Lehrstuhl für Elektroakustik
der Technischen Universität München
Arbeitsgruppe Akustische Kommunikation

Gehörgerechte Repräsentation von Audiosignalen durch das Teiltonzeitmuster

Wolfgang Heinbach

Vollständiger Abdruck der von der Fakultät für
Elektrotechnik und Informationstechnik der
Technischen Universität München
zur Erlangung des akademischen Grades eines
Doktor-Ingenieurs
genehmigten Dissertation

Vorsitzender: Univ.-Prof. Dr.-Ing. Dr.-Ing. E.h. H. Marko
1. Prüfer: Univ.-Prof. Dr.-Ing. E. Terhardt
2. Prüfer: Univ.-Prof. Dr.-Ing. Th. Einsele

Die Dissertation wurde am 15.10.1987 bei der
Technischen Universität München eingereicht und durch
die Fakultät für Elektrotechnik und Informationstechnik
am 23.12.1987 angenommen.

Tag der Promotion: 20.1.1988

Meiner Familie

Inhaltsverzeichnis

Seite

Seite

1. Einleitung

In der vorliegenden Arbeit wird eine neue Methode zur gehörgerechten Gewinnung der wesentlichen informationstragenden Merkmale von Audiosignalen vorgestellt.

Die akustische Kommunikation (Sprache, Musik, Geräusche) spielt eine außerordentlich wichtige Rolle im menschlichen Leben. Der physikalische Träger der Nachrichten ist das Schallsignal. Dieses wird auf seinem Weg von einer Schallquelle zum Empfänger entsprechend den Eigenschaften der Übertragungsstrecke verändert und durch andere Schallquellen gestört.

Der Empfänger kann die Nachricht nur dem veränderten und gestörten Schallsignal entnehmen. Er muß in der Lage sein, die zur Nachricht gehörenden Signalmerkmale von denen, die von der Veränderung und Störung hervorgerufen wurden, zu trennen. Das Gehör ist dazu in hohem Maße befähigt.

Da die akustische Kommunikation eine so wichtige Rolle in unserem Leben spielt, besteht beispielsweise der Wunsch,

- auch mit Maschinen in einer uns gewohnten Weise zu kommunizieren,
- Audiosignale mit möglichst geringem Aufwand und in großer Menge zu speichern und zu übertragen,
- die in Audiosignalen enthaltene Information in kontrollierter Weise zu verändern und
- mit Hörhilfen Gehörgeschädigten eine möglichst gleichwertige Kommunikation zu ermöglichen.

Um diese Ziele zu erreichen, ist es notwendig, die Information im Signal zu erfassen und zu verarbeiten. Dies erfordert Kenntnisse über die Eigenschaften der Schallquellen, über die Veränderungen des Schallsignals durch die Übertragungsstrecke und über die Art und Weise, wie das Gehör Information aufnimmt. Darüber wird im zweiten Kapitel durch Zusammenfassen der aus der Literatur bekannten Erkenntnisse und Probleme bei der Informationserfassung und -verarbeitung von Audiosignalen berichtet. Weiter werden das der vorliegenden Arbeit zugrunde liegende Untersuchungskonzept und die Vorgehensweise beschrieben.

Die aus der Literatur bekannte Vorgehensweise bei der Nachbildung der auditiven Informationsaufnahme beruht in erster Linie auf einer möglichst detaillierten Nachbildung von Gehöreigenschaften auf der Grundlage von psychoakustischen und neurophysiologischen Daten. Das Ziel der vorliegenden Arbeit ist es dagegen, den wesentlichen Vorgang der Informationsaufnahme, der aus einem Entscheidungsprozess besteht, nachzubilden. Das zeitliche und spektrale

Auflösungsvermögen des Gehörs wird im Rahmen einer vergleichsweise einfachen Vorverarbeitung (Spektralanalyse) berücksichtigt, ohne Gehöreigenschaften wie beispielsweise Mithörschwellen detailliert nachbilden zu wollen.

Eine Methode zur Bestimmung der spektralen Feinstruktur durch einen Entscheidungsprozess auf der Basis einer zeitvariablen Spektralanalyse ist Gegenstand des dritten Kapitels.

Die Beschreibung der zeitvariablen spektralen Feinstruktur erfolgt mit zeitvariablen Teiltönen, die in ihrer Gesamtheit als Teiltonzeitmuster bezeichnet werden. Das Teiltonzeitmuster wird im vierten Kapitel definiert und bezüglich seiner Eigenschaften bei der Abbildung von mathematisch einfach beschreibbaren Signalen untersucht. Weiter wird eine graphische Darstellung des Teiltonzeitmusters, das Maxigramm, vorgestellt und ein Verfahren zur Resynthese von Audiosignalen aus dem Teiltonzeitmuster beschrieben.

Die Eignung des Teiltonzeitmusters zur Informationserfassung und -verarbeitung hängt entscheidend davon ab, inwieweit es genau diejenigen Signalmerkmale (als wesentliche Information bezeichnet) enthält, welche vom Gehör ausgewertet werden. Im fünften Kapitel wird deshalb über Untersuchungen berichtet, die mit aus dem Teiltonzeitmuster resynthetisierten Sprach- und Musiksignalen durchgeführt wurden. Zugleich werden Anwendungen wie Datenreduktion oder Signalbearbeitung beschrieben.

Im sechsten Kapitel sind die Ergebnisse der vorliegenden Arbeit zusammengefasst dargestellt.

2. Problemstellung und Untersuchungskonzept

Dieses Kapitel ist in vier Teile gegliedert. Der erste Teil handelt von den Grundprinzipien der akustischen Kommunikation. Die wichtigsten bekannten Eigenschaften des Gehörs bei der Informationsaufnahme sind Gegenstand des zweiten Teils dieses Kapitels. Der dritte Teil enthält eine Übersicht über die bekannten Methoden der Gewinnung von Information aus Audiosignalen. Im vierten und letzten Teil wird das Konzept und die Vorgehensweise bei der vorliegenden Arbeit dargelegt.

2.1 Akustische Kommunikation

Zur Kommunikation über einen Nachrichtenkanal im Sinne der Informations- und Kommunikationstheorie nach Shannon [70] gehören: eine Nachrichtenquelle X, ein Coder C, der Nachrichtenkanal K mit Störsignalquelle, ein Decoder D und die Nachrichtensenke (Empfänger) Y, wie in Fig. 2.1-1 dargestellt (nach [49]).

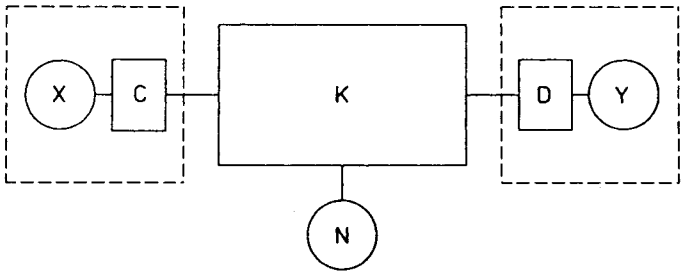


Fig. 2.1-1: Blockschaltbild des verallgemeinerten Nachrichtenkanals der Informations- und Kommunikationstheorie nach Shannon [49], [70].
Nachrichtenquelle X, Coder C, Nachrichtenkanal K mit Störsignalquelle, Decoder D und Nachrichtensenke Y.

Bei der idealisierten akustischen Kommunikation sendet die Schallquelle ein Schallsignal (Nutzsignal) aus, das eine Nachricht trägt. Dieses wird durch die Luft oder andere Medien übertragen. Im allgemeinen wird das Nutzsignal entsprechend den Eigenschaften der Übertragungsstrecke verändert. Zusätzlich erfolgt eine Störung z.B. aufgrund der Eigenschaften der Übertragungsstrecke (Verzerrungsprodukte, Echos) oder durch andere Schallquellen. Der Schallempfänger empfängt das Nutzsignal zusammen mit den überlagerten Störsignalen und gewinnt daraus eine Nachricht. Aufgrund der Veränderungen und Störungen kann die empfangene Nachricht von der gesendeten abweichen.

Bei der realen akustischen Kommunikation läßt sich die Unterscheidung zwischen Nutzsignal und Störsignal nicht so einfach wie im zuvor dargestellten Fall treffen. So ist das Klingeln des Telephons während eines Gesprächs oder

beim Anhören von Musik gleichzeitig Störsignal und Nutzsignal. Zum einen wird die gerade stattfindende Aufnahme von Sprache oder Musik gestört, zum anderen kann jenes Klingeln von besonderer Wichtigkeit für den Empfänger sein, dessen Wahrnehmung vom Sprach- oder Musiksignal gestört wird. Im allgemeinen Fall muß beim Vorhandensein mehrerer überlagerter Signale der Empfänger entscheiden, welches die Nutz- und welches die Störsignale sind.

Auf die reale akustische Kommunikation trifft daher eher das in Fig. 2.1-2 dargestellte Blockdiagramm zu. Es sind mehrere Nachrichtenquellen vorhanden, die ihre Signale über im allgemeinen unterschiedliche Übertragungsstrecken mit unterschiedlichen Störungen zum Empfänger senden. Dieser hat in gewissem Umfang die Fähigkeit, die einzelnen Nachrichten quasi gleichzeitig dem Gesamtsignal zu entnehmen.

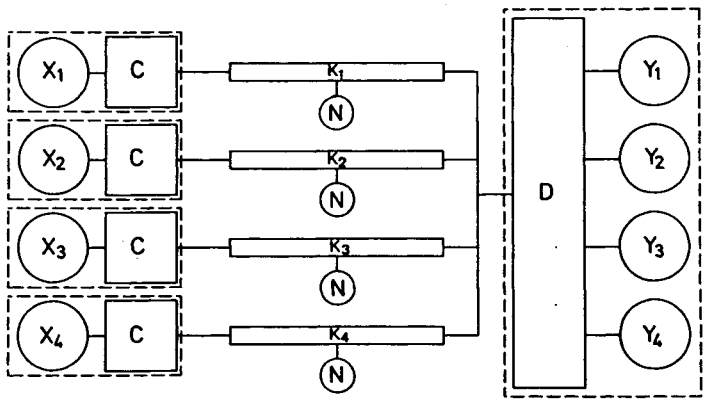


Fig. 2.1-2: Blockschaltbild der realen akustischen Kommunikation mit mehreren Signalquellen und Trennung derselben im Empfänger.

2.1.1 Schallquellen

Schallquellen wie die menschliche Stimme oder Musikinstrumente senden zeitlich und spektral strukturierte Signale aus. Die Nachricht ist sowohl in der Zeitstruktur als auch in der Spektralstruktur enthalten. Zwei wesentliche Signalarten lassen sich unterscheiden: regelmäßige, tonale Signale und regellose, geräuschhafte Signale.

Das Spektrum tonaler Signale besteht aus einzelnen Teiltönen. Bei periodischen Zeitsignalen stehen die Teiltonfrequenzen in ganzzahligen (harmonischen) Verhältnissen zueinander. Bei stimmhaften Sprachlauten (Vokalen) ist dies beispielsweise weitgehend der Fall. Bei anderen Schallquellen die tonale Signale aussenden, wie z.B. Glocken, können die Teiltonfrequenzen auch inharmonische Zahlenverhältnisse aufweisen [86]. Die Information kann in der Anordnung der Teiltöne, ihren Amplitudenverhältnissen (Hüllkurve), der Grundfrequenz harmonischer Klänge und im zeitlichen Verlauf dieser Parameter enthalten sein.

Das Spektrum regelloser Signale ist kontinuierlich. Das heißt, die Energie ist nicht auf sehr enge Bereiche konzentriert, wie bei den Teiltönen tonaler Signale, sondern über einen vergleichsweise weiten Frequenzbereich verteilt. Informationstragend ist die spektrale Hüllkurve und deren zeitliche Veränderung.

2.1.2 Übertragung

Die akustische Übertragungsstrecke läßt sich weitgehend als ein lineares System auffassen. Die Veränderungen des Schallsignals lassen sich somit systemtheoretisch durch die Faltung der Impulsantwort der Übertragungsstrecke mit dem Signal der Quelle beschreiben. Im freien Schallfeld ohne Reflexionen tritt nur die frequenzabhängige Dämpfung der Luft auf. Bei der Übertragung in Räumen treten zusätzlich Reflexionen an den Wänden auf. Die zugehörige Übertragungsfunktion entspricht der eines linearen Systems mit einer großen Anzahl statistisch verteilter, dicht benachbarter Eigenfrequenzen [67]. Der genaue Verlauf der Übertragungsfunktion hängt sehr stark von den räumlichen Positionen von Schallsender und Schallempfänger ab.

Die vom akustischen Signal durchlaufene Übertragungsstrecke bewirkt somit eine unsystematische, ortsabhängige Veränderung der spektralen Verteilung der Amplituden und Phasen eines Signals. Der einzige Signalparameter, der von ihr nicht verändert wird, ist die Frequenz diskreter Teiltöne [92].

2.1.3 Empfänger

Das am Empfänger ankommende Signal besteht aus der Überlagerung der veränderten Signale der einzelnen Schallquellen. Der Empfänger kann die Aufgabe der Trennung der einzelnen Schallsignale dann am besten durchführen, wenn er sich an ihre Eigenschaften und die Art der Signalveränderung bei der Übertragung anpaßt. Folgende Merkmale können in der Regel zur Trennung ausgenutzt werden:

- die unterschiedlichen Orte der Schallquellen (Richtwirkung);
- das ganzzahlige Frequenzverhältnis der Teiltöne periodischer Signale;
- der gemeinsame zeitliche Verlauf der Teiltonfrequenzen periodischer Signale bei zeitlichen Änderungen der Grundfrequenz (trifft meistens auch für aperiodische, tonale Signale zu);
- die Abbildung der zeitlichen Makrostruktur, besonders des Signalanfangs, in zeitvariablen spektralen Merkmalen;
- eine regelmäßige zeitliche Struktur (Rhythmus).

Zur Auswertung dieser Merkmale muß der Empfänger in der Lage sein, sowohl spektrale als auch zeitliche Strukturen möglichst fein aufzulösen.

Bei der Informationsaufnahme ergibt sich aber noch ein weiteres Problem. Die Information, die ein Sender dem Signal aufprägt, ist diskret bzw. kategorial und in ihrer Menge begrenzt. Das Signal dagegen ist stetig und kontinuierlich und enthält - nicht zuletzt aufgrund von Veränderungen und Störungen - im Prinzip unendlich viel Information. Der Empfänger kann diese Menge an Information gar nicht verarbeiten. Er ist also gezwungen, auszuwählen, zu reduzieren. Informationsaufnahme bedeutet eine Reduktion des Signals auf Symbole.

Informationsverarbeitung besteht aus Entscheidungsprozessen. Informationsverarbeitende Systeme sind hierarchisch aufgebaut. Jede Stufe baut auf den Ergebnissen der vorherigen auf. Bei der Verarbeitung darf keine Information verlorengehen, die von höheren Ebenen noch benötigt wird, da diese sonst unwiederbringlich verloren ist. Die erste Stufe befindet sich dort, wo das Signal zum ersten Mal diskretisiert, also durch Symbole beschrieben wird.

Diskretisieren bedeutet immer weglassen von Information. Es ist daher notwendig, daß der Empfänger bei der Diskretisierung zwischen wesentlicher und unwesentlicher Information unterscheidet und sich auf die Erfassung der wesentlichen Information beschränkt. Im Fall der akustischen Kommunikation entspricht die wesentliche Information den Signalmerkmalen, die

- von einer Nachrichtenquelle zum Zweck der Informationsübermittlung generiert wurden und
- die der Empfänger im Rahmen der physikalischen Gegebenheiten (Frequenz- und Zeitauflösungsvermögen) auch erfassen kann.

2.2 Informationsaufnahme durch das Gehör

Die Informationsaufnahme durch das Gehör läßt sich in zwei Bereiche unterteilen:

- Umformung der Schallsignale in Nervenimpulse, eine Aufgabe des peripheren Gehörs,
- Verarbeitung und Auswertung der Nervenimpulse im zentralen Gehör.

Die Umformung der Schallsignale in Nervenimpulse entspricht der im vorigen Abschnitt erwähnten Diskretisierung. Das Schallsignal mit seiner unendlichen Menge an Information wird in eine endliche Anzahl von Nervenimpulsen gewandelt. Die Verarbeitung und Auswertung durch das zentrale Gehör ist somit auf die Merkmale begrenzt, die vom peripheren Gehör aufgenommen werden.

Durch psychoakustische Messungen ist schon seit langem bekannt, welche Anteile eines Schalles wahrgenommen werden können (z.B [96]). Mit den meisten dieser Messungen werden Wahrnehmungsgrenzen bestimmt, die auf Eigenschaften

des peripheren Gehörs zurückzuführen sind. Sämtliche Wahrnehmungsgrenzen sind frequenzabhängig, die meisten stehen in einer Beziehung zur Frequenzgruppenbreite [96],[101]. Wichtige Wahrnehmungsgrenzen sind beispielsweise:

- der hörbare Frequenzbereich [96], [101];
- Frequenz-, Amplituden-, und Phasenunterschiedsschwellen [96], [101];
- die Zeitkonstanten des Gehörs [98];
- das Frequenzauflösungsvermögen (Trennung spektraler Anteile) [57], [78].

Die Wahrnehmungsgrenzen begrenzen die Informationsaufnahme und -verarbeitung. Andererseits ist eine Verarbeitung erst möglich, wenn der Informationsfluß durch Wahrnehmungsgrenzen eingeschränkt wird. Damit aber Information überhaupt übertragen werden kann, muß eine gewisse Anpassung zwischen Sender und Empfänger bezüglich der informationstragenden Merkmale bestehen. Vergleicht man die Wahrnehmungsgrenzen des Gehörs mit den Eigenschaften des Sprachsignals hinsichtlich der zeitlichen und spektralen Struktur, so ist offensichtlich, daß eine sehr hohe Anpassung zwischen Spracherzeugung und Sprachwahrnehmung besteht, die sich generell auf die Rezeption von Schallen auswirkt [84], [88].

Durch zahlreiche Untersuchungen mit Musik- und Sprachsignalen wurde nachgewiesen, daß das Gehör in gewissem Umfang in der Lage ist, verschiedene Quellen zu unterscheiden und die Nachricht einer Quelle selektiv aufzunehmen Dazu werden verschiedene Merkmale ausgewertet, wie beispielsweise:

- Ortung durch binaurales Hören [13], (s.a. Übersicht in [20]);
- unterschiedlicher zeitlicher Einsatz verschiedener Quellen [18], [63];
- unterschiedliche Grundfrequenzen [65], [103];
- zeitliche und spektrale Nähe (Streaming-Effekte) [10], [52].

Im allgemeinen überlagern sich die Spektren verschiedener Quellen. Aufgrund von zeitlichen Veränderungen spektraler Anteile sind Ort und Art von Überschneidungen nicht konstant. Vielmehr tritt ein "Durchkreuzen" der spektralen Anteile mehrerer Quellen auf. Als Folge der begrenzten zeitlichen und spektralen Analysefähigkeit des Gehörs können überlagerte bzw. dicht benachbarte Signalanteile nicht vollständig voneinander getrennt werden; sie maskieren sich gegenseitig [101].

Das zentrale Gehör ist in der Lage, fehlende oder verdeckte Signalmerkmale aus vorhandenen in gewissen Grenzen zu rekonstruieren (s. z.B. Übersicht in [20]). Im Gegensatz zur rein analytischen Funktion des peripheren Gehörs besitzt das zentrale Gehör somit auch einen synthetischen Wahrnehmungsmodus [81]. Ein vielzitiertes Beispiel dafür ist die Wahrnehmung der "richtigen" Tonhöhe eines Sprechers auch dann, wenn das Sprachsignal nach Hochpassfilterung die

zugehörige Spektralkomponente der Glottisschwingung nicht mehr enthält [79]. Die Strategie der Analyse und Synthese, die Bestandteil der Gestaltwahrnehmung ist, ermöglicht es dem Gehör, auch bei starker Störung die gewünschte Signalinformation aufzunehmen. Die Wahrnehmung 'akustischer Gestalten' scheint hierbei das wesentliche Prinzip zu sein [84], [90], [91], [92]. Besonders deutlich zeigt sich diese Gestaltwahrnehmung beim Hören von Musik [20], [50], [91], bei der Wahrnehmung von virtuellen Tonhöhen [79], [81] oder bei der Gruppierung zeitlicher Ereignisse [10], [11], [28], [54].

Die beschriebenen Leistungen des Gehörs lassen sich nur durch Auswertung der zeitvariablen spektralen Merkmale des Signals erklären. Die fundamentale Bedeutung der ersten Verarbeitungsstufe besteht eben darin, diese Merkmale zu bestimmen und zur Verfügung zu stellen.

Die spektralen Merkmale lassen sich in solche der Grob- und Feinstruktur einteilen. Empfindungen wie Lautheit und Schärfe beruhen hauptsächlich auf der Auswertung der spektralen Grobstruktur oder Hüllkurve [6], [101]. Diese ist auch maßgeblich für die Sprachwahrnehmung [26], zumindest im ungestörten Fall. Ebenso bestimmt sie zusammen mit der Feinstruktur die Klangfarbenwahrnehmung [3], [4], [6].

Die Fähigkeiten des Gehörs zur Wahrnehmung der Feinstruktur wurden am intensivsten im Rahmen der Tonhöhenwahrnehmung untersucht. Die Wahrnehmung der Feinstruktur hängt in erster Linie von der Frequenzselektivität des peripheren Gehörs ab. So werden zwei Sinustöne mit gleichem Pegel dann als zwei getrennte Tonhöhen wahrgenommen, wenn ihr Frequenzabstand mehr als 50-75% der Frequenzgruppenbreite beträgt [57], [78]. Bei komplexen Schallen mit vielen Teiltönen, wie z.B. Vokalen, lassen sich nicht alle Teiltöne selektiv wahrnehmen. In Abhängigkeit von der Versuchsmethode und der spektralen Hüllkurve der Schalle kann die Wahrnehmung von 8-12 Teiltönen nachgewiesen werden, unabhängig davon, ob die Teiltöne harmonisch sind oder nicht [57], [73].

Die Teiltonfrequenzen eines Signals werden durch die Übertragungsstrecke oder durch andere Quellen nicht verändert. Bei periodischen Signalen besteht ein fester Zusammenhang zwischen der Frequenz eines Teiltons und der zugehörigen Grundfrequenz. Die Teiltonfrequenzen eignen sich daher sehr gut zur Unterscheidung mehrerer Quellen. Das Gehör scheint diesen Sachverhalt zur Trennung auszunutzen: Zwei überlagerte Vokale werden besonders gut getrennt (d.h. einzeln erkannt), wenn sich ihre Grundfrequenz um etwa ein bis drei Halbtöne unterscheidet. Werden die gleichen Vokale jedoch stimmlos dargeboten, so sinkt die Erkennungsrate deutlich ab, obwohl die spektrale Hüllkurve die gleiche ist [65].

Aus den Ergebnissen solcher Versuche läßt sich der Schluß ziehen, daß das Gehör dadurch sehr gut an seine akustische Umgebung angepasst ist, daß es zur Informationsaufnahme und -verarbeitung die spektrale Feinstruktur auswertet. Jene muß somit bei einer Nachbildung der Informationsaufnahme des Gehörs gebührend berücksichtigt werden, damit kein Informationsverlust auftritt.

2.3 Nachbildung der Informationsaufnahme

Die Berücksichtigung der Eigenschaften des Gehörs ist immer dann von Bedeutung, wenn

- Schallsignale in anderen Medien übertragen bzw. gespeichert werden (z.B. Telefon, Schallplatte);
- Audiosignale bearbeitet werden (z.B. Verbesserung alter Schallaufnahmen);
- akustische Information erfaßt und verarbeitet wird (z.B. Spracherkennung).

So wird beispielsweise bei der Übertragung und Speicherung von Audiosignalen wie selbstverständlich eine Begrenzung entsprechend dem hörbaren Frequenzbereich vorgenommen. Durch weitere Bandbegrenzung, wie beim Telefon, erreicht man eine Verringerung der notwendigen Übertragungskapazität - aber nur wenn dabei die Eigenschaften des Gehörs berücksichtigt werden [27]. Ein Übertragungssystem mit gehörgerechter Bandbegrenzung bildet die Informationsaufnahme des Gehörs auf einer sehr niedrigen Stufe nach, da es nur Informationen in einem Frequenzbereich aufnimmt und weitergibt, in dem dieses auch vom Gehör bevorzugt durchgeführt wird. Die begrenzte Wahrnehmung ist auch ein wichtiger Punkt bei der Entwicklung von Verfahren zur Datenreduktion von Audiosignalen [45], [93]. Solche Verfahren können derzeit nur mit Hörtests überprüft werden, da es bislang noch kein Meßverfahren gibt, mit dem z.B. die Güte eines Signals objektiv beurteilt werden könnte [25], [75].

Eine grafische Darstellung der Signalmerkmale ist sehr wichtig für die Nachbildung und Untersuchung der an der Informationsaufnahme beteiligten Mustererkennungsprozesse, insbesondere für die Trennung mehrerer Schallquellen [94]. Ein Beispiel ist der schon seit langem bekannte "Sound Spectrograph" und das mit ihm gewonnene Spektrogramm [44]. Das Spektrogramm stellt das Sprachsignal in so anschaulicher Weise dar, daß es geübten Spektrogrammlesern möglich ist, diesem den gesprochenen Text zu 80 - 90% zu entnehmen [15]. Für die kontrollierte Untersuchung und Veränderung von Klängen ist eine grafische Darstellung ebenfalls von Bedeutung [76]. Die Entwicklung gehörgerechter Signaltransformationen und Darstellungen ist deshalb schon lange das Ziel von Forschergruppen, die sich mit der Spracherkennung befassen.

Die Entwicklung solcher Transformationen erfolgt auf der Grundlage psychoakustischer und neurophysiologischer Daten [16], [39], [99], [102]. Man erhofft

sich davon die Entwicklung einer verbesserten Formantextraktion und einer gehörgerechten Metrik zur Beschreibung und Bestimmung der Formantverläufe [7], [8], [12], [16], [39], [41], [42], [69], [77]. Die erste Stufe (Vorverarbeitung) besteht in der Regel aus einer Filterbank, deren Filter entsprechend den Frequenzgruppen angeordnet sind [39]. Zum Teil werden nichtlineare Eigenschaften des Gehörs einbezogen [16], [102]. Alle Autoren werten in erster Linie die spektrale Grobstruktur, also die Hüllkurve aus. Durch Bestimmung der in einer Frequenzgruppe vorherrschenden Anregungsfrequenz wird versucht, Merkmale der spektralen Feinstruktur, z.B. zur Formantbestimmung [42], einzubeziehen [16], [69]. Es ist jedoch nicht gesichert, daß die Auswertung der spektralen Grobstruktur, vor allem die Bestimmung der Formanten, auf dieser ersten, der Peripherie entsprechenden Stufe stattfindet [7].

Der wesentliche Vorgang bei der Informationsaufnahme des Gehörs sind Entscheidungsprozesse. Das Ziel dieser Arbeit ist es, jene Entscheidungsprozesse im Prinzip nachzubilden und zu untersuchen. Das zeitliche und spektrale Auflösungsvermögen des Gehörs wird nur im Rahmen einer vergleichsweise einfachen Vorverarbeitung berücksichtigt. Die detaillierte Nachbildung oder Erklärung psychoakustischer oder neurophysiologischer Messungen wird nicht angestrebt.

2.4 Untersuchungskonzept

Zur Definition von "Schnittstellen" und zur Abgrenzung von Fragestellungen bei der Beschreibung und Untersuchung der Informationsaufnahme des Gehörs wurde von Terhardt das von Popper & Eccles [59] entwickelte Konzept der "Drei Welten" auf die akustische Kommunikation übertragen [92]:

- Welt 1: Physikalische Welt, z.B. Audiosignal;
- Welt 2: Subjektives Empfinden, z.B. Lautheit, Klangfarbe;
- Welt 3: Information, z.B. Sprachlaute, Wörter, musik. Töne, Akkorde

Die bereits mehrfach erwähnte Diskretisierung bzw. Kategorisierung ist das Grundprinzip der "Welt 3" und macht den wesentlichen Unterschied zwischen der Welt 1, dem Signal, und der Welt 2, den Empfindungen aus. Ein Beispiel für Kategorisierung ist die Klassifikation eines Sprachsignals als /a/ oder /e/.

Im Visuellen ist der erste elementare Entscheidungsprozess die Wahrnehmung von Konturen. Aus diesen ergeben sich Gestalten wie Schriftzeichen, Gesichter oder Gegenstände. Die visuellen Konturen stellen somit den Schlüssel zur optischen Informationsverarbeitung dar.

Als auditives Gegenstück zu den visuellen Konturen bezeichnet Terhardt die Spektraltonhöhen. Das von Terhardt 1972 [79] vorgestellte Konzept der Spektraltonhöhen und das daraus resultierende Spektraltonhöhenmuster [83], [74] stellt die psychophysikalische Grundlage der vorliegenden Arbeit dar.

Nach diesem Konzept läßt sich jede Spektraltonhöhe unmittelbar auf eine physikalische Ursache zurückführen, aber nicht jede spektrale Komponente eines Schalles erzeugt eine Spektraltonhöhe. Das Spektraltonhöhenmuster ist die Basis für die Gesamttonhöhenwahrnehmung, die virtuelle Tonhöhe. Ohne Spektraltonhöhen gibt es keine virtuellen Tonhöhen [79], [81]. Ebenso ist es geeignet zur Beschreibung der tonalen Eigenschaften eines Klanges und eines großen Teils der Klangfarbenwahrnehmung [3], [4], [74].

Das gegenüber [79] erweiterte Modell erklärt die Spektraltonhöhen zu den Trägern der wesentlichen Information und weist ihnen damit eine Schlüsselrolle bei der Erfassung und Verarbeitung von Audiosignalen zu [90], [92]. Insbesondere sollen sie das Ergebnis des ersten, informationsgewinnenden Entscheidungsprozesses sein. Auf die Richtigkeit dieser Auffassung lagen schon bisher zahlreiche Hinweise vor. Mit der vorliegenden Arbeit wird sie unmittelbar nachgewiesen.

Zur Untersuchung, welche Signalmerkmale informationstragend sind, hat sich in vielen Fällen die Methode der Analyse und Resynthese bewährt [2], [31], [64]. Die zu untersuchenden Merkmale werden durch eine entsprechende Analyse aus dem Signal gewonnen. Direkt oder nach gezielter Veränderung wird aus ihnen ein Zeitsignal resynthetisiert. Hörtests erlauben sodann die Beurteilung des Beitrags der untersuchten Merkmale an der Gesamtempfindung. Die Methode der Analyse und Resynthese wird deshalb auch in der vorliegenden Arbeit angewendet.

Es wird eine Methode zur gehörgerechten Bestimmung eines zeitvariablen Musters von Teiltönen als physikalisches Äquivalent des Spektraltonhöhenmusters entwickelt und durch die Methode der Analyse und Resynthese gezeigt, daß dieses Muster tatsächlich die wesentliche Information trägt. Dies wird in folgenden Schritten durchgeführt:

- Kurzzeitspektralanalyse unter Beachtung des zeitlichen und spektralen Auflösungsvermögens des Gehörs;
- Reduktion des Kurzzeitspektrums auf Teiltöne;
- Untersuchung einer Beschreibungsform und Darstellung der Teiltöne, die eine weitere Verarbeitung und Auswertung im Hinblick auf Gestaltwahrnehmung ermöglicht;
- Resynthese der Teiltöne zu einem Zeitsignal und dessen auditive Überprüfung;
- Voruntersuchungen zur Anwendung bei der Analyse und Verarbeitung von Sprache und Musik.

3. Spektralanalyse

Die Bestimmung einer zeitvariablen spektralen Feinstruktur setzt die Beschreibung des Zeitsignals durch ein zeitvariables Spektrum voraus. In Abschnitt 3.1 werden deshalb die Grundlagen der Kurzzeitspektralanalyse kurz erläutert und in Abschnitt 3.2 die wesentlichen Eigenschaften der Fourier-t-Transformation beschrieben. Abschnitt 3.3 befaßt sich mit der Auswertung der spektralen Feinstruktur; Abschnitt 3.4 mit der Parameterwahl und numerischen Berechnung.

3.1 Kurzzeitspektralanalyse

Der Zusammenhang zwischen Zeitfunktion und Spektrum wird durch die beiden Integrale der Fourier-Transformation beschrieben:

$$P(\omega) = \int_{-\infty}^{+\infty} p(t) \cdot e^{-j\omega t} \, dt \qquad (3.1.1a)$$

$$p(t) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} P(\omega) \cdot e^{j\omega t} \, d\omega \qquad (3.1.1b)$$

Die Fourier-Transformation hat sich als ein außerordentlich wichtiges und leistungsfähiges Werkzeug zur Analyse, Synthese und Verarbeitung von Signalen auf vielen Gebieten erwiesen. Die in Gl. (3.1.1) angegebenen Transformationen werden praktisch von allen Autoren verwendet [14], [26], [48], [53].

Im Gegensatz zu analytisch beschreibbaren Signalen kann das Fourier -Integral formal nicht zur Analyse von realen, kausalen Zeitsignalen verwendet werden, da das Signal nur bis zum Beobachtungszeitpunkt t_a, der 'Gegenwart', bekannt ist. Die Transformation Gl. (3.1.1a) muß daher so modifiziert werden, daß mit ihr auch die Analyse zeitlich begrenzter Signale möglich ist. Sie wird dann als Kurzzeitspektralanalyse bezeichnet. Sie stellt ein wichtiges Werkzeug zur Verarbeitung von Signalen, insbesondere von Audiosignalen dar [1], [14], [22], [53], [60], [61], [62], [66]. Die Art und Weise, in der eine Kurzzeitspektralanalyse verwendet wird, ist jedoch nicht einheitlich. Je nach Anwendungsfall werden eine Vielzahl verschiedener Analysefenster oder -filter und Analyseintervalle vorgeschlagen, die jeweils ganz bestimmte Vor- und Nachteile aufweisen [40].

Anmerkung: Der Einfachheit halber werden in diesem Kapitel alle zu logarithmierenden Größen auf ihre jeweilige Grundeinheit normiert betrachtet. Dementsprechend ist etwa die Amplitude A im Ausdruck L=20·log(A)dB auf 1V normiert.

Nach Terhardt [85] lassen sich Kurzzeitspektralanalysen prinzipiell anhand ihres Analyseintervalls einteilen. Als Analyseintervall wird das Zeitintervall bezeichnet, innerhalb dessen sich das Zeitsignal aus dem Spektrum wieder vollständig herstellen läßt. Insbesondere muß man zwischen endlichen und unendlichen Analyseintervallen unterscheiden. Drei wichtige Begriffe werden im Zusammenhang mit Kurzzeitspektralanalysen gebraucht: das Zeitauflösungsvermögen, das Frequenzauflösungsvermögen und das sogenannte BT-Produkt.

Das Zeitauflösungsvermögen beschreibt die Reaktion des Kurzzeitspektrums auf zeitliche Änderungen des Signals, wie z.B. bei Frequenz- oder Amplitudenmodulation. Das Frequenzauflösungsvermögen beschreibt die Grenze, bei der zwei benachbarte Spektralkomponenten gleicher Amplitude noch voneinander getrennt werden können. Das BT-Produkt bestimmt den Zusammenhang zwischen Frequenz- und Zeitauflösungsvermögen; es wird meist ein möglichst kleiner Wert angestrebt.

Das komplexe Kurzzeitspektrum $P(\omega_n,t_a)$, das man mit einem endlichen Analyseintervall T zum Beobachtungszeitpunkt t_a erhält, ist nur bei ganzzahligen Vielfachen der Frequenz $f_1 = 1/T$ definiert:

$$P(\omega_n,t_a) = \int_{t_a-T}^{t_a} p(t) \cdot e^{-j\omega nt}\, dt \qquad (3.1.2)$$

mit $\omega_n = 2\pi n/T; \quad n = 1, 2, 3...$

Da das Analyseintervall T das Zeitauflösungsvermögen bei allen Analysefrequenzen ω_n bestimmt, ist dieses bei hohen Analysefrequenzen relativ ungünstiger als bei tiefen Frequenzen. Das BT-Produkt ist bei allen Analysefrequenzen gleich Eins. Da die Analysebandbreite B bei allen Frequenzen f_n gleich groß ist, steigt die Güte $Q_n = f_n/B$ eines vergleichbaren Filters mit steigender Analysefrequenz f_n an.

Natürliche Spektralanalysatoren, wie z.B. das menschliche Gehör, weisen näherungsweise einen konstanten Verlauf der Güte über der Frequenz auf. Das heißt: bei hohen Frequenzen ist das absolute Frequenzauflösungsvermögen schlechter als bei tiefen Frequenzen, dagegen ist das absolute Zeitauflösungsvermögen besser. Mit einer Transformation wie in Gl. (3.1.2) ist es prinzipiell nicht möglich, unmittelbar ein Spektrum zu bestimmen, dessen relatives Frequenz- *und* Zeitauflösungsvermögen frequenzunabhängig ist. Dies ist aber eine Voraussetzung zur Durchführung einer gehörgerechten Spektralanalyse. Darauf wird auch von anderen Autoren hingewiesen (beispielsweise [16], [39]). Zusätzlich werden insbesondere zeitliche Änderungen von Signalmerkmalen innerhalb des Analyseintervalls T nur unzureichend oder verfälscht wiedergegeben [72].

Trotzdem sind Verfahren, die auf einem endlichen Analyseintervall beruhen, sehr verbreitet, da mit der Fast- Fourier-Transformation (FFT) ein schneller und leistungsfähiger Algorithmus zur Berechnung des Kurzzeitspektrums zur Verfügung steht [14], [17], [53].

Bei beliebigen Analysefrequenzen können die Werte des Kurzzeitspektrums nur bei unendlichem Analyseintervall bestimmt werden [87]. Daraus folgt jedoch nicht, daß beide Grenzen des Transformationsintegrals unendlich sein müssen. Vielmehr wird die Obergrenze durch den Beobachtungszeitpunkt t_a bestimmt und die Untergrenze durch den Zeitpunkt t = 0, in dem das Signal p(t) eingeschaltet wird. Das heißt, für $t < 0$ hat p(t) den Wert Null. Läßt man das Analyseintervall bei $t = -\infty$ beginnen, so enthält das komplexe Spektrum $P(\omega,t)$ explizit die Information, daß das Signal für $t < 0$ gleich Null ist. Macht man den Beobachtungszeitpunkt t_a gleitend und gleich dem aktuellen Zeitpunkt t, so erhält man die Transformation:

$$P(\omega,t) = \int_0^t p(x) \cdot e^{-j\omega x} \, dx \qquad ; t > 0 \qquad . \tag{3.1.3}$$

Mit zunehmender Zeit t wächst der Betrag des Spektrums immer mehr an, so daß die jüngsten Werte des Zeitsignals p(t) in Bezug zum Gesamtwert $P(\omega,t)$ immer weniger beitragen. Aus diesem Grund wird eine Bewertungsfunktion g(t-x) eingeführt, die das Zeitsignal für weiter zurückliegende Zeitpunkte bedämpft [26], [29]. Damit das Analyseintervall tatsächlich unendlich lang bleibt, darf sich g(t-x) nur asymptotisch dem Nullpunkt nähern. Terhardt [87] schlägt dazu eine Exponentialfunktion $g(t-x) = e^{-a(t-x)}$ vor, entsprechend einem Tiefpass erster Ordnung:

$$P(\omega,t) = \int_0^t p(x) \cdot e^{-a(t-x)} \cdot e^{-j\omega x} \, dx \qquad ; t > 0 \; . \tag{3.1.4}$$

Die Transformationskonstante a ist gleich dem Kehrwert der Zeitkonstanten, mit der die Bewertungsfunktion für zurückliegende Zeitpunkte abnimmt.

Kurzzeitspektralanalysen entsprechend Gl. (3.1.4) - wenn auch mit anderen, aber ähnlichen Bewertungsfunktionen - wurden bereits von mehreren Autoren für eine gehörgerechte Spektralanalyse vorgeschlagen [9], [25], [29], [68]. Transformationen diesen Typs entsprechen einer Spektralanalyse durch reale, das heißt kausale, lineare Systeme [22], [30], [66]. Zur Beschreibung des Ausgangssignals

eines linearen Systems zum Zeitpunkt t_a ist es notwendig, den Verlauf des gesamten Eingangssignals bis zum Zeitpunkt t_a zu kennen [87].

In der vorliegenden Arbeit wird ausschließlich die Transformation Gl. (3.1.4) verwendet. Sie wird nach einem Vorschlag von Terhardt [87] als Fourier-t-Transformation (FTT) bezeichnet.

3.2 Fourier-t-Transformation

Die wichtigsten Eigenschaften der FTT, wie in [24] und [87] beschrieben, werden hier zusammengefaßt wiedergegeben, da sie zum Verständnis der nachfolgenden Abschnitte wichtig sind.

Bei der Analyse eines Zeitsignals $p_T(t) = A_T \cdot \cos(\omega_T \cdot t + \varphi_T)$ erhält man mit Gl. (3.1.4) und den Näherungen $\omega + \omega_T \gg \omega - \omega_T$ und $\omega + \omega_T \gg a$ das normierte Leistungsspektrum* [24]:

$$F(\omega,t) = \left|2a \cdot P(\omega,t)\right|^2 = \underbrace{\frac{A_T^2}{\left(1 + \left(\frac{\omega - \omega_T}{a}\right)^2\right)}}_{F_S(\omega)} \cdot \underbrace{\left(1 - 2 \cdot \cos(\omega - \omega_T)t \cdot e^{-at} + e^{-2at}\right)}_{F_T(\omega,t)} . \tag{3.2.1}$$

Dieses besteht aus einem transienten Teil $F_T(\omega,t)$, der für $t \to \infty$ verschwindet, und einem stationären Teil $F_S(\omega)$. Der Term $F_S(\omega)$ in Gl. (3.2.1) entspricht der quadrierten Übertragungsfunktion eines einfachen Bandpasses mit der Bandbreite $B = a/\pi$. Der zweite Term $F_T(\omega,t)$ bestimmt das Einschwingverhalten des Leistungsspektrums bei der Frequenz $\omega = \omega_T$, welches durch die Zeitkonstante $T = 1/a$ beschrieben werden kann. Zum Zeitpunkt $t = 1/a$ hat das Leistungsspektrum Gl. (3.2.1) seinen Endwert bis auf 4dB erreicht. Nach [87] gilt somit für das BT-Produkt:

$$BT = \frac{a}{\pi} \cdot \frac{1}{a} = \frac{1}{\pi} \quad . \tag{3.2.2}$$

Das Spektrum $F(\omega,t)$ eines zum Zeitpunkt $t = 0$ eingeschalteten Sinustons ist in Fig. 3.2-1 für mehrere Zeitpunkte dargestellt (nach [87]). Die Lage der Nebenmaxima und -minima wird durch den cos - Ausdruck im zweiten Teil von Gl. (3.2.1) bestimmt. Mit zunehmender Zeit t wird der Abstand der Nebenmaxima kleiner und ihre Ausgeprägtheit nimmt ab. Darauf wird in Abschnitt 3.3 näher eingegangen.

---

*) Rein formal ist $F(\omega,t)$ das Betragsquadratspektrum. In Anlehnung an den vor allem in der englischsprachigen Literatur üblichen Sprachgebrauch (Powerspectrum), wird $F(\omega,t)$ als Leistungsspektrum bezeichnet (s.a. [14]).

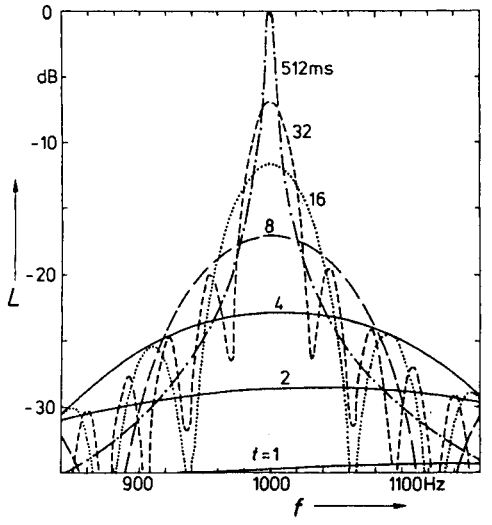


Fig. 3.2-1: Leistungsspektren eines bei t=0 eingeschalteten 1kHz-Sinustons zu mehreren Zeitpunkten (nach [87]).
$L = 10 \cdot \log [F(\omega,t)/F(2\pi \cdot 1kHz, 512ms)]dB$

Bei einer Analysezeit t » 1/a bildet sich ein stationäres Hauptmaximum aus. Für die weitab von ω_T liegenden Anteile des Leistungsspektrum gelten die für Gl. (3.2.1) verwendeten Näherungen nicht mehr. Dort tritt ein 'Schaukeln' des Spektrums mit der doppelter Signalfrequenz ω_T auf [87],

3.3 Bestimmung von Teiltönen aus der spektralen Feinstruktur

Aus Untersuchungen zur Klangfarben- und Tonhöhenwahrnehmung geht hervor, daß diese Wahrnehmungen nicht, oder nur in sehr geringem Maße, von der Phasenlage der spektralen Komponenten beeinflußt werden [58], [55]. Daraus läßt sich der Schluß ziehen, daß zur wesentlichen Informationsaufnahme des Gehörs, die auch einohrig stattfinden kann, die Phasenlage nicht ausgewertet wird. Es wird daher untersucht, ob zur Auswertung der spektralen Feinstruktur allein das FTT-Leistungsspektrum ohne Berücksichtigung der Phasenbeziehungen verwendet werden kann.

3.3.1 Abbildung einfacher Signale im zeitvariablen FTT-Leistungsspektrum

Das eingeschwungene FTT-Leistungsspektrum eines Sinustons (t » 1/a) wird von dessen Frequenz und Amplitude sowie der Transformationskonstante a bestimmt (vergl. 3.2). Fig. 3.3-1 zeigt die Spektren eines 1kHz-Sinustons im eingeschwungen Zustand zu verschiedenen Zeitpunkten mit zwei verschiedenen Analysebandbreiten. Die Lage des spektralen Maximums ist jeweils dieselbe, ebenso der zeitliche Verlauf der Schwankung der Flanken. Die Bandbreite bzw. Transformationskonstante beeinflußt lediglich die absolute Lage und die Steigung der Flanken. Dieses Spektrum weist somit eine Redundanz auf, da es bis auf die absolute Lage von Frequenz und Amplitude vollständig von der Transformationskonstanten abhängt. Die spektrale Lage des Maximums bleibt konstant, sie entspricht der Frequenz des Sinustons. Das Spektrum eines Sinustones kann somit durch ein einziges diskretes Wertepaar repräsentiert werden: Frequenz und Amplitude des Maximums.

Bei komplexen Klängen, wie sie in vielen natürlichen Schallen enthalten sind, müssen mehrere Fälle unterschieden werden. Fig. 3.3-2 zeigt die Teiltöne eines komplexen Klanges und das zugehörige FTT - Leistungsspektrum $F(\omega,t)$ eines bestimmten Zeitpunktes $t \gg 1/a$. Nicht jeder der Teiltöne bildet im *Leistungsspektrum* $F(\omega,t)$ ein Maximum aus, obwohl die Information über sämtliche Teiltöne im *komplexen* Spektrum $P(\omega,t)$ enthalten ist.

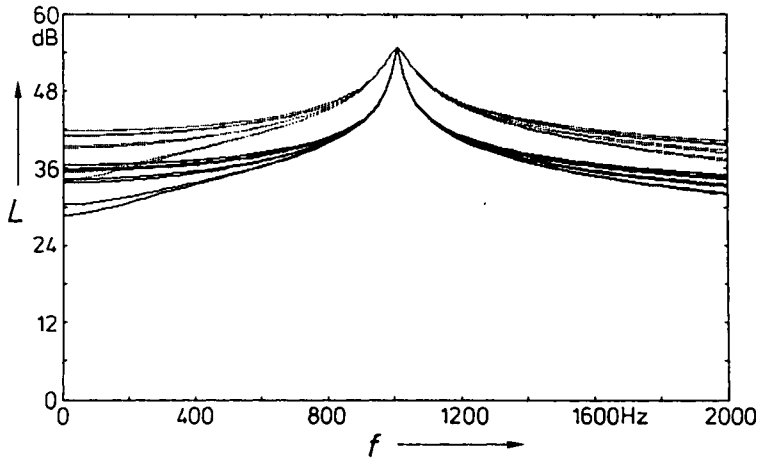


Fig. 3.3-1: Spektren eines 1kHz-Sinustons zu verschiedenen Zeitpunkten $t \gg 1/a$. Durchgezogen: Analysebandbreite $B = 15Hz$; gepunktet: $B = 50Hz$. Für die Ordinate gilt: $L=10 \cdot \log(F(2\pi f,t))dB$. Die Stufen im Verlauf des Spektrums sind eine Folge der verwendeten Rastergrafik und nicht eine Eigenschaft des Spektrums. Dies gilt in gleicher Weise auch für die folgenden Abbildungen (s.a. Anmerkung S. 12).

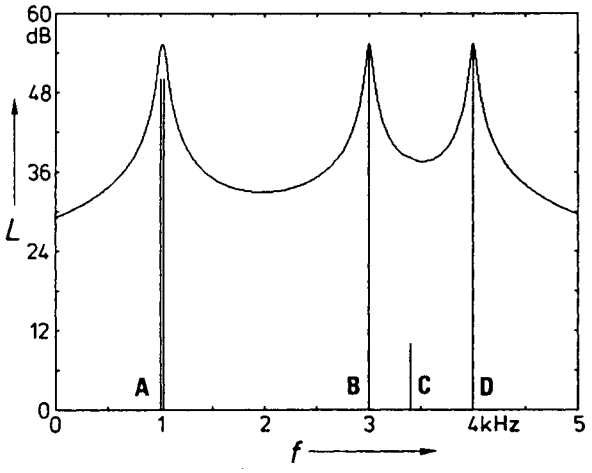


Fig. 3.3-2: Schematisches FTT-Leistungsspektrum eines komplexen Klanges zu einem Zeitpunkt $t \gg 1/a$. Die Teiltöne des Klanges sind als senkrechte Linien eingezeichnet. A-D: siehe Text. Für diese Figur und die folgenden gilt ebenfalls: $L = 10 \cdot \log(F(2\pi f,t))dB$.

Das Maximum 'A' in Figur 3.3-2 wird von zwei Teiltönen hervorgerufen. Ihr Frequenzabstand ist kleiner als die Analysebandbreite. Am Leistungsspektrum eines Zeitpunktes läßt sich nicht erkennen, ob das Maximum 'A' von einem oder mehreren Teiltönen hervorgerufen wird. Im Gegensatz zum zeitlich konstanten Maximum eines Teiltons verändert sich das Maximum 'A' zeitlich in seiner spektralen Lage und Höhe. In Fig. 3.3-3 sind die Leistungsspektren zweier dicht benachbarter Sinustöne gleicher Amplitude über eine Periode $T_{12} = 1/(f_1 - f_2)$ übereinandergezeichnet. Der Vergleich mit Fig. 3.3-1 zeigt, wie sich neben den Flanken auch das Maximum zeitlich verändert.

Die Information über mehrere dicht benachbarte Teiltöne läßt sich also dem Leistungsspektrum eines Zeitpunktes nicht entnehmen. Diese Information ist aber in den zeitlichen Veränderungen des Leistungsspektrums enthalten.

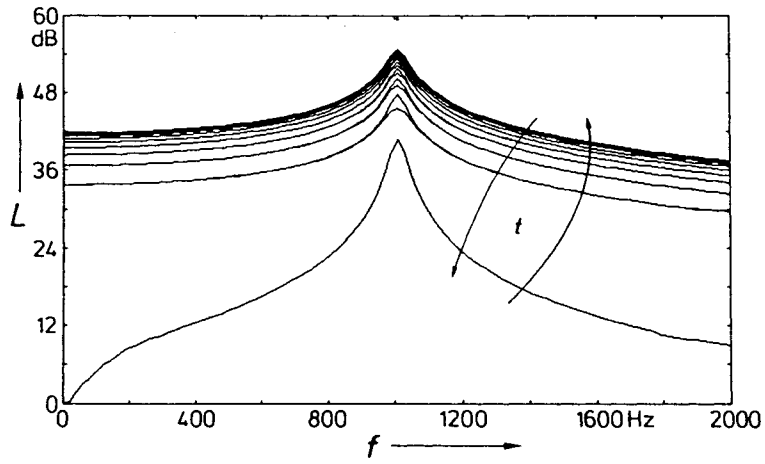


Fig. 3.3-3: Zeitliche Veränderung des FTT-Leistungsspektrums zweier dicht benachbarter Sinustöne (999Hz und 1001Hz) gleicher Amplitude im eingeschwungenen Zustand ($t \gg 1/a$). Analysebandbreite $B = 15Hz$, zeitlicher Abstand der einzelnen Spektren $\Delta t = 25ms$.

Ähnlich verhält es sich beim Gehör: Ist der Frequenzabstand zweier spektraler Komponenten (Sinustöne) so gering, daß sie nicht mehr spektral aufgelöst werden können, so werden in bestimmten Grenzen die zeitlichen Fluktuationen (Schwebungen) als Rauhigkeit oder Schwankung wahrgenommen. Im Bereich der Schwankungswahrnehmung werden beide Sinustöne als eine einzige Spektraltonhöhe mit zeitlich schwankender Lautstärke wahrgenommen. Bei zunehmendem Abstand der beiden Sinustöne nehmen Schwankungsstärke und Schwankungsperiode ab, bis statt Schwankung die sogenannte Rauhigkeit wahrgenommen wird [78], [80].

Die beiden Teiltöne B und D bilden im Leistungsspektrum in Fig. 3.3-3 zwei getrennte Maxima aus, da ihr Frequenzabstand groß gegenüber der Analysebandbreite ist. Diese Maxima sind dann zeitlich konstant, wenn die jeweiligen Anteile der Teiltöne B und D groß gegenüber denen anderer Teiltöne sind. Ebenso werden vom Gehör zwei Sinustöne bei entsprechend großem Abstand als zwei Spektraltonhöhen ohne gleichzeitige Rauhigkeit wahrgenommen [78].

Der Teilton C bildet im Spektrum Fig. 3.3-2 kein Maximum aus. Sein Anteil am Spektrum an der Stelle C ist kleiner als die Summe der Anteile von B und D. Der Teilton C wird durch die Teiltöne B und D verdeckt. Die Verdeckung schwächerer Teiltöne durch stärkere Teiltöne mit anderer Frequenz tritt auch beim Gehör auf. Der Grad der Verdeckung wird durch den Verlauf der Mithörschwellen beschrieben [96], [101].

3.3.2 Teiltonkonzept

Im folgenden wird das Konzept der Reduktion einer spektralen Verteilung auf diskrete Elemente vorgestellt. Entsprechend den Teiltönen eines komplexen Klanges werden diese Elemente Teiltöne genannt.

Damit das Leistungsspektrum im eingeschwungenen Zustand ein relatives Maximum aufweisen kann, muß eine spektrale Anregung bei der entsprechenden Frequenz vorhanden sein. Dem Maximum wird deshalb ein Teilton mit der Frequenz und dem Pegel des Maximums zugeordnet. Verändert sich das Maximum zeitlich, so soll sich der zugeordnete Teilton in gleichem Maße verändern. Treten mehrere Maxima gleichzeitig auf, so wird jedem ein Teilton zugeordnet. Jeder Teilton ist somit ein Repräsentant der im allgemeinen unbekannten spektralen Anregung, die zu einem Maximum im Spektrum führt. Verdeckten Anteilen des Signals werden keine Teiltöne zugeordnet, da sie keine Maxima ausbilden. Bei Kenntnis der Analysebandbreite läßt sich der Verlauf des Leistungsspektrums aus den extrahierten Teiltönen näherungsweise zurückgewinnen.

Die Beschreibung des Signals bzw. des Leistungspektrums durch Teiltöne bezog sich bislang ausschließlich auf Spektren im eingeschwungenen Zustand. Im allgemeinen ist jedoch zum Zeitpunkt t nicht bekannt, ob die Bedingung des eingeschwungenen Zustands zutrifft. Fig. 3.2-1 (Seite 16) zeigt die Leistungsspektren mehrerer Zeitpunkte eines zum Zeitpunkt t = 0 eingeschalteten Sinustons. Diese Spektren weisen je nach Zeitpunkt mehrere Nebenmaxima auf. Entsprechend dem Teiltonkonzept würde jedem dieser Nebenmaxima ein Teilton zugeordnet werden, obwohl das Signal nur aus einem Sinuston besteht. Es ist deshalb notwendig, daß sich die Nebenmaxima von den Maxima des eingeschwungenen Zustands unterscheiden lassen.

3.3.3 Unterdrückung der Nebenmaxima

Die Nebenmaxima in Fig. 3.2-1 sind eine Folge der endlichen Signaldauer bei unendlichem Analyseintervall. Mit zunehmender Analysezeit t_a bewegen sich alle Nebenmaxima und -minima auf das von $p_T(t)$ hervorgerufene Hauptmaximum zu; gleichzeitig nimmt ihr Frequenzabstand ab. Der Betrag des Spektrums bei einer Analysefrequenz $\omega \mp \omega_T$ ist somit nicht zeitlich konstant, sondern bewegt sich um eine Ruhelage.

Zur Unterdrückung der Nebenmaxima schlägt Terhardt [89] eine zeitliche Glättung des Leistungsspektrums mit einem Tiefpass 1. Ordnung und einer Zeitkonstanten von 100ms vor. Bei Glättung mit dieser Zeitkonstanten werden aber unter Umständen die Vorteile des kleinen BT-Produkts wieder zunichte gemacht. Feldtkeller [24] konnte empirisch zeigen, daß die Glättungszeitkonstante zur Vermeidung von Nebenmaxima deutlich niedriger gewählt werden kann und schlägt einen Wert von etwa 10ms bei allen Analysefrequenzen vor. Bei frequenzabhängiger Wahl der Analysebandbreite wäre es jedoch von Vorteil, wenn sich die dadurch entstehenden unterschiedlichen Einschwingzeiten durch angepaßte

Glättung auch ausnutzen ließen. Dazu ist es notwendig, das Zustandekommen der Nebenmaxima analytisch zu beschreiben.

Unter Beachtung der Näherungen bei der Bestimmung von Gl. (3.2.1) läßt sich das zeitliche Verhalten des Leistungsspektrums eines im Zeitpunkt $t = 0$ eingeschaltenen sinusförmigen Testtons $p_T(t)$ durch den transienten Anteil $F_T(\omega,t)$ beschreiben:

$$F_T(\omega,t) = 1 - 2 \cdot \cos(\omega - \omega_T)t \cdot e^{-at} + e^{-2at} \quad . \qquad (3.3.1)$$

Mit Gl. (3.3.1) läßt sich sowohl die Lage der spektralen Maxima und Minima in Abhängigkeit von der Analysefrequenz ω als auch der zeitliche Einschwingvorgang bei derselben berechnen. Die Frequenz des Einschwingvorgangs hängt von der Differenz zwischen Analysefrequenz und Testtonfrequenz ab. Fig. 3.3-4 zeigt zwei typische Verläufe. Bei der punktierten Kurve ist der Frequenzabstand um den Faktor 5 gegenüber der durchgezogenen Kurve erhöht.

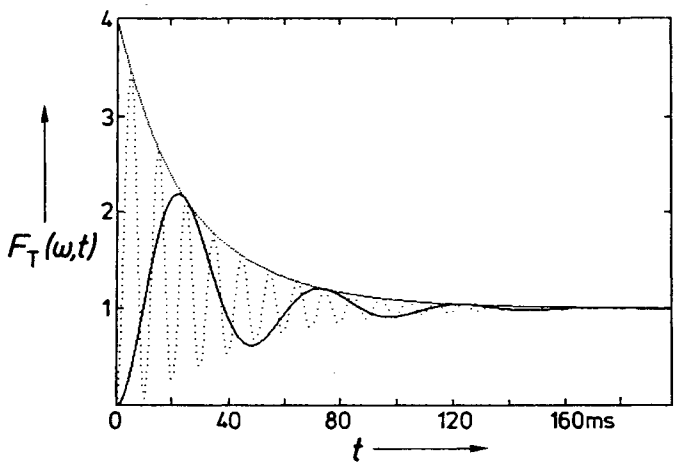


Fig. 3.3.-4: Zeitlicher Verlauf des Wechselanteils $F_T(\omega,t)$ des FTT -Leistungsspektrums. Analysebandbreite B = 10Hz. Durchgezogen: $f - f_T = 20Hz$, weitpunktiert: $f - f_T = 100Hz$. eng punktiert: $Max(F_T(\omega,t))$.

Die Ausgeprägtheit der Nebenmaxima hängt in erster Linie von der maximalen Amplitude des cos - Anteils in Gl. (3.3.1) ab. Die Zuordnung von Teiltönen zu Nebenmaxima kann verhindert werden durch

- die Erweiterung der Teiltonbestimmung um ein zusätzliches Schwellenkriterium und
- die Dämpfung des Wechselanteils in $F_T(\omega,t)$.

Als Schwellenkriterium dient die Ausgeprägtheit eines Maximums. Als Ausgeprägtheit wird der minimale Pegelunterschied L_A zwischen dem Maximum und den unmittelbar benachbarten Minima definiert. Einem Maximum wird nur dann ein Teilton zugeordnet, wenn es genügend ausgeprägt ist.

Das Prinzip der Teiltonbestimmung mit zusätzlichem Schwellenkriterium ist in Fig. 3.3-5 schematisch dargestellt. Ein relatives Maximum ist durch den Vorzeichenwechsel des Differentials $dF(\omega,t)/d\omega\big|_{t_a}$ in negativer Richtung definiert.

Ein Teilton wird dem Maximum aber nur dann zugordnet, wenn das Schwellenkriterium zu beiden Seiten des Maximums erfüllt ist. Für das Maximum 'A' in Fig. 3.3-5 ist die Bedingung erfüllt; für Maximum 'B' jedoch nicht.

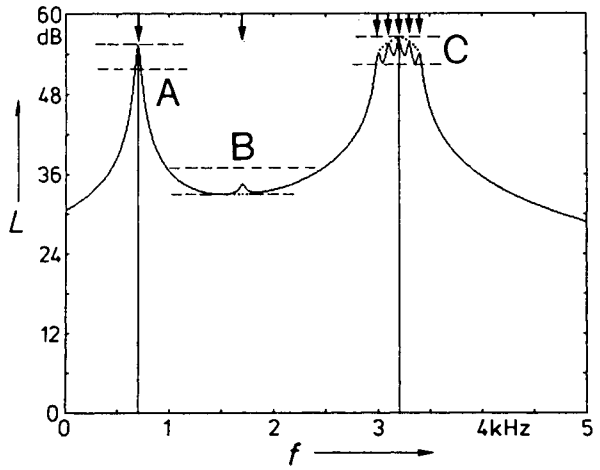


Fig. 3.3-5: Prinzip der Teiltonbestimmung. Durchgezogen: Pegelspektrum, punktiert: ohne Berücksichtigung kleiner Änderungen. Gestrichelt: Schwellenwerte, Pfeile: rel. Maxima. A, B, C: siehe Text.

Etwas komplizierter sind die Verhältnisse im Bereich des Maximums 'C'. In diesem Fall bilden mehrere schwach ausgeprägte Maxima zusammen ein globales Maximum, einen Formanten. Für keines dieser Maxima wird das Schwellenkriterium beidseitig durch die unmittelbar benachbarten Minima erfüllt. Damit solchen Verteilungen zumindest ein Teilton zugeordnet wird, werden Änderungen des Kurvenverlaufs zu größeren oder kleineren Werten ignoriert, wenn die Änderung kleiner als der Schwellenwert ist. Dadurch ergibt sich der in Fig. 3.3-5 punktiert eingezeichnete Verlauf. Dem globalen Maximum 'C' wird nun an seiner höchsten Stelle ein Teilton zugeordnet.

Eine Zuordnung von Spektraltonhöhen zu globalen Maxima wird auch beim Gehör beobachtet. Beispiele dafür sind die Tonhöhenwahrnehmungen bei Schmalbandrauschen [5], [23] oder bei höheren Vokalformanten [73].

Der größte Wert, den $F_T(\omega,t)$ annehmen kann, hängt ebenfalls vom Abstand $\Delta\omega$ der Analysefrequenz ω zur Testtonfrequenz ω_T ab:

$$\mathrm{Max}(F_T(\omega,t)) = \left(1 + e^{-a\pi/|\omega - \omega_T|}\right)^2 \quad . \qquad (3.3.2)$$

Das heißt, daß sich das Leistungsspektrum $F(\omega,t)$ bei weitabliegenden Analysefrequenzen ($\Delta\omega \gg 1/a$) mit starkem Überschwingen dem stationären Wert nähert und infolgedessen auch stark ausgeprägte Nebenmaxima entstehen (vergl. Fig. 3.2-1 und Fig. 3.3-4). In der Umgebung der Testtonfrequenz tritt jedoch kein Überschwingen auf. Die Abhängigkeit des Überschwingens bzw. der Ausgeprägtheit der Nebenmaxima vom Abstand $\Delta\omega$ der Analysefrequenz zur Testtonfrequenz legt nahe, einen Tiefpaß zur Begrenzung der Schwingungsamplitude von $F_T(\omega,t)$ und damit zur Reduzierung der Ausgeprägtheit einzusetzen. Die Zuordung

von Teiltönen zu den verbleibenden Nebenmaxima kann dann durch relativ kleine Schwellen vermieden werden.

Die Grenzfrequenz des Tiefpasses wird so gewählt, daß das Produkt aus maximal auftretendem Überschwingen $\mathrm{Max}(F_T(\omega,t))$ und Übertragungsfunktion des Tiefpasses nicht größer als der Schwellenwert wird. Dies geschieht in folgenden Schritten:

1. Wahl einer Ausgeprägtheitsschwelle ΔL_A. Sinnvolle Werte für ΔL_A sind beispielsweise $\Delta L_A = 1\mathrm{dB}$ oder $\Delta L_A = 3\mathrm{dB}$.
2. Bestimmung des Frequenzabstandes $\Delta\omega$, für den $\mathrm{Max}(F_T(\omega,t))$ den dem Schwellenwert entsprechenden Faktor bis auf den Faktor 0,7 erreicht, durch den Ansatz:

$$\left(1 + e^{-a\pi/\Delta\omega}\right)^2 = 10^{\Delta L_A/10\mathrm{dB}} \cdot \sqrt{2} \quad . \qquad (3.3.3)$$

3. Gleichsetzen der Tiefpaßgrenzfrequenz mit dem aus Gl. (3.3.3) ermittelten Frequenzabstand $\Delta\omega$ durch Wahl der Zeitkonstante

$$T_G = 1/\Delta\omega \quad . \qquad (3.3.4)$$

Die Zeitkonstante des so ermittelten Tiefpasses erster Ordnung wird im folgenden als Glättungszeitkonstante T_G bezeichnet. Ein Tiefpass erster Ordnung ist zur Begrenzung des Wechselanteils $F_T(\omega,t)$ ausreichend, da dessen Dämpfung in Abhängigkeit von $\Delta\omega$ steiler ansteigt als $\mathrm{Max}(F_T(\omega,t))$.

Für eine Ausgeprägtheitsschwelle ΔL_A=3dB erhält man mit Gl. (3.3.3) und Gl. (3.3.4) eine Glättungszeitkonstante T_G =0,12/a; für ΔL_A=1dB erhält man T_G = 0,33/a; für ΔL_A = 0dB (keine Nebenmaxima) T_G = 0,54/a.

Der mit (3.3.3) und (3.3.4) näherungsweise ermittelte Zusammenhang zwischen Glättungszeitkonstante und Transformationskonstante gilt streng genommen nur dann, wenn das Zeitsignal aus einem einzigen Sinuston besteht. Untersuchungen mit komplexen Zeitsignalen zeigten jedoch, daß mit einer Glättungszeitkonstanten T_G = 0,2/a und einem Maximumkriterium von ΔL_A= 3dB im Einschwingvorgang keine Nebenmaxima auftreten.

Das zeitlich geglättete Leistungsspektrum $G(\omega,t)$ wird mit der Formel

$$G(\omega,t) = \frac{1}{T_G} \int_{-\infty\ (0)}^{t} F(\omega,x) \cdot e^{-(t-x)/T_G}\, dx \qquad ;\ t>0 \qquad (3.3.5)$$

berechnet. Der genaue Verlauf der Glättungszeitkonstanten in Abhängigkeit von der Frequenz wird in Abschnitt 3.4.1 festgelegt.

3.4 Numerische Berechnung

3.4.1 Wahl der Transformationsparameter

Zur Bestimmung der spektralen Feinstruktur muß die Analysebandbreite klein genug sein, damit zwei benachbarte Sinustöne mit einem Frequenzabstand Δf entsprechend dem Frequenzauflösungsvermögen des Gehörs voneinander getrennt werden können. Dazu ist es notwendig, daß im Leistungsspektrum zwei genügend ausgeprägte Maxima entstehen. Wenn der Pegel zwischen beiden Maxima im eingeschwungenen Zustand 3dB unter den Maxima liegen soll, dann muß die Analysebandbreite B zu $B = \Delta f/2$ gewählt werden. Die Grenze der Wahrnehmung von zwei Tonhöhen gegenüber einer Tonhöhe bei Darbietung zweier Sinustöne verläuft bei tiefen Frequenzen bei etwa 0,2 Bark und steigt bei hohen Frequenzen auf 0,5 Bark an [78]. Für die Analysebandbreite wird deshalb ein Wert von 0,1 Bark gewählt. Die zeitlichen Verläufe der Bewertungsfunktionen nach dieser Dimensionierung bei fünf Analysefrequenzen zeigt Fig. 3.4-1.

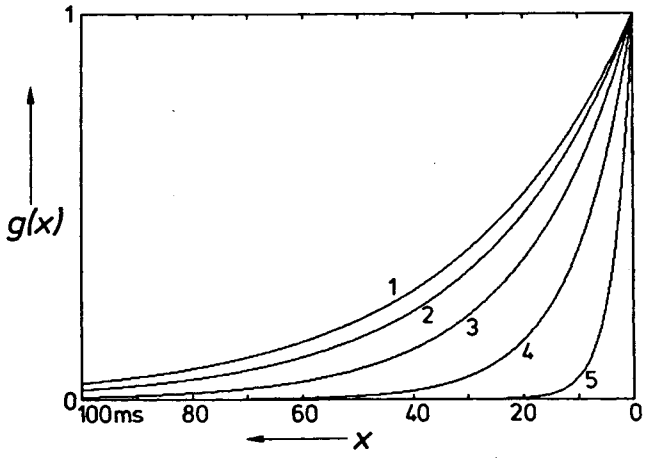


Fig. 3.4-1: Zeitlicher Verlauf der Bewertungsfunktionen in Abhängigkeit von der Analysefrequenz. Analysebandbreite B = 0,1Bark. Analysefrequenzen: 1 = 100Hz, 2 = 500Hz, 3 = 1kHz, 4 = 2kHz, 5 = 5kHz.

Aufgrund des unendlichen Analyseintervalls kann das Spektrum der FTT an jeder beliebigen Analysefrequenz bestimmt werden. Zur numerischen Berechnung erscheint es ausreichend, die spektralen Stützwerte im Abstand des eben wahrnehmbaren Frequenzunterschieds des Gehörs zu bestimmen. Nach Zwicker [96], [102] hat dieser für frequenzmodulierte Sinustöne den festen Wert von 0,04Bark.

Aus psychoakustischen Messungen über die Wahrnehmung von zeitlichen Schalländerungen ist eine minimale Grenzdauer (Zeitkonstante) von 2ms bekannt [98]. Schalle mit zeitlichen Strukturen unter 2ms können vom Gehör nicht mehr aufgelöst werden, unabhängig davon, ob sie periodisch oder statistisch sind [97]. Terhardt [78] ermittelte für die Wahrnehmungsgrenze der Rauhigkeit amplitudenmodulierter Sinustöne einen ähnlichen Wert. Diese Grenze entspricht bis 3kHz der Frequenzgruppenbreite, bleibt ab 3kHz jedoch konstant. Es bietet

sich daher an, die Grenzfrequenz des Glättungstiefpasses proportional zur Wahrnehmungsgrenze der Rauhigkeit zu wählen. Mit der oben angegebenen Dimensionierung von $T_G = 0{,}2/a$ entspricht die Grenzfrequenz einem Viertel der Frequenzgruppenbreite bis 3kHz und bleibt oberhalb dieser Frequenz konstant. Sie hat dort den Betrag $T_G = 1{,}25$ms.

3.4.2 Diskrete Fourier-t-Transformation

Liegt das Zeitsignal in zeitdiskreten Stützwerten $p(nT_S)$ vor, so kann das Spektrum $P(\omega,t)$ nach (3.1.4) rekursiv zu diskreten Zeitpunkten $t = nT_S$ berechnet werden [24], [87]:

$$P(\omega,nT_S) = P(\omega,(n-1)T_S) \cdot e^{-aT_S} + 2aT_S \cdot e^{-j\omega nT_S} \cdot p(nT_S)\ ;\ P(\omega,0)=0\ . \quad (3.4.1)$$

Mit T_S wird das Abtastintervall bezeichnet, während mit n, beginnend mit n=0, die einzelnen Abtastwerte fortlaufend durchnumeriert werden.

Da nur das normierte Leistungsspektrum $F(\omega,nT_S)$ ausgewertet wird, vereinfacht sich die Berechnung wie folgt (vergl. auch [26]).
Mit

$$Q(\omega,nT_S) = 2a \cdot P(\omega,nT_S) \cdot e^{j\omega nT_S} \quad (3.4.2)$$

erhält man aus (3.4.1):

$$Q(\omega,nT_S) = Q(\omega,(n-1) \cdot T_S) \cdot e^{(j\omega-a)T_S} + 2aT_S \cdot p(nT_S)\ . \quad (3.4.3)$$

Da $F(\omega,nT_S) = |Q(\omega,nT_S)|^2 = |2a \cdot P(\omega,nT_S)|^2$, kann zur Bestimmung von $F(\omega,nT_S)$ Gl. (3.4.3) verwendet werden, mit dem Vorteil, daß im Argument der Exponentialfunktion die Nummer n des aktuellen Zeitpunktes nicht enthalten ist.

Zur Unterscheidung von der im Prinzip kontinuierlichen Analysefrequenz ω werden die bei der realen Berechnung verwendeten Analysefrequenzen als w_i bezeichnet. Mit dem Index i soll verdeutlicht werden, daß die Bestimmung des Leistungsspektrums an mehreren Frequenzstützpunkten erfolgt.

Soweit nicht anders angegeben, werden in der vorliegenden Arbeit bei allen Berechnungen des zeitdiskreten, geglätteten Leistungsspektrums $G(w_i,nT_S)$ und bei der Bestimmung der Teiltöne die in Tabelle 3.4.1 angegebenen Parameter verwendet.

Genaue Berechnungsvorschriften für die Berechnung der Analyseparameter und die Rekursionschritte zur Berechnung des geglätteten Leistungsspektrums können dem Anhang entnommen werden.

Abtastintervall T_S	:	1/12,8kHz
Analysefrequenzbereich	:	20Hz bis 5kHz
Frequenzabstand Δz der Analysefrequenzen	:	0,05Bark
Analysebandbreite B	:	0,1Bark
Glättungszeitkonstante T_G	:	$0{,}2/\pi \cdot B$ bis 3kHz 1,25ms ab 3kHz
Ausgeprägtheitsschwelle ΔL_A	:	3dB

Tab. 3.4.1: Parameter zur Berechnung des zeitlich geglätteten Leistungsspektrums und zur Bestimmung der Teiltöne.

Das Leistungsspektrum $F(\omega,nT_S)$ läßt sich auch mit Hilfe der Fast-Fourier -Transformation (FFT) [14], [17], [53] bestimmen. Die rekursive Berechnung nach Gl. (3.4.3) weist jedoch folgende Vorteile auf.

- Damit die Glättung $G(w,nT_S)$ durchgeführt werden kann, muß das Spektrum $F(w,nT_S)$ zu jedem Stützwert n vorliegen. Bei einer FFT mit N Analysefrequenzen kann das Analyseintervall mit einer Länge von $T = 2NT_S$ Zeitsignalstützwerten also jeweils nur um einen Stützwert verschoben werden. Für jedes zu einem Zeitpunkt nT_S bestimmte Spektrum sind somit etwa $N \cdot \mathrm{ld}\, N$ komplexe Multiplikationen (ohne Betragsbildung) notwendig, bei rekursiver Berechnung jedoch nur N komplexe Multiplikationen [43]. Die Akkumulierung der Rundungs- und Diskretisierungsfehler, die bei rekursiver Berechnung auftritt [43], wird durch die exponentiell abklingende Bewertungsfunktion vermieden.

- Die rekursive Berechnung erlaubt unmittelbar beliebig verteilte Analysefrequenzen [43] mit jeweils eigener Bewertungsfunktion, welche sich bei exponentiellem Abfall besonders effektiv realisieren läßt.

- Ein im Prinzip unendliches Analyseintervall, wie bei der FTT gefordert, kann bei rekursiver Berechnung unmittelbar realisiert werden.

Eine Bestimmung des Leistungsspektrums F(w,t) ist mit Filteranordnungen, wie in [26], [66] beschrieben und im Anhang abgebildet, möglich. Die vielfach verwendete Anordnung Bandpaß - Gleichrichter - Tiefpaß ist bezüglich des stationären Verhaltens mit F(w,t) vergleichbar. Im Gegensatz zu F(w,t) lassen sich jedoch die bei Einschwingvorgängen entstehenden Nebenmaxima nicht unterdrücken (siehe z.B. Abbildungen in [16]).

4. Teiltonzeitmuster

Auf der Grundlage des im vorigen Kapitel vorgestellten Teiltonkonzepts wird das Teiltonzeitmuster definiert und seine graphische Darstellung erläutert. Weiter wird untersucht, wie mathematisch einfach beschreibbare Signale im Teiltonzeitmuster abgebildet werden. Abschließend wird ein Verfahren zur Synthese eines Zeitsignals aus dem Teiltonzeitmuster beschrieben.

4.1 Definition des Teiltonzeitmusters

Das Ensemble der Teiltöne, die zu einem bestimmten Zeitpunkt t im zeitlich geglätteten Betragsquadratspektrum G(ω,t) vorhanden sind, wird als *Teiltonmuster* TTM(t) bezeichnet:

$$TTM(t) := \{ f_1, L_1;\ldots; f_j, L_j;\ldots; f_m, L_m \} \quad . \qquad (4.1.1)$$

Mit m wird in Gl. (4.1.1) die Anzahl der Teiltöne bezeichnet; mit j=1....m werden die Teiltöne durchnumeriert.

Mit der zeitlichen Veränderung des Spektrums verändert sich auch das zugehörige Teiltonmuster. Das *Teiltonzeitmuster* TTZM(t) enthält die zeitabhängige Veränderung der Teiltonmuster bzw. der einzelnen Teiltöne bis zum Zeitpunkt t:

$$TTZM(t) := \{ TTM(0),\ldots\ldots, TTM(t) \} \qquad (4.1.2)$$

$$= \{ f_1(t), L_1(t);\ldots; f_j(t), L_j(t);\ldots; f_{m(t)}(t), L_{m(t)}(t) \}$$

Fig. 4.1-1 zeigt schematisch die Überführung eines diskreten Zeitsignals $p(nT_S)$ in das ebenfalls diskrete Teiltonzeitmuster $TTZM(nT_S)$. Zu jedem Stützwert des Zeitsignals $p(nT_S)$ wird das zugehörige frequenz- und zeitdiskrete, geglättete Leistungsspektrum $G(w_i, nT_S)$ berechnet. Die schraffierten Flächen entsprechen dem jeweiligen Pegel von $G(w_i, nT_S)$. Zu jedem Abtastzeitpunkt werden den genügend ausgeprägten Maxima Teiltöne zugeordnet, im Bild durch schwarze Quadrate gekennzeichnet. Mit dem offenen Quadrat wird beispielhaft ein ungenügend ausgeprägtes Maximum markiert. Darunter ist eine mögliche Repräsentation dieses Teiltonzeitmusters eingezeichnet. Es ist eine Matrix, entsprechend derjenigen von $G(w_i,nT_S)$, in der aber alle die Elemente zu Null gesetzt sind, denen in $G(w_i,nT_S)$ kein Teilton zugeordnet wurde. Diese Art der Repräsentation ist vorteilhaft für die Resynthese, beschrieben in Abschnitt 4.4, und die Verarbeitung von Teiltonzeitmustern, dargestellt in Kapitel 5.3.

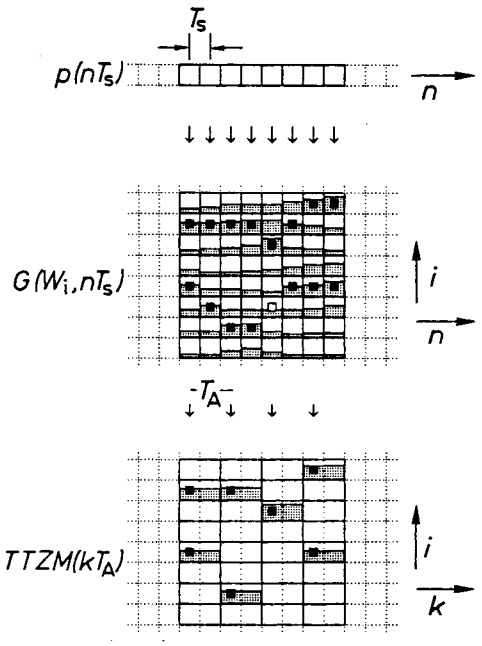


Fig. 4.1-1: Schematische Darstellung der Bestimmung des zeitdiskreten Teiltonzeitmusters $TTZM(kT_A)$ aus dem zeit- und frequenzdiskreten, zeitlich geglätteten Leistungsspektrum $G(w_i, nT)$. Schraffiert: Pegel des Leistungsspektrums; geschlossene Quadrate: genügend ausgeprägte Maxima (Teiltöne); offenes Quadrat: ungenügend ausgeprägtes Maximum. T_S : Abtastintervall; n: lfd. Nummer des Abtastzeitpunktes; w_i : Analysefrequenzen; k: lfd. Nummer des Auswertezeitpunktes; T_A : Auswerteintervall. Hier als Beispiel $T_A = 2 \cdot T_S$.

Nach Gl. (4.1.1) bzw. Gl. (4.1.2) wird jedem Zeitpunkt t (oder nT_S bei zeitdiskreten Signalen) ein Teiltonmuster zugeordnet. Die zeitliche Änderung von $G(\omega,t)$ bzw. $G(w_i,nT_S)$ ist aber durch die von Null verschiedene und endliche Transformationskonstante und die zeitliche Glättung begrenzt. Unter diesem Aspekt ist es daher möglich, den zeitlichen Verlauf des Teiltonzeitmusters innerhalb eines Zeitintervalls T_A - im folgenden als Auswerteintervall T_A bezeichnet - zu vernachlässigen. Das heißt, daß mehrere Teiltonmuster innerhalb des Auswerteintervalls T_A durch ein einziges ersetzt werden. Die Länge dieses Intervalls hängt davon ab, welche Änderungen zugelassen werden. Da die maximal möglichen zeitlichen Veränderungen des geglätteten Leistungsspektrums von T_G abhängen, reicht es in der Regel aus, die Länge des Auswerteintervalls T_A gleich der kleinsten vorkommenden Glättungszeitkonstanten zu wählen. Für den im vorigen Abschnitt angegebenen Verlauf der Glättungszeitkonstanten ergibt sich ein Auswerteintervall T_A = 1,25ms.

Die praktische Bedeutung des Auswerteintervalls besteht darin, daß die Teiltöne nur im Abstand des Auswerteintervalls bestimmt werden müssen. Das gleiche gilt für die Speicherung und Verarbeitung der Teiltöne. Ein weiterer Vorteil ergibt sich aus dem Umstand, daß das Auswerteintervall unabhängig von der Abtastfrequenz des Signals gewählt werden kann.

4.2 Darstellung des Teiltonzeitmusters

Von den Darstellungen eines zeitabhängigen Spektrums als Spektrogramm ist bekannt, daß die meisten informationstragenden Merkmale auch visuell zu

erkennen sind. Eine adäquate graphische Darstellung des Teiltonzeitmusters ist zu folgenden Zwecken notwendig:

- Überprüfung der Teiltonextraktion durch graphische Auswertung bekannter Signale;
- Visueller Vergleich mit Spektogrammen bezüglich der Repräsentation informationstragender Merkmale;
- Untersuchung der Analogie visueller und auditiver Gestaltwahrnehmung;
- Entwicklung und Überprüfung von Algorithmen zur Bearbeitung des Teiltonzeitmusters.

Zur vollständigen Abbildung des Teiltonzeitmusters sind drei Achsen notwendig: Frequenz, Pegel und Zeit. Von einer Darstellung des Teiltonzeitmusters in einer pseudo-3-dimensionalen Form ('Wasserfall'-Diagramm) wurde abgesehen, da diese bei Teiltönen sehr unübersichtlich wird. Statt dessen erfolgt die Darstellung des Teiltonzeitmusters in ähnlicher Form wie bei Spektrogrammen üblich: Die Abszisse entspricht der Zeit und die Ordinate der Frequenz. Für jeden Teilton eines Teiltonzeitmusters wird an der entsprechenden Stelle ein horizontaler Strich gezeichnet, dessen Länge gleich dem Auswerteintervall T_A ist. Diese Darstellung des Teiltonzeitmusters wird im folgendem als 'Maxigramm' bezeichnet, da sie aus dem geglätteten Leistungsspektrum durch Maximumbestimmung hervorgeht. Die bei Spektrogrammen mittels Schwärzungsgrad übliche Andeutung der Amplitude bzw. des Pegels wird durch ein symmetrisches Verbreitern des gezeichneten Punktes bzw. Striches parallel zur Frequenzachse ersetzt. In vielen Fällen ist es jedoch ausreichend, den zeitlichen Verlauf der Teiltonfrequenzen zu kennen (vergl. [34], [35]).

Bei den Maxigrammen der vorliegenden Arbeit ist die Frequenzachse entsprechend der Tonheitsskala eingeteilt. Die Umrechnung der Frequenz in die Tonheit erfolgt gemäß der Formel [100]:

$$\frac{z}{\text{Bark}} = 13 \cdot \arctan\left(0{,}76 \frac{f}{\text{kHz}}\right) + 3{,}5 \cdot \arctan\left(\frac{f}{7{,}5\,\text{kHz}}\right)^2 \quad . \qquad (4.2.1)$$

In Fig. 4.2-1 sind vier Maxigramme der Worte "Er geht" (männlicher Sprecher) mit drei verschiedenen Einteilungen der Frequenzachse bei gleichem Frequenzbereich (20Hz - 6kHz) dargestellt. Die Einteilung nach der Tonheitsskala in Bark (4.2-1c) erweckt den visuell ausgewogensten Eindruck. Der wesentliche Bereich der beiden ersten Formanten (500Hz bis 2,5 kHz) liegt etwa in der Mitte und überstreicht die Hälfte der gesamten Ordinate. Die vierte Abbildung (Fig.4.2-1d) zeigt ein Maxigramm, bei dem auf eine Umsetzung des Teiltonpegels in Strichbreite verzichtet wurde. Weitere Maxigramme von Sprache und Musik sind in Kapitel 5 zu finden.

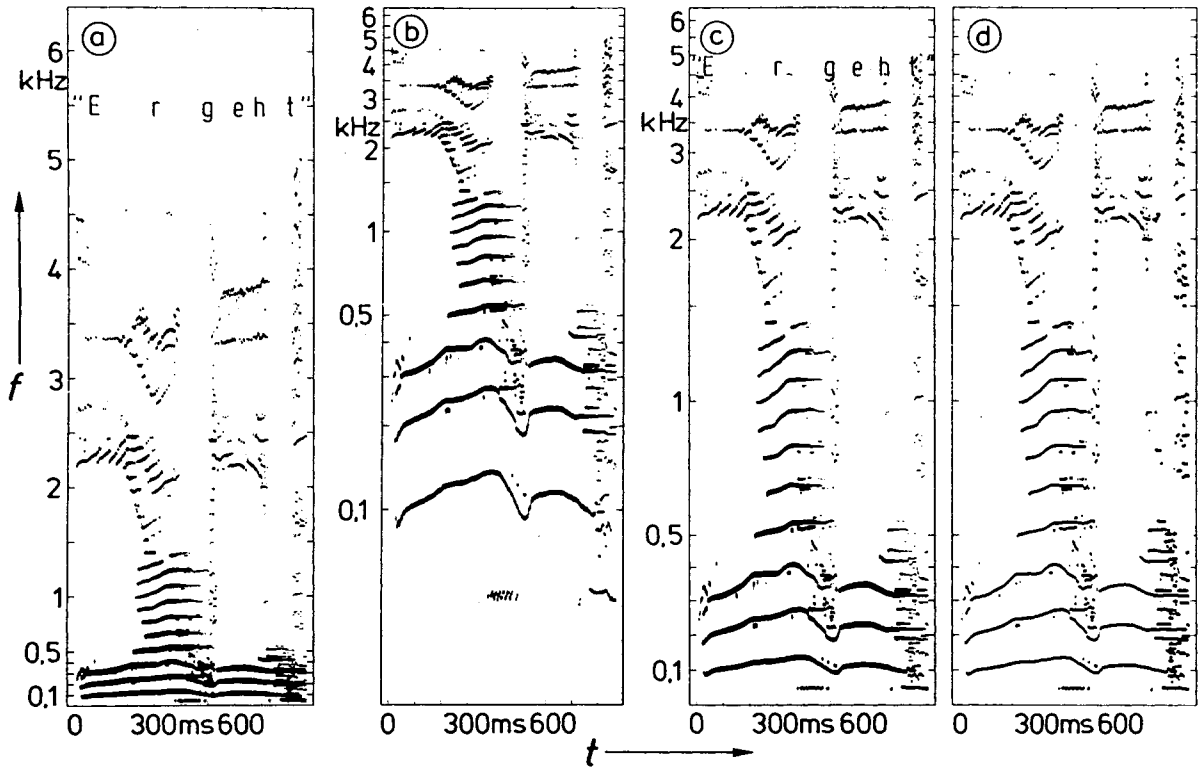


Fig. 4.2-1: Maxigramme mit verschiedenen Einteilungen der Ordinate: a) linear, b) logarithmisch, c) Tonheit (Bark). d) Ohne Abhängigkeit der Strichbreite vom Teiltonpegel. Dargestellt sind die Worte "Er geht". Auswerteintervall T_A = 1,25ms, Frequenzbereich 20Hz bis 6kHz.

4.3 Abbildung einfacher Signale im Teiltonzeitmuster

Dieser Abschnitt hat das Ziel, die Eigenschaften des Teiltonzeitmusters bei 'einfachen' Testsignalen (Sinustöne, Komplexe Töne, Modulierte Töne, also mathematisch eindeutig beschreibbar) zu demonstrieren. Dazu gehören das Verhalten von Teiltonfrequenz und - pegel bei Ein- und Ausschaltvorgängen, die Frequenzselektivität und das Verhalten bei modulierten Signalen.

4.3.1 Sinustonimpulse

Das Teiltonzeitmuster eines Sinussignals besteht, unabhängig von dessen Dauer, aus einem einzigen Teilton. Wird das Sinussignal zum Zeitpunkt Null eingeschaltet, so daß

$$p_T(t) = \begin{cases} A_T \cdot \sin(\omega_T \cdot t + \varphi_T & t > 0, \\ 0 & t \le 0 \end{cases} \quad , \qquad (4.3.1)$$

so kann jener Teilton erst dann auftreten, wenn das zugehörige Maximum von

G(ω,t) genügend ausgeprägt ist. Wegen der Frequenzabhängigkeit der Analysebandbreite schwingt das Spektrum zuerst bei Frequenzen oberhalb der des Testtons ein. Bei entsprechend großer Steigung der Bandbreite über der Frequenz tritt als erstes ein Teilton bei höheren Frequenzen auf. Mit zunehmender Zeit nähert sich die Teiltonfrequenz der Testtonfrequenz an. Ähnliches gilt für den Teiltonpegel. Dieser nimmt den Wert des Spektrums bei der Frequenz an, bei der der Teilton erstmals auftritt, und nähert sich asymptotisch dem Pegel des Testtons.

Die Messung des Einschwingverhaltens erfolgte mit Sinustönen von $f_T = \omega_T/2\pi$ = 250Hz, 1kHz, 4kHz; φ_T = 0 und $\pi/2$; $L_T = 20\log(A_T) = 50$dB. Die Analyseparameter der Messungen entsprechen, bis auf den Abstand der Analysefrequenzen mit Δw_i = 2,5Hz, denen der Tabelle 3.4.1 . Der Abstand der Analysefrequenzen wurde so klein gewählt, um den Verlauf der Teiltonfrequenz möglichst genau bestimmen zu können.

Die Verläufe von Teiltonfrequenz und -pegel der drei Testfrequenzen zeigen Fig. 4.3-1 bis 4.3-3. Abweichende Verläufe bei $\varphi_T = \pi/2$ sind punktiert eingezeichnet.

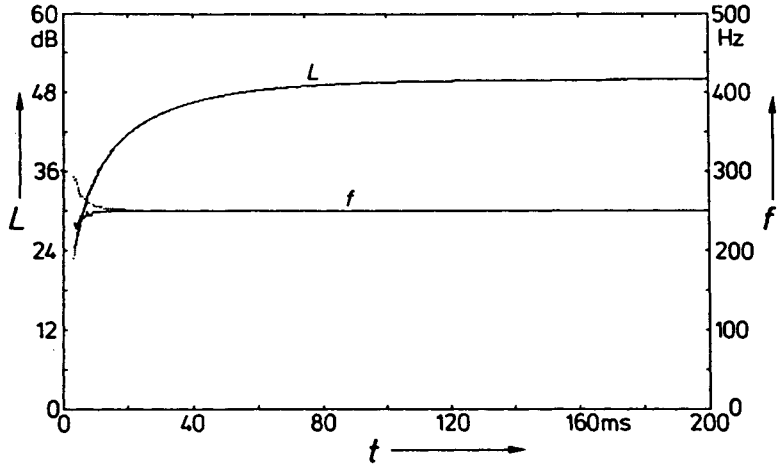


Fig. 4.3-1: Einschwingverhalten des Teiltonpegels L_1 und der Teiltonfrequenz f_1 eines sinusförmigen Testons mit f_T = 250Hz. Durchgezogen: Sinusphase, punktiert: Cosinusphase.

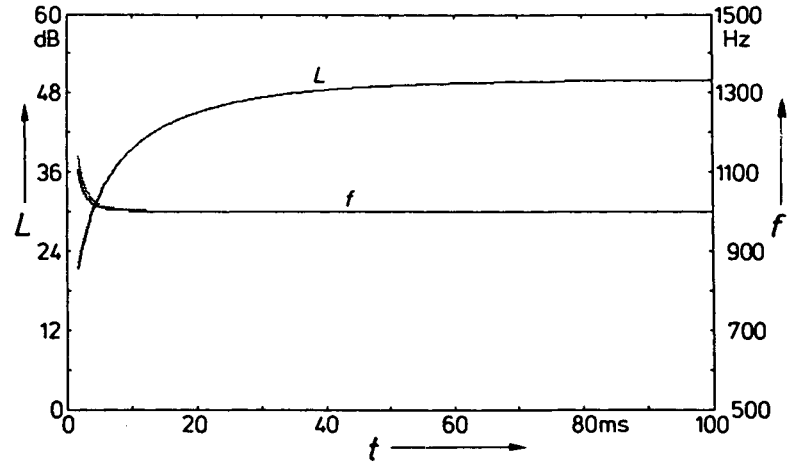


Fig. 4.3-2: wie Fig. 4.3-1, jedoch mit f_T = 1kHz. Man beachte den unterschiedlichen Zeitmaßstab.

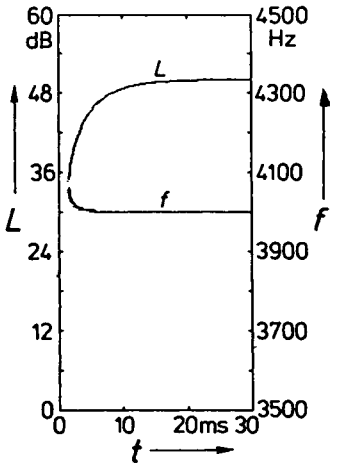


Fig. 4.3-3: f_T = 4kHz sonst wie Fig. 4.3-1.

Der Anstieg des Teiltonpegels hängt allein von der effektiven Fensterlänge T = 1/a bei der jeweiligen Teiltonfrequenz ab. Die zeitliche Glättung bewirkt bei Teiltonfrequenzen unter 6kHz eine zeitliche Verzögerung um T_G gegenüber dem ungeglätteten Spektrum an der Teiltonfrequenz, ohne jedoch die Steigung des Einschwingvorgangs abzuflachen. Nach der Zeit $t = 1/a + T_G$ hat der Teiltonpegel seinen Endwert bis auf 4dB erreicht, nach $t = 2{,}9/a + T_G$ bis auf 0,5dB.

Bei Sinustonimpulsen

$$p_T(t) = \begin{cases} 0 & ; t \le 0 \\ A_T \cdot \sin(\omega_T \cdot t + \varphi_T) & ; 0 < t \le T_P \\ 0 & ; T_P < t \end{cases} \qquad (4.3.2)$$

mit relativ großer Impulsdauer ($T_P > 3/a$) erreicht der Teilton seinen stationären Wert bevor der Sinuston abgeschaltet wird. Wie Fig. 4.3-4 bis 4.3-6 zeigen, fällt die Teiltonamplitude mit der Zeitkonstanten 1/a ab, solange 1/a größer als T_G ist. Für Teiltonfrequenzen unter etwa 6kHz kann der Einfluß der Glättungszeitkonstanten nach Tab. 3.4.1 vernachlässigt werden.

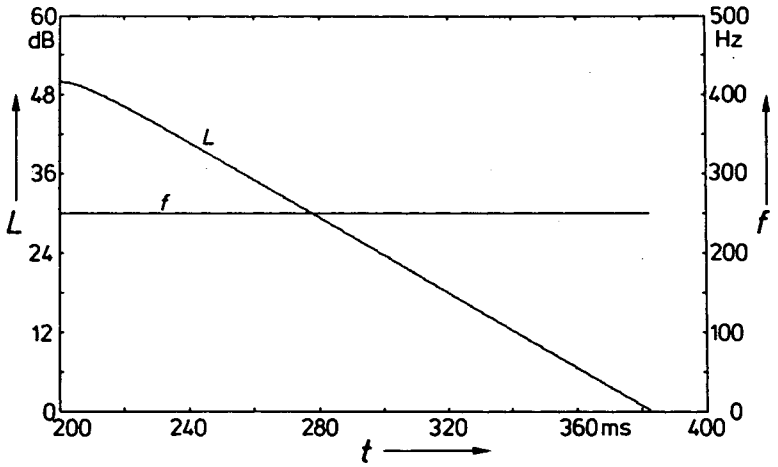


Fig. 4.3-4: Abklingverhalten des Teiltonpegels L_1 und der Teiltonfrequenz f_1 eines sinusförmigen Testons mit f_T = 250Hz, Sinus- und Cosinusphase. Impulsdauer T_P = 200ms.

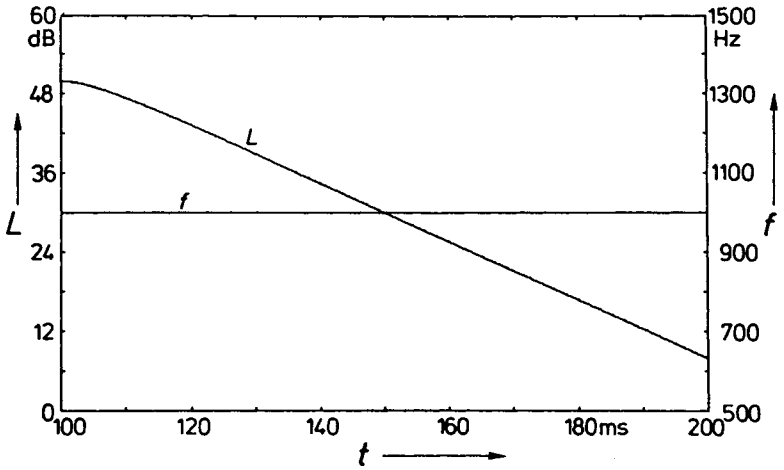


Fig. 4.3-5: wie Fig. 4.3-4, jedoch mit f_T = 1kHz, T_P = 100ms. Man beachte den unterschiedlichen Zeitmaßstab.

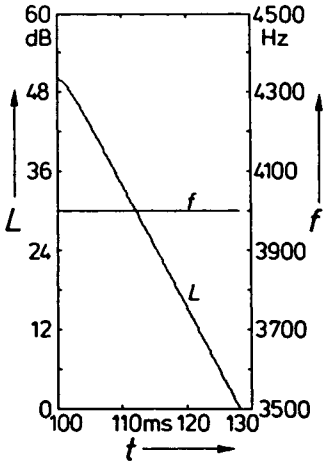


Fig. 4.3-6: f_T = 4kHz, sonst wie Fig. 4.3-5.

Der Teiltonpegel erreicht seinen Endwert nicht, wenn die Impulsdauer T_P des Sinustons kleiner als etwa 3/a ist. Der Verlauf des Einschwingvorgangs entspricht genau dem zuvor beschriebenen. Nach Ende des Sinustonimpulses fällt die Teiltonamplitude mit der Zeitkonstante 1/a ab.

Während des Abklingens des Teiltonpegels können weitere Teiltöne bei den Frequenzen $f = f_T \pm (2k-1)/t_1$ mit $k = 2,3,4...$ auftreten. Jene sind die Folge von Nebenmaxima, die aufgrund des fehlenden Eingangssignals nicht mehr "weggeglättet" werden können. Der Glättungstiefpass kann Nebenmaxima nur dann wegfiltern, wenn sie sich zeitlich 'bewegen', d.h., wenn sich das Spektrum an der Analysefrequenz zeitlich verändert (s.a. Abschnitt 3.3). Dies ist jedoch nach Abschalten des Sinustons nicht mehr der Fall. Das Auftreten von Nebenmaxima, bzw. die Zuordnung von Teiltönen zu ihnen hängt davon ab, inwieweit die Minima bei den Frequenzen $f = f_T \pm k/T_P$ mit $k = 1,2,3...$ infolge eines kleinen relativen Frequenzabstandes zu den benachbarten Maxima 'aufgefüllt' werden.

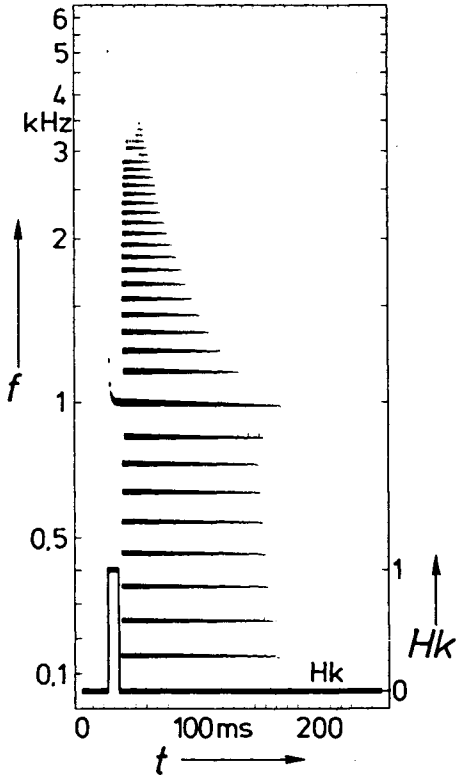


Fig. 4.3-7: Maxigramm 1kHz–Sinuston-impuls, rechteckförmige Hüllkurve (Hk), T_P = 10ms.

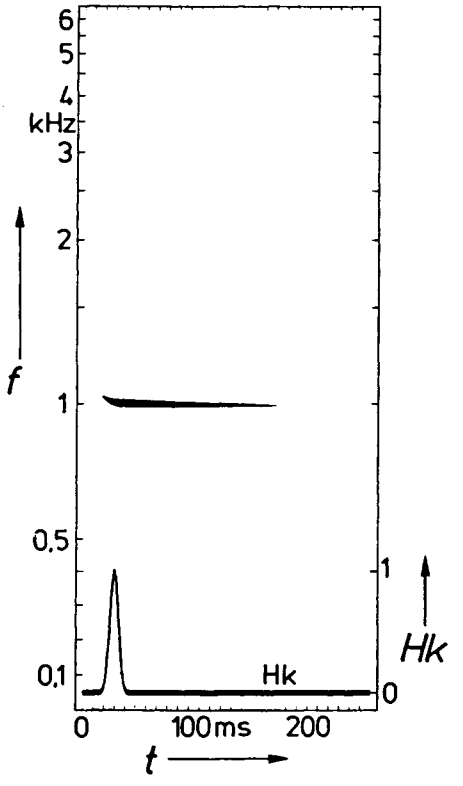


Fig. 4.3-8: Maxigramm 1kHz-Sinuston-impuls, gaussförmige Hüllkurve (Hk), T_P = 10ms.

Das Fourierspektrum eines Sinustonimpulses besitzt eine (sin x)/x - förmige Hüllkurve. Im Teiltonzeitmuster werden den spektralen Maxima dieser Hüllkurve Teiltöne zugeordnet, wenn diese weit genug auseinander liegen. Bemerkenswert

ist allerdings, daß jene 'Neben'- Teiltöne erst dann auftreten, wenn feststeht, daß es sich tatsächlich um einen Sinuston*impuls* handelt. Vorher wird dem Sinustonimpuls immer nur ein Teilton zugeordnet. In Fig. 4.3-7 ist als Beispiel das Teiltonzeitmuster eines 1- kHz-Sinustonimpulses mit 10ms Dauer dargestellt. Die 'Neben'- Teiltöne klingen mit der der jeweiligen Analysefrequenz entsprechenden Zeitkonstante T = 1/a ab.

Bei Sinustonimpulsen mit entsprechender Hüllkurve (z.B. Gaussförmig) treten keine Nebenmaxima und damit auch keine zusätzlichen Teiltöne auf, wie in Fig. 4.3-8 für einen 1kHz-Gauss-Tonimpuls mit 10ms Impulsdauer dargestellt ist.

4.3.2 Komplexe Töne

Ein komplexer Ton besteht im allgemeinen aus mindestens zwei Sinustönen verschiedener Frequenz, Phase und Amplitude:

$$p(t) = \sum_{i=1}^{n} A_i \cdot \sin(\omega_i t + \varphi_i) \qquad . \quad (4.3.3)$$

Der einfachste Fall liegt vor, wenn der komplexe Ton aus zwei Sinustönen gleichen Pegels besteht. In Abhängigkeit vom Frequenzabstand in Relation zur Analysebandbreite B treten drei verschiedene Teiltonzeitmuster auf, und zwar je nachdem, ob

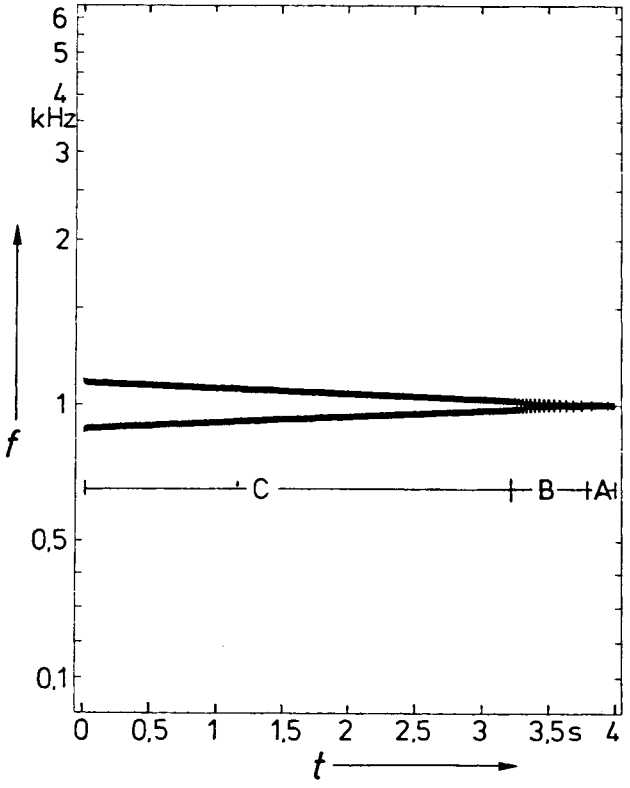


Fig. 4.3-9: Maxigramm zweier Sinustöne, deren Frequenzabstand innerhalb von 4s von 200Hz auf 2Hz abnimmt. A, B, C: siehe Text.

A) $|f_1 - f_2| \ll B$,

B) $|f_1 - f_2| \approx B$, oder

C) $|f_1 - f_2| \gg B$.

Im Fall A tritt nur ein Teilton auf, dessen Amplitude periodisch mit $T = 1/|f_1 - f_2|$ schwankt. In Fall B erscheinen in periodischen Abständen ein oder zwei Teiltöne, da das Spektrum zwischen den beiden Sinustönen mit der Differenzfrequenz schwebt. Liegen die Sinustöne weit auseinander (Fall C), so treten zwei stationäre Teiltöne auf, die sich gegenseitig nicht beeinflussen. In Fig. 4.3-9 ist das Teiltonzeitmuster zweier Sinustöne dargestellt, deren Frequenzabstand von 200Hz auf 2Hz innerhalb von 4s verringert wird. Die verschiedenen Phasen sind deutlich zu erkennen.

Die gerade noch mögliche Wahrnehmung zweier Tonhöhen bei der Darbietung zweier dicht benachbarter Sinustöne, wie sie von z.B von Terhardt [78] bestimmt wurde, stellt eine wichtige Grenze zur Beschreibung des spektralen Auflösungsvermögens des Gehörs dar. Anhand dieser Grenze wurden die Analyseparameter ausgewählt (s. Kap. 3.4). Fig. 4.3-10 zeigt, welche Grenze sich aus dem Teiltonzeitmuster ergibt, wenn das Kriterium darin besteht, daß innerhalb von 50ms nach Beginn des komplexen Tons zwei Teiltöne auftreten. Bei Verwendung dieses Kriteriums verläuft die Grenze um den Faktor 2 - 3 unterhalb der in [78] genannten Grenzkurve. Allerdings hängt die ermittelte Grenze sehr stark vom verwendeten Kriterium ab.

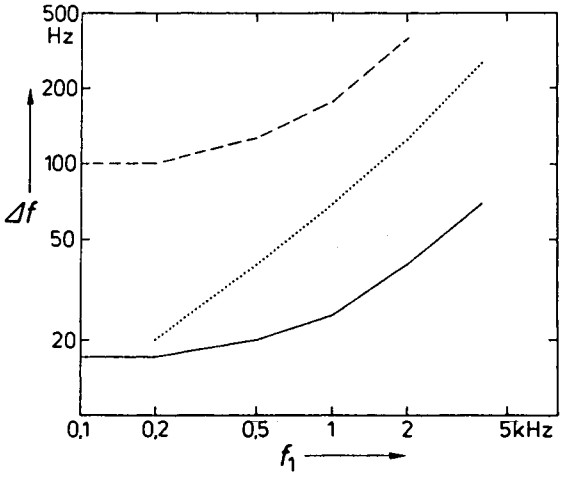


Fig. 4.3-10: Trennung zweier Sinustöne gleichen Pegels mit Frequenzabstand Δf in Abhängigkeit von der Frequenz f_1 des tieferen. Durchgezogen: Zwei Teiltöne innerhalb der ersten 50ms. Punktiert: Grenze nach [78], gestrichelt: Frequenzgruppenbreite [96].

Neben der eben genannten Wahrnehmungsgrenze wird das spektrale Auflösungsvermögen des Gehörs vor allem durch Mithörschwellen [96], [101] beschrieben. In Fig. 4.3-11 sind die berechneten Mithörschwellen eines Sinustons f_T dargestellt, mit f_T = 250Hz bzw. 1kHz in der Mitte zweier Sinustöne in Abhängigkeit vom Frequenzabstand Δf der beiden verdeckenden Sinustöne. Das Auftreten von drei Teiltönen zu einem beliebigen Zeitpunkt wurde als Kriterium gewählt.

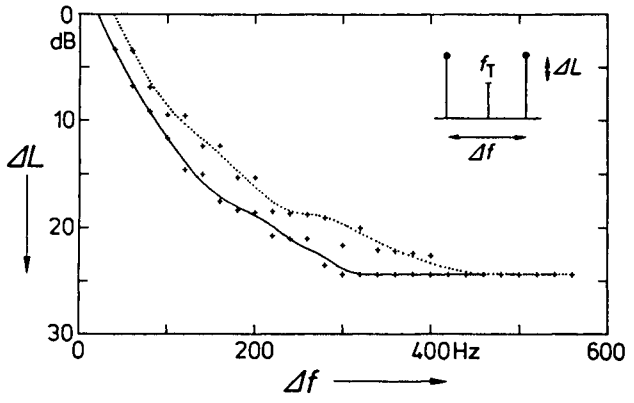


Fig. 4.3-11: Berechnete Mithörschwellen eines Sinustons in der Mitte zwischen zwei Sinustönen mit Frequenzabstand Δf. Durchgezogen: f_T = 250Hz, punktiert: f_T = 1kHz.

Der Verlauf der in Fig. 4.3.-11 dargestellten Kurven entspricht der nach Gl. (3.2.1) erwarteten Frequenzselektivität.

Für die Abbildung von komplexen Tönen, die sich aus mehreren Sinustönen entsprechend Gl. (4.3.3) zusammensetzen, gilt im wesentlichen dasselbe wie für zwei Sinustöne: Bei zu kleinem Frequenzabstand der einzelnen Komponenten treten zeitlich veränderliche Teiltonpegel oder eine veränderliche Anzahl von Teiltönen auf. Ist der Frequenzabstand der Komponenten groß im Verhältnis zur Bandbreite, so entstehen zeitlich konstante Teiltöne. Als Beispiele sind in Fig. 4.3-12 die Teiltonzeitmuster zweier komplexer Töne mit 19 Harmonischen dargestellt. In den Teilbildern a und b handelt es sich um einen harmonischen Klang mit f_b = 250Hz während in Fig. 4.3-12c und d das Teiltonzeitmuster eines unharmonischen komplexen Tons $f_i = f_b \cdot i + 3 \cdot \sqrt{f_b + i - 1}$ mit i = 1...19 und f_b = 250Hz abgebildet ist. Beide komplexe Töne wurden aus vorgegebenen Teiltonzeitmustern (Fig. 4.3-12a und c) mit der in Abschnitt 4.4 angegebenen Methode synthetisiert. Alle Komponenten der Klänge wiesen denselben Pegel auf. Die in Fig. 4.3-12b und d dargestellten Teiltonzeitmuster entstanden durch Analyse aus den synthetisierten komplexen Tönen. Bei beiden Maxigrammen (b und d) ist der Einschwingvorgang zu Beginn zu erkennen. Die Frequenz der einzelnen Teiltöne wurde bei der Teiltonbestimmung mit einer Genauigkeit von ± 0,02Bark bestimmt.

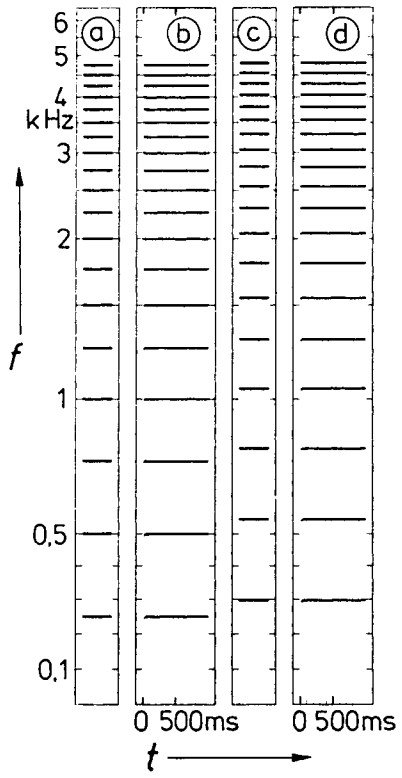


Fig. 4.3-12: Maxigramme komplexer Töne.
a) Teiltonzeitmuster zur Synthese des harmonischen Tones;
b) Teiltonzeitmuster nach Analyse des Tones aus a);
c) Teiltonzeitmuster zur Synthese des inharmonischen Tones;
d) Teiltonzeitmuster nach Analyse des Tones aus c).

4.3.3 Modulierte Signale

Weitere wichtige Kenngrößen des Gehörs sind die eben wahrnehmbaren Modulationen der Frequenz und der Amplitude von Sinustönen [96], [101]. In Abhängigkeit von der Mittenfrequenz gehen die Schwellen der eben wahrnehmbaren

Amplituden- und Frequenzmodulation ab einer bestimmten Modulationsfrequenz in einander über. Diese Frequenz wird Phasengrenzfrequenz genannt.

Im Teiltonzeitmuster tritt ein ähnlicher Effekt auf: Bei Amplitudenmodulation mit niedrigen Modulationsfrequenzen kann der Teiltonpegel den Schwankungen des Signals zeitlich folgen. Mit zunehmender Modulationsfrequenz ist dies immer weniger gut möglich. Zusätzlich treten Seitenlinien, d.h., zusätzliche Teiltöne, in periodischen Abständen auf. Bei großen Modulationsfrequenzen sind die Seitenlinien soweit von der Mittenfrequenz entfernt, daß sich drei konstante Teiltöne ausbilden.

Bei Frequenzmodulation mit kleinem Modulationsindex und geringer Modulationsfrequenz kann der Teilton den Frequenzänderungen zeitlich folgen. Mit zunehmender Modulationsfrequenz erscheinen periodisch Seitenlinien und Änderungen des Teiltonpegels. Bei hohen Modulationsfrequenzen bilden sich konstante Teiltöne aus. Im Teiltonzeitmuster lassen sich somit frequenz- und amplitudenmodulierte Sinustöne dann nicht mehr unterscheiden, wenn sie an der Modulationsschwelle zum gleichen Teiltonzeitmuster führen.

Die Phasengrenzfrequenz beim Gehör wird durch getrennte Messung der eben wahrnehmbaren Frequenz- und Amplitudenmodulation bestimmt. Eine ähnliche Messung wurde am Teiltonzeitmuster durchgeführt.

Als Schwellenwerte zur Unterscheidung einer AM oder FM vom unmodulierten Sinuston wurden Pegeländerungen des Teiltonpegels um 1dB, Frequenzänderungen um 0,05Bark und das Auftreten von mehr als einem Teilton verwendet. Die Ergebnisse, im Vergleich zu den vom Gehör bekannten Verläufen [96], sind für Trägerfrequenzen von 250Hz, 1kHz und 4kHz in den Fig. 4.3-13 bis 4.3-15 dargestellt.

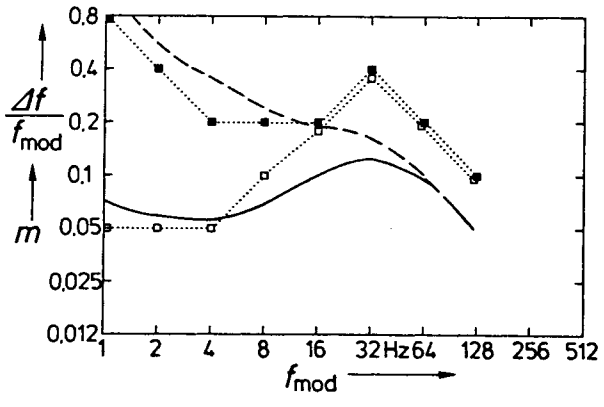


Fig. 4.3-13: Modulationsschwellen für amplituden- und frequenzmodulierten 250Hz-Sinuston. Quadrate: AM-Schwelle (offen) und FM-Schwelle (geschlossen) des TTZM. AM-Schwellen (durchgezogen) und FM-Schwellen (gestrichelt) des Gehörs nach [96].

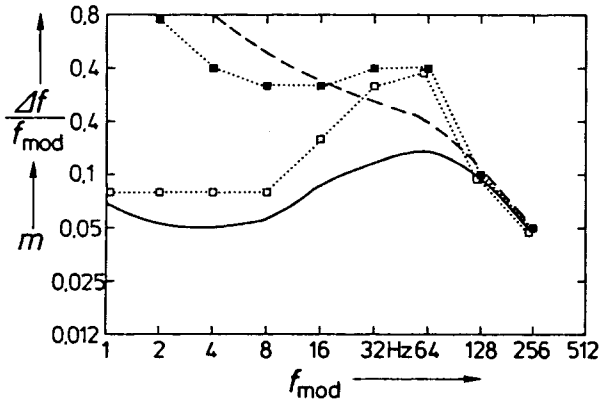


Fig. 4.3-14:
Modulationsschwellen für amplituden- und frequenzmodulierten 1kHz-Sinuston. Quadrate: AM-Schwelle (offen) und FM-Schwelle (geschlossen) des TTZM. AM-Schwellen (durchgezogen) und FM-Schwellen (gestrichelt) des Gehörs nach [96].

Die aus dem Teiltonzeitmuster ermittelten Verläufe der eben wahrnehmbaren Amplituden- und Frequenzmodulation stimmen recht gut mit den vom Gehör bekannten Daten überein. Fig. 4.3-16 zeigt den Verlauf der Phasengrenzfrequenz

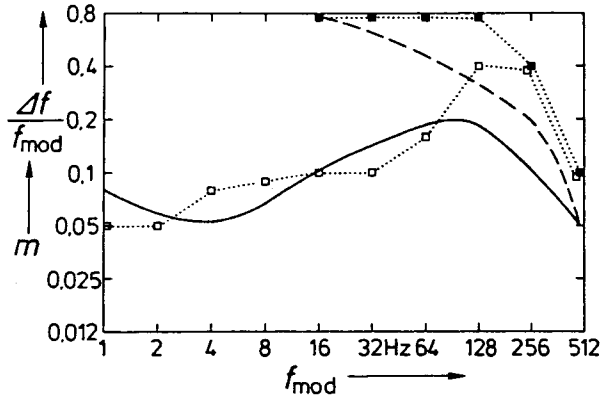


Fig. 4.3-15:
Modulationsschwellen für amplituden- und frequenzmodulierten 4kHz-Sinuston. Quadrate: AM-Schwelle (offen) und FM-Schwelle (geschlossen) des TTZM. AM-Schwellen (durchgezogen) und FM-Schwellen (gestrichelt) des Gehörs nach [96].

beim Gehör nach [96]. Die aus dem Teiltonzeitmuster ermittelten Werte sind durch Quadrate gekennzeichnet. Die Phasengrenzfrequenz des Teiltonzeitmusters verläuft etwas unterhalb der beim Gehör ermittelten; der qualitative Verlauf ist aber derselbe. Diese Ergebnisse sind insofern bemerkenswert, als sie allein durch Auswertung des Leistungsspektrums zustandekommen.

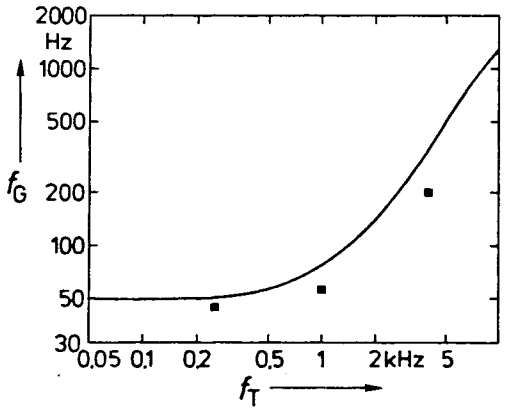


Fig. 4.3-16:
Phasengrenzfrequenz f_G des Gehörs und des Teiltonzeitmusters. Durchgezogene Kurve: Phasengrenzfrequenz des Gehörs nach [96]. Quadrate: Phasengrenzfrequenz aus Teiltonzeitmuster.

Bei Sprache und Musik treten oft schnelle Änderungen der Grundfrequenz und damit der Frequenzen der einzelnen Komponenten auf. Es wurde daher untersucht, wie gut Sinustonsweeps im Teiltonzeitmuster abgebildet werden.

Dazu wurden drei Testsignale mit linearem Frequenzanstieg, beginnend bei 100Hz zum Zeitpunkt t = 0, mit Anstiegsgeschwindigkeiten von 18kHz/s, 45kHz/s und 90kHz/s und einer oberen Grenze von 5kHz, berechnet. Ausgewertet wurden diese Testsignale im Frequenzbereich 100Hz bis 1kHz. Dieser Frequenzbereich wurde von den Testsignalen in 50ms, 20ms und 10ms durchlaufen. In den Fig. 4.3-17 bis 4.3-19 sind die Teiltonverläufe in der gleichen Weise wie in Fig. 4.3-1 bis 4.3-3 dargestellt. Die Momentanfrequenz des Signals zum Zeitpunkt t ist punktiert eingetragen.

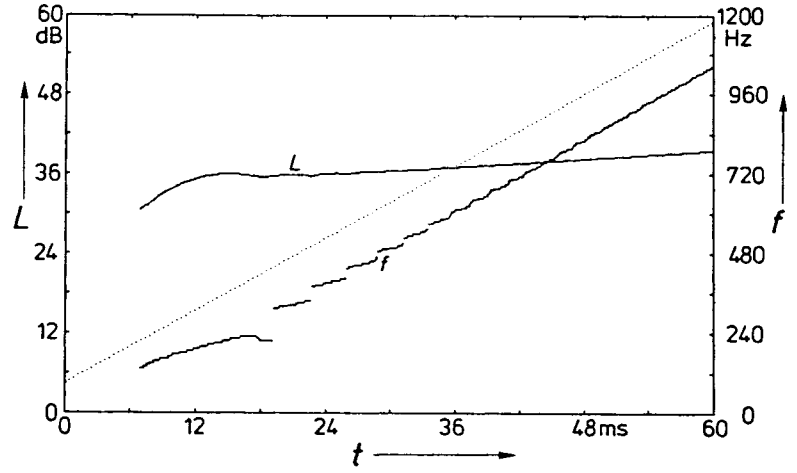


Fig. 4.3-17: Verlauf der Teiltonfrequenz und des Teiltonpegels eines Sinustonsweeps von 100Hz bis 1kHz in 50ms. Durchgezogen: Teiltonfrequenz bzw. -pegel. Punktiert: Momentanfrequenz zum Zeitpunkt t.

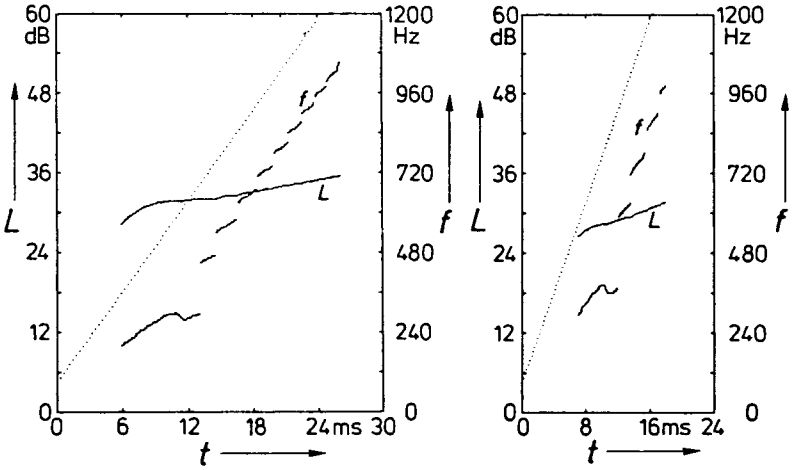


Fig. 4.3-18 (links): wie Fig. 4.3-19, jedoch mit Anstieg in 20ms.

Fig. 4.3-19 (rechts): wie Fig. 4.3-19, jedoch mit Anstieg in 10ms.

Der Zeitpunkt, zu dem der Teilton nach Beginn der Analyse auftritt, hängt von der Anstiegsgeschwindigkeit ab: 3,3ms bei 90kHz/s; 4,1ms bei 45kHz/s und 5,6ms bei 18kHz/s. Der Teilton tritt somit früher auf als bei stationären Sinussignalen. Die zeitliche Verzögerung zwischen Augenblicksfrequenz und Teiltonfrequenz hängt ebenfalls von der Anstiegsgeschwindigkeit ab: 3,4ms bei 90kHz/s; 4,9ms bei 45kHz/s und 6,4ms bei 18kHz/s.

Der Verlauf der Teiltonfrequenz zeigt besonders bei großen Anstiegsgeschwindigkeiten Sprünge. Der Grund dafür liegt im schnellen Einschwingen aber langsa-

men Abklingen des geglätteten Leistungsspektrums (vergl. Fig. 4.3-1). Dadurch bildet sich ein nach hohen Frequenz sehr steil, nach tiefen Frequenzen jedoch sehr flach abfallendes Maximum aus. Aufgrund des Schwellenkriteriums bei der Teiltonbestimmung springt die Lage des Teiltons, sobald sich die Ausgeprägtheiten entsprechend verändern. Besonders groß sind die Sprünge bei Fig. 4.3-18 und 4.3-19 an der Stelle, an der die Analysebandbreite entsprechend dem Verlauf der Frequenzgruppenbreite vom frequenzunabhängigen in den frequenzproportionalen Verlauf übergeht.

4.4 Resynthese

Anhand der in den letzten Abschnitten vorgestellten Daten läßt sich schwer einschätzen, ob die informationstragenden Merkmale eines Sprach- oder Musiksignals im Teiltonzeitmuster ausreichend repräsentiert werden. Eine Aussage darüber kann letztlich nur durch das Gehör selbst gemacht werden. Um im weiteren das Konzept der Analyse und Resynthese auf das Teiltonzeitmuster anwenden zu können, wurde ein Verfahren zur Synthese eines Zeitsignals aus dem Teiltonzeitmuster entwickelt.

Das Verfahrens basiert auf einer Überlagerung der Zeitsignale aller Teiltöne jeweils eines Teiltonmusters [33], [35]:

$$q(t) = \sum_{j=1}^{m(t)} A_j(t) \cdot \sin(\Theta_j(t)) \quad . \qquad (4.4.1)$$

Mit $A_j(t)$ wird in (4.4.1) die im allgemeinen zeitvariable Teiltonamplitude bezeichnet. Die Synthesephase $\Theta_j(t)$ hängt von der Teiltonfrequenz $f_j(t)$ und einer Korrekturphase Ω ab:

$$\Theta_j(t) = 2\pi f_j(t) \cdot t + \Omega_j \quad . \qquad (4.4.2)$$

Die Korrekturphase Ω dient dazu, Phasensprünge bei Änderungen der Teiltonfrequenzen zu vermeiden. Besonders wichtig ist die Korrekturphase, wenn das Teiltonzeitmuster aus zeitdiskreten Teiltonmustern mit Gültigkeitsbereich T_A besteht, da bei diesen auch bei stetigen Frequenzänderungen diskrete Frequenzstufen auftreten.

Ändert sich die Frequenz eines Teiltons von einem Zeitpunkt $t-\Delta t$ zum nächsten Zeitpunkt t (beispielsweise an der Grenze zweier Teiltonmuster)

innerhalb eines frequenzabhängigen Bereiches v, so wird die Korrekturphase nach der Formel

$$\Omega_j = \begin{cases} \Omega_x & ; \quad f_x(t-\Delta t) = f_j(t) \\ \Theta_x(t) - 2\pi f_j(t) \cdot t & ; \quad |f_x(t-\Delta t) - f_j(t)| \leq v \\ 0 & ; \quad |f_x(t-\Delta t) - f_j(t)| > v \end{cases} \qquad (4.4.3)$$

berechnet. Mit $\Theta_x(t)$ wird die Synthesephase des in der Frequenz nächst gelegenen Teiltons x des Teiltonmusters vor der Änderung bezeichnet, die diese zum Zeitpunkt t hätte:

$$\Theta_x(t) = 2\pi f_x(t-\Delta t) \cdot t + \Omega_x \qquad (4.4.4)$$

Nach Bedingung (4.3.3) bleibt die Korrekturphase im Fall der gleichen Teiltonfrequenz unverändert. Die Korrekturphase wird zu Null gesetzt, wenn sich im Bereich v um den neuen Teilton im vorhergehenden Teiltonmuster kein Teilton befand.

Aufgrund der frequenzabhängigen Wahl der Analysefrequenzabstände und -bandbreiten können sich Teiltonfrequenzen bei hohen Frequenzlagen in einem größeren Frequenzbereich ändern als bei tiefen Frequenzen. Es ist daher sinnvoll, den Bereich v, innerhalb dessen die Korrekturphase bestimmt wird, an die Analysefrequenzen anzupassen. Ein empirisch ermittelter Wert von $v = \pm 0{,}15$ Bark hat sich als günstig erwiesen.

Bei anderen Verfahren zur Resynthese von Zeitsignalen aus Spektren werden synthetisierte Signalabschnitte mit einem Fenster bewertet und überlappend aneinandergefügt (beispielsweise [1], [45]. Diese Methode wird beim Verfahren der Teiltonüberlagerung bewußt nicht angewendet, damit eventuelle Fehler bei der Teiltonbestimmung gut hörbar werden.

Ein nach dem oben beschriebenen Verfahren aus dem Teiltonzeitmuster resynthetisiertes Zeitsignal unterscheidet sich im allgemeinen in seinem Verlauf vom Originalsignal bezüglich zeitlicher Hüllkurve, Phasenlage und spektraler Zusammensetzung. Ob diese Unterschiede hörbar sind und ob dadurch informationstragende Merkmale verloren gehen, wird anhand von Ergebnissen der Analyse und Resynthese von natürlichen Schallsignalen im folgenden Kapitel beschrieben.

5. Sprache und Musik im Teiltonzeitmuster

Dieses Kapitel bildet den Haupteil der vorliegenden Arbeit. Mit den darin beschriebenen Untersuchungen wird der Nachweis erbracht, daß die für die akustische Kommunikation wesentliche Information tatsächlich im Teiltonzeitmuster enthalten ist. Das Kapitel ist in drei Abschnitte unterteilt. Im Abschnitt 5.1 (Sprache) wird über qualitative und quantitative Untersuchungen an Einzelvokalen sowie ein- und mehrstimmiger fließender Sprache berichtet; im Abschnitt 5.2 (Musik) über qualitative Untersuchungen an Musikausschnitten und im Abschnitt 5.3 (Audiosignalverarbeitung) über Anwendungen des Teiltonzeitmusters bei der Verarbeitung von Audiosignalen.

5.1 Sprache

5.1.1 Vokale

5.1.1.1 Ziele und Methode der Untersuchungen

Der wesentliche Parameter zur Beschreibung von Vokalen ist die Lage der Formanten. Die Formanten sind jedoch im Spektrum nicht unmittelbar vorhanden, sondern nur durch ihre Stützwerte, die harmonischen Teiltöne des Vokals repräsentiert. Aus Messungen und Modellierungen des Spektraltonhöhenmusters von Vokalen ist bekannt, daß vor allem diejenigen Teiltöne im Bereich eines Formanten vom Gehör selektiv wahrgenommen werden können. Dies soll anhand von Fig. 5.1-1 verdeutlicht werden. Die Punkte bezeichnen die Frequenzen der einzelnen Harmonischen des Vokals /a/, während die spektrale Hüllkurve gestrichelt eingezeichnet ist. Eine ähnliche Kurve stellt der Verlauf des Erregungspegels L_E dar, in dessen Bildung die Frequenzselektivität des Gehörs eingeht [96], [101]. Darunter sind die Überschußpegel LX der einzelnen Komponenten des berechneten Spektraltonhöhenmusters eingetragen, das sehr gut mit psychoakustischen Messungen übereinstimmt [74], [83]. Eine Beschreibung des Vokals ist durch jeden der dargestellten Verläufe möglich.

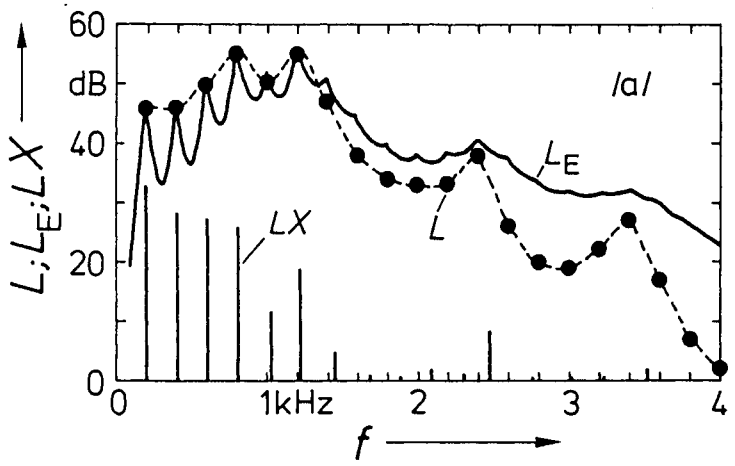


Fig. 5.1-1: verschiedene spektrale Repräsentationen des Sprachvokals /a/. Punkte: Harmonische Teiltöne des Vokals; durchgezogen: Verlauf des Erregungspegels L_E; gestrichelt: spektrale Hüllkurve entlang der Harmonischen; Linien: Überschußpegel LX (nach [73]).

Die Repräsentation von Vokalen durch das Teiltonzeitmuster wird im Hinblick auf folgende Fragestellungen untersucht:

1) Beschreibt das Teiltonzeitmuster Vokale auditiv ausreichend?
2) Welches sind die wesentlichen informationstragenden Parameter?
3) Ist die Darstellung durch das Teiltonzeitmuster mit der Beschreibung durch Formanten vergleichbar?

Zur Klärung dieser Fragen wurden zwölf Verständlichkeitsmessungen mit verschiedenen Realisationen natürlicher Einzelvokale in vier Gruppen durchgeführt:

(A) Grundverständlichkeit (Bezugswerte)
(B) Beschränkung der Teiltonanzahl
(C) Veränderung des Teiltonpegels
(D) Veränderung der Teiltonfrequenz

Es wurden die Einzelvokale /a/, /e/, /i/, /o/, /u/ von vier männlichen und vier weiblichen Sprechern verwendet, so daß insgesamt 40 verschieden gesprochene Vokalrealisationen zur Verfügung standen. Die Vokale wurden in einer schallgedämmten Meßkabine auf Tonband aufgenommen und hatten eine Dauer zwischen 300 und 700ms. Die Sprecher wurden angewiesen, die Vokale einzeln und deutlich in normaler Lautstärke zu sprechen. Besondere Anforderungen an die Artikulation wie beispielsweise eine gleichbleibende Grundfrequenz während der Aussprache wurden nicht gestellt. Zur weiteren Bearbeitung wurden die natürlichen Vokalrealisationen digitalisiert (Tiefpaßbegrenzung f_g=5,6kHz, T_s=1/12,8kHz, 12bit Quantisierung). Dabei wurden die Signalamplituden so normiert, daß alle Realisationen näherungsweise diesselbe Lautheit besaßen.

Zur Beschreibung der Hörversuche wird folgende Terminologie definiert:

- Testschall: jede Einzeldarbietung in einem Hörversuch;
- Testschallversion: Gruppe von Testschallen, die gegebenenfalls auf dieselbe Art und Weise verändert wurden;
- natürliche Vokalrealisation (NVR): Zeitsignal eines natürlichen Vokals, das bis auf die Digitalisierung nicht verändert wurde;
- resynthetisierte Vokalrealisation (RVR): Zeitsignal, das aus einem gegebenenfalls bearbeiteten Teiltonzeitmuster (gewonnen aus einer NVR) resynthetisiert wurde.

Ein Testschall kann aus einer NVR (natürlichen Vokalrealisation) oder einer RVR (resynthetisierten Vokalrealisation) bestehen. Eine Testschallversion enthält aber ausschließlich NVR oder RVR, letztere mit gleichartiger Bearbeitung. Einen Überblick über die mit verschiedenen Testschallversionen als Verständlichkeitsmessungen durchgeführten Hörversuche gibt Tabelle 5.1.1.

Gruppe	Versuch	Beschreibung der Testschallversion
A	V1	digitalisierte Originalzeitsignale (natürliche Vokalrealisation)
	V2	Resynthese aus Teiltonzeitmuster ohne Bearbeitung (RVR)
B	V3	RVR unter Verwendung der 10 Teiltöne mit größtem Pegel
	V4	RVR unter Verwendung der 5 Teiltöne mit größtem Pegel
	V5	RVR unter Verwendung der 3 Teiltöne mit größtem Pegel
	V6	RVR unter Verwendung der 5 Teiltöne mit größtem Gewicht
C	V7	RVR wie V3, Teiltonpegel eines Teiltonmusters jeweils gleich
	V8	RVR wie V6, Teiltonpegel entsprechend Formantbandbreite
D	V9	RVR wie V3, nur ein Teiltonmuster (200ms nach Beginn)
	V10	RVR wie V2, Teiltonfrequenzen inharmonisch
	V11	RVR wie V2, mit zugesetztem Rauschen
	V12	RVR wie V10, mit zugesetztem Rauschen

Tab. 5.1.1.: Übersicht über Verständlichkeitsmessungen mit Vokalrealisationen. Die Bezeichnung der Testschallversionen mit V1 bis V12 entspricht der der Messungen. RVR: resynthetisierte Vokalrealisation.

Die Berechnung der Teiltonzeitmuster und der resynthetisierten Vokalrealisationen erfolgte mit den in Tabelle 3.4.1 angegebenen Parametern, einem Auswerteintervall T_A = 5ms und einer Phasenanbindung v = 0,15Bark.

5.1.1.2 Verständlichkeitsmessungen

Aus den 40 Vokalrealisationen einer Testschallversion wurde ein Tonband mit 80 Testschallen in zufälliger Reihenfolge so zusammengestellt, daß jede Realisation zweimal dargeboten wurde. Die einzelnen Testschalle waren durch eine Pause von drei Sekunden voneinander getrennt; nach jeweils zehn Testschallen wurde ein Markierton zur Orientierung der Versuchsperson (VP) eingefügt. Die Testschalle wurden in einer schallgedämmten Meßkabine über freifeldentzerrte Kopfhörer diotisch mit Schallpegeln zwischen 60dB und 70dB dargeboten.

An den Verständlichkeitsmessungen nahmen neun normalhörende VPn im Alter zwischen 25 und 35 Jahren teil. Zwei VPn waren Sprecher von Vokalrealisationen gewesen, vier VPn hatten geringe oder keine Erfahrung in Hörversuchen. Nach Darbietung eines Testschalls mußten die VPn innerhalb der Pause von drei Sekunden diesen auf einem Protokollblatt einer der fünf Vokalkategorien /a/, /e/, /i/, /o/, /u/ zuordnen. Andere Antworten oder Enthaltungen wurden nicht zugelassen. Den VPn wurde erst nach der Messung mitgeteilt, welche Version ihnen dargeboten worden war. Zusätzlich wurden die Versuchspersonen nach ihrem Eindruck bezüglich der Wiedergabe sprechertypischer Eigenschaften und der Sprachgüte befragt.

5.1.1.3 Grundverständlichkeit (A)

Mit Messung V1 wird die Verständlichkeit der digitalisierten NVR ermittelt, um einen Bezugswert zur Beurteilung der Veränderungen bei resynthetisierten Vokalrealisationen zu erhalten. Mit Messung V2 wird die Verständlichkeit der ohne weitere Bearbeitung resynthetisierten Vokale bestimmt. In Fig. 5.1-2a ist das Maxigramm des Teiltonzeitmusters des Vokals /i/ eines männlichen Sprechers

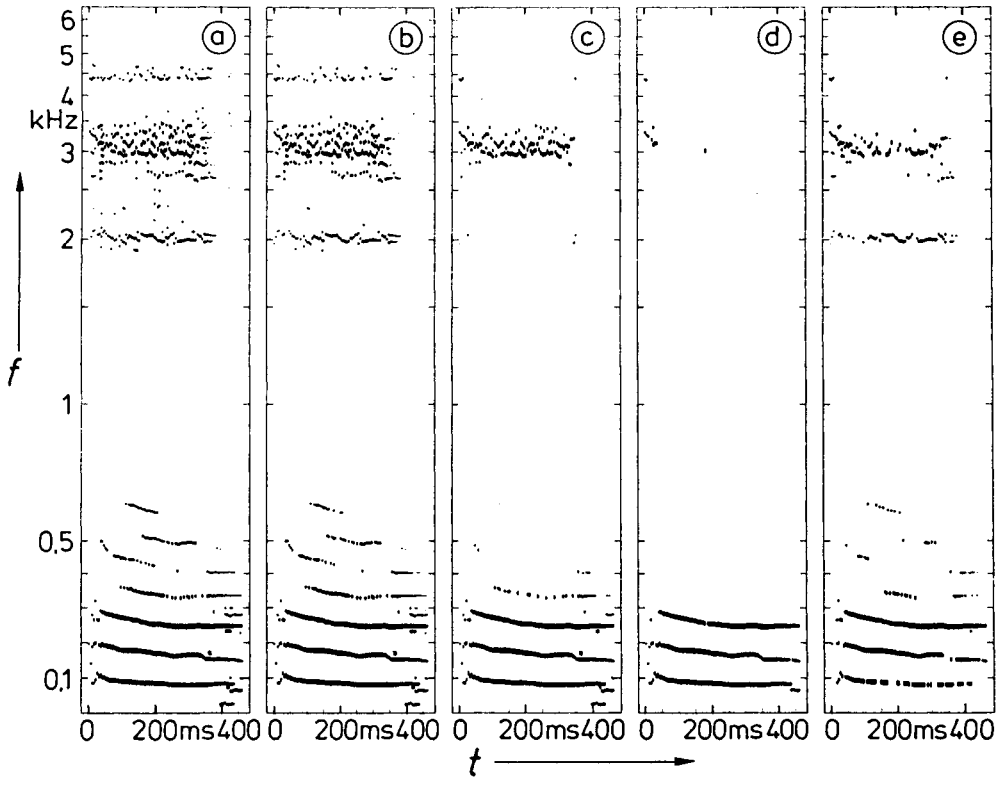


Fig. 5.1-2: Maxigramme des Vokals /i/ (Männerstimme). a) gesamtes Teiltonzeitmuster; b) begrenzt auf 10 Teiltöne, c) fünf Teiltöne; d) drei Teiltöne; e) fünf Teiltöne nach Spektralgewicht.

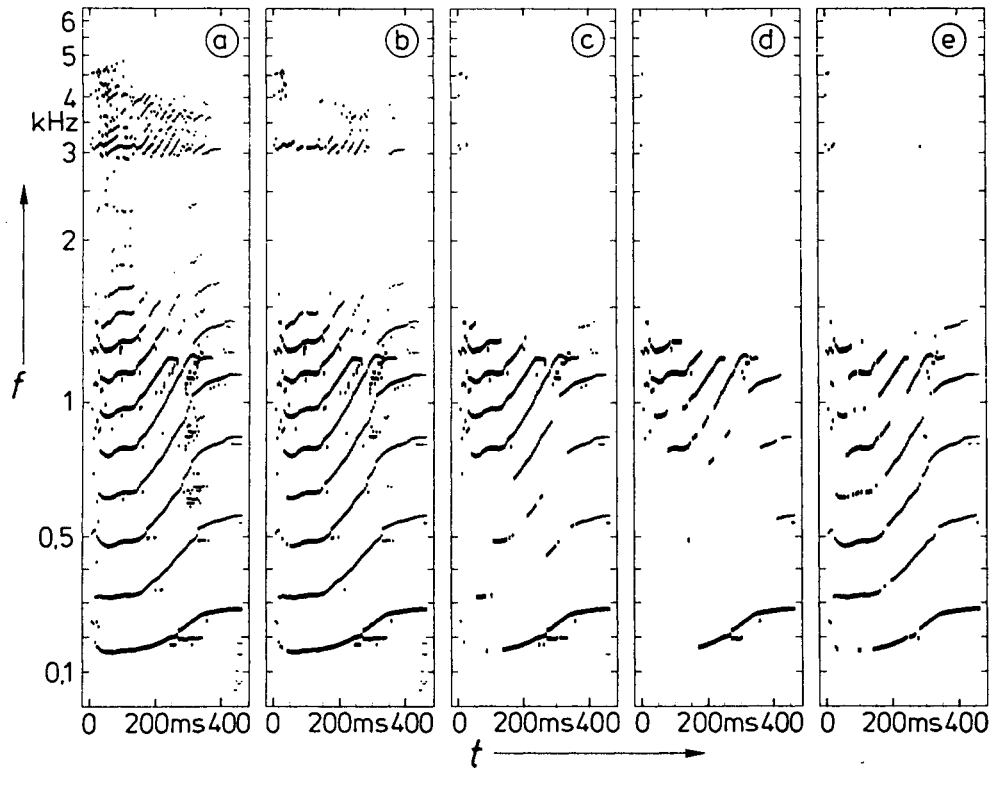


Fig. 5.1-3: Maxigramme des Vokals /a/ (Frauenstimme). a) gesamtes Teiltonzeitmuster; b) begrenzt auf 10 Teiltöne, c) fünf Teiltöne; d) drei Teiltöne; e) fünf Teiltöne nach Spektralgewicht.

und in Fig. 5.1-3a das Maxigramm des Vokals /a/ eines weiblichen Sprechers dargestellt. Die Formanten sind schon anhand der Teiltondichte deutlich zu erkennen, wobei in Fig. 5.1-3 erster und zweiter Formant verschmelzen.

Die Ergebnisse der Messungen sind in Tabelle 5.1.2 aufgeführt. Alle Angaben sind gemittelte und gerundete Fehler, d.h. die Summe der aufgetretenen Fehler dividiert durch die Anzahl der Versuchspersonen. Die linke Hälfte stellt die Verwechslungsmatrix dar. Die Zahlenwerte geben an, wie oft im Mittel einer der in der linken Spalte angegebenen Vokale mit den anderen verwechselt wurde. So wurde beispielsweise im Versuch 1 bei Darbietung eines /e/ im Mittel in 1,6 Fällen mit einem /i/ geantwortet. In der Spalte 'mittl. Fehler' ist der insgesamt bei Darbietung eines Testschalles einer Vokalkategorie aufgetretene gemittelte Fehler angegeben. Dieser Wert ist die Summe der Angaben in der entsprechenden Zeile der Verwechslungsmatrix. Die rechte Spalte gibt den Bereich der Standardabweichung um den zugehörigen mittleren Fehler an. In der obersten Zeile ist der Gesamtfehler mit Standardabweichung angegeben. Er gibt an, wieviele Fehler im Mittel bei allen Einzeldarbietungen aufgetreten sind. Bei Messung V1 also 3,2 +/- 0,8 Fehler.

Versuch 1 Digitalisierte Originalzeitsignale (NVR)							
Dargeboten	Verwechslung (gemittelt)					mittl. Fehler	+/-
	A	E	I	O	U	Σ 3,2	0,8
A						0	
E			1,6			1,6	0,6
I		0,4				0,4	0,7
O					0,2	0,2	0,3
U				1,0		1,0	0,3
Versuch 2 Resynthese aus Teiltonzeitmuster ohne Bearbeitung (RVR)							
Dargeboten	Verwechslung (gemittelt)					mittl. Fehler	+/-
	A	E	I	O	U	Σ 3,4	1,3
A						0	
E			1,3			1,3	0,9
I		0,1				0,1	0,2
O					0,6	0,6	0,7
U				1,4		1,4	1,0

Tab. 5.1.2: Ergebnisse der Versuchsgruppe A. Erläuterung im Text und auf Seite 47.

Wie die in Tabelle 5.1.2 aufgeführten Ergebnisse zeigen, besteht kein signifikanter Unterschied (95%-Niveau) zwischen dem Gesamtfehler von NVR und demjenigen von unbearbeiteten RVR. Ebenso erkärten die VPn auf Befragen, keinen Unterschied bezüglich Klangfarbe oder der Wiedergabe sprechertypischer

Eigenschaften zu bemerken. Im allgemeinen erkannten die VPn die resynthetisierte Version nur an leichten Knistergeräuschen, die bei starken Grundfrequenzänderungen auftraten (vergl. Fig. 5.1-3).

5.1.1.4 Beschränkung der Teiltonanzahl (B)

Nachdem sich unbearbeitete (d.h. hinsichtlich des TTZM unveränderte) RVR von den NVR nicht signifikant unterscheiden, wurde mit den Testschallen der Gruppe B untersucht, welchen Einfluß eine Verminderung der Teiltonanzahl hat. In den Versionen V3 bis V5 wurde deshalb die maximale Anzahl der zur Resynthese verwendeten Teiltöne schrittweise reduziert. Die Auswahl dieser Teiltöne erfolgte anhand des Teiltonpegels: Die Teiltöne der einzelnen Teiltonmuster wurden nach fallendem Pegel geordnet und nur die ersten N (N = 10, 5, 3), also diejenigen mit größtem Pegel, zur Resynthese verwendet. Die Wirkung dieser Beschränkung zeigen die Maxigramme in Fig. 5.1-2b-d und Fig. 5.1-3b-d. Mit abnehmender Anzahl der Teiltöne verschwinden zuerst die Teiltöne, die wenig zu einem Formanten beitragen (Fig. 5.1-2b gegenüber 5.1-2a bzw. Fig. 5.1-3b gegenüber 5.1-3a). Bei weiterer Reduktion verschwinden die zu schwachen Formanten gehörenden Teiltöne bei 2kHz und 4,5kHz (Fig. 5.1-2c) und schließlich bei 3kHz (Fig. 5.1-2d). Ähnlich verläuft die Reduktion beim Teiltonzeitmuster des in Fig. 5.1-3 dargestellten Vokals. Über die Reduktion der Teiltonanzahl im Hinblick auf Datenreduktion wurde in [34] berichtet.

Wie die Ergebnisse in Tabelle 5.1.3 zeigen, verändert sich der Gesamtfehler bei Begrenzung auf zehn Teiltöne gegenüber den Messungen V1 und V2 nicht. Auch hier erkannten die VPn die resynthetisierten Vokale an Störgeräuschen. Diese Störgeräusche entstehen, wenn bei geringen Pegelunterschieden zwischen Teiltönen durch die Auswahl der N Teiltöne in aufeinanderfolgenden Teiltonmustern schnelle Wechsel auftreten.

Bei Begrenzung auf fünf Teiltöne zeigt sich ein geringer, aber signifikanter Anstieg des Gesamtfehlers. Bei Begrenzung auf drei Teiltöne wird er drastisch erhöht. Dieser starke Anstieg geht mit dem Verschwinden der Teiltöne im Bereich des zweiten Formanten einher. Dies läßt sich auch der Verwechslungsmatrix entnehmen, die die typischen Verwechslungen /e/ mit /o/ und /i/ mit /u/ aufweist.

Die Anzahl von fünf Teiltöne stellt demnach hinsichtlich der Beeinträchtigung der Vokalverständlichkeit eine recht deutliche Grenze dar. Für diesen Fall wurde deshalb der Einfluß eines anderen Auswahlkriteriums untersucht. Zur Synthese der Testschallversion V6 wurden die Teiltöne anhand eines Spektralgewichtes ausgewählt. Dieses Gewicht setzt sich aus der Ausgeprägtheit L_A eines Teiltons und einer Frequenzbewertung zusammen, analog der Berechnung

Versuch 3 RVR unter Verwendung der 10 Teiltöne mit größtem Pegel								
Dargeboten	Verwechslung (gemittelt)					mittl. Fehler	+/-
	A	E	I	O	U	Σ 3,2	1,6
A						0	
E			1,4			1,4	0,6
I		0,1				0,1	0,3
O					0,4	0,4	0,6
U				1,2		1,2	0,9
Versuch 4 RVR unter Verwendung der 5 Teiltöne mit größtem Pegel							
Dargeboten	Verwechslung (gemittelt)					mittl. Fehler	+/-
	A	E	I	O	U	Σ 5,5	2,3
A						0	
E			1,9	0,1	0,1	2,1	0,3
I		0,1				0,1	0,3
O					1,2	1,2	1,3
U		0,1	0,1	1,9		2,1	1,8
Versuch 5 RVR unter Verwendung der 3 Teiltöne mit größtem Pegel							
Dargeboten	Verwechslung (gemittelt)					mittl. Fehler	+/-
	A	E	I	O	U	Σ 18,9	2,3
A					0,1	0,1	0,3
E				5,9	3,0	8,9	3,7
I					5,4	5,4	2,3
O	0,1				0,4	0,5	0,7
U		0,2	0,1	3,7		4,0	1,8
Versuch 6 RVR unter Verwend. der 5 Teiltöne mit größtem Gewicht							
Dargeboten	Verwechslung (gemittelt)					mittl. Fehler	+/-
	A	E	I	O	U	Σ 3,7	1,2
A						0	
E			2,0	0,4	0,1	2,5	0,7
I						0	
O					0,8	0,8	0,5
U	0,1			0,3		0,4	0,6

Tab. 5.1.3: Ergebnisse der Versuchsgruppe B (Beschränkung der Teiltonanzahl). Entsprechende Maxigramme in Fig. 5.1-2 und Fig. 5.1-3 sind b (Versuch 3), c (Versuch 4), d (Versuch 5) und e (Versuch 6). Der linke Teil der Tabelle enthält die Verwechslungsmatrix; die Zahlenwerte geben an, wie oft im Mittel bei Darbietung der in der Spalte 'Dargeboten' aufgeführten Vokalrealisationen mit einer der anderen Vokalkategorien geantwortet wurde. Im rechten Teil sind der mittlere Gesamtfehler (oberste Zeile) und dessen Anteile bei den einzelnen Vokalkategorien mit den Bereichen der Standardabweichung angegeben. Alle Angaben sind gemittelte Fehler.

des Spektraltonhöhengewichts im Tonhöhenberechnungsverfahren nach [85]:

$$G = \frac{1 - e^{-L_A/15\,dB}}{\sqrt{1 + 0{,}07 \cdot \left(\frac{f}{0{,}7kHz} - \frac{0{,}7kHz}{f}\right)^2}} \quad . \qquad (5.1.1)$$

Das Spektraltonhöhengewicht (Gl. 12a in [85]) setzt sich aus dem Überschußpegel LX und einer Frequenzbewertung entsprechend dem dominanten Frequenzbereich zusammen. Bei Ersetzen des Überschußpegels LX durch die Ausgeprägtheit L_A erhält man Gl. (5.1.1). Die Ausgeprägtheit L_A ist die kleinste Pegeldifferenz zwischen dem einem Teilton zugehörigen Maximum in $G(\omega,t)$ und den benachbarten Minima (vergl. Abschnitt 3.3.3). Die Maxigramme Fig. 5.1-2e und 5.1-3e zeigen die Wirkung dieses Auswahlkriteriums auf das Teiltonzeitmuster der Vokale /i/ und /a/.

Die Ergebnisse in Tabelle 5.1.3 zeigen eine deutliche Verringerung des mittleren Gesamtfehlers. Ein signifikanter Anstieg des Fehlers liegt beim Vokal /e/ vor. Dieser entstand infolge der Verwendung eines Testschalles, der bei allen Messungen häufig mit /i/ verwechselt wurde. Die Auswahl nach der Ausgeprägtheit verstärkte diese Tendenz.

5.1.1.5 Veränderung des Teiltonpegels (C)

Allein aus der Verteilung der Teiltonfrequenz, vor allem bei Beschränkung der Teiltonanzahl, kann man auf die grobe Form der spektralen Hüllkurve schließen. Mit den Testschallversionen V7 und V8 sollte näher untersucht werden, wie wichtig die Einhaltung des exakten Wertes des Teiltonpegels bei der Repräsentation eines Vokals ist.

Die Testschalle V7 wurden aus dem Teiltonzeitmuster von V3 (Beschränkung auf zehn Teiltöne) berechnet, indem alle Teiltonpegel auf den mittleren Pegel des jeweiligen Teiltonmusters gesetzt wurden. In Fig. 5.1-4a ist dies schematisch für ein Teiltonmuster dargestellt.

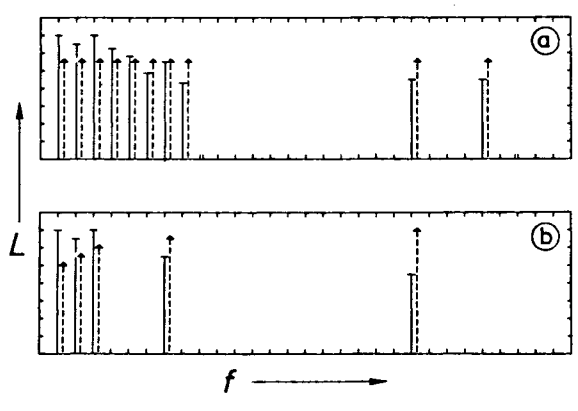


Fig. 5.1-4: Veränderungen des Teiltonpegels bei den Testschallen der Versuche V7 und V8.
a) Konstanter Teiltonpegel. Durchgezogen: Originalpegel, gestrichelt: Konstant.
b) Nach Formantbandbreite. Durchgezogen: Originalpegel, gestrichelt: bewertet.

Infolgedessen stieg der Fehler (Tab. 5.1.4) deutlich an, hauptsächlich durch Verwechslung ähnlicher Vokale. Bei 16 Vokalrealisationen trat jedoch kein Fehler auf, bei weiteren drei nur ein Fehler. Nach Aussage der Versuchspersonen blieben sprechertypische Merkmale trotz geringer Sprachgüte weitgehend erhalten.

Versuch 7 RVR wie V3, Teiltonpegel jeweils gleich								
Dargeboten	Verwechslung (gemittelt)					mittl. Fehler	+/-
	A	E	I	O	U	Σ 23	4,5
A							
E			0,9			0,9	0,9
I		6,8				6,8	2,1
O	4,8	0,2			3,1	8,1	2,6
U	1,4			5,8		7,2	1,6
Versuch 8 RVR wie V6, Teiltonpegel entsprechend Formantbandbreite							
Dargeboten	Verwechslung (gemittelt)					mittl. Fehler	+/-
	A	E	I	O	U	Σ 22,2	6,2
A				0,1		0,1	0,3
E			1,4	0,4		1,8	1,2
I		4,3		0,1	0,8	5,2	1,4
O	4,7				2,9	7,6	1,9
U	0,9			6,6		7,5	2,4

Tab. 5.1.4: Ergebnisse der Versuchsgruppe C (Veränderung des Teiltonpegels). Erläuterungen siehe Tab. 5.1.2, Seite 47.

Bei den in V7 verwendeten Teiltönen ist die Unterscheidung eng benachbarter Formanten erschwert. Mit den Testschallen V8 wurde deshalb untersucht, ob sich bei einer weiteren Beschränkung der Teiltonanzahl eine Verbesserung ergibt. Verwendet wurden die Teiltonzeitmuster von V6 (Beschränkung auf fünf Teiltöne mit größtem Spektralgewicht). Im Unterschied zu V7 erfolgte eine Bewertung entsprechend dem durchschnittlichen Verlauf der Formantbandbreiten [21]. Letztere wurden näherungsweise nach der Formel

$$B_F(f) = 0{,}05 \cdot \left(1 + 0{,}1 \cdot \left(\frac{f}{\mathrm{kHz}}\right)^3\right) \mathrm{kHz} \quad . \quad (5.1.2)$$

berechnet. Der Pegel jedes Teiltons wurde aus dem mittleren Pegel $\overline{L}$ des Teiltonmusters und einer aus Bandbreite und Teiltonfrequenz ermittelten Resonanzüberhöhung berechnet:

$$L_j = \overline{L} + 20 \cdot \log(f_j / B_F(f_j)) \quad . \quad (5.1.3)$$

Fig. 5.1-4b zeigt schematisch ein Teiltonmuster, das nach diesem Verfahren erzeugt wurde. Die Ergebnisse in Tabelle 5.1.4 zeigen, daß sich keine signifikante Änderung im Vergleich zu Versuch V7 ergibt.

5.1.1.6 Veränderung der Teiltonfrequenz (D)

Die Testschalle der Gruppe D dienten der Untersuchung der Auswirkung von kleinen Teiltonfrequenzänderungen auf die Verständlichkeit. Dazu vier Messungen.

Mit den Testschallen der Version V9 wurde untersucht, wie sich die Verständlichkeit verändert, wenn die Teiltonfrequenzen zeitlich konstant gehalten werden. Dazu wurden 400ms andauernde Vokalrealisationen aus dem Teiltonmuster 200ms nach Beginn der jeweiligen Teiltonzeitmuster von V3 resynthetisiert. Die Ergebnisse (Tabelle 5.1.5) zeigen keine signifikante Änderung der Verständlichkeit gegenüber V3.

Erhöht man die Teiltonfrequenzen der einzelnen Teiltonzeitmuster um einen konstanten Wert, so werden die Teiltöne inharmonisch. Die spektrale Hüllkurve bleibt aber bei kleinen Änderungen erhalten. Wie sich die Verständlichkeit dabei verändert, wurde mit Messung V10 ermittelt. Die Testschalle wurden aus den Teiltonzeitmustern V2 nach Erhöhen der Teiltonfrequenzen um 30Hz synthetisiert. Aus den Ergebnissen (Tab. 5.1.5) läßt sich keine signifikante Änderung der Verständlichkeit ableiten.

Bei den Testschallen V10 blieb die spektrale Hüllkurve erhalten, da die Frequenzen der Teiltöne nur um einen geringen Betrag verschoben wurden. Die spektrale Hüllkurve verändert sich jedoch durch Überlagerung eines breitbandigen Störsignals. Zur Erkennung des Vokals muß deshalb die spektrale Feinstruktur ausgewertet werden.

Mit den Verständlichkeitsmessungen V11 und V12 wurde deshalb untersucht, ob sich ein Unterschied in der Verständlichkeit von harmonischen gegenüber inharmonischen Vokalrealisationen bei gleichzeitiger Störung ergibt und ob die spektrale Feinstruktur im Teiltonzeitmuster entsprechend abgebildet wird.

Die Vokalanteile der Testschalle V11 und V12 wurden aus den Teiltonzeitmustern V2 resynthetisiert. Bei V12 wurden die Teiltonfrequenzen um 30Hz erhöht. Für jeden Testschall von V11 wurde ein Normierungsfaktor zur Normierung auf den gleichen Spitzenwert der Signalamplitude bestimmt. Als Störsignal wurde Weißes Rauschen in der gleichen Weise wie die Originalvokale digitalisiert und für alle Testschalle in der gleichen Weise verwendet. Anhand der Normierungsfaktoren wurden die Zeitsignalstützwerte der Vokalanteile von V11 und V12 bewertet, um 6dB gedämpft und zu den entsprechenden Stützwerten des Rauschens addiert. Die beiden Versionen unterschieden sich somit bezüglich ihrer spektralen Feinstruktur nur in der Lage der Teiltonfrequenzen.

Die Ergebnisse der Verständlichkeitsmessungen (Tab. 5.1.5) zeigen einen signifikanten Unterschied zwischen V11 und V12.

Versuch 9 RVR wie V3, nur ein Teiltonmuster (200ms nach Beginn)								
Dargeboten	Verwechslung (gemittelt)					mittl. Fehler	+/-
	A	E	I	O	U	Σ 3,1	1,6
A				0,1		0,1	0,3
E			1,6	0,1		1,7	0,7
I	0,1	0,2				0,3	0,4
O					0,4	0,4	0,6
U				0,6		0,6	0,7
Versuch 10 RVR wie V2, Teiltonfrequenzen inharmonisch							
Dargeboten	Verwechslung (gemittelt)					mittl.Fehler	+/-
	A	E	I	O	U	Σ 3,2	1,2
A							
E			0,7			0,7	0,9
I		0,6				0,6	0,9
O					1,0	1,0	1,3
U				0,9		0,9	1,4
Versuch 11 RVR wie V2, mit zugesetztem Rauschen							
Dargeboten	Verwechslung (gemittelt)					mittl. Fehler	+/-
	A	E	I	O	U	Σ 12	4,1
A							
E			1,3	0,4	0,3	2,0	0,8
I	0,1			0,1	2,5	2,7	2,8
O	0,5	1,8			0,9	3,2	0,7
U	0,1	1,2	2,2	0,6		4,1	2,0
Versuch 12 RVR wie V10, mit zugesetztem Rauschen							
Dargeboten	Verwechslung (gemittelt)					mittl. Fehler	+/-
	A	E	I	O	U	Σ 16,8	4,6
A							
E			1,8	0,9	0,5	3,2	2,0
I	0,1	0,5		0,1	3,0	3,7	2,8
O	0,4	2,2	0,2		1,0	3,8	1,5
U	0,1	1,4	3,3	1,3		6,1	2,2

Tab. 5.1.5: Ergebnisse der Versuchsgruppe D (Veränderung der Teiltonfrequenz)

Durch Berechnung der Teiltonzeitmuster der Testschalle V11 und V12 wurde überprüft, welche Gemeinsamkeiten und Unterschiede in den Teiltonzeitmustern zu finden sind. Als Beispiel dafür dient die Vokalrealisation lol eines männlichen Sprechers. Diese Realisation wurde in den Messungen V1 bis V6 immer richtig erkannt, weist jedoch in Messung V12 einen signifikanten Fehlerzuwachs gegenüber V11 auf.

Die Maxigramme der verschiedenen Vokalrealisationen sind in Fig. 5.1-5a bis e dargestellt. Teilbild a zeigt das Maxigramm der NVR ohne Rauschen. Die beiden nächsten Maxigramme zeigen die RVR mit Rauschen (b) und die RVR mit verschobenen Teiltonfrequenzen und Rauschen (c). Die beiden letztgenannten Maxigramme wurden aus den resynthetisierten Zeitsignalen berechnet. Durch Übereinanderlegen der Maxigramme b und c erhält man zum einen alle Teiltöne, die sowohl in b als auch in c enthalten sind (d), und zum anderen alle Teiltöne von c, die nicht mit b übereinstimmen (e). Aus den Maxigrammen ist ersichtlich, daß die Teiltöne des Rauschens im Bereich der ersten beiden Formanten des Vokals verdeckt werden und daß die Teiltonzeitmuster V11 und V12 bis auf die verschobenen Teiltöne der Vokalrealisation praktisch identisch sind.

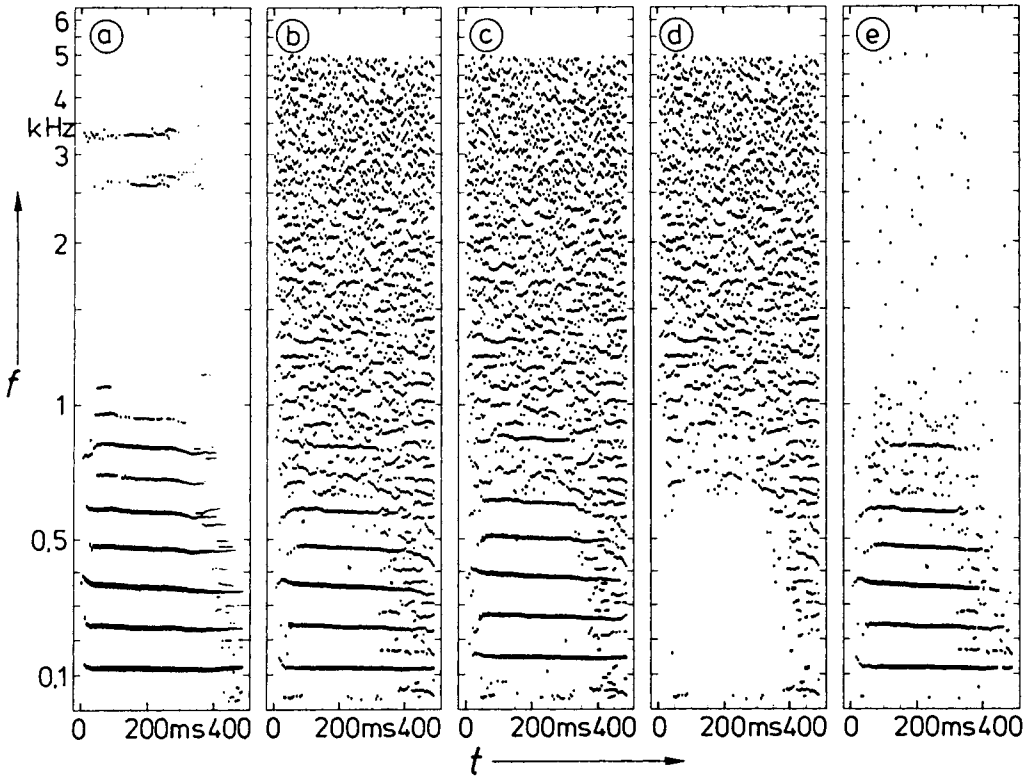


Fig. 5.1-5: Gegenüberstellung der Maxigramme einer VR lol aus den Versuchen: a) V2; b) V11; c) V12; d) Gemeinsames in b und c; e) Unterschiede zwischen b und c.

Das Zunehmen des Fehlers ist ein Hinweis darauf, daß die Erkennung bei Störung von der spektralen Feinstruktur (harmonisch gegenüber inharmonisch) abhängt. Dieses wurde auch bei anderen Untersuchungen zur Erkennung von Vokalen beobachtet [92].

5.1.1.7 Repräsentation von Formanten

Zwischen den Ergebnissen der Messung V6 (fünf Teiltöne mit größter Ausgeprägtheit) und denen der Messung V1 (natürliche Vokalrealisationen) besteht kein signifikanter Unterschied. Wird die Pegelinformation der Teiltöne durch einen mittleren (V7) oder frequenzabhängigen Wert (V8) ersetzt, so werden immer noch 70% der Vokalrealisationen erkannt. Es wurde deshalb

untersucht, inwieweit die Formanten der Vokale durch die Frequenzen der Teiltöne repräsentiert werden.

Dazu wurden die Häufigkeitsverteilungen des Auftretens von Teiltonfrequenzen in den Teiltonzeitmustern von V6 für jede der fünf Vokalkategorien bestimmt. Jede Häufigkeitsverteilung beruht somit auf den Vokalrealisationen von vier männlichen und vier weiblichen Sprechern. Aufgrund der unterschiedlichen Sprachgrundfrequenzen sind die Häufigkeitsverteilungen (Punkte in Fig. 5.1.-6a bis e) sehr unübersichtlich. Sie wurden deshalb mit einem 0,5Bark breiten Fenster geglättet (durchgezogene Kurven). Die Bereiche der beiden ersten

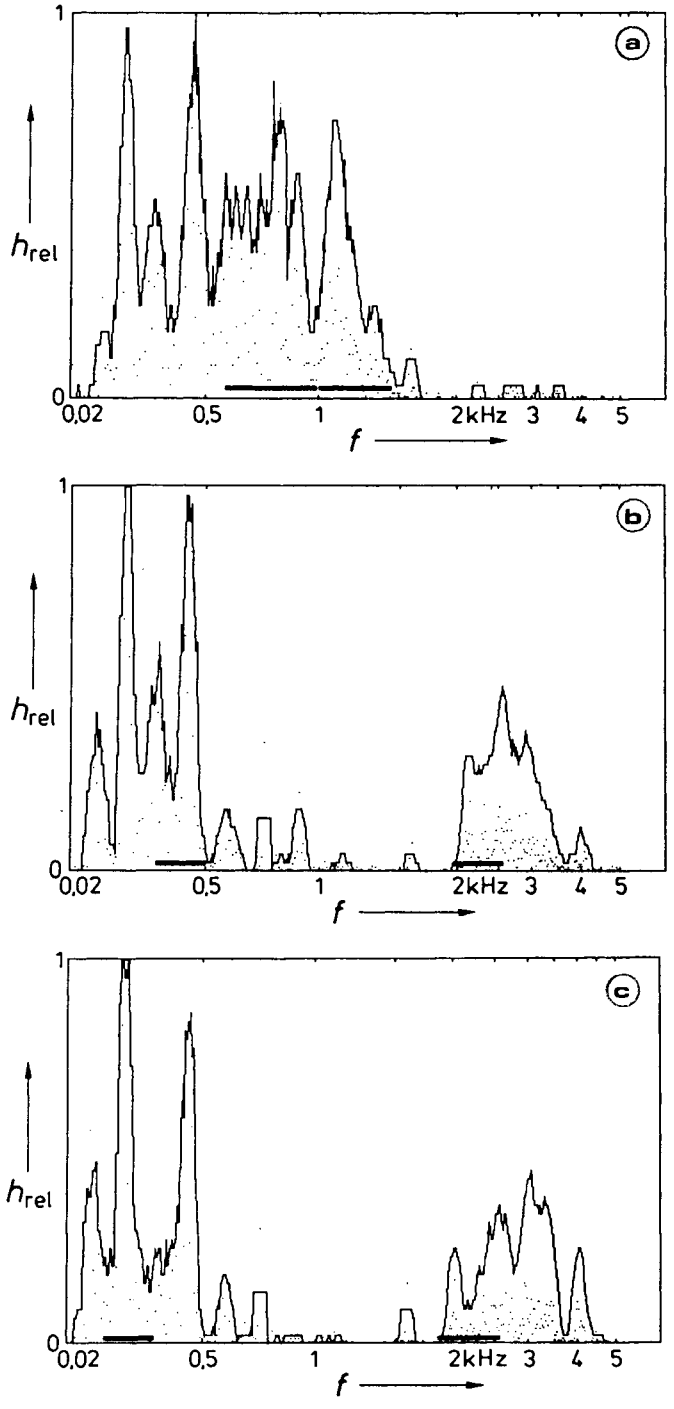


Fig. 5.1-6:
Häufigkeitsverteilungen der Frequenzen der fünf Teiltöne mit größtem Spektralgewicht (V6). Die Punkte geben die Stützwerte der Verteilungen an. Die durchgezogene Verteilungskurve wurde durch Glätten mit einem 0,5Bark breiten Fenster gewonnen. Die Frequenzachse ist entsprechend der Tonheit eingeteilt. Die Balken an der Abszisse geben die Formantlagen langer Einzelvokale an (nach [36]).

a) Vokal /a/;
b) Vokal /e/;
c) Vokal /i/;

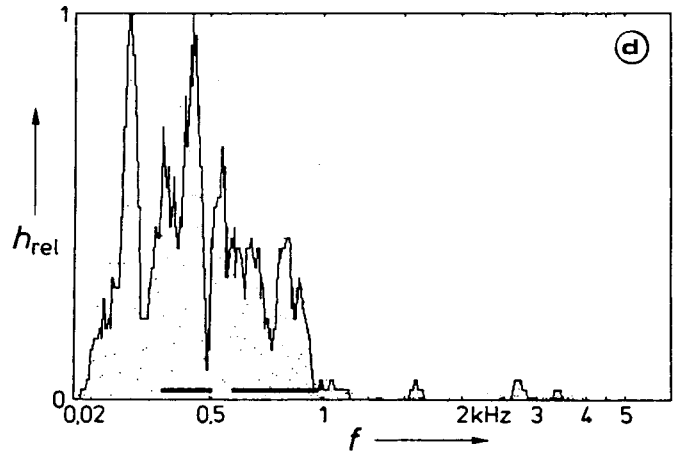


Fig. 5.1-6d:
Vokal /o/;
Erläuterung:
siehe vorherige Seite.

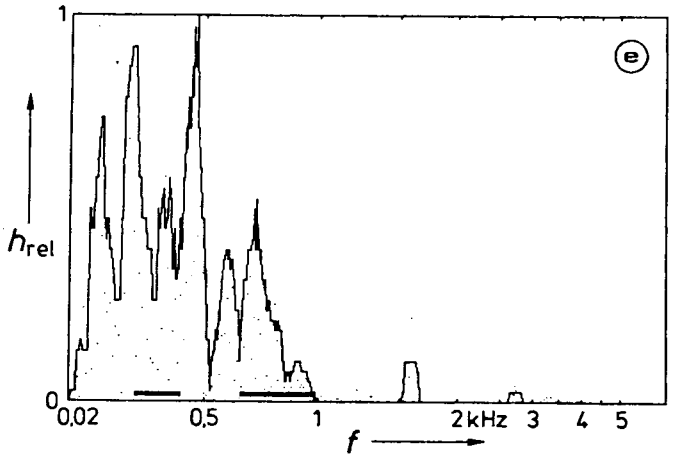


Fig. 5.1-6e:
Vokal /u/;
Erläuterung:
siehe vorherige Seite.

Formanten langer Vokale der deutschen Sprache (nach [36]) sind als schwarze Balken eingezeichnet. Die Bereiche des gehäuften Auftretens von Teiltonfrequenzen stimmen sehr gut mit den Formantbereichen überein. Dies trifft ebenfalls auf die Lücken zwischen den Formantbereichen zu, in denen praktisch keiner der gewichtigsten (d.h. größte Ausgeprägtheit besitzenden) Teiltöne auftritt.

5.1.1.8 Diskussion der Ergebnisse

In Fig. 5.1-7 sind die Gesamtfehler der Verständlichkeitsmessungen V1-V12 als Zentralwerte (Karos) mit den Bereichen der Wahrscheinlichen Schwankung dargestellt. Darüber hinaus sind die Korrelationen der Fehlerverteilungen der Messungen zu denjenigen von V1 (geschlossene Quadrate) und V2 (offene Quadrate) aufgetragen. Als Fehlerverteilung wird die Verteilung der insgesamt aufgetretenen Fehler (d.h. von allen VP) über die einzelnen Vokalrealisationen einer Testschallversion bezeichnet. Eine hohe Korrelation bei unterschiedlichem Gesamtfehler zweier Messungen ist gleichbedeutend mit einer gleichartigen Zu- oder Abnahme des Fehlers über allen Vokalrealisationen. Eine geringe Korrelation bedeutet eine ungleichmäßige Veränderung der Fehlerverteilung. Die Korrelation zwischen zwei Fehlerverteilungen ist somit ein Maß dafür, ob sich eine Bearbeitung gleichmäßig auf die Verständlichkeit von Vokalrealisationen auswirkt oder nicht.

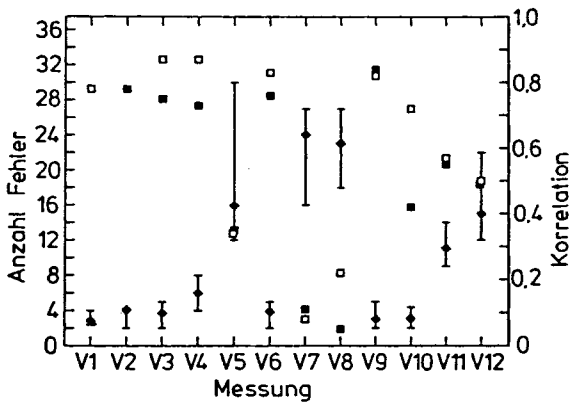


Fig. 5.1-7: Gesamtergebnis der Verständlichkeitsmessungen. Karos: Zentralwerte des Gesamtfehlers mit Wahrscheinlichen Schwankungen. Quadrate: Korrelationen zu V1 (geschlossen) und V2 (offen)

Aus Fig. 5.1-7 ist ersichtlich, daß sich die Gesamtfehler der Messungen V1 (natürliche VR) und V2 (resynthetisierte VR) nicht signifikant unterscheiden. Der Korrelationskoeffizient von 0,78 (95%-Vertrauensintervall 0,6-0,87) zwischen den beiden Fehlerverteilungen erreicht denselben Wert wie die Korrelationskoeffizienten der Fehlerverteilungen von V1 untereinander. Aus den Werten des Gesamtfehlers und der Korrelation geht hervor, daß isolierte Vokale bezüglich der Verständlichkeit durch ihr unbearbeitetes Teiltonzeitmuster vollkommen ausreichend repräsentiert werden. Sogar für die Wiedergabe von sprechertypischen Merkmalen und die Sprachgüte trifft dies nach Aussage der VPn mit sehr guter Näherung zu. Dies ist insofern bemerkenswert, als im Teiltonzeitmuster keine Information über die Beziehungen zwischen den Teiltonphasen vorhanden ist und bei der Berechnung des Teiltonzeitmusters die einzelnen Teiltöne unabhängig voneinander bestimmt werden (im Gegensatz zu einer grundfrequenzsynchronen Analyse wie z.B [31], [36]).

Bei Beschränkung auf die fünf bis zehn Teiltöne mit größtem Pegel bzw. größtem Spektralgewicht lassen sich keine oder nur geringe, d.h. gerade noch signifikante Unterschiede in der Verständlichkeit bzw. Korrelation feststellen, obwohl die Sprachgüte nach Aussage der Vpn zum Teil deutlich abnimmt. Es kann daher angenommen werden, daß die verbleibenden Teiltöne die wesentliche Information über den Vokal enthalten. Diese Teiltöne treten fast ausschließlich im Bereich der aus der Literatur bekannten Formantgebieten auf. Die Ergebnisse bei Veränderung des Teiltonpegels (V7 und V8) zeigen trotz ihres hohen Gesamtfehlers, daß etwa 70% der Vokalinformationen allein durch die Frequenzen der Teiltöne repräsentiert werden.

Den Einfluß der Teiltonfrequenz auf die Wahrnehmung von Vokalen zeigen die Ergebnisse der Messungen V10 bis V12. Während RVR mit inharmonischen Teiltonfrequenzen ohne signifikante Änderung des Gesamtfehlers gegenüber V1 oder V2 erkannt werden, zeigt sich eine signifikante Erhöhung des Gesamtfehlers bei Zusatz von Rauschen gegenüber harmonischen Vokalrealisationen, obwohl die spektrale Feinstruktur bis auf die Frequenzverschiebung der Vokalanteile die gleiche ist.

5.1.2 Natürliche Sprache

Da das Spektrum von Vokalen aufgrund ihres quasiperiodischen Zeitsignals praktisch aus harmonischen Teiltönen besteht, überrascht es nicht, daß sie im Teiltonzeitmuster (im folgenden als TTZM abgekürzt) so gut repräsentiert werden. In diesem Abschnitt sollen deshalb die Eigenschaften des Teiltonzeitmusters bezüglich der Repräsentation von natürlicher, fließender Sprache untersucht werden. Im ersten Abschnitt (5.1.2.1) werden Beispiele für die Abbildung von Einzelstimmen vorgestellt und danach die Repräsentation von Stimmengemischen (5.1.2.2) erläutert. Die Ergebnisse eines Reimtests zur Ermittlung der Verständlichkeit resynthetisierter Einzelworte werden in Abschnitt 5.1.2.3 vorgestellt.

5.1.2.1 Einzelstimmen

Zur Demonstration der Abbildung natürlicher, fließender Sprache wurde bis auf eine Ausnahme der Satz 'Und nun das Wetter in Bayern bis morgen früh' verwendet, im folgenden als Testsatz bezeichnet. Dieser Satz wurde in mehreren Varianten aufgenommen und digitalisiert (Rundfunknachrichten, männlicher Sprecher; Aufnahme bei Wiedergabe im Raum; Männerstimme geflüstert; Frauenstimme), sowie das TTZM mit den in 5.1.1.1 angegebenen Parametern berechnet.

Um zu überprüfen, ob die informationstragenden Merkmale in den jeweiligen TTZM noch enthalten sind, wurde aus diesen wieder ein Audiosignal synthetisiert und auditiv beurteilt. Bis auf eine leichte Halligkeit waren keine Unterschiede wahrzunehmen. Bei Resynthese unter Verwendung der zehn Teiltöne mit dem größten Pegel blieb die Sprachinformation ebenfalls erhalten; die Güte verschlechterte sich etwas infolge von Klangfarbenänderungen und leichten Störgeräuschen.

Fig. 5.1-8a zeigt das TTZM einer Männerstimme (Rundfunksprecher), Fig. 5.1-8b das Spektrogramm dieses Schalles. Auf den ersten Blick wirkt das Spektrogramm übersichtlicher. Das TTZM enthält jedoch wesentlich mehr Information, wie im folgenden erläutert wird.

VOKALE lassen sich im TTZM durch ihren regelmäßigen Aufbau und den synchronen Zeitverlauf der zugehörigen Teiltöne bei Grundfrequenzänderungen leicht erkennen (vergl. Fig. 5.1-2 oder 5.1-3, S. 44). Diese Darstellung entspricht bis auf die verdeckten Harmonischen dem tatsächlichen physikalischen Aufbau der Vokale. Im Spektrogramm sind die Vokale entweder an der größeren Intensität im Vergleich zu den Konsonanten oder an der zeitlichen Mikrostruktur, hervorgerufen durch die Glottisimpulse, zu erkennen. Beispiele für diese Unterschiede sind /i/ und /s/ in *'bis'* oder /ü/ in *'früh'*. Die Teiltöne des TTZM mit größtem Pegel entsprechen den im Spektrogramm sichtbaren Formanten.

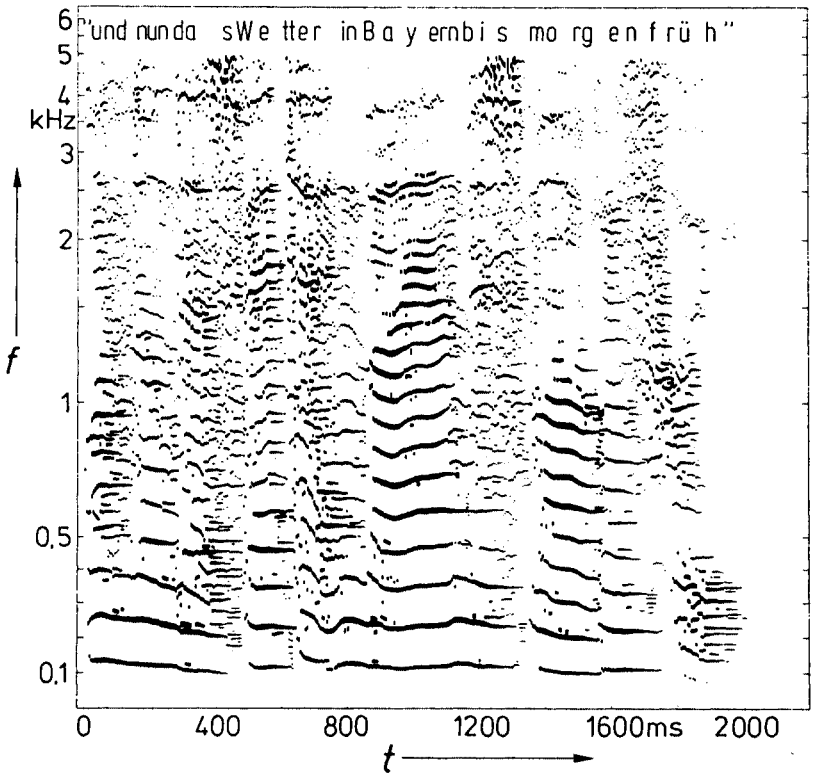


Fig. 5.1–8a: Maxigramm des Testsatzes "Und nun das Wetter in Bayern bis morgen früh" (männlicher Sprecher):

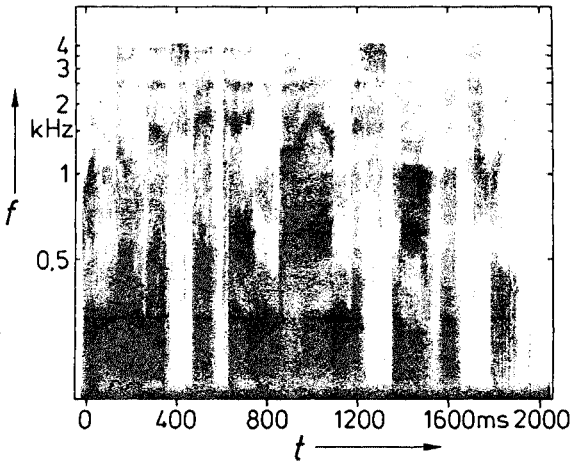


Fig. 5.1-8b: Spektrogramm desselben Signals wie im Teilbild a.

FRIKATIVE oder stimmlose Laute im allgemeinen sind im TTZM an den unregelmäßigen Frequenzverläufen und der hohen Teiltondichte erkennbar. Auch hier entsprechen die Teiltöne mit größtem Pegel den Formanten. Im Spektrogramm sind die Frikative an der 'verwaschenen' Struktur und dem Fehlen von Energie bei tiefen Frequenzen erkennbar.

LIQUIDE wie das |r| in *'Bayern'* oder in *'früh'* sind im Spektrogramm nur sehr schwer zu erkennen. Im TTZM sind Liquide an dicht benachbarten Teiltönen zu erkennen, die neben den Teiltönen, hervorgerufen durch die Grundfrequenz, auftreten. Dies ist die Folge einer Art Amplitudenmodulation der Glottisschwingung bzw. von unregelmäßigen Abständen der Glottisimpulse. Ein ähnlicher Effekt tritt immer dann auf, wenn vor Verschlußlauten die Glottisschwingung nicht abrupt aufhört, sondern noch einige Impulse folgen. Beispiele sind der Übergang 'n -> d' in 'und', 'e -> t' in *'Wetter'* und 'r -> g' in *'morgen'*.

PLOSIVE weisen wie Frikative ein unregelmäßiges TTZM auf, dem jedoch eine Pause vorangeht.

PAUSEN sind zur Wahrnehmung, Erkennung und Unterscheidung von Plosiven und Frikativen besonders wichtig (z.B. [26]). Im Spektrogramm lassen sich Pausen nur dann erkennen, wenn das Hintergrundrauschen nicht zu einer Schwärzung führt, der Rauschabstand also größer als der darstellbare Dynamikbereich ist. So ist die kurze Pause zwischen /u/ und /d/ in *'und'* praktisch nicht erkennbar. Im TTZM lassen sich dagegen Pausen an mehreren Merkmalen erkennen:

- 'Auflaufen' des TTZMs auf dem Hintergrundrauschen, sichtbar an einer deutlichen Erhöhung der Teiltondichte mit fast parallelen Teiltonverläufen, die plötzlich verschwinden (Pause zwischen *'bis'* und *'morgen'*; für das /m/ muß der Mund geschlossen werden);
- Unterbrechung von Teiltonverläufen besonders im mittleren Frequenzbereich (Pause zwischen /n/ und /d/ in *'und'*);
- Versatz von Teiltonfrequenzen infolge einer geänderten Grundfrequenz (Pause *'We-tter'*).

Besonders die letzten beiden Merkmale ermöglichen die Erkennung einer Pause auch bei kleinem Rauschabstand.

Fig. 5.1-9 zeigt das Maxigramm (a) bzw. das Spektrogramm (b) des selben Sprachsignals wie in Fig. 5.1-8 nach Wiedergabe und gleichzeitiger Aufnahme in einem Raum. Der Raum (Laborraum, keine Teppiche etc), hatte eine Größe von etwa 4x4x3m; das Mikrofon des Aufnahmegeräts war etwa 1,5m vom Wiedergabegerät entfernt.

Durch Vergleich der Spektrogramme Fig. 5.1-8b und Fig. 5.1-9b wird das starke Verschleifen der Zeitstruktur des Signals durch den Raumhall deutlich. Selbst deutliche Pausen wie in *'We-tter'* sind nicht mehr zu erkennen. Die zeitliche Strukturierung des Spektrogramms wird vor allem von Echos bestimmt, die sich in den auffälligen vertikalen Streifenmustern wiederspiegeln. Die Lage der Formanten bleibt erhalten, sie werden aber infolge des Raumhalls verlängert. Dies kann dazu führen, daß keine Unterscheidung zwischen stimmhaften und stimmlosen Lauten möglich ist (z.B. /f/ in *'früh'*). Im Spektrogramm ist nicht zu erkennen, welche Anteile des Signals vom Direktschall und welche vom Raumhall stammen.

Das TTZM in Fig. 5.1-9a hat sich gegenüber dem in Fig. 5.1-8a ebenfalls deutlich verändert. Die zeitliche Struktur ist insofern verschieden, als zusätzliche Teiltöne durch den Raumhall entstehen oder bestehende Teiltöne verlängert werden. Ein deutliches Beispiel für ein Echo ist durch einen Pfeil gekennzeichnet. Dieser Teilton entspricht einem um 70ms verzögerten Anteil des Signals.

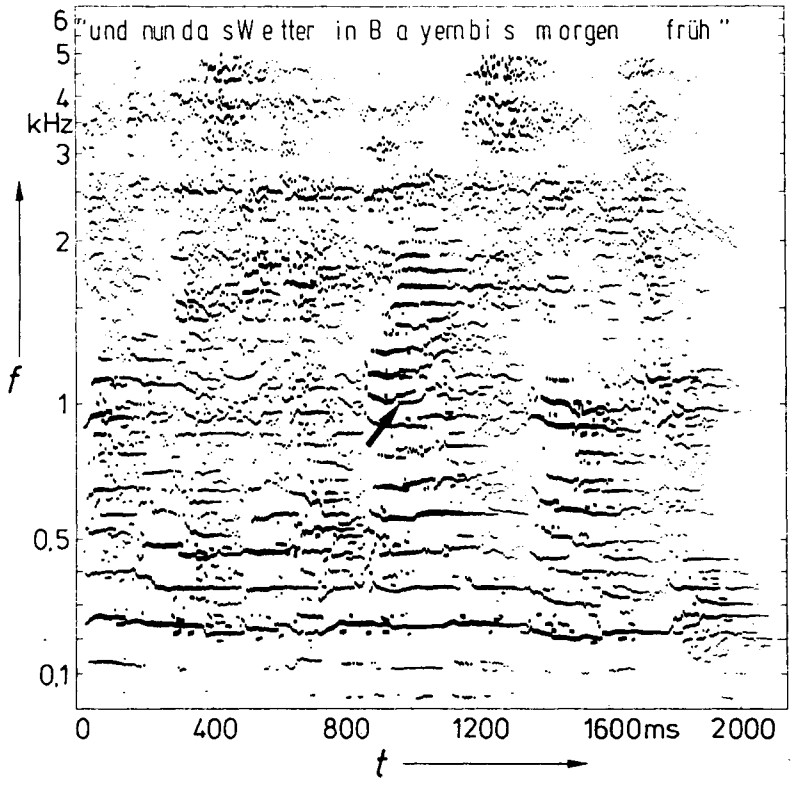


Fig. 5.1–9a: Maxigramm des in Fig. 5.1-8 abgebildeten Sprachsignals nach Wiedergabe und Aufnahme in einem Laborraum. Pfeil: Teilton, der von Echo hervorgerufen wurde.

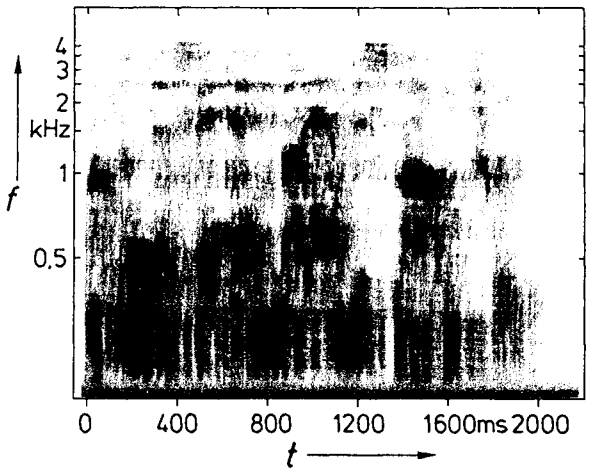


Fig. 5.1–9b: Spektrogramm desselben Signals wie im Teilbild a.

Die Verlängerung von Teiltönen, besonders bei tiefen Frequenzen, kann dazu führen, daß Teiltonverläufe bei Grundfrequenzänderungen verfälscht wiedergegeben werden (Übergang *'Wetter'* - *'in'*). Pausen lassen sich gerade noch am Versatz der Teiltonfrequenzen und dem Auftreten von 'Seitenlinien' erkennen. Stimmlose und stimmhafte Laute sind recht gut anhand der Regelmäßigkeit des Teiltonaufbaus und den gleichartigen Teiltonverläufen zu unterscheiden.

Der von einem anderen männlichen Sprecher *geflüsterte* Testsatz ist in Fig. 5.1-10 dargestellt. Die Teiltöne in Fig. 5.1-10a, die den Formanten im Spektrogramm Fig. 5.1-10b entsprechen, haben einen Verlauf, der denen von periodischen Signalen ähnelt. Auch die zeitliche Veränderung von Formanten (z.B. in *'Bayern'*) läßt sich sehr gut erkennen. Dies ist insofern bemerkenswert, als auf den ersten Blick der Verlauf der spektralen Hüllkurve, wie im

Spektrogramm dargestellt, gerade bei stimmlosen Lauten die einzige sinnvolle Beschreibung zu sein scheint.

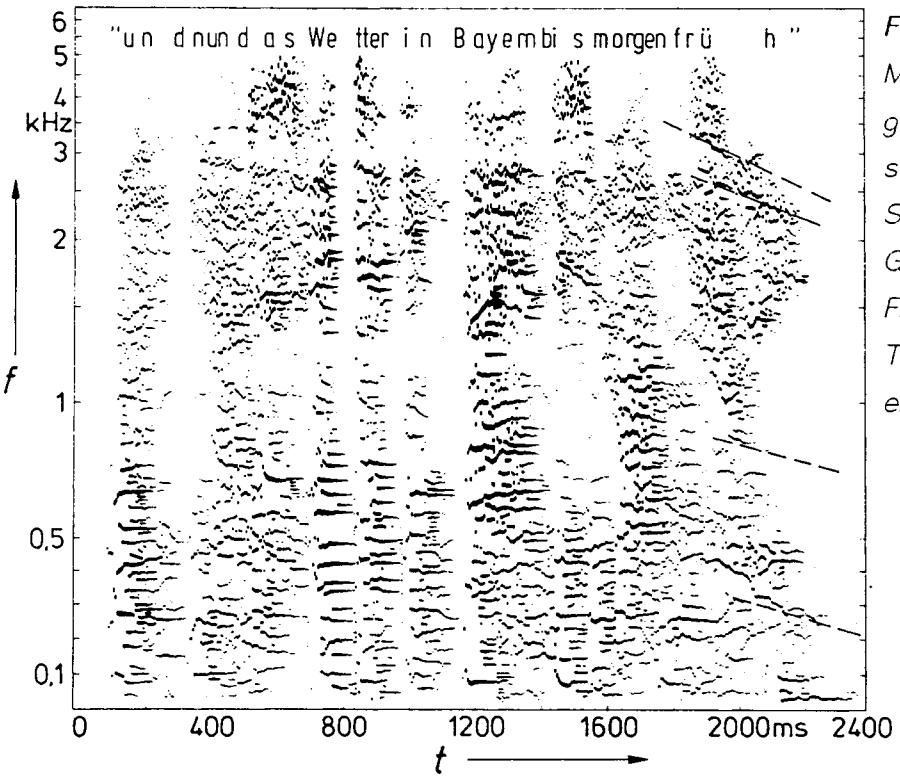


Fig. 5.1–10a: Maxigramm des geflüsterten Testsatzes (männlicher Sprecher). Gestrichelt: in der Frequenz abfallende Teiltöne zum Satzende.

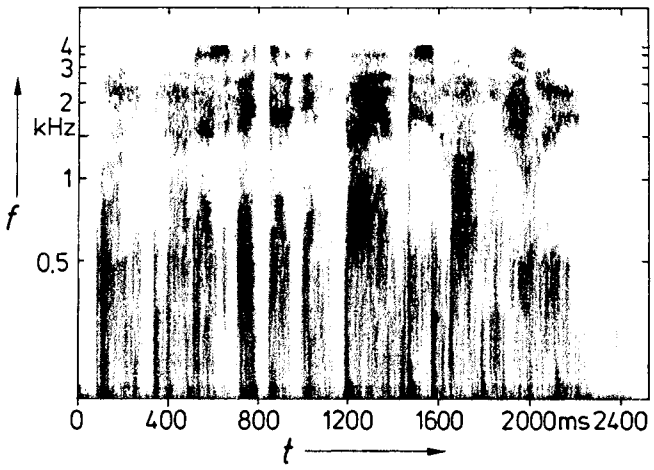


Fig. 5.1–10b: Spektrogramm desselben Signals wie im Teilbild a.

Aus einer Arbeit von Meyer-Eppler [51] ist bekannt, daß prosodische Merkmale, die normalerweise in Grundfrequenzverläufen enthalten sind, auch bei geflüsterter Sprache wiedergegeben werden können. Mit Hilfe des TTZM scheint es möglich zu sein, dazu weitere Untersuchungen durchzuführen. So beachte man die fallenden Teiltonverläufe am Ende von *'früh'*.

Maxigramm und Spektrogramm des Testsatzes, gesprochen von einer Frau, zeigen Fig. 5.1-11 a und b. Durch die hohe Grundfrequenz werden auch im Spektrogramm die einzelnen Teiltöne aufgelöst, wodurch allerdings die Bestimmung der Formanten erschwert wird. Am lul in *'nun'* oder am lal in *'das'* wird deutlich, daß Formanten bei hohen Grundfrequenzen mit einzelnen Harmonischen identisch

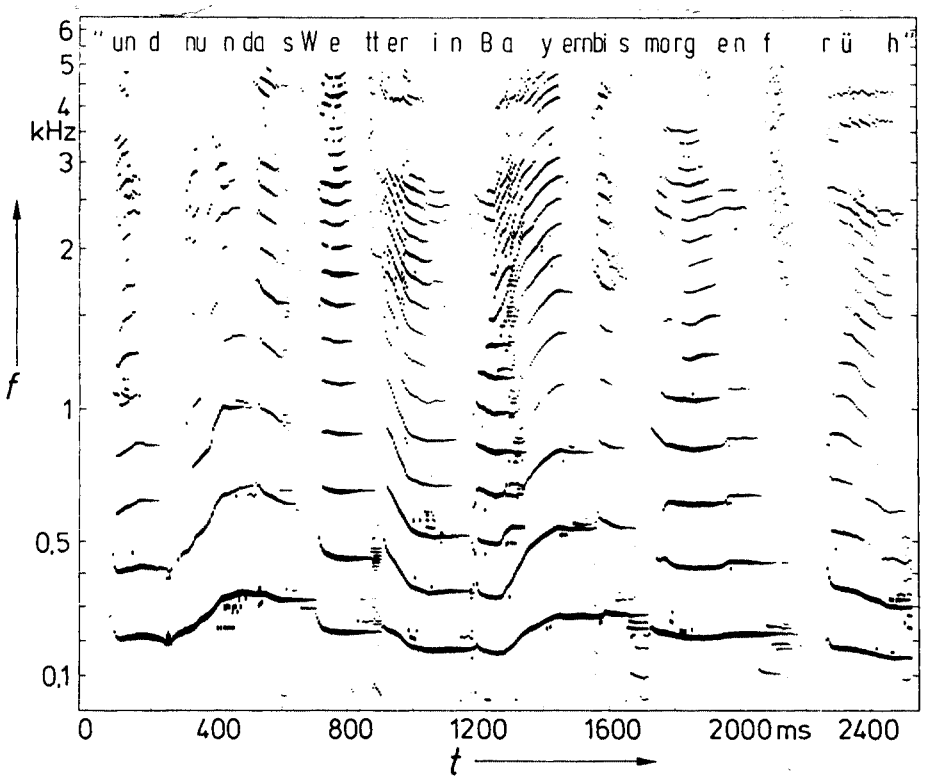


Fig. 5.1–11a: Maxigramm des von einer Frau gesprochenen Testsatzes.

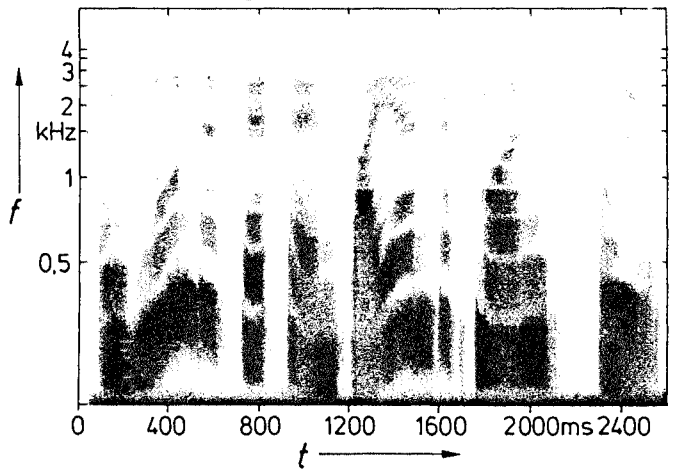


Fig. 5.1–11b: Spektrogramm desselben Signals wie im Teilbild a.

sind. Dies hat zur Folge, daß bei entsprechendem Grundfrequenzverlauf die Formanten bei etwas anderen Frequenzen als den sonst üblichen auftreten können. Die Sprachwahrnehmung wird dadurch nicht gestört (im Gegensatz zu automatischen Spracherkennungsystemen [36], [41], [42]).

Für das TTZM in Fig. 5.1-11a gilt im wesentlichen das gleiche wie für die anderen TTZM. Einige artikulatorische Feinheiten lassen sich noch erkennen, wie z.B. das nur angedeutete /g/ in *'morgen'*, das mehr als *'morjen'* ausgesprochen wurde (vergl. 5.1-8a, Seite 57).

Die Wortfolge *'Ein rotes'*, gesprochen von einem vierjährigen Kind, ist in Fig. 5.1-12a als Maxigramm und in Fig. 5.1-12b als Spektrogramm dargestellt. Die in der zweiten Hälfte im Maxigramm auftretenden tiefen Teiltöne stammen von Lüftergeräuschen in der Aufnahmekabine, die aufgrund des niedrigen Signalpegels nicht mehr verdeckt werden. Die Information über eine zweite Schallquelle läßt

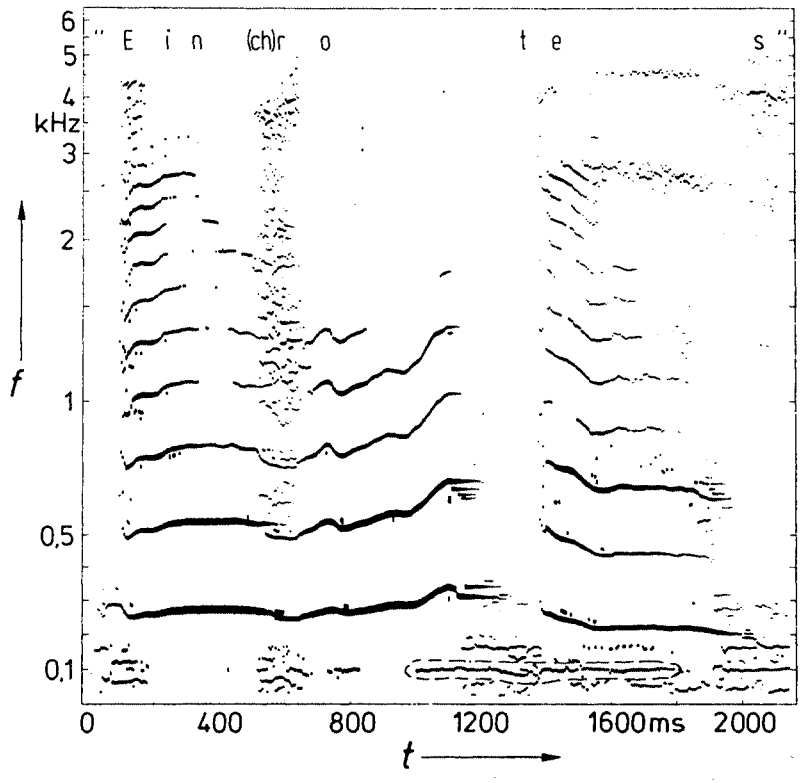


Fig. 5.1–12a: Maxigramm der Wortfolge "Ein rotes", gesprochen von einem vierjährigen Kind. Gestrichelt: Anteile eines Lüftergeräusches im Aufnahmeraum.

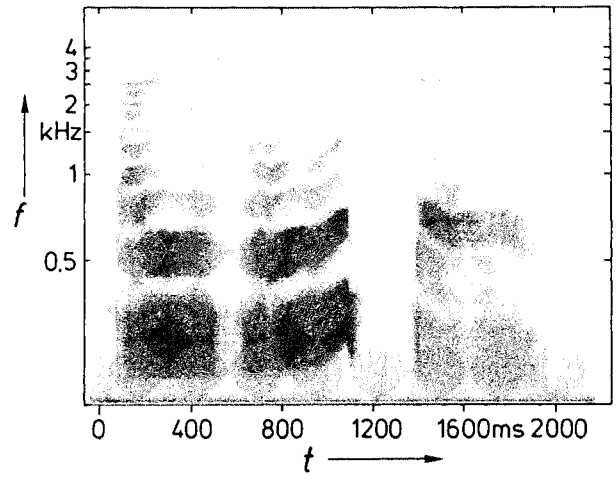


Fig. 5.1–12b: Spektrogramm desselben Signals wie im Teilbild a.

sich dem Spektrogramm in dieser Deutlichkeit nicht entnehmen. Die Anregung im lel in *'rotes'* verändert sich etwa 300ms nach Beginn von stimmhaft nach stimmlos. Dies wird in beiden Darstellungen sichtbar. Im Maxigramm kann jedoch zwischen Geräuschanteilen des Lüfters und denen der Stimme aufgrund der Fortsetzung des Formantverlaufs besser unterschieden werden.

5.1.2.2 Mehrere Stimmen

Die Sprachsignale des vorherigen Abschnitts stellen einen Sonderfall dar, da sie nur von einem Sprecher stammen und weitgehend frei von Störungen sind. In diesem Abschnitt werden deshalb die TTZM und Spektrogramme dreier

Schalle vorgestellt, die der Wahrnehmung unter realen Bedingungen näherkommen:

- Männer- und Frauenstimme gleichzeitig mit dem gleichen Satz;
- Männer- und Frauenstimme mit verschiedenen Sätzen;
- Vier Männerstimmen gemeinsam den gleichen Satz.

In Fig. 5.1-13 a und b sind das Maxigramm bzw. Spektrogramm des Testsatzes aus 5.1.2.1, gemeinsam von einer Frau und einem Mann gesprochen, dargestellt. Die Aufnahme erfolgte über ein einziges Mikrofon, beide Sprecher bemühten sich, den Satz synchron zu sprechen.

Das Spektrogramm weist eine hohe Ähnlichkeit zu dem in Fig. 5.1-9b (Raumhall, S. 59) gezeigten auf. Eine Information über die Anzahl der Sprecher läßt sich dem Spektrogramm nicht entnehmen. Dies gilt auch für das hier nicht abgebildete Spektrogramm mit 50Hz Analysebandbreite. Die zeitliche Strukturierung

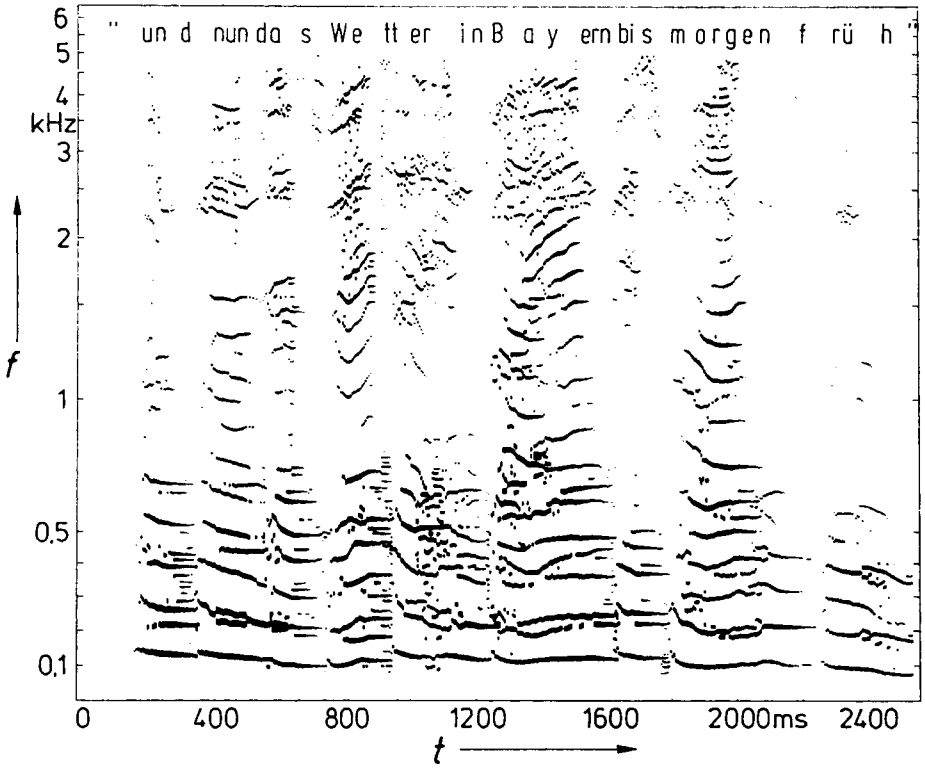


Fig. 5.1–13a: Maxigramm des Testsatzes, gleichzeitig von einer Frau und einem Mann gesprochen.

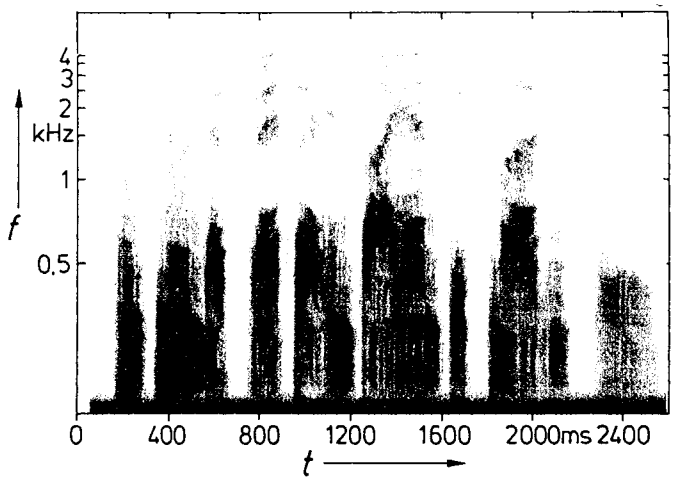


Fig. 5.1–13b: Spektrogramm desselben Signals wie im Teilbild a.

(vertikale Streifen) bei den stimmhaften Lauten ist auf die Schwebung von Spektralanteilen innerhalb der Bandbreite des Analysefilters zurückzuführen. Die Formanten sind klar erkennbar.

Im Maxigramm läßt sich an den Zeitverläufen der Teiltöne feststellen, daß zwei Sprecher beteiligt waren. Da die Grundfrequenzen nie exakt gleichzeitig verändert wurden, lassen sich die entsprechenden zueinander gehörenden Harmonischen am gleichen Zeitverlauf erkennen. Beispiele dafür sind *'nun'* - der Sprecher senkt die Grundfrequenz ab, während die Sprecherin sie konstant hält - und *'Bayern'*, wo im Übergang 'a-y' gegenläufige Änderungen der Grundfrequenz auftreten. An der zuletzt genannten Stelle ist auch sehr gut zu sehen, wie sich einzelne Harmonische gegenseitig verdecken und zum Teil nicht vorhanden sind. Trotzdem bleibt die 'Gestalt' des Verlaufs weitgehend erhalten.

Die Unterschiede zwischen Originalsignal und resynthetisiertem Signal sind gering und betreffen nicht die Information bezüglich des Inhalts und der beiden Sprecher. Bei Beschränkung auf die zehn Teiltöne mit größtem Pegel tritt beim resynthetisierten Signal der Effekt auf, daß sich die beiden Sprecher abwechselnd in den 'Vordergrund' schieben. Beim Anhören des Originalsignal ist die 'Vordergrund'- und 'Hintergrund'- Wahrnehmung durch Konzentration auf die entsprechende Stimme möglich.

Bei der realen Kommunikation kommt es häufig vor, daß mehrere unterschiedliche Quellen vorhanden sind. Dieser Fall ist für zwei Stimmen in Fig. 5.1-14a als Maxigramm bzw. als Spektrogramme mit 300Hz (Fig. 5.1-14b) und 50Hz Analysebandbreite (Fig. 5.1-14c) dargestellt. Gesprochen wurde *'Und nun das Wetter in Bayern bis morgen früh'* (Männerstimme) und *'Die Mensa bietet heute an'* (Frauenstimme). Die Aufnahme beider Stimmen erfolgte gleichzeitig über ein einziges Mikrofon.

Das Spektrogramm ähnelt den anderen Spektrogrammen des Testsatzes. Die Spektren der beiden Sprachsignale sind so miteinander vermischt, daß ihre Trennung praktisch nicht möglich ist. Auch die Auswertung des Spektrogramms mit 50Hz Bandbreite hilft nicht weiter.

Im Maxigramm lassen sich die stimmhaften Anteile beider Sprecher durch ihre unterschiedlichen Grundfrequenzverläufe relativ gut voneinander trennen. Aus der gut sichtbaren Gundfrequenz des männlichen Sprechers läßt sich eine harmonische 'Maske' generieren, mit der sich dessen Teiltöne von denen der Frauenstimme weitgehend trennen lassen. Das Verfahren wird in Abschnitt 5.3 genauer beschrieben. Als Resultat erhält man die beiden Maxigramme 5.1-15a (Teiltöne der Männerstimme) und 5.1-15b (Teiltöne der Frauenstimme). Das Maxigramm der 'Maske' ist in Fig. 5.1-15c dargestellt. Die Laute der beiden

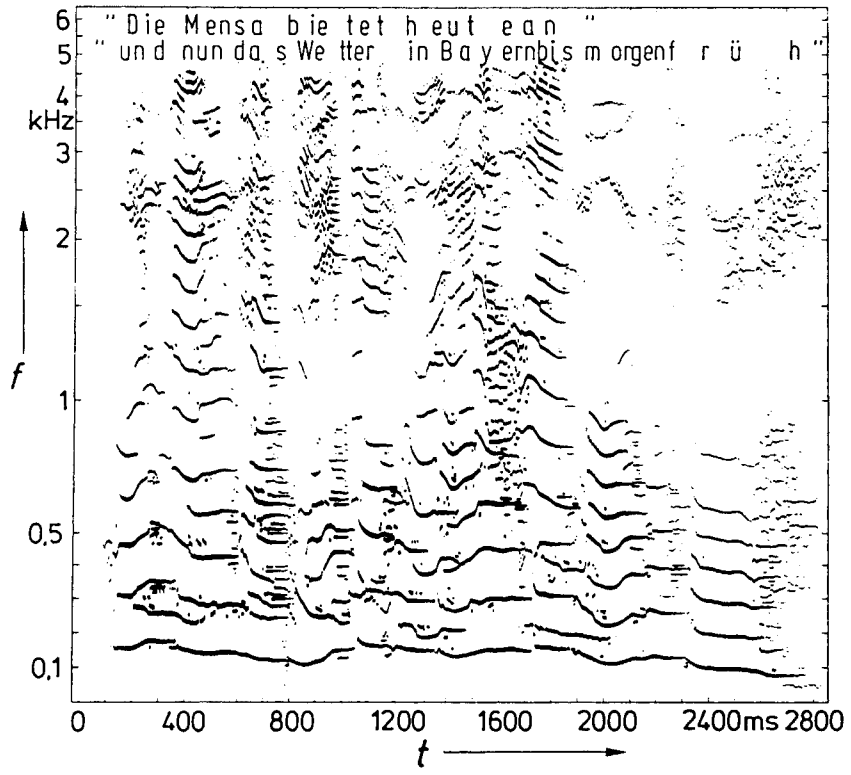


Fig. 5.1–14a (oben): *Maxigramm der gleichzeitig gesprochenen Sätze "Und nun das Wetter in Bayern bis morgen früh" (Männerstimme) und "Die Mensa bietet heute an" (Frauenstimme).*

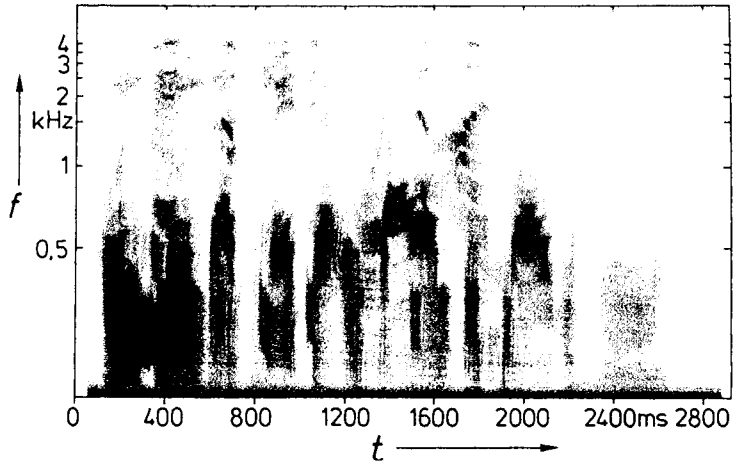


Fig. 5.1–14b (links oben): *Spektrogramm des genannten Sprachsignals mit 300Hz Analysebandbreite.*

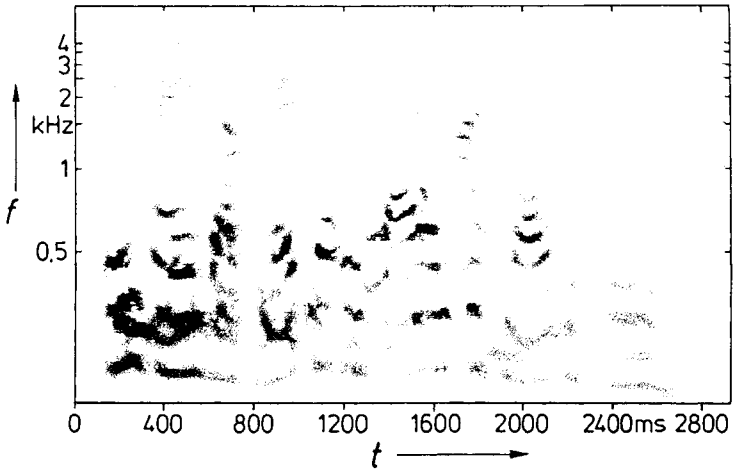


Fig. 5.1–14c: *Spektrogramm wie Fig. 5.1-14b, jedoch mit 50Hz Analysebandbreite.*

gesprochenen Sätze lassen sich den beiden Maxigrammen weitgehend entnehmen. Bei Resynthese der TTZM von Fig. 5.1.1-15 a und b erhält man zwei Sprachsignale, die zwar eine schlechte Sprachgüte aufweisen, aber dennoch als das jeweilige Sprachsignal verständlich sind.

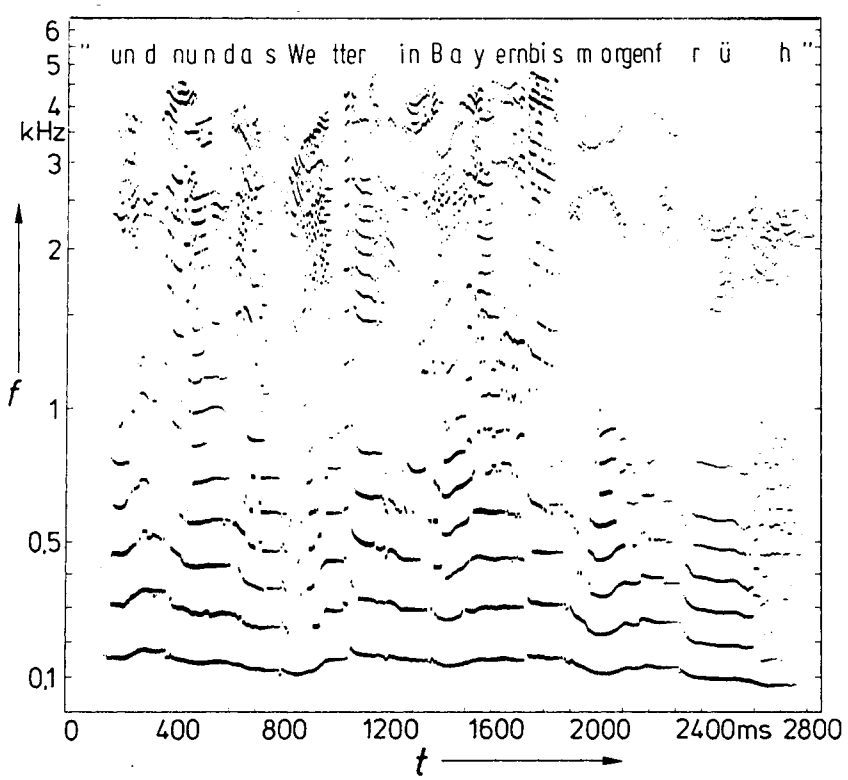


Fig. 5.1–15a: Anteil der Männerstimme in Fig. 5.1-14a.

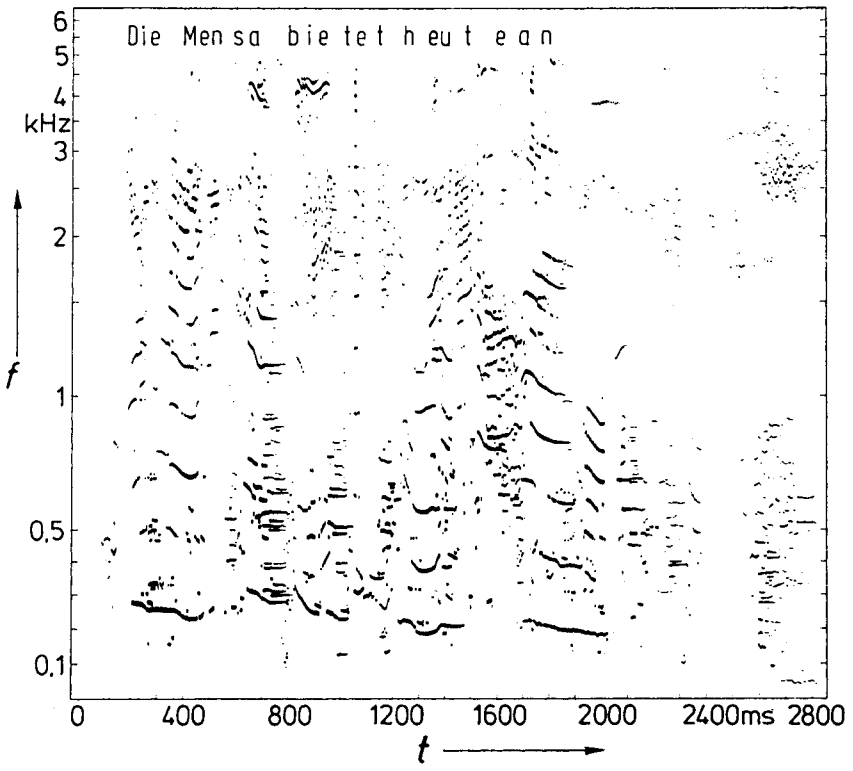


Fig. 5.1–15b: Anteil der Frauenstimme aus Fig. 5.1-14a. Das Maxigramm enthält alle Teiltöne, die sich in Fig. 5.1-14a nicht mit der in Fig. 5.1-15c gezeigten Maske in einem Bereich von 0,1 Bark überdecken.

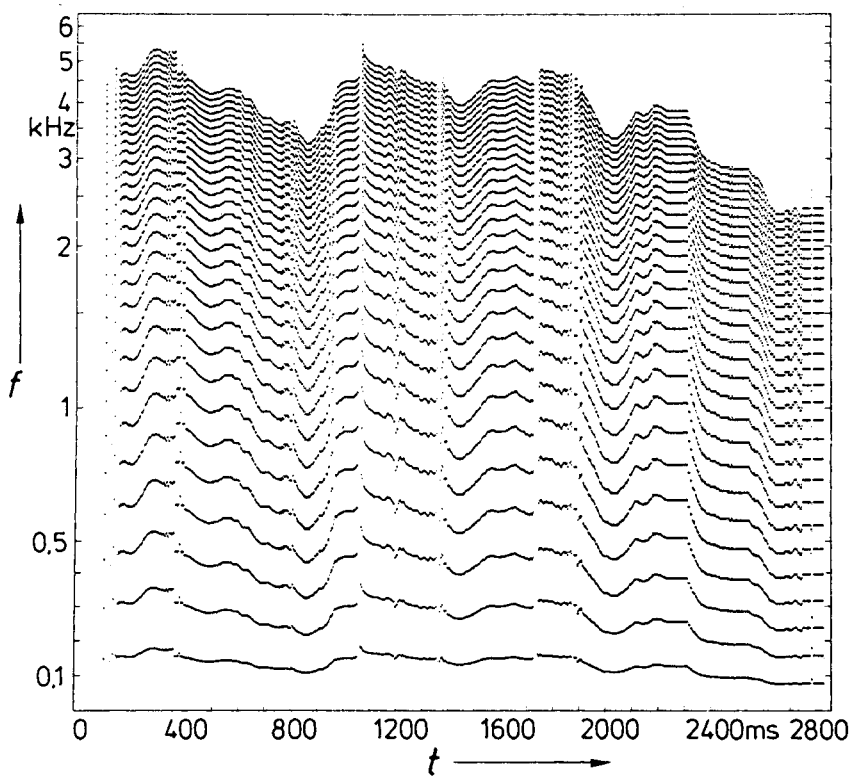


Fig. 5.1–15c: Maske zur Trennung der Teiltöne der Männerstimme in Fig. 5.1-14a von denen der Frauenstimme. Die Maske entsteht durch ganzzahliges Vervielfachen des Verlaufs des Teiltones bei 100-150Hz (Fig. 5.1-14a)

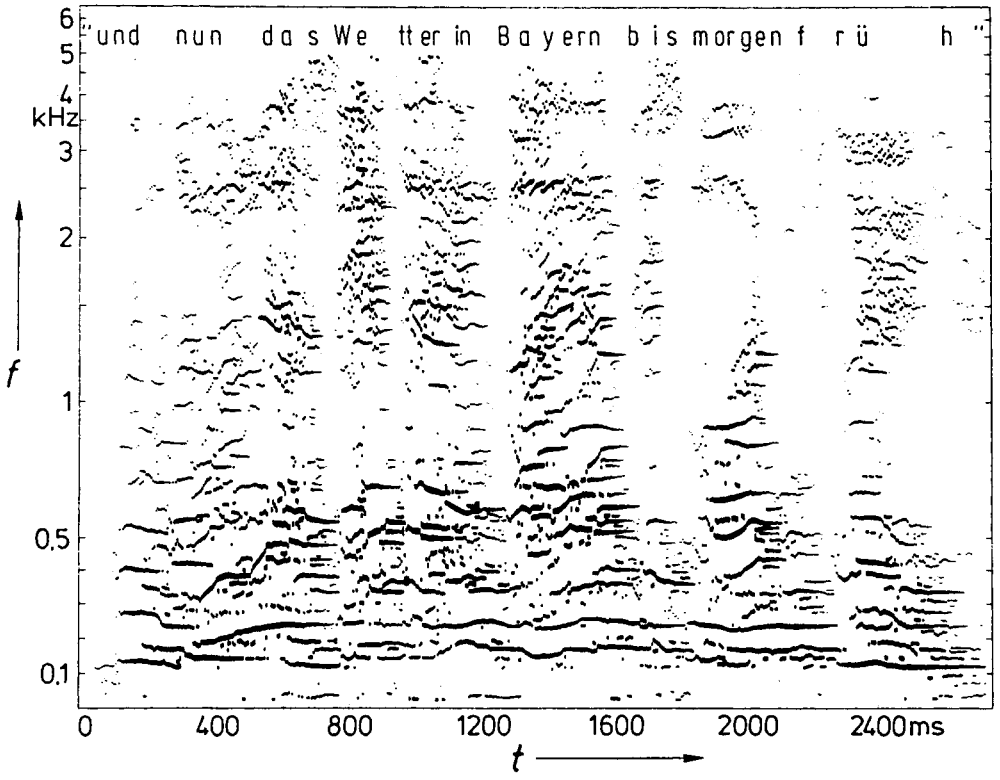


Fig. 5.1–16a: Maxigramm des Testsatzes, gleichzeitig gesprochen von vier männlichen Sprechern.

Ein weiteres Beispiel für überlagerte Stimmen zeigen Fig. 5.1-16a und b. Vier männliche Sprecher sprachen gleichzeitig den Testsatz. Das Spektrogramm

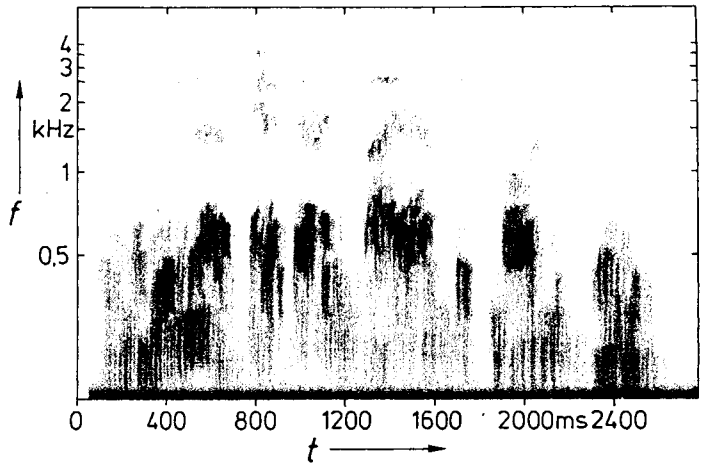


Fig. 5.1-16b ähnelt Fig. 5.1-13b; die zeitliche Strukturierung aufgrund von Schwebungen ist in Fig. 5.1-16 viel ausgeprägter. Auch diesem Spektrogramm ist keine Information über Anzahl und Grundfrequenzverläufe der Sprecher zu entnehmen.

Fig. 5.1-16b: Spektrogramm desselben Signals wie im Teilbild a.

Sie ist jedoch im Maxigramm an einigen Stellen sichtbar. Zwei Grundfrequenzverläufe sind deutlich zu erkennen, die anderen beiden werden teilweise verdeckt, was auch dem auditiven Eindruck entspricht. Die Trennung der einzelnen Sprecher im Maxigramm ist immer dann möglich, wenn sich ihre Grundfrequenzverläufe unterscheiden und nicht alle Harmonischen eines Sprechers verdeckt werden. Die Trennung ist z.B. bei 'nun' und 'früh' möglich.

5.1.2.3 Verständlichkeit

Die resynthetisierten TTZM der in den Abschnitten 5.1.2.1 und 5.1.2.2 untersuchten Sprachsignale wurden nur qualitativ von zwei bzw. drei Personen in ihrer Verständlichkeit und Güte beurteilt. Um zu quantitativen Aussagen über die Verständlichkeit der resynthetisierten Sprache zu kommen, wurden Verständlichkeitsmessungen mit einem deutschen Reimtest [56] durchgeführt. Die jeweils 100 Wörter der Testlisten 1 und 2 (Originalkopien des Forschungsinstituts der Deutschen Bundespost, Berlin) wurden in der üblichen Weise (s. Abschnitt 5.1.1) digitalisiert und die TTZM unter Verwendung der Parameter in Tab. 3.4.1 mit dem Auswerteintervall T_A = 2,5ms berechnet.

Acht normalhörende Versuchspersonen im Alter zwischen 25 und 30 Jahren nahmen an Messungen mit den digitalisierten Originalsignalen und den resynthetisierten Testwörtern teil. Der Ablauf einer einzelnen Messung wurde von einem Mikrorechner gesteuert und bestand aus:

- Laden des jeweiligen Testschalles von Diskette,
- Darbietung des Ankündigungssatzes 'Bitte markieren Sie das Wort',
- Darbietung des Testwortes,
- Anzeige der Antwortauswahl auf dem Bildschirm der Vp,
- Abfrage der Antworttasten und
- Speichern der Antwort.

Die Darbietung erfolgte nach Wandlung durch einen 12bit Digital-Analog--Wandler und anschliessender Tiefpassfilterung (f_g= 5,6kHz) diotisch mit einem freifeldentzerrten Kopfhörer in einer schallgedämmten Meßkabine. Der Ankündigungssatz wurde immer als digitalisiertes Originalsignal ausgegeben, d.h. er war von Veränderungen durch Analyse und Resynthese nicht betroffen. Nach Darbietung des Testworts wurden sechs Wörter zur Auswahl auf dem Bildschirm der Vp angezeigt. Als Antwort mußte die Vp die Taste desjenigen Wortes drücken, das sie zu hören glaubte. Dazu hatte sie beliebig lange Zeit.

Verständlichkeit resynthetisierter Sprachsignale

Reimtestliste	Originale	Resynthesen
1	99%	99%
2	100%	100%

Tab. 5.1.7: Reimtestverständlichkeiten resynthetisierter Einzelworte.

Die Ergebnisse in Tab. 5.1.7 zeigen, daß kein signifikanter Unterschied in der Verständlichkeit zwischen den beiden Testschallarten besteht. Über weitere Untersuchungen zur Verständlichkeit und Sprachgüte wird im folgenden berichtet.

5.1.3 Datenreduktion

Durch die zunehmende Verbreitung der digitalen Übertragung und Speicherung von Audiosignalen gewinnen Verfahren zur Datenreduktion immer mehr an Bedeutung. Die bekannten Verfahren der datenreduzierenden Signalcodierung lassen sich in drei Klassen einteilen:

1) Signalformcodierung
2) Transformationscodierung
3) Parametrische Codierung

Bei der Signalformcodierung werden statistische Bindungen im Zeitsignal ausgenutzt. Mit den bekannten Verfahren wie DPCM, ADPCM, Deltamodulation (DM) und ADM werden Datenraten im Bereich von einigen 100kbit/s bis herunter auf 16kbit/s erzielt [25]. Gehöreigenschaften werden in der Regel nicht berücksichtigt. Die Störgeräusche, die bei niedrigen Bitraten durch die unvermeidlichen Abweichung des Signals von der Schätzfunktion und die Datenreduktion entstehen, sind in der Regel breitbandig und entsprechend gut wahrnehmbar, da sie im allgemeinen durch das Signal nicht vollständig verdeckt werden.

Die Transformationscodierungen erzielen eine Datenreduktion durch grobe Quantisierung des Kurzzeitspektrums. Die Quantisierung wird entweder aufgrund statistischer Bindungen oder aufgrund von Gehöreigenschaften, wie z.B Mithörschwellen, durchgeführt. Vor allem in neuerer Zeit wurden Verfahren zur

Datenreduktion bei hochwertigen Audiosignalen (Hifi-Qualität) vorgestellt, bei denen Datenraten von 100-200kbit/s erzielt werden [45], [47], [93].

Zur datenreduzierten Speicherung bzw. Übertragung von Sprachsignalen mittlerer Qualität mit Datenraten zwischen 3 und 16kbit/s werden parametrische Codierungen verwendet. Dabei wird das Sprachsignal durch Parameter wie stimmhaft/stimmlos, Grundfrequenz oder Formantfrequenzen beschrieben. Diese Verfahren, wie beispielsweise Vocodersysteme, orientieren sich sehr stark an der Spracherzeugung [25], [26]. Zum Teil wird bei der Frequenzbandaufteilung von Vocodersystemen auch die Frequenzselektivität des Gehörs berücksichtigt [37], [46].

Im allgemeinen muß davon ausgegangen werden, daß kein einzelnes, ungestörtes Sprachsignal zur Verfügung steht. Dies bereitet bei der nach wie vor problematischen Grundfrequenzbestimmung sehr große Schwierigkeiten. Die Fähigkeit des Gehörs, in gewissem Umfang Quellen aufgrund ihrer spektralen Feinstruktur zu trennen, wird durch die beiden möglichen Anregungsarten 'periodisch' (mit nur einer möglichen Grundfrequenz) oder 'geräuschhaft' nicht ausgenutzt und führt zu auditiv gut wahrnehmbaren Störungen.

Das Teiltonzeitmuster stellt eine parametrische Beschreibung des Audiosignals dar. Eine Verwendung des TTZM zur Datenreduktion bietet sich aus folgenden Gründen an:

- das Signal ist durch diskrete Elemente (Teiltöne) repräsentiert;
- die wesentliche Information ist in etwa zehn Teiltönen enthalten;
- keine Grundfrequenzbestimmung oder 'stimmlos/stimmhaft'-Entscheidungen notwendig;
- keine Beschränkung auf periodische Signale (harm. Teiltonaufbau);
- völlige Unabhängigkeit von Signalquellenmodellen;
- die Datenreduktion orientiert sich an Gehöreigenschaften.

Es wurde deshalb auf der Grundlage des TTZMs ein Verfahren entwickelt, mit dem eine Datenreduktion von Audiosignalen bei mittlerer Qualität (vergleichbar der des Telefons) erzielt werden kann [35].

5.1.3.1 Verfahren zur Datenreduktion

Die Auswahl, Berechnung und Quantisierung der zu übertragenden Teiltöne steht im Vordergrund der folgenden Beschreibung, während der optimalen Codierung dieser Elemente weniger Aufmerksamkeit gewidmet wird.

Codiert man jeden Teilton eines Teiltonmusters unabhängig von den anderen, so erhält man die zugehörige momentane Datenrate mit der Formel:

$$d = m \cdot \left(n_f + n_L\right) / T_A \quad \text{bit/s} \qquad . \qquad (5.1.4)$$

Mit n_f und n_L werden in (5.1.4) die Wortlängen der zur Codierung von Pegel bzw. Frequenz notwendigen Datenworte bezeichnet, mit m die Anzahl der Teiltöne und mit T_A die Länge des Auswerteintervalls.

Wie die Verständlichkeitsmessungen mit Vokalen in 5.1.1 zeigten, reichen die 5-10 Teiltöne mit größtem Pegel zur Beschreibung der wesentlichen Information aus. Die Anzahl m der zu übertragenden Teiltöne wird deshalb auf zehn begrenzt. Überträgt man konstant zehn Teiltöne, so erhält man eine konstante Datenrate, wenn T_A und n_f bzw. n_L nicht zeitvariabel sind. Wie die statistischen Daten der TTZM der Reimtestwörter zeigen (Tab. 5.1.8), liegt die mittlere Anzahl der Teiltöne deutlich über zehn. Bei Begrenzung auf zehn Teiltöne sinkt der Mittelwert auf etwa 9,6 Teiltöne ab. Dieser geringe Unterschied rechtfertigt nicht den höheren Aufwand zur Codierung der Teiltonanzahl. Treten weniger als zehn Teiltöne auf, so werden Teiltöne mit Frequenz und Pegel gleich Null hinzugefügt.

Statistische Verteilung der Teiltonanzahl pro Teiltonmuster		
Reimtestvokabular:	Liste 1	Liste 2
mittlere Teiltonanzahl	22,72 ± 11,8	21,62 ± 11,6
mittlere Teiltonanzahl nach Begrenzung auf max. 10 Teiltöne	9,69 ± 1,1	9,60 ± 1,2

Tab. 5.1.8: Mittlere Anzahl der Teiltöne pro Teiltonmuster.

Da die Analysefrequenzen zur Berechnung des geglätteten Leistungsspektrums (Abschnitt 3.3) entsprechend dem Frequenzauflösungsvermögen des Gehörs gewählt sind, bietet es sich an, die Teiltonfrequenz durch den Index der zugehörigen Analysefrequenz anzugeben. Vergrößert man den Abstand der Analysefrequenzen auf 0,07 Bark und legt den zu übertragenden Frequenzbereich auf 100Hz - 5,2kHz fest, so werden 256 Analysefrequenzen benötigt. Daraus resultiert eine Wortlänge von 8bit zur Angabe der Teiltonfrequenz. Die durch ungenaue Frequenzangabe hervorgerufene Inharmonizität bei Signalen mit harmonischem Teiltonaufbau wird erst ab einem Abstand der Analysefrequenzen von etwa 0,28 Bark (64 Frequenzstützwerte) wahrgenommen. Bei einem einzelnen, ungestörten Sprachsignal wären diese Inharmonizitäten noch tolerierbar, bei komplexeren Signalen wird jedoch die spektrale Feinstruktur nur unzureichend wiedergegeben.

Vom Gehör werden Pegelunterschiede von 1dB gerade noch wahrgenommen. Eine Quantisierung des Teiltonpegels in 4dB-Schritten ist im A-B Vergleich wahrnehmbar, wirkt sich aber nicht auf die Verständlichkeit aus. Mit 15 Pegel-

stufen erhält man einen Dynamikbereich von 60dB, die zugehörige Wortlänge n_L beträgt 4bit.

Die Anzahl der Daten zur Beschreibung des Teiltonpegels läßt sich noch weiter reduzieren. In Fig. 5.1-17a ist ein typisches TTM mit in 4dB-Schritten quantisierten Teiltonpegeln dargestellt (durchgezogen). Ordnet man diese Teiltöne nach fallendem Pegel, wie in Fig. 5.1-17b, so läßt sich die 'Hüllkurve' der Teiltonpegel durch eine Gerade annähern. Die in Fig. 5.1-17b eingezeichnete Gerade wird vom größten und kleinsten Teiltonpegel bestimmt. Bestimmt man die Pegel der dazwischenliegenden Teiltöne aus dieser Geraden, so erhält man das in Fig. 5.1-17a gestrichelt eingezeichnete TTM. Dieses sehr einfache Verfahren liefert für Sprachsignale brauchbare Ergebnisse. Zur Angabe der Geraden sind insgesamt acht Bit notwendig, bei zehn Teiltönen reduziert sich somit der rechnerische Wert von n_L auf 8/10 bit.

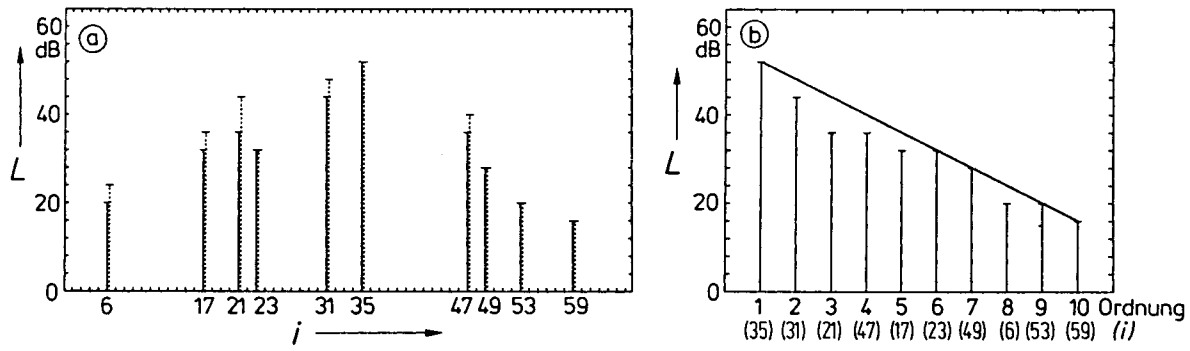


Fig. 5.1–17: angenäherte Beschreibung der Teiltonpegel durch eine Interpolationsgerade. a) (links): Teiltonmuster mit Originalpegeln (durchgezogen) und Pegeln aus der Interpolationsgeraden. b) Bestimmung der Interpolationsgeraden durch Ordnen der Teiltöne nach fallenden Pegeln.

Nach Formel 5.1.4 wird die Datenrate auch von der Länge des Auswerteintervalls T_A bestimmt. Die Verlängerung des Auswerteintervalls hat zur Folge, daß Teiltöne mit kürzerer Dauer fälschlicherweise verlängert werden, was als leichtes Klingeln wahrgenommen wird. Dadurch wird die Verständlichkeit aber nicht beeinträchtigt, da sich die Teiltöne an Stellen befinden, an denen eine Anregung stattfinden soll. Bei Verlängerung des Auswerteintervalls über 25ms werden zeitliche Änderungen des Sprachsignals störend verfälscht und Rauschanteile klingen tonal.

Die datenreduzierte Übertragung von Sprache mit Hilfe des TTZM wurde für Datenraten von 4kbit/s und 16kbit/s untersucht. Die entsprechenden Parameter sind in Tabelle 5.1.9 angegeben.

Parameter zur datenreduzierten Übertragung von Sprachsignalen				
Berechnung des Teiltonzeitmusters:		Datenrate in kbit/s	16	4
Anzahl der Analysefrequenzen	256	Auswerteintervall T_A	7,5ms	20ms
Abstand der Analysefrequenzen	0,07 Bark	Wortlänge Frequenz n_f	8bit	8bit
Frequenzbereich	20Hz - 5,2kHz	Wortlänge Pegel n_L	4bit	8/10bit
Ausgeprägtheitsschwelle ΔL_A	3dB	Teiltonanzahl m	10	10
Analysebandbreite B	0,1 Bark			
Glättungszeitkonstante (< 3kHz)	0,063/B			
(> 3kHz)	1,25ms			

Tab. 5.1.9: Analyse- und Codierungsparameter zur datenreduzierten Übertragung von Sprachsignalen.

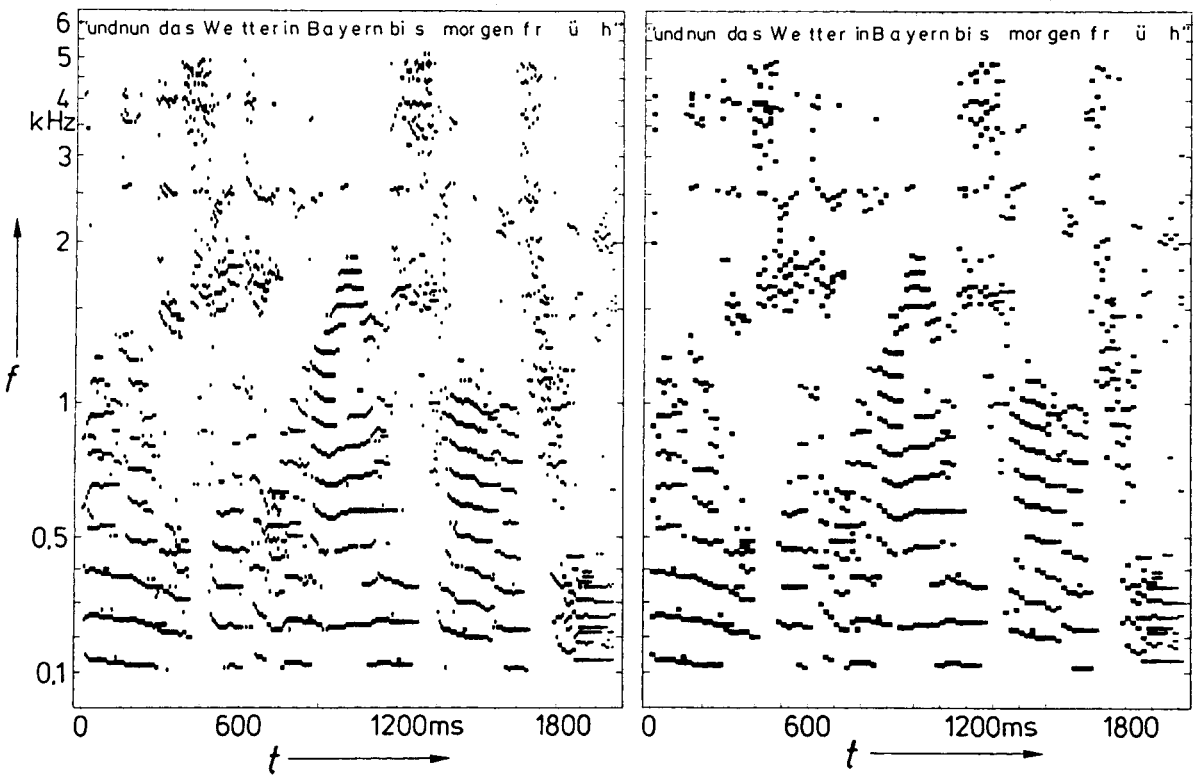


Fig. 5.1–18:
Maxigramm des auf 16kbit/s datenreduzierten Teiltonzeitmusters des Testsatzes.

Fig. 5.1–19:
Maxigramm des auf 4kbit/s datenreduzierten Teiltonzeitmusters des Testsatzes.

In Fig. 5.1-18 und Fig. 5.1-19 sind die TTZM des Satzes 'Und nun das Wetter in Bayern bis morgen früh' nach Datenreduktion auf 16kbit/s (5.1-18) und 4kbit/s (5.1-19) dargestellt. Bei Vergleich mit Fig. 5.1-8 (S. 57) fällt besonders die Wirkung des verlängerten Auswerteintervalls auf.

5.1.3.2 Verständlichkeit und Güte der datenreduzierten Sprache

Zur Beurteilung der Verständlichkeit und Güte der datenreduzierten Sprachsignale wurden Messungen durchgeführt. Als Testschalle wurden die Wörter der Testlisten 1 und 2 des Reimtests verwendet (s.a. Abschnitt 5.1.2.3). Nach Digitalisierung wurden die TTZM mit den in Tab 5.1.9 (linke Seite) angegebenen Transformationsparametern und einem Auswerteintervall T_A = 2,5ms berechnet. Aus diesen TTZM wurden jeweils drei Versionen von Testschallen resynthetisiert, eine weitere Version bestand aus den digitalisierten Originalsignalen:

A: keine Beschränkung der Teiltonanzahl
B: Datenrate 16kbit/s, wie in Tab. 5.1.9
C: Datenrate 4kbit/s, wie in Tab. 5.1.9
O: digitalisierte Originalsignale

Die Verständlichkeitsmessungen wurden mit dem Reimtest, wie in Abschnitt 5.1.2 beschrieben, durchgeführt. An den Messungen nahmen acht normalhörende Versuchspersonen im Alter zwischen 25 und 30 Jahren teil. Die vier Versionen der Testschalle wurden jeder Vp jeweils nur einmal dargeboten. Die Messungen mit den einzelnen Versionen erfolgte bei Testliste 1 in der Reihenfolge O-C-B-A und bei Testliste 2 in der Reihenfolge C-B-A-O, da bei begrenztem Testmaterial (2 Listen) starke Lerneffekte auftreten können. Es wurde daher ein Taubtest ohne Testschalldarbietung durchgeführt. Bei diesem mußten einige Vpn aus den sechs visuell vorgegebenen Testworten eines Reimtestensembles eines auswählen. Dabei wurde bei der häufig verwendeten Liste 1 eine Trefferquote von 64% erzielt, bei Liste 2 nach Ende aller Messungen 30%. Durch die angegebene Reihenfolge wurde bei Liste 2 vor allem vermieden, daß sich die Verständlichkeit der am meisten datenreduzierten Version C durch Lerneffekte erhöht.

Die Ergebnisse der Verständlichkeitsmessungen zeigt Fig. 5.1-20. Dargestellt sind die Zentralwerte und Wahrscheinlichen Schwankungen der Reimtestverständlichkeiten von Liste 1 (geschlossene Quadrate) und Liste 2 (offene Quadrate). Mit abnehmender Datenrate nimmt auch die Verständlichkeit ab. Trotzdem ist die Verständlichkeit der datenreduzierten Sprache, vor allem in Anbetracht des einfachen Verfahrens, sehr hoch.

Aus Verständlichkeitsmessungen läßt sich bei hohen Verständlichkeitswerten nur bedingt auf die Sprachgüte schliessen [71]. Deshalb wurde eine Versuchsreihe

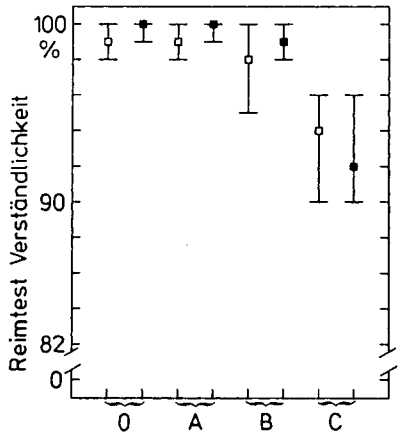


Fig. 5.1–20: Zentralwerte und Wahrscheinliche Schwankungen der Reimtestverständlichkeit. Offene Quadrate: Liste 1 (8 Vp), geschlossene Quadrate: Liste 2 (7Vp). O: Originale, A: komplettes TTZM, B: 16Kbit/s, C: 4Kbit/s.

zur Beurteilung der Sprachgüte durchgeführt. Aus je 50 Wörtern der Testschallversionen O, A, B und C der Liste 1 wurden zwei Testlisten zu hundert Wörtern zusammengestellt. Die Reihenfolge der Wörter bzw. Versionen war zufällig; alle vier Versionen eines Wortes waren jeweils in einer Liste. Diese Testschalle wurden 14 Vpn in der gleichen Weise wie beim Reimtest dargeboten. Nach jeder Einzeldarbietung mußten die Vpn die Sprachgüte mit den Zahlen 0 (=sehr schlecht) bis 10 (=sehr gut) beurteilen und ihre Bewertung auf einem Protokollblatt notieren.

Die Häufigkeitsverteilungen der Antworten in Abhängigkeit von der Testschallversion sind in Fig. 5.1-21b dargestellt, die entsprechenden Zentralwerte und Wahrscheinlichen Schwankungen in Fig. 5.1-21a. Jede der vier Häufigkeitsverteilungen beruht auf 700 Antworten. Der Abfall der Sprachgüte mit abnehmender Datenrate ist signifikant. In Anbetracht der erzielten Datenreduktion um den Faktor 10 bzw. 40 (153kbit/s auf 16kbit/s bzw. 4kbit/s) und der hohen Verständlichkeit ist die Abnahme der Sprachgüte akzeptabel.

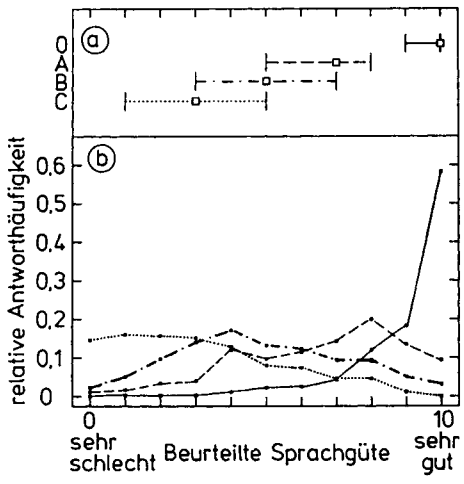


Fig. 5.1–21: Bewertung der Sprachgüte durch 12 Vp. Durchgezogen: Originale (O), gestrichelt: alle Teiltöne (A), strichpunktiert: 16Kbit/s (B), punktiert: 4Kbit/s (C). a) Zentralwerte und Wahrscheinliche Schwankung, b) Häufigkeitsverteilung.

Das vorgestellte Verfahren ist in dieser Form sicher noch nicht zu einer praktischen Verwendung geeignet, da der Rechenaufwand gegenüber bestehenden Verfahren, wie z.B. LPC-Vocodern um den Faktor 10-100 größer ist. Andererseits bereitet die Übertragung komplexer Schalle, wie die Beispiele in Abschnitt 5.1.2 zeigen, keine Schwierigkeiten, da keine Grundfrequenzbestimmung notwendig ist und die wesentlichen Merkmale der spektralen Feinstruktur erhalten bleiben.

5.1.4 Diskussion

Aus sämtlichen Messungen und Untersuchungen von Sprachsignalen, die aus TTZMn resynthetisiert wurden, geht hervor, daß Sprachsignale im TTZM vollständig gehörgerecht repräsentiert sind. Durch die Beschränkung auf genügend ausgeprägte Maxima des zeitlich geglätteten Leistungsspektrums $G(\omega,t)$ und den Verzicht auf eine Phaseninformation gehen die informationstragenden Merkmale nicht verloren. Dies gilt vor allem auch dann, wenn das Nutzsignal durch weitere Schallquellen gestört oder bei der Übertragung vom Schallsender zum Empfänger verändert wird. Gerade in diesen Fällen ist die Erhaltung der Sprachinformation und des allgemeinen Höreindruckes von großer Bedeutung.

Die Repräsentation von Sprachsignalen im TTZM steht nicht im Widerspruch zu üblichen Beschreibungen des Sprachsignals durch Anregung und spektrale Hüllkurve, da Teiltöne bevorzugt im Bereich der Maxima der spektralen Hüllkurve, den Formanten, auftreten. Bei stimmhafter Anregung ist dies auch zu erwarten, da das Spektrum tatsächlich im wesentlichen einem Linienspektrum entspricht. Die Ergebnisse stimmen mit denen eines anderen Verfahrens zur datenreduzierten Übertragung von Sprache bzw. zur Synthese natürlich klingender Sprache auf der Basis von Teiltönen weitgehend überein [32]. Die Übereinstimmung zwischen Teiltönen und Formanten trifft jedoch auch bei stimmloser Anregung zu, wie z.B. bei geflüsterter Sprache. Der geringe auditive Unterschied zwischen dem Original und der resynthetisierten geflüsterten Sprache ist insofern bemerkenswert, als Rauschsignale zunächst als nicht geeignet für eine Beschreibung durch Teiltöne erscheinen.

Das in Abschnitt 5.1.3 vorgestellt Verfahren zur Datenreduktion ist ein Beispiel für die vielfältigen Anwendungsmöglichkeiten des TTZMs. Insbesondere der vergleichsweise einfache Ansatz und die durch gehöradäquate Datenreduktion erzielte niedrige Datenrate sind von Vorteil. Die Schwierigkeiten, die bei Vocodersystemen durch die Grundfrequenzbestimmung und die zwei möglichen Anregungsarten (periodisch oder Rauschen) auftreten, entstehen beim beschriebenen Verfahren nicht, da die Bestimmung der Teiltöne des TTZM völlig unabhängig von Signalquellenmodellen mit ihren im allgemeinen nicht praxisgerechten Annahmen ist. Die Ergebnisse der Verständlichkeitsmessungen bei Reduktion der Teiltonanzahl stimmen weitgehend mit den Daten überein, die mit einem Vocodersystem ermittelt wurden, dessen Syntheseteil aus auf die Mittenfrequenzen der Bandpässe abgestimmten Sinusgeneratoren bestand [95]. Die Art und Weise, in der die Datenreduktion beim vorgestellten Verfahren möglich ist (Begrenzung der Anzahl der Teiltöne, grobe Quantisierung des Pegels und des zeitlichen Verlaufs), bestätigt die Annahme, daß vor allem die Frequenzen der Teiltöne wichtig sind, deren Amplituden weniger und die Phasenbeziehungen überhaupt nicht.

5.2 Musik

Sprache ist aufgrund ihrer relativ einfachen spektralen Struktur (harmonisch oder geräuschhaft) sehr unempfindlich gegenüber Veränderungen. Die spektrale Feinstruktur von Musik ist dagegen meist wesentlich komplizierter und im allgemeinen nicht allein durch harmonische Anregung oder Beschreibung der spektralen Hüllkurve nachzubilden. Eine Veränderung der spektralen Feinstruktur musikalischer Signale, z.B. bei Verzerrung oder Modulation, macht sich außerordentlich störend bemerkbar. Musikalische Signale sind somit 'kritische' Signale zur Überprüfung des Teiltonzeitmusters.

In diesem Abschnitt wird dargestellt, wie Zeitsignale von einzelnen und von mehreren zusammenklingenden Instrumenten im TTZM repräsentiert werden. Die Analyse der Musiksignale erfolgte mit dem Auswerteintervall T_A = 2,5ms. Die Quellenangaben der Musikstücke sind im Anhang zu finden. Sämtliche Signale wurden nach Resynthese des entsprechenden TTZM auditiv von drei Personen daraufhin überprüft, wie gut ihre wesentlichen Merkmale erhalten blieben. Es wurden jedoch keine Hörversuche zur Beurteilung der Qualität durchgeführt.

5.2.1 Einzelinstrumente

In den Fig. 5.2-1 bis 5.2-6 sind die TTZM von sechs Instrumenten als Maxigramme dargestellt: Violine, Klavier, Gitarre, Orgel, Gesangstimme männlich und Schlagzeug.

Bei dem in Fig. 5.2-1 dargestellten Musikstück handelt es sich um eine Paganini-Variation, bei der zwei Töne im Oktavabstand gleichzeitig und mit Vibrato gespielt werden. Die durch das Vibrato entstehende Frequenzmodulation ist deutlich erkennbar, ebenso gleitende Frequenzänderungen. Die Unterbrechung der Saitenschwingung beim Anspielen eines neuen Tons oder beim Wechsel der Streichrichtung ist deutlich an den unterbrochenen Teiltonverläufen und dem Auftreten kurzer Teiltöne zu erkennen. Diese Maxigramme ähneln denen, die bei sprachlichen Verschlußlauten entstehen.

An einer Stelle ist erkennbar, daß die beiden gestrichenen Töne durch das Vibrato in ihrer Frequenz nicht gleichmäßig verändert wurden. Dadurch lassen sich die Harmonischen der beiden Töne visuell weitgehend voneinander trennen (durch a und b gekennzeichnet).

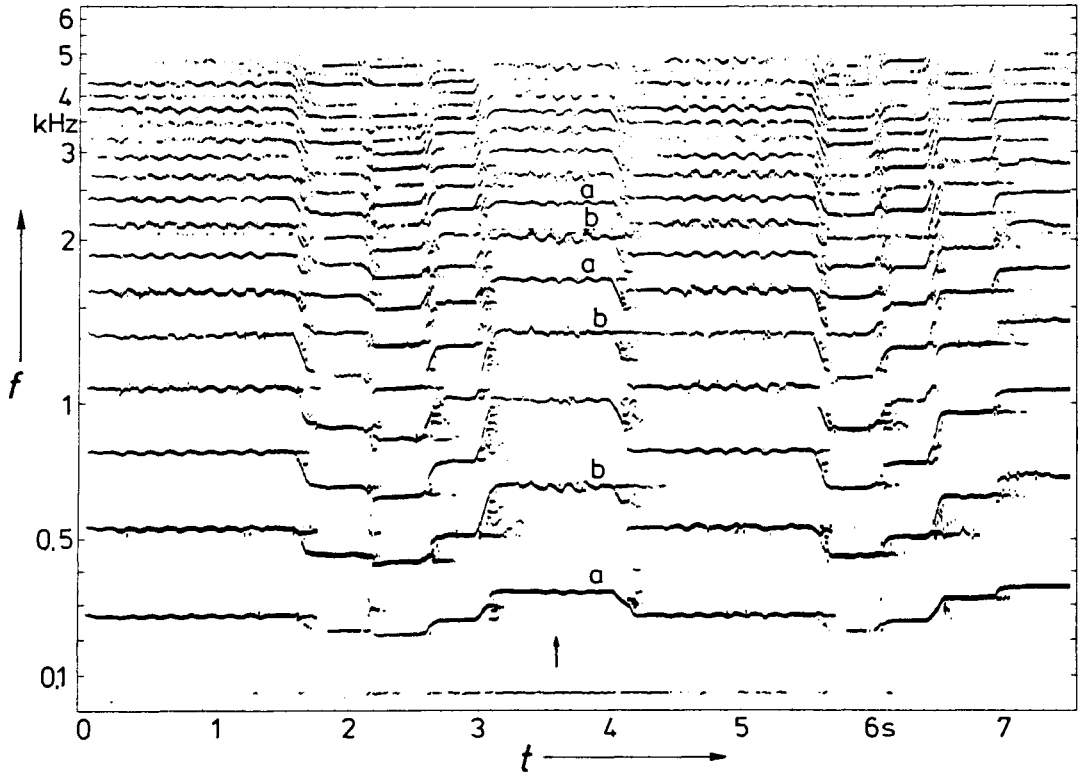


Fig. 5.2–1: Maxigramm eines Violinstückes (Paganini-Variation). Alle Töne werden als Oktaven gegriffen. An einer Stelle (Pfeil) ist deutlich die unterschiedliche Modulation zweier Töne einschließlich der zugehörigen Harmonischen zu erkennen.

Fig. 5.2-2 zeigt das Maxigramm eines Klavierstücks (Bach: Goldberg-Variationen, 1. Variation. Pianist: Glenn Gould). Typisch für Klavierklänge sind die zur Zeitachse parallelen Teiltonverläufe. Bei einigen treten jedoch auch Frequenzmodulationen auf. Nach den bisher bekannten Eigenschaften des TTZM bezüglich der genauen Wiedergabe tonaler Anteile sind jene Frequenzänderungen auf das Musiksignal zurückzuführen. Inwiefern aus solchen Teiltonverläufen Aussagen über die Klangeigenschaften des Klavie-

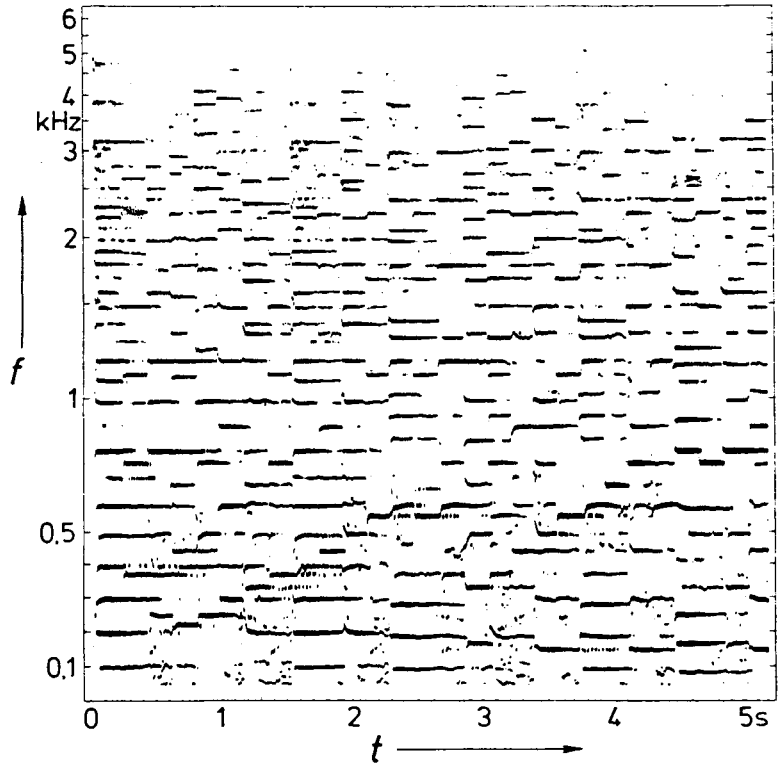


Fig. 5.2–2: Maxigramm Klavierstück

res gemacht werden können, muß noch geklärt werden.

Einen Ausschnitt aus einem Gitarrenstück (Pepe Romere: Rosita, Konzertgitarre) zeigt das Maxigramm in Fig. 5.2-3. Es zeigt viel mehr Geräuschanteile, die durch das Anzupfen der Saiten und das Greifen entstehen, als das Klavierstück in Fig. 5.2-2. In der ersten Hälfte wird ein ständiger Wechsel zwischen einem Baßton und einem Akkord gespielt, dann folgt ein schneller Einzelnotenlauf (zuerst aufwärts, dann abwärts), der nach einem kräftig angeschlagenen Ton von einem Akkord beendet wird.

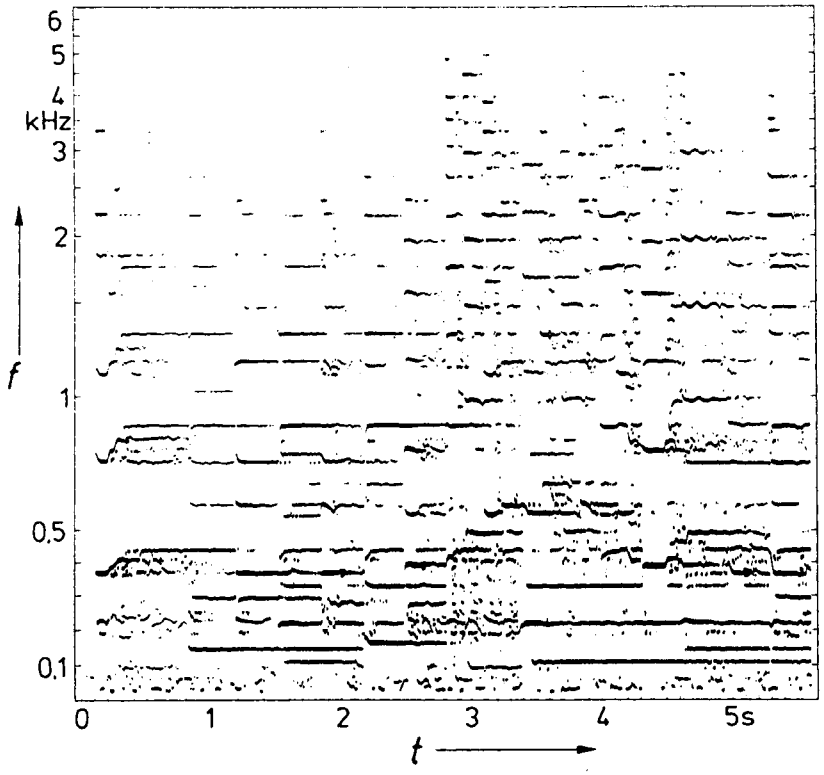


Fig. 5.2–3: Maxigramm eines Gitarrenstückes.

Fig. 5.2-4 zeigt das Maxigramm eines Orgelstückes (Bach: Toccata & Fuge). Wie schon beim Klavier zeichnen sich auch die Teiltöne der Pfeifenorgel durch einen parallelen Verlauf zur Zeitachse aus. Im Vergleich zum Klavier treten jedoch nur sehr wenige Frequenzmodulationen auf.

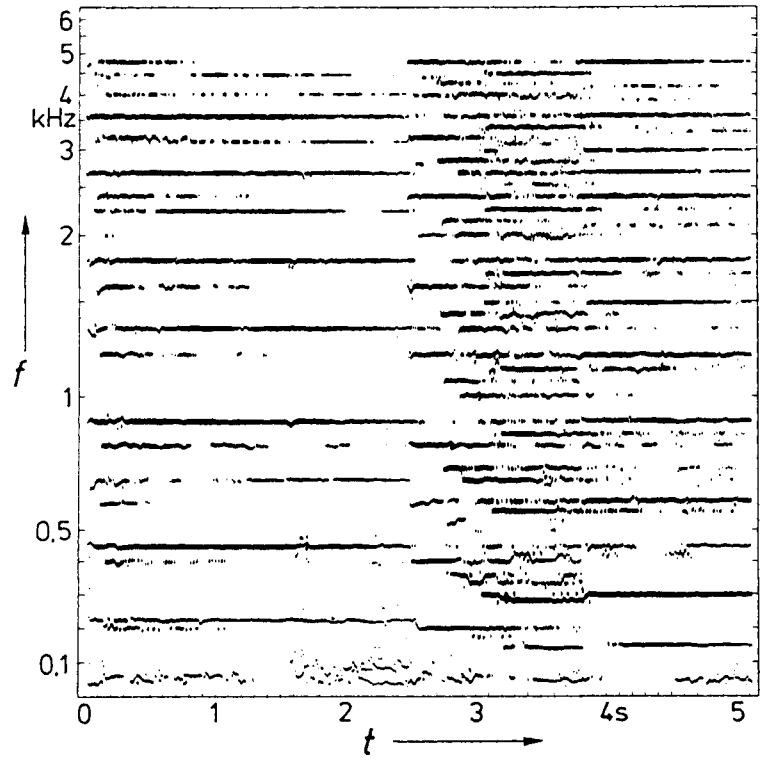


Fig. 5.2–4: Maxigramm eines Orgelstückes (Pfeifenorgel).

Das Maxigramm in Fig. 5.2-5 zeigt die Gesangsstimme des Vorsängers des Gregorianischen Chorals, der im nächsten Abschnitt abgebildet ist. Hier zeigt sich der Einfluß des Raumhalls in der schlechten Wiedergabe des Vibratos der beiden tiefsten Teiltöne am Anfang des Stückes. Sehr deutlich ist zu sehen, daß der gezeigte Ausschnitt nur Klangfarbenänderungen durch verschiedene Laute, aber praktisch keine (musikalisch relevanten) Grundfrequenzänderungen enthält.

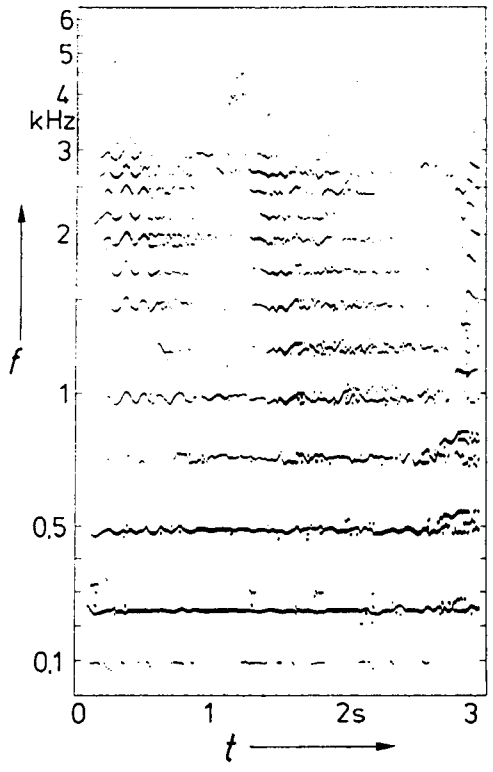
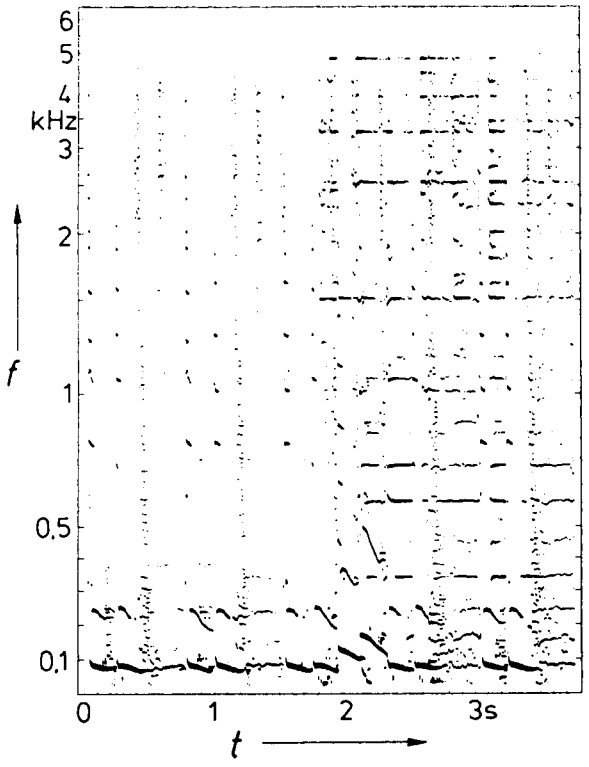


Fig. 5.2–5 (oben): Maxigramm einer männlichen Gesangsstimme.

Das letzte Maxigramm dieses Abschnittes zeigt den Ausschnitt eines Schlagzeugstückes. (Pat Metheny Group: Barcarole). Hier zeigt sich, daß die vermeintliche Baßtrommel keine ist, sondern ein wiederholter kurzer, tiefer Ton. Die regelmäßigen Strukturen mit sehr vielen Teiltönen stammen von einem 'trockenen' kurzen Schlag. Die etwas länger andauernden Teiltöne in der zweiten Hälfte des Maxigrammes stammen von glockenähnlichen Klängen.



Die resynthetisierten Zeitsignale der abgebildeten TTZM unterschieden sich auditiv nur gering von den digitalisierten Originalschallen. Die resynthetisierten Schalle klangen etwas halliger, im Bereich großer Frequenzänderungen traten leichte Störgeräusche auf. Eigenschaften wie Klangfarbe, Spielweise oder Intonation blieben jedoch erhalten.

Fig. 5.2–6:
Maxigramm eines Schlagzeuges.

5.2.2 Mehrere Instrumente

Bei einem einzelnen Instrument ist es sehr einfach, Teiltöne zuzuordnen, da nur diese einzige Schallquelle existiert. Im folgenden werden deshalb die Teilonzeitmuster einer Frauenstimme mit Klavierbegleitung, eines Männerchors ohne Begleitung, eines Kammerorchesters und einer Popgruppe vorgestellt.

Das TTZM der weiblichen Gesangsstimme mit Klavierbegleitung ist in Fig. 5.2-7 zu sehen. Die erste Hälfte des Maxigramms zeigt das Singen einer ansteigenden Linie ohne Begleitung. Mit Begleitung durch das Klavier wird eine absteigende Folge mit Vibrato am Ende gesungen, während mit dem Klavier eine ansteigende Melodie gespielt wird.

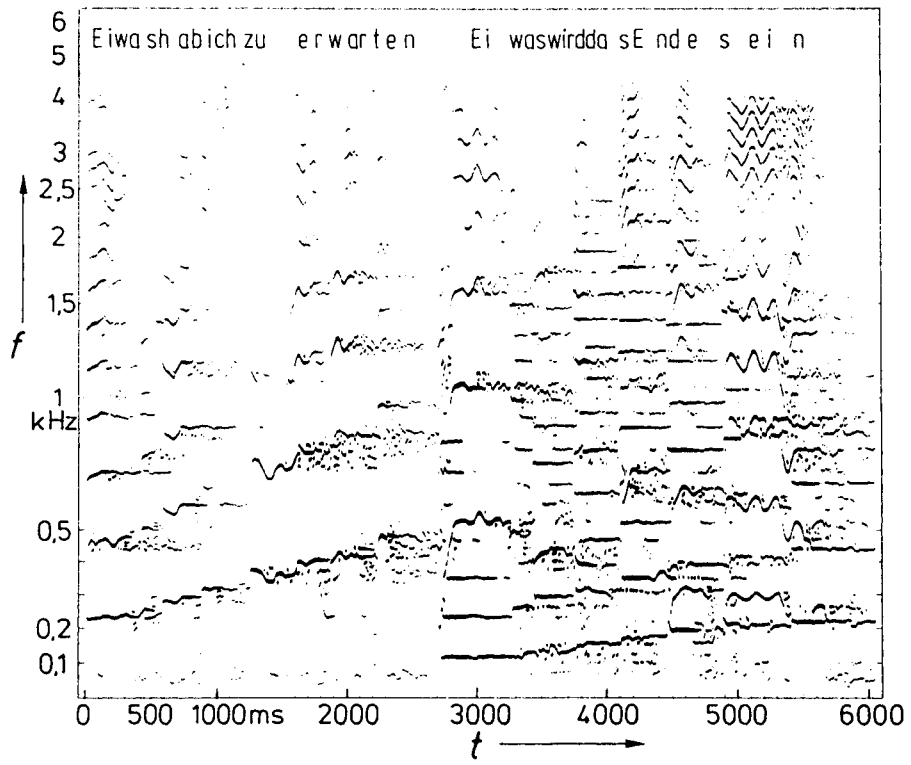


Fig. 5.2–7: Maxigramm weibliche Gesangsstimme mit Klavierbegleitung.

Der Gesang des im vorhergehenden Abschnitt erwähnten Gregorianischen Männerchors wird im Maxigramm Fig. 5.2-8 dargestellt. Auch hier ist, ebenso wie beim Vorsänger, keine größere Veränderung der Tonhöhe festzustellen.

Das Anfangsthema von Mozarts 'Kleiner Nachtmusik', gespielt von einem Kammerorchester, zeigt Fig. 5.2-9. Hier ist sehr deutlich zu sehen, wie bei den eine Quinte tiefer liegenden Tönen einfach zusätzliche Teiltöne zwischen bereits vorhandenen auftreten. Ebenso sind vier Teiltonverläufe zu erkennen (bei etwa 200Hz, 600Hz, 1200Hz und 2400Hz), die sich wie Harmonische des musikalischen Grundtons 'G' durch das gesamte dargestellte Stück ziehen.

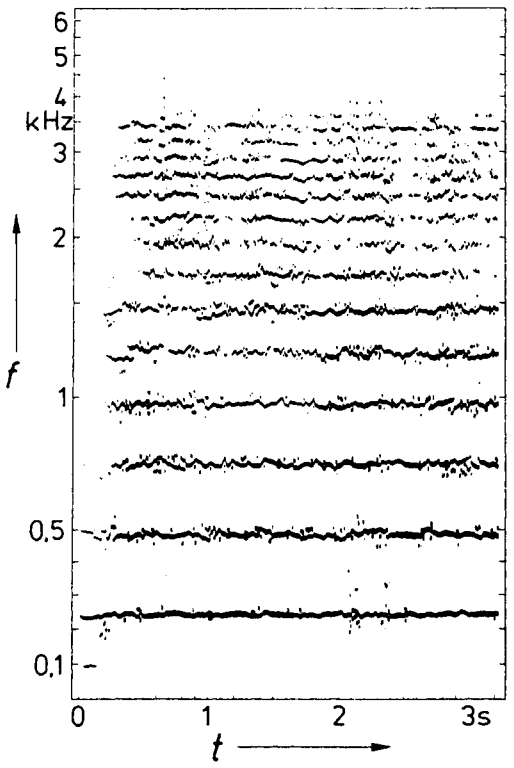


Fig. 5.2–8: Maxigramm des Gesanges eines Männerchores (Gregorianischer Choral).

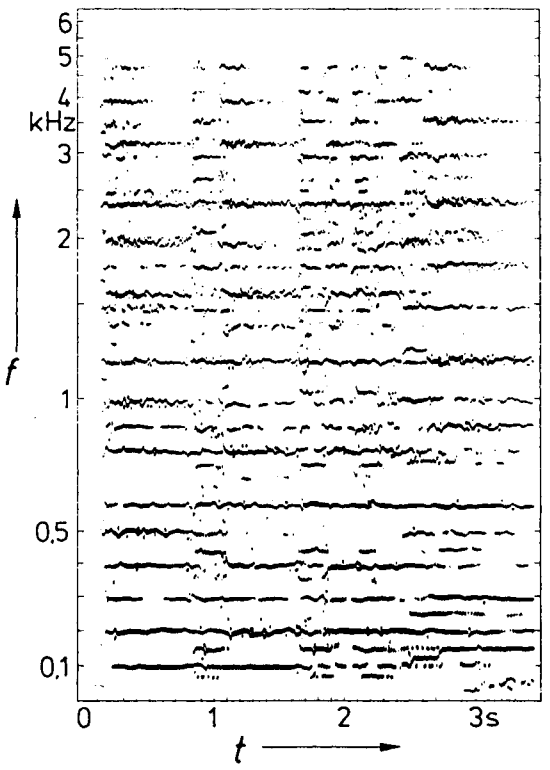


Fig. 5.2–9: Maxigramm von Mozart's "Kleiner Nachtmusik", gespielt von einem Kammerorchester.

Fig. 5.2-11 zeigt das TTZM eines Popmusikstücks (Sade) mit der Instrumentierung Bass, Schlagzeug, Klavier und Saxophon. Die vom Saxophon stammenden

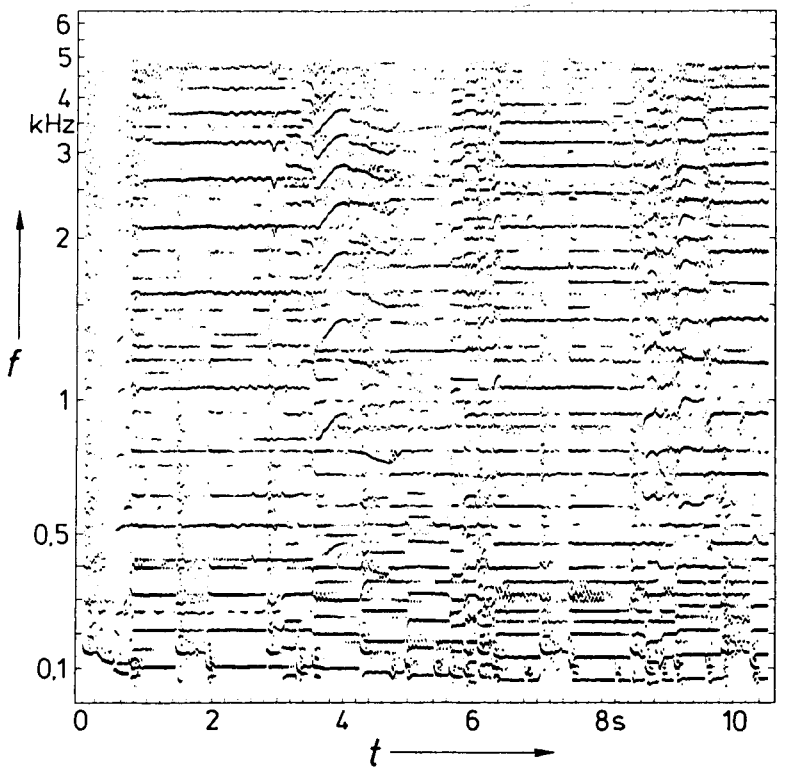


Fig. 5.2–10: Maxigramm des Musikstückes einer Popgruppe.

Anteile lassen sich aufgrund der lang ausgehaltenen Töne gut erkennen. Gut sichtbar ist auch ein gleitender Übergang bei t=4s, da die zum Saxophon gehörenden Töne alle den gleichen zeitlichen Teiltonverlauf aufweisen.

Für die hier ausgemachten Unterschiede zwischen resynthetisierten Signalen und den digitalisierten Originalsignalen gilt das gleiche wie für die Einzelinstrumente.

5.2.3 Diskussion

Die hohe Güte der aus dem TTZM resynthetisierten Zeitsignale und die zum Teil verblüffend klare Darstellung von Instrumentenklängen und Melodien im Maxigramm zeigen, daß auch bei Musiksignalen die wesentliche Information in den Zeitverläufen der Teiltöne enthalten ist.

Durch das TTZM und seine Resynthese läßt sich eine Gestalterkennung bei Musik visuell verdeutlichen und akustisch überprüfen. Weiter besteht die Möglichkeit, die Teiltöne einzelner Instrumentenklänge im TTZM zu verändern und dadurch beispielsweise deren Einfluß auf die Gesamtwahrnehmung zu untersuchen (siehe Abschnitt 5.3). Ebenso lassen sich Modelle der Gesamttonhöhenwahrnehmung, wie z.B. das Tonhöhenberechnungsverfahren nach Terhardt, Stoll und Sewann [85] auditiv überprüfen, indem nur solche Teiltöne zur Resynthese verwendet werden, die Harmonische einer bestimmten Grundtonhöhe sind.

Wie bei Sprachsignalen läßt sich auf der Grundlage des TTZMs sicherlich auch bei Musiksignalen ein Verfahren zur gehörbezogenen Datenreduktion entwickeln, da prinzipiell kein Unterschied in der Repräsentation von Sprach- und Musiksignalen im Teiltonzeitmuster besteht. Bei dem in 5.1 beschriebenen Datenreduktionsverfahren geht zuviel Information verloren, die Wiedergabegüte ist für Musiksignale zu gering.

Eine weitere Anwendungsmöglichkeit des TTZMs besteht in der Musikforschung, da die musikalische Information anhand der Teiltonfrequenzen sehr genau bestimmt werden kann, besonders auch dann, wenn noch andere Instrumente oder Töne vorhanden sind. Ein Beispiel dafür ist das Violinstück aus Fig. 5.2-1, bei dem die Intonation jeder Stimme gemessen werden kann. Da beide Stimmen im Oktavabstand gespielt werden, lassen sich ihre Frequenzen nicht aus dem Abstand der Zeitfunktions-Nulldurchgänge bestimmen. Im TTZM dagegen läßt sich auch ein unterschiedlich gespieltes Vibrato der beiden Stimmen festellen.

5.3 Audiosignalverarbeitung mit dem Teiltonzeitmuster

Sinn und Zweck der Verarbeitung von Audiosignalen ist es, Signale so zu verändern, daß gewünschte Anteile besser und unerwünschte weniger hörbar werden. Diese umfaßt beispielsweise:

- Veränderungen des Frequenzganges (z.B. Klangregler);
- Herausfiltern von Frequenzbändern (Bandpass/sperre);
- Veränderung der Signaldynamik (Dynamikkompressor);
- Frequenzänderung ('Harmonizer'), bzw. -verzerrungen;
- Veränderung des Zeitablaufs (zeitliche Verzerrungen);
- Erzeugung zusätzlicher Signale in Abh. vom Vorhandenen;
- Signaltrennung, Störbefreiung, Hallbefreiung.

Ein Großteil der Audiosignalverarbeitung wird mit idealerweise linearen Systemen durchgeführt, z.B. analoge und digitale Filter, und resultiert in einer Veränderung der spektralen Hüllkurve und der Phasenbeziehungen. Eine Bearbeitung des Signals durch Veränderung der Frequenz- oder Zeitachse ist mit hohem Aufwand verbunden [62]. Dasselbe gilt für eine Trennung überlagerter Signale wie beispielsweise die Trennung zweier Sprachsignale [54].

Bei einer musterorientierten Signalverarbeitung, wie sie größtenteils in der Bildverarbeitung existiert, erfolgt die Beeinflussung des Signals nicht durch Frequenzgänge, Modulation usw., sondern durch Veränderung bestimmter Merkmale des Signals. Dazu ist es zum einen notwendig, diese Merkmale zu erkennen und zu bestimmen, und zum anderen, nur sie zu wandeln oder herauszugreifen, ohne andere Signalparameter zu beeinflussen. Ein Beispiel hierfür ist die Trennung zweier Sprecher (vergl. Abschnitt 5.1.2.2). Dazu ist es notwendig, die Signalanteile des jeweiligen Sprechers im Gemisch zu erkennen und sie aus dem Signal herauszufiltern, ohne die des anderen Sprechers zu beeinflussen. Eine musterorientierte Signalverarbeitung kann darum in hohem Maße nichtlineare Eigenschaften aufweisen.

Das TTZM ist für eine musterorientierte Signalverarbeitung sehr gut geeignet, da das Audiosignal durch diskrete Elemente, die Teiltöne, repräsentiert wird. Sie sind jeweils durch ihre Frequenz, Amplitude und Dauer charakterisiert. Wie in den vorhergehenden Abschnitten 5.1 und 5.2 gezeigt wurde, lassen sich die informationstragenden Merkmale eines Signals weitgehend bestimmten Teiltönen oder Teiltongruppen zuordnen. Dies ist eine wichtige Voraussetzung zur Musterextraktion und -veränderung.

Die Anwendungsgebiete und Aufgabenstellungen einer musterorientierten Signalverarbeitung sind fast unübersehbar. Die Beschreibung einer solchen

Signalverarbeitung mit Hilfe des TTZMs wird deshalb auf die wichtigsten und einfachsten Fälle beschränkt. Im einzelnen sind dies:

- Bestimmung einfacher Merkmale;
- Veränderung von Teiltonmustern.
- Filterung durch Mustervergleich;

Die prinzipielle Durchführung einer Signalverarbeitung mit dem TTZM ist in Fig. 5.3-1 als Blockschaltbild dargestellt. Zuerst wird das Teiltonzeitmuster des Audiosignals berechnet. Anhand bestimmter Vorgaben werden Muster- bzw. Merkmalserkennungsprozesse durchgeführt. Diese Prozesse führen zu einem zweiten, zeitlichen parallelen Teiltonzeitmuster. Dieses zweite TTZM dient als 'Zeiger' auf die zu verändernden oder zu bearbeitenden Teiltöne des eigentlichen TTZM. Dadurch ist es möglich, Transformationen oder Manipulationen selektiv auf bestimmte Teiltöne anzuwenden oder auch neue Teiltöne dazuzufügen. Auf diese Weise lassen sich aus einem TTZM mehrere 'Abkömmlinge' erzeugen, die jeweils getrennt voneinander resynthetisiert werden können.

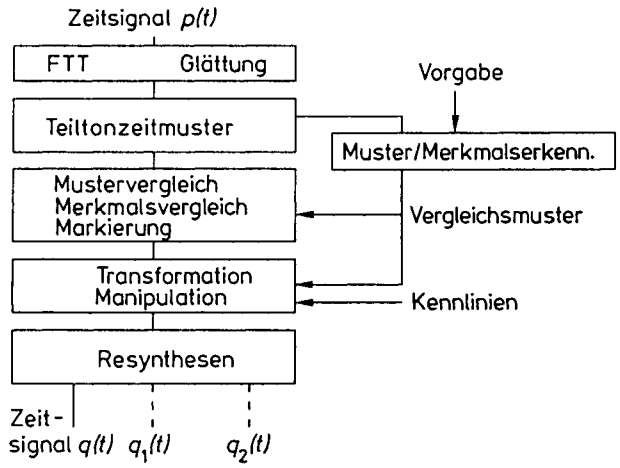


Fig. 5.3–1: Blockschaltbild der Signalverarbeitung mit Hilfe des Teiltonzeitmusters. Erklärung im Text.

5.3.1 Extraktion einfacher Muster

Als einfache Muster werden

a) Teiltöne mit bestimmter Frequenz;
b) Teiltöne mit bestimmtem Pegel;
c) Teiltöne mit bestimmter Dauer;
d) der Frequenzverlauf eines Teiltons über der Zeit oder
e) der Pegelverlauf eines Teiltons über der Zeit

bezeichnet. Die Erkennung all dieser Muster erfolgt nach dem in Fig. 5.3-2 dargestellten Prinzip. Die Darstellung des TTZM erfolgt in einer zweidimensionalen Matrix, ähnlich Fig. 4.1-1 in Abschnitt 4.1, die als Matrix 'A' bezeichnet wird. Jede Zeile dieser Matrix entspricht einem bestimmten Frequenzband, jede Spalte einem bestimmten Auswertezeitpunkt. Die Anzahl und Abstufung der

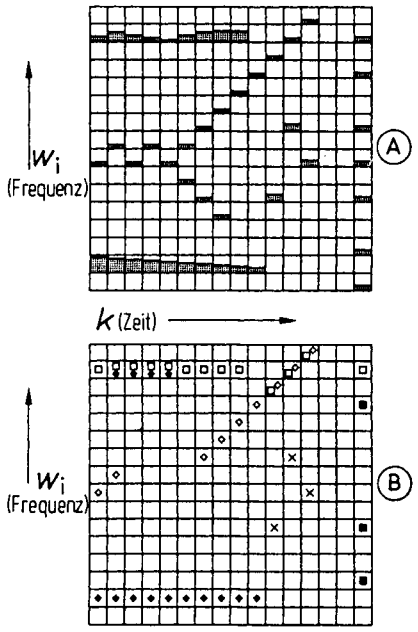


Fig. 5.3-2: Matrizen 'A' und 'B' zur Erkennung und Verarbeitung einfacher Muster.
Matrix 'A': Teiltonzeitmuster.
Jedes gefüllte Kästchen repräsentiert einen Teilton. Die schraffierte Fläche ist ein Maß für den Teiltonpegel.
Matrix 'B': Vergleichsmuster, in dem die Elemente markiert werden, deren zugehöriger Teilton in Matrix 'B' eine bestimmte Bedingung erfüllt (siehe Text).
Symbole in Matrix 'B':
Offene Quadrate: Teiltöne mit bestimmter Frequenz; geschlossene Quadr.: Teiltöne bestimmten Pegels; offene Rauten: bestimmter Frequenzanstieg; geschlossene Rauten: bestimmter Pegelabfall; Kreuze: einzelner Teilton (bestimmte Dauer).

Zeilen bzw. Frequenzbänder hängt vom Frequenzbereich und der notwendigen Genauigkeit ab. Das Zeitintervall, in dem die Mustererkennung stattfinden soll, bestimmt die Anzahl der Spalten. In eine zweite, gleichartige Matrix 'B' werden alle Teiltöne übernommen, die ein bestimmtes Auswahlkriterium erfüllen.

Die Frequenz eines Teiltons kann nur mit einer endlichen Genauigkeit bestimmt werden. Es ist daher sinnvoll, die Mustererkennung in einem bestimmten Toleranzbereich um das interessierende Frequenzband durchzuführen, anstatt auf einen genauen Wert abzuprüfen. Dazu ist es zweckmäßig, die Mittenfrequenzen der Matrixzeilen entsprechend den Analysefrequenzen zu wählen und den Toleranzbereich etwa auf ± 1 Zeile um die gewünschte Zeile herum festzulegen.

Die Bestimmung von Teiltönen bestimmter Frequenz bzw. eines bestimmten Frequenzbereiches ist sehr einfach. Entweder wird direkt der Frequenzwert abgeprüft oder es wird in einem bestimmten Frequenzbereich in 'A' nach Teiltönen gesucht. Werden die Mittenfrequenzen der Matrixzeilen mit w_i, $i=0..i_{max}$ bezeichnet, so wird der Index I der zur gesuchten Frequenz f gehörenden Zeile nach der Vorschrift:

$$I = \mathrm{Min}\left(\left| f - w_i \right| \right) \quad \text{mit } i := 1....i_{max} \tag{5.3.1}$$

bestimmt. Alle Teiltöne der Matrix 'A' im Bereich der Zeile I ± 1 werden in die Matrix 'B' übernommen:

$$B(I \pm 1,\ k) = A(I \pm 1,\ k)\ . \tag{5.3.2}$$

Die übernommen Elemente sind in Fig. 5.3-2 mit offenen Quadraten gekennzeichnet. Bei Resynthese der Teiltöne in Matrix 'B' entspräche dies näherungsweise einer Bandpaßfilterung (Amplitudengang).

Nach einem ähnlichen Prinzip werden Teiltöne mit einem bestimmten Pegel in die Matrix 'B' übernommen. Als Beispiel soll das letzte Teiltonmuster in Fig. 5.3-2 dienen. Alle Teiltöne, deren Pegel größer als ein bestimmter Schwellenwert L_S sind, werden im Matrix 'B' übernommen:

$$B(i,k) = \begin{cases} A(i,k) & ; A(i,k) > L_S \\ 0 \quad \text{sonst} & \end{cases} \quad . \qquad (5.3.3)$$

Die nach dieser Vorschrift übernommenen Teiltöne werden in Fig. 5.3-2 durch geschlossene Quadrate gekennzeichnet. Die Wirkung entspricht in etwa der eines 'noise gate'. Auf die gleiche Weise lassen sich alle die Teiltöne übernehmen, deren Pegel eine Schwelle unterschreiten oder in einen bestimmten Bereich fallen, z.B alle Teiltöne mit Pegeln zwischen 50dB und 60dB. Ebenso können statt absoluten Schwellen auch adaptive Schwellen verwendet werden, in dem z.B die Schwelle L_S in Gl. (5.3.3) in einer bestimmten Relation zum größten vorkommenden Teiltonpegel des jeweiligen Teiltonmusters gewählt wird.

Mit offenen Kreisen sind in Matrix 'B', Fig. 5.3-2 alle Teiltöne markiert, die, ausgehend von einem Startwert, in ihrer Frequenz um jeweils eine Zeile gegenüber dem vorhergehenden ansteigen. Als erstes wird ein Startwert I_S vorgegeben. Erfüllt ein Teilton in Matrix 'A' (Fig. 5.3-2) die Bedingung

$$A(I_S + 1, k) > 0 \quad ; k = 1 \qquad (5.3.4)$$

so wird er in Matrix 'B' übernommen. Anschließend werden I_S und k um eins erhöht. Für einen Teilton im nächsten Teiltonmuster trifft Bedingung (5.3.4) ebenfalls zu, er wird übernommen und I_S und k werden um eins erhöht. In den folgenden vier Teiltonmustern trifft Bedingung (5.3.4) für keinen Teilton zu; es wird nur k erhöht. Bei Erreichen der aufsteigenden Teiltonfolge ist Bedingung (5.3.4) wieder erfüllt, es erfolgt eine Übernahme in Matrix B.

In vielen Fällen ist es notwendig, den zeitlichen Verlauf eines bestimmten Signalanteils zu verfolgen (Tracking). Die aus der Literatur bekannten Methoden sind sehr aufwendig, da bei diesen mit spektralen Verteilungen gearbeitet wird [38]. Mit Hilfe des TTZM ist dies auf folgende Art und Weise möglich. Ausgehend von einem Startwert I_S sucht man im nächsten Teiltonmuster in einem bestimmten Bereich nach einem Teilton. Ist dieser dort vorhanden, so wird er übernommen und seine Frequenz als neuer Startwert verwendet.

Neben zeitlichen Frequenzänderungen lassen sich auch Pegeländerungen erkennen, wie bei den in Fig. 5.3-2 mit geschlossenen Kreisen gekennzeichneten Teiltönen. Hier wurde als Kriterium ein Pegelabfall bei konstanter Frequenz gewählt:

$$B(i,k) = \begin{cases} A(i,k) & ; A(i,k) > A(i,k+1) \\ 0 & \text{sonst} \end{cases} \quad . \qquad (5.3.5)$$

Auf ähnliche Weise lassen sich im Pegel ansteigende Teiltonverläufe mit bestimmten Steigungen des Teiltonpegels bestimmen. Mit solchen Algorithmen wäre vielleicht auch eine Nachhallbefreiung möglich.

Bei beiden bisher beschriebenen Auswahlverfahren von Teiltönen mit bestimmten Eigenschaften spielt der Zeitverlauf der Teiltöne keine Rolle, die Auswahl wird in jeder Spalte der Matrix 'A' unabhängig von den anderen vorgenommen. Bei der Suche nach Teiltönen bestimmter Dauer ist dies jedoch nicht der Fall. Grundsätzlich hat jeder Teilton eine Dauer entsprechend dem Auswerteintervall bzw. einem ganzzahligen Vielfachen davon. Die zeitlichen Grenzen eines Teiltons bzw. eines Teiltonverlaufs hängen auch davon ab, ob und in welchem Maß Frequenzänderungen als zu einem bestimmten Teiltonverlauf gehörend zugelassen werden. Mit Kreuzen sind in Fig. 5.3-2 alle Teiltöne markiert, die nur in einem Teiltonmuster vorhanden sind und denen im Frequenzband k ± 1 kein Teilton vorausgeht oder nachfolgt. Die zugehörige Übernahmebedingung lautet:

$$B(i,k) = \begin{cases} A(i,k) & ; A(i \pm 1, k-1) = 0 \wedge A(i \pm 1, k+1) = 0 \\ 0 & \text{sonst} \end{cases} \quad . \qquad (5.3.6)$$

5.3.2. Veränderung des Teiltonzeitmusters

Jedes Element des TTZMs wird durch drei Parameter beschrieben: Frequenz, Pegel und Dauer. Jeder dieser Parameter läßt sich praktisch unabhängig von den anderen verändern. So gilt für die Veränderung einer Teiltonfrequenz f :

$$f_j^* = X_f(f_j) \quad . \qquad (5.3.7)$$

X_f wird als Frequenzoperator bezeichnet. Dieser kann in beliebiger Weise von anderen Teiltonfrequenzen, -pegeln oder der Zeit abhängen. Im folgenden werden zwei Beispiele für Frequenzoperatoren beschrieben.

Die Verschiebung der Teiltonfrequenzen in Abschnitt 5.1.1.6 erfolgt durch:

$$X_f(f) : \; f + 30\,\text{Hz} \quad . \qquad (5.3.8)$$

Mit (5.3.7) erhält man :

$$f_j^* = f_j + 30\,\text{Hz} \quad . \qquad (5.3.9)$$

Eine zeitabhängige Veränderung der Frequenz erhält man mit dem Operator:

$$X_f(f): \; f \cdot \left(0{,}75 + 0{,}5 \cdot t/2{,}5s\right) \qquad ; 0 < t \leq 2{,}5s \qquad (5.3.10)$$

Wendet man diesen Operator beispielsweise auf Musiksignale an, so ergibt sich der Höreindruck einer langsam anlaufenden Schallplatte, jedoch ohne Veränderung der zeitlichen Merkmale wie z.B. dem Rhythmus.

Für die Veränderung des Pegels gilt in ähnlicher Weise wie für die Frequenz:

$$L_j^* = X_L(L_j) \qquad . \qquad (5.3.11)$$

Auch der Pegeloperator X_L kann in beliebiger Weise von Frequenz, Pegel, usw. abhängen. Beispiele finden sich in Abschnitt 5.1.1.5 bei der Veränderung der Teiltonpegel natürlicher Einzelvokale. Eine Veränderung der spektralen Hüllkurve, wie sie durch Filter stattfindet, erhält man durch die Abhängigkeit des Pegeloperators von der Teiltonfrequenz. Fig. 5.3-3 zeigt eine Bewertung der spektralen Hüllkurve, wie sie mit dem folgenden Pegeloperator erzielt wird:

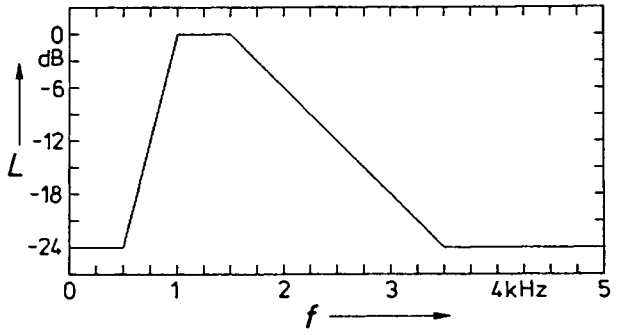


Fig. 5.3-3: Frequenzgang der mit Gl. (5.3.12) erzeugten Bewertung der spektralen Hüllkurve.

$$X_L(L) = \begin{cases} L = \dfrac{f/\mathrm{Hz} - 1000}{125} \cdot 6\mathrm{dB} & ; 500\mathrm{Hz} < f < 1000\mathrm{Hz} \\ L & ; 1000\mathrm{Hz} \leq f < 1500\mathrm{Hz} \\ L = \dfrac{1500 - f/\mathrm{Hz}}{500} \cdot 6\mathrm{dB} & ; 1500\mathrm{Hz} \leq f < 3500\mathrm{Hz} \\ L - 24\mathrm{dB} & ; \text{sonst} \end{cases} \cdot \qquad (5.3.12)$$

Durch Messung der Rauschleistungsdichte eines derart bearbeiteten TTZM von Weißem Rauschen läßt sich zeigen, daß die in Fig. 5.3-3 dargestellte Bewertung tatsächlich erzielt werden kann. Ein Operator gemäß Gl. (5.3.12) läßt sich demnach als Filterung im weitesten Sinne auffassen. Allerdings läßt sich nur der Amplitudengang, also die spektrale Hüllkurve beeinflussen. Da im TTZM keine Phaseninformation enthalten ist, kann über den Phasengang eines solchen Filters keine Aussage gemacht werden.

Eine Veränderung der Zeitachse, die das gesamte TTZM betrifft, ist sehr einfach durchzuführen, indem die Dauer des Syntheseintervalls, die sonst dem

Auswerteintervall entspricht, verändert wird. Die Veränderung der Dauer bestimmter Teiltonmuster erfolgt durch entsprechende Vervielfachung dieser Teiltonmuster. Um bestimmte Teiltöne zu verlängern, müssen diese in nachfolgende Teiltonmuster übernommen werden.

5.3.3 Filterung durch Mustervergleich

Im Gegensatz zu einer globalen Veränderung des Teiltonpegels oder der Teiltonfrequenz, wie in Abschnitt 5.3.1 beschrieben, können im TTZM auch ganz bestimmte Teiltöne oder Teiltongruppen verändert werden. Dies erfolgt durch Mustervergleich, indem Frequenz- oder Pegeloperatoren nur bei den Teiltönen angewandt werden, die sowohl im TTZM als auch in einem Vergleichsmuster vorhanden sind. Das Vergleichsmuster wird entweder vorgegeben oder aus dem TTZM nach den in 5.3.1 beschriebenen Prinzipien gewonnen. Am Beispiel der weitgehenden Trennung der Stimmen zweier Sprecher wird die Filterung durch Mustervergleich beschrieben. Folgende Schritte sind dazu notwendig:

- Bestimmung des Grundfrequenzverlaufs eines Sprechers
- Generierung eines harmonischen Vergleichsmusters auf der Basis dieses Grundfrequenzverlaufes
- Herausfiltern der Teiltöne des Teiltonzeitmusters, die mit denen des Vergleichsmusters korrespondieren.

Die Bestimmung des Grundfrequenzverlaufs eines Sprechers erfolgt durch Anwendung des in Abschnitt 5.3.1 beschriebenen 'Tracking'algorithmus auf das TTZM, dessen Maxigramm in Fig. 5.1-14a (S. 65) dargestellt ist. Es handelt sich dabei um den Satz "Und nun das Wetter in Bayern bis morgen früh" (Männerstimme) und "Die Mensa bietet heute an" (Frauenstimme). Da die Männerstimme immer eine tiefere Grundfrequenz aufweist, kann diese Grundfrequenz näherungsweise durch Verfolgen des Teiltonverlaufs bei etwa 100Hz ermittelt werden. Dieses Verfahren zur Grundfrequenzbestimmung ist jedoch nicht ideal, da infolge der kleinen Analysebandbreite, entsprechend einer großen effektiven Fensterlänge, schnelle Grundfrequenzänderungen, wie sie z.B. bei 'morgen' auftreten,fehlerhaft wiedergegeben werden (vergl. Kap. 4).

Sehr viel besser wäre eine Grundfrequenzbestimmung, die sich an der Grundfrequenzbestimmung des Gehörs orientiert bzw. diese modelliert [82], [85]. Zur Demonstration ist jedoch die Bestimmung aus dem tiefsten Teilton ausreichend.

Der zeitliche Verlauf dieser Teiltonfrequenz wird zur Generierung des in Fig. 5.1-15c dargestellten Vergleichsmusters verwendet, indem zusätzlich 29 Teiltöne, deren Frequenzen ganzzahlige Vielfache der Grundfrequenz sind, hinzugefügt werden. Bei undefinierter Grundfrequenz (Unterbrechungen des Teiltonverlaufs)

werden keine Harmonischen erzeugt. Das Vergleichsmuster enthält somit eine Information über die möglichen Teiltonfrequenzen des männlichen Sprechers bei stimmhafter Anregung, jedoch keine über deren Pegel.

Aus dem ursprünglichen TTZM werden mit Hilfe des Vergleichsmusters zwei neue TTZM (Fig. 5.1-15a und 15b, S. 66) erzeugt. Das Verfahren ist schematisch in Fig. 5.3-4 dargestellt. Wie bei der Mustererkennung in Abschnitt 5.3.1 werden auch hier die einzelnen Teiltöne in die Zeilen einer Matrix übernommen und zum Mustervergleich ein Toleranzbereich von einer Zelle verwendet.

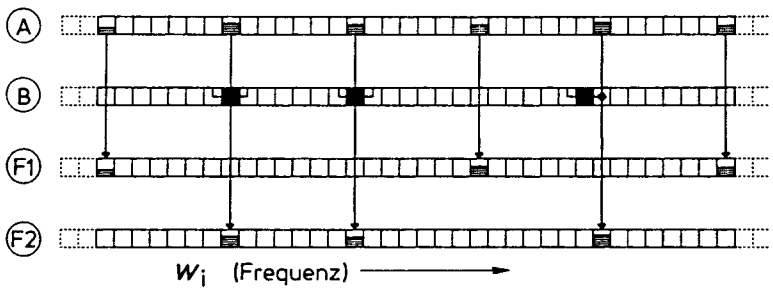


Fig. 5.3-4: Schematische Darstellung der Filterung durch Mustervergleich: Die Teiltöne (gerastert) des Teiltonmusters (TTM) 'A' können dann nach 'F1' gelangen, wenn die entsprechende Zelle im TTM 'B' leer ist. Ansonsten gelangt der Teilton nach 'F2'.

Anstatt der direkten Übernahme in andere TTZM ist es auch möglich, Frequenz- oder Pegeloperatoren bei einer Koinzidenz von TTZM und Vergleichsmuster anzuwenden. Bezogen auf das Beispiel der Stimmentrennung ließen sich durch die Anwendung eines Pegeloperators beispielsweise die Teiltöne, die mit dem harmonischen Vergleichsmuster übereinstimmen, um 10dB dämpfen.

Eine weitere Anwendung der Filterung durch Mustervergleich ist in Fig. 5.3-5a dargestellt. Bei gleichzeitiger Beschallung eines Raumes mit einem bekannten Signal soll der Anteil eines Störgeräusches 'S' bestimmt werden. In der Regel wird der Störanteil in den Pausen des bekannten Signals 'A' ermittelt. Wenn dieses jedoch keine Pausen aufweist, oder der Abstand der Pausen zu groß ist, dann muß das Mikrofonsignal vom Anteil des Signals 'A' befreit werden. Dies ist

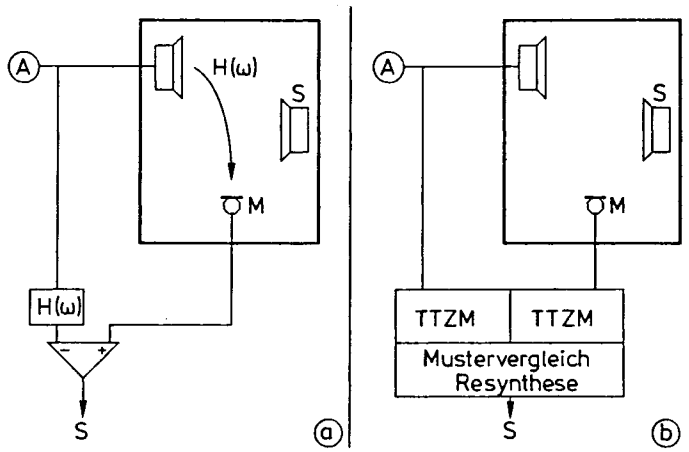


Fig. 5.3-5: Aufgabenstellung der Bestimmung eines Störsignals bei gleichzeitiger Beschallung mit einem bekannten Signal.

auf einfache Weise möglich, wenn die Übertragsfunktion H(ω) zwischen Lautsprechersignal 'A' und Mikrofonsignal exakt bestimmbar und zeitinvariant ist. In den meisten Fällen sind diese Bedingungen jedoch nicht erfüllt.

Eine Lösungsmöglichkeit zur näherungsweisen Bestimmung von 'S' zeigt Fig. 5.3-5b: Es werden die beiden TTZM der Signale 'A' und 'M' berechnet. Durch Vergleich der beiden Muster lassen sich alle Teiltöne in 'M' bestimmen, die nicht in 'A' enthalten sind. Bei entsprechender Wahl des Auswerteintervalls und des Toleranzbereiches können auch Anteile des Raumhalls von 'A' mit einbezogen werden. Die Teiltöne, die nicht von 'A' hervorgerufen werden, repräsentieren dann im wesentlichen den Störer 'S' (einschließlich der Veränderungen im Raum), wenn

- 'S' nicht von 'A' nahezu vollständig verdeckt wird (aufgrund der Eigenschaften der Analyse) und umgekehrt;
- der Hallanteil von 'A' schnell abklingt und nicht in derselben Größenordnung wie 'S' liegt;
- die spektrale Struktur von 'A' weitgehend verschieden von der von 'S' ist.

Fig. 5.3-6 zeigt die Ergebnisse eines Versuchs: Ein digital gespeichertes Musiksignal (Fig. 5.3-6a) und ein Fahrzeuggeräusch (Fig. 5.3-6b) wurden digital gemischt. Dieses Mischsignal wurde in einer schallgedämmten Meßzelle (3x2x2m) über einen Lautsprecher abgestrahlt und gleichzeitig mit einem Mikrofon aufgezeichnet. Das Maxigramm des durch den Raum veränderten Mischsignals zeigt Fig. 5.3-c. Es ist deutlich zu erkennen, wie sich Anteile des Musiksignals und des Fahrgeräusches gegenseitig verdecken. Das Ergebnis des Mustervergleichs ist in Fig. 5.3-6d zu sehen. Das zugehörige TTZM enthält im wesentlichen die Teiltöne des Fahrgeräusches, die durch das Musiksignal nicht verdeckt werden. Durch Resynthese kann diese Festellung auch auditiv bestätigt werden.

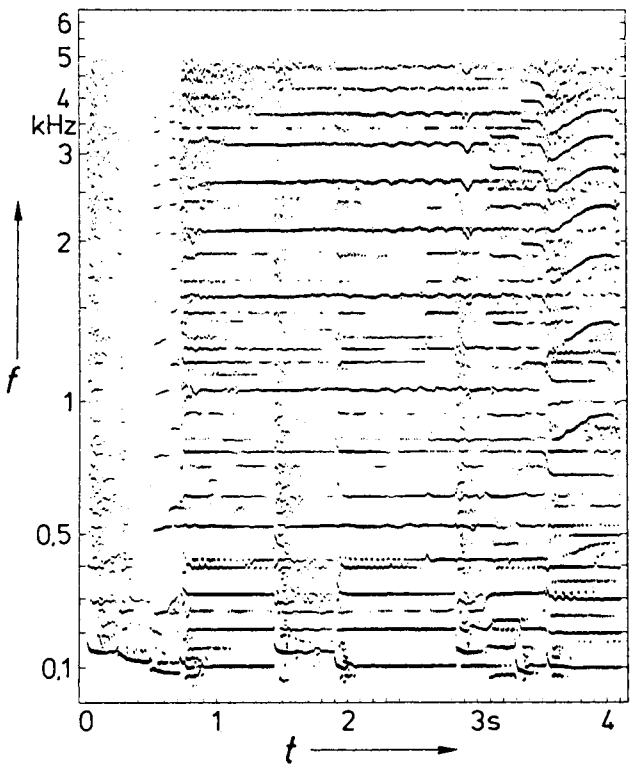


Fig. 5.3–6a: Maxigramm des Musiksignals.

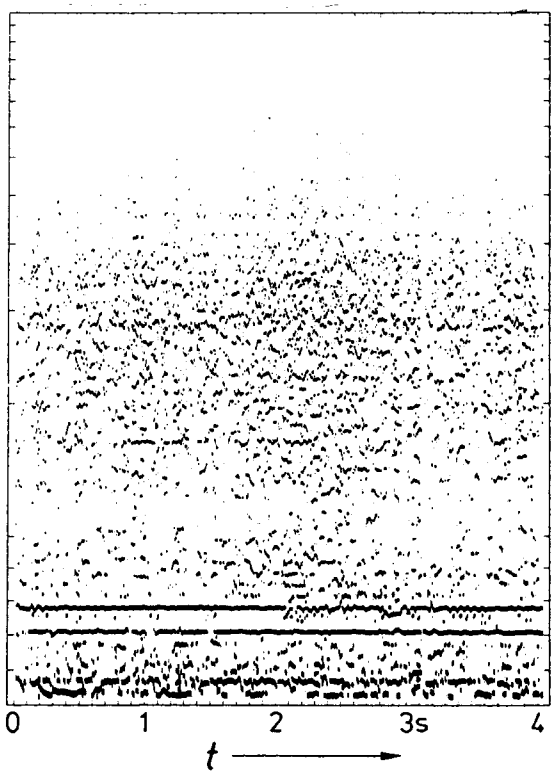


Fig. 5.3–6b: Maxigramm des Fahrgeräusches.

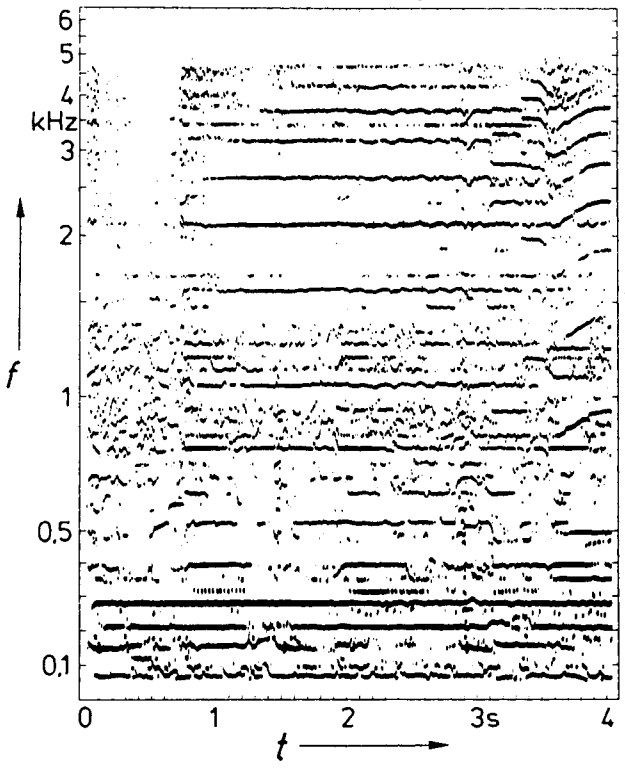


Fig. 5.3–6c: Maxigramm des Mischsignals.

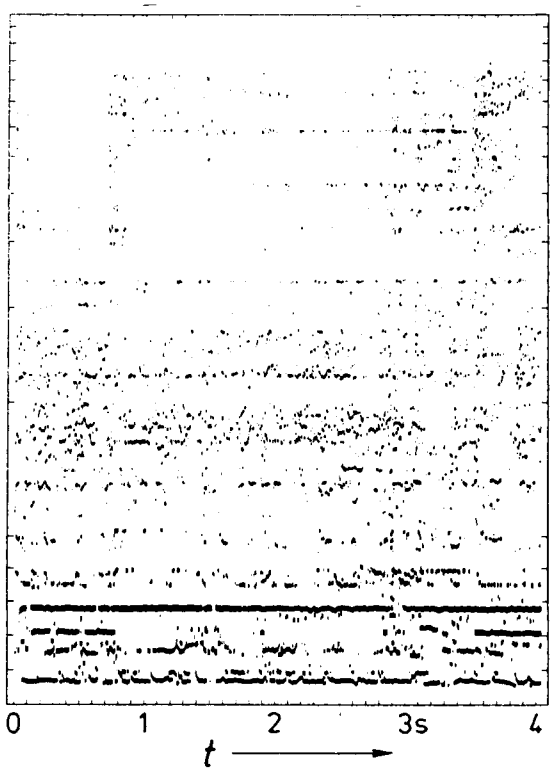


Fig. 5.3–6d: Maxigramm des herausgerechneten Fahrgeräusches.

5.3.4 Diskussion

Wie die letzten Abschnitte (5.3.1 bis 5.3.3) zeigen, läßt sich das TTZM als Ausgangsbasis zur Lösung vieler Aufgabenstellungen bei der Signalverarbeitung verwenden. Dies wird durch die folgenden Eigenschaften ermöglicht:

- Repräsentation des Signals bzw. informationstragender Merkmale durch diskrete Elemente, die Teiltöne;
- sowohl zeitliche als auch spektrale Beschreibung des Signals;
- jeweils unabhängiger Zugriff auf Frequenz, Pegel und Dauer von Teiltönen;
- dadurch einfache Algorithmen zur Bestimmung und Veränderung von Signalanteilen;
- einfache auditive Überprüfung von Verarbeitungsschritten durch Resynthese.

Es muß jedoch betont werden, daß die Möglichkeiten der Signalverarbeitung durch die Eigenschaften der Analyse begrenzt sind. Dazu gehört vor allem die relativ geringe Dynamik innerhalb eines TTMs, wodurch Unterschiede der Teiltonpegel von mehr als etwa 40dB nicht möglich sind.

Eine intelligente Signalverarbeitung, wie sie zur Bildverarbeitung bereits verwendet wird (z.B. Erkennen von Gegenständen in beliebiger Lage und bei teilweiser Verdeckung), scheint mit Hilfe des TTZMs realisierbar zu sein. Die 'Konturisierung', d.h. die Reduktion einer stetigen Verteilung auf eine Linie, spielt eine wesentliche Rolle sowohl in der Bildverarbeitung als auch bei der Bestimmung des TTZMs. Die Repräsentation informationstragender Merkmale durch die Lage und den zeitlichen Verlauf dieser Linien (Teiltöne) legt nahe, ähnliche Algorithmen wie sie in der Bildverarbeitung angewandt werden, zur Verarbeitung des TTZMs zu verwenden.

5.4 Gehörgerechte Repräsentation von Audiosignalen

Aus sämtlichen zuvor beschriebenen Untersuchungen des Teiltonzeitmuster durch Resynthese und auditive Beurteilung geht hervor, daß die wesentliche Information über Audiosignale jeglicher Art in ihm bewahrt wird. Insbesondere ist die Repräsentation im TTZM unabhängig davon, ob es sich um tonale oder geräuschhafte, harmonische oder inharmonische, von einer oder von mehreren Quellen stammende Signale handelt. Die wahrgenommenen Unterschiede zwischen Originalsignal und resynthetisiertem Signal, die beispielsweise zur Abnahme der beurteilten Sprachgüte führen (Abschnitt 5.1.3), sind im wesentlichen eine leichte Halligkeit und Störgeräusche, die dem Knistern einer verschmutzten Schallplatte ähneln.

Die vor allem bei Sprache und percussiven Klängen (Schlagzeug in Abschnitt 5.2) wahrgenommene Halligkeit ist eine Folge der Verschleifung der zeitlichen Struktur, wie sie auch bei den Sinustonimpulsen im Abschnitt 4.3 festgestellt wurde. Selbst unter der Annahme, die Spektralanalyse wiese das gleiche Zeitverhalten wie das Gehör auf, nähme man sehr wahrscheinlich eine Halligkeit wahr. Die nach Analyse und Resynthese veränderte Zeitstruktur würde bei der Analyse durch das Gehör noch weiter verschliffen werden, ähnlich einem Signal, das zwei Tiefpässe gleicher Grenzfrequenz durchläuft. Da die Art des zeitlichen Abfalls des Teiltonpegels - und damit die Halligkeit - vom Wert der Transformationskonstanten a abhängt, erscheint es möglich, Teiltöne dann zu unterdrücken, wenn ihr Pegel einen ganz bestimmten Zeitverlauf aufweist. Dazu wurden jedoch bisher keine Untersuchungen durchgeführt.

Soweit bisher ermittelt werden konnte sind die Störgeräusche eine Folge des Syntheseverfahrens, bei dem bewußt auf die Verwendung von Bewertungsfunktionen zwischen zeitlich aufeinanderfolgenden Syntheseabschnitten verzichtet wird. Sie treten vor allem bei großen Änderungen der Teiltonfrequenzen zwischen zwei Teiltonmustern auf. Bei Verkleinern des Auswerteintervalls T_A unter 1ms verschwinden die Störgeräusche.

Aus den Ergebnissen geht klar hervor, daß das Frequenz- und Zeitauflösungsvermögen der Analyse im wesentlichen den Anforderungen des Gehörs genügt. Ebenso geht aus ihnen hervor, daß die von den Teiltonparametern des TTZM repräsentierte Information für die bei der akustischen Kommunikation wesentlichen Hörwahrnehmungen ausreichend ist. Aus den Verständlichkeitsmessungen mit resynthetisierten, veränderten Vokalrealisationen und der Art und Weise, in der die Datenflußreduktion bei Sprache möglich ist, folgt, daß die genauen Werte der Teiltonfrequenzen eine große , die der Teiltonpegel eine geringere und die der Phasenlage der Teiltöne überhaupt keine Rolle spielen.

Die Teiltöne des Teiltonzeitmusters lassen sich als physikalisches Äquivalent zu den psychoakustisch beobachteten Spektraltonhöhen auffassen, d.h. jeder dieser Teiltöne ist ein Kandidat für eine Spektraltonhöhenwahrnehmung. Die Erhaltung der wesentlichen Information in den (diskreten) Teiltönen des TTZM stützt die These, daß die Spektraltonhöhen das Ergebnis der ersten Stufe der Informationsgewinnung sind. Sie sind somit die Ausgangsbasis für die weitere Informationsverarbeitung im Rahmen der Höheren Wahrnehmungen wie Erkennen von Stimmen und Geräuschen oder Trennen von Schallquellen. Das Teiltonzeitmuster läßt sich somit als ein vergleichsweise einfaches, aber wirkungsvolles Modell der Informationsgewinnung des Gehörs auffassen.

Insbesondere deuten die Ergebnisse darauf hin, daß der Entscheidungsprozeß bei der Teiltonbestimmung der wesentliche Schritt bei der Informationsgewinnung ist. Die Vorverarbeitung in Gestalt der Spektralanalyse kann vergleichsweise einfach, d.h. ohne Nichtlinearitäten und exakte Nachbildung von psychoakustischen oder neurophysiologischen Daten durchgeführt werden, wenn die wesentlichen Eigenschaften des Gehörs bezüglich des Frequenz- und Zeitauflösungsvermögens berücksichtigt werden. Durch Einbeziehung von Nichtlinearitäten kann die Umsetzung des Signals in das TTZM sicherlich noch verbessert werden.

Den zahlreichen Darstellungen des TTZM als Maxigramme sind die wesentlichen Merkmale auch visuell zu entnehmen. Vor allem bei Signalen, die aus mehreren Quellen stammen, wie beispielsweise das Gesangsstück mit Klavierbegleitung in Fig. 5.2-7, ist die Information über die einzelnen Quellen dem Maxigramm sehr gut entnehmbar. Die Sprachlaute können identifiziert und die Frequenzen der Klaviertöne bestimmt werden. Der visuelle Eindruck bestärkt die Auffassung, daß die Spektraltonhöhen bzw. die Frequenzen der Teiltöne die auditiven Konturen des Schalles repräsentieren. Es ist durchaus vorstellbar, daß im zentralen Gehirn Schalle in ähnlicher Weise repräsentiert werden und die Verarbeitung ähnlich der des visuellen Systems erfolgt.

Mit Hilfe des TTZM und der Resynthese lassen sieh Modellvorstellungen der Höheren Verarbeitung auf der Basis des Spektraltonhöhenmusters sehr gut überprüfen, da die Ergebnisse von Verarbeitungsschritten hör- und sichtbar gemacht werden können. Als Beispiel sei hier die Trennung zweier überlagerter Sprachsignale allein aufgrund der Frequenzinformation, wie in den Abschnitt 5.1.2 und 5.3 beschrieben, genannt.

Die Fülle der Anwendungsmöglichkeiten wie Datenreduktion, Quellentrennung oder die Umsetzung akustischer in visuelle Information zeigt, daß das TTZM generell zur informationstechnischen Verarbeitung von Audiosignalen verwendbar ist und nicht nur ein spezielles Verfahren zur Lösung eines genau abgegrenzten Problemes darstellt.

6. Zusammenfassung

Gegenstand der vorliegenden Arbeit ist die Darstellung von beliebigen Audiosignalen durch das Teiltonzeitmuster (TTZM). Es stellt eine informationsreduzierte Beschreibung der für das Gehör wesentlichen Signalmerkmale durch die Frequenzen und Amplituden zeitvariabler Teiltöne ohne Auswertung der Phasenbeziehungen dar. Aus einer im Prinzip sehr großen Menge von Teiltönen enthält das Teiltonzeitmuster nur diejenigen, von denen angenommen wird, daß sie vom Gehör als Träger der wesentlichen Information ausgewertet werden. Diese Teiltöne werden durch Maximumdetektion aus dem zeitlich geglätteten Leistungsspektrum einer gehörangepaßten Kurzzeitspektralanalyse gewonnen. Durch Verständlichkeitsmessungen und qualitative Beurteilungen von resynthetisierten Zeitsignalen (Sprache und Musik) wird nachgewiesen, daß das Teiltonzeitmuster tatsächlich die für die akustische Kommunikation wesentliche Information enthält. Neben einem Verfahren zur Reduktion des Datenflusses von Sprachsignalen werden einige Anwendungen des Teiltonzeitmusters im Bereich der Audiosignalverarbeitung beschrieben.

Die Informationsaufnahme durch das Gehör erfolgt im allgemeinen in einer Umgebung mit mehreren Schallquellen, deren Signale durch die Eigenschaften des Raumes verändert werden. Das Gehör ist in der Lage, die einzelnen Quellen weitgehend zu trennen und die jeweilige Information auch bei starken Verzerrungen durch den Raum aufzunehmen. Es gibt zahlreiche Hinweise, daß das Gehör dadurch an die akustische Umgebung angepaßt ist, daß es vor allem die Frequenzen der Signale auswertet, während die Amplituden eine geringere und die Phasenbeziehungen praktisch keine Rolle spielen. Es wird angenommen, daß die Fähigkeit des Gehörs zur Informationsaufnahme unter Störung auf einer Auswertung der spektralen und zeitlichen Feinstruktur des Gesamtsignals beruht.

Das Teiltonzeitmuster stellt ein Modell der vom Gehör durchgeführten Auswertung der zeitvariablen spektralen Feinstruktur dar. Als psychophysikalische Grundlage dient das von Terhardt entwickelte Konzept der Spektraltonhöhe. Die Spektraltonhöhen lassen sich als 'auditive Konturen' analog zu den visuellen Konturen interpretieren. Die visuellen Konturen tragen bekanntlich die wesentliche Information über Gegenstände. Entsprechend tragen die Spektraltonhöhen die wesentliche Information über die Schallquellen.

Die aus der Literatur bekannten Methoden zur Informationsgewinnung aus Audiosignalen beruhen dagegen fast ausschließlich auf der Auswertung der spektralen Grobstruktur. Es wird versucht, die Signalinformation durch die Parameter einer spektralen Verteilung wie z.B. Formantfrequenzen und -bandbreiten zu repräsentieren. Dadurch geht die Information über verschiedene Quellen im allgemeinen verloren.

Eine wesentliche Voraussetzung für den Erfolg des Teiltonzeitmusters liegt in der Art der Spektralanalyse. Von ihr wird gefordert, daß sie den zeitlichen Änderungen des Signals in dem Maße folgt, wie es dem menschlichen Gehör möglich ist, und zugleich eine vergleichbare spektrale Auflösung besitzt. Die Fourier-t-Transformation (FTT) erfüllt beide Forderungen in hohem Maße. Sie befreit insbesondere von Einschränkungen wie konstante Analysebandbreite und Zeitauflösung der üblicherweise verwendeten Spektralanalyseverfahren mit endlichem Analyseintervall. Dadurch ist eine Anpassung des Frequenz- und Zeitauflösungsvermögens an die Eigenschaften des Gehörs realisierbar. Die ebenfalls angepaßte zeitliche Glättung unterdrückt störende spektrale Nebenmaxima und ermöglicht dadurch erst die Teiltonbestimmung durch Maximumdetektion.

Das Teiltonzeitmuster besteht aus den zeitvariablen Amplituden und Frequenzen von Teiltönen, die genügend ausgeprägten Maxima im zeitlich geglätteten Leistungsspektrum zugeordnet werden. Die graphische Darstellung des Teiltonzeitmusters erfolgt als Maxigramm. Dieses zeigt den Verlauf der Teiltonfrequenzen über der Zeit und ist mit einem Spektrogramm vergleichbar: die Abszisse entspricht der Zeit und die Ordinate der Frequenz. Der Teiltonpegel wird durch die Breite einer Teiltonlinie im Maxigramm angedeutet. Die Untersuchung der Abbildung einfacher Signale im Teiltonzeitmuster zeigt, daß deren wesentliche spektrale und zeitliche Eigenschaften sehr gut wiedergegeben werden. Berechnete Mithör- und Modulationsschwellen dokumentieren die Verträglichkeit des Teiltonzeitmusters mit den entsprechenden Gehöreigenschaften.

Mit Hilfe der Resynthese läßt sich aus dem Teiltonzeitmuster wieder ein Zeitsignal gewinnen, das sich aufgrund der Informationsreduktion im allgemeinen von dem des Originalsignals unterscheidet. Der auditive Vergleich des Originalsignals mit dem aus dem Teiltonzeitmuster resynthetisierten Signal erlaubt eine Beurteilung des verbleibenden Informationsgehalts.

Durch umfangreiche Verständlichkeitsmessungen mit verschiedenen resynthetisierten Vokalrealisationen natürlicher Einzelvokale wird untersucht, ob und welche Anteile des Teiltonzeitmusters die wesentliche Information über die Vokale enthalten. Dabei zeigt sich, daß fünf bis zehn Teiltöne zur Repräsentation der Vokale ausreichen. Werden die fünf gewichtigsten Teiltöne mit jeweils gleichem Pegel dargeboten, so sind Versuchspersonen in der Lage, 70% der Vokalrealisationen richtig zu klassifizieren. Das heißt, daß allein die Frequenzen die wesentlichsten Merkmale enthalten. Die Frequenzbereiche, in denen jene fünf Teiltöne bei allen Vokalrealisationen gehäuft auftreten, weisen eine hohe Übereinstimmung zu den aus der Literatur bekannten Formantgebieten auf.

Eine qualitative Beurteilung eines von unterschiedlichen Sprechern (auch mehrstimmig) in verschiedenen Sprechweisen gesprochenen Testsatzes zeigt, daß die sprechertypischen Merkmale bei ein- und mehrstimmiger Sprache durch das Teiltonzeitmuster erhalten bleiben. Durch die entsprechenden Maxigramme wird dokumentiert, daß eine Vielzahl von Merkmalen visuell zu erkennen sind. Die vollständige Repräsentation fließender Sprache im Teiltonzeitmuster wird anhand von Verständlichkeitsmessungen, durchgeführt mit den resynthetisierten Wortrealisationen eines deutschen Reimtests, quantitativ abgesichert.

Mit Auschnitten ein- und mehrstimmiger Musik wird gezeigt, daß auch die wesentlichen Merkmale musikalischer Signale im Teiltonzeitmuster erhalten bleiben und im Maxigramm sichtbar sind. Wie bei den Sprachsignalen sind zwischen den Originalsignalen und den aus den Teiltonzeitmustern resynthetisierten Musiksignalen nur geringe Unterschiede wahrzunehmen.

Der wichtigste Parameter eines Teiltons ist seine Frequenz. Die Frequenz, der spektrale Ort eines Teiltons, enthält mehr Information als seine Amplitude oder gar seine Phasenlage. Es wird eine Methode zur Datenflussreduktion von Sprachsignalen vorgestellt, die diesen Sachverhalt ausnutzt. Sie beruht auf einer Reduktion der Teiltonanzahl und einer groben Quantisierung sowohl der Teiltonamplituden als auch des zeitlichen Verlaufs der verbleibenden Teiltöne. Die Sprachgüte nimmt durch die Datenreduktion deutlich ab, die Verständlichkeit jedoch bleibt erhalten. Da eine Bestimmung der Sprachgrundfrequenz im Gegensatz zu anderen bekannten Verfahren nicht notwendig ist, treten auch bei komplexen Sprachsignalen, wie dem zweier gleichzeitiger Sprecher, keine Störungen auf.

Einige Anwendungen wie die Störsignalbefreiung oder die Trennung zweier überlagerter Sprachsignale weisen auf die universelle Verwendbarkeit des Teiltonzeitmusters bei der Verarbeitung von Audiosignalen hin. Durch die voneinander unabhängige Veränderung der Parameter Frequenz, Pegel und Dauer der einzelnen Teiltöne lassen sich Signale kontrolliert verändern.

Die Wichtigkeit der Teiltonfrequenz und die Beschreibung der Signalinformation durch relativ wenige Teiltöne bestätigen das von Terhardt vertretene Modell der Spektraltonhöhenwahrnehmung. Insbesondere ist die Interpretation der Spektraltonhöhen als 'auditive Konturen' mit einer den visuellen Konturen vergleichbaren Bedeutung gerechtfertigt.

Das Teiltonzeitmuster stellt eine kategoriale, konturisierte Repräsentation der informationstragenden Merkmale von Audiosignalen dar und bildet damit eine universelle Grundlage zur informationstechnischen Verarbeitung von Audiosignalen.

Abschließend möchte ich mich bei allen Kollegen und Mitarbeitern des Lehrstuhls für Elektroakustik bedanken, die mich bei der Durchführung dieser Arbeit sowohl fachlich als auch moralisch unterstützten. Insbesondere danke ich Herrn Dipl.-Ing. D. Jurzitza für die sorgfältige und kritische Durchsicht des Manuskripts und Frau A. Kabierske für die Beschriftung und fotografische Reproduktion der Bilder.

Durch ihre Tätigkeit als Diplomanden, Studienarbeiter und Werkstudenten wurde mir die Arbeit durch Herrn cand. Ing. U. Baumann, Herrn Dipl.-Ing. S. Buckreuß, Herrn Dipl.-Ing. T. Gülec und Herrn Dipl.-Ing. M. Mummert sehr erleichtert. Ihnen sei herzlicher Dank dafür.

Herrn Prof. Dr.-Ing. E. Zwicker danke ich für die Möglichkeit, die vorliegende Arbeit an seinem Institut durchzuführen sowie für fruchtbare Anregungen und Kritik.

Der Deutschen Forschungsgemeinschaft sei für die Unterstützung dieser Forschungsarbeit im Rahmen des Sonderforschungsbereichs 204 "Gehör" gedankt. Dem Bund der Freunde der Technischen Universität München danke ich für die Bereitstellung eines hochauflösenden Laserdruckers, durch den erst der Ausdruck der Maxigramme in der vorliegenden Form ermöglicht wurde.

Mein ganz besonderer Dank gebührt Herrn Prof. Dr.-Ing. E. Terhardt, in dessen Arbeitsgruppe "Akustische Kommunikation" die vorliegende Arbeit entstand. Ohne seine ständige Gesprächsbereitschaft, seine hilfreichen und fruchtbaren Anregungen und Ideen wäre diese Arbeit in der vorliegenden Form nicht zustandegekommen.

Literatur

[1] **Allen, J.B.** (1977): Short term spectral analysis, synthesis, and modification by discrete Fourier transform. IEEE ASSP-25, 235-238

[2] **Bell,C.G.; Fujisaki,H.;Heinz,J.M.;Stevens,K.N.;House,A.S.** (1961): Reduction of speech spectra by analysis-by-synthesis techniques. J. Acoust. Soc. Am. 33, 1725-1736

[3] **Benedini, K.** (1978): Psychoakustische Messungen der Klangfarbenähnlichkeit harmonischer Klänge und Beschreibung der Zusammenhänge zwischen Amplitudenspektrum und Klangfarbe durch ein Modell. Dissertation Technische Universität München

[4] **Benedini, K.** (1979): Ein Funktionsschema zur Beschreibung von Klangfarbenunterschieden. Biological Cybernetics 34, Springer-Verlag, Berlin, Heidelberg, New York, 111-117

[5] **Bilsen, F.A.** (1977): Pitch of noise signals: evidence for a central spectrum. J. Acoust. Soc. Am. 61, 150-161

[6] **von Bismarck, G.** (1974): Sharpness as an attribute of the timbre of steady sounds. Acustica 30, S. Hirzel Verlag, Stuttgart, 159-172

[7] **Bladon, A.** (1982): Arguments against formants in the auditory representation of speech. 'The representation of speech in the peripheral auditory system' (R. Carlson and B. Granström eds.), Elsevier Biomedical Press, 95-102

[8] **Blomberg, M; Carlson, R.; Elenius, K.; Granström, B.** (1983): Auditory models and isolated word recognition. STL-QPSR 4, 1-15

[9] **Bosquet, J.** (1978): A synthetic model of the monaural auditory function. Biosciences Communications, S. Karger, Basel, 160-174

[10] **Bregman, A.S.** (1978): Auditory streaming: Competition among alternative organizations. Perception & Psychophysics 23, 391-398

[11] **Bregman, A.S.; Pinker, S.** (1978): Auditory streaming and the building of timbre. Canad. J. Psychol. 32, 19-31

[12] **Carlson, R.; Granström, B.** (1982): Towards an auditory spectograph. 'The representation of speech in the peripheral auditory system' (R. Carlson and B. Granström eds.), Elsevier Biomedical Press, 109-114

[13] **Cherry, E.C.** (1953): Some experiments on the recognition of speech, with one and with two ears. J. Acoust. Soc. Am. 25, 975-979

[14] **Childers, D.; Durling, A.** (1975): 'Digital filtering and signal processing.' West Publishing Co., New York

[15] **Cole, R.; Rudnicky, A.; Zue, V.; Reddy, D.R.** (1978): Speech as patterns on paper. 'Perception and production of fluent speech' (R. Cole, ed.), Hillside, NJ, Erlbaum Assoc.

[16] **Cooke, M.P.** (1986): A computer model of peripheral auditory processing incorporating phase-locking, suppression and adaptation effects. Speech Communication 5, 261-281

[17] **Cooley, J.W.; Tukey, J.W.** (1965): An algorithm for the machine calculation of complex Fourier series. Math. Computation 19, 297-301

[18] **Darwin, C.J.** (1981): Perceptual grouping of speech components differing in fundamental frequency and onset-time. Quart. J. Exp Psychology 33A, 185-207

[19] **Darwin, C.J.; Gardener, R.B.** (1987): Perceptual separation of speech from concurrent sounds. Proc. on 'The psychophysics of speech perception' (M.E.H. Schouten ed.), Utrecht, 112-124

[20] **Deutsch, D.** (1982): Grouping mechanisms in music. 'The psychology of music' (D. Deutsch ed.), Academic Press, New York, 99-134

[21] **Dunn, H.K.** (1961): Methods of measuring vowel formant bandwidths. J. Acoust. Soc. Am. 33, 1737-1746

[22] **Fano, R.M.** (1950): Short-time autocorrelation functions and power spectra. J. Acoust. Soc. Am. 22, 546-550

[23] **Fastl, H.** (1971): Über Tonhöhenempfindungen bei Rauschen. Acustica 25, 350-354

[24] **Feldtkeller, M.** (1985): Fourier-t-Transformation als gehörbezogene Spektralanalyse. Diplomarbeit am Lehrstuhl für Elektroakustik der Technischen Universität München

[25] **Fellbaum, K.** (1982): 'Sprachverarbeitung und Sprachübertragung.' Springer Verlag, Berlin, Heidelberg, New York

[26] **Flanagan, J.L.** (1972): 'Speech analysis, synthesis and perception.' 2nd edition, Springer Verlag Berlin, Heidelberg, New York

[27] **Fletcher, H.; Galt, R.H.** (1950): The perception of speech and its relation to telephony. J. Acoust. Soc. Am. 22, 89-151

[28] **Fraisse, P.** (1975): Is rhythm a gestalt? 'Gestalttheorie in der modernen Psychologie' (S.Ertel, L. Kemmler, Stadler eds.), Steinkopff, Darmstadt, 227-232

[29] **Gambardella, G.** (1968): Time scaling and short-time spectral analysis. J. Acoust. Soc. Am. 44, 1745-1747

[30] **Gambardella, G.** (1971): A contribution to the theory for short-time spectral analysis with nonuniform bandwidth filters. IEEE Transactions on circuit theory CT-18, 4, 455-460

[31] **Grey, J.M.; Moorer, J.A.** (1972): Perceptual evaluations of synthesized musical instrument tones. J. Acoust. Soc. Am. 62, 454-462

[32] **Hedelin, P.** (1982): A representation of speech with partials. 'The representation of speech in the peripheral auditory system' (R. Carlson and B. Granström eds.), Elsevier Biomedical Press, 247-250

[33] **Heinbach, W.** (1986): Untersuchung einer gehörbezogenen Spektralanalyse mittels Resynthese. 'Fortschritte der Akustik - DAGA '86', Bad Honnef, DPG-GmbH, 453-456

[34] **Heinbach, W.** (1987): Verständlichkeitsmessungen mit datenreduzierten natürlichen Einzelvokalen. 'Fortschritte der Akustik - DAGA '87', Bad Honnef, DPG-GmbH, 665-668

[35] **Heinbach, W.** (1987): Datenreduktion von Sprache unter Verwendung von Gehöreigenschaften. ntz-Archiv 9, 327-333

[36] **Hess, W.** (1972): Digitale grundfrequenzsynchrone Analyse von Sprachsignalen als Teil eines automatischen Spracherkennungssystems. Dissertation Technische Universität München

[37] **Höge, H.** (1984): New filter design for a channel vocoder based on the perceptual properties of the human ear. Siemens Forschungs- u. Entwickl.-Berichte. Bd. 13, Springer Verlag Berlin, Heidelbg., NY, 68-73

[38] **Hsu, F.M.; Giordano, A.A.** (1977): Line tracking using autoregressive spectral estimates. IEEE ASSP-25, 510-519

[39] **Kates, J. M.** (1983): An auditory spectral analysis model using the chirp z-transformation. IEEE Transactions ASSP 31, 148-156

[40] **Kay, S. M.; Marple, S. L. jr.** (1981): Spectrum analysis - a modern perspective. Proceedings of the IEEE, Vol 69, 1380-1419

[41] **Klatt, D.H.** (1982): Speech processing strategies based on auditory models. 'The representation of speech in the peripheral auditory system' (R. Carlson and B. Granström eds.) Elsevier Biomedical Press, 181-196

[42] **Klatt, D.H.** (1986): Representation of the first formant in speech recognition and in models of the auditory periphery. Proc. of Montreal symposium on speech recognition, Canad. Acoust. Assoc., 5-7

[43] **Klemm, R.** (1977): Zur rekursiven Berechnung der diskreten Fouriertransformation. ntz 30, 159

[44] **Koenig, W.; Dunn, H.K.; Lacey, L.J.** (1946): The sound spectrograph. J. Acoust. Soc. Am. 18 , 19-49

[45] **Krahé D.** (1986): Ein Verfahren zur Datenredukton bei digitalen Audiosignalen unter Ausnutzung psychoakustischer Phänomene. Rundfunktechnische Mitteilungen, 117-123

[46] **Krüger, E.; Strube, H.W.** (1986): Adaptive prädiktive Sprachkodierung mit Anpassung an die Barkskala. 'Fortschritte der Akustik - DAGA '86', Bad Honnef, DPG-GmbH, 509-512

[47] **Lovis, G.** (1986): Untersuchungen zu datenreduzierten Codierverfahren für diskrete Musiksignale. Dissertation Bergische Universität - Gesamthochschule Wuppertal

[48] **Marko, H.** (1977): 'Methoden der Systemtheorie.' Springer Verlag, Berlin, Heidelberg, New York

[49] **Marko, H.** (1981): Informationstheorie und Kommunikationstheorie. Frequenz 35(1) 2-7

[50] **McAdams, S.** (1984): The auditory image. A metaphor for musical and psychological research on the auditory organization. 'Cognitive processes in the perception of art' (Crozier & Chapman eds.), Elsevier Science Publishers B.V. (North Holland), 289-323

[51] **Meyer-Eppler, W.** (1957): Realization of prosodic features in whispered speech. J. Acoust. Soc. Am. 29, 104-106

[52] **van Noorden, L.P.A.S** (1975): Temporal coherence in the perception of tone sequences. Doctoral thesis Technische Hogeschool Eindhoven

[53] **Papoulis, A.** (1986): 'Signal analysis.' International Student Edition, McGraw-Hill Book Company, Singapore

[54] **Parsons, T.W.** (1976): Separation of speech from interfering speech by means of harmonic selection. J. Acoust. Soc. Am. 60, 911-918

[55] **Patterson, R.D.** (1973): The effects of relative phase and the number of components on residue pitch. J. Acoust. Soc. Am. 53, 1565-1572

[56] **Pfeiffer, B.H.; Sotschek, J.** (1984): Versuch zur Sprachaudiometrie bei Lärmschwerhörigkeit mit einem Reimtest aus der Nachrichtentechnik. BIA-Report 1/84 Berufsgenossenschaftliches Institut für Arbeitssicherheit - BIA, Sankt Augustin

[57] **Plomp, R.** (1964): The ear as a frequency analyzer. J. Acoust. Soc. Am. 36, 1628-1636

[58] **Plomp, R.; Steeneken, H.J.M.** (1969): Effect of phase on the timbre of complex tones. J. Acoust. Soc. Am. 69, 409-421

[59] **Popper, K.R.; Eccles, J.C.** (1984): 'Das Ich und sein Gehirn.', Piper Verlag, München

[60] **Portnoff, M.R.** (1980): Time-frequency representation of digital signals and systems based on short-time Fourier analysis. IEEE ASSP-28, 55-69

[61] **Portnoff, M.R.** (1981): Short-time Fourier analysis of sampled speech. IEEE ASSP-29, 364-373

[62] **Portnoff, M.R.** (1981): Time-scale modification of speech based on short-time Fourier analysis. IEEE ASSP-29, 374-390

[63] **Rasch, R.A.** (1978): Perception of simultaneous notes such as in polyphonic music. Acustica 40, 21-33

[64] **Risset, J-C.; Wessel, D.L.** (1982): Exploration of timbre by analysis and synthesis. 'The psychology of music' (D. Deutsch ed.), Academic Press, New York, 25-58

[65] **Scheffers, M.T.M** (1983): Sifting vowels: auditory pitch analysis and sound segregation. Dissertation Universität Groningen

[66] **Schroeder, M.R.; Atal, B.S.** (1962): Generalized short-time power spectra and autocorrelation functions. J. Acoust. Soc. Am. 34, 1679-1683

[67] **Schroeder, M.R.; Kuttruff, H.** (1962): On frequency response curves in rooms. Comparison of experiment, theoretical, and Monte Carlo results for the average frequency spacing between maxima. J. Acoust. Soc. Am. 34, 76-80

[68] **Searle, C.L.** (1982): Speech perception from an auditory and visual viewpoint. Canad. Journ. Psychol. 36, 402-409

[69] **Seneff, S.** (1984): Pitch and spectral estimation of speech based on auditory synchrony model. Working papers IV, May 1984, Speech communication group, Res. lab of Electronics, MIT, 43-56

[70] **Shannon, C.E.** (1948): Mathematical theory of communication. Bell Syst. Techn. Journal 27, 379-423, 623-652

[71] **Sotschek, J.** (1986): Sprachgüteuntersuchungen an einem Sprachsynthetisator mittels Reimtest-Verständlichkeitsmessungen. Proc. of 6.FASE - Sopron, Hungary, 189-192

[72] **Sreenivas, T.V.; Rao, P.V.S** (1982): Analysis of non-stationary voiced segments in speech signals. 'The representation of speech in the peripheral auditory system' (R. Carlson and B. Granström eds.), Elsevier Biomedical Press, 235-240

[73] **Stoll, G.** (1980): Psychoakustische Messungen der Spektraltonhöhenmuster von Vokalen. 'Fortschritte der Akustik - DAGA '80', VDE-Verlag, Berlin, 631-634

[74] **Stoll, G.** (1982): Spectral-pitch pattern. 'Music, mind and brain, The neuropsychology of music' (M. Clynes ed.), Plenum Press - New York, 271-278

[75] **Stoll, G.; Theile, G.** (1986): Neue digitale Tonübertragungsverfahren: Wie erfolgt die Beurteilung der Tonqualität?. "Bericht 14. Tonmeistertagung, München", Bildungswerk des Verbandes Deutscher Tonmeister, Berlin, 472-495

[76] **Strawn, J.** (1987): Editing time-varying spectra. Journ. Audio Eng. Soc. 35, 337-352

[77] **Syrdal, A.K.; Gopal, H.S.** (1986): A perceptual model of vowel recognition based on the auditory representation of English vowels. J. Acoust. Soc. Am. 79, 1086-1100

[78] **Terhardt, E.** (1968): Über die durch amplitudenmodulierte Sinustöne hervorgerufene Hörempfindung. Acustica 20, 210-214

[79] **Terhardt, E.** (1972): Zur Tonhöhenwahrnehmung von Klängen, I. Psychoakustische Grundlagen. Acustica 26, 173-186 und Zur Tonhöhenwahrnehmung von Klängen; II. Ein Funktionsschema. Acustica 26, 187-199

[80] **Terhardt, E.** (1974): On the perception of periodic sound fluctuations (roughness). Acustica 30, 201-213

[81] **Terhardt, E.** (1974): Pitch, consonance, and harmony. J. Acoust. Soc. Am. 55, 1061-1069

[82] **Terhardt, E.** (1979): Calculating virtual pitch. Hearing Research, 1, Elsevier/North Holland Biomedical Press, 155-182

[83] **Terhardt, E.** (1979): On the perception of spectral information in speech. 'Experimental brain research supplementum II: Hearing mechanisms and speech' (Creutzfeld,O.; Scheich,H.; Schreiner,Chr. eds.), Springer Verlag, Berlin, Heidelberg, New York, 281-291

[84] **Terhardt, E.** (1981): Sprachsignal, Sprechvorgang und Hörwahrnehmung: Eine vergleichende Übersicht. Audiologische Akustik 20, 96-126

[85] **Terhardt, E.; Stoll, G.; Seewann, E.** (1982): Algorithm for extraction of pitch and pitch salience from complex tonal signals. J. Acoust. Soc. Am. 71, 679-688

[86] **Terhardt, E.; Seewann, M.** (1984): Auditive und objektive Bestimmung der Schlagtonhöhe von historischen Kirchenglocken. Acustica 54, 129-144

[87] **Terhardt, E.** (1985): Fourier transformation of time signals: conceptual revision. Acustica 57, 242-256

[88] **Terhardt, E.** (1985): Some psycho-physical analogies between speech and music. 'Music in medicine' (R. Spintge, R. Droh eds.), Mayer Verlag GmbH, Miesbach, 89-102

[89] **Terhardt, E.** (1985): Verfahren zur gehörbezogenen Frequenzanalyse. 'Fortschritte der Akustik - DAGA '85', Bad Honnef, DPG-GmbH, 811-814

[90] **Terhardt, E.** (1986): Pitch perception and frequency analysis. Proc. of 6. FASE - Sopron Hungary, 221-228

[91] **Terhardt, E.** (1987): Gestalt principles and music perception. 'Auditory processing of complex sounds.' (W.A. Yost, C.S. Watson eds.) Lawrence Erlbaum Assoc. Hillsdale, USA, 157-166

[92] **Terhardt, E.** (1987): Psychophysics of audio signal processing and the role of pitch in speech. Proc. on 'The psychophysics of speech perception' (M.E.H. Schouten ed.), Utrecht, 271-283

[93] **Theile, G.; Stoll, G.** (1987): Low-bit rate coding of high quality audio signals. Proc. of 82nd Convention of AES, London

[94] **Weintraub, M.** (1987): Sound separation and auditory perceptual organization. Proc. on 'The psychophysics of speech perception' (M.E.H. Schouten ed.), Utrecht, 125-134

[95] **Zollner, M.** (1979): Verständlichkeit der Sprache eines einfachen Vocoders. Acustica 43, 272-272

[96] **Zwicker, E.; Feldtkeller, R.** (1967): 'Das Ohr als Nachrichtenempfänger.' 2. Auflage, Hirzel, Stuttgart

[97] **Zwicker, E.; Schütte, H.** (1973): On the time pattern of the threshold of tone impulses. Acustica 29, 343-347

[98] **Zwicker, E.** (1974): Die Zeitkonstanten (Grenzdauern) des Gehörs. Zeitschrift für Hörgeräte-Akustik 13, 82-102

[99] **Zwicker, E.; Terhardt, E.; Paulus, E.** (1979): Automatic speech recognition using psychoacoustic models. J. Acoust. Soc. Am. 65, 487-498

[100] **Zwicker, E.; Terhardt, E.** (1980): Analytical expressions for critical-band rate and critical bandwith as a function of frequency. J. Acoust. Soc. Am. 68, 1523-1525

[101] **Zwicker, E.** (1982): 'Psychoakustik.' Springer Verlag, Berlin, Heidelberg, New York

[102] **Zwicker, E.** (1986): Peripheral preprocessing in hearing and psychoacoustics as guidelines for speech recognition. 'Units and their representation in speech recognition', Proc. of Montreal symposium on speech recognition, Canad. Acoustic Assoc., 1-4

[103] **Zwicker, U.T.** (1984): Auditory recognition of diotic and dichotic vowel pairs. Speech communication 3, 265-277

Verzeichnis häufig verwendeter Formelzeichen und Abkürzungen

a	Transformationskonstante
a_i	Transformationskonstante bei der Analysefrequenz f_i
A	Amplitude (Spannung)
$A(i,k)$	Matrixelement Teiltonzeitmuster
B	Analysebandbreite
B_F	Formantbandbreite
$B(i,k)$	Matrixelement Vergleichsmuster
d	Datenrate
ΔL_A	Ausgeprägtheitsschwelle
Δf, $\Delta\omega$	Frequenzabstand, Kreisfrequenzabstand
Δf_G	Frequenzgruppenbreite
Δt	Zeitdifferenz
Δz	Abstand der Analysefrequenzen in Bark
f	Frequenz
f_i	Analysefrequenz
f_j	Teiltonfrequenz
f_G	Grenzfrequenz
f_T	Testtonfrequenz
FFT	Fast-Fourier-Transformation
$F_S(\omega)$	stationärer Anteil des FTT-Leistungsspektrums
FTT	Fourier-t-Transformation
$F_T(\omega,t)$	transienter Anteil des FTT-Leistungsspektrums
$F(\omega,t)$	FTT-Leistungsspektrum
φ	Phasenwinkel
$g(t)$	zeitliche Bewertungsfunktion
G	Spektralgewicht
$G(\omega,t)$	zeitlich geglättetes FTT-Leistungsspektrum
i	laufende Nummer von Analysefrequenz oder Frequenzband
j	laufende Nummer eines Teilton in einem Teiltonmuster
k	laufende Nummer eines Teiltonmusters
L	Pegel
L_A	Ausgeprägtheit
L_E	Erregungspegel
L_j	Teiltonpegel
L_T	Testtonpegel
LX	Überschußpegel
m	Anzahl der Teiltöne pro Teiltonmuster

n	laufende Nummer der Abtastzeitpunkte
n_f	Wortlänge Frequenz
n_L	Wortlänge Pegel
NVR	natürliche Vokalrealisation
Ω	Korrekturphase
$p(t)$	Zeitsignal
$p_T(t)$	Zeitsignal Testton
$P(\omega,t)$	Kurzzeitspektrum (komplex)
$q(t)$	resynthetisiertes Zeitsignal
$Q(\omega,t)$	modifiziertes Kurzzeitspektrum
RVR	resynthetisierte Vokalrealisation
Θ	Synthesephase
t	Zeit
t_a	Beobachtungszeitpunkt
T	Länge des effektiven Analyseintervalls
T_A	Auswerteintervall
T_G	Glättungszeitkonstante
T_{G_i}	Glättungszeitkonstante bei der Analysefrequenz f_i
T_P	Impulsdauer
T_s	Abtastintervall
TTM	Teiltonmuster
TTZM	Teiltonzeitmuster
v	Übernahmebereich der Synthesephase
ω	Kreisfrequenz $2\pi f$
w_i	Analysefrequenz oder Frequenzband $2\pi f_i$
X_f	Frequenzoperator
X_L	Pegeloperator
z	Tonheit

Anhang

Der Anhang besteht aus:

I. Rekursive Berechnung des geglätteten Leistungsspektrums

Die Berechnung der Spektren $Q(w_i,nT_s)$, $F(w_i,nT_s)$ und $G(w_i,nT_s)$ kann gemäß den folgenden Rekursionsgleichungen erfolgen:

$$\begin{aligned} Re(w_i,nT_s) &= X \cdot Re[w_i,(n-1)T_s] - Y \cdot Im[w_i,(n-1)T_s] + C \cdot p(nT_s) \\ Im(w_i,nT_s) &= X \cdot Im[w_i,(n-1)T_s] + Y \cdot Re[w_i,(n-1)T_s] \\ F(w_i,nT_s) &= Re(w_i,nT_s) \cdot Re(w_i,nT_s) + Im(w_i,nT_s) \cdot Im(w_i,nT_s) \\ G(w_i,nT_s) &= D \cdot F(w_i,nT_s) + E \cdot G[w_i,(n-1)T_s] \end{aligned} \qquad (I.1)$$

wobei

$$\begin{aligned} X &= e^{-a_i T_s} \cdot \cos(w_i T_s) \\ Y &= e^{-a_i T_s} \cdot \sin(w_i T_s) \\ C &= 2a_i T_s \\ D &= T_s / T_{G_i} \\ E &= e^{-D} \\ Re(w_i,nT_s) &= \mathrm{Re}\{Q(w_i,nT_s)\} \\ Im(w_i,nT_s) &= \mathrm{Im}\{Q(w_i,nT_s)\} \\ w_i &= 2\pi f_i \end{aligned} \qquad (I.2)$$

und

$$Re(w_i,0)=0,\ Im(w_i,0)=0,\ F(w_i,0)=0,\ G(w_i,0)=0 \quad . \qquad (I.3)$$

II. Berechnung der Transformationsparameter

Die Werte der Analysefrequenzen f_i bzw. w_i im Bereich $i=2...i_{max}$ bei Verteilung entlang der Barkskala erhält man ausgehend von einem Startwert mit beispielsweise 20Hz gemäß

$$f_i = f_{i-1} + \frac{\Delta z}{Bark} \cdot \Delta f_G(f_{i-1}), \qquad w_i = 2\pi f_i \qquad (II.1)$$

mit

$$\Delta f_G(f) = 25 + 75 \cdot \left[1 + 1{,}4 \cdot \left(\frac{f}{kHz}\right)^2\right]^{0{,}69} Hz \qquad (II.2)$$

und

$$f_1 = 20Hz \quad . \qquad (II.3)$$

Mit Gl. (II.2) wird näherungsweise die einem Bark entsprechende Frequenzgruppenbreite in Abhängigkeit von der Frequenz f berechnet [100]. Der Faktor Δz in Gl. (II.1) bestimmt den Abstand der Analysefrequenzen zueinander. Für den in Tab. 3.4.1 angegebenen Frequenzabstand gilt $\Delta z = 0{,}05\text{Bark}$.

Die zur Analysefrequenz f_i bzw. w_i zugehörige Transformationskonstante a_i wird nach der Formel

$$a_i = \frac{B}{\text{Bark}} \cdot \pi \cdot \Delta f_G(f_i) \qquad \text{(II.4)}$$

berechnet. Gemäß Tab. 3.4.1 gilt B=0,1Bark. Die Berechnung der Glättungszeitkonstanten T_G an der Stelle der Analysefrequenz f_i entsprechend den Angaben in Tab. 3.4.1 erfolgt mit

$$T_{G_i} = \begin{cases} 0{,}2/a_i & ;f_i \le 3\text{kHz} \\ 1{,}25\text{ms} & ;f_i > 3\text{kHz} \end{cases} . \qquad \text{(II.5)}$$

III. Analoge Bestimmung des geglätteten FTT-Leistungsspektrum

Neben der digitalen rekursiven Berechnung besteht auch die Möglichkeit, das zeitlich geglättete FTT-Leistungsspektrum mit der in Fig. III-1 angegebenen Anordnung zu berechnen (s.a [26], [66]).

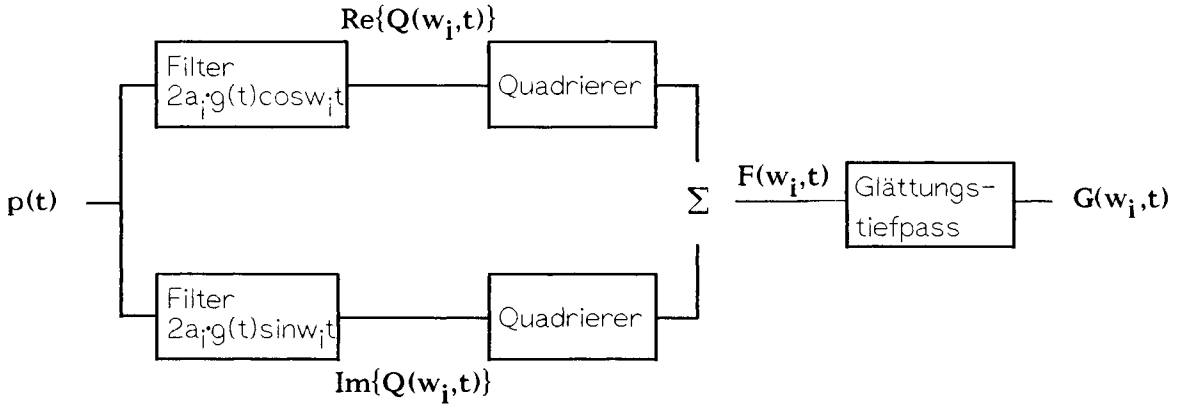


Fig. III-1: Analoge Bestimmung des zeitlichen geglätteten Leistungsspektrums $G(w_i,t)$. Die Bewertungsfunktion $g(t)$ entspricht der Bewertungsfunktion $g(x)$ in Kapitel 3, $Q(w_i,t)$ ist das modifizierte Kurzzeitspektrum von Seite 24, a_i ist die Transformationskonstante bei der Analysefrequenz w_i.

IV. Quellenangabe der Musikstücke in Kapitel 5.2

Im folgenden sind die Quellen der Musikstücke, von denen jeweils ein Ausschnitt als Maxigramm in Kapitel 5.2 abgebildet ist, bis auf die der Paganini-Variation (Fig. 5.2-1) angegeben. Das letztgenannte Musikstück stammt von einer Tonbandaufnahme, über die keine näheren Einzelheiten bekannt sind.

Fig. 5.2-2:
Interpret: Glenn Gould
Titel: Variation 1
aus: "J.S. Bach: The Goldberg Variations, BWV 988 (Aria & 30 Variations)", Compact Disc (CD) CBS Nr. 37779

Fig. 5.2-3:
Interpret: Pepe Romero
Titel: Rosita von Francisco Tarrega
aus: "Recuerdos de La Alhambra Jeux interdits Asturias", CD Philips Nr. 411 033-2

Fig. 5.2-4:
Interpret: Marie-Claire Alain
Titel: Toccata & Fuge in d-moll
aus: "J.S. Bach: Toccata & Fugue, Passacaglia", CD Errato ECD 88004

Fig. 5.2-5 und 5.2-8
Interpret: Mönchsschola der Erzabtei St. Ottilien
Leitung: P. Johannes Berchmans Göschl OSB
Titel: Dritter Adventssonntag - Vesper
aus: "Gregorianische Gesänge", CD Calig Nr. 50 858

Fig. 5.2-6:
Interpret: Pat Metheny Group
Titel: Barcarole
aus: "Offramp", CD ECM Nr. 817 138-2

Fig. 5.2-7:
Interpret: Margaret Price, Sopran und James Lockhart, Piano
Titel: Die Kartenlegerin
aus: "Robert Schumann: Ausgewählte Lieder", CD Orfeo Nr. C 031821 A

Fig. 5.2-9:
Interpret: Orpheus Chamber Orchestra
Titel: Eine kleine Nachtmusik in G major, K. 525
aus: "Wolfgang A. Mozart: Eine kleine Nachtmusik"
CD Deutsche Grammophon Nr. 419 192-2

Fig. 5.2-10:
Interpret: Sade
Titel: Is it a crime
aus: "Promise", CD EPIC Nr. 86318