. (5.7) berücksichtigt die Pegelabhängigkeit der Tonhöhe, der zweite den Einfluß der Maskierung durch Komponenten, die eine kleinere Frequenz f_v besitzen als die untersuchte Frequenz f_μ , und der dritte den Einfluß der Maskierung durch Komponenten die größere Frequenzen f_v besitzen als f_μ .

Die so berechnete Tonhöhe H_μ wird als *wahre spektrale Tonhöhe* bezeichnet [44]. Sie ist identisch zu der Frequenz (in Hz) eines reinen Tones mit einem Schalldruckpegel von 60 dB, die die gleiche Tonhöhe hervorruft wie die untersuchte Komponente (f_μ , L_μ), maskiert durch alle anderen Komponenten (f_v , L_v).

Zur Unterscheidung dazu wird die *nominelle spektrale Tonhöhe* definiert als Tonhöhe die man erhält, wenn Tonhöhenverschiebungen unberücksichtigt bleiben [44]. In diesem Fall ist die Tonhöhe in pu identisch der Frequenz der betrachteten Komponente in Hz.

Die Variable LX'_μ in Gl. (5.7) steht für den Überschußpegel der Frequenz f_μ , der ausschließlich von der Maskierung durch Frequenzen f_v abhängt, die kleiner als die untersuchte Frequenz f_μ sind: $f_v < f_\mu$. LX''_μ berücksichtigt den Überschußpegel für größere Frequenzen f_v als die untersuchte Frequenz f_μ : $f_v > f_\mu$:

$$(5.8) \quad LX'_\mu = L_\mu - 20 \cdot \log\left(\sum_{v=1}^{\mu-1} 10^{L_{E_v}(f_\mu)/(20[dB])} \right) [dB]$$

$$(5.9) \quad LX''_\mu = L_\mu - 20 \cdot \log\left(\sum_{v=\mu+1}^N 10^{L_{E_v}(f_\mu)/(20[dB])} \right) [dB]$$

5.2.4 Gewichtung der spektralen Komponenten

Der Beitrag der einzelnen Komponenten $\mu = 1 \dots R$ zur Tonhöhenwahrnehmung hängt von ihrer Frequenz f_μ sowie von ihrem Überschußpegel LX_μ ab. In der hier vorliegenden Arbeit wird die Frequenzabhängigkeit an das Prinzip der spektralen Dominanz von Plomp bzw. Ritsma [nach Terhardt, 50] angelehnt. Der Überschußpegel einfluß wird in dem vorliegenden Tonhöhenberechnungsalgo-

rithmus so konstruiert, daß mit wachsendem Überschußpegel LX_μ der Einfluß der μ ten Komponente steigt und etwa für 20 dB Überschußpegel in eine Sättigung übergeht [50].

Das *Gewicht der spektralen Tonhöhe* WS_μ nach [50], welches ein Maß für den Beitrag zur Tonhöhenwahrnehmung ist, wird in dem hier implementierten Tonhöhenberechnungsalgorithmus wie folgt numerisch beschrieben:

$$(5.10) \quad WS_\mu = \frac{1 - \exp\left(\frac{-LX_\mu}{15[\text{dB}]}\right)}{\sqrt{1 + 0.07 \cdot \left(\frac{f_\mu}{0.7[\text{kHz}]} - \frac{0.7[\text{kHz}]}{f_\mu}\right)^2}}$$

Dieser Ansatz der Gewichtsfunktion, Gl (5.10), approximiert laut Terhardt [50] die experimentellen Daten gut und kann somit als geeignete Grundlage der Komponentengewichtung dienen.

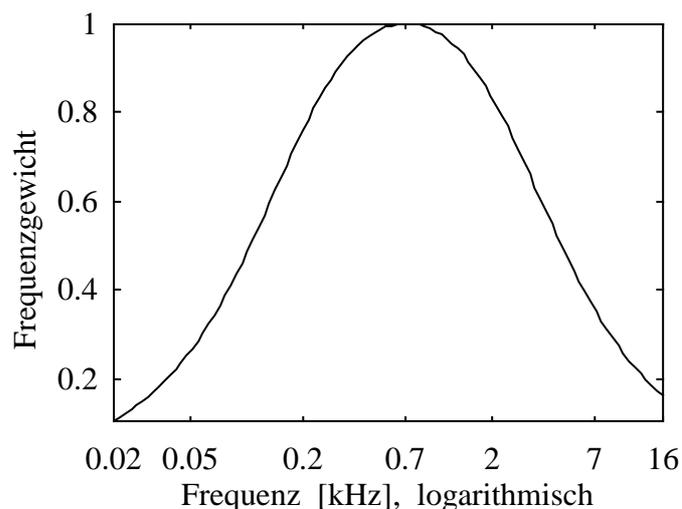


Abb. 5.4: *Frequenzabhängigkeit der Gewichtsfunktion. Dargestellt ist das Frequenzgewicht, d.h. der frequenzabhängige Anteil von Gl. (5.10), über der logarithmierten Frequenz. Zu erkennen ist das Maximum bei 700 Hz und der Abfall zu größeren bzw. kleineren Frequenzen hin. Einen sehr ähnlichen Verlauf besitzt auch die von Fletcher und Gault gefundene Funktion, die den Einfluß der Frequenz auf die Sprachverständlichkeit beschreibt [nach Terhardt, 50].*

Abb. 5.4 stellt graphisch die Frequenzabhängigkeit von Gl. (5.10) dar, Abb. 5.5 verdeutlicht die Überschußpegelabhängigkeit, d.h. den Zähler von Gl. (5.10).

Das Ergebnis dieser Wichtung wird mittels des *Spektraltonhöhenmusters* nach [50] dargestellt. In ihm werden die WS_μ über der Tonhöhe in äquivalenter Frequenz aufgetragen. Zusätzlich enthält es die Tonhöhenverschiebungen v_μ , die jedoch i.a. im Spektraltonhöhenmuster nicht dargestellt werden. Abb. 5.9c zeigt ein Beispiel für ein Spektraltonhöhenmuster, welches aus dem Überschußpegelmuster, Abb. 5.9b, des Beispiels aus Kapitel 5.3.2 berechnet wird.

Nach Gl. (5.10) liegen die WS_μ für physikalisch sinnvolle Frequenzen f_μ und Pegelüberschüsse LX_μ im Intervall $[0;1]$, vgl. auch Abb. 5.9c.

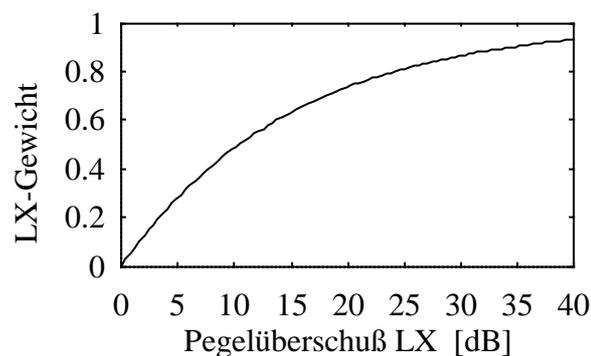


Abb. 5.5: Überschußpegelabhängigkeit der Gewichtsfunktion. Aufgetragen ist das Überschußpegelgewicht, d.h. der Zähler von (5.10), gegen den Überschußpegel LX . Es ist abzulesen, daß das LX -Gewicht für $LX < 15$ dB stark ansteigt, dann in eine Sättigung übergeht und sich für $LX > 30$ dB kaum verändert.

5.2.5 Virtuelle Tonhöhen-Extraktor (VTE)

Ein vereinfachtes Modell des VTE wird in Abb. 5.2 erläutert. In diesem Abschnitt wird der VTE-Algorithmus nach [44], [50] hergeleitet und diskutiert. In dem vorliegenden Tonhöhenberechnungsalgorithmus ist der VTE von Terhardt implementiert, so wie im folgenden besprochen.

Die Eingangsgrößen für den VTE bestehen aus Komponentenfrequenz f_i , Tonhöhenverschiebung v_i und dem Gewicht der spektralen Tonhöhe WS_i . Diese Größen werden aus dem Spektraltonhöhenmuster entnommen. Die Anzahl der Komponenten im Spektraltonhöhenmuster beträgt R (vgl. Abschnitt 5.2.2).

Aus Kapitel 2 folgt, daß jede mögliche virtuelle Tonhöhe als Subharmonische einer der Frequenzen des Spektraltonhöhenmusters dargestellt werden kann.

Bleiben die Tonhöhenverschiebungen v_i in erster Näherung unberücksichtigt, so erhält man die *nominelle virtuelle Tonhöhe* \underline{H}_{im}^{10} [44]:

$$(5.11) \quad \underline{H}_{im} = \frac{f_i}{[\text{Hz}]} \frac{1}{m} [\text{pu}]$$

\underline{H}_{im} repräsentiert die m te Subharmonische der i ten Komponente aus dem Spektraltonhöhenmuster. Die Genauigkeit von \underline{H}_{im} kann durch Berücksichtigung der Tonhöhenverschiebungen v_i und Oktavspreizungen (vgl. z.B [41], [42]) verbessert werden. In diesem Fall ergibt sich die *wahre virtuelle Tonhöhe* zu:

$$(5.12) \quad \underline{H}_{im} = \frac{f_i}{[\text{Hz}]} \frac{1}{m} \cdot \left\{ 1 + v_i - \text{sgn}(m-1) \cdot 0,001 \cdot \left(18 + 2,5m - \frac{(50-7m)}{m} \frac{f_i}{[\text{kHz}]} + 0,1 \cdot \left(m \frac{[\text{kHz}]}{f_i} \right)^2 \right) \right\} [\text{pu}]^{(11)}$$

Dieser Zusammenhang wird von Terhardt [44] anhand experimenteller Daten abgeleitet. Gl. (5.12) nähert die wahre virtuelle Tonhöhe für $\underline{H}_{im} \leq 500$ pu. Für größere Frequenzen wird die Näherung zunehmend schlechter. Dies jedoch keine große Einschränkung, da virtuelle Tonhöhen oberhalb von 500 pu nur selten vorkommen [50].

Mit (5.12) bzw. (5.11) erhält man auf diese Weise $N \cdot R$ virtuelle Tonhöhen \underline{H}_{im} . Ziel des Algorithmus ist es, die ausgeprägtesten dieser virtuellen Tonhöhen herauszufinden. Dafür muß ein Maß der Ausprägtheit, hier als Gewicht W_{im} der virtuellen Tonhöhe \underline{H}_{im} bezeichnet, definiert werden:

Das Gewicht W_{im} der virtuellen Tonhöhe \underline{H}_{im} soll nach [50]

1. mit der Anzahl nahezu übereinstimmender virtueller Tonhöhen \underline{H}_{im} steigen,
2. mit der Genauigkeit der Übereinstimmung nach 1. steigen,
3. mit dem Gewicht WS_i der spektralen Tonhöhen f_i steigen und
4. mit der Nummer m der Subharmonischen abfallen.

Bedingung drei legt nahe, die Eingangsgrößen (f_i , v_i , WS_i) für den VTE nach ihrem Gewicht WS_i zu ordnen, um die virtuellen Tonhöhen \underline{H}_{im} mit dem größten Gewicht W_{im} möglichst schnell aufzufinden. Dieser Sortierprozeß findet in der vorliegenden Implementation des VTE als erstes statt.

¹⁰ Virtuelle Tonhöhen werden durch Unterstreichung des Formelzeichens von spektralen Tonhöhen unterschieden.

¹¹ Die sgn- Funktion ist definiert als: $\text{sgn}(x) = -1$ für $x < 0$; 0 für $x = 0$; $+1$ für $x > 0$.

Abb. 5.6 zeigt ein Flußdiagramm des VTE-Algorithmus. In ihm ist als erster Schritt die Sortierung der Eingangsgrößen für den VTE- Algorithmus (f_i , v_i , WS_i) nach ihrem Gewicht WS_i dargestellt.

Der weitere Ablauf des VTE- Algorithmus wird im folgenden ausführlich beschrieben.

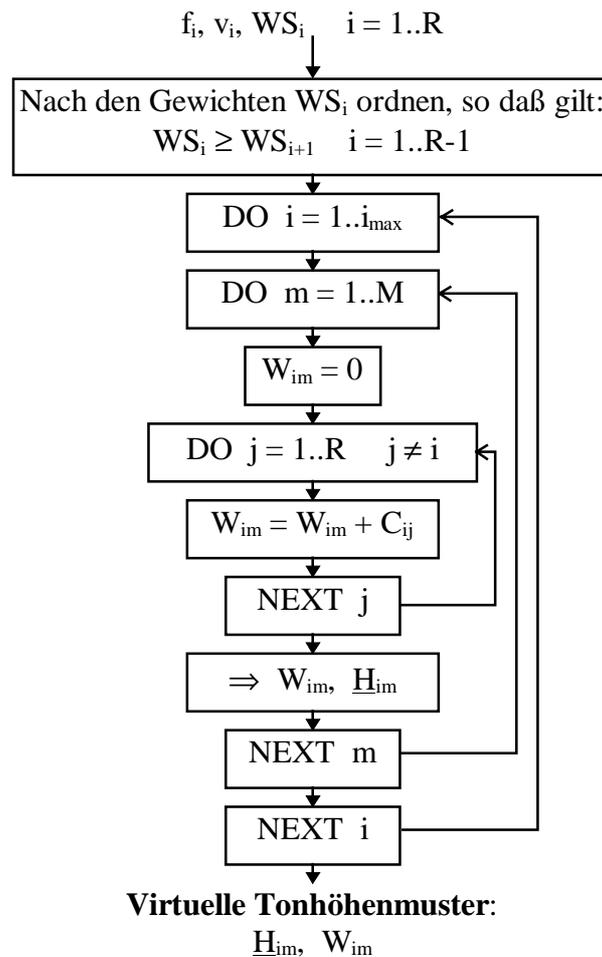


Abb. 5.6: Flußdiagramm des in dieser Arbeit implementierten VTE. Input des VTE ist das Spektraltonhöhenmuster (f_i ; v_i ; WS_i), Output ist das virtuelle Tonhöhenmuster (H_{im} ; W_{im}). Darstellung ähnlich [50], weitere Erläuterung siehe Text.

In den nächsten Schritten wird für die erste spektrale Tonhöhe (f_i , v_i , WS_i) mit $i = 1$ die Übereinstimmung ihrer ersten Subharmonischen ($m = 1$) mit allen Subharmonischen der zweiten Komponente ($j = 2$) getestet. Der Übereinstimmungskoeffizient C_{ij} nach [50] wird dafür als Maß der Übereinstimmung der

beiden momentan untersuchten Subharmonischen m und n der spektralen Tonhöhen i und j definiert:

$$(5.13) \quad C_{ij} = \begin{cases} \sqrt{\frac{WS_i \cdot WS_j}{m \cdot n}} \cdot \left(1 - \frac{\gamma}{\delta}\right) & \text{für } \gamma \leq \delta \\ 0 & \text{für } \gamma > \delta \text{ und / oder } n > 20 \end{cases}$$

Hierbei muß die Nummer der Subharmonischen n jeweils neu berechnet werden:

$$(5.14) \quad n = NINT\left(\frac{m \cdot f_j}{f_i}\right) \quad (12)$$

Der Nenner unter dem Wurzelzeichen aus Gl. (5.13) repräsentiert die vierte Forderung der Definition des Gewichts der virtuellen Tonhöhe. Der Zähler unter dem Wurzelzeichen beinhaltet die dritte und der Faktor $(1 - \gamma/\delta)$ steht für die zweite Forderung.

γ in Gl. (5.13) gibt die relative Abweichung der Frequenzen i und j an:

$$(5.15) \quad \gamma = \left| \frac{n \cdot f_i}{m \cdot f_j} - 1 \right|$$

γ muß kleiner sein als ein vorgegebenes *Übereinstimmungsintervall* der relativen Breite δ , damit die Subharmonischen m und n der Frequenzen j und i als übereinstimmend detektiert werden, vgl. Gl. (5.13). Bei $\gamma=0$ sind die untersuchten Subharmonischen identisch.

Die erste Forderung der Definition des Gewichts der virtuellen Tonhöhe W_{im} wird durch Summation der Übereinstimmungskoeffizienten C_{ij} für $j = 1 \dots R$ berücksichtigt. Für die virtuelle Tonhöhe \underline{H}_{im} ergibt sich das Gewicht W_{im} zu [50]:

$$(5.16) \quad W_{im} = \beta(\underline{H}_{im}) \sum_{j=1, j \neq i}^R C_{ij} \cdot$$

In Gl. (5.16) wird das *Grundfrequenzgewicht* $\beta(\underline{H}_{im})$ eingeführt, um den Existenzbereich der virtuellen Tonhöhe, Abb. 2.7, nicht zu verlassen [50]:

$$(5.17) \quad \beta(\underline{H}_{im}) = \frac{1}{1 + (\underline{H}_{im}/800[pu])^4} \cdot$$

Den grundsätzlichen Verlauf des Grundfrequenzgewichts zeigt Abb. 5.7:

¹² 'NINT' in Gl. (5.14) bedeutet, daß die nächstbenachbarte ganze Zahl genommen wird, analog dem FORTRAN- Befehl NINT.

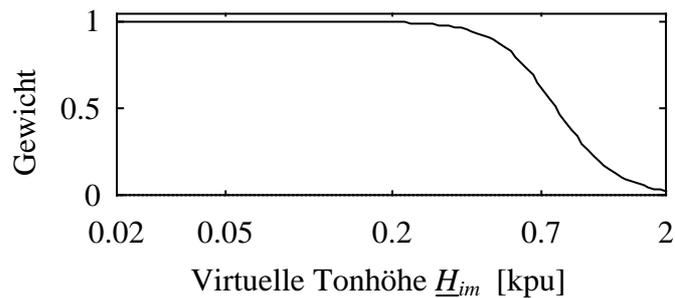


Abb. 5.7: Grundfrequenzgewicht $\beta(\underline{H}_{im})$. Aufgetragen ist das Grundfrequenzgewicht $\beta(\underline{H}_{im})$ über der virtuellen Tonhöhe \underline{H}_{im} , (logarithmisch, in Einheiten von 1000 pu). Zu erkennen ist das Tiefpaßverhalten des Grundfrequenzgewichtes $\beta(\underline{H}_{im})$.

Nachdem mit diesem Verfahren, Gl. (5.13) bis (5.17), die erste Subharmonische ($m = 1$) der ersten Spektraltonhöhe f_i ($i = 1$) mit allen Subharmonischen der zweiten Spektraltonhöhe f_j ($j = 2$) verglichen wurde, wird der Vorgang für die zweite Subharmonische ($m = 2$) der ersten Spektraltonhöhe f_i ($i = 1$) wiederholt. Anschließend wird die Ordnung der Subharmonischen m sukzessive erhöht, bis der Maximalwert M erreicht ist. Der Maximalwert M kann vom Benutzer festgelegt werden. Der größte Wert, den M annehmen kann ist 20, wie in Kapitel 2.4 abgeleitet wird. Nach Terhardt et al. [50] ist 12 ein typischer, d.h. sinnvoller, Wert für die maximale Anzahl der zu berücksichtigenden Subharmonischen M .

Wenn $m = M$ erreicht ist, wurden M Wertepaare $(\underline{H}_{im}; W_{im})$ mit $i = 1$ für die Spektraltonhöhe mit dem größten Frequenzgewicht WS_i berechnet. Die gleiche Rechnung wird nun für die Spektraltonhöhe f_i ($i = 2$) mit dem zweitgrößten Frequenzgewicht WS_i durchgeführt. i wird daraufhin solange sukzessive erhöht, bis alle Spektraltonhöhen abgearbeitet sind, d.h. bis $i = R$ ist.

Es zeigt sich, daß für Komponenten, die ein kleineres spektrales Gewicht WS_i besitzen als ein bestimmter Schwellenwert WS_{imax} , keine weiteren virtuellen Tonhöhen \underline{H}_{im} mehr auftreten, die zu den größeren der vorkommenden virtuellen Tonhöhen gehören [50]. Der Algorithmus kann also abgebrochen werden, wenn $WS_i \leq WS_{imax}$ gilt, d.h. i muß nur von 1 bis i_{max} erhöht werden.

Terhardt et al. [50] haben empirisch ermittelt, daß $WS_i \geq 0.7 \cdot WS_1$ ein Wert ist, der einerseits hohe Wahrscheinlichkeit dafür birgt die größten virtuellen Tonhöhen \underline{H}_{im} zu detektieren, andererseits jedoch den Rechenaufwand beträchtlich

senkt. Es werden daher nur diejenigen Komponenten berücksichtigt, deren spektrales Gewicht WS_i mindestens 70% des größten Gewichts WS_l beträgt. Nachdem alle relevanten Spektraltonhöhen wie im Flußdiagramm Abb. 5.6 berücksichtigt wurden, erhält man das *virtuelle Tonhöhenmuster* [50]. In ihm sind die Gewichte der virtuellen Tonhöhen W_{im} über den virtuellen Tonhöhen \underline{H}_{im} aufgetragen. Im virtuelle Tonhöhenmuster können identische virtuelle Tonhöhen \underline{H}_{im} mit verschiedenen Gewichten W_{im} auftreten. Ist dies der Fall, so werden nur die virtuellen Tonhöhen weiter betrachtet, die das größte Gewicht besitzen [50].

Ein Beispiel für ein virtuelles Tonhöhenmuster zeigt Abb. 5.9d für das Sprachsignal des schon aus Abschnitt 5.3.2 und 5.3.4 bekannten Vokals /a/. Das virtuelle Tonhöhenmuster ist eine Art Histogramm, in dem die Wahrscheinlichkeit (d.h. das Gewicht) über den möglichen virtuellen Tonhöhen aufgetragen wird. Die Darstellung kann daher nicht eindeutig sein. Diese Tatsache ist jedoch keine Unzulänglichkeit, sondern spiegelt den Umstand wieder, daß Versuchspersonen unterschiedliche (virtuelle) Tonhöhen wahrnehmen können, je nach dem auf welche Tonhöhen sie sich konzentrieren. Ein Beispiel hierfür ist die Oktavunsicherheit oder Mehrdeutigkeit, vgl. z.B. [39], [40], [41], [42] oder [50]. Auf die gleiche Weise muß auch das Spektraltonhöhenmuster aufgefaßt werden, denn in der Regel können Versuchspersonen durch 'analytisches Hören' ihre Aufmerksamkeit auf bestimmte spektrale Tonhöhen richten und diese von den anderen, gleichzeitig wahrgenommenen spektralen Tonhöhen trennen. Der Wertebereich für das Gewicht der virtuellen Tonhöhe W_{im} ist prinzipiell unendlich, wird jedoch durch Maskierungseffekte (Abschnitt 5.3.2) und damit indirekt durch die Anzahl der relevanten Komponenten begrenzt. In der Praxis liegen die W_{im} im Intervall $[0 ; \text{ca. } 3]$, vgl. [50].

5.2.6 Vergleich von Spektral- und virtuelle Tonhöhenmuster

Das Ziel aller oben ausgeführten Rechnungen ist die ausgeprägteste Tonhöhe eines Schalles zu ermitteln, d.h. den Tonhöheindruck des Schalles vorherzusagen, den eine Versuchsperson beim Anhören desgleichen wahrnimmt. Dabei ist es unerheblich, ob es sich hierbei um eine spektrale oder virtuelle Tonhöhe handelt.

Das Problem dabei liegt darin, die Ausgeprägtheiten, bzw. die Gewichte der spektralen Tonhöhen WS_i und der virtuellen Tonhöhen W_{im} , miteinander zu vergleichen. Dies ist nicht unmittelbar möglich, da jedes der Gewichte nur dafür konzipiert wurde, die Ausgeprägtheiten der spektralen, bzw. der virtuellen, Tonhöhen untereinander zu vergleichen.

Aus diesem Grund müssen die Skalen der Gewichte der spektralen WS_i und der virtuellen Tonhöhe W_{im} so abgeändert werden, daß gleiche Gewichte WS_i bzw. W_{im} auch die gleiche Ausprägtheit der Tonhöhen hervorrufen. In empirischen Untersuchungen konnten Terhardt et al. zeigen [50], daß bei einem Verhältnis von 1:2 des Spektraltonhöhengewichts zum Gewicht der virtuellen Tonhöhen beide Phänomene den gleichen Tonhöheneindruck bei Versuchspersonen hervorrufen. D.h. wenn die spektralen Tonhöhengewichte WS_i mit dem Faktor 0,5 multipliziert werden, so sind sie direkt mit den Spektraltonhöhengewichten WS_i vergleichbar. In dieser Arbeit werden, wenn nicht ausdrücklich anders betont, die Spektraltonhöhengewichte stets mit 0,5 multipliziert, um eine leichte Vergleichbarkeit zwischen spektralen und virtuellen Tonhöhenmustern zu gewährleisten.

Dabei bleibt allerdings der situative Kontext unberücksichtigt, d.h. die persönlichen Erfahrungen, die Erwartungshaltung, die akustische Vorbildung, das Befinden und die äußeren Umstände der Versuchspersonen werden aufgrund ihrer Komplexität vernachlässigt.

Auf dieser Basis kann eine Beurteilung der Tonhöhen stattfinden: Die mit 0,5 multiplizierten spektralen und die virtuellen Tonhöhen werden nun direkt miteinander verglichen, d.h. nach abfallender Größe ihres Gewichtes sortiert. Die Tonhöhe mit dem größten Gewicht ist die Tonhöhe (egal ob virtuell oder spektral), die eine Versuchsperson wahrscheinlich wahrnimmt (vgl. die Bemerkungen zum situativen Kontext).

5.2.7 Implementierter Tonhöhenberechnungsalgorithmus

Das hier implementierte Modell, so wie in den Abschnitten 5.3.1 bis 5.3.6 vorgestellt, wird durch Abb. 5.8 zusammengefaßt:

Im ersten Schritt wird der TTLZM- Algorithmus, wie in Kapitel 4 beschrieben, angewendet. Für die tonalen Komponenten des TTLZM werden anschließend die Maskierungseinflüsse (Kap 5.2.2) und die Tonhöhenverschiebungen (Kap. 5.2.3) berücksichtigt. Im dritten Schritt werden die Gewichte der Spektraltonhöhen berechnet. Auf diese Weise erhält man das Spektraltonhöhenmuster. Aus diesem berechnet der VTE- Algorithmus schließlich die virtuellen Tonhöhen und deren Gewichte.

Abb. 5.9 zeigt zusammenfassend die Zwischen- sowie die Endergebnisse des hier implementierten Modells für das exemplarisch ausgewählte Sprachsignal des Abschnitts 5.3.2, in dem eine männliche Person den Vokal /a/ spricht. Diese

Abbildung erfüllt lediglich die Aufgabe, die einzelnen Stufen des Tonhöhenberechnungsalgorithmus und deren Zwischenergebnisse zu verdeutlichen.

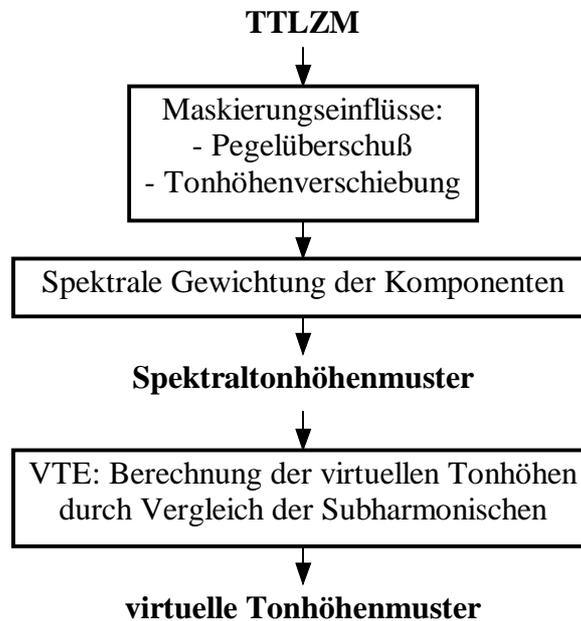


Abb. 5.8: *In dieser Arbeit implementiertes Modell zur Tonhöhenbestimmung. Zusammenfassung der Vorgehensweise, die in den Abschnitten 5.3.1 bis 5.3.6 näher erläutert wird.*

5.3 Zeitkontinuierliche Verarbeitung

Im Abschnitt 5.3 wird das Berechnungsmodell der virtuellen Tonhöhe für einen beliebigen aber festen Zeitpunkt t_k dargelegt. An dieser Stelle soll eine Erweiterung auf eine quasikontinuierliche, d.h. eine in kleinen, äquidistanten Zeitschritten erfolgende Berechnung mit dem in Abschnitt 5.3 entwickelten Algorithmus erfolgen. Dazu wird der Zeitpunkt t_k sukzessive verschoben und so die Zeitabhängigkeit berücksichtigt.

Die zweidimensionalen Auftragungen, Abb. 5.9a bis Abb. 5.9d bekommen somit eine zusätzliche Dimension: die Zeit. Daher muß im weiteren eine dreidimensionale Auftragung erfolgen. Diese kann z.B. durch eine Wasserfall- oder eine Oberflächendarstellung erfolgen. Im Rahmen dieser Arbeit werden ausschließlich Oberflächendarstellungen folgender Art verwendet:

Die erste unabhängige Größe Zeit wird auf der Abszisse abgetragen, die zweite unabhängige Größe Frequenz (wahlweise in Hz oder Bark), bzw. äquivalente Tonhöhe in pu, wird auf der Ordinate dargestellt. Die abhängigen Größen wie die Ausprägtheit der spektralen und virtuelle Tonhöhen oder der Überschuß-

pegel werden farbig skaliert über der durch Abszisse und Ordinate aufgepannten Ebene dargestellt, analog dem TTLZM und den FTT-Leistungsspektren, vgl. Kap. 3 und 4.

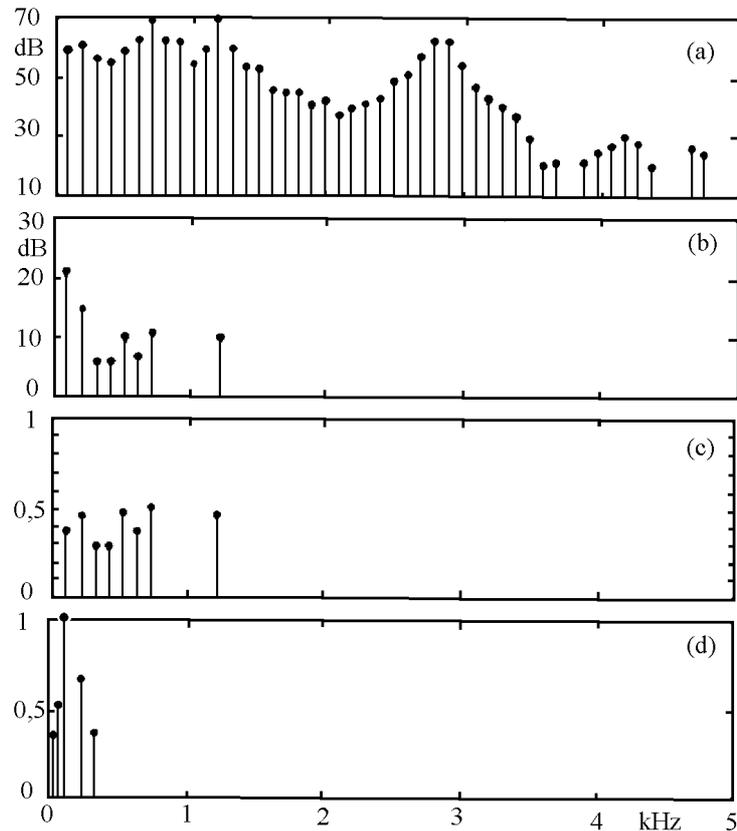


Abb. 5.9: Tonhöhenanalyse eines Sprachschalles. Dargestellt ist das Sprachsignal des Vokales /a/, gesprochen von einer männlichen Person zu einem beliebigen, aber festen Zeitpunkt t_k . Die Grundfrequenz beträgt ca 99 Hz. Die Darstellung ist ähnlich zu [50]

a.) Teiltonlinienzeitmuster TTLZM: Aufgetragen sind die Teiltöne über der Frequenz.

b.) Überschußpegelmuster: Dargestellt sind die Überschußpegel LX ($LX > 0$) gegen die Frequenz.

c.) Spektraltonhöhenmuster: Gewicht WS der spektralen Komponenten über der Tonhöhe in äquivalenter Frequenz (Einheit pu). Nicht dargestellt sind die den spektralen Komponenten zugeordneten Tonhöhenverschiebungen v .

d.) Virtuelle Tonhöhenmuster: virtuelle Tonhöhen über äquivalenter Frequenz. Abgebildet sind die 5 Komponenten mit der größten virtuellen Tonhöhe. Bemerkenswert ist die Tatsache, daß die Tonhöhe, die der Grundfrequenz entspricht, durch eine virtuelle Tonhöhe hervorgerufen wird.

Um die zeitkontinuierlichen Ergebnisse von denen zu einem festen Zeitpunkt t_k zu unterscheiden, werden die entsprechenden Abbildungen nicht mehr mit ‘Muster’, sondern mit ‘Diagramm’ bezeichnet.

Abb. 5.10 zeigt das Spektraltonhöhendigramm für das Beispielsignal aus Kapitel 2.9. Es wird aus dem TTLZM (Abb. 4.9) berechnet. Die spektrale Tonhöhe ist dabei nur im Bereich unterhalb von 800 pu dargestellt, um den Vergleich mit dem virtuellen Tonhöhendigramm, Abb. 5.11, zu erleichtern. Die hier gegen die Zeit und die äquivalente Frequenz der spektralen Tonhöhe dargestellte Größe ‘Ausgeprägtheit der spektralen Tonhöhe’ ist dem in Abschnitt 5.3.4 eingeführten ‘Gewicht der spektralen Tonhöhe’ äquivalent. Dieses Spektraltonhöhendigramm ist das einzige in der vorliegenden Arbeit, welches nicht mit dem Faktor 0,5 multipliziert wurde, um die einzelnen Stufen des Tonhöhenberechnungsalgorithmus besser zu verdeutlichen.

Aus Abb. 5.10 können die Grundfrequenz und die Oberschwingungen der männlichen Stimme abgelesen werden. Es zeigt sich, daß nicht die Grundfrequenz, sondern die erste Harmonische der Grundfrequenz die ausgeprägteste spektrale Tonhöhe hervorruft. Auch die zweite und dritte Harmonische sind, soweit vorhanden, stärker ausgeprägt als die Grundfrequenz. Prinzipiell weist das spektrale Tonhöhendigramm die gleichen Formantstrukturen auf wie das FFT-Spektrum (Abb. 3.6) oder das TTLZM (Abb. 4.9). Lediglich kleine Tonhöhenverschiebungen (siehe auch Abschnitt 5.3.3) sind ablesbar. Beim Vergleich des TTLZM und des spektralen Tonhöhendigrammes fällt auf, daß einige Teiltöne des TTLZM keine spektrale Tonhöhe hervorrufen. Das liegt an der gegenseitigen Maskierung der Teiltöne, vgl. Abschnitt 5.3.2.

In Abb. 5.11 ist das virtuelle Tonhöhendigramm für die spektralen Tonhöhen nach Abb. 5.10 dargestellt. Die äquivalente Frequenz der virtuellen Tonhöhe ist auf der linken Seite abgetragen. Ihre Ausgeprägtheit läßt sich anhand der farblichen Skalierung ermitteln. Gut zu erkennen ist die Oktavunsicherheit (vgl. Abschnitt 5.3.6) bei der Berechnung von virtuellen Tonhöhen: neben der ausgeprägtesten virtuellen Tonhöhe von etwa 100 bis 110 pu sind weitere virtuelle Tonhöhen bei der halben (etwa 50 pu) und doppelten (ca. 220 pu) äquivalenten Frequenz zu erkennen. An den Stellen der Abbildung, an denen keine stimmhaften Laute analysiert werden, tritt auch keine virtuelle Tonhöhe großer Ausprägung hervor. Das gilt zum Beispiel für den Frikativ ‘ch’ im Zeitbereich um ca. 1,0 sec so wie für die Plosive ‘k’ bei etwa 0,1 sec und ‘t’ bei ca. 0,8 sec.

Folgende Parameterwerte werden benutzt, um dieses Diagramm zu berechnen:

Übereinstimmungsintervall δ nach Gleichung (5.15): 0,08

Max. Anzahl der Subharmonischen M nach Abschnitt 5.3.5: 12

Anzahl der größten zu berücksichtigten virtuellen Tonhöhen: 5

Für einen Vergleich des Tonhöhereindrucks der virtuellen und spektralen Tonhöhe sind in Abb. 5.12 beide Größen in ein Diagramm eingezeichnet. Die virtuelle Tonhöhe ist durch Kreuze, die spektrale durch waagerechte Striche gekennzeichnet. Die Ausprägtheit von virtueller und spektraler Tonhöhe wird nach Abschnitt 5.3.6 wie 1:2 ins Verhältnis gesetzt. Dazu wird das Gewicht der spektralen Tonhöhe nach Abb. 5.11 durch den Faktor 2 geteilt. Die wahrgenommene Tonhöhe von Signalen kann dann einfach ermittelt werden, indem die ausgeprägteste Tonhöhe, egal ob virtuell oder spektral, aufgesucht wird.

Für das untersuchte Sprachsignal ergibt sich je nach Satzmelodie eine (virtuelle) Tonhöhe von etwa 100 bis 130 pu.

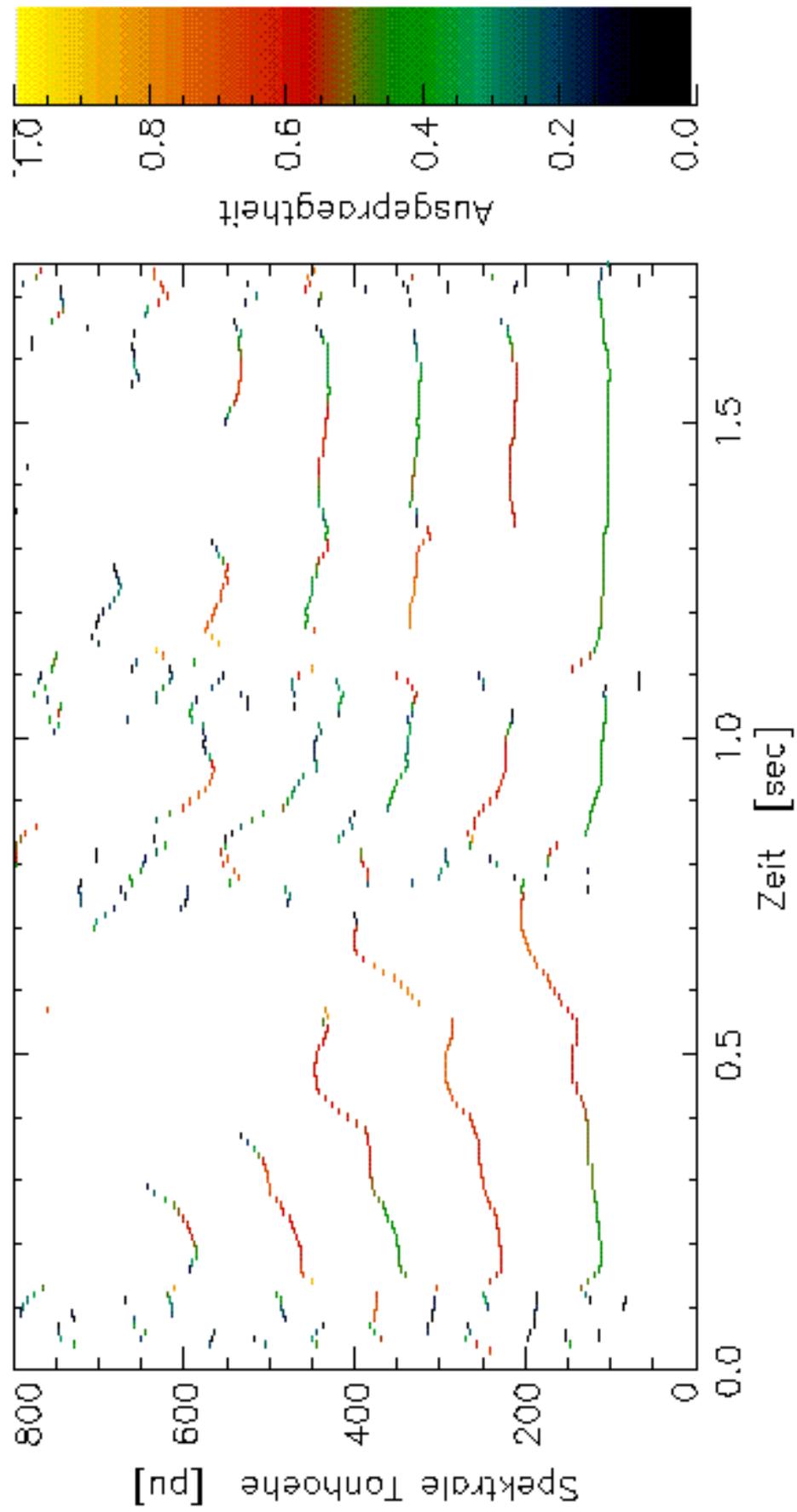


Abb 5.10: Spektraltonhöhendigramm für das Beispielsignal aus Kapitel 3.9.

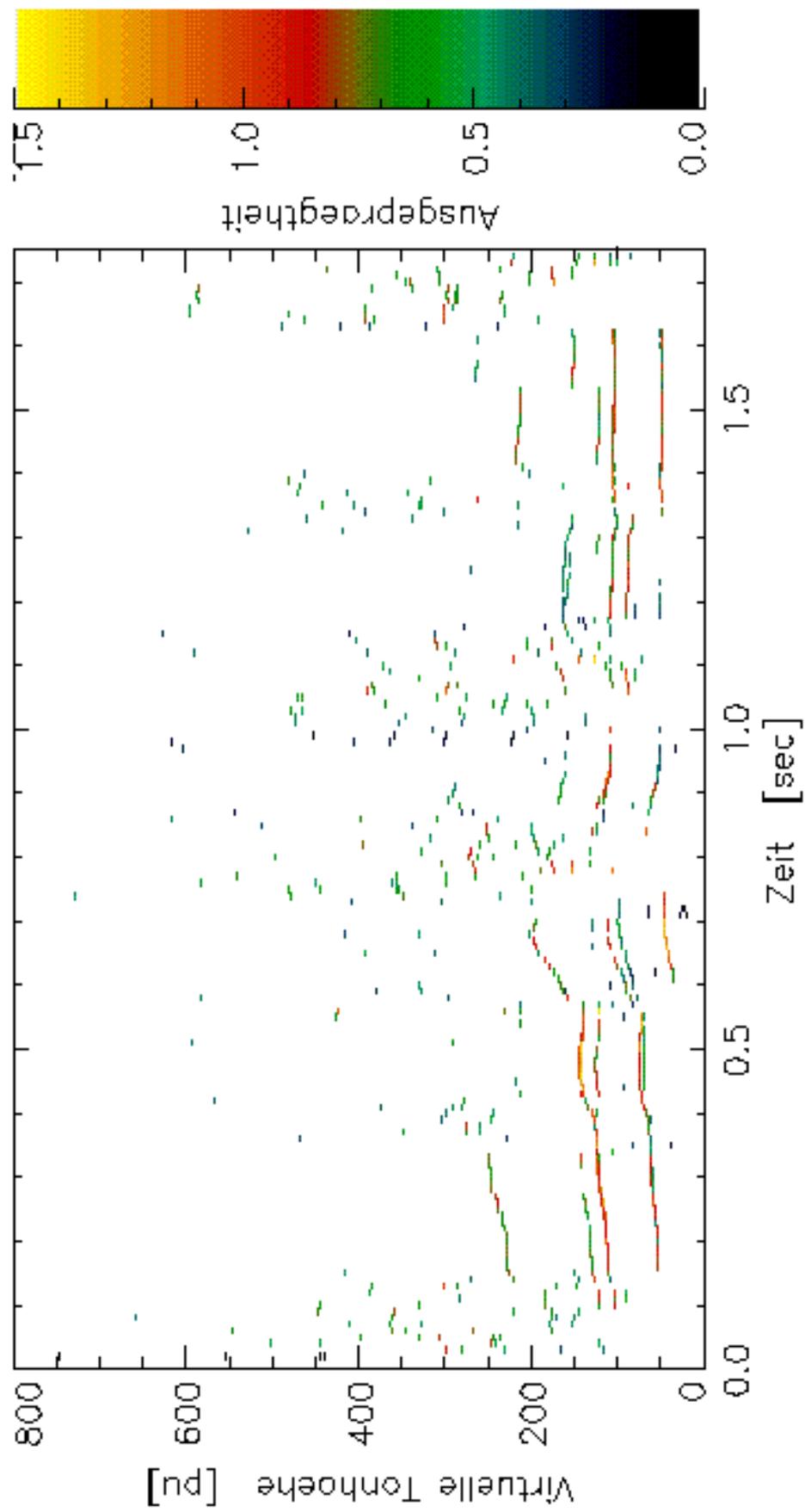


Abb 5.11: Virtuelle Tonhöhendigramm für das Beispielsignal aus Kapitel 3.9.

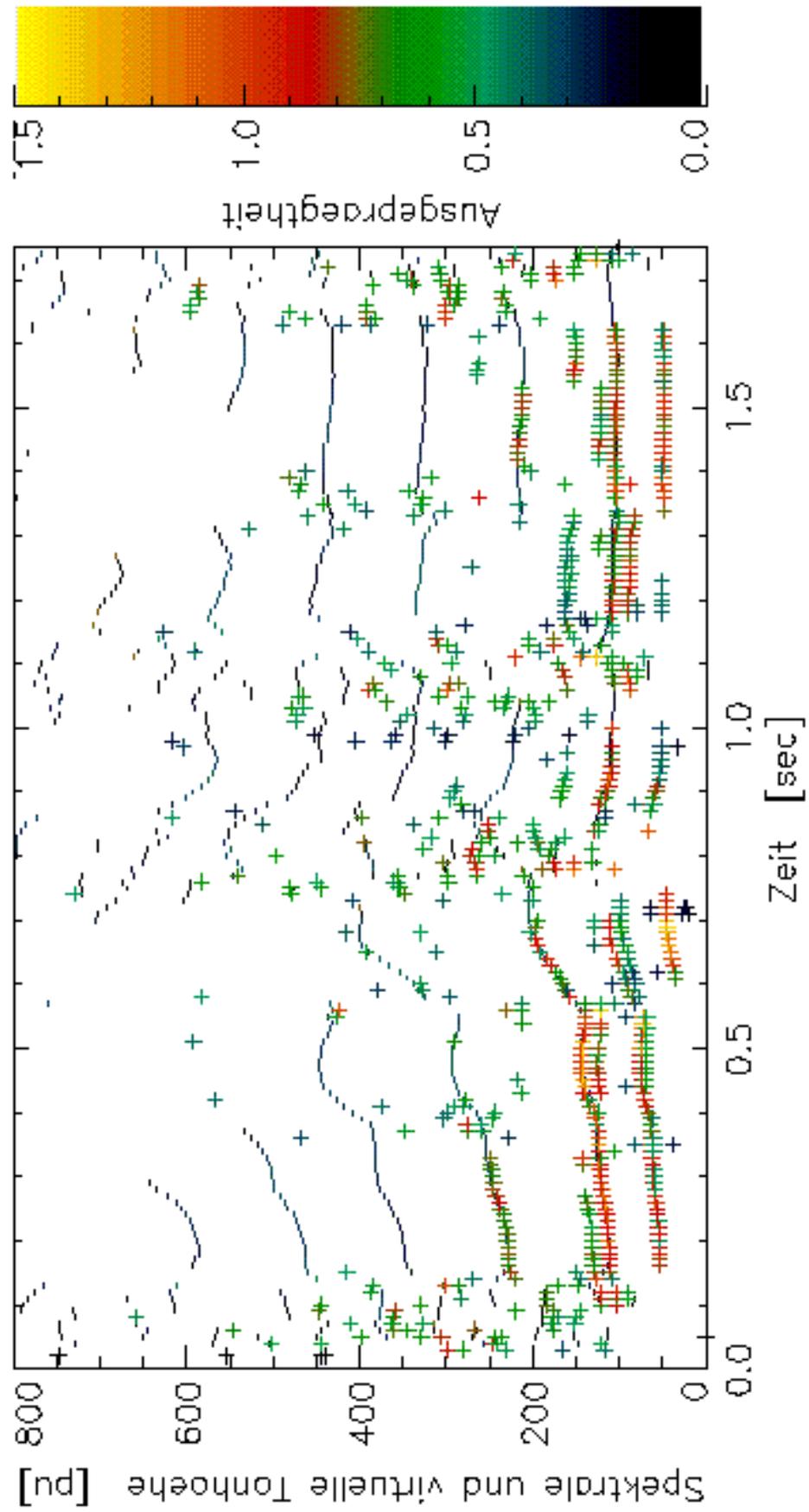


Abb 5.12: Virtuelle & spektrale Tonhöhendigramm für das Beispielsignal aus Kapitel 3.9.

6 PROGRAMMODULE, WÄHLBARE PARAMETER UND PROGRAMMTESTS

Die Modelle der Kapitel 2 bis 5 wurden auf einer Workstation implementiert, um die virtuelle und spektrale Tonhöhe von Umweltgeräuschen berechnen zu können. Alle Modelle sind in der Programmiersprache FORTRAN umgesetzt und sind mit einer Ausnahme auf einer SUN-Workstation funktionsfähig. Diese Ausnahme ist ebenfalls in FORTRAN programmiert, allerdings ist das Programm nur auf einem PC mit MICROSOFT FORTRAN 5.0-Compiler einzusetzen. Anstelle eines großen, allumfassenden Programms werden mehrere kleine Programme erstellt, die nacheinander aufgerufen werden müssen. Dies hat aus programmieretechnischer Sicht einige Vorteile.

Das Programmpaket besteht aus sieben einzelnen Programmen, die in der Regel selbsterklärend sind. Eine ausführliche Darstellung findet sich in Anhang D, dort wird auch ihre Bedienung erläutert. Eine Kurzdarstellung der Programme erfolgt im nächsten Abschnitt.

Die Programme lesen jeweils die Daten des zuvor ausgeführten Programmes von Festplatte ein und speichern die neu berechneten Daten wieder auf Festplatte ab. So ist jeder Schritt der Berechnung von Tonhöhen im nachhinein nachvollziehbar. Zusätzlich wird eine *Kontrolldatei* angelegt, in der die Eigenschaften des Signals und die verwendeten Parameter der einzelnen Programme festgehalten werden. Die Kontrolldatei ist daran zu erkennen, daß sie als **.dm* auf Festplatte gespeichert wird.

Die mittels dieses Programmpakets zu untersuchenden Signale müssen im WAVE-Format vorliegen und mit 44,1 bzw. 48 kHz gesampelt sein. In diesem Fall erkennt das Programm die Eigenschaften der Datei in denen das Signal gespeichert ist selbständig und legt die Kontrolldatei an. Es ist ebenfalls möglich reine ASCII-Dateien zu bearbeiten. Dafür muß die Kontrolldatei jedoch manuell erstellt werden.

Erweitert wird das Programmpaket durch drei Hilfsprogramme, die nicht zur Berechnung von Tonhöhen gebraucht werden, aber für die graphische Darstellung der FFT-Spektren, der TTLM und der TTLZM nützlich sind. Die Hilfsprogramme werden im Anhang E dargestellt.

6.1 Die einzelnen Programmodule

Es folgt eine Kurzbeschreibung der im Rahmen dieser Arbeit erstellten Programmodule zur Berechnung der Tonhöhe. Eine ausführliche Beschreibung enthält Anhang D.

Nummer	Name	Zweck
1	wav2msc4.0	Umwandlung von WAVE- Dateien in das hier verwendete interne Format, welches die folgenden Programmodule weiterbearbeiten können.
2a	ftt1.5d	Berechnung von FTT- Spektren 1. oder 2. Ordnung, geglättet oder ungeglättet. Das frequenzabhängige Auswertintervall $T_A(f)$ wird nicht frequenzabhängig berücksichtigt, sondern ausschließlich konstant über den ganzen Frequenzbereich verwendet !!
2b	ftt3.1c	Berechnung von geglätteten FFT- Spektren 2. Ordnung. Das frequenzabhängige Auswertintervall $T_A(f)$ wird berücksichtigt.
3	maxi3.0c	Bestimmung des TTZM.
4	netz4.0a	Vernetzung eines TTZM zum TTLZM.
5	vp2.3a	Berechnung von spektralen und virtuellen Tonhöhen sowie deren Ausgeprägtheit.
-	spc2pvwav9.0	Hilfsprogramm: wandelt interne Darstellung der FTT- Spektren so um, daß sie mit der Visualisierungssoftware PVWAVE als Oberflächendiagramm darstellbar sind
-	tzm2pvwav3.0	Hilfsprogramm: wandelt interne Darstellung eines TTZM so um, daß es mit dem Visualisierungspaket PVWAVE darstellbar ist.
-	ttz2pvwav3.0	Hilfsprogramm: wandelt interne Darstellung eines TTLZM so um, daß es mit dem Visualisierungspaket PVWAVE darstellbar ist.

6.2 Parameter der Modelle

Nachfolgend sind alle vom Benutzer zu verändernden Parameter der einzelnen Programme aufgelistet. Die nicht dargestellten oder als ‘feste Parameter’ gekennzeichneten Parameter sind nicht vom Benutzer zu beeinflussen. Sie werden so gewählt wie in den Abschnitten angegeben, in denen sie eingeführt wurden.

Samplingfrequenz f_s :

Es können nur solche WAVE-Dateien analysiert werden, die mit 44,1 oder 48 kHz gesampelt sind.

Fourier- Zeit- Transformation:

- 1. oder 2. Ordnung, geglättet oder ungeglättet, jeweils mit konstantem Auswertintervall T_A .
- ‘Audio-FTT’, d.h. FTT 2. Ordnung, geglättet, mit frequenzabhängigem Auswertintervall $T_A(f)$.

Analysebandbreite b :

Die Analysebandbreite nach Gl. (3.4) sollte zu $b = 0,1$ Bark für eine FTT erster Ordnung [12] und zu $b = 0,25$ Bark für eine FTT zweiter Ordnung [12] gewählt werden, um optimale Ergebnisse zu erzielen. Sie kann im vorliegenden Programmpaket zur Berechnung der Tonhöhen im folgenden Rahmen variiert werden: $0 < b < 1$.

Unterste Analysefrequenz f_u :

f_u wird in Gl. (3.6) definiert. Nach Heldmann [14] muß $f_u \geq 100$ Hz sein, das Programm akzeptiert Werte aus dem Bereich zwischen 20 und 15000 Hz.

Größte Frequenz f_g :

Aus der Differenz $f_g - f_u$ wird die Gesamtzahl der Analysefrequenzen N berechnet, vgl. (3.6) und Kap. 3.2. f_g kann im Rahmen des hörbaren Frequenzbereiches gewählt werden: $f_u < f_g < 16$ kHz.

Frequenzabstand der Analysefrequenzen Δz :

Der Frequenzabstand der Analysefrequenzen nach Gl. (3.6) wird laut [14] am besten durch $\Delta z = 0,03$ Bark approximiert. Er kann in diesem Programm wie folgt variiert werden: $0 < \Delta z < 1$.

Länge des Analyseintervalls T_A :

T_A ist definiert in Gl. (4.1). Für die gehörgerechte ‘Audio-FTT’ wird damit die frequenzabhängige Länge des Analyseintervalls $T_A(f)$ berechnet. Für die ‘normale’ FTT ist T_A konstant über alle Frequenzen. Der optimale Wert für T_A beträgt 1 msec [14]. T_A kann folgendermaßen gewählt werden:

$$(f_s)^{-1} \leq T_A \leq \text{Signallänge.}$$

Schwellenkriterium ΔL :

Das Schwellenkriterium für die Bestimmung der Teiltöne der Teiltonmuster wird in Abschnitt 4.1.5 definiert. Werte aus dem Bereich $0,1 \leq \Delta L \leq 10$ können eingegeben werden, üblich ist $\Delta L = 1$ oder 3 dB [13].

Mindestlänge für 100 Hz l_m :

Die frequenzabhängige Mindestlänge der Teiltonlinien im TTLZM wird anhand der Mindestlänge für 100 Hz berechnet, siehe Kap. 4.2, Gl. (4.2). Heldmann [14] gibt als Anhaltspunkt einen Wert von $l_m = 35$ msec an, der jedoch stark signalabhängig ist: für Signale mit starken zeitlichen Strukturen muß l_m kleiner ($l_m \cong 15$ msec), für Signale mit dominanten tonalen Anteilen größer ($l_m \cong 40$ msec) gewählt werden. Der erlaubte Bereich von l_m für dieses Programm liegt zwischen 10 und 500 msec.

Maximalabstand d_{max} :

Der Maximalabstand $d_{max} := \Delta i \Delta z$ wird in Kapitel 4.2.1 definiert. Ähnlich wie die Mindestlänge ist das optimale d_{max} signalabhängig. Ein typischer Wert für d_{max} liegt bei 0,27 Bark [14]. Für überwiegend tonale Signale empfiehlt Heldmann ca. 0,33 Bark und für Signale mit starken zeitlichen Strukturen ungefähr 0,24 Bark. Im Rahmen dieses Programmes kann d_{max} aus dem folgenden Intervall gewählt werden: [0,03 ; 0,4].

Übereinstimmungsintervall δ :

Die Definition Gl. (5.19) wird in Abschnitt 5.2.5 gegeben. Terhardt et al. [50] benutzen 8% als Übereinstimmungsintervall. Das Programm läßt Werte zwischen 1% und 15% zu.

Maximalwert der Subharmonischen M :

Der größte theoretisch zulässige Wert ist $M = 20$, vgl. Abschnitt 5.2.5 und 2.4. Ein Wert von $M = 12$ ist laut [50] typisch und ausreichend, um die virtuellen Tonhöhen zu berechnen. Das Programm akzeptiert Werte für M im Bereich von 3 bis 20.

Schwellenwert WS_{imax} :

Der Schwellenwert wird in Kapitel 5.2.5 definiert, dort wird er auch eingehend besprochen. Laut Terhardt [50] ist ein Schwellenwert WS_{imax} von 70% des Gewichts der größten spektralen Tonhöhe ausreichend, um die ausgeprägtesten virtuellen Tonhöhen zu detektieren. Die vorliegende Implementation des Tonhöhenberechnungsalgorithmus erlaubt Werte von 0%, d.h. alle spektralen Tonhöhen werden berücksichtigt, bis zu 90%.

Anzahl der größten zu berücksichtigenden virtuellen Tonhöhen:

Das Programm zur Berechnung der virtuellen Tonhöhe ordnet die virtuellen Tonhöhen nach ihrer Ausprägung: die Tonhöhe mit der größten Ausprägung zuerst, die mit der kleinsten Ausprägung zuletzt. Mittels dieses Parameters ist es möglich, die Anzahl der größten virtuellen Tonhöhen festzulegen, die ausgegeben werden soll. Alle kleineren Tonhöhen werden nicht berücksichtigt. Je nach Zweck der Untersuchung sind verschiedene Werte für diesen Parameter zu benutzen, ein typischer Wert ist 5.

6.3 Test der Modelle

Nachfolgend werden die in Kapitel 6.1 beschriebenen Programmmodule mit den 'optimalen' Parametern aus Kapitel 6.2 getestet. Dafür werden zuerst stationäre Testsignale verwendet, anschließend wird zu zeitabhängigen Signalen übergegangen.

6.3.1 Stationäre Testsignale

6.3.1.1 Sinuston

Als einfachstes Testsignal wird ein reiner Ton, d.h. ein Sinussignal mit einem Pegel von 70 dB und einer Frequenz von 500 Hz verwendet (siehe Abb. 6.1). Nach Kap. 2 ruft dieses Testsignal keine virtuelle Tonhöhe hervor, dafür aber eine sehr ausgeprägte spektrale Tonhöhe [73, Kapitel 5.7].

Die Programme zur Berechnung der virtuellen Tonhöhe berechnen das spektrale Tonhöhengewicht nach Gl. (5.9) zu 0,49 bei einer Frequenz von 499,8 pu. Der

Wert des Tonhöhengewichts ist bereits durch 2 geteilt, vgl. Kap. 5.2.6, d.h. die spektrale Tonhöhe ist als sehr ausgeprägt zu bezeichnen. (Der Maximalwert beträgt 0,5.) Das Programmpaket detektiert keine virtuelle Tonhöhe, wie es der Theorie nach zu erwarten ist.

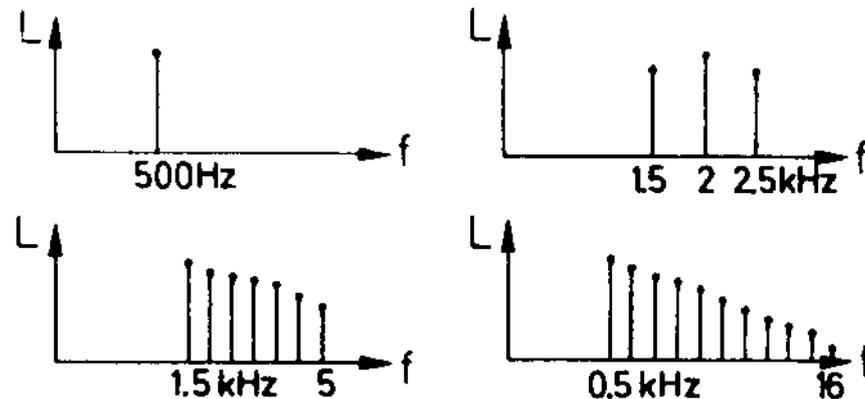


Abb. 6.1: Schema der verwendeten Testsignale. Die größte Spektrallinie innerhalb jedes der vier Darstellungen hat einen Pegel von 70 dB.

oben links: 500 Hz Sinuston

oben rechts: Amplitudenmodulierter Ton, Modulationsgrad 1, Trägerfrequenz 2 kHz, Modulationsfrequenz 500 Hz.

unten links: Residualton, kleinste vorhandene Frequenz 1,5 kHz, Abstand der Frequenzkomponenten 500 Hz, Pegelabfall 6 dB/Oktave, Cut-off bei 5 kHz, entspricht 8 Spektrallinien

unten rechts: vollständiger, harmonischer Klang, Grundfrequenz 500 Hz, Pegelabfall-6 dB/Oktave .

6.3.1.2 Amplitudenmodulierter Ton

Als nächstes wird ein amplitudenmodulierter Ton mit dem Modulationsgrad eins, einer Trägerfrequenz von 2 kHz und einer Modulationsfrequenz von 500 Hz untersucht, siehe Abb. 6.1. Dieses Signal wird als 'stationär' bezeichnet, weil das Gehör nicht in der Lage ist die zeitlichen Fluktuationen, d.h. Schwankungsstärke oder Rauigkeit, wahrzunehmen [7], [52]. Nach der Tonhöhentheorie sollte dieses Signal eine virtuelle Tonhöhe unterhalb von 500 pu hervorrufen, sowie drei spektrale Tonhöhen für 1,5, 2 und 2,5 kpu, deren Gewicht zu höheren Frequenzen kleiner wird [66].

Das Programmpaket liefert folgende spektrale Tonhöhen und zugehörigen Gewichte W : 1501,4 pu mit einem Gewicht von $W = 0,44$, 2031,7 pu mit $W = 0,33$ und 2547,8 pu mit $W = 0,20$. Außerdem werden virtuelle Tonhöhen bei 494,8 und 242,1 pu detektiert. Die erste virtuelle Tonhöhe hat ein Gewicht von 0,32, die zweite von 0,10. Dies ist ein typischer Fall der Oktavunsicherheit, vgl. Kap 5.2.5. Die erste Tonhöhe hat eine deutlich größere Ausprägung (Faktor 3) als die zweite und etwa die doppelte Tonhöhe.

Beim Vergleich der theoretischen Vorhersagen mit den Ergebnissen des Modells kann gute Übereinstimmung festgestellt werden.

6.3.1.3 Residuum

Ein Residualklang mit einer niedrigsten Frequenz von 1,5 kHz, einem Komponentenabstand von 500 Hz, einem Pegelabfall von 6 dB/Oktave, einer Komponentenanzahl von 8 und einem Pegel von 70 dB für die Frequenz von 1,5 kHz wird als Testsignal benutzt, siehe Abb. 6.1.

Das Tonhöhenberechnungsmodell berechnet folgende spektrale Tonhöhen:

Tonhöhe [pu]	Gewicht
1500,4	0,45
2039,1	0,29
2565	0,20
3093,2	0,13
3629,0	0,08
4166,7	0,03
4720,3	0,01
5098,6	0,01

Die virtuelle Tonhöhe wird zu 494,4 pu mit einem Gewicht von 0,55 bestimmt. Wiederum ergibt sich eine Oktavunsicherheit bei 241,9 pu mit dem Gewicht 0,31. Die ausgeprägteste Tonhöhe ist demnach eine virtuelle. Nach den Ausführungen der Kapitel 2 und 5 entspricht dies den Daten für die Tonhöhe von Residualklängen. Beim Anhören des Signals ruft die virtuelle Tonhöhe den dominanten Tonhöhen Eindruck hervor, aber auch die spektrale Tonhöhe, hervorgerufen durch die niedrigste Frequenz des Residualklanges, kann aus dem Testsignal herausgehört werden.

6.3.1.4 Vollständiger harmonischer Klang

In diesem Abschnitt wird ein vollständiger, harmonischer Klang untersucht. Er besitzt eine Grundfrequenz von 500 Hz, welche einen Pegel von 70 dB hat. Der Pegelabfall zu größeren Frequenzen beträgt 6 dB/Oktave. Das Testsignal zeichnet sich dadurch aus, daß alle im Hörbereich liegenden Oberschwingungen im Spektrum vorhanden sind. Daraus ergibt sich eine Gesamtzahl von 32 Spektrallinien im Frequenzbereich zwischen 0,5 und 16 kHz.

Das implementierte Programmpaket berechnet neun spektrale Tonhöhen:

Spektrale Tonhöhe [pu]	Gewicht
499,2	0,48
1005,5	0,44
1516,2	0,38
2037,2	0,30
2558,7	0,22
3083,5	0,15
3613,3	0,10
4145,4	0,06
4702,3	0,02

Weiterhin wird eine virtuelle Tonhöhe von 499,2 pu mit einem Gewicht von 2,03, sowie deren erste und zweite Subharmonische mit 245,7 pu bzw. 162,4 pu bei einem Gewicht von 1,16 bzw. 0,67 berechnet. Die beiden Tonhöhen von 245,7 und 162,4 pu sorgen für eine Oktavunsicherheit. Im Sinne der hier angewandten Theorie wird die Tonhöhe des Signals durch eine virtuelle Tonhöhe beschrieben, da ihr Gewicht wesentlich größer ist als das der spektralen Tonhöhen. Virtuelle Tonhöhen können also auch dann einen entscheidenden Beitrag zur Tonhöhenwahrnehmung leisten, wenn die Grundfrequenz die am stärksten ausgeprägte spektrale Tonhöhe ist. Diese Ergebnisse stimmen gut mit den Daten von Terhardt [66] überein.

Im weiteren wird sukzessive die jeweils kleinste Frequenzkomponente aus dem Spektrum entfernt und jedesmal werden die spektralen und virtuellen Tonhöhen neu berechnet. Die so ermittelten Gewichte der spektralen und virtuellen Tonhöhe zeigt Abb. 6.2.

Aus Abb. 6.2 geht hervor, daß keine virtuelle Tonhöhe mehr erzeugt wird, wenn mehr als acht Komponenten aus dem Spektrum des Testsignals entfernt werden. Dieses Ergebnis deckt sich gut mit Abb. 2.7 des Kapitels 2, in der der Existenz-

bereich der virtuellen Tonhöhe angegeben wird. Für das hier untersuchte Testsignal gilt: Grundfrequenz = 500 Hz, kleinste im Spektrum vorhandene Frequenz = $8 \cdot 500 \text{ Hz} = 4000 \text{ Hz}$. Das Signal fällt also knapp in den Bereich, in dem es eine virtuelle Tonhöhe hervorrufen kann. Wird eine weitere Komponente weggelassen, so ist die kleinste im Spektrum vorhandene Frequenz 4500 Hz und liegt damit nicht mehr innerhalb des Existenzbereiches der virtuellen Tonhöhe, Abb. 2.7.

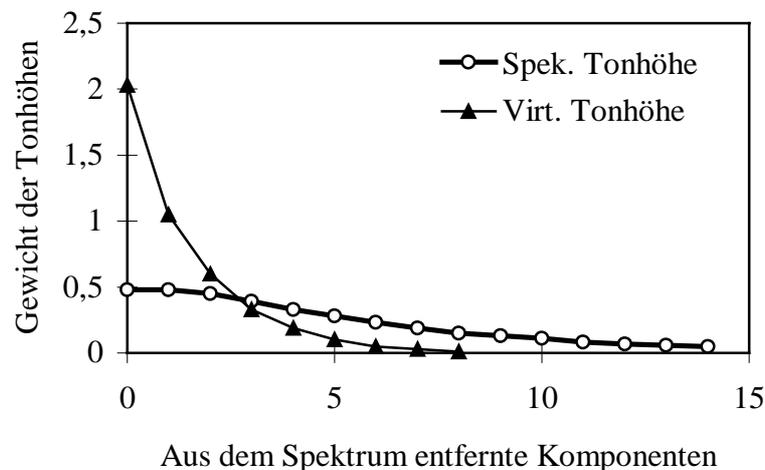


Abb. 6.2: *Gewicht der virtuellen und spektralen Tonhöhen über der Anzahl der aus dem Spektrum des vollständigen Klanges entfernten Komponenten. Abzulesen ist für dieses Signal, daß bei 3 entfernten Spektrallinien die virtuelle und die spektrale Tonhöhe in etwa gleiche Gewichte hervorrufen. Werden mehr Komponenten entfernt, dann überwiegt die spektrale Tonhöhe. Werden weniger Komponenten aus dem Spektrum herausgenommen, ist die virtuelle Tonhöhe dominant.*

Die virtuelle Tonhöhe eines Signals wird wesentlich mehr durch das Entfernen tieffrequenter Spektralanteile beeinflußt als die spektrale Tonhöhe. Deshalb kommt es zu dem Effekt, daß Klänge, die kleine Harmonische der Grundfrequenz im Spektrum enthalten, fast immer virtuelle Tonhöhen hervorrufen, während Klänge, die die kleinen Harmonischen nicht enthalten, eher spektrale Tonhöhen evozieren. Die detektierte virtuelle Tonhöhe in Abhängigkeit von der Anzahl der aus dem Spektrum entfernten Frequenzkomponenten zeigt Abb. 6.3. Nach Abb. 6.3 ist zu erkennen, daß die virtuelle Tonhöhe mit Fortlassen der unteren Spektralkomponenten umso mehr abnimmt, desto mehr Spektrallinien fortgelassen werden. Dieser Effekt wird durch Intervallstreckungen hervorge

rufen und durch Theorie und Messungen bestätigt [53]-[56], [59]. Lediglich für die achte Komponente ergibt sich eine Abweichung für die monoton fallende Kurve in Abb. 6.3. Das Gewicht dieser virtuellen Tonhöhe ($< 0,01$) ist jedoch viel kleiner als das Gewicht aller anderen Tonhöhen, so daß sie nicht weiter betrachtet werden muß.

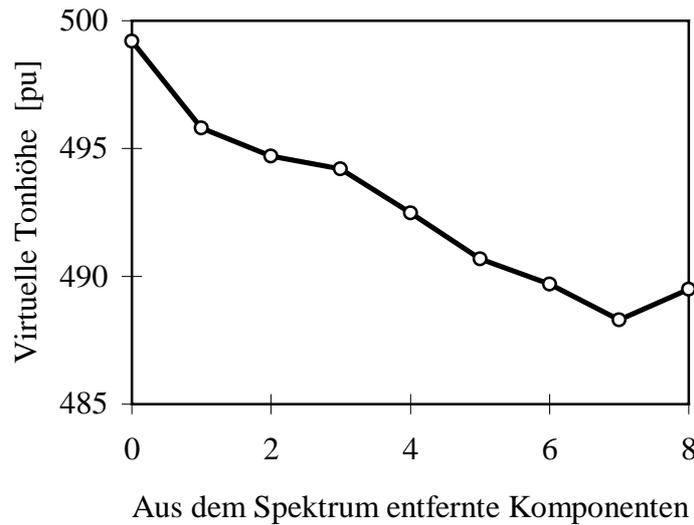


Abb. 6.3: *Einfluß der Anzahl der aus dem Spektrum des Testsignals entfernten Spektrallinien auf die berechnete virtuelle Tonhöhe.*

6.3.1.5 Residuum mit zusätzlichen Tönen

Zu dem Signal aus 6.2.1.3 wird eine zusätzliche Spektrallinie addiert und die berechneten Tonhöhen werden mit denen des Originalsignals verglichen. Die zusätzliche Spektralkomponente wird willkürlich aus dem Frequenzbereich zwischen 500 und 4000 Hz gewählt, der zugehörige Pegel wird aus dem Intervall zwischen 40 und 70 dB ausgesucht.

Es zeigt sich, daß der dominante Tonhöhen Eindruck dabei nur unwesentlich verändert wird. Die bestimmende Tonhöhe bleibt die virtuelle Tonhöhe bei 494,4 pu, die lediglich minimal ($< 1\%$) variiert. Das Gewicht der virtuellen Tonhöhe schwankt in einem größeren Bereich, jedoch bleibt es immer größer als das Gewicht der größten spektralen Tonhöhe. Die spektralen Tonhöhen selber ändern sich aufgrund von Maskierungseffekten (vgl. Kap. 5.2.2 und 5.2.3) stärker als die virtuelle Tonhöhe. Die Rangfolge der Spektraltonhöhen gewichte hängt von der genauen Frequenz und dem Pegel des zusätzlichen Tones ab.

6.3.1.6 Zufällig angeordnete Sinustöne

Als letztes wird ein völlig inharmonisches Testsignal untersucht. Dafür wird das Testsignal aus folgenden Frequenzen zusammengestellt, die zufällig ausgewählte Primzahlen zwischen 500 und 4000 Hz sind: 421 Hz, 743 Hz, 1153 Hz, 1429 Hz, 1871 Hz, 2713 Hz, 2833 Hz, 3313 Hz. Der kleinsten Frequenz wird ein Pegel von 70 dB zugeordnet, alle anderen Pegel werden mit -6 dB/Oktave hieraus bestimmt.

Das Programmpaket berechnet die folgenden spektralen Tonhöhen:

Tonhöhe [pu]	Gewicht
419,2	0,47
754,7	0,38
1169,4	0,40
1469,4	0,27
1898,9	0,31
2799,1	0,16
3380,2	0,08

Die drei ausgeprägtesten virtuellen Tonhöhen ergeben sich zu 101,1 pu, 205,9 pu und 80,2 pu. Ihre Gewichte betragen 0,26, 0,18 und 0,09.

Die dominantesten Tonhöhen des Testsignals sind daher die spektralen Tonhöhen von 419,2, 1169,4 und 754,7 pu. Dieses Ergebnis stimmt gut mit der Theorie des Kapitels 2 überein.

An diesem Beispiel wird der Einfluß des Übereinstimmungsintervalls δ nach Gl (5.19) deutlich. Je größer δ gewählt wird, desto mehr steigt das Gewicht der virtuellen Tonhöhen an. Das Übereinstimmungsintervall δ ist daher von entscheidender Bedeutung für die Detektion von virtuellen Tonhöhen. Besonders ihr Gewicht, und damit die Vergleichbarkeit der virtuellen und spektralen Tonhöhen, hängt erheblich von der Wahl des Übereinstimmungsintervalls ab. Im Extremfall erreichen die Gewichte der virtuellen Tonhöhen dieses Beispiels größere Werte als die Gewichte der spektralen Tonhöhen und tragen somit zum dominanten Tonhöhen Eindruck bei. Um diese Effekte zu vermeiden, müssen die Parameter des Kapitels 6.2 stets als Gesamtheit betrachtet werden. In diesem Fall müßte bei Vergrößerung von δ das Verhältnis von virtuellem und spektralem Tonhöhengewicht von 1:2 ebenfalls abgeändert werden.

6.3.2 Nichtstationäre Testsignale

Die Testsignale des vorangehenden Abschnitts sind allesamt stationäre Signale, jetzt soll das Modell für zeitlich schwankende Signale überprüft werden.

6.3.2.1 Nichtharmonische Klänge

Ein nichtharmonischer Klang entsteht, wenn die einzelnen Spektrallinien eines harmonischen Klanges um einen bestimmten Betrag verschoben werden, der kein Vielfaches der Grundfrequenz ist.

Als ein Beispiel werden die Spektrallinien eines harmonischen Klanges, der fünf Spektrallinien enthält, zeitabhängig verschoben. Die Spektrallinien haben zu Beginn des Testsignals Frequenzen von 1200, 1500, 1800, 2100 und 2400 Hz. Alle fünf Komponenten werden gemeinsam mit 200 Hz/sec zu größeren Frequenzen hin verschoben. Das Testsignal wird über drei Sekunden lang generiert, d.h. die Spektrallinien haben am Ende des Signals Frequenzen von 1800, 2100, 2400, 2700 und 3000 Hz.

Für dieses Signal berechnet das Programmpaket die virtuellen Tonhöhen, die in Abb. 6.4 dargestellt sind. Aufgetragen sind dort die Tonhöhen über der kleinsten im Spektrum des Signals vorhandenen Frequenz f_k . Zu dieser gelangt man durch eine lineare Transformation der Zeit t nach folgender Formel:

$$f_k = 1200 + 200 \cdot t_{[\text{sec}]} \text{ [Hz]}.$$

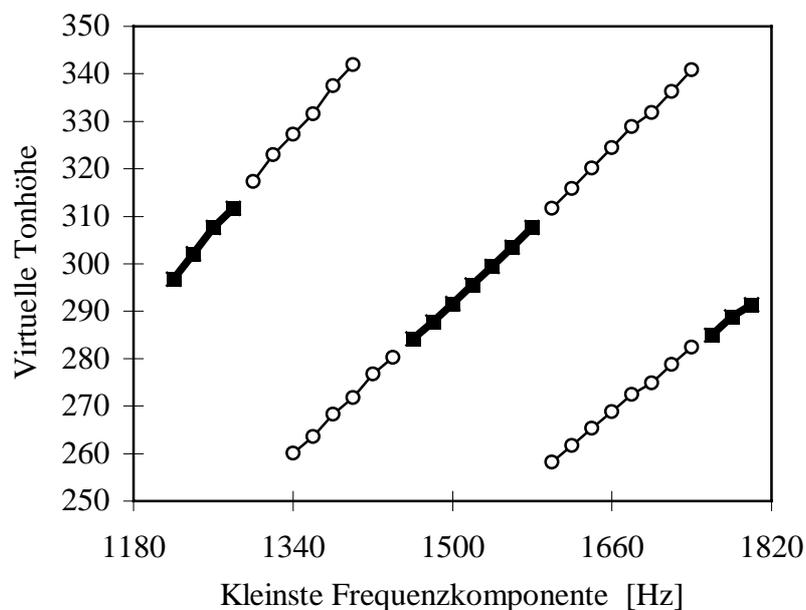


Abb. 6.4: Virtuelle Tonhöhe über der kleinsten Frequenzkomponente des Spektrums für das Testsignal dieses Abschnittes. Das Symbol ■ steht für eindeutig bestimmte virtuelle Tonhöhen, ○ bedeutet nicht eindeutig bestimmte virtuelle Tonhöhen, siehe Text.

Zu Beginn der Analyse wird ein harmonischer Klang untersucht, da die Verschiebung der Spektrallinien noch Null beträgt. Es ergibt sich eine virtuelle Tonhöhe von 296,8 pu mit einem Gewicht von 0,44. Diese Tonhöhe ist die dominante Tonhöhe zu diesem Zeitpunkt, d.h. sie ist eindeutig bestimmt. Solche Tonhöhen sind in Abb. 6.4 durch ■ bezeichnet. Mit fortschreitender Zeit wird die Verschiebung des inharmonischen Tones größer. Dies hat zunächst nur den Einfluß, daß die virtuelle Tonhöhe etwas ansteigt und das Gewicht abfällt. Nach 0,4 sec beträgt die Verschiebung 80 Hz, die virtuelle Tonhöhe beträgt 317,3 pu bei einem Gewicht von 0,3. Die nächst ausgeprägte virtuelle Tonhöhe ergibt sich zu 252,2 pu mit einem Gewicht von 0,19. In diesem Fall ist das Gewicht der ersten virtuellen Tonhöhe noch deutlich größer als das der zweiten. Bei 0,5 sec, oder einer Verschiebung von 100 Hz, wird die virtuelle Tonhöhe bei 323,0 pu detektiert, ihr Gewicht beträgt 0,25. Die nächstgrößere virtuelle Tonhöhe hat bereits ein Gewicht von 0,2 bei 256,7 pu. In diesem Fall ist der Unterschied der Tonhöhengewichte sehr gering und die Entscheidung, welcher der beiden der dominante ist, unsicher. Solche Tonhöhen werden in Abb. 6.4 durch ein ○ gekennzeichnet. Bei 0,7 sec, d.h. die kleinste Frequenz des Testsignals beträgt 1340 Hz, werden Tonhöhen bei 331,6 und 263,6 pu berechnet. Die Gewichte haben einen Werte von 0,22 bzw 0,21. Diese Gewichte sind so ähnlich, daß eine Entscheidung, welche Tonhöhe die dominantere ist, nicht möglich ist. Deshalb sind für diesen Zeitpunkt zwei virtuelle Tonhöhen in Abb. 6.4 eingetragen. Dieser Zustand, in dem keine der beiden größten virtuellen Tonhöhe deutlich dominanter ist als die andere, hält bis 1,0 sec (entspricht 1400 Hz in Abb. 6.4) an. Bei 1,1 sec werden folgende Tonhöhen ermittelt: 232,3 pu (Gewicht 0,17) und 280,3 pu (Gewicht 0,26). Die dominantere Tonhöhe ist in Abb. 6.4 als 'nicht eindeutig' gekennzeichnet. Ab diesem Zeitpunkt läuft die beschriebene zeitliche Entwicklung rückwärts, bis nach 1,5 sec (das entspricht einer Frequenz von 1500 Hz der kleinsten Frequenzkomponente) wieder ein harmonischer Klang erreicht wird. Im Zeitbereich von 1,5 bis 3 sec wiederholt sich der Ablauf wie erläutert. Diese Ergebnisse korrelieren gut mit den in [58] beschriebenen Daten.

7 ANWENDUNG DES MODELLS AUF UMWELTGERÄUSCHE

In diesem Kapitel wird der Algorithmus zur Bestimmung virtueller und spektraler Tonhöhen auf Umweltschalle angewendet. Die Umweltschalle sind verschiedenen CDs entnommen, siehe entsprechende Quellenhinweise bei den Geräuschen. Da der Originalpegel der verschiedenen Geräusche auf den CDs nicht bekannt ist, wird für jedes Signal der maximale Pegel des Leistungsspektrums auf 70 dB kalibriert. Das FTT-Spektrum sowie die spektralen und virtuellen Tonhöhen aller Signale werden mit den Parametern des Kapitels 2.2 berechnet.

Als erstes Geräusch wird eine Autohupe [60] untersucht. Beim Anhören des Signals ist eine deutliche Tonhöhe wahrzunehmen, die etwa 1 sec andauert. Genauere Informationen können dem FTT-Leistungsspektrum des Signals, Abb. 7.1, entnommen werden. Das FTT-Spektrum beinhaltet ca. zehn deutlich ausgeprägte Spektrallinien, die alle zeitlich stationär über etwa 1,1 sec vorhanden sind. Zwei Spektrallinien sind wesentlich stärker ausgeprägt als alle anderen, ihre Frequenzen betragen ca. 300 Hz (3 Bark), bzw. 3150 Hz (16 Bark). Anhand des FTT-Spektrums ist die Tonhöhe des Hupsignals aufgrund der großen Anzahl der Spektrallinien nicht vorhersagbar.

Abb. 7.2 stellt die berechnete virtuelle und spektrale Tonhöhe des Hupsignals dar. In dem Zeitbereich von 0,1 bis 1,15 sec ist die Tonhöhe weitgehend zeitunabhängig, daher lassen sich die drei ausgeprägtesten virtuellen und die ausgeprägteste spektrale Tonhöhe für diesen Zeitbereich folgendermaßen angeben:

Tonhöhe	Art	Gewicht
283 pu	virtuell	2,39
138 pu	virtuell	1,04
570 pu	virtuell	0,77
590 pu	spektral	0,44

Aus dieser Auflistung geht hervor, daß die dominante Tonhöhe des Hupsignals bei ca. 283 pu liegt, da diese die Tonhöhe bei weitem größte Gewicht besitzt. In dem Moment, in dem die Hupe abgestellt wird, d.h. nach etwa 1,15 sec in Abb. 7.1, bzw. Abb. 7.2, sinkt die Tonhöhe zunächst etwas ab, bevor sie ganz verschwindet. Diese Ergebnisse entsprechen gut dem subjektiven Höreindruck des Signals.

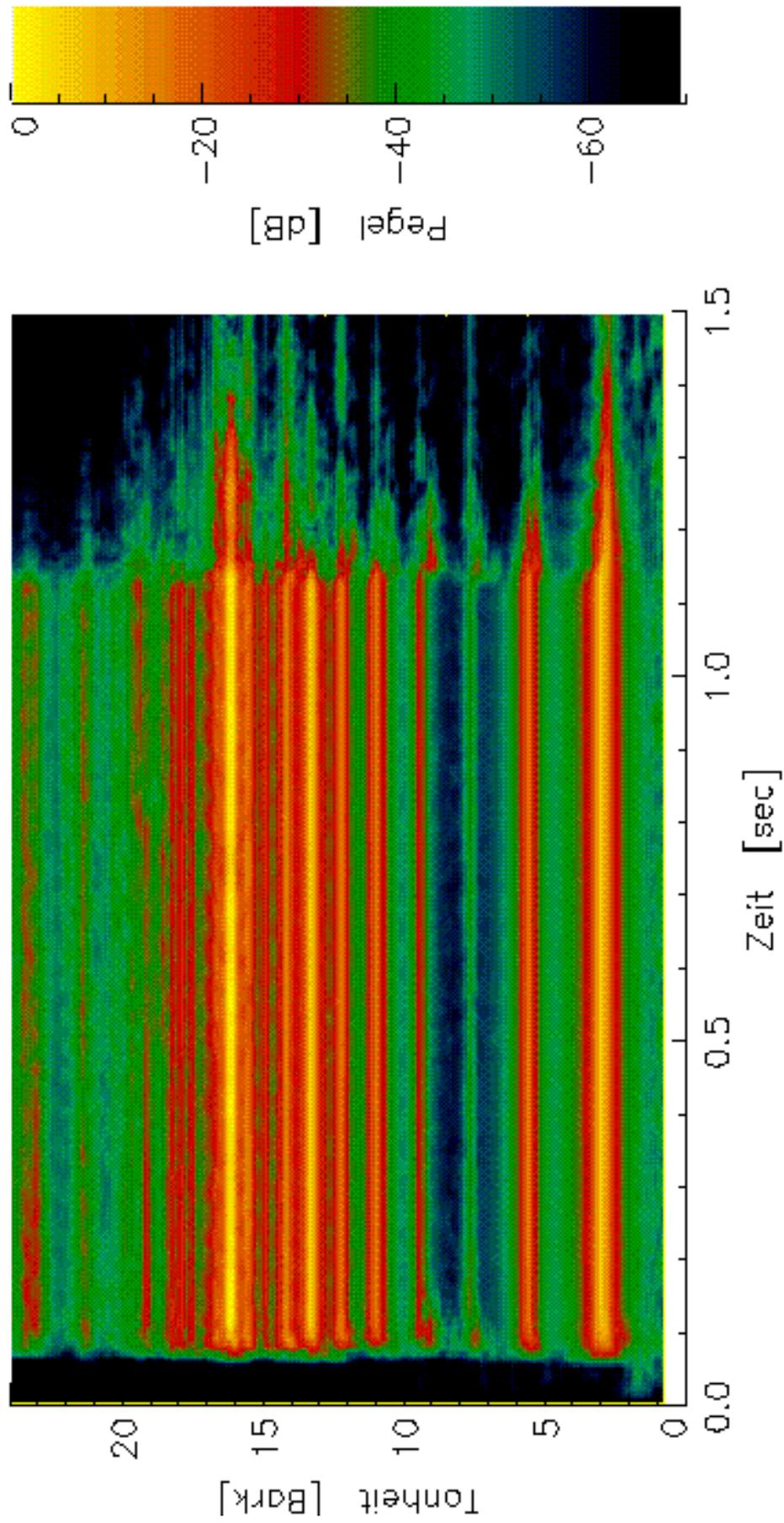


Abb. 7.1:FTT-Spektrum für das Hupsignal.

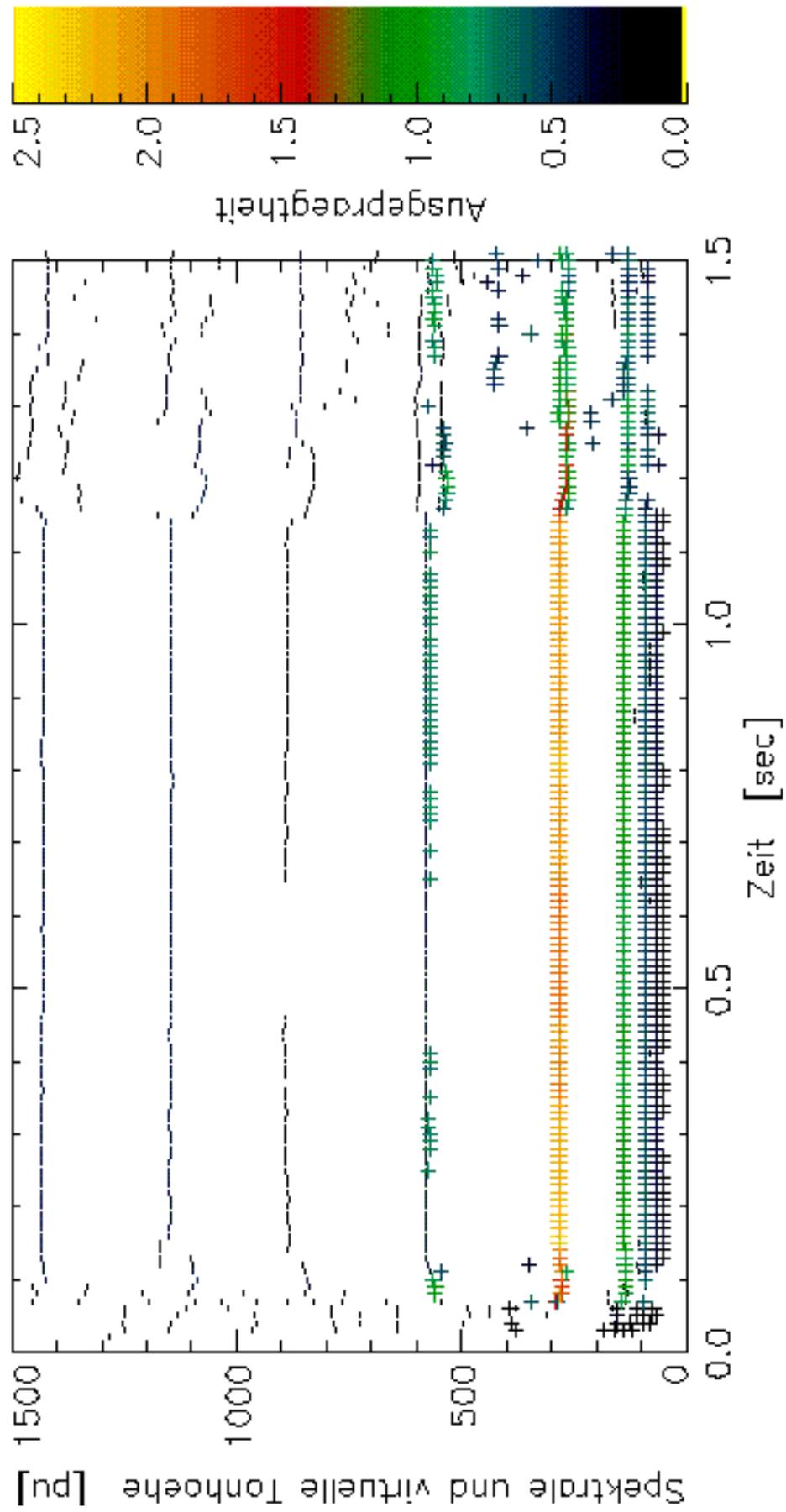


Abb. 7.2: Virtuelle und spektrale Tonhöhen-Diagramm für das Hupsignal.

Als nächstes wird das Signal einer Sirene untersucht, die das Signal 'Luftangriff' gibt [61]. Aufgenommen sind die ersten 9,2 sec nach dem Beginn des Signals. Der Zuhörer nimmt ein periodisch auf- und abschwellendes Geräusch wahr, welches seine Tonhöhe zyklisch ändert. Diese Signaleigenschaften sind auch im FTT-Spektrum, Abb. 7.3, sichtbar. Den zeitlichen Verlauf der Tonhöhen zeigt Abb. 7.4. Zu Beginn des Signals bei 0,1 sec ist die Tonhöhe noch sehr undeutlich. Es existieren mehrere, eng benachbarte Tonhöhen, die alle gleiche, sehr geringe Gewichte besitzen. Aus diesem Grund wird dem Sirensignal zu dieser Zeit keine dominante Tonhöhe zugeordnet. Im Zeitbereich von 0,1 bis ca. 2,0 sec bildet sich langsam eine dominante (virtuelle) Tonhöhe aus. Die Tonhöhe wächst in diesem Zeitbereich von ca. 90 auf ca. 347 pu an. Bei etwa 2,0 sec besitzt die Tonhöhe eine Wendestelle, d.h. mit fortschreitender Zeit sinkt die Tonhöhe wieder ab, bis sie bei 4,3 sec erneut ansteigt. Dieser Vorgang wiederholt sich zyklisch mit dem An- und Abschwellen des Sirengeräusches.

Während des ersten Anstiegs der Tonhöhe im Zeitbereich von 0,1 bis 2 sec steigt die Ausprägtheit der dominanten Tonhöhe kontinuierlich an und erreicht bei 2 sec einen Wert von ca. 2,1. Im Zeitbereich nach 2,0 sec ändert sich zwar die Tonhöhe des Sirensignals, ihre Ausprägung bleibt jedoch ziemlich konstant bei einem Wert um 2,2.

Die Ergebnisse der Tonhöhenanalyse werden durch den akustischen Eindruck des Sirensignals bestätigt.

Das FTT-Spektrum eines näherkommenden Martinshornes [62] zeigt Abb. 7.5. Aus ihm kann eine Frequenzmodulation des Hornsignals mit etwa 14 Hz abgelesen werden. Die beiden unterschiedlichen Heultöne, die jeweils 0,7 sec lang sind und sich periodisch abwechseln, können im Spektrum identifiziert werden. Die Intensitätszunahme der spektralen Komponenten mit fortschreitender Zeit ergibt sich durch das Näherkommen des Horns. Um die Tonhöhe des Signals zu ermitteln, wird das Tonhöhendigramm, Abb. 7.6, berechnet. Aus diesem werden zwei dominante (virtuelle) Tonhöhen bestimmt. Diese Tonhöhen können den beiden Heultönen der Sirene zugeordnet werden. Die größere Tonhöhe hat einen Wert zwischen 570 und 590 pu, die kleinere zwischen 425 und 445 pu. Die Tonhöhen schwanken aufgrund der Frequenzmodulation, die im FTT-Spektrum bereits zu erkennen war. Damit ergeben sich folgende Werte für die beiden ausgeprägtesten (virtuellen) Tonhöhen:

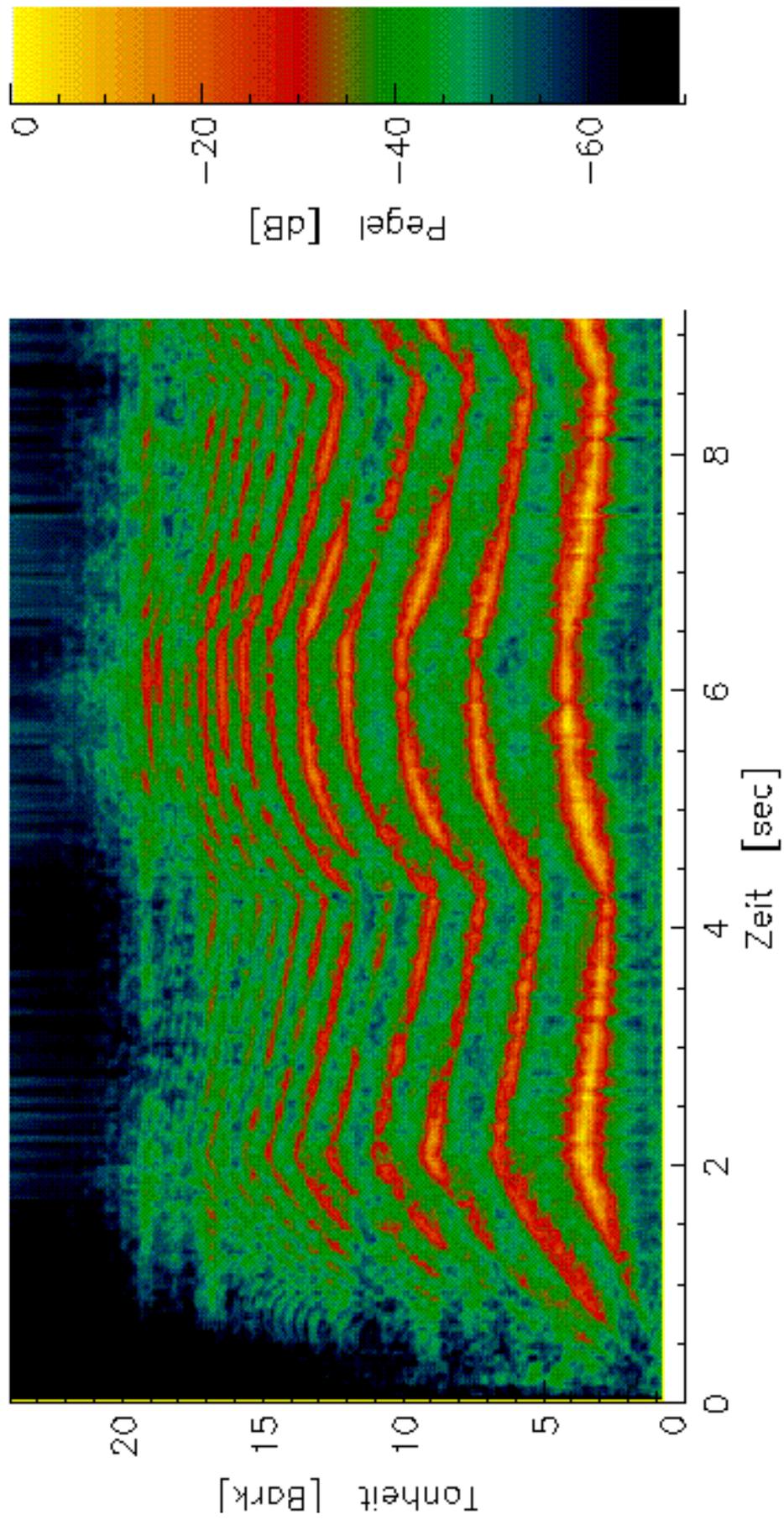


Abb. 7.3: FTT-Spektrum für das Signal der Sirene.

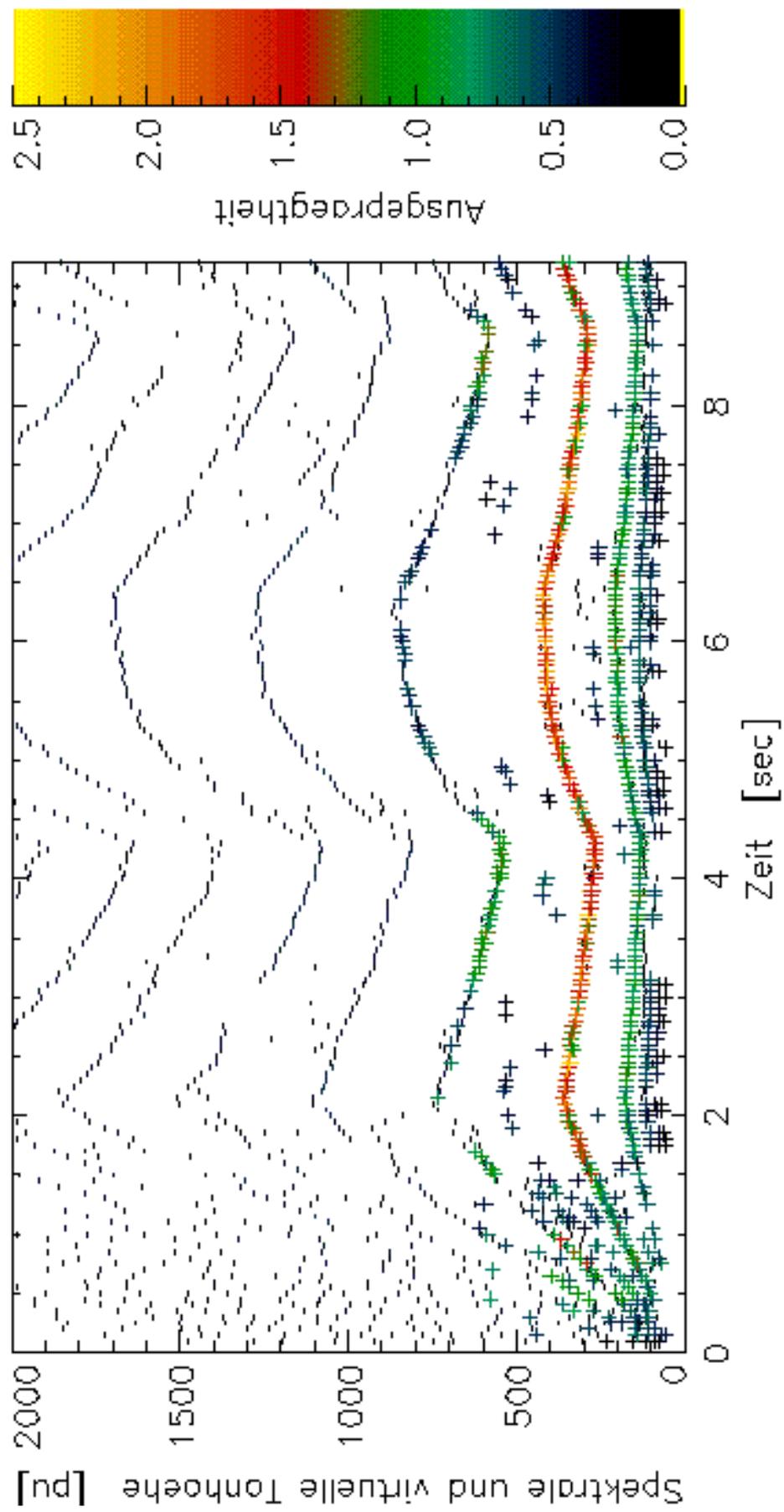


Abb. 7.4: Virtuelle und spektrale Tonhöhen-Diagramm für das Signal der Sirene.

Heulton 1		Heulton 2	
Tonhöhe [pu]	Gewicht	Tonhöhe [pu]	Gewicht
570-590	≈1,3	425-445	≈1,3
ca. 290	≈0,75	ca. 215	≈0,65

Hier ist das Gewicht der Tonhöhe bei 290 pu von ca. 0,75 größer als das Gewicht der Tonhöhe 215 pu von ca. 0,65.

Die dargestellten Ergebnisse werden durch Anhören der Schalle des Martinshornes bestätigt: Das Martinshorn ruft zwei ausgeprägte, sich abwechselnde Tonhöhen hervor.

Als nächstes wird die Tonhöhe von Kirchenglocken [63] analysiert. Dazu steht ein Schall zur Verfügung, der drei Glockenschläge verschiedener Glocken nacheinander enthält. Beim Anhören des Signals der drei Glocken nimmt die Tonhöhenwahrnehmung von Schlag zu Schlag ab, die Lautstärke zu. Die Lautstärkezunahme ist auch im FTT-Spektrum, Abb. 7.7 nachzuvollziehen. Die Abnahme der Tonhöhe kann nicht unmittelbar aus dem Spektrum entnommen werden, da eine große Anzahl Spektrallinien im Spektrum vorhanden ist und aus diesem Grund die Bestimmung der dominanten Tonhöhe nicht möglich ist. Wird als erste Näherung der Schwerpunkt der Spektrallinien für jeden einzelnen Glockenschlag betrachtet, so ist eine Verschiebung zu tieferen Tonhöhen erklärbar. Um genauere Aussagen zu ermöglichen, die dem Gehör zudem besser angepaßt sind, wird das Tonhöhendigramm, Abb. 7.8, für das Glockensignal berechnet. Aus ihm können dem ersten Glockenschlag die folgenden drei ausgeprägtesten virtuellen Tonhöhen zugeordnet werden:

Tonhöhe [pu]	Gewicht
510	≈1,15
162	≈0,87
345	≈0,8

Im Bereich von 0,3 bis 0,7 Sekunden werden etwas andere Tonhöhen detektiert. Die ausgeprägteste Tonhöhe bei 227 pu hat ein Gewicht von 0,82, die nächstausgeprägten Tonhöhen bei 344 und 516 pu haben ein Gewicht von 0,79

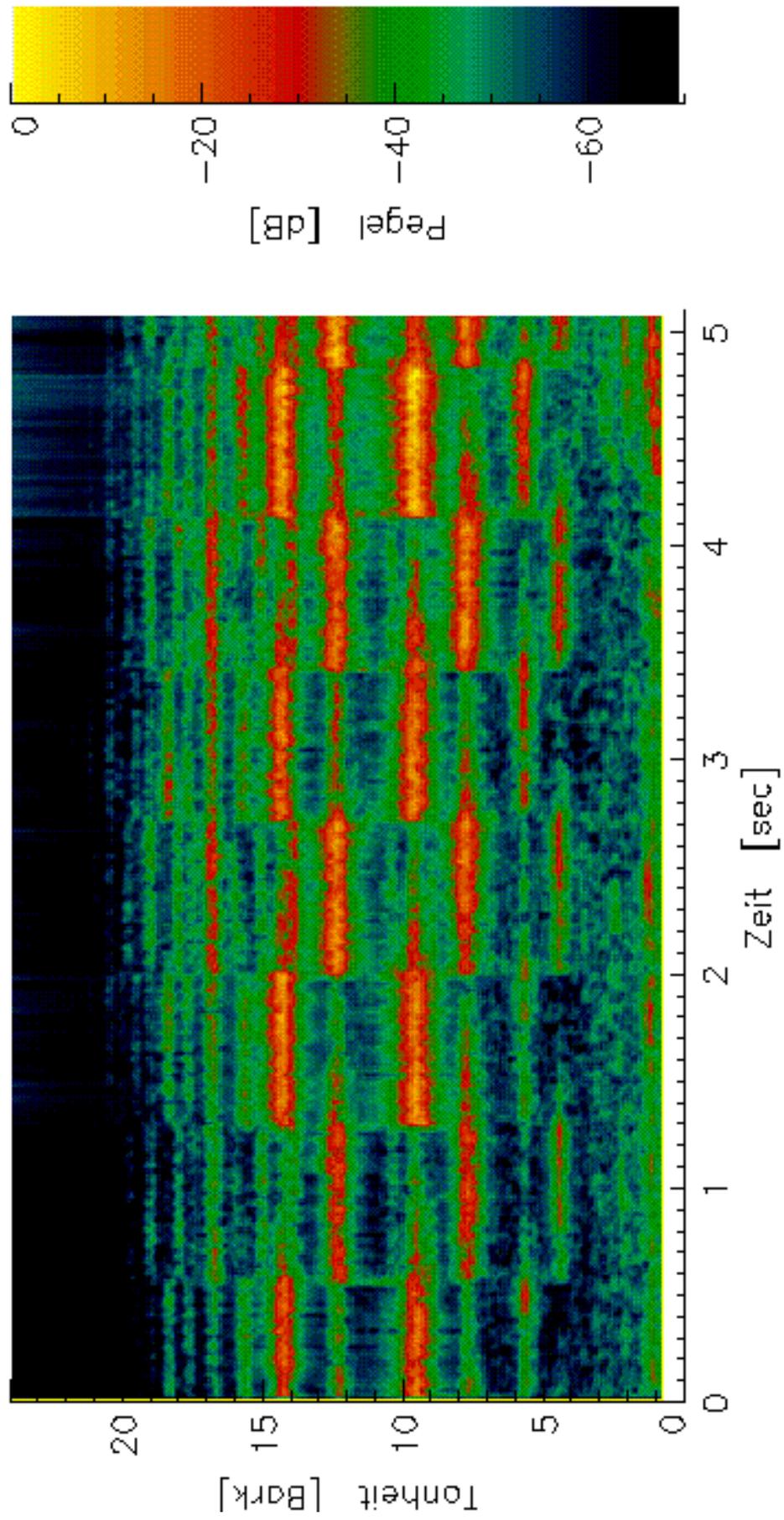


Abb. 7.5: FTT-Spektrum für das Signal des Martinshorns..

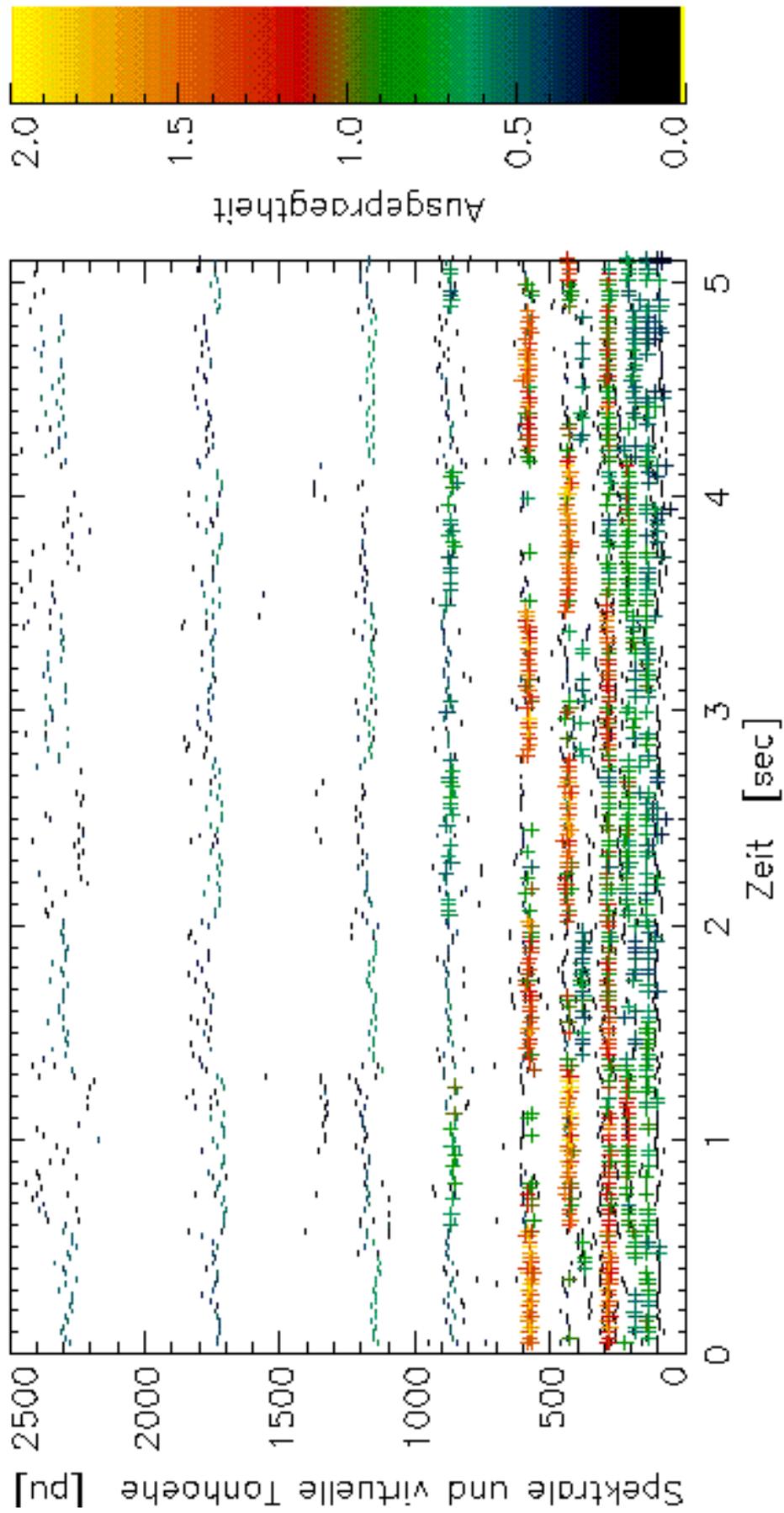


Abb. 7.6: Virtuelle und spektrale Tonhöhen-Diagramm für das Signal des Martinshorns.

bzw. 0,65. Im Zeitsignal ist allerdings durch Anhören des Signals keine Änderung der dominanten Tonhöhe zu bemerken.

Der zweite Glockenschlag besitzt vier etwa gleich ausgeprägte Tonhöhen:

Tonhöhe [pu]	Gewicht
347	≈0,7
171	≈0,7
ca. 433	≈0,65
ca. 256	≈0,65

Diese Verhältnisse sind jedoch so zeitabhängig, daß keine Rangfolge der Tonhöhen über die gesamte Dauer des zweiten Glockeschlages erfolgen kann. Aus diesem Grund existiert keine dominante Tonhöhe. Dieser Effekt, daß mehrere Tonhöhen mit nahezu gleicher Ausprägung auftreten ist für Glockenklänge bekannt [34]. Er ist eine Folge der Oktavunsicherheit, die für Glocken häufig vorkommt [34]. Das menschliche Gehör ist jedoch in der Lage diesem Glockenschlag eine eindeutige virtuelle Tonhöhe zuzuordnen

Der dritte Glockenschlag hat eine Tonhöhe von ca. 373 pu bei einem Gewicht von etwa 1,05. Die nächst ausgeprägte Tonhöhe liegt bei 246 pu mit einem Gewicht von 0,74. Durch den großen Unterschied der Gewichte kann die dominante Tonhöhe eindeutig als 373 pu identifiziert werden.

Eine ähnliche Aufgabe wie die Vorhersage von Tonhöhen ist die Berechnung der Tonhaltigkeit, vgl. [7]. Die psychoakustische Empfindungsgröße Tonhaltigkeit gibt an, ob und wie stark einzelne Töne oder tonale Komponenten in einem Testsignal wahrnehmbar sind. Tonhaltige Geräuschmissionen, z.B. von Windenergiekonvertern, sind in der Regel besonders beeinträchtigend. Daher sind in Meßverfahren für Geräuschmissionen Tonzuschläge zum äquivalenten Dauerschallpegel vorgesehen. Der korrekten Erfassung der Tonhaltigkeit kommt deshalb eine wichtige Rolle zu [7].

Als letztes soll das vorliegende Tonhöhenberechnungsprogramm dazu eingesetzt werden, die Tonhöhe des von einem Windenergiekonverter hervorgerufenen Schalles zu berechnen.

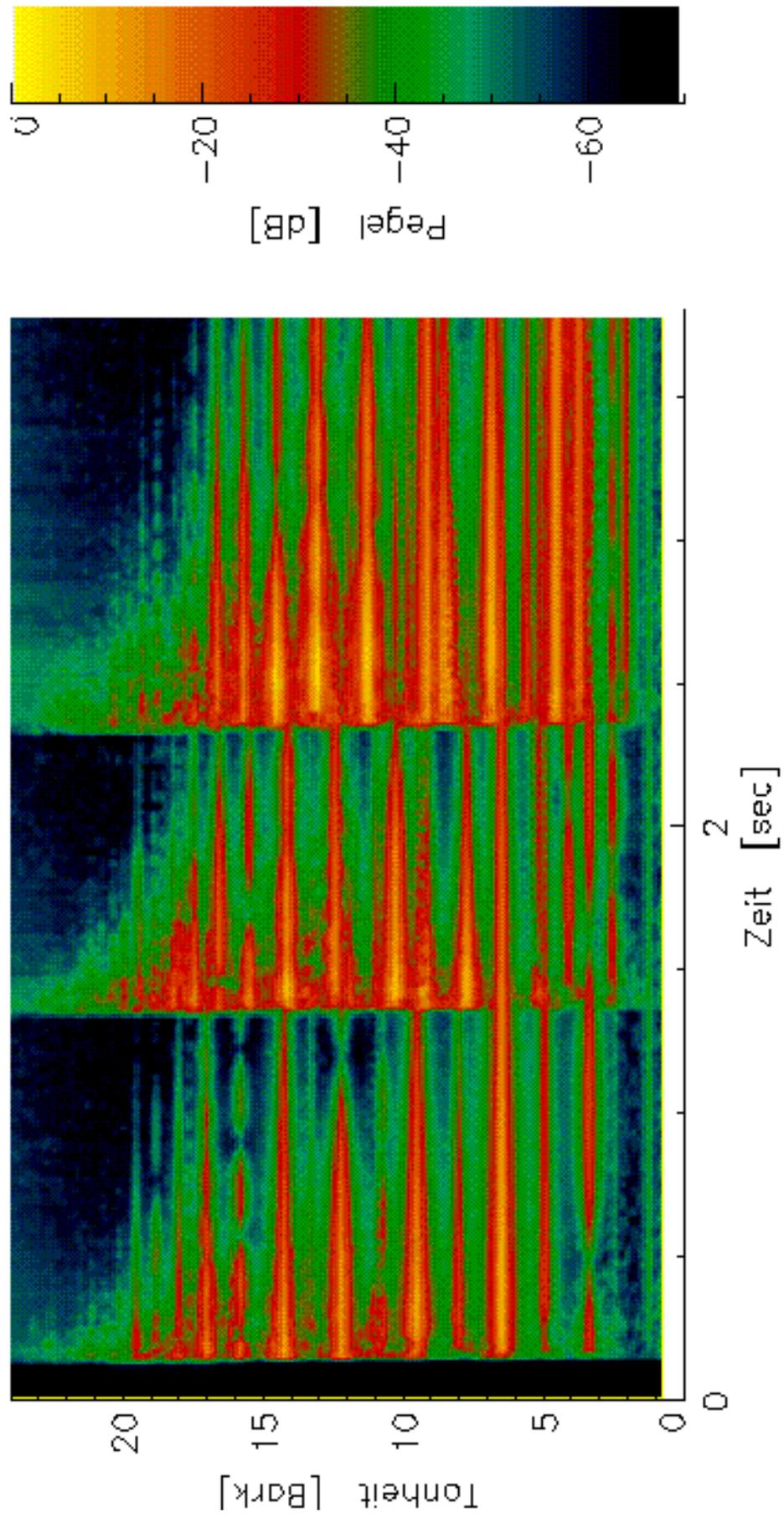


Abb. 7.7: FTT-Spektrum für das Signal der Glocke.

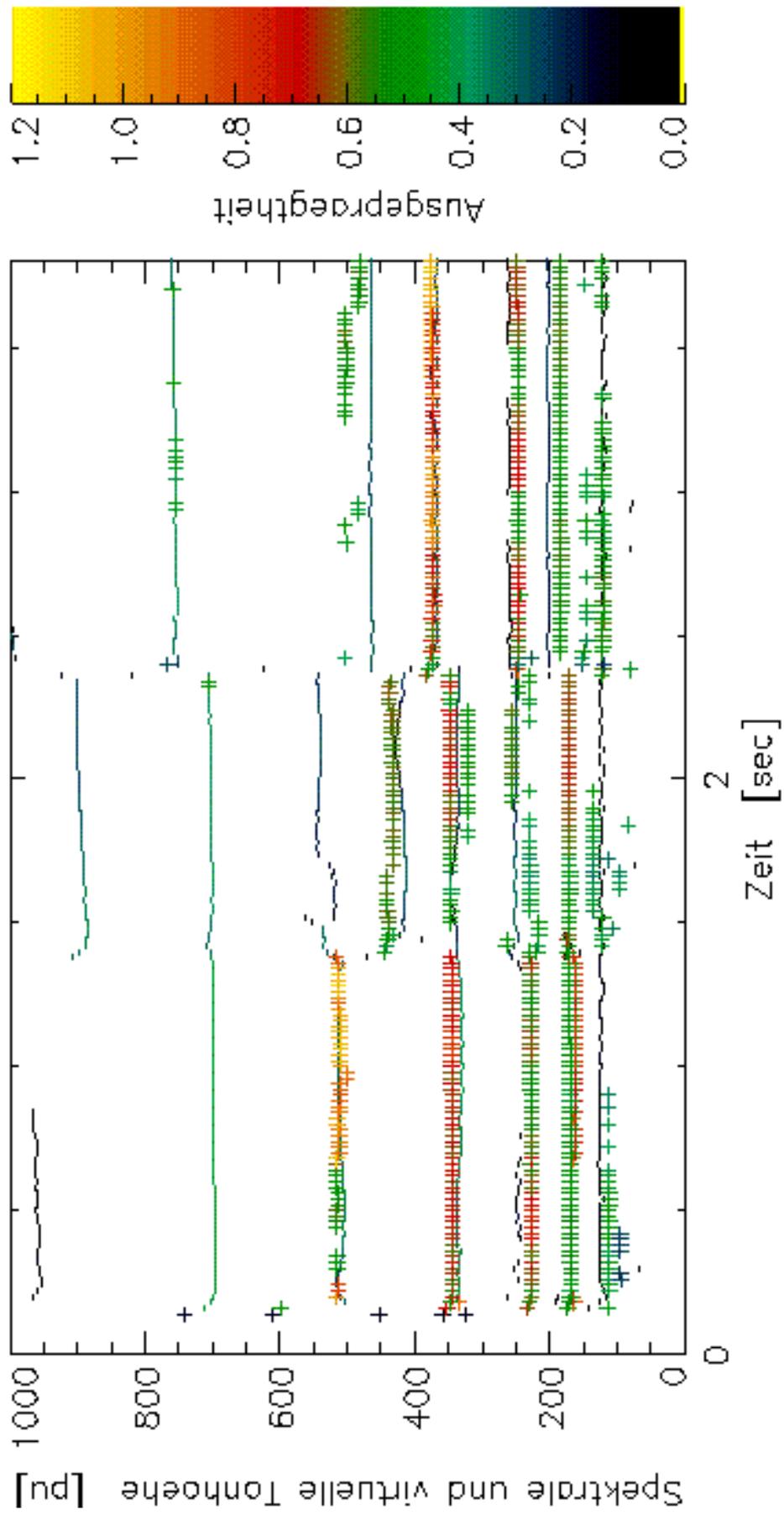


Abb. 7.8: Virtuelle und spektrale Tonhöhen-Diagramm des Signals der Glocke.

Es werden die Tonhöhen der Geräuschmissionen eines Windenergiekonverters mit dem hier vorgestellten Tonhöhenberechnungsmodell analysiert. Das FTT-Leistungsspektrum dieses Signals zeigt Abb. 7.9. In dieser Abbildung ist eine tonale Komponente bei etwa 5,5 Bark (ca. 550 Hz) zu erkennen. Ihr Pegel schwankt mit einer Frequenz von ungefähr 1 Hz. Eine Pegelabnahme von tiefen zu hohen Frequenzen ist über den ganzen analysierten Zeitraum abzulesen.

Die tonale Komponente, welche im FTT- Spektrum zu beobachten ist, kann bei genauem Zuhören als zeitlich schwankende, d.h. instationäre tonale Komponente in den Geräuschmissionen wahrgenommen werden.

Die berechneten virtuellen und spektralen Tonhöhen sind aus dem Tonhöhendigramm, Abb. 7.10 zu entnehmen. Die dominante (virtuelle) Tonhöhe liegt etwa im Bereich zwischen 550 und 600 pu. Ihre Ausprägung ist nicht besonders stark. Auch in Abb. 7.10 ist die Modulation mit etwa einem Hertz zu erkennen, die sich in diesem Diagramm als Schwankung in der Ausgeprägtheit des Signals offenbaren. Durch die Oktavunsicherheit treten in Abb. 7.10 zusätzlich tonale Komponenten bei etwa 290 und 190 pu hervor, die Ausgeprägtheit dieser Komponenten ist aber deutlich kleiner als die der tonale Komponente bei 575 Hz sind.

Dieses Beispiel läßt vermuten, daß eine Berücksichtigung der hier vorgestellten Verfahren für die Bestimmung der Tonhaltigkeit lohnend sein könnte. In jedem Fall scheinen weitere Untersuchungen der Tonhaltigkeit in Verbindung mit der Fourier-Zeit-Transformation und dem Teiltonlinienzeitmuster vielversprechende Ansätze zu bieten, um die Tonhaltigkeit aus beliebigen Signalen quantitativ berechnen zu können.

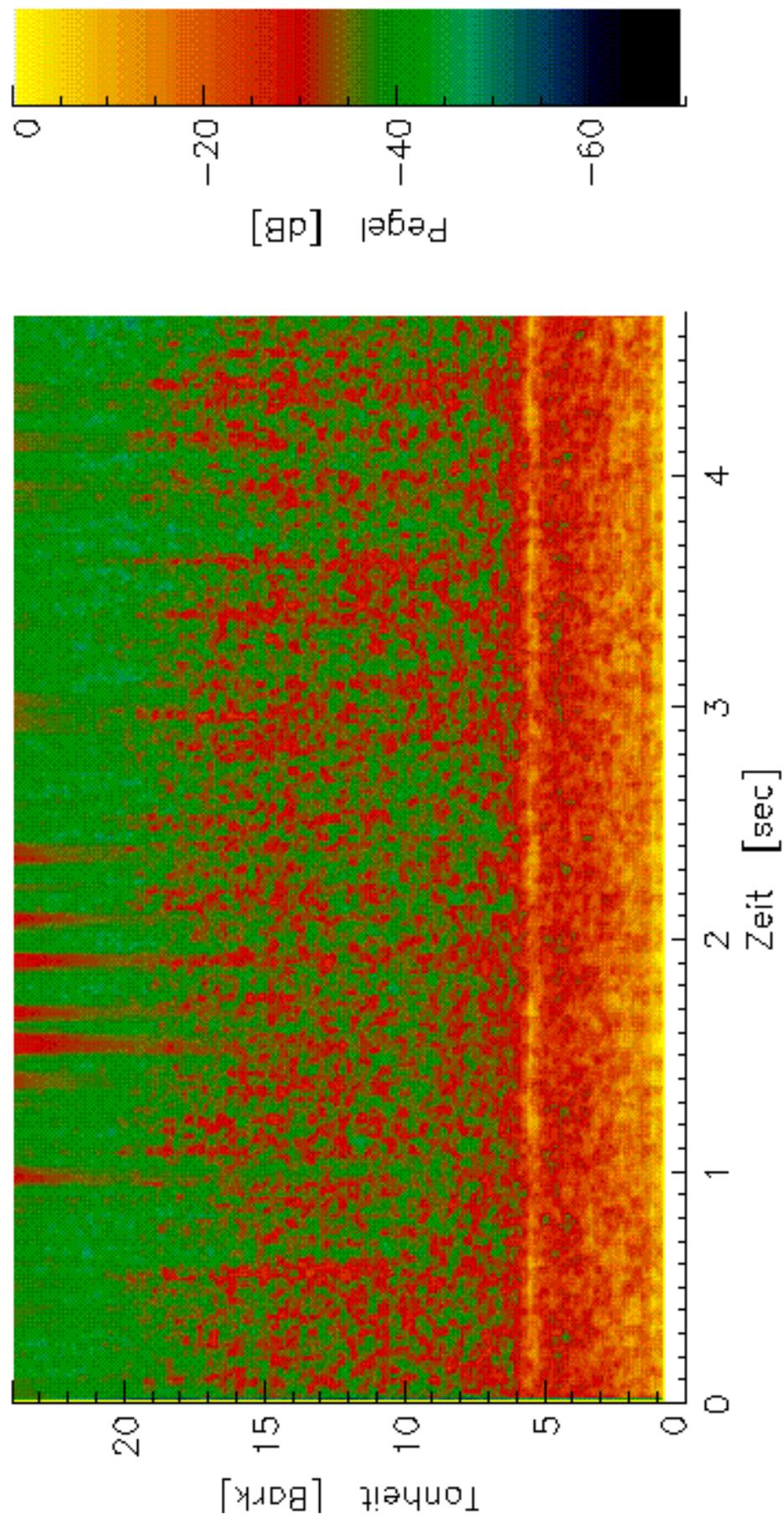


Abb. 7.9: FFT-Spektrum für das Signal des Windenergieconverters.

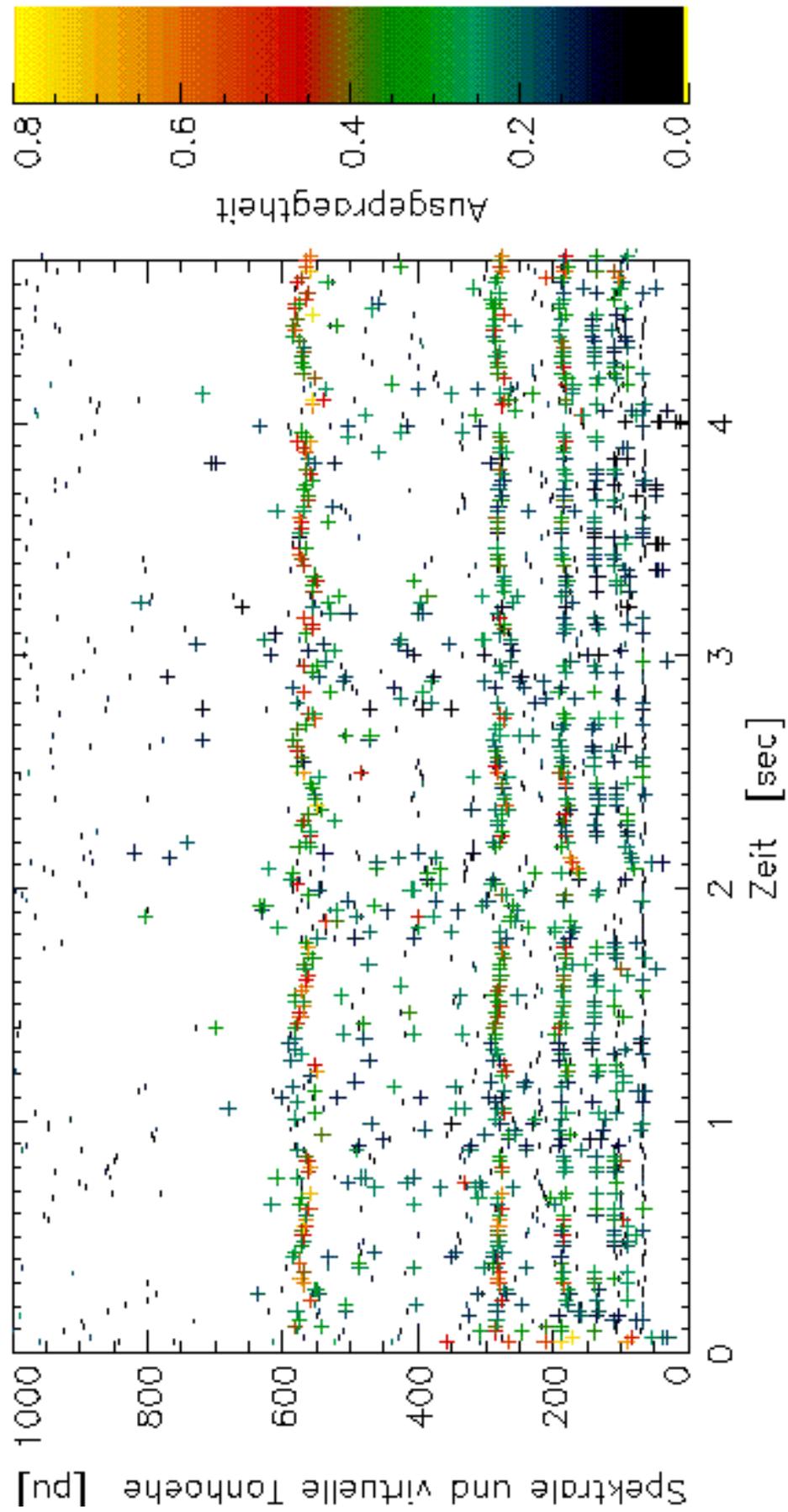


Abb. 7.10: Virtuelle und spektrale Tonhöhen-Diagramm des Signals des Windenergiekonverters.

8 ZUSAMMENFASSUNG UND AUSBLICK

In dieser Arbeit wird ein Tonhöhenberechnungsmodell vorgestellt und implementiert. Das Modell ist in der Lage die virtuelle und spektrale Tonhöhe sowie deren Ausprägtheit von Umweltschallen vorherzusagen.

Durch eine dem Gehör angepaßte Signalverarbeitung wird eine Datenreduktion erzielt. Die in dieser Arbeit vorgestellten Algorithmen berücksichtigen nur die Größen, die zur Ermittlung von Tonhöhen und deren Ausprägtheit notwendig sind, alle überflüssigen Daten werden vernachlässigt.

Im ersten Schritt des Modells wird eine Fourier- Zeit- Transformation (FTT) nach Terhardt [46], [47] mit einer Fensterfunktion zweiter Ordnung nach [33] durchgeführt. Aus dem FFT- Spektrum wird das FTT- Leistungsspektrum berechnet. Dieses wird mittels eines Tiefpasses erster Ordnung geglättet, um von der Fensterfunktion hervorgerufene Einschwingeffekte zu minimieren. Die Auswertung des geglätteten Leistungsspektrums findet im weiteren nur zu frequenzabhängigen, diskreten Auswertzeitpunkten statt [15]. Für hohe Analysefrequenzen ist das frequenzabhängige Auswertintervall, d.i. der Zeitraum zwischen zwei Auswertzeitpunkten, kürzer als für tiefe Frequenzen. Durch diese Art der Auswertung erreicht die in dieser Arbeit implementierte Transformation eine Frequenz- und Zeitauflösung, die dem Gehör angepaßt ist.

Im darauffolgenden Schritt des Tonhöhenberechnungsmodells wird das Teiltonzeitmuster TTZM [12], [13] gebildet. Die Bestimmung der tonalen Komponenten aus dem FTT- Leistungsspektrum erfolgt hierbei durch eine Maximumdetektion mit Schwellenkriterium nach Heinbach [13]. Diese Art der Maximumbestimmung berücksichtigt neben lokalen Maxima in den FTT- Leistungsspektren auch globale Maxima. Für die nachfolgenden Berechnung werden so die informationstragenden Merkmale der Tonhöhe detektiert, alle nicht tonalen Anteile werden vernachlässigt. Dadurch wird eine erhebliche Datenreduktion erreicht.

Aus dem TTZM wird im dritten Schritt das Teiltonlinienzeitmuster (TTLZM) nach Heldmann [14] berechnet. Dafür wurde ein 'Vernetzungsalgorithmus' zur Mustererkennung entwickelt, der in der Lage ist, die Teiltonlinienverläufe über der Zeit zu erkennen. Der Mustererkennungsalgorithmus sucht dabei nur diejenigen Teiltonlinienverläufe auf, die eine frequenzabhängige, zeitliche Mindestlänge aufweisen. Über einen weiteren Parameter, den Maximalfrequenzabstand, kann die höchstens zulässige Frequenzänderung der Teiltonlinie

von einem Auswertzeitpunkt zum nächsten festgelegt werden. Das TTLZM gewährleistet, daß nur diejenigen Frequenzen berücksichtigt werden, die grundsätzlich vom Gehör wahrnehmbar sind. Auf diese Weise wird eine Rauschbefreiung durchgeführt und eine weitere Datenreduktion ohne Informationsverlust für die Berechnung der virtuellen und spektralen Tonhöhe erreicht. Dieses wurde durch Analyse und Resynthese qualitativ überprüft.

Im letzten Schritt werden die Tonhöhen berechnet. Dazu wird zuerst die gegenseitige Maskierung der einzelnen Teiltöne aus dem Teiltonlinienzeitmuster berücksichtigt. Für jeden Teilton wird ein Überschußpegel berechnet und so ermittelt ob, bzw. inwieweit, der Teilton zur Tonhöhe des Signals beiträgt. Die Teiltöne mit einem positiven Überschußpegel tragen zur Tonhöhe bei, alle anderen werden vernachlässigt. Aus den Pegelüberschüssen werden die Tonhöhenverschiebungen aller Teiltöne berechnet. Auf diese Weise erhält man die spektralen Tonhöhen des Signals, die sich aus den Teiltonfrequenzen mit positiven Überschußpegeln und den zugehörigen Tonhöhenverschiebungen ergeben. Die spektralen Tonhöhen werden mittels einer Gewichtsfunktion bewertet, die ihren frequenz- und überschußpegelabhängigen Beitrag zum Tonhöhereindruck berücksichtigt. Auf diese Weise wird jeder spektralen Tonhöhe ein Gewicht zugeordnet, welches ein Maß für die Ausprägtheit der spektralen Tonhöhe ist. Durch Vergleich der Tonhöhengewichte wird die ausgeprägteste spektrale Tonhöhe ermittelt. Die spektralen Tonhöhen und deren Gewichte dienen im weiteren der Extraktion von virtuellen Tonhöhen. Dies geschieht mit dem ‘virtuelle Tonhöhen-Extraktor’ [44], [50]. Dieser ermittelt die virtuellen Tonhöhen durch Vergleich der Subharmonischen aller spektralen Tonhöhen. Der VTE ordnet jeder virtuellen Tonhöhe ein Gewicht zu, welches aus dem Gewicht der beteiligten spektralen Tonhöhen, der Genauigkeit und der Anzahl der übereinstimmenden Subharmonischen der spektralen Tonhöhen und der Ordnung der Subharmonischen gebildet wird.

Durch Vergleich der Tonhöhengewichte der spektralen und virtuellen Tonhöhen können Vorhersagen über die Art und Ausprägtheit der Tonhöhen von Signalen gemacht werden.

Das vorgestellte Modell liefert für die untersuchten, nichtstationären Umweltgeräusche gute Vorhersagen der Tonhöhe. Dies wurde qualitativ überprüft. Der Tonhöhenberechnungsalgorithmus ermöglicht ebenfalls eine Unterscheidung zwischen virtuellen und spektralen Tonhöhen, sowie einen quantitativen Vergleich dieser beiden Tonhöhereindrücke über die entsprechenden Gewichte.

Neben der Modellierung des Tonhöhenberechnungsalgorithmus und der Implementation desgleichen wurden im Rahmen dieser Arbeit Scripte erstellt, die es ermöglichen mit dem Visualisierungspaket PVWAVE die Leistungsspektren der Fourier-Zeit-Transformation, Teiltonzeitmuster und Teiltonlinienzeitmuster sowie Tonhöhendigramme darzustellen.

Ausblick

Die Untersuchung der Tonhaltigkeit, d.i. die Ausgeprägtheit von Tonhöhen, führte in Kapitel 7 zu Ergebnissen die darauf hinweisen, daß das vorliegende Modell erfolgreich zur Berechnung der Tonhaltigkeit einsetzbar ist. Es sollte daher in weiteren Studien überprüft werden, inwieweit die mittels dieses Modells berechneten Ausgeprägtheiten der Tonhöhen mit den subjektiven Tonhaltigkeitsurteilen von Versuchspersonen korrelieren. Dazu können die Daten von [7] herangezogen werden.

Weiterhin sollte überprüft werden, ob die Ergebnisse der bekannten Tonhaltigkeitsberechnungsverfahren besser als bisher mit den Versuchspersonendaten von [7] übereinstimmen, wenn die hier erläuterten Prinzipien wie die Fourier-Zeit-Transformation und das Teiltonzeitmuster, bzw. das Teiltonlinienzeitmuster in den Tonhaltigkeitsalgorithmen implementiert werden.

ANHANG A FOURIER-TRANSFORMATION

Die folgenden Ausführungen sind angelehnt an [1], [4] und [37], die mathematische Theorie wird im Taschenbuch der Mathematik [5] behandelt.

A.1 Fourier-Reihe

Jede periodische Funktion $f(t)$ mit der Periode T_0 kann als sogenannte *Fourier-Reihe* ausgedrückt werden:

$$(A.1) \quad f(t) = \frac{a_0}{2} + \sum_{n=1}^{\infty} (a_n \cos(n\omega_0 t) + b_n \sin(n\omega_0 t))$$

mit den beiden *Fourier-Koeffizienten* a_n und b_n

$$(A.2) \quad a_n = \frac{2}{T_0} \int_{t_0}^{t_0+T_0} f(t) \cos(n\omega_0 t) dt \quad n=0,1,2,\dots$$

$$(A.3) \quad b_n = \frac{2}{T_0} \int_{t_0}^{t_0+T_0} f(t) \sin(n\omega_0 t) dt \quad n=0,1,2,\dots$$

wobei die Kreisfrequenz ω_0 gegeben ist durch

$$(A.4) \quad \omega_0 = \frac{2\pi}{T_0} = 2\pi f_0$$

Gl. (A.1) lautet in mathematisch äquivalenter, komplexer Schreibweise¹³:

$$(A.5) \quad f(t) = \sum_{n=-\infty}^{\infty} \bar{A}_n e^{in\omega_0 t}$$

mit komplexen Fourierkoeffizienten \bar{A}_n

$$(A.6) \quad \bar{A}_n = \frac{1}{T_0} \int_{t_0}^{t_0+T_0} f(t) e^{-in\omega_0 t} dt$$

A.2 Fourier-Transformation

Nichtperiodische Funktionen $f(t)$ können in eine Fourier-Reihe entwickelt werden, wenn die Periodendauer T_0 von $f(t)$ als unendlich lang angesehen wird. Für

¹³ Komplexe Größen werden durch einen Strich über ihrem Formelzeichen dargestellt.

absolut integrierbare Funktionen $f(t)$ ergibt sich die *Fourier-Transformierte* analog zu Gl. (A.6). Ist $f(t)$ quadratintegabel, so ist die Transformation umkehrbar eindeutig:

$$(A.7) \quad \bar{F}(\omega) = \int_{-\infty}^{\infty} f(t)e^{-i\omega t} dt$$

und die *inverse Fourier-Transformierte* wird dargestellt durch

$$(A.8) \quad f(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \bar{F}(\omega)e^{i\omega t} dt$$

Das *Fourier-Spektrum* $\bar{F}(\omega)$ für nichtperiodische Funktionen ist kontinuierlich, während das Fourier-Spektrum für periodische Funktionen, welches durch a_n und b_n repräsentiert wird, Gl. (A.2) und (A.3), diskret ist.

A.3 Diskrete Fourier-Transformation (DFT)

Die Zeitfunktion $f(t)$ wird durch äquidistante *Abtastung* mit dem *Abtast-* oder *Sampleintervall* T_s zu einer zeit- und wertdiskreten Folge von Abtastwerten. Diese (reellen) Abtastwerte der Funktion $f(t)$ liegen an den Stellen $f(nT_s)$ vor, im folgenden als f_n bezeichnet.

Ein Satz dieser Abtastwerte $\{f_n\}$ läßt sich daher nach Gl. (A.7) umkehrbar eindeutig auf den entsprechenden Satz der Fourierkoeffizienten $\{\bar{F}_m\}$ abbilden:

$$(A.9) \quad \bar{F}_m = \sum_{n=0}^{N-1} f_n e^{-i(2\pi mn/N)} \quad m=0,1,2,\dots,N-1$$

\bar{F}_m steht als Abkürzung für $\bar{F}(m\omega) = \bar{F}(2\pi m / NT_s)$. N bedeutet die Anzahl der Abtastwerte. Gl. (A.9) wird als *diskrete Fourier-Transformation (DFT)* bezeichnet. Die *inverse diskrete Fourier-Transformation (IDFT)* ist gegeben durch:

$$(A.10) \quad f_n = \frac{1}{N} \sum_{m=0}^{N-1} \bar{F}_m e^{i(2\pi mn/N)} \quad n=0,1,2,\dots,N-1.$$

Von den $2N$ Werten des komplexen Fourierspektrums \bar{F}_m sind N Werte unabhängig voneinander, die andere Hälfte ist redundant. Das kann mit Hilfe der beiden *Symmetrieeigenschaften*,

$$(A.11) \quad \bar{F}_m = \bar{F}_{-m}^*$$

$$(A.12) \quad \bar{F}_m = \bar{F}_{m-N}$$

die unmittelbar aus (A.9) folgen, bewiesen werden: Wegen $\overline{F}_{-m}^* = \overline{F}_{m-N} \Leftrightarrow \overline{F}_{N-m}^* = \overline{F}_m$ ist die Behauptung gezeigt. Durch diesen Sachverhalt kann die Berechnung von \overline{F}_m auf $N/2$ Werte eingeschränkt werden (f_n reell!).

I Zusammenhang zwischen diskreter und kontinuierlicher Fourier-Transformation

Die DFT ist einerseits eine eigenständige lineare Transformation wie in Gl. (A.9) definiert, andererseits kann sie als Näherung der Fourier-Transformation Gl. (A.7) dienen:

a.) Zeitbegrenzte Signale:

Die Zeitfunktion $f(t)$ soll zeitbegrenzt mit der Länge T_0 sein. Dadurch reduziert sich die Integration in Gl. (A.7) auf den Zeitbereich von 0 bis T_0 :

$$(A.13) \quad \overline{F}(\omega) = \int_{-\infty}^{\infty} f(t)e^{-i\omega t} dt \Rightarrow \int_0^{T_0} f(t)e^{-i\omega t} dt$$

Wird $f(t)$ in nullter Ordnung durch die Abtastwerte $[f(t)]$ mit $T_0 = NT_S$ ersetzt, vereinfacht sich die Integration zu einer Summation. Nach Division durch T_0 ergibt sich:

$$\Rightarrow \sum_{n=0}^{N-1} f_n e^{-i\omega n T_S} \Rightarrow \sum_{n=0}^{N-1} f_n e^{-i(2\pi mn/N)} = \overline{F}_m$$

Die Zeitfunktion $f(t)$, die als periodisch mit der Periode T_0 betrachtet werden kann, wird so in ein Linienspektrum analog zu Gl. (A.9) zerlegt.

b.) Beliebige Signale:

Im Fall eines unbegrenzten Zeitsignals $f(t)$ kann die Deltafunktion zur Abtastung herangezogen werden:

$$(A.14) \quad f_{\delta}(t) = \sum_{n=0}^{N-1} f_n \delta(t - nT_S).$$

Hierdurch wird das Signal $f(t)$ zeitlich begrenzt, oder *gefenstert*¹⁴:

$$(A.15) \quad \overline{F}(\omega) = \int_{-\infty}^{\infty} f_{\delta}(t)e^{-i\omega t} dt \Rightarrow \sum_{n=0}^{N-1} f_n \int_0^{T_0} \delta(t - nT_S)e^{-i\omega t} dt$$

$$\Rightarrow \sum_{n=0}^{N-1} f_n e^{-i\omega n T_S} \Rightarrow \sum_{n=0}^{N-1} f_n e^{-i(2\pi mn/N)} = \overline{F}_m$$

Hier wird, wie bei a.), ein Linienspektrum berechnet.

II Zusammenhang zwischen DFT und Fourier-Reihe

Für einen (periodisch fortsetzbaren) Abtastsatz $[f_n]$ der Periode N einer Fourier-Reihe von $f(t)$, dargestellt durch Gl. (A.5), werde eine DFT nach Gl. (A.9) durchgeführt:

$$(A.16) \quad f_n = f(nT_s) = \sum_{k=-\infty}^{\infty} \bar{A}_k e^{ik\omega_0 nT_s}$$

$$(A.17) \quad \bar{F}_m = \sum_{n=0}^{N-1} f_n e^{-i(2\pi mn/N)} \quad m=0,1,2,\dots,N-1.$$

(A.5) in (A.9) einsetzen und vertauschen der Summationen ergibt:

$$\bar{F}_m = \sum_{k=-\infty}^{\infty} \bar{A}_k \sum_{n=0}^{N-1} e^{-i2\pi n(m-k)/N}.$$

Die innere Summe liefert N falls $k-m = l \cdot N$ ist; l repräsentiert hierbei eine ganze Zahl. Für $k-m \neq l \cdot N$ wird die innere Summe (geom. Reihe) umgeformt mit [5, Seite 114]:

$$\sum_{n=0}^{N-1} e^{-i2\pi n(m-k)/N} = \sum_{n=0}^{N-1} \left(e^{-i2\pi(m-k)/N} \right)^n = \frac{e^{-i2\pi(m-k)} - 1}{e^{-i2\pi(m-k)/N} - 1}.$$

Es zeigt sich, daß für diesen Fall der Nenner ungleich, der Zähler gleich Null ist. Der gesamte Ausdruck liefert also nur einen Beitrag, wenn $k = l \cdot N + m$ ist. Das bedeutet für \bar{F}_m :

$$\bar{F}_m = N \sum_{l=-\infty}^{\infty} \bar{A}_{m+l \cdot N}$$

Wird das Abtasttheorem $m < N/2$ (siehe z. B. [4, Seite 104ff]) angewendet, vereinfacht sich obige Darstellung zu

$$(A.18) \quad \bar{F}_m = N \bar{A}_m.$$

Aus Gl. (A.18) wird ersichtlich, daß die DFT auch als diskrete Fourier-Reihe angesehen werden kann. In diesem Fall wird die inverse DFT gegeben durch:

$$(A.19) \quad f_n = \frac{1}{N} \sum_{|m| \leq N/2} \bar{F}'_m e^{i2\pi mn/N} \quad [37, \text{Gl. (5.28)}]:$$

mit

$$(A.20) \quad \bar{F}'_m = \begin{cases} \frac{\bar{F}_{N/2}}{2} & \text{für } |m| = N/2 \\ \bar{F}_m & \text{für } |m| < N/2 \end{cases}.$$

für gerade N , für ungerade N wird $N/2 \neq m$.

¹⁴ Den Zweck und den Einfluß von Fensterfunktionen in Zusammenhang mit der Fourier-Transformation beschreiben z.B [4], [21], [31] und [35].

III Schnelle Fourier-Transformation (FTT)

Die *schnelle Fourier-Transformation* ist eine Umsetzung der DFT in einen geschwindigkeitsoptimierten Rechner-Algorithmus. Die Grundlagen dafür wurden in den sechziger Jahren von Cooley und Tuckey erarbeitet. Die verwendeten mathematischen Methoden finden sich z.B. in [1], [4], [31] oder [37].

Die FFT senkt die Bearbeitungszeit gegenüber einer DFT proportional zum Faktor

$$(A.21) \quad \frac{\log_2(N)}{N} \quad [1, \text{Seite 112}].$$

Dabei muß die Anzahl der Samples N nach Gl. (A.9) notwendigerweise durch eine Zweierpotenz gegeben sein: $N = 2^k, k = 1, 2, 3, \dots$

IV Eigenschaften & Rechenregeln der Fourier-Transformation

Im Anhang B sind einige Eigenschaften und Rechenregeln der Fourier-Transformation angegeben. Mit den zuvor diskutierten Zusammenhängen können diese auch auf die DFT übertragen werden.

A.4 Zeit- und Frequenzauflösung der Fourier-Transformation

Für die DFT Gl. (A.9) ist die Zeitauflösung ΔT gleich der Länge des (periodisch fortsetzbaren) transformierten Signals: $\Delta T = N \cdot T_s$. Die Frequenzauflösung Δf ergibt sich zu $\Delta f = (\Delta T)^{-1}$. Eine Vergrößerung der Frequenzauflösung bewirkt demnach eine Verschlechterung der Zeitauflösung und umgekehrt. Das Produkt aus Frequenz- und Zeitauflösung beträgt stets eins.

Bei signaltheoretischer Betrachtung zeigt sich, daß eine Verallgemeinerung auf die Fouriertransformation nicht direkt möglich ist, da ein Signal und sein zugehöriges Spektrum nicht gleichzeitig von endlicher Dauer und Breite sein können (Unschärferelation der Signalverarbeitung).

Um eine endliche Signaldauer zu erzielen, werden die zu untersuchenden Signale im Zeitbereich gefenstert. Die Signallänge wird hierdurch auf die Fensterlänge begrenzt. Bei zeitlich unendlichen Fenstern wird die *effektive Fensterlänge* T_{eff} definiert. Die Definition ist in der Literatur nicht einheitlich. In dieser Arbeit wird die effektive Fensterlänge folgendermaßen vereinbart:

- Die effektive Fensterlänge T_{eff} eines Signals ist der Zeitbereich, in dem die Fensterfunktion $h(t)$ um nicht mehr als ein e -tel, bezogen auf ihr absolutes Maximum, abgefallen ist:

$$(A.22) \quad T_{eff} := |h(t_1) - h(t_2)| \quad \text{mit} \quad h(t_1) = h(t_2) = \frac{\max(h(t))}{e}, \quad t_1 \neq t_2$$

Durch Fensterung im Zeitbereich ergibt sich ein unendliches, kontinuierliches, fensterabhängiges Spektrum. Um Aussagen über das Frequenzauflösungsvermögen dieser Transformation zu machen, wird die *effektive Bandbreite* B_{3dB} definiert. Diese Definition richtet sich nach der üblichen Konvention der Signalverarbeitung:

- Die effektive Bandbreite B_{3dB} ist derjenige Spektralbereich des Signals, in dem das Spektrum um nicht mehr als 3 dB, bezogen auf sein absolutes Maximum, abgesunken ist:

$$(A.23) \quad B_{3dB} := |\bar{F}(\omega_1) - \bar{F}(\omega_2)| \quad \text{mit}$$

$$\bar{F}(\omega_1) = \bar{F}(\omega_2) = \frac{\max(\bar{F}(\omega))}{\sqrt{2}}, \quad \omega_1 \neq \omega_2$$

Das $B_{3dB}T_{eff}^{15}$ Produkt ist die analoge Größe der Fouriertransformation für prinzipiell unendlich lange Signale zum Produkt aus Frequenz- und Zeitauflösung der DFT. Es eignet sich als Qualitätsmaß der benutzten Transformation in folgender Weise:

- T_{eff} sollte möglichst klein sein, um Frequenzänderungen über der Zeit korrekt wiederzugeben und
- B_{3dB} sollte möglichst gering sein, um die Frequenzbestimmung so genau wie möglich durchzuführen [47].

Je nach Wahl der Fensterfunktion erhält man für das BT Produkt der Fouriertransformation Werte zwischen 0,8 und 2 [52].

¹⁵ Im weiteren wird die effektive Bandbreite B_{3dB} nur noch durch B bezeichnet, die effektive Fensterlänge T_{eff} mit T .

ANHANG B EIGENSCHAFTEN DER FOURIER-TRANSFORMATION

Die folgende Tabelle ist zum Teil entnommen aus Stearns [37], für eine umfassende Darstellung siehe Skudrzyk [35].

Eigenschaft	Funktion <small>16</small>	Fourier-Transformierte
Fourier-Transformation	$f(t)$	$\bar{F}(\omega)$
Multiplikation mit reeller Konstanten a	$a \cdot f(t)$	$a \cdot \bar{F}(\omega)$
Linearität	$f(t) + g(t)$	$\bar{F}(\omega) + \bar{G}(\omega)$
Differentiation	$\frac{df(t)}{dt}$	$i\omega\bar{F}(\omega)$
Integration	$\int_{-\infty}^t f(\tau) d\tau$	$\frac{\bar{F}(\omega)}{i\omega}$
Zeit-Multiplikation	$t \cdot f(t)$	$i \frac{d\bar{F}(\omega)}{d\omega}$
Frequenzverschiebung (Modulation) mit fester Frequenz ω_0	$e^{-i\omega_0 t} f(t)$	$\bar{F}(\omega + \omega_0)$
Zeitverschiebung um feste Zeit t_0	$f(t-t_0)$	$e^{-i\omega t_0} \bar{F}(\omega)$
Zeit- bzw. Frequenzskalierung mit reeller Konstanten a	$f\left(\frac{t}{a}\right)$	$a \cdot \bar{F}(a\omega)$
Faltung ¹⁷ im Zeitbereich	$f(t) * g(t)$	$\bar{F}(\omega) \cdot \bar{G}(\omega)$
Faltung im Frequenzbereich	$f(t) \cdot g(t)$	$\bar{F}(\omega) * \bar{G}(\omega)$
Parsevallsches Theorem	$\int_{-\infty}^{+\infty} f^2(t) dt =$	$\frac{1}{2\pi} \int_{-\infty}^{+\infty} \bar{F}(\omega) ^2 d\omega$

Tabelle B.1: Eigenschaften der Fourier-Transformation

¹⁶ Diese Funktionen sind stets reell angenommen, da physikalische Zeitsignale i.A. reell sind.

¹⁷ Zur Einführung der Faltung siehe z.B. Lücke [21].

ANHANG C EFFEKTIVE FENSTERLÄNGE UND BANDBREITE DER FTT

Die effektive Fensterlänge T und Bandbreite B einer FTT berechnen sich wie folgt beschrieben. Die Frequenzabhängigkeit der Transformationskonstanten $a(\omega)$ wird übersichtshalber nicht explizit mitgeführt: $a := a(\omega)$.

- FTT 1. Ordnung

Die effektive Fensterlänge T von $h_1(t - t') = \eta_1 \cdot \exp\{-a_1(t - t')\}$ mit $t \geq t' \geq 0$ wird durch die Substitution $x := t - t'$ berechnet. Die Fensterfunktion $h_1(x)$ nimmt ihr Maximum von η_1 bei $x_{max} = 0$ an. Bei $x_1 = 1/a$ ist der Wert der Fensterfunktion $h_1(x_1)$ auf η_1/e abgefallen. Für $x_2 = 0$ besitzt die Fensterfunktion eine Sprungstelle, an der ihr Wert von η_1 auf 0 abfällt. Damit ergibt sich die effektive Fensterlänge als Differenz von x_1 und x_2 : $T_1 = |x_1 - x_2| = 1/a$.

Zur Ermittlung der Bandbreite wird das Betragsquadratspektrums $|H(\omega)|^2$ der Fensterfunktion $h_1(x)$ mittels Fourier-Sinus- und Cosinustransformation [5, Tab. 4.4.2.2] berechnet:

$$\left. \begin{array}{l} H_{\sin}(\omega) \sim a \frac{\omega}{\omega^2 + a^2} \\ H_{\cos}(\omega) \sim a \frac{a}{\omega^2 + a^2} \end{array} \right\} \Rightarrow |H(\omega)|^2 \sim \frac{a^2}{\omega^2 + a^2} = \frac{1}{1 + \frac{\omega^2}{a^2}}$$

Das (einseitige) Betragsquadratspektrum $|H(\omega)|^2$ zeigt Abb. 3.3. Es besitzt ein Maximum von 1. Der Abfall auf die Hälfte (wegen der Quadrierung des Betragsspektrums) ist erreicht bei $\omega_1 = -a$ und $\omega_2 = a$. Die Bandbreite B_1 errechnet sich demnach zu

$$B_1 = \left| \frac{\omega_1 - \omega_2}{2\pi} \right| = \frac{a}{\pi}$$

Diese Werte decken sich mit den Berechnungen von [46], [13], [14].

- FTT 2. Ordnung

Die Rechnungen werden analog wie für die FTT 1. Ordnung durchgeführt. Die Fensterfunktion $h_2(x) = x \cdot \exp(-a \cdot x)$ mit $x := t - t'$, $t \geq t' \geq 0$ hat ein Maximum von $(e \cdot a)^{-1}$. Bei $x_1 = 0,15859/a$ und $x_2 = 3,14615/a$ (mit Newton-

Verfahren) ist das Maximum auf seinen halben Wert abgefallen. Die effektive Fensterlänge T_2 ergibt sich zu: $T_2 = 2,98756/a$. Heldmann [14] gibt für T_2 einen Wert von $\sqrt{5}/a = 2,23607/a$ an.

Die Bandbreite wird wiederum über das Betragsquadratspektrum $|H(\omega)|^2$, nach [5, Tab. 4.4.2.2] bestimmt:

$$\left. \begin{aligned} H_{\sin}(\omega) &\sim a^2 \frac{\sin\left(2 \arctan\left(\frac{\omega}{a}\right)\right)}{\omega^2 + a^2} \\ H_{\cos}(\omega) &\sim a^2 \frac{\cos\left(2 \arctan\left(\frac{\omega}{a}\right)\right)}{\omega^2 + a^2} \end{aligned} \right\} \Rightarrow |H(\omega)|^2 \sim \frac{a^4}{(\omega^2 + a^2)^2}$$

Das (einseitige) Betragsquadratspektrum $|H(\omega)|^2$ ist in Abb. 3.3 dargestellt. Der Abfall von $|H(\omega)|^2$ auf die Hälfte seines maximalen Wertes tritt bei $\omega_1 = a \cdot \sqrt{\sqrt{2}-1}$, bzw. $\omega_2 = -a \cdot \sqrt{\sqrt{2}-1}$ auf. Die Bandbreite B_2 errechnet sich demnach zu:

$$B_2 = \left| \frac{\omega_1 - \omega_2}{2\pi} \right| = \sqrt{\sqrt{2}-1} \cdot a/\pi \approx 0,64 \cdot a/\pi$$

Dieser Wert deckt sich mit den Berechnungen von Heldmann [14].

ANHANG D PROGRAMME ZUR BERECHNUNG DER VIRTUELLEN TONHÖHE

Programm 1: wav2msc4.0

- Beschreibung: Dieses Programm ermittelt die Eigenschaften eines zu untersuchenden Signals, welches als WAVE-Datei vorliegen muß und wandelt die Datei ins ASCII-Format um. Zur Erläuterung soll die Vorgehensweise am Beispiel des Signals, *'beispiel.wav'* erklärt werden: **wav2msc4.0** ermittelt die Dateieigenschaften des Signals *'beispiel.wav'* wie z.B. den Dateinamen, die Samplingfrequenz, die Kanalzahl, die Länge des Signals und die Anzahl der Bits pro Sample. Diese werden in der Kontrolldatei *'beispiel.dtn'* abgelegt. (Der Benutzer kann anstelle von *'beispiel.dtn'* auch einen anderen Namen *'*.dtn'* wählen). Bei mehrkanaligen Aufnahmen wird der Benutzer nach dem Kanal gefragt, welcher berücksichtigt werden soll. (Die weitere Verarbeitung ist nur für einkanalige Dateien konzipiert.) Dieser Kanal wird anschließend von der internen Hexadezimal-Darstellung der WAVE-Datei in ASCII-Darstellung transformiert und in *'beispiel.msc'* auf Festplatte abgelegt.
- Programmaufruf: **wav2msc4**
- Programm fragt ab:
- 1.) den Namen *name1* der zu untersuchenden Datei *'name1.wav'*,
 - 2.) den Namen *name2* der neu anzulegenden Dateien *'name2.dtn'* und *'name2.msc'*,
 - 3.) bei mehrkanaligen WAVE-Dateien den Kanal *chan* der berücksichtigt werden soll.
- Programm liest ein: *'name1.wav'* (zu untersuchendes Signal).
- Programm berechnet:
- 1.) die Kontrolldatei *'name2.dtn'*,
 - 2.) *'name2.msc'*, d.h. die ASCII-Darstellung des Kanals *chan* der Datei *'name1.wav'*.
- Kommandozeilenversion: **wav2msc4 name1 name2 chan**

Bemerkungen: Programm geschrieben in MS-FORTRAN 5.1, daher nur lauffähig auf einem PC mit dem entsprechenden Compiler.

Programm 2a: ftt1.5d

Beschreibung: Dieses Programm berechnet eine FTT wie in Kapitel 3 beschrieben. Dazu liest es die Datei *'beispiel.msc'* ein. Zunächst kann der Benutzer die Art (durch Tiefpaß geglättete Leistungsspektren oder nicht durch Tiefpaß geglättete Leistungsspektren) und die Ordnung (erste oder zweite Ordnung der Fensterfunktion) der FTT wählen. Die Parameter Analysebandbreite b und Abstand der Analysefrequenzen Δz sind daraufhin vom Benutzer festzulegen, ebenso der Frequenzbereich der Transformation (Ober- und Untergrenze der Frequenzen) und der zeitliche Bereich der Analyse (Start- und Endzeitpunkt). Mit diesen Angaben wird die FTT durchgeführt und das berechnete FTT-Leistungsspektrum als Datei *'beispiel.spc'* abgespeichert.

Programmaufruf:

ftt15d

Programm fragt ab:

- 1.) Name *name1* der zu transformierenden Datei *'name1.msc'*,
- 2.) Analysebandbreite, Frequenzabstand der Analysefrequenzen, Länge des Auswertintervalls (entspricht dem zeitlichen Anstand, zu dem aufeinanderfolgende Einzelspektren gespeichert werden),
- 3.) Frequenzbereich der Transformation,
- 4.) Zeitlicher Bereich der Transformation.

Programm liest ein:

'name1.dtn', *'name1.msc'*,

Programm berechnet:

- 1.) *'name1.spc'* FTT-Leistungsspektrum
- 2.) *'name1.brk'* Diskrete Frequenzen [Bark] innerhalb des Frequenzbereiches der Transformation.
- 3.) *'name1.frq'* Diskrete Frequenzen [Hz] innerhalb des Frequenzbereiches der Transformation.

Feste Parameter: Transformationskonstante nach Gl. (3.4)
Frequenzgruppenbreite nach Gl. (3.5)
Glättung der Leistungsspektren nach Gl. (3.18)
Glättungszeitkonstante nach Gl. (3.12)

Kommandozeilenversion: **ftt15d** *parameterdatei*

Bemerkungen: Die *parameterdatei* für die Kommandozeilenversion muß eine ASCII-Datei sein, die wie folgt aufgebaut ist:

1. Zeile: zu transformierende Datei: *name1* (ohne Endung !)
2. Zeile: FTT-Spektrum: *name2* (ohne Endung !)
3. Zeile: *Art* = 1,2,3,4 (1= Fensterfunktion 1.Ordnung, keine Tiefpaßglättung; 2= Fensterfunktion 1.Ordnung, Tiefpaßglättung; 3= Fensterfunktion 2.Ordnung, keine Tiefpaßglättung; 4= Fensterfunktion 2.Ordnung, Tiefpaßglättung).
4. Zeile: Startzeitpunkt der Analyse relativ zum Signalanfang in Sekunden.
5. Zeile: Endzeitpunkt der Analyse relativ zum Signalanfang in Sekunden.
6. Zeile: Länge des Auswertintervalls in Samples.
7. Zeile: Frequenzabstand der Analysefrequenzen in Bark.
8. Zeile: kleinste zu transformierende Freq. [Hz]
9. Zeile: größte zu transformierende Frequenz [Hz]
10. Zeile: Analysebandbreite [Bark]

Programm 2b: ftt3.1c

Beschreibung: Dieses berechnet eine FTT wie in Kapitel 3 beschrieben; die Fensterfunktion ist von zweiter Ordnung, es wird eine Tiefpaßglättung wie in Kapitel 3 durchgeführt. Zusätzlich erfolgt die Auswertung nur innerhalb der in Kapitel 4.1.4 beschriebenen Auswertintervalle, wodurch die Zeitauflösung hörgerecht ausgeführt wird.

Der Vorteil, die frequenzabhängigen Auswertintervalle aus dem Programm zur Berechnung des TTZM in das Programm zur Berechnung des FTT-Leistungsspektrums zu ziehen liegt darin, daß die Berechnung des Teiltonzeitmusters in diesem Falle für die Routinen **ftt1.5d** und **ftt3.1c** gleich ausgeführt werden kann.

Zuerst liest das Programm **ftt3.1c** eine Datei '*beispiel.msc*' ein. Die Parameter Analysebandbreite b und Frequenzabstand der Analysefrequenzen Δz sind daraufhin vom Benutzer festzulegen, ebenso der Frequenzbereich der Transformation (Ober- und Untergrenze der Frequenzen) und der zeitliche Bereich der Analyse (Start- und Endzeitpunkt). Mit diesen Angaben wird die FTT durchgeführt und das berechnete FTT-Leistungsspektrum als Datei '*beispiel.spc*' abgespeichert.

Programmaufruf:	ftt31c
Programm fragt ab:	<ol style="list-style-type: none"> 1.) Name <i>name1</i> der zu transformierenden Datei '<i>name1.msc</i>' 2.) Analysebandbreite, Abstand der Analysefrequenzen, Länge des frequenzabhängigen Auswertintervalls für 24 Bark. 3.) Frequenzbereich der Transformation. 4.) Zeitlicher Bereich der Transformation.
Programm liest ein:	' <i>name1.dtn</i> ', ' <i>name1.msc</i> ',
Programm berechnet:	<ol style="list-style-type: none"> 1.) '<i>name1.spc</i>' FTT-Leistungsspektrum 2.) '<i>name1.brk</i>' Diskrete Frequenzen [Bark] innerhalb des Frequenzbereiches der Transformation. 3.) '<i>name1.frq</i>' Diskrete Frequenzen [Hz] innerhalb des Frequenzbereiches der Transformation.
Feste Parameter:	Transformationskonstante nach Gl. (3.4) Frequenzgruppenbreite nach Gl. (3.5) Glättung der Leistungsspektren nach Gl. (3.18) Glättungszeitkonstante nach Gl. (3.12)
Kommandozeilenversion:	ftt31c <i>parameterdatei</i>
Bemerkungen:	Die Parameterdatei für die Kommandozeilenversion muß eine ASCII-Datei sein, die wie folgt aufgebaut ist:

1. Zeile: zu transformierende Datei: *name1* (ohne Endung !)
2. Zeile: FTT-Spektrum: *name2* (ohne Endung !)
3. Zeile: wird nicht ausgewertet, d.h. Leerzeile
4. Zeile: Startzeitpunkt der Analyse relativ zum Signalanfang in Sekunden
5. Zeile: Endzeitpunkt der Analyse relativ zum Signalanfang in Sekunden
6. Zeile: Länge des Auswertintervalls in Samples für 24 Bark
7. Zeile: Frequenzabstand der Analysefrequenzen in [Bark]
8. Zeile: kleinste zu transformierende Freq. [Hz]
9. Zeile: größte zu transformierende Frequenz [Hz]
10. Zeile: Analysebandbreite in [Bark]

Programm 3: maxi3.0c

Beschreibung:

Programm zur Berechnung des Teiltonzeitmusters mit Schwellenkriterium nach Heinbach, wie in Kap. 4.1.5 erläutert.

Zuerst fragt das Programm den Wert für das Schwellenkriterium ab. Anschließend wird der maximale Pegel, der im FTT-Leistungsspektrum auftritt, ermittelt. Dieser Pegel kann auf einen beliebigen, vom Benutzer einzugebenden Pegelwert umskaliert werden. Dies ist z.B. nützlich, wenn Schalle von Geräusch-CDs untersucht werden sollen. Auf diese Weise können die Pegel der Geräusche von CDs den jeweiligen Bedürfnissen angepasst werden.

Um das eigentliche TTZM zu erhalten, wird das von **ftt1.5d** oder **ftt3.1c** berechnete FTT-Leistungsspektrum '*beispiel.spc*' eingelesen und daraus die Maxima wie in Kapitel 4.1 ermittelt.

Programmaufruf:

maxi30c

- Programm fragt ab: 1.) Name '*name*' des FTT-Leistungsspektrums '*name.spc*', für welches das TTZM berechnet werden soll.
2.) Schwellenkriterium ΔL für die Maximumbestimmung
3.) Pegel, auf den das Maximum des FTT-Leistungsspektrum umskaliert wird.
- Programm liest ein: '*name1.dtn*', '*name1.spc*'
- Programm berechnet: Teiltonlinienzeitmuster '*name1.ttz*'

Programm 4: netz4.0a

- Beschreibung: Dieses Programm vernetzt ein TTZM '*name.ttm*' zum TTLZM '*name.ttz*' wie in Kapitel 4.2 beschrieben. Der Benutzer kann hierbei das Maximalabstandskriterium $\Delta i \cdot \Delta z$ in Vielfachen Δi des Frequenzabstandes der Analysefrequenzen Δz angeben. Außerdem kann er die Mindestlänge l_m der Teiltonlinien im TTLZM für 100 Hz wählen.
- Programmaufruf: **netz40a**
- Programm fragt ab: 1.) Name '*name1*' des Teiltonzeitmusters '*name1.ttm*', für welches das TTZM vernetzt werden soll.
2.) Maximalabstand Δi in Vielfachen des Frequenzabstandes Δz .
3.) Mindestlänge der Teiltonlinien l_m .
- Programm liest ein: '*name1.dtn*', '*name1.ttm*'
- Programm berechnet: vernetztes Teiltonzeitmuster TTLZM: '*name1.ttz*'

Programm 5: vp2.3a

- Beschreibung: Dieses Programm berechnet die spektralen und virtuellen Tonhöhen sowie ihre Gewichte nach Kapitel 5. Dazu liest es ein TTLZM '*name1.ttz*' ein. Der Benutzer kann das Übereinstimmungsintervall δ , den Maximalwert der Subharmonischen M , den Schwellenwert WS_{imax} und die Anzahl der größten zu berücksichtigenden virtuellen Tonhöhen vorgeben.

- Programmaufruf: **vp23a**
- Programm fragt ab:
- 1.) Name '*name1*' des TTLZM '*name1.ttz*', für welches die virtuellen und spektralen Tonhöhen berechnet werden sollen.
 - 2.) Übereinstimmungsintervall δ nach Kap. 5.2.5
 - 3.) Maximalwert der Subharmonischen M wie definiert in Kap. 5.2.5.
 - 4.) Schwellenwert WS_{imax}
 - 5.) Anzahl der größten zu berücksichtigenden virtuellen Tonhöhen.
- Programm liest ein: '*name1.dtn*', '*name1.brk*', '*name1.frq*', '*name1.ttz*'
- Programm berechnet:
- 1.) Das Spektraltonhöhendigramm '*name1.spt*'
 - 2.) Das virtuelle Tonhöhendigramm '*name2.vtt*'

ANHANG E HILFSPROGRAMME

Hilfsprogramm 1: spc2pvwav9.0

Beschreibung:	Dieses Hilfsprogramm dient dazu, die interne Darstellung der Datei ' <i>beispiel.spc</i> ' so umzusetzen, daß sie mit dem Programm PVWAVE (oder anderen vergleichbaren Visualisierungspaketen) graphisch darstellbar ist. Dafür wird die Datei ' <i>beispiel.spc</i> ' in die Datei ' <i>beispiel.fgn</i> ' transformiert, in der jeder diskrete Datenpunkt des Leistungsspektrums explizit als Zahlentriplett aus Zeit, Frequenz und Pegel repräsentiert wird.
Programmaufruf:	spc2pvwav9
Programm fragt ab:	1.) Name <i>name1</i> der zu wandelnden Datei ' <i>name1.spc</i> ' 2.) Skalierung der Frequenzachse: in Bark oder Herz
Programm liest ein:	' <i>name1.dtn</i> ', ' <i>name1.spc</i> ', ' <i>name1.frq</i> ', ' <i>name1.brk</i> '
Programm berechnet:	' <i>name1.fgn</i> '
Bemerkungen:	Durch die Transformation von ' <i>beispiel.spc</i> ' zu ' <i>beispiel.fgn</i> ' vergrößert sich die Dateigröße erheblich!

Hilfsprogramm 2: ttm2pvwav3.0

Beschreibung:	Dieses Hilfsprogramm dient dazu, die interne Darstellung der Datei ' <i>beispiel.ttm</i> ' so umzusetzen, daß sie mit dem Programm PVWAVE (oder anderen vergleichbaren Visualisierungspaketen) graphisch darstellbar ist. Dafür wird die Datei ' <i>beispiel.ttm</i> ' in die Datei ' <i>beispiel.ptm</i> ' transformiert, in der jeder diskrete Datenpunkt des Teiltonzeitmusters explizit als Zahlentriplett aus Zeit, Frequenz und Pegel repräsentiert wird.
Programmaufruf:	ttm2pvwav3
Programm fragt ab:	1.) Name <i>name1</i> der zu wandelnden Datei ' <i>name1.ttm</i> ' 2.) Skalierung der Frequenzachse: in Bark oder Herz
Programm liest ein:	' <i>name1.dtn</i> ', ' <i>name1.ttm</i> ', ' <i>name1.frq</i> ', ' <i>name1.brk</i> '
Programm berechnet:	' <i>name1.ptm</i> '

Hilfsprogramm 3: ttz2pvwav3.0

Beschreibung:	Dieses Hilfsprogramm dient dazu, die interne Darstellung der Datei 'beispiel.ttz' so umzusetzen, daß sie mit dem Programm PVWAVE (oder anderen vergleichbaren Visualisierungspaketen) graphisch darstellbar ist. Dafür wird die Datei 'beispiel.ttz' in die Datei 'beispiel.ptz' transformiert, in der jeder diskrete Datenpunkt des vernetzten Teiltonzeitmusters explizit als Zahlentriplett aus Zeit, Frequenz und Pegel repräsentiert wird.
Programmaufruf:	ttz2pvwav3
Programm fragt ab:	1.) Name <i>name1</i> der zu wandelnden Datei 'name1.ttz' 2.) Skalierung der Frequenzachse: in Bark oder Herz
Programm liest ein:	'name1.dtn', 'name1.ttz', 'name1.frq', 'name1.brk'
Programm berechnet:	'name1.ptz'

ANHANG F LITERATURVERZEICHNIS

- [1] Achilles, D.: „*Die Fourier-Transformation in der Signalverarbeitung*“ Springer-Verlag Berlin Heidelberg New York, 1. Auflage 1978.
- [2] Attia, F.S. & Kirsch, C.: „*Vergleich der gehörorientierten FFT (Fourier-Time-Transformation und der FFT (Fast-Fourier-Transformation) in der Fehlerdiagnose*“ Fortschritte der Akustik, DAGA '92, Bad Honnef: DPG-GmbH 1992, S. 901-904.
- [3] Aures, W.: „*Berechnungsverfahren für den Wohlklang beliebiger Schallsignale, ein Beitrag zur gehörbezogenen Schallanalyse*“ Dissertation an der TU München 1989.
- [4] Brigham, E.O.: „*FFT - Schnelle Fourier-Transformation*“ R. Oldenbourg Verlag München Wien, 3. verbesserte Auflage 1987.
- [5] Bronstein, I.N. & Semendjajew, K.A.: „*Taschenbuch der Mathematik*“ Verlag Harri Deutsch, Thun und Frankfurt/Main, 24. Auflage 1989, herausgegeben von Grosche, G., Ziegler, V. und Ziegler, D.
- [6] Daniel, P.: „*Berechnungsmodell für die psychoakustische Rauhigkeit*“ Diplomarbeit an der Universität Oldenburg 1992.
- [7] Dreesen, T.: „*Vergleich von Tonhaltigkeitsberechnungsverfahren auf der Grundlage subjektiver Urteile*“ Diplomarbeit an der Universität Oldenburg 1995
- [8] Eargle, J.M.: „*Music, Sound and Technology*“ Van Nostrand Reinhold New York, 1st edition 1990.
- [9] Fastl, H.: „*Pitch strength of pure tones*“ 2nd Proc. 13th ICA Belgrade 1989, Vol. 3, S. 11-14.
- [10] Fastl, H. & Wiesmann, N.: „*Ausgeprägtheit der virtuellen Tonhöhe von AM- und QFM-Tönen*“ Fortschritte der Akustik, DAGA '90, Bad Honnef: DPG-GmbH 1990, S. 759-762.
- [11] Grubert, A.: „*Tonhöhen harmonischer und inharmonischer komplexer Töne*“ Fortschritte der Akustik, DAGA '87, Bad Honnef: DPG-GmbH 1987, S. 573-576.
- [12] Heinbach, W.: „*Aurally adequate signal representation: The Part-Tone-Time-Pattern*“ Acustica Vol. 67, 1988, S. 113-121.
- [13] Heinbach, W.: „*Gehörgerechte Repräsentation von Audiosignalen durch das Teiltonzeitmuster*“ Dissertation TU München, 1988.

-
- [14] Heldmann, K.: „Wahrnehmung, gehörgerechte Analyse und Merkmalsextraktion technischer Schalle“, VDI-Verlag 1994, Fortschritt-Berichte, Reihe 17, Nr. 109.
- [15] Heldmann, K. & Keiper, W.: „Analyse von instationären technischen Geräuschen“ Fortschritte der Akustik, DAGA '91, Bad Honnef: DPG-GmbH 1991, S. 761-764.
- [16] Heldmann, K. & Keiper W.: „Merkmalsextraktion an technischen Geräuschen mittels Teiltonzeitmuster“ Fortschritte der Akustik, DAGA '92, Bad Honnef: DPG-GmbH 1992, S. 897-900.
- [17] Heldmann, K. & Keiper, W.: „Impulserkennung im Teiltonzeitmuster industrieller Schalle“ Fortschritte der Akustik, DAGA '93, Bad Honnef: DPG-GmbH 1993, S. 804-807.
- [18] Hesse, A.: „Zur Ausgeprägtheit der Tonhöhe gedrosselter Sinustöne“ Fortschritte der Akustik, DAGA '85, Bad Honnef: DPG-GmbH 1985, S. 535-538.
- [19] Hesse, A.: „Beschreibung der Pegelabhängigkeit der Spektraltonhöhe von Sinustönen anhand von Mithörschwellen“ Fortschritte der Akustik, DAGA '86, Bad Honnef: DPG-GmbH 1986, S. 445-448.
- [20] Hesse, A.: „Ein Funktionsschema der Spektraltonhöhe von Sinustönen“ Acustica Vol. 63 1987, S. 1-16.
- [21] Lüke, H.D.: „Signalübertragung“ Springer-Verlag Berlin Heidelberg New York, 2. überarbeitete und erweiterte Auflage 1983
- [22] Meddis, R. & Hewitt, M. J.: „Virtual pitch and phase sensitivity of a computer model of the auditory periphery. I: Pitch identification“ J. Acoust. Soc. Am. Vol. 89(6) 1991, S. 2866-2882.
- [23] Meddis R. & Hewitt, M. J.: „Virtual pitch and phase sensitivity of a computer model of the auditory periphery. II: Phase sensitivity“ J. Acoust. Soc. Am. Vol. 89(6) 1991, S. 2883-2894.
- [24] Meyer, E. & Guicking, D.: „Schwingungslehre“ Vieweg-Verlag Braunschweig, 1. Auflage 1974.
- [25] Meyer, E. & Neumann, E.-G.: „Physikalische und Technische Akustik“ Vieweg-Verlag Braunschweig, 3. neubearbeitete und erweiterte Auflage 1979.
- [26] Mummert, M.: „Trennung von tonalen und geräuschhaften Anteilen im Sprachsignal“ Fortschritte der Akustik, DAGA '90, Bad Honnef: DPG-GmbH 1990, S. 1047-1050.

-
- [27] Mummert, M.: „Rücktransformation des Kurzzeitspektrums der Fourier-t-Transformation und Ansatz für eine gehörgerechte Transformationskodierung“ Fortschritte der Akustik, DAGA '91, Bad Honnef: DPG-GmbH 1991, S. 753-756.
- [28] Patterson, R.D.: „The effects of relative phase and the number of components on residue pitch“ J. Acoust. Soc. Am. Vol. 53(6) 1973, S. 1565-1572.
- [29] Plomp, R., Steeneken H.J.M.: „Effect of phase on the timbre of complex tones“ J. Acoust. Soc. Am. Vol. 69(2) 1969, S. 409-421.
- [30] Press, W.H., Flannery, B.P., Teukolsky, S.A., Vetterling, W.T.: „Numerical Recipes - Fortran Version“ Cambridge University Press, 1. edition 1989.
- [31] Randall, R.B., Tech, B.B.: „Frequency Analysis“ Brüel & Kjøer 1987.
- [32] Reckhardt, C.: „Beurteilung von Geräuschqualitäten mit einem Analyse durch Syntheseverfahren“ Diplomarbeit an der Universität Oldenburg 1995.
- [33] Schlang, M. & Mummert, M.: „Die Bedeutung der Fensterfunktion für die Fourier-t-Transformation als gehörgerechte Spektralanalyse“ Fortschritte der Akustik, DAGA '90, Bad Honnef: DPG-GmbH 1990, S. 1043-1046.
- [34] Seewann, M. & Terhardt, E.: „Messungen der wahrgenommenen Tonhöhe von Glocken“ Fortschritte der Akustik, DAGA '80, Bad Honnef: DPG-GmbH 1980, S. 635-638.
- [35] Skudrzyk, E.: „The Foundations of Acoustics“ Springer-Verlag Wien New York, 1. Auflage 1971.
- [36] Sonntag, B.: „Zur Tonhöhenverschiebung gedrosselter Sinustöne“ Fortschritte der Akustik, DAGA '81, Bad Honnef: DPG-GmbH 1981, S. 729-732.
- [37] Stearns, S.D.: „Digitale Verarbeitung analoger Signale“ R. Oldenbourg Verlag München Wien, 4. verbesserte Auflage 1988.
- [38] Terhardt, E.: „Über die durch amplitudenmodulierte Sinustöne hervorgerufene Hörempfindung“ Acustica Vol. 20, 1968, S. 210-214.
- [39] Terhardt, E.: „Pitch shifts of harmonic, an explanation of the octave enlargement phenomenon“ Proc. 7th ICA Budapest 1971, Vol. 3, S. 621-624.
- [40] Terhardt, E.: „Die Tonhöhe Harmonischer Klänge und das Oktavintervall“ Acustica Vol. 24, 1971, S. 126-136.

-
- [41] Terhardt, E.: „Zur Tonhöhenwahrnehmung von Klängen I. Psychoakustische Grundlagen“ *Acustica* Vol. 26, 1972, S. 173-186.
- [42] Terhardt, E.: „Zur Tonhöhenwahrnehmung von Klängen II. Ein Funktionsschema“ *Acustica* Vol. 26, 1972, S. 187-199.
- [43] Terhardt, E.: „Pitch, Consonance and Harmony“ *J. Acoust. Soc. Am.* Vol. 55(5), 1974, S. 1061-1069.
- [44] Terhardt, E.: „Calculating virtual pitch“ *Hearing Research* 1, 1979, S. 155-182.
- [45] Terhardt, E.: „Absolute and relative pitch revisited on psychoacoustic grounds“ *Proc. 11th ICA Paris 1983*, Vol. 4, S. 427-430.
- [46] Terhardt, E.: „Fourier Transformation of Time Signals: Conceptual Revision“ *Acustica* Vol. 57, 1985, S. 242-256.
- [47] Terhardt, E.: „Verfahren zur gehörbezogenen Frequenzanalyse“ *Fortschritte der Akustik, DAGA '85*, Bad Honnef: DPG-GmbH 1985, S. 811-814.
- [48] Terhardt, E. & Grubert, A.: „Factors affecting pitch judgement as a function of spectral composition“ *Perception & Psychophysics*, Vol. 42(6), 1987, S. 511-514.
- [49] Terhardt, E., Stoll, G. & Seewann, M.: „Pitch of complex signals according to virtual-pitch theory: Tests, examples and predictions“ *J. Acoust. Soc. Am.* Vol. 71(3), 1982, S. 671-678.
- [50] Terhardt, E., Stoll, G. & Seewann, M.: „Algorithm for extraction of pitch salience from complex tonal signals“ *J. Acoust. Soc. Am.* Vol. 71(3), 1982, S. 679-688.
- [51] Vormann, M., Weber, R.: „Gehörgerechte Darstellung von instationären Umweltgeräuschen mittels Fourier-Time-Transformation (FTT)“ *Fortschritte der Akustik, DAGA '94*, Bad Honnef: DPG-GmbH, im Druck.
- [52] Walliser, K.: „Über die Abhängigkeit der Tonhöhenempfindung von Sinustönen vom Schallpegel, von überlagertem drosselndem Störschall und von der Darbietungsdauer“ *Acustica* Vol. 21, 1969, S. 211-221.
- [53] Walliser, K.: „Zusammenhänge zwischen dem Schallreiz und der Periodentonhöhe“ *Acustica* Vol. 21, 1969, S. 319-329.
- [54] Wighthman, F.L.: „The pattern-transformation model of pitch“ *J. Acoust. Soc. Am.* Vol. 54(2) 1973, S. 407-416.
- [55] Zwicker, E.: „Subdivision of the audible frequency range into critical bands (Frequenzgruppen)“ *J. Acoust. Soc. Am.* Vol. 33(2), 1961, S. 248.

-
- [56] Zwicker, E.: „*Psychoakustik*“ Springer-Verlag Berlin, 1. Auflage 1982.
- [57] Zwicker, E. & Feldkeller, R.: „*Das Ohr als Nachrichtenempfänger*“ Hirzel-Verlag Stuttgart, 2. neubearbeitete Auflage 1967.
- [58] Zwicker, E. & Fastl, H.: „*Psychoacoustics*“ Springer-Verlag Berlin, 1. Auflage 1990.
- [59] Zwicker, E. & Terhardt, E.: „*Analytical expressions for critical-band rate and critical bandwidth as a function of frequency*“ J. Acoust. Soc. Am. Vol. 68(5), 1980, S. 1523-1525.
- [60] CD: „*Geräusche Digital*“ Intercord 1986, Nr. 72: Hupe.
- [61] CD: „*Ultraschall - Geräusche Digital*“ Intercord 1988, Nr. 45: Heulton 'Luftangriff'.
- [62] CD: „*Geräusche Digital*“ Intercord 1986, Nr. 24: Näherkommendes Martinshorn.
- [63] CD: „*Authentic Sound Effects Volume 1*“ Elektra Records, Nr. 45: Chimes Tolling Hour.
- [64] DIN 1320: „*Akustik Grundbegriffe*“ Okt. 1969

ANHANG G DANKSAGUNG UND ERKLÄRUNG

Danksagung:

Ich möchte mich an dieser Stelle bei allen, die tatkräftig am Zustandekommen dieser Arbeit mitgewirkt haben, bedanken.

Erklärung:

Hiermit versichere ich, daß ich diese Arbeit selbständig verfaßt und keine anderen als die angegebenen Quellen verwendet habe.