Lehrstuhl für Mensch-Maschine-Kommunikation
der Technischen Universität München

Sprachcodierung durch Konturierung eines gehörangepaßten Spektrogramms und ihre Anwendung zur Datenreduktion

Markus Mummert

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.techn. J. Swoboda

Prüfer der Dissertation:

1. Univ.-Prof. Dr.-Ing. E. Terhardt
2. Univ.-Prof. Dr.-Ing. J. Hagenauer

Die Dissertation wurde am 23.5.1997 bei der Technischen Universität München eingereicht und durch die Fakultät für Elektrotechnik und Informationstechnik am 27.11.1997 angenommen.

Für die Hilfsbereitschaft und die angenehme, freundschaftliche Atmosphäre in unserer Arbeitsgruppe möchte ich mich bei meinen Kollegen, den Herren Dr.-Ing. U. Baumann, Dr.-Ing. W. Heinbach, Dr.-Ing. K. Pfaffelhuber, Dipl.-Ing. C. von Rücker, Dr.-Ing. S. Wartini und bei meiner Kollegin, Frau Dipl.-Ing. M. Valenzuela, bedanken. Dieser Dank gebührt auch allen anderen Mitarbeitern des Lehrstuhls für Mensch-Maschine-Kommunikation und des damaligen Lehrstuhls für Elektroakustik.

Weiterhin gilt mein Dank Herrn Dipl.-Ing. E. Weidinger für die Programmierung der Trennung von tonalen und nichttonalen Anteilen in den Frequenzkonturen. Nicht nur für meine Arbeit hat er damit einen wichtigen Beitrag geleistet. Bei Herrn O. Riesener bedanke ich mich für die Hilfe bei der Einrichtung einer eigenen UNIX-Workstation.

Mein Freund und Kollege Dipl.-Ing. T. Horn hat nicht nur das Manuskript mit großer Sorgfalt durchgesehen. Er ist auch ein überaus wichtiger Gesprächspartner gewesen. Sein Interesse für die Einzelheiten dieser Arbeit war für mich ein wesentlicher Ansporn, und seine Fähigkeit zur klaren Reflektion komplizierter Sachverhalte wie auch seine eigenen Erkenntnisse haben diese Arbeit ohne Zweifel beeinflußt. Ich bin ihm sehr dankbar.

Herrn Prof. Dr.-Ing. J. Hagenauer danke ich für die Übernahme des Mitberichts.

Mein besonderer Dank gilt Herrn Prof. Dr.-Ing. E. Terhardt, der die Dissertation über den langen Zeitraum betreut hat. Von seinen Modellen und Denkansätzen habe ich sehr profitiert, auch über den fachlichen Horizont hinaus. Ohne seine Anregungen wäre die Arbeit in der vorliegenden Form nicht entstanden.

Bei Frau Caroline Hohmann M.A.phil. bedanke ich mich herzlich für die abschließende gründliche Durchsicht des Manuskripts.

Ein Teil dieser Arbeit wurde von der Deutschen Forschungsgemeinschaft im Rahmen des Sonderforschungsbereiches 204 ‘Gehör’ gefördert.

I wish to thank the countless people on the Net who provided and helped to provide the excellent free software that I have relied on throughout my work.

http://www.mmk.e-technik.tu-muenchen.de/~rue/mum/
http://home.t-online.de/home/Markus.Mummert/

Meinem Großvater
Dr.-Ing. Walter Schilling

Inhaltsverzeichnis

Einleitung

Codierverfahren für Audiosignale, zu denen Sprachsignale zählen, sind in der Informationstechnik wichtig für die Übertragung und Aufzeichnung. Codierung bedeutet zunächst, daß das von Natur aus kontinuierliche, 'analoge' Schallsignal vorübergehend in einen Strom von diskreten, 'digitalen' Daten gewandelt wird. Um teure Übertragungs- und Speicherkapazität einzusparen, befreit man den Strom mit Hilfe von Rechenverfahren von redundanten und irrelevanten Daten. Die vorliegende Arbeit befaßt sich mit den Grundlagen und der Anwendung eines speziellen Rechenverfahrens, das sich an einem Modell der Informationsverarbeitung im menschlichen Gehör orientiert. Eine solche Orientierung hilft, diejenigen Daten auszuwählen, die für den Zuhörer wesentlich sind.

Ähnlich einem Prisma zerlegt die Basilarmembran im Innenohr das Schallsignal in seine spektralen Bestandteile. Beispielsweise werden Sinustöne verschiedener Frequenz, wenn auch verschwommen, auf verschiedene Orte der Membran abgebildet. Das Ergebnis dieses Vorgangs nähert ein spezielles *Spektrogramm* an, das man sich nach Bild 1 als 'Gebirge' vorstellen kann. Seine Grundfläche wird durch eine Zeit- und eine Frequenzachse aufgespannt, seine Höhe gibt den Energiedichtepegel des Signals über diesen beiden Dimensionen wieder. Konturierung des Spektrogramms bedeutet näherungsweise, daß nur noch die 'Gratlinien' des Gebirges verarbeitet werden, wie in Bild 1 angedeutet. Die zeitparallelen Konturlinien geben scharf die hörbaren Sinustöne im Signal wieder, die sich vorher im Spektrogramm wie auch auf der Basilarmembran nur verschwommen abzeichneten.

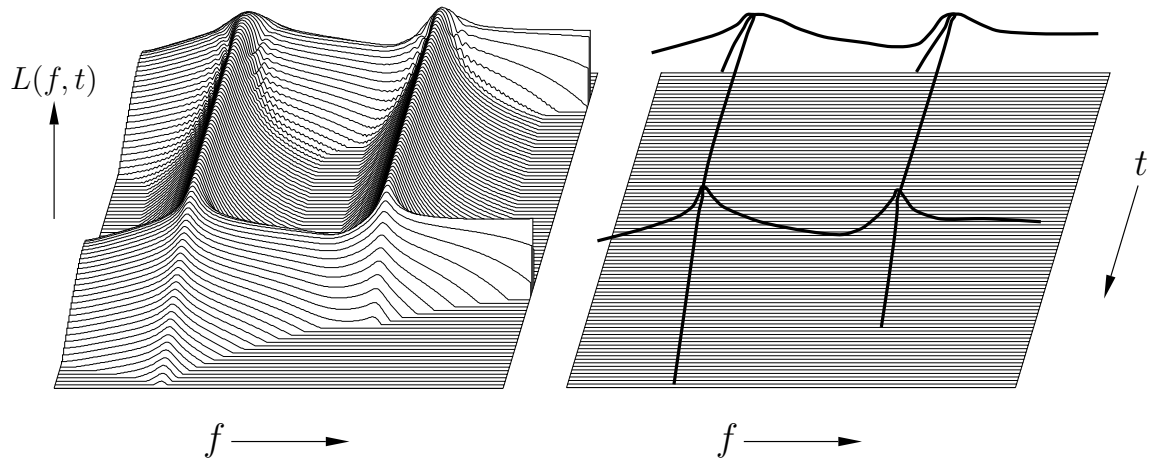

Bild 1: *Konturierung des Spektrogramms entspricht geometrisch etwa der Reduktion eines 'Pegelgebirges' $L(f,t)$, dargestellt über der Zeit t und der Frequenz f (links), auf seine 'Gratlinien' (rechts). Das zugrunde liegende Signal setzt sich aus zwei Sinustönen unterschiedlicher Frequenz zusammen, die gleichzeitig ein- und wieder ausgeschaltet werden.*

Die Grundlage für ein solches Verarbeitungsprinzip stammt aus einem Modell der Informationsaufnahme durch das Gehör nach Terhardt [Ter92]. Dort stellt sich der Hörvorgang als eine hierarchische Anordnung von autonomen Entscheidungsprozessen dar, welche als Konturierungsvorgänge realisiert werden. Der Begriff 'Konturierung' ist dabei in Anlehnung an den visuellen Begriff als 'Reduktion auf das Wesentliche' zu verstehen. Den elementarsten Entscheidungsprozeß verkörpert die Konturierung des Spektrogramms, dessen spezielle Gehöranpassung eine wichtige Rolle spielt. Es wird über die von Terhardt vorgeschlagene *Fourier-t-Transformation* (FTT) gewonnen, einer Kurzzeitspektraltransformation, die eine frequenzabhängige Analysebandbreite aufweist [Ter85].

Daß sich bestimmte Konturen eines FTT-Spektrogramms zur gehörorientierten Signalrepräsentation eignen, konnte bereits Heinbach mit seinem *Teiltonzeitmuster*(TTZM)-Verfahren zeigen [Hei88a]. Teiltöne entsprechen darin lokalen Maxima in Schnitten parallel zur Frequenzachse, die in ihrer zeitlichen Entwicklung verfolgt werden. Damit sind solche 'Gratlinien' erfaßt, die nicht parallel zur Frequenzachse verlaufen und die später als *Frequenzkonturen* bezeichnet werden. Um das Signal zurückzugewinnen, werden sie als zeitvariante Sinustöne interpretiert und überlagert. Hierin besteht eine entfernte Verwandtschaft zu sogenannten *Sinustonrepräsentationen* (z.B. [Mca86]), die sich aber nicht am Gehör orientieren. Das rekonstruierte Signal klingt im großen und ganzen wie das Original, auch wenn die Qualität nicht als Grundlage für hochqualitative Sprachcodierung ausreicht. Heinbach stellte auch Verfahrensvarianten vor, die das Teiltonzeitmuster mit Hilfe von einfachen Maßnahmen auf wenige, wesentliche Teiltöne reduzieren. Sie erzielen Datenraten bis herab zu etwa 4 kbit/s. Die Verständlichkeit von Sprache wird dabei nicht sonderlich beeinträchtigt, ihre Qualität sinkt allerdings auf ein wenig akzeptables Niveau.

Anfänglich bestand das Ziel der Arbeit darin, aus den Heinbachschen Verfahrensansätzen ein Sprachcodierverfahren mit niedriger Datenrate und akzeptabler Qualität zu entwickeln. Bald zeigte sich jedoch, daß Teiltöne alleine wenig geeignet sind, um alle wahrnehmbaren Sprachanteile datenreduziert zu repräsentieren. Auf der anderen Seite gab es auch keine befriedigende Erklärung, warum selbst ohne Datenreduktion des Teiltonzeitmusters wahrnehmbare Verfälschungen auftreten. Aus dieser Situation heraus schien es fruchtbarer, das TTZM-Verfahren zu analysieren und zu überarbeiten. Daraus wurde die Vorstellung von erweiterten Signalrepräsentationen entwickelt, die sich weiterhin an den Prinzipien des Terhardtschen Modells orientieren. Im Mittelpunkt steht die 'Gratlinien'-Repräsentation des FTT-Spektrogramms, mit der prinzipiell hochqualitative Audiocodierung erreichbar ist. Auf dieser erweiterten Grundlage wurde schließlich erneut eine datenreduzierende Sprachcodierung angestrebt. Die Arbeit ist folgendermaßen angelegt:

Kapitel 1 beschreibt als Grundlagenkapitel das Terhardtsche Modell, die FTT und das Heinbachsche TTZM-Verfahren. Dazu wird ein allgemeiner Überblick über Audiocodierungen, über die Bedeutung gehörorientierter Verfahrensweisen und über die Verwandtschaft zu bekannten Sinustonrepräsentationen gegeben.

Kapitel 2 analysiert das Heinbachsche TTZM-Verfahren und seine datenreduzierende Variante als Systeme zur Verarbeitung von Audiosignalen, die bestimmte wahrnehmbare Verfälschungen verursachen. Indem die wesentlichen Verfälschungen und ihre Ursachen herausgearbeitet werden, zeichnen sich grundsätzliche Beschränkungen der Heinbachschen Verfahrensansätze ab. Diese sollen in den weiteren Kapiteln umgangen werden. (Zusammenfassung S. 54)

Kapitel 3 erweitert das Konturierungskonzept, damit nun alle 'Gratlinien' des FTT-Spektrogramms, oder genauer, des zeitvarianten FTT-Pegelspektrums [1] erfaßt werden. So kommen die sogenannten *Zeitkonturen* hinzu, welche transiente Spektralanteile repräsentieren und mit impulshaft wahrgenommenen Signalanteilen in Verbindung stehen. Danach wird untersucht, wie die von Heinbach verwendete FTT modifiziert werden kann, so daß sie keine Verfälschungen mehr verursacht und möglichst gut mit dem erweiterten Konzept zusammenspielt. Transformations- und Konturierungsparameter werden subjektiv optimal eingestellt. (Zusammenfassung S. 96)

Kapitel 4 führt eine neue gehörorientierte Repräsentationsform ein, die direkt auf Konturen aufbaut. Dazu werden Konturen, deren genauer Verlauf für die Wahrnehmung nicht wichtig ist, insgesamt als *Textur* bezeichnet. Mit diesem Kontur/Textur-Konzept ist eine getrennte Codierung von tonalen und geräuschhaften Signalanteilen möglich. Probleme der Heinbachschen Datenreduktion können später damit umgangen werden. (Zusammenfassung S. 114)

Kapitel 5 entwickelt Verfahren zur Signalrekonstruktion aus Konturen und Kontur/Textur-Repräsentationen. Im erweiterten Konturierungskonzept macht es keinen Sinn mehr, Konturen von vornherein als zeitvariante Sinustöne zu betrachten. Die Verfahren lassen sich dadurch gewinnen, daß Codierung mit Konturen schrittweise aus einer verfälschungsfreien Codierung des zeitvarianten, komplexen FTT-Spektrums abgeleitet wird. (Zusammenfassung S. 154)

Kapitel 6 untersucht, wie leicht sich mit Konturen und Kontur/Textur-Repräsentationen datenreduzierende Sprachcodierungen aufbauen lassen. Mit einfachen Maßnahmen sind niedrige Datenraten nur dann zu erreichen, wenn man Anwendungen mit eher niedrigen Qualitätsanforderungen anstrebt – etwa im Bereich der Telefonsprache. Die vorgestellten Verfahren werden mit bekannten etablierten Verfahren, etwa mit CELP, aber auch mit der datenreduzierenden Variante des TTZM-Verfahrens verglichen. (Zusammenfassung und Ausblick S. 184)

Eine Zusammenfassung aller wesentlichen Ergebnisse findet sich ab Seite 187.

[1] Der Begriff 'Spektrogramm' wird als Kurzform für den Begriff 'zeitvariantes Kurzzeitbetragsspektrum' verwendet. Zeitvarianz ist darin deshalb zu unterstreichen, weil das 'Gebirge' im Raum, nicht aber nur seine frequenzparallelen Schnitte konturiert werden. Der Hauptteil der Arbeit umgeht die Kurzform und präzisiert von Fall zu Fall, weil auch öfter das (komplexe) Kurzzeitspektrum ohne Betragsbildung anzusprechen ist.

Kapitel 1

Grundlagen

Dieses Kapitel behandelt ausgewählte allgemeine und spezielle Grundlagen der Arbeit. Zunächst folgt eine grobe Übersicht über das etablierte Feld von Audiocodierungen. Dann wird erörtert, wie der hier angewandte gehörorientierte Ansatz zur Sprachcodierung gegenüber einem quellenorientierten Ansatz zu bewerten ist. Das qualitative Gehörmodell nach Terhardt und sein Konturierungsprinzip werden danach vorgestellt. Diese Abschnitte verdeutlichen, weshalb Sprachverarbeitung mit Konturen sinnvoll ist.

Gehörorientiertes Vorgehen setzt eine kurzzeitspektrale Betrachtungsweise des Signals voraus. Hier wird die sogenannte Fourier-t-Transformation (FTT) verwendet, die besonders an das Auflösungsvermögen des Gehörs angepaßt ist. Zusammen mit der anschließenden Einführung in das Heinbachsche Teiltonzeitmuster(TTZM)-Verfahren bildet sie den Schwerpunkt des Kapitels. Abschließend wird ein Überblick über verwandte Verfahren gegeben.

1.1 Anwendungsbereiche und Bewertungskriterien von Audiocodierungen

Audiocodierungen, so auch Sprachcodierungen, dienen dazu, Schallsignale geeignet über einen digitalen Nachrichtenkanal zu übertragen. Auf der einen Seite sollen sie möglichst die gesamte hörbare Information unverfälscht übertragen, auf der anderen Seite müssen sie die Beschränkungen des Kanals berücksichtigen. Je nach Anwendung und zulässigem Aufwand werden hier unterschiedliche Kompromisse geschlossen. Extrembeispiele dafür sind die Übertragung von Musik in CD-Qualität oder von gerade eben verständlicher Sprache in einem Kommunikationssystem. Der Begriff des Kanals schließt neben der räumlichen Übertragung auch die Möglichkeit von Aufzeichnung, Speicherung und Wiedergabe als ‘zeitliche Übertragung’ ein.

In der Theorie der Nachrichtenübertragung unterteilt man Codierungen in Quellencodierung und Kanalcodierung [Ste82, Jay84]. Letztere befindet sich näher am Zugang beziehungsweise Abgriff des Übertragungsmediums und stellt an den Schnittstellen zur umgebenden Quellencodierung einen gekapselten, digitalen Kanal bereit. Die Aufgabe der Anpassung an ein Medium und insbesondere die Sicherung gegen Übertragungsstörungen

ist dadurch idealisierend abgetrennt, wodurch allerdings potentiell Codierungseffizienz verschenkt wird [Cox91]. Die Quellencodierung hat die Aufgabe, das zu übertragende Signal in einen Bitstrom mit möglichst niedriger Rate zu wandeln. Die Aufgabe der sender- und empfängerseitigen Analog/Digital-Wandlung des Schallsignals wird dabei normalerweise nicht mehr beachtet. In diesem Sinne sind übliche Verfahren zur Audiocodierung im wesentlichen Quellencodierungen, auch wenn oft zusätzliche Bits zur Absicherung gegen Kanalstörungen verwendet werden. [1]

Man kann hauptsächlich drei Anwendungsbereiche für Audiocodierungen angeben [Nol95], in deren letzten beiden sich die Zielsetzung der vorliegenden Arbeit bewegt:

Breitbandaudio: originalgetreue, zweikanalige Codierung beliebiger Audiosignale im Frequenzbereich bis 20 kHz bei Quelldatenraten von 512 bis 768 kbit/s pro Kanal. Der Stand der Technik, beispielsweise der ISO-MPEG Audiostandard, läßt bei Raten von 128 kbit/s pro Kanal keine Beeinträchtigungen mehr hörbar werden, d.h. die Codierung wirkt 'transparent' [Bra94].

Breitbandsprache: hochqualitative Sprachkommunikation im Frequenzbereich bis 7 kHz bei Quelldatenraten von 128 kbit/s. Hier existieren Verfahren mit Datenraten zwischen 16 bis 64 kbit/s, die aber bereits von Varianten aus dem Gebiet Breitbandaudio bedrängt werden [MPE95].

Telefonsprache: Sprachkommunikation im Frequenzbereich 300 bis 3400 Hz bei Quelldatenraten von 64 kbit/s. Existierende Verfahren erlauben praktisch transparente Codierungen bei 16 kbit/s [LDC95]. Die Entwicklung konzentriert sich, je nach Qualitätsanforderung, auf Datenraten von 8 kbit/s und darunter. Für abhörsichere Kommunikation, bei der die Qualitätsanforderungen noch niedriger liegen, sind Verfahren mit einer Datenrate von 2,4 kbit/s bereits seit einiger Zeit im Gebrauch [Jay90, Tre82].

Ein ausführlicher Überblick über bekannte Verfahren in diesen Bereichen findet sich in [Ger94]. Neben Qualität und Datenrate sind noch weitere Kriterien zu berücksichtigen, beispielsweise Rechenaufwand, Verzögerungszeit sowie Unempfindlichkeit gegenüber Kanalstörungen und Mehrfachverarbeitung. Bei niedrigratiger Sprachcodierung treten zum Qualitätsbegriff Aspekte wie Verständlichkeit, Sprecheridentifizierbarkeit, Natürlichkeit sowie Lästigkeit von codierungsbedingten Störungen (Artefakten) hinzu. Artefakte sind beispielsweise bei Datenreduktion mit dem Heinbachschen TTZM-Verfahren ein wesentliches Problem. Weiterhin wird die Frage interessant, wie sich Störungen im Quellsignal, etwa Umgebungsgeräusche, unterschiedliche Sprecher und Sprechweisen oder die Einspeisung von nichtsprachlichen Signalen auswirken. Die Unempfindlichkeit des Codierverfahrens gegen solche Einflüsse wird allgemein als Robustheit bezeichnet.

[1] Der mißverständliche Begriff 'Quellencodierung' impliziert nicht, daß nur Eigenschaften der Quelle statt Eigenschaften des Empfängers ausgenutzt werden.

1.2 Bedeutung des gehörorientierten Ansatzes bei Datenreduktion

Zur Datenreduktion gibt es zwei methodische Ansätze: Redundanzreduktion und Irrelevanzreduktion (z.B. [Jay84]). Idealerweise werden beim ersten zunächst vorhersagbare Signalanteile entfernt (Dekorrelation). Anschließend werden für die übrigbleibenden, informationstragenden Anteile anhand ihrer statistischen Verteilung optimal platzsparende Codes gewählt (Entropiecodierung). [2] Dabei ist ein Modell nötig, welches Eigenschaften des Quellsignals beschreibt. Bei Beschränkung auf Sprache kann die eingeschränkte Signalvielfalt ausgenutzt werden, wodurch allerdings die Robustheit der Codierung leidet (vgl. Abschnitt 6.3.5).

Während ideale Redundanzreduktion reversibel ist, entfernen übliche Sprachcodierungen später durch Quantisierung einen gewissen Anteil an Information, der sich für den Empfänger als mehr oder weniger irrelevant erweist. Damit gibt es in solchen Codierungen auch Irrelevanzreduktion, obwohl sie hier zunächst nicht gezielt eingesetzt wurde. Die Gruppe der Linearen-Prädiktions-Codierungen (LPC) hat sich in dieser Weise entwickelt [Tre82, Var88, Cam90, Iye91, Che93, CSA95], ihr liegt ein Sprachproduktionsmodell zugrunde (z.B. [Rab78]). Inzwischen ist ein starker Trend zur gezielten Berücksichtigung der Irrelevanz zu beobachten [Nol95]. Für Breitbandaudio-Codierungen birgt Redundanzreduktion nur einen beschränkten Gewinn, weil die Vielfalt der Quellsignale nur begrenzt vorhersagbar und damit kaum modellierbar ist.

Irrelevanzreduktion betrachtet das Signal aus der Sicht des Empfängers und entfernt für ihn uninteressante Signalanteile. Ein solcher Ansatz benötigt ein Modell der Empfängereigenschaften. Es beherbergt bei Audiocodierungen idealerweise das Wissen darüber, was am Signal oder an dosierten, codierungsbedingten Störungen für das Gehör mit welcher Ausgeprägtheit wahrnehmbar ist. Dieses Modell steuert die Quantisierung und formt dadurch das Quantisierungsgeräusch derart, daß es vom Nutzsignal im Gehör verdeckt wird ('noise shaping') [Sch79, Tri79]. Irrelevanzreduktion ist im Gegensatz zur Redundanzreduktion von vornherein irreversibel, existiert aber ebenso nur in Mischformen mit der letzteren. Maximale Datenreduktion kann grundsätzlich nur bei voller Ausschöpfung von Redundanz- *und* Irrelevanzreduktion erreicht werden.

Audiocodierungen unter besonderer Berücksichtigung der Irrelevanzreduktion, sogenannte 'perzeptive' Codierungen, ermöglichten große Fortschritte bei der Breitbandaudio-Codierung [Kra85, Sto86, Bra87, Joh88]. In diesem Einsatzbereich darf aber nur 'perzeptive' Irrelevanz entfernt werden, also nur das, was wirklich nicht wahrgenommen werden kann. Dagegen kann der Begriff der Irrelevanz bei Telefonsprache unter Umständen viel weiter gefaßt werden, wenn für bestimmte Anwendungen bloße Verständlichkeit ausreichend ist. Das Empfängermodell benötigt für solche Einsätze allerdings mehr Wissen. Zu beachten ist, daß dann auch das Potential der Redundanzreduktion steigt, weil Quellenmodelle von Sprache wieder sinnvoll angewendet werden können.

Der Ansatz dieser Arbeit ist die Irrelevanzreduktion, Modelle der Quelleneigenschaften und insbesondere Sprachproduktionsmodelle werden nicht berücksichtigt. Dies gründet

[2] Die Zweistufigkeit des Vorgangs reflektiert die Unterscheidung zwischen Bindungs- und Verteilungsredundanz. Die erste bezeichnet die Korrelation aufeinanderfolgender Zeichen. Eine Reduktion der ersten beeinflußt immer auch die zweite, welche durch unangepaßte Codelängen entsteht.

sich erstens darauf, daß Irrelevanzreduktion im Rahmen der Fortschritte bei der Breitband-audio-Codierung ein sehr großes Potential bewiesen hat. Zweitens zeigen die vielversprechenden Ergebnisse von Heinbach, daß das Terhardtsche Modell der Informationsaufnahme durch das Gehör einen effizienten Ansatz für Codierungen bietet, insbesondere bei Verzicht auf strenge Bewahrung von perzeptiver Relevanz. Die Vernachlässigung von Methoden der Redundanzreduktion bedeutet, daß unter Umständen großzügig mit vorhandener Redundanz umgegangen wird. Möglicherweise wird sogar an bestimmten Stellen noch zusätzliche Redundanz eingebracht, etwa durch ungünstige Codezuweisung.

1.3 Konturierung und Informationsaufnahme durch das Gehör

Der Ansatz der Irrelevanzreduktion benötigt Wissen über die Signalverarbeitung im Gehör. Die Formulierung dieses Wissens geschah bisher mit Hilfe von Modellen oder Funktionsschemata, die mehr oder weniger zahlreiche Charakteristika der Hörwahrnehmung quantitativ nachbilden. Dazu gehört insbesondere das Funktionsschema der Maskierung nach Zwicker und Feldtkeller [Zwi67, Zwi82, Zwi90], von dem die Verfahren zur Transparentcodierung profitieren. Statt der wesentlichen informationsverarbeitenden Eigenschaften werden dadurch allerdings nur Begleiterscheinungen erfaßt.

Einen qualitativen, aber umfassenden Ansatz bietet dagegen das Modell der Informationsaufnahme durch das Gehör nach Terhardt [Ter87, Ter91, Ter92]. Es geht von der Grundfragestellung aus, wie das Gehör Information aus einem Sprachsignal extrahiert oder, anders ausgedrückt, wie es Irrelevanz daraus entfernt. Dabei werden evolutions-, wahrnehmungs- und informationstheoretische Prinzipien berücksichtigt.

‘Kontur’ ist ein Begriff, der normalerweise der visuellen Wahrnehmung zugeordnet wird, seine Anwendung innerhalb der auditiven Wahrnehmung erscheint ungewöhnlich. Terhardt wies jedoch nach [Ter87], daß sich beispielsweise die Kategorien der Tonhöhenwahrnehmung [Ter72a, Ter72b, Ter79] analog zu visuellen Konturen verhalten. Konturierung, also die Extraktion von Konturen, repräsentiert demnach ein Verfahrensprinzip der Wahrnehmung. Weil Konturierung gleichzeitig Beschränkung auf das Wesentliche bedeutet, impliziert dieses Prinzip Irrelevanzreduktion.

In Terhardts Modell wird die Informationsverarbeitung im Gehör durch eine hierarchische Anordnung von einzelnen, autonomen Entscheidungsprozessen dargestellt. Am Eingang des Modells steht eine gehörangepaßte Kurzzeitspektraltransformation, die die Verarbeitung durch das Innenohr verkörpert. Tiefere Ebenen des Modells reichen das Ergebnis einer selbständigen Entscheidung an die nächsthöhere weiter. Dabei entspricht Entscheidung in der Informationsverarbeitung Konturierung in der Wahrnehmung, die gesamte Hierarchie drückt einen Gestaltsfindungsprozeß aus. Darin formt sich die Wahrnehmung von ‘auditiven Objekten’ hierarchisch aus ‘Sub-Objekten’. Auf höherem Hierarchieniveau beispielsweise spiegelt sich darin die Zusammensetzung von Worten aus Silben wider, die wiederum aus Phonemen bestehen und so weiter. Das Wissen der auditiven Informationsverarbeitung verteilt sich dadurch über die gesamte Hierarchie. Die Hierarchie ist nach oben offen, denn das Bewußtsein hat wahlfreien Zugriff zu den einzelnen Ebenen, die ihre Ergebnisse eine begrenzte Zeit zwischenspeichern. Ein tieferer Zugriff bedeutet mehr

analytisches, ein höherer mehr synthetisches Hören.

Als monaural relevantes Ergebnis der Kurzzeitspektraltransformation sieht Terhardt das zeitvariante Betragsspektrum der Fourier-t-Transformation (FTT) ohne weitere Phaseninformation an [Ter92]. Zwar gibt es Gehörmodelle, die eine primäre Phasenmessung durch die Basilarmembran einbeziehen ('phase synchrony' oder 'phase locking', [Sen84, Coo86, Coo93, Pat95]). Andererseits reicht zur Erklärung von monauralen Phaseneffekten bei Tonkomplexen die zeitvariante Hüllkurve in den (überlappenden) Frequenzgruppen aus (z.B. [Zwi82]).

1.4 Spektralanalyse durch Fourier-t-Transformation (FTT)

Hauptcharakteristikum der peripheren Signalverarbeitung im Gehör ist die Frequenz/Ort-Transformation auf der Basilarmembran [Zwi82]. Mit ihrer Hilfe realisiert das Gehör einen natürlichen Spektralanalysator, dessen grundsätzliches Verhalten sich durch eine Kurzzeitspektraltransformation beschreiben läßt [Fla72]. Solche Transformationen entwickelten sich historisch aus den Methoden zur Messung von Leistungsspektren (z.B. [Fan50, Sch62]), sie bieten aber zusätzlich die Möglichkeit der Signalrückgewinnung durch Rücktransformation. Ohne Angabe einer Analyse formulierte schon Gabor eine Signalrepräsentation mit Hilfe kurzzeitiger, spektral begrenzter Einzelelemente [Gab46]. Spektrales und zeitliches Auflösungsvermögen, die fundamentalen Eigenschaften der Transformation, werden durch Wahl einer Fensterfunktion festgelegt. Das nachzubildende Auflösungsvermögen des Gehörs, das mit Hilfe der Frequenzgruppenbreite [Zwi82] modelliert wird, ändert sich allerdings über der Frequenz. Die Möglichkeit frequenzabhängiger Fensterfunktionen wurde bereits in [Gam71] hervorgehoben, wobei sich aber eine Rücktransformation als problematisch erweist.

Ein konkretes Konzept einer gehörgerechten Kurzzeitspektraltransformation wurde erstmals von Terhardt vorgestellt [Ter85]. Die Fourier-t-Transformation (FTT) bildet die Grundlage der weiteren spektralen Betrachtungen und Verfahrensweisen. Spezielle Effekte, wie etwa die pegelabhängige Anhebung der oberen Mithörschwellenflanke (z.B. [Zwi82]) bleiben allerdings unberücksichtigt. Die Realisierung als zeitdiskretes System greift Abschnitt 3.3.1 auf, die allgemeine Formulierung einer Rücktransformation behandelt Abschnitt 5.1.1.

1.4.1 Definition der FTT

Sei $s(t)$ ein kausales Zeitsignal und, mit ω als Frequenz, $h(\omega, t)$ eine kausale, frequenzabhängige Fensterfunktion. Dann ist das komplexe, zeit- und frequenzabhängige FTT-Spektrum von $s(t)$ definiert als

$$s(\omega, t) = \int_0^t s(\tau) h(\omega, t - \tau) e^{-j\omega\tau} d\tau. \tag{1.1}$$

Übliche Definitionen von Kurzzeitspektraltransformationen [Fla72] kennen keine Frequenzabhängigkeit in der Fensterfunktion und schreiben auch keine Kausalität vor. Der Faktor

$h(\omega, t-\tau)$, betrachtet bei festgehaltenem ω, kann als zeitverschiebliches *Analysefenster* interpretiert werden. Sinnvolle Fensterfunktionen stellen die Impulsantwort eines kausalen Tiefpasses dar. Ihre effektive Dauer entspricht der Analysefensterlänge, die einen Richtwert für das zeitliche Auflösungsvermögen verkörpert. Die komplexe 3dB-Bandbreite $B_{3dB} = 2f_{3dB}$ des Tiefpasses, definiert über $f = \frac{\omega}{2\pi}$, wird als *Analysebandbreite* bezeichnet und bietet entsprechend einen Richtwert für das spektrale Auflösungsvermögen. Das Produkt dieser beiden Werte ist unabhängig von der Frequenz und steht für die Zeit/Frequenz-Unschärfe des Kurzzeitspektrums. Man wählt die Analysebandbreite derart, daß sie sich proportional zur Frequenzgruppenbreite des Gehörs Δf_G verhält. Hierbei kommt die Formel

$$\Delta f_G \;=\; \left\{ 25 \;+\; 75 \left[1 + 1,4 \left(\frac{f}{\mathrm{kHz}} \right)^2 \right]^{0,69} \right\} \mathrm{Hz} \;\;\hat{=}\;\; 1 \text{ Bark} \tag{1.2}$$

nach [Zwi80] zur Anwendung. Bis auf weiteres wird die ursprünglich vorgeschlagene Fensterfunktion verwendet. Mit $a(\omega)$ als frequenzabhängiger Fensterkonstante und einer in [Fel85] eingeführten Normierung lautet sie

$$h(\omega, t) = 2a(\omega) e^{-a(\omega)t}. \tag{1.3}$$

Dies entspricht bei festem ω der Impulsantwort eines Tiefpasses erster Ordnung. Seine 3dB-Bandbreite B_{3dB}, also die zuvor definierte Analysebandbreite, hängt wie folgt von der Fensterkonstante ab:

$$B_{3dB} = \frac{a}{\pi}. \tag{1.4}$$

1.4.2 Systeminterpretation der FTT und ihre Anwendung

Die Formulierung der FTT nach Gl. (1.1) erweist sich als recht unhandlich, nicht nur für eine Deutung als Spektralanalysator, sondern auch für eine praktische Realisierung oder zur Herleitung weiterer Konsequenzen. Deshalb wird hauptsächlich die folgende formale Interpretation benutzt, mit der übliche Methoden der Systemtheorie anwendbar werden (z.B. [Mar82]). Sie liegt der Filterbankinterpretation von Analyse/Synthese-Systemen mit Kurzzeitspektren zugrunde [Por80], ist aber zur Messung von Kurzzeitleistungsspektren schon länger bekannt [Sch62]. Man betrachtet den Zeitverlauf des FTT-Spektrums aus Gl. (1.1) an einer festen *Analysefrequenz* ω_A. Dieser kann als Faltung der Fensterfunktion $h(\omega_A, t) = h_{\omega_A}(t)$ mit dem Produkt von Signal $s(t)$ und komplexer Schwingung $e^{-j\omega_A t}$ angesehen werden:

$$s_{\omega_A}(t) = s(\omega, t)|_{\omega=\omega_A} \;=\; \int_0^t \left(s(\tau) e^{-j\omega_A \tau} \right) h_{\omega_A}(t-\tau) d\tau \tag{1.5}$$

$$\;=\; \left(s(t) \cdot e^{-j\omega_A t} \right) * h_{\omega_A}(t). \tag{1.6}$$

Formal kommen hierin nur noch Zeitsignale vor, ω_A ist nun ein beigeordneter Parameter. Als System interpretiert, entspricht Gl. (1.6) einem linearen, aber zeitvarianten *Analysefilter*, mit $s(t)$ als Eingangs- und $s_{\omega_A}(t)$ als Ausgangssignal. Wie in Bild 1.1 dargestellt besteht es aus einem komplexen Modulator für die Multiplikation mit $e^{-j\omega_A t}$ und einem

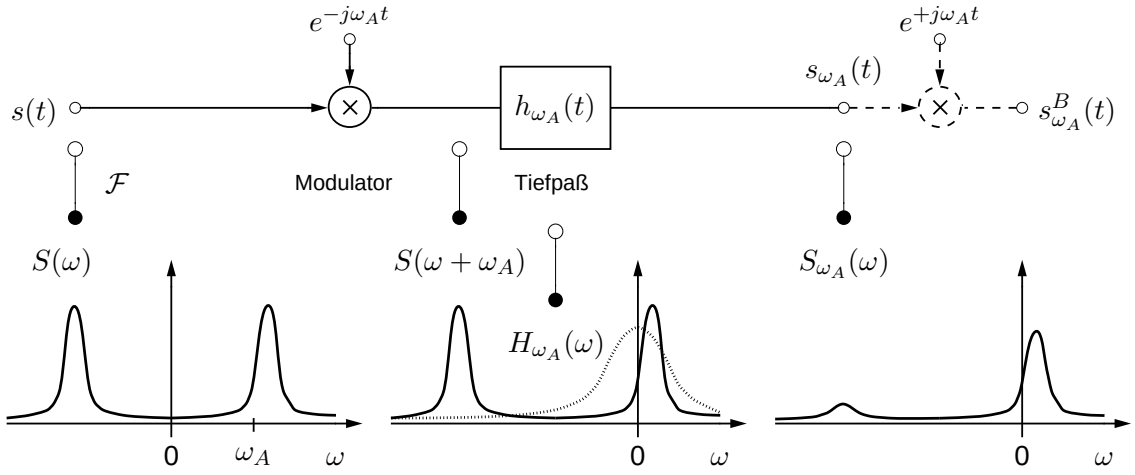

Bild 1.1: *Interpretation der FTT an einer festen Analysefrequenz ω_A als System, das sogenannte Analysefilter (oben). Demonstration des Analysevorganges anhand von Fourier-Betragsspektren (unten). Der gestrichelte Teil oben würde das System zum komplexen Bandpaß erweitern.*

Tiefpaß mit der Impulsantwort $h_{\omega_A}(t)$, dem *Analysetiefpaß*. Die Wirkungsweise als Spektralanalysator erklärt sich, wenn auf Gl. (1.6) die Fourier-Transformation $\mathcal{F}$ angewendet wird: [3]

$$S_{\omega_A}(\omega) = \mathcal{F}\{s_{\omega_A}(t)\} = \mathcal{F}\{s(t)e^{-j\omega_A t}\} \cdot \mathcal{F}\{h_{\omega_A}(t)\} \tag{1.7}$$

$$= S(\omega+\omega_A) \cdot H_{\omega_A}(\omega). \tag{1.8}$$

Die Skizzen von Fourier-Betragsspektren im unteren Teil von Bild 1.1 veranschaulichen das Ergebnis der Umformung. Die komplexe Modulation verschiebt das Fourier-Spektrum $S(\omega)$ eines hier willkürlich gewählten Eingangssignals um ω_A nach links (links bzw. Mitte unten). Die nachfolgende Tiefpaßfilterung entspricht einer Multiplikation mit der Systemfunktion $H_{\omega_A}(\omega)$ (Mitte bzw. rechts unten). Insgesamt manifestiert sich die Spektralanalyse dadurch, daß ein um ω_A zentrierter Teil des Fourier-Spektrums am Eingang etwa mit der Analysebandbreite ausgewählt und in Basisbandlage verschoben wird. Die Leistung des Basisbandsignals $s_{\omega_A}(t)$ repräsentiert somit die Leistung im ausgewählten Analyseband. Die Anpassung an die Frequenzgruppenbreite vergrößert die Breite des Bandes bei höheren Analysefrequenzen.

Im Kontext einer Signalverarbeitung durch kausale Systeme ist oft die Laplace-Transformation $\mathcal{L}$ anstelle der Fourier-Transformation günstiger. Allerdings entfällt dann die unmittelbare Deutbarkeit als Spektrum. Beispielsweise gilt Gl. (1.8) in vergleichbarer Form auch für die Laplace-Transformierten:

$$S_{\omega_A}(p) = \mathcal{L}\{s_{\omega_A}(t)\} = S(p+j\omega_A) \cdot H_{\omega_A}(p) \tag{1.9}$$

Manchmal stört die Zeitvarianz des Analysefilters infolge des Modulators. Formal verschwindet sie nach gegensinniger Modulation mit $e^{j\omega_A t}$ am Ausgang, ohne daß sich dabei der Betrag des Ausgangssignals ändert. So wird das gefilterte Band in Bild 1.1 wieder an

[3] Großschreibung mit Argument ω markiert die Fourier-Transformierte eines kleingeschriebenen Zeitsignals, Großschreibung mit Argument p seine Laplace-Transformierte. Die Variable ω der Fourier-Transformation hat *nichts* mit derjenigen in der FTT-Definition (1.1) zu tun.

die ursprüngliche Bandlage zurückgeschoben. Das Gesamtsystem realisiert die Filterung eines Eingangssignals $s(t)$ mit einem komplexen Bandpaß der Impulsantwort $h_{\omega_A}(t) \cdot e^{j\omega_A t}$. Man erhält unter Verwendung von Gl. (1.6) das *FTT-Bandpaßsignal*

$$s^B_{\omega_A}(t) = s_{\omega_A}(t) \cdot e^{j\omega_A t} \tag{1.10}$$

$$= s(t) * \left(h_{\omega_A}(t) \cdot e^{j\omega_A t}\right). \tag{1.11}$$

Weil es sich hierbei näherungsweise um ein analytisches Signal handelt ($S^B_{\omega_A}(\omega) \approx 0$ für $\omega < 0$), liefert sein Betrag $|s^B_{\omega_A}(t)|$ die zeitliche Hüllkurve des betrachteten Bandes (z.B. [Pap86]). Dies gilt dann natürlich auch schon für den betragsgleichen Zeitverlauf des FTT-Spektrums $|s_{\omega_A}(t)|$ nach Gl. (1.6).

Bei Systeminterpretation geht die Frequenz als kontinuierliche Dimension des FTT-Spektrums nur scheinbar verloren. Einerseits kann man sich jederzeit unendlich viele, in ω_A infinitesimal versetzte Analysefilter nach Bild 1.1 vorstellen. Andererseits weisen die später eingesetzten Formen des FTT-Spektrums bestimmte Verlaufseigenschaften in Frequenzrichtung auf, durch die beliebige Zwischenwerte interpolierbar sind (vgl. Anhang C.3). Die praktische Auswertung des FTT-Spektrums mit Systemen an diskreten Analysefrequenzen (Filterbank) kann formal als Abtastung eines rekonstruierbaren, kontinuierlichen Frequenzverlaufes angesehen werden. Deshalb wird später meistens die Darstellung nach Gl. (1.6) als (zeitvariantes) FTT-Spektrum angesetzt.

Oftmals wird das Verhalten der FTT bei stationären Tönen benötigt, etwa bei einem Sinuston

$$s(t) = A\cos(\omega_T t - \phi_0) \tag{1.12}$$

$$= \frac{Ae^{-j\phi_0}}{2} \cdot e^{j\omega_T t} + \frac{Ae^{j\phi_0}}{2} \cdot e^{-j\omega_T t} \tag{1.13}$$

mit Amplitude A, Frequenz ω_T und Phasenlage ϕ_0. Mit Hilfe von Gl. (1.7) ergibt sich das 'eingeschwungene', trotzdem aber zeitvariante FTT-Spektrum [4]

$$\lim_{t\to+\infty} s_{\omega_A}(t) = \frac{Ae^{-j\phi_0}}{2} H_{\omega_A}(\omega_T - \omega_A) \cdot e^{j(\omega_T-\omega_A)t} + \frac{Ae^{j\phi_0}}{2} H_{\omega_A}(-\omega_T - \omega_A) \cdot e^{j(-\omega_T-\omega_A)t} \tag{1.14}$$

$$\approx \frac{Ae^{-j\phi_0}}{2} H_{\omega_A}(\omega_T - \omega_A) \cdot e^{j(\omega_T-\omega_A)t} \quad \text{gültig für} \quad |\omega_T - \omega_A| \ll |\omega_T + \omega_A|. \tag{1.15}$$

Das exakte Ergebnis entspricht der Summe zweier rotierender, komplexer Zeiger. Die spektrale Selektion durch den Tiefpaß, welche von der betrachteten Analysefrequenz abhängt, bestimmt ihr Längenverhältnis. Die Näherung gilt für den Normalfall, bei dem nur die

[4] Der Grenzübergang ist formal zur Eliminierung der Einschwingeffekte erforderlich, da in der FTT-Definition (1.1) kausale Signale vorgeschrieben sind. Auch im weiteren Verlauf der Arbeit werden aus Gründen der Übersichtlichkeit Signal- und Fensterdefinitionen verwendet, in denen auf explizite Nullsetzung für $t < 0$ verzichtet wird. Ausnahmen sind Signaldefinitionen, bei denen ein Übergang bei $t = 0$ wesentlich ist.

Positivseite des Eingangsspektrums maßgeblich selektiert wird. Wenn nur das Betragsspektrum interessiert, ist allein die Länge der Zeigersumme maßgeblich. Im Normalfall ist der Betrag also konstant, weist exakt betrachtet aber einen kleinen Anteil auf, der mit der Differenzfrequenz $2\omega_T$ der beiden Zeiger schwingt. Sind im Eingangssignal $s(t)$ weitere Sinustöne zu einem Tonkomplex überlagert, so ergibt sich $s_{\omega_A}(t)$ infolge der Linearitätseigenschaften des Analysefilters als Überlagerung von Zeigern nach Gl. (1.14) oder (1.15).

1.5 Teiltonzeitmuster(TTZM)-Verfahren nach Heinbach

Das von Heinbach entwickelte Verfahren realisiert eine gehörorientierte Repräsentation von Audiosignalen durch zeitvariante Sinustöne [Hei88b, Hei88a, Hei87b, Hei87a, Hei86]. Sie werden als Teiltöne bezeichnet und bilden in ihrer Gesamtheit das Teiltonzeitmuster (TTZM). Sein Gewinnungsprozeß, nachfolgend Teiltonanalyse genannt, orientiert sich an niedrigen Hierarchieebenen des Modells von Terhardt. Deshalb ist keine exakte Nachbildung psychoakustischer Befunde beabsichtigt, vielmehr steht die Erfassung der für das Gehör wesentlichen Information im Vordergrund. Nach der einleitenden Übersicht werden die einzelnen Operationen, soweit später benötigt, genauer besprochen.

Bei der *Teiltonanalyse* wird das Signal mit Hilfe der FTT zunächst in ein zeitvariantes Pegelspektrum überführt. Daran schließt sich ein Konturierungsvorgang an, der ausgeprägte lokale Pegelmaxima über der Frequenz auswählt. Sie repräsentieren in ihrer jeweiligen zeitlichen Entwicklung die Teiltöne. Das Teiltonzeitmuster läßt sich anschaulich darstellen, wie in Bild 1.2 oben für ein Sprachsignal geschehen. Linien verkörpern die Teiltonverläufe in Zeit und Frequenz, ihre Dicke markiert den Pegel. Die Frequenzachse wird proportional zur gehörgerechten Tonheitsskala z gewählt, deren Einheit Bark der Frequenzgruppenbreite an der jeweiligen Frequenz entspricht. [5] Ausgeprägtere Teiltonverläufe spiegeln mitunter einzeln wahrnehmbare Töne im Sinne von Spektraltonhöhen [Ter72a] wider. Gebiete mit dichtliegenden, kurzen und unregelmäßigen Verläufen repräsentieren beispielsweise Rauschanteile.

Zur Rückwandlung des Teiltonzeitmusters in ein hörbares Signal, hier *Teiltonsynthese* genannt, [6] werden einfach Sinustöne überlagert, deren Parameter Amplitude und Frequenz von den Teiltönen unmittelbar vorgegeben sind. Phasenlagen untereinander bleiben dabei unberücksichtigt, motiviert durch die qualitative Anpassung der Teiltonanalyse an die 'Phasenempfindlichkeit' des Gehörs. Gegenüber der authentischen Teiltonsynthese existiert mittlerweile eine verbesserte Variante.

[5] In TTZM-Darstellungen kommt zwischen der Frequenz f und der Tonheit z die untenstehende Transformationsformel nach [Zwi80] zur Anwendung, ansonsten sind Frequenzabstände oder Bandbreiten in Bark immer von Gl. (1.2) abgeleitet:

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

[6] Heinbach benutzt den Begriff 'Resynthese'.

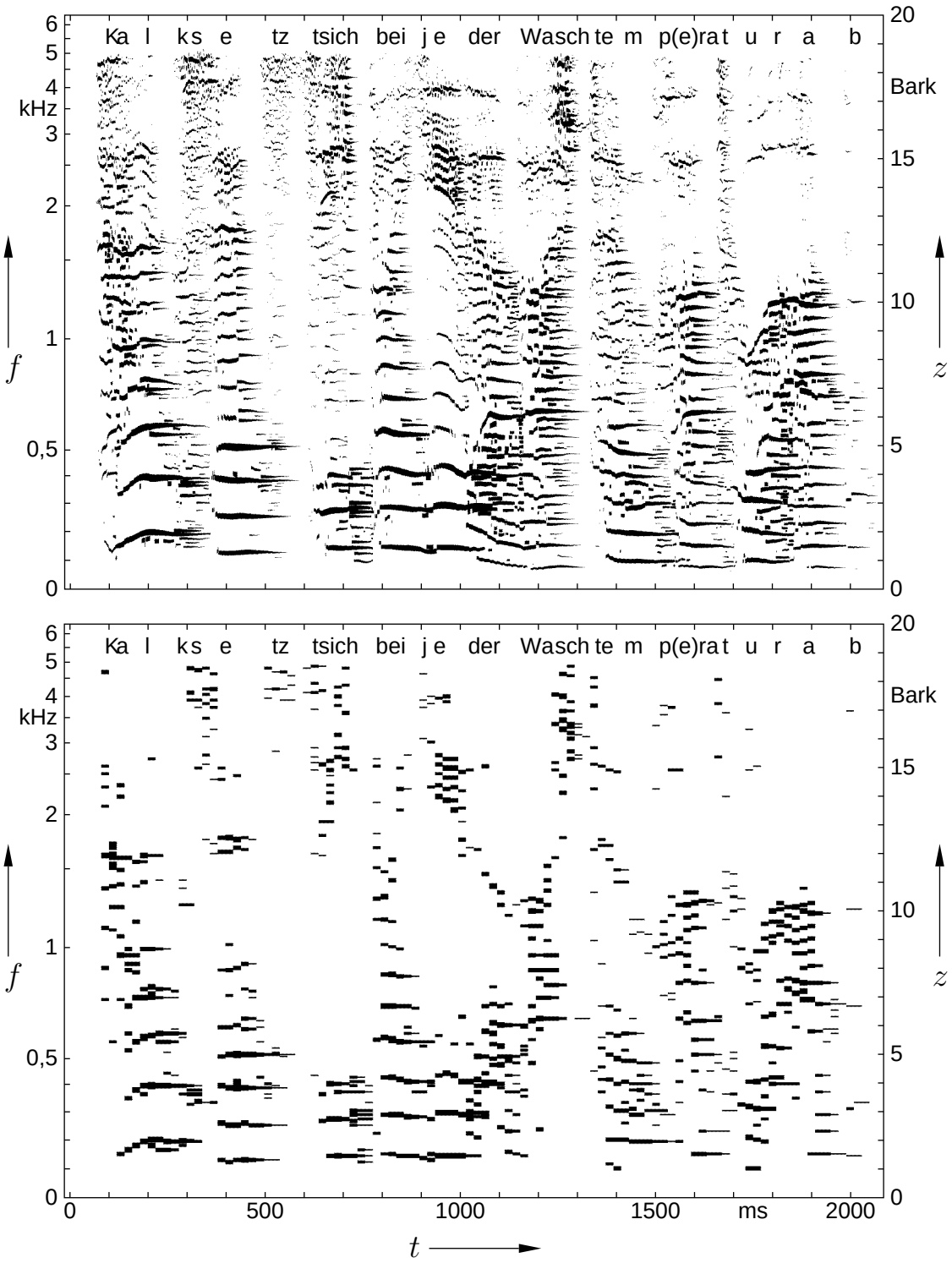

Bild 1.2: *Teiltonzeitmuster von Sprache eines männlichen Sprechers (a) und die auf etwa* 4 kbit/s *datenreduzierte Variante (b), jeweils nach Heinbach [Hei88a]. Die Liniendicke markiert den Teiltonpegel über einen Dynamikbereich von rund* 60 dB.

Unter dem Begriff *TTZM-Verfahren* sollen zukünftig Teiltonanalyse und Teiltonsynthese zusammengefaßt verstanden werden. Als Grundlage für Codierverfahren impliziert dies für sich allein jedoch noch keine Datenreduktion. Dafür sind besondere Maßnahmen zu ergreifen, die die Quantisierung des Teiltonzeitmusters vergröbern, wie in Bild 1.2 unten verdeutlicht.

1.5.1 Berechnung des FTT-Pegelspektrums

Ausgangspunkt ist die Systeminterpretation der FTT aus Gl. (1.6) oder (1.11). Sie repräsentiert das zeitvariante FTT-Spektrum an einer bestimmten Analysefrequenz ω_A durch das komplexe Signal $s_{\omega_A}(t)$ oder dessen betragsgleiche Variante $s^B_{\omega_A}(t)$. Heinbach legte dafür die Analysefensterfunktion aus Gl. (1.3) mit einer 3dB-Bandbreite $B_{3dB} = 0{,}1$ Bark zugrunde. Den Übergang auf das Signal $p_{\omega_A}(t)$ des FTT-Leistungsspektrums beschreiben die untenstehenden beiden Formeln. Speziell die erste erlaubt eine Aussage über die Vererbung von Signaleigenschaften: Nach Gl. (1.8) ist das Fourier-Spektrum von $s_{\omega_A}(t)$ infolge des Analysetiefpasses $H_{\omega_A}(\omega)$ näherungsweise bandbegrenzt. Multiplikation mit dem Konjugierten führt zu einer Art Selbstfaltung im Fourier-Spektrum. Schlimmstenfalls verdoppelt sich also die Signalbandbreite beim Übergang.

$$p_{\omega_A}(t) = s_{\omega_A}(t) \cdot s^*_{\omega_A}(t) \tag{1.16}$$

$$= |s_{\omega_A}(t)|^2 = |s^B_{\omega_A}(t)|^2. \tag{1.17}$$

Heinbach ermittelte das Pegelspektrum aus einer zeitlich geglätteten Form $p^G_{\omega_A}(t)$ des Leistungsspektrums. Sie entspringt der Filterung mit einem Tiefpaß erster Ordnung, der die Impulsantwort $h^G_{\omega_A}(t)$ aufweist:

$$p^G_{\omega_A}(t) = p_{\omega_A}(t) * h^G_{\omega_A}(t), \tag{1.18}$$

$$h^G_{\omega_A}(t) = \frac{1}{T_G} \cdot e^{-\frac{t}{T_G}}. \tag{1.19}$$

Abhängig von der Analysefrequenz gilt als Glättungszeitkonstante

$$T_G = \begin{cases} 0,2\ (\pi B_{3dB})^{-1} & \text{für} \quad \frac{|\omega_A|}{2\pi} \leq 3\ \text{kHz}, \\ T_G\big|_{\frac{\omega_A}{2\pi} = 3\ \text{kHz}} & \text{sonst.} \end{cases} \tag{1.20}$$

Ihre Dimensionierung hängt mit der Wahl der Ausgeprägtheitsschwelle $\Delta L_A = 3$ dB im nächsten Unterabschnitt zusammen. Die kombinierte Maßnahme von Glättung und Ausgeprägtheitsschwelle ist für das Funktionieren des Heinbachschen TTZM-Verfahrens wesentlich. Andernfalls können bei Tönen störende Nebenmaxima auftreten. Darüberhinaus begründet Heinbach die Glättung auch als Anpassung an das Verhalten der Rauhigkeitswahrnehmung [Ter68a, Ter74]. Hiervon ist auch die Frequenzgrenze von 3 kHz in Gl. (1.20) motiviert, ab der T_G nicht weiter ansteigt. Unter dieser Grenze ist die Bandbreite des Glättungstiefpasses fünfmal so groß wie die Analysebandbreite.

Die Berechnung des FTT-Pegelspektrums mit vorheriger Glättung $L^G(f,t)$ beziehungsweise ohne Glättung $L(f,t)$ geschieht durch Logarithmierung in Relation zu einer Referenzleistung $p_{ref} \mathrel{\hat{=}} 0$ dB. Weil Konturierung später über kontinuierliche Dimensionen

definiert wird, wechselt die Notation wieder auf eine Form, bei der Zeit und Frequenz gleichrangige Variablen sind:

$$L^{(G)}(f,t)\big|_{2\pi f\,=\,\omega_A} = 10\,\lg\left(\frac{p^{(G)}_{\omega_A}(t)}{p_{ref}}\right)\ \text{dB}. \tag{1.21}$$

Sofern nicht anders vermerkt, werden Pegelspektren an $N = 380$ Analysefrequenzen im Abstand $\frac{\Delta\omega_A}{2\pi} = 0{,}05$ Bark ausgewertet. Sie beginnen bei $\frac{\omega_{A_0}}{2\pi} = 20$ Hz und reichen bis etwa 5,5 kHz. Die Abtastrate des zu analysierenden Signals $s(t)$ beträgt für die späteren Experimente immer $f_a = 12{,}8$ kHz. Berechnungen des (geglätteten) Leistungsspektrums erfolgen mit der gleichen Rate, dagegen orientiert sich der Übergang auf Pegelwerte an der Nachfrage der nachfolgenden Operationen. In späteren Pegeldarstellungen entspricht die Referenzleistung p_{ref} immer der Spitzenleistung des jeweils analysierten Signals im Zeitdiskreten.

1.5.2 Teiltonextraktion/Frequenzkonturierung

Zu regelmäßigen Auswertezeitpunkten werden lokale Maxima des Pegelspektrums über der Frequenz gesucht. Davon werden diejenigen ausgewählt, die mit einer Mindestausgeprägtheit aus der Nachbarschaft hervorragen, etwa gegenüber benachbarten Minima. Sie repräsentieren die momentan vorhandenen Teiltöne. Heinbach benutzte als Ausgeprägtheitsschwelle einen Wert von $\Delta L_A = 3$ dB, der auf die Glättungszeitkonstante T_G nach Gl. (1.20) abgestimmt ist.

Der Vorgang der Teiltonextraktion wird später in Frequenzkonturierung umbenannt, Teiltöne heißen dann Frequenzkonturen. Seine exakte Definition formuliert Anhang A.1. Charakteristisch für ihn ist, daß er im ‘Gebirge’ des zeitvarianten Pegelspektrums nur diejenigen Konturen erkennt, die in frequenzparallelen Schnitten durch lokale Maxima auf sich aufmerksam machen.

Der Analysefrequenzabstand $\frac{\Delta\omega_A}{2\pi} = 0{,}05$ Bark stellt für die Teiltonextraktion eine Frequenzauflösung sicher, die dem eben wahrnehmbaren Frequenzunterschied von Sinustönen gleicht. Eine feinere Frequenzauflösung steigert allerdings die Qualität der bildlichen TTZM-Darstellung, weil in den Teiltonverläufen keine Frequenzstufen mehr irritieren.

Da die Analysebandbreite $B_{3dB} = 0{,}1$ Bark merklich über dem Analysefrequenzabstand liegt, kann der Verlauf des Leistungsspektrums zwischen den Analysefrequenzen gut rekonstruiert werden (Anhang C.3). Diesen Sachverhalt machen sich zwei Approximationsverfahren zunutze, um die Auflösung der Maxima und damit der Teiltonfrequenzen zu verbessern (Anhang B.1). Das Feldtkeller-Verfahren wurde ohne besondere Erwähnung schon von Heinbach angewandt [HeiPK], weshalb es bis auf weiteres beibehalten wird. In später modifizierten Analyseverfahren kommt dagegen standardmäßig Parabel-Approximation zum Einsatz. Letztere wirkt sich praktisch wie ein verringerter Analysefrequenzabstand aus, das Feldtkeller-Verfahren weist zusätzlich in bestimmten Situationen eine glättende Wirkung in Frequenzrichtung auf. Deshalb unterscheiden sich die Verfahren bei Modulationssignalen etwas in der Teiltonzuweisung, womit geringfügige Verarbeitungsunterschiede hörbar werden können. Bei der Verarbeitung von Sprachsignalen hört man keine Unterschiede.

1.5.3 Codierung des Teiltonzeitmusters und Datenreduktion

Frequenzen und Pegel aller jener Teiltöne, die zu einem Auswertezeitpunkt vorhanden sind, bilden einen Satz von Wertepaaren, der *Teiltonmuster* genannt wird. Den Abstand der Auswertezeitpunkte und damit die Zeitquantisierung legt das *Auswerte-* oder *Analyseintervall* T_A fest. Die Abfolge von Teiltonmustern, jeweils in Frequenzen und Pegeln quantisiert, ergeben das codierte Teiltonzeitmuster. Zwar sind die Wertepaare nicht explizit zu Verläufen von zeitvarianten Teiltönen assoziiert: Dies kann aber jederzeit mittels Wertevergleich zwischen aufeinanderfolgenden Teiltonmustern erreicht werden, solange Frequenz- und Zeitquantisierung nicht zu grob gewählt wurden. Eine solche Linienassoziation ist spätestens zur Teiltonsynthese unumgänglich, Anhang A.2 beschreibt sie formal.

Die Quantisierung des Teiltonzeitmusters ohne Datenreduktion ist für spätere Experimente zunächst so eingestellt, daß die maximal mögliche Qualität gewährleistet ist. Das Auswerteintervall beträgt $T_A = 1{,}25$ ms, Pegelwerte sind mit einer Dynamik von rund 80 dB auf 0,5 dB genau codiert. Die Frequenzwerte, die nunmehr feiner als die Analysefrequenzen approximiert sind, werden mit 0,2 Hz im bis 5,5 kHz reichenden Nutzfrequenzbereich aufgelöst.

Bei den Verfahren zur Datenreduktion verwendete Heinbach vier spezielle Maßnahmen, um die Datenrate auf 16 beziehungsweise 4,4 kbit/s zu beschränken [Hei87b, Hei88a]. [7] Sie scheinen insgesamt die Vorstellung zu bestätigen, daß man auch mit wenigen Tönen die wesentliche gehörrelevante Information repräsentieren kann, selbst wenn die Amplituden nicht so exakt zu rekonstruieren sind. Die Maßnahmen wirken als Irrelevanzreduktion im erweiterten Sinne von Abschnitt 1.2, da sie Quelleigenschaften nicht in Betracht ziehen:

- Das Auswerteintervall wird auf 7,5 ms bei der Variante mit 16 kbit/s und auf 20 ms bei der mit 4,4 kbit/s erhöht.
- Die Anzahl der pro Teiltonmuster codierten Teiltöne beschränkt sich auf maximal zehn der pegelstärkeren.
- Der Dynamikbereich berücksichtigter Teiltonpegel reduziert sich auf 60 dB und wird in Stufen von 4 bit quantisiert. Speziell bei der 4,4 kbit/s-Variante tritt noch eine grobe, aber wirkungsvolle Pegelcodierung hinzu. Dabei werden die Teiltöne eines Teiltonmusters nach Pegel sortiert übertragen, von den Pegelwerten aber nur der stärkste und der schwächste codiert. Zur Decodierung werden die fehlenden Pegel zwischen den Eckwerten linear über dem Sortierindex interpoliert.
- Die Teiltonfrequenzen innerhalb von 0,1 bis 5,2 kHz werden in 256 Stufen mit rund 0,07 Bark quantisiert. [8]

Wie sich diese Maßnahmen auf das unbearbeitete Teiltonzeitmuster auswirken, veranschaulicht Bild 1.2 unten für die Variante von 4,4 kbit/s. Neben der verminderten Anzahl gleichzeitiger Teiltöne wird die zeitliche Quantisierung infolge des groben Auswerteintervalls erkennbar. Das Konstanthalten der Pegel- und Frequenzwerte über seine Dauer

[7] Heinbach berechnet in Tabelle 5.1.9 in [Hei88a] irrtümlicherweise 4 statt 4,4 kbit.
[8] Die Tabelle 5.1.9 in [Hei88a] gibt fälschlicherweise eine Untergrenze von 20 Hz an, vgl. dort S. 71.

stellt eine spezielle, wenn auch primitive Verlaufsrekonstruktion dar. Sie wurde in dieser Form auch von Heinbach stillschweigend vorausgesetzt. Tatsächlich sind die codierten Teiltonwerte nur Abtastwerte, aus denen ein verbesserter Decoder beispielsweise sanfte Linienübergänge rekonstruieren könnte, bevor sie der Teiltonsynthese angeboten werden.

1.5.4 Teiltonsynthese (TTSR und TTSD)

Beim Heinbachschen Syntheseprinzip werden Teiltonverläufe als zeitvariante Sinusschwingungen aufgefaßt, die zu überlagern sind. Mit formalen Gemeinsamkeiten beginnend werden zwei Syntheseverfahren beschrieben, die sich hinsichtlich ihres sogenannten *Synthesefensters* unterscheiden. Das erste Verfahren, die Teiltonsynthese mit Rechteckfenster (TTSR), entspricht dem authentischen von Heinbach [Hei88a]. Das zweite, die Teiltonsynthese mit Dreieckfenster (TTSD), vermeidet auf einfache Weise die offenkundigen Probleme des ersten. Weil damit eine Qualitätsverbesserung zu erzielen ist, gilt in späteren Untersuchungen des Heinbachschen TTZM-Verfahren *stillschweigend TTSD angewandt*, soweit nichts anderes vermerkt ist.

Vorbereitend muß man im codierten Teiltonzeitmuster die Linienassoziation der abgetasteten Teiltonverläufe herstellen. Dazu werden Paarungen von Teiltonwerten in angrenzenden Teiltonmustern gebildet. Ihr Frequenzabstand darf maximal $\pm\Delta f_\Phi$ betragen und muß gegenüber möglichen Konkurrenzpaarungen minimal sein. Aus aufeinanderfolgenden Paarungen entstehen Ketten, die über ihre Lebenszeit jeweils eine Synthesesinusschwingung in Frequenz und Amplitude kontrollieren. Amplitudenwerte erhält man direkt aus den delogarithmierten Teiltonpegeln.

An die Phasenfortschreibung einer allgemein zeitkontinuierlichen Synthesesinusschwingung sind zwei Bedingungen geknüpft. Sie soll erstens stetig sein. Zweitens soll sie bei null beginnen, oder, wenn im Abstand bis $\pm\Delta f_\Phi$ eine andere Sinusschwingung vorhanden ist, mit deren Momentanphase. Die vorher schon bei Linienassoziation maßgebliche Grenze Δf_Φ bestimmt demnach folgendes Phasenübergabeprinzip: Existieren innerhalb von $\pm\Delta f_\Phi$ eine oder mehre Fortsetzungsmöglichkeiten einer Synthesesinusschwingung, dann sind alle phasenstetig anzuschließen. Für TTSR wählte Heinbach einen Wert $\Delta f_\Phi = 0{,}15$ Bark, der in TTSD zur Kompatibilität mit späteren Analyseverfahren auf 0,25 Bark erhöht wird.

Um schließlich einer Kette eine passende zeitkontinuierliche Sinusschwingung zuzuweisen, spielt das Synthesefenster $h^S(t)$ eine zentrale Rolle. In einem sehr einfachen Teiltonzeitmuster bilde die Menge (f_i, A_i) eine Kette von Frequenz/Amplituden-Wertepaaren, die Auswertezeitpunkten $i_B \le i \le i_E$ im Abstand von T_A zugeordnet sind. Mit Hilfe einer Übergabephase ϕ_i, welche die Phasenstetigkeit zwischen den Auswerteintervallen sicherstellt, erzeugt die Kette das Synthesesignal

$$\hat{s}(t) = \sum_{i=i_B}^{i_E} A_i \cdot h^S(t - iT_A) \cdot \sin\left(2\pi f_i(t - iT_A) + \phi_i\right), \tag{1.22}$$

$$\text{wobei} \qquad \phi_i = \begin{cases} 0 & \text{für} \quad i = i_B, \\ \phi_{i-1} + 2\pi f_{i-1} T_A & \text{sonst.} \end{cases} \tag{1.23}$$

In normalerweise komplizierteren Teiltonzeitmustern ist zusätzlich über alle vorkommenden Ketten zu summieren. Außerdem kann man dann nicht mehr bei allen Ketten von

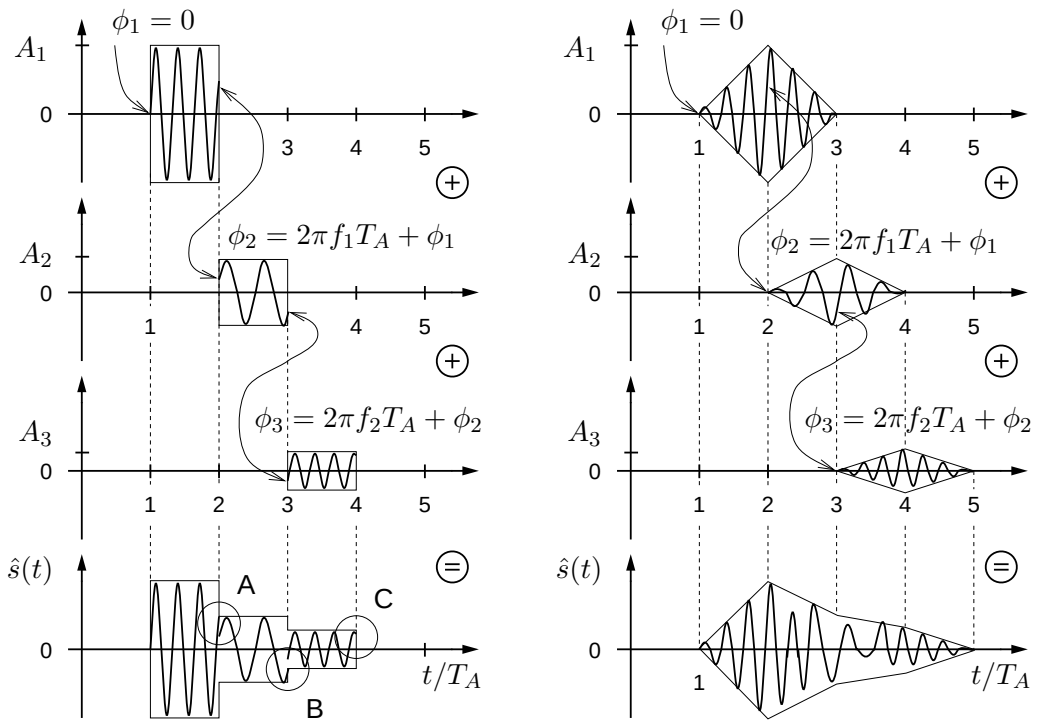

Bild 1.3: *Teiltonsynthese mit Rechteckfenster nach Heinbach [Hei86] (links) und neu mit Dreieckfenster (rechts) am Beispiel einer Kette von drei Teiltonwerten (siehe Text). Trotz Phasenübergabe zwischen den Sinustonimpulsen entstehen beim Rechteckfenster Sprünge im Synthesesignal $\hat{s}(t)$ (Kreise A, B und C).*

einer Nullinitialisierung der Startphase ausgehen.

1.5.4.1 Ursprüngliche Teiltonsynthese mit Rechteckfenster (TTSR)

Heinbach hält Amplitude und Frequenz der zeitkontinuierlichen Synthesesinusschwingung über der Dauer eines Auswerteintervalls konstant. Dies impliziert ein Rechteck als Synthesefenster:

$$h^S(t) = \begin{cases} 1 & \text{für} \quad 0 \le t < T_A, \\ 0 & \text{sonst.} \end{cases} \tag{1.24}$$

Jeder Summand in Gl. (1.22) bildet nun einen Sinustonimpuls mit rechteckiger Hüllkurve, die zeitlich exakt an die des Kettennachbarn anstößt. Bild 1.3 links verdeutlicht dies am Beispiel einer Kette über drei Auswerteintervalle, von $i_B = 1$ bis $i_E = 3$. Ein Sinustonimpuls beginnt jeweils mit der Sinusphase ϕ_i nach Gl. (1.23) und schließt deshalb stetig an den Phasenstand an, den die vorige Sinusfunktion zuletzt erreicht hat. Aneinandergereiht ergeben die Tonimpulse das Synthesesignal $\hat{s}(t)$.

Mit einem Rechteckfenster sind jedoch Nachteile verbunden. Üblicherweise ändert sich die Amplitude im Verlauf einer Kette. Deshalb können trotz Phasenstetigkeit an den Auswerteintervallgrenzen Sprünge im Signalverlauf auftreten (Kreise A und B in Bild 1.3 links). Zu Beginn führt selbst das Einschalten der Amplitude bei Nullphase auf einen relativ abrupten Übergang. Am Ende ist ein Abschalten im Phasennulldurchgang unwahrscheinlich und damit meist ein Sprung die Folge (Kreis C). Abhängig vom Signal und von

der Länge des Auswerteintervalls T_A entstehen deshalb Störungen, die als überlagertes Rauschen, Knistern oder als einzelne Knacke wahrzunehmen sind. Weil die Sprünge an die Auswertezeitpunkte gekoppelt sind, kann sich die Auswerteintervallrate $1/T_A$ sogar als tonale Störung durchschlagen.

Eine Verkleinerung von T_A auf das Signalabtastintervall $1/f_a$ kann die Störungen nicht eliminieren. Einerseits schlägt sich die Pegelquantisierung des Teiltonzeitmusters weiterhin in Form von Amplitudensprüngen nieder, andererseits bleiben die Effekte zu Beginn und Ende der Kette bestehen.

1.5.4.2 Neue Teiltonsynthese mit Dreieckfenster (TTSD)

Die Synthesequalität kann verbessert werden, wenn man mittensymmetrische Dreiecke mit der Halbwertsbreite T_{6dB} als Synthesefenster wählt:

$$h^S(t) = \begin{cases} \frac{T_A}{T_{6dB}} \left(1 - \left|\frac{t}{T_{6dB}} - 1\right|\right) & \text{für} \quad 0 \leq t \leq 2T_{6dB}, \\ 0 & \text{sonst.} \end{cases} \tag{1.25}$$

Die Dimensionierung $T_{6dB} = 1{,}25$ ms hat sich als günstig erwiesen, sie stimmt beim nichtreduzierenden TTZM-Verfahren mit T_A überein. Für diese Verhältnisse ist in Bild 1.3 rechts die Umsetzung des Beispiels zu sehen. Die Nachteile beim Rechteckfenster werden offensichtlich vermieden. Bei konstanten Frequenzen f_i erfährt nun die Synthesesinusschwingung eine lineare Amplitudeninterpolation zwischen den Auswertezeitpunkten, zusätzlich entstehen lineare Rampen am Beginn und am Ende.

Strenggenommen kann man das Syntheseergebnis nun aber nicht mehr als eine Sinusschwingung ansehen, deren Parameter exakt den Teiltonverlauf widerspiegeln. Dies liegt zum einen an den hinzuinterpretierten Rampen. Zum anderen können bei größeren Frequenzunterschieden oder längeren Auswerteintervallen beziehungsweise Halbwertsbreiten zusätzliche Modulationen in den Überlappungsbereichen entstehen, wie in Bild 1.3 rechts übertrieben angedeutet. Beim kurzen Auswerteintervall des TTZM-Verfahrens ohne Datenreduktion ist dies jedoch nicht wesentlich. Das Syntheseergebnis ist vielmehr mit einer Sinusschwingung gleichzusetzen, deren Parameterübergänge in Amplitude und Frequenz leicht geglättet wurden.

1.5.4.3 Kombiniert spektral/zeitliche Betrachtung

Weitere Einblicke in die Wirkung des Synthesefensters gibt eine spektrale Betrachtung des Synthesevorgangs. Dazu wendet man die Fourier-Transformation auf Gl. (1.22) an, indem man $\sin(x) = \frac{1}{j2}(e^{jx} - e^{-jx})$ einsetzt und Verschiebungssätze in Zeit und Spektrum berücksichtigt:

$$\hat{S}(\omega) \quad = \quad \sum_{i=i_B}^{i_E} \frac{A_i}{2j} e^{-j\omega i T_A} \Big(H^S(\omega - 2\pi f_i) e^{j\phi_i} - H^S(\omega + 2\pi f_i) e^{-j\phi_i} \Big). \tag{1.26}$$

Jeder Summenterm korrespondiert mit der Fourier-Transformierten eines Sinustonimpulses, dessen zeitliche Lage und Übergabephase in Form von rein phasendrehenden Exponentialfunktionen berücksichtigt ist. Da $\hat{s}(t)$ rein reell ist, gibt es Spektralanteile auf der

positiven und negativen Frequenzhalbachse, entsprechend der linken und rechten Hälfte des Summenterms. Wesentlich ist, daß sich jeder Tonimpuls mit der Breite der Fourier-Transformierten $H^S(t)$ des Synthesefensters im Spektrum des Synthesesignals niederschlagen kann. Die eigentlich schmalbandige Information eines Teiltons kann auf diese Weise spektral 'verschmiert' werden. Je breiter $H^S(\omega)$ ausfällt, desto wahrscheinlicher sind wahrnehmbare Störungen.

Unter diesem spektralen Blickwinkel erweist sich ein Rechteck als ungünstiges Synthesefenster. Seine Fourier-Transformierte beschreibt eine si-Funktion, deren Flanken schwach abfallen. Dagegen entspricht die Fourier-Transformierte des Dreieckfensters in der Standarddimensionierung genau der quadrierten des Rechtecks. Sie verläuft deshalb schmaler und mit steileren Flanken. Wegen der Reziprozität von spektraler und zeitlicher Ausdehnung könnte man eine weitere Verschmälerung noch über eine zeitliche Streckung erreichen. Unter einem zeitlichen Blickwinkel bedeutet dies allerdings, daß die zeitliche Glättung des Teiltonverlaufs zunimmt, was sich ebenfalls störend auswirkt. Eine optimale Form und Skalierung für das Synthesefenster bleibt deshalb zu untersuchen.

1.6 Verwandtschaft mit Sinustonrepräsentationen

Weil das TTZM-Verfahren ein Sinussynthesemodell verwendet, weist es Gemeinsamkeiten zu sogenannten Sinustonrepräsentationen (STR) auf. Sie haben sich im wesentlichen seit den achtziger Jahren entwickelt und können grob in drei Gruppen unterteilt werden:

Quellorientierte STR: Hier orientierte man sich an Produktionsmodellen von einstimmigen Musikinstrumenten [Gre77] oder stimmhafter Sprache [Alm82a]. Die Beschränkungen dieser Ansätze lagen darin, daß Quasistationarität und Harmonizität vorausgesetzt wurden. Zwar konnten Almeida und Tribolet mit dem Formalismus der 'verallgemeinerten Harmonischen' exakte Repräsentation beliebiger Signale erzwingen [Alm83]. Bei nichtstimmhaften Sprachanteilen schnellt jedoch die Komplexität in die Höhe, so daß diese besser als Residualsignal ausgegliedert und mit anderen Mitteln repräsentiert werden. Oder es müssen pauschale und somit fehlerträchtige stimmhaft/stimmlos-Unterscheidungen getroffen werden [Mar89]. Auf diesen Grundlagen entstanden Sprachcodierverfahren unter dem Begriff 'Harmonic Coding' [Alm82b, Tra88, Mar90]. Die Probleme der pauschalen stimmhaft/stimmlos-Unterscheidung wurden erstmals im 'Multiband Excitation Vocoder' von Griffin und Lim vermieden, bei dem in einzelnen Frequenzbändern unterschieden wird [Gri88]. Verbesserte Varianten erreichen Datenraten bis herab zu 2,4 kbit/s [Kon94].

Formale STR: Die zweite Gruppe verwendet einen auf Hedelin zurückreichenden, rein formalen Ansatz. Er stellt das Signal als eine Überlagerung von zeitvarianten, nicht notwendigerweise harmonischen Sinusschwingungen dar [Hed82]. McAulay und Quatieri entwickelten ein vielbeachtetes Verfahren, welches die Sinusparameter ähnlich wie beim TTZM-Verfahren per Maximumsextraktion und Linienassoziation sucht [Mca86, Mca89b]. Das zugrunde gelegte Kurzzeitspektrum ist aber nicht gehörangepaßt, weshalb Phasen ein wichtiger Bestandteil der Repräsentation sind [Mca89ba]. Unter dem Begriff 'Sinusoidal Transform Coding' entwickelten sich daraus Sprachcodierungsverfahren mit Datenraten bis herab zu 2,4 kbit/s [Mca95, Mca85, Mca87,

Mca88, Mca91]. Ein weiteres Verfahren wurde von Feiten und Becker zur Modellierung von Instrumentenklängen vorgestellt [Fei90].

Die Verarbeitung nichtstimmhafter Anteile stellte in dieser Gruppe zunächst kein Problem dar. Für eine ausreichende Qualität benötigt man allerdings eine höhere Anzahl von Sinusschwingungen. Die STR ist hier unhandlich, weshalb zunehmend versucht wird, Rauschanteile gesondert zu repräsentieren. Hierzu zählen die Mischrepräsentationen von Serra und Smith [Ser90, Ser96] und Marques et al. [Mar88, Mar91] mit ihren datenreduzierenden Varianten [Mar94]. In dieser Gruppe gilt allgemein, daß die datenreduzierenden Varianten wieder den quellorientierten STR nahe rücken, unter anderem weil sie Harmonizität voraussetzen.

Gehörorientierte STR: Schon früh gab es Anregungen, Gehöreigenschaften zur Aufstellung einer STR zu nutzen [Hed82]. Die Tatsache, daß beispielsweise wenige Harmonische des stimmhaften Sprachsignals tatsächlich hörbar sind, ist durch die Tonhöhentheorie schon länger bekannt [Ter72a, Ter72b, Ter79]. Bisher schwach vertreten sucht diese Gruppe weder die Nähe zu Produktionsmodellen, noch erhebt sie den Anspruch, mit den überlagerten Sinusschwingungen tatsächlich das Zeitsignal zu rekonstruieren. Stattdessen sollen sich Original und STR nur aus der Sicht der Hörwahrnehmung möglichst wenig unterscheiden. Ghitza betrachtete die Ebene des Hörnervs [Ghi87], um das Verfahren von McAulay und Quatieri zu modifizieren. Ein bereits dem TTZM-Verfahren ähnlicher Ansatz existiert von Ellis [Ell92], wenngleich dieser mit einer frequenzproportionalen Analysebandbreite arbeitet.

Das Heinbachsche TTZM-Verfahren scheint also zur letzten Gruppe zu zählen. Doch nimmt die vorliegenden Arbeit einen anderen Standpunkt ein. Der wesentliche Beitrag des Teiltonzeitmusters und später der Konturen wird in der Repräsentation gesehen, die als Ergebnis eines gehörorientierten Aufbereitungsmodells entsteht. Unmittelbare Interpretation der Teiltonverläufe als Sinusparameter läßt sich zwar als ‘erste Näherung’ einer Signalrekonstruktion ansehen. Für eine verbesserte Rekonstruktionsqualität fällt sie aber möglicherweise wesentlich komplizierter aus. Deshalb wird eine allgemeinere Kategorie als die der Sinustonrepräsentationen beansprucht, nämlich die der gehörorientierten und synthesefähigen Audiorepräsentationen. Bislang in dieser Kategorie bekannte Verfahren arbeiten mit aufwendigen, detaillierten Gehörmodellen [Huk89, Coo93, Sla94].

Kapitel 2

Grenzen des Heinbachschen TTZM-Verfahrens

Bei Heinbach stand die Eigenschaft des Teiltonzeitmusters im Vordergrund, das Verhalten von Verarbeitungsprozessen im Gehör zu reflektieren. Teiltonsynthese und Verbund mit der Teiltonanalyse zum TTZM-Verfahren dienten zunächst nur zur Kontrolle [Hei86]. Auch die datenreduzierenden Verfahrensvarianten und weitere, in [Hei88a] vorgestellte Anwendungen haben eher exemplarischen Charakter, um die Eignung des Teiltonzeitmusters zur gehörgerechten Signalverarbeitung zu demonstrieren. Ohne Rücksicht auf diese Aspekte wird das TTZM-Verfahren nun als signalverarbeitendes System untersucht. In dieser Eigenschaft weist es eine Reihe von qualitätsmindernden Verfälschungen auf, die den praktischen Einsatz zur Sprachcodierung beeinträchtigen. Ziel des Kapitels ist es, wesentliche Verfälschungseffekte und ihre Ursachen zu erarbeiten.

Überwiegend geht es dabei um das Verfahren ohne Datenreduktion. Es wird beobachtet, wie eine Reihe von synthetischen Testsignalen verarbeitet werden. Ob bestimmte Veränderungen wahrnehmbar sind, läßt sich bei den verwendeten Signalen leicht mit bekannten Gesetzmäßigkeiten der Psychoakustik objektivieren (z.B. [Zwi67, Zwi82, Zwi90]). Ebenso lassen sich die Ursachen leicht ergründen. Bei manchen Effekten sind noch besondere Auswirkungen bei Sprachverarbeitung zu diskutieren.

Zum Schluß werden die zusätzlichen Verfälschungen durch Datenreduktion betrachtet. Hierfür werden von vornherein Sprachsignale verwendet. Die Ursachen der mitunter sehr deutlichen Verfälschungseffekte hängen hier natürlich direkt mit den einzelnen Reduktionsmaßnahmen zusammen. Deshalb stützt sich ihre Erforschung darauf, daß verschiedene Maßnahmenkombinationen gegeneinander ausgespielt werden.

2.1 Verarbeitung transienter Signale

Unter transienten Signalen könnten solche verstanden werden, die Impulse, Sprünge oder sonstige ‘vorübergehende’ Eigenschaften im Zeitbereich vorweisen. Da das Signal aber nur über sein FTT-Spektrum beobachtet wird, bestimmt zweckmäßigerweise dessen Sicht, ob ein Signal transienten oder, als Gegenpol, quasistationären Charakter hat. Die Eigenschaft ‘transient’ bezieht sich also auf die Eigenschaft seiner Abbildung in einem bestimmten

Zeit/Frequenz-Gebiet des FTT-Spektrums. Es wird deshalb präzisierend von transienten oder quasistationären Signalanteilen die Rede sein. Erstere zeigen sich als breitbandige und flache, letztere als schmalbandige Bereiche im Momentanspektrum. 'Transiente Signale' sind nun also solche, die aus der Sicht des FTT-Spektrums transiente Signalanteile enthalten.

Zuerst wird die TTZM-Verarbeitung eines isolierten Dirac-Impulses analysiert. Dieser Fall läßt sich gut analytisch behandeln. Ein realistischerer Fall ist eine Folge von Impulsen, wie sie zur Modellierung der Glottisschwingung in Sprachproduktionsmodellen verwendet wird (z.B. [Rab78, Osh87]). Beide Untersuchungen verdeutlichen, daß die Heinbachsche Teiltonextraktion als Konturierungsvorgang prinzipbedingt nicht alle gehörrelevanten Eigenschaften des FTT-Spektrums erfassen kann. Eine Untersuchung der fehlenden Knackrepräsentation bei geschalteten Signalen offenbart darüber hinaus prinzipielle Beschränkungen der von Heinbach verwendeten FTT-Spektraldarstellung. Daß sich diese Effekte nicht als besonders kritisch für die Verarbeitung von Sprachsignalen erweist, liegt an einer Reihe von günstigen Umständen, die abschließend angesprochen werden.

2.1.1 Unzureichende Repräsentation von Impulsen

Als Eingangssignal $s(t)$ für das TTZM-Verfahrens dient ein Dirac-Impuls $\delta(t)$. FTT-Spektrum, Teiltonzeitmuster und Synthesesignal sollen unter zulässigen Vereinfachungen untersucht werden. Aus der FTT-Definition (1.1) und der Fensterfunktion (1.3) ergibt sich als FTT-Spektrum

$$s(\omega, t) = 2ae^{-at}, \quad \text{mit} \quad a = a(\omega). \tag{2.1}$$

Die Modulator/Tiefpaß-Interpretation der FTT nach Bild 1.1 auf S. 10 veranschaulicht, daß der Impuls den Modulator unverändert passiert und daß deshalb der Zeitverlauf des FTT-Spektrums die Impulsantwort des Analysetiefpasses wiedergibt. Höhe und Dauer der Impulsantwort sind frequenzabhängig, weil die 3dB-Analysebandbreite $\frac{a}{\pi}$ an die Frequenzgruppenbreite nach Gl. (1.2) gekoppelt ist. Das ungeglättete FTT-Pegelspektrum in Bild 2.1 links verkürzt sich deshalb zu höheren Frequenzen hin in der Dauer, steigt aber in der zeitlich maximal erreichten Höhe.

Da keine spektralen Nebenmaxima existieren, kann für eine TTZM-Berechnung die Wirkung der Glättung nach Gln. (1.18), (1.19) vernachlässigt werden. Um außerdem alle theoretisch möglichen Teiltöne zu extrahieren, wird mit $\Delta L_A = 0$ keine Forderung nach einer Mindestausgeprägtheit erhoben. Teiltöne entsprechen dann direkt den lokalen Maxima des FTT-Pegelspektrums $L(f, t)$ über der Frequenz. Deren Orte (ω_{max}, t) erfüllen die Bedingung

$$\left|\frac{\partial s(\omega, t)}{\partial \omega}\right|_{\omega=\omega_{max}} = \left|2(1-at)\frac{da(\omega)}{d\omega}e^{-at}\right|_{\omega=\omega_{max}} = 0. \tag{2.2}$$

Nach Gl. (1.2) ist $a(\omega)$ für die betrachteten positiven ω eine monoton steigende Funktion, so daß nur der geklammerte Term den Wert null herbeiführen kann. Die zentrale Bestimmungsgleichung für relative Maxima lautet deshalb

$$a(\omega_{max}) = 1/t, \quad \text{mit} \quad \omega_{max} = \omega_{max}(t). \tag{2.3}$$

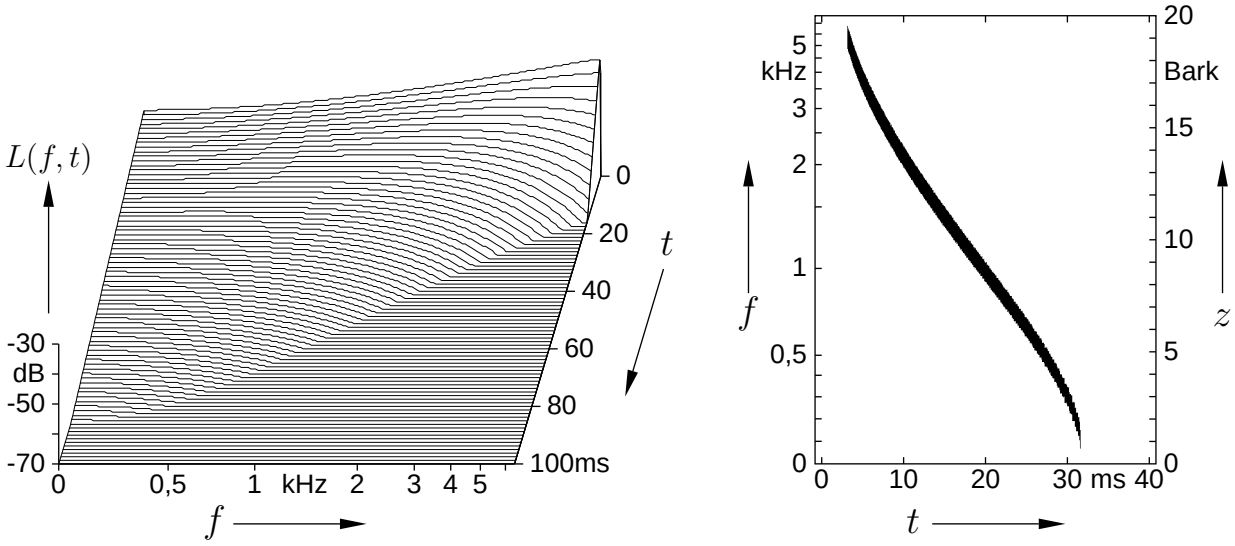

Bild 2.1: *FTT-Pegelspektrum (links) und Teiltonzeitmuster (rechts) eines Dirac-Impulses bei $t = 0$. Zeitliche Glättung und Ausgeprägtheitsschwelle wurden vernachlässigt. Die Rechnung zeigt, daß der Teiltonverlauf nur einen Bruchteil der hörbaren Energie im betrachteten Frequenzbereich repräsentieren kann.*

Sie gibt implizit den Ort eines einzigen, zeitveränderlichen Maximums vor. Es würde bei $t = 0$ bei $\omega = \infty$ auftauchen, um sich dann monoton zu niedrigeren Frequenzen hin bis herab zu null zu bewegen. Üblicherweise beschränken sich die betrachteten Analysefrequenzen aber auf einen positiven Bereich von ω_{A_0} bis ω_{A_N}. Deshalb erscheint erst ab dem Zeitpunkt t_1 ein relatives Maximum, um später bei $t_2 > t_1$ wieder zu verschwinden. Das resultierende Teiltonzeitmuster zeigt Bild 2.1 rechts. Die genaue Verlaufsdefinition des Spektralmaximums lautet:

$$s(\omega_{max}(t), t) = \frac{2}{t}e^{-1} \quad \text{und} \quad a(\omega_{max}(t)) = 1/t \quad \text{für} \quad t_1 < t < t_2, \tag{2.4}$$

$$\text{wobei} \quad t_1 = \frac{1}{a(\omega_{A_N})}, \quad t_2 = \frac{1}{a(\omega_{A_0})}. \tag{2.5}$$

Die Teiltonsynthese mit Rechteckfenster liefert, in Abwandlung von Gl. (1.22) für verschwindend kurzes Auswerteintervall T_A, das Synthesesignal

$$\hat{s}(t) = |s(\omega_{max}(t), t)| \cdot \sin(\phi(t)), \tag{2.6}$$

$$\text{wobei} \quad \frac{d\phi}{dt} = \omega_{max}(t). \tag{2.7}$$

Die Momentanphase $\phi(t)$ ist hier weiter nicht von Bedeutung. Der Eingangsimpuls wird also in einen Gleitsinuston umgewandelt. Aus der Sicht einer gehörorientierten Repräsentation wäre dagegen prinzipiell nichts einzuwenden, wenn einerseits seine Frequenzänderung unhörbar schnell wäre. Andererseits müßte seine Energie mit der des Eingangsimpulses im betrachteten Analysefrequenzbereich ω_{A_0} bis ω_{A_N} übereinstimmen. Tatsächlich läßt sich zeigen, daß nur ein Bruchteil der Energie übertragen wird, obwohl die Leistung von stationären Sinustönen durch die Skalierung von Gln. (1.1), (1.3) und (2.6) unverändert

bleibt und obwohl keine Ausgeprägtheit bei der Teiltonextraktion gefordert wurde. Dazu wird die Energie E_δ des unverarbeiteten Impulses zwischen ω_{A_0} und ω_{A_N} über sein Fourier-Leistungsspektrum $|\mathcal{F}\{\delta(t)\}|^2 = 1$ bestimmt:

$$E_\delta = \frac{1}{\pi}\int_{\omega_{A_0}}^{\omega_{A_N}} 1 d\omega = \frac{\omega_{A_N} - \omega_{A_0}}{\pi}. \tag{2.8}$$

Die Energie $E_{\hat{s}}$ der Synthesesinusschwingung nach Gl. (2.6) schätzt man im Zeitbereich ab. Ihre in Gl. (2.4) definierte Hüllkurve wird, mit $1/\sqrt{2}$ multipliziert, als zeitvarianter Effektivwert angesetzt. Integration seines Quadrates im Existenzbereich nach Gl. (2.5) ergibt

$$E_{\hat{s}} \approx \int_{t_1}^{t_2} \left|\frac{s(\omega_{max}(t), t)}{\sqrt{2}}\right|^2 dt = 2e^{-2}(a(\omega_{A_N}) - a(\omega_{A_0})). \tag{2.9}$$

Gegeben sei nun ein Analysefrequenzbereich von $\frac{\omega_{A_0}}{2\pi} = 20\,\text{Hz}$ bis $\frac{\omega_{A_N}}{2\pi} = 5{,}5\,\text{kHz}$ bei einer Analysebandbreite $\frac{a}{\pi} = 0{,}1\,\text{Bark}$. Unter Verwendung der Näherung $1\,\text{Bark} \hat{=} 100$ Hz für Analysefrequenzen von $f < 500$ Hz und $1\,\text{Bark} \hat{=} 0,2f$ darüber [Zwi82] erhält man $\pi^{-1}a(\omega_{A_0}) \approx 10$ Hz und $\pi^{-1}a(\omega_{A_N}) \approx 100$ Hz. Eingesetzt in Gln. (2.8), (2.9) ergibt sich, daß die übertragene Energie um etwas mehr als 20 dB sinkt:

$$\frac{E_{\hat{s}}}{E_\delta} \approx \frac{2e^{-2}(\pi 100 - \pi 10)}{\pi^{-1}(2\pi 5500 - 2\pi 20)} \approx 0,007. \tag{2.10}$$

Die deutliche Unterrepräsentation läßt den Schluß zu, daß diejenigen Signalanteile nicht ausreichend im Teiltonzeitmuster vertreten sind, die flache Momentanspektren aufweisen. Diese werden von transienten Signalanteilen hervorgerufen, für die sich keine quasistationären FTT-Spektren mehr ausbilden. Ein Teilton, der dann aus dem Wert eines lokalen spektralen Maximums berechnet wird, kann eben nicht die Energie erfassen, die sich über die Breite des Spektrums verteilt. Natürlich verkörpert die Ausbildung eines Gleitsinustons nur eine spezielle Situation. Es gibt aber durchaus eine noch ungünstigere: Man stelle sich anhand von Bild 2.1 einen Eingangsimpuls vor, der zu hohen Frequenzen hin zunehmend gedämpft wurde. Dann kann es sogar passieren, daß das lokale Maximum außerhalb des betrachteten Frequenzbereichs zu liegen kommt und das Synthesesignal null bleibt.

2.1.2 Klangverfälschte Repräsentation von Impulsfolgen

Transiente Signalanteile treten üblicherweise nicht als zeitlich hervorgehobene Inseln und über die volle spektrale Breite ausgedehnt im FTT-Spektrum auf. Besonders in Sprachsignalen laufen sie mit denjenigen Anteilen zusammen, die gut durch Teiltöne repräsentierbar sind. Bei einer Impulsfolge zeigt sich der Übergang zwischen beiden Anteilen besonders deutlich. Sie sei zunächst als stationäre Folge von Dirac-Impulsen im Abstand Δt definiert:

$$s(t) = \sum_{n=-\infty}^{+\infty} \delta(t - n\Delta t). \tag{2.11}$$

Durch Ansatz der Rücktransformation der Fourier-Transformierten erscheint $s(t)$ als eine Summe komplexer Exponentialschwingungen mit dem Frequenzabstand $\frac{1}{\Delta t}$ [Mar82]:

$$s(t) = \sum_{k=-\infty}^{+\infty} \frac{1}{\Delta t} e^{j2\pi k \frac{t}{\Delta t}}. \tag{2.12}$$

Das Signal stellt sich somit gleichermaßen als Folge diskreter Impulse oder als Summe von diskreten, amplitudenkonstanten Harmonischen dar. Mit einem Kurzzeitspektralanalysator nach Art der FTT kann man nur auf eine einzige, gemischte Darstellung zugreifen. Ob diese dann zeitliche oder spektrale Signalmerkmale wiedergibt, hängt vom Verhältnis von Merkmalsdauer zu Analysefensterlänge ab. Ist das Analysefenster kurz im Vergleich zum Impulsabstand Δt im obigen Signal, so muß sich für Zeiten in der Nähe eines Impulses ein flaches Momentanspektrum einstellen. Es entsteht dort eine Formation wie in Bild 2.1 im vorigen Abschnitt. Wird das Analysefenster länger als der Impulsabstand, dann tritt die zeitliche Struktur zugunsten der spektralen Struktur in den Hintergrund. Das Betragsspektrum gibt dann mit einer Unschärfe die diskreten Harmonischen aus Gl. (2.12) wieder. Der Umschlagpunkt kann auch spektral durch das Verhältnis Analysebandbreite B zu Harmonischenabstand $\Delta f = \frac{1}{\Delta t}$ ausgedrückt werden. Die Besonderheit bei der FTT besteht nun eben darin, daß die Analysebandbreite durch die Frequenzgruppenanpassung zu höheren Analysefrequenzen ansteigt.

Eine Folge von Impulsen $s(t)$ mit zunehmend geringerem Abstand ist zusammen mit seinem nach Heinbach berechneten Teiltonzeitmuster in Bild 2.2 dargestellt. Die Impulsfolgefrequenz beginnt bei 20 Hz, um sich über eine Dauer von 2 s gleichmäßig auf 200 Hz zu erhöhen. Damit wird auch der Grundfrequenzbereich eines männlichen Sprechers abgedeckt. Es wird angenommen, daß der Frequenzanstieg langsam genug ist, so daß die zuvor angestellten Betrachtungen für die stationäre Impulsfolge weiterhin angewandt werden können.

Das Teiltonzeitmuster teilt sich in zwei Hauptbereiche, die durch einen unterschiedlich breiten Übergangsbereich voneinander getrennt sind. Im einen Hauptbereich finden sich in den Teiltönen die Harmonischen wieder. Da die Impulsfolgefrequenz zunimmt, steigen sie in der Frequenz an. Hier übersteigt die wirksame Analysefensterlänge den Impulsabstand deutlich. Anders ausgedrückt ist hier die Analysebandbreite deutlich geringer als der Harmonischenabstand. Im anderen Hauptbereich, der weißen Fläche, sind die Verhältnisse umgekehrt. Dort werden isolierte Impulsspektren wie in Bild 2.1 links aufgelöst, deren Ausdehnung parallel zur Frequenzachse im wesentlichen unrepräsentiert bleibt. Daß nicht wenigstens Teiltonverläufe wie in Bild 2.1 rechts zu sehen sind, liegt unter anderem an der Wirkung von Glättung, Ausgeprägtheitsschwelle und an der Auswerteintervalldauer.

Im Übergangsbereich kommen flächendeckend Teiltonlinien vor, die in der Frequenz abzufallen scheinen. Dies ist allerdings eine optische Täuschung, tatsächlich bleiben die Teiltonlinien zwischen den Impulsen konstant, um beim nächsten Impuls in der Frequenz nach oben zu springen. Die Annahme, daß die Änderung der Impulsfolgefrequenz langsam genug ist, gilt hier ausnahmsweise nicht. Der Bereich zieht sich zu höheren Frequenzen und größeren Zeiten zu einem unregelmäßigen, schmalen Band zusammen. Der gebogene Verlauf der Bereichsgrenzen ergibt sich aus der Zunahme der Impulsfolgefrequenz in Kombination mit dem Anstieg der Analysebandbreite zu hohen Frequenzen hin und der Bark-Skalierung der Frequenzachse.

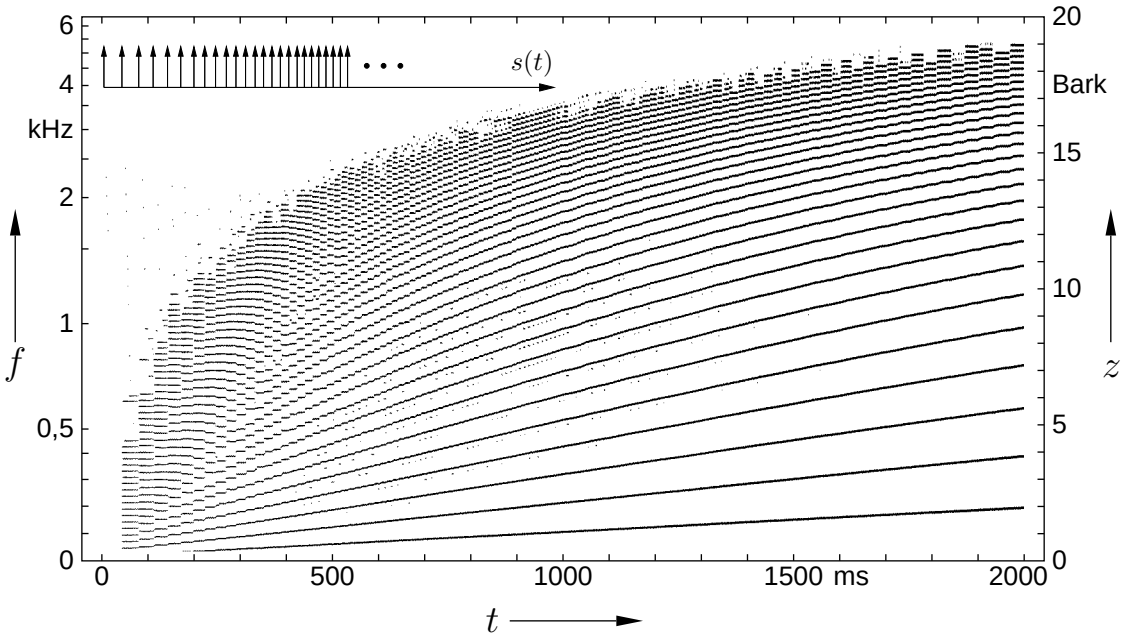

Bild 2.2: *Teiltonzeitmuster einer Impulsfolge $s(t)$, in der die Folgefrequenz von* 20 *auf* 200 Hz *linear über der Zeit ansteigt. Die wachsende obere Grenzfrequenz im Muster ist nach Teiltonsynthese deutlich wahrnehmbar und demonstriert, daß die darüber hörbaren Signalanteile nicht repräsentiert sind.*

Das Synthesesignal zeigt eine charakteristische Klangverfälschung: Während die Schärfe des Originalsignals, ohne weitere Hörempfindungen zu berücksichtigen, durchgängig gleich empfunden wird, beginnt das Synthesesignal dumpf, um sich im Verlauf langsam aufzuhellen. Der in der Hörempfindung fehlende Anteil korrespondiert mit dem beschriebenen unausgefüllten Bereich des Teiltonzeitmusters in Bild 2.2.

Das Beispiel der Impulsfolge verdeutlicht, daß transiente Signalanteile in Spektralbereichen verschieden stark repräsentiert sein können. Manche Bereiche des Momentanbetragsspektrums weisen die ausgeprägten lokalen Maxima auf, die für eine Repräsentation durch Teiltöne ausreichend sind. Zur gleichen Zeit kann in anderen Spektralbereichen ein flaches Momentanbetragsspektrum vorherrschen. Meist ist dies zu höheren Frequenzen hin zunehmend der Fall, aber durchaus nicht immer, wie Abschnitt 3.2.2 demonstrieren wird. Da solche Bereiche dann unterrepräsentiert sind, kann es Klangverfälschungen geben.

2.1.3 Fehlende Knackrepräsentation geschalteter Signale

Es ist bekannt, daß Töne bei Schaltvorgängen die Empfindung eines Knackgeräusches hervorrufen, was mit der spektralen Verbreiterung des Kurzzeitspektrums erklärt wird [Zwi82, S. 68]. Im Teiltonzeitmuster zeigt sich aber beispielsweise beim Ein- oder Ausschalten eines stationären Sinustons nur ein Anklingen bzw. Abklingen des zugehörigen Teiltons [Hei88a, S. 30ff]. Es werden keine zusätzlichen Teiltöne erkannt, die auf eine spektrale Verbreiterung hinweisen. Auch der Hörvergleich von Original- und Synthesesignal zeigt, daß bei geschalteten Signalen die Knackempfindung verloren geht. Anders als bei den oben behandelten Impulsen hängt dies nicht allein mit der Teiltonextraktion zusammen, die bestimmte Energieverteilungen im Spektrum nicht erfassen kann. Vielmehr

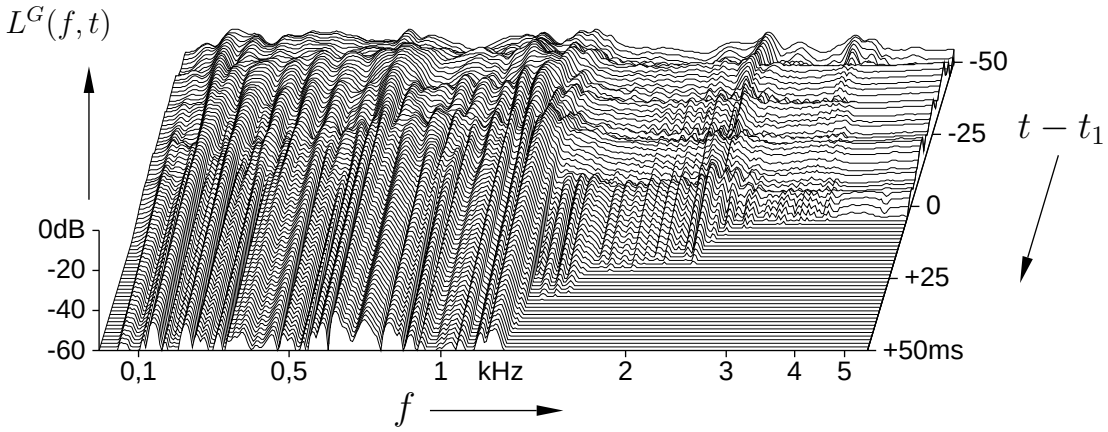

Bild 2.3: *Erstarren und Zusammensinken des von Heinbach verwendeten, geglätteten FTT-Pegelspektrums nach dem Ausschalten eines Sprachsignals. Es handelt sich um das Sprachsignal aus Bild 1.2, das bei $t_1 = 1{,}25$ s 'hart' abgeschaltet wird.*

ist der Grund bereits im FTT-Spektrum zu finden. Zur Demonstration ist der Abschaltvorgang am besten geeignet. Sei hierzu $x(t)$ ein beliebiges kausales Signal, welches zum Zeitpunkt $t_1 > 0$ abgeschaltet werde. Das Eingangssignal s(t) der Spektralanalyse lautet dann

$$s(t) = \begin{cases} x(t) & \text{für} \quad 0 < t \leq t_1, \\ 0 & \text{sonst.} \end{cases} \tag{2.13}$$

Für Zeiten $t \geq t_1$ nach dem Abschalten ergibt sich mit Gl. (1.1) und der Fensterfunktion nach Gl. (1.3) das FTT-Spektrum

$$s(\omega, t) = \int_0^{t_1} 2ax(\tau) e^{a(t-\tau)} e^{-j\omega\tau} d\tau. \tag{2.14}$$

Der Integralbeitrag zwischen den Grenzen t_1, t fällt weg, weil $s(t)$ dort zu null wird. Die Integralgrenzen sind dadurch zeitunabhängig. Durch Substitution von $t = t_1 + t'$ läßt sich ein zeitabhängiger Exponentialfaktor vor das Integral ziehen, welches sich als Spektrum zum Zeitpunkt t_1 erweist:

$$s(\omega, t_1 + t') = e^{-at'} s(\omega, t_1), \quad \text{für} \quad t' \geq 0. \tag{2.15}$$

Man erkennt, daß das FTT-Spektrum beim Abschalten erstarrt und in sich zusammensinkt. Insbesondere zeigt es keine spektrale Verbreiterung an, woran auch die Betragsbildung und Glättung nichts ändert. Dies unterstreicht das Beispiel in Bild 2.3. Wohl erfolgt das Absinken bei hohen Frequenzen schneller als bei tiefen, weil sich die Fensterkonstante a nach Gl. (1.2) proportional zur Frequenzgruppe verhält. Dennoch wird deutlich, daß sich an der Verteilung der lokalen Pegelmaxima über der Frequenz kaum noch etwas ändern kann. Somit weisen auch keine zusätzlichen Teiltöne auf einen besonderen Abschalteffekt hin.

Auch beim Einschalten gibt es keine kurzzeitspektrale Verbreiterung, was in Abschnitt 3.3.5 noch genauer erläutert wird. Der einzige Schalteffekt ist hier das vorübergehende Auftreten von Nebenmaxima. Durch Glättung und Ausgeprägtheitsschwelle hindurch

können sie sich aber normalerweise nicht als Teiltöne etablieren. [1] Die Beobachtungen lassen sich auch auf Verlaufssprünge und allgemein auf schnelle Hüllkurvenänderungen im Signalverlauf übertragen. Die von Heinbach verwendete Fensterfunktion weist folglich den Mangel auf, daß sie die bei solchen Signalen wahrnehmbaren Effekte nicht darstellen kann.

2.1.4 Zusammenstellung günstiger Effekte bei Sprache

Unterrepräsentation von transienten Anteilen im Heinbachschen TTZM-Verfahren tritt in der Praxis von Sprachsignalen meist nicht in der Deutlichkeit auf, wie das die bisher behandelten Signale demonstrieren. Dafür können vorgreifend einige begünstigende Effekte angegeben werden, die meist im weiteren Verlauf der Arbeit in speziellerem Kontext wieder auftauchen:

- Transiente Anteile können durch vorherige Raumübertragung aus der Sicht des FTT-Spektrums in quasistationäre Anteile umgewandelt worden sein. Sehr gut funktioniert dies bei schlechter Raumakustik, wenn wenig Direktschall übertragen wird und nicht zu kurze Nachhallzeiten vorliegen. Die Raumübertragung verzerrt dann beispielsweise das FTT-Spektrum des Impulses in Bild 2.1 derart, daß jederzeit genügend lokale Maxima zur Teiltonrepräsentation gefunden werden. Gleichzeitig erfolgt eine zeitliche Streuung der Hauptenergiekonzentrationen in verschiedenen Spektralbereichen. Flache, über die spektrale Breite verteilte Energiekonzentrationen verschwinden.

- Transiente Anteile des FTT-Spektrums sind meist stark mit quasistationären ineinander verwoben. Unrepräsentierte transiente Anteile lassen deshalb genügend quasistationäre übrig, die die spektral/zeitliche Hüllfläche des Spektrums noch gut beschreiben. Damit kann die wesentliche Information durch Teiltöne allein dargestellt werden (Abschnitt 3.2.3, Bild 3.4). Ausgedehnte spektrale Bereiche wie bei der Impulsfolge fallen selten weg. Verfälschungen durch kleine Lücken sind dennoch unvermeidbar, wohl aber von subtilerer Natur (Tonalisierung, Abschnitt 2.5)

- Kleine Analysebandbreiten, bezogen auf die Frequenzgruppenbreite des Gehörs, benachteiligen das Entstehen von transienten Anteilen im FTT-Spektrum. Vereinfacht zeigt sich dies in Bild 2.2 darin, daß der Bereich ohne Teiltöne bei sprachrelevanten Grundfrequenzen bereits relativ klein ist. Allgemein werden Verwebungen derart getrimmt, daß spektral ausgedehnte transiente Bereiche ohne eingeflochtene quasistationäre Anteile und damit ohne Teiltöne unwahrscheinlicher werden. Um andere Nachteile zu vermeiden, muß allerdings die Analysebandbreite viel größer gemacht werden (Abschnitt 3.4.3.2).

- Störungen aus der Teiltonsynthese helfen, unrepräsentierte transiente Anteile zu verschleiern. Erstens können sie die hinterlassenen spektralen Lücken etwas auffüllen.

[1] Zwar treten bei Heinbach Nebenteiltöne im Teiltonzeitmuster eines 10 ms langen, rechteckförmig geschalteten 1 kHz–Sinustonimpulses auf [Hei88a, S. 32]. Sie setzen aber erst *nach* dem Abschalten im Verlauf der Abklingphase des Spektrums ein, so daß ihre repräsentierte Energie viel zu gering ist.

Zweitens ermöglichen sie zusammen mit bestimmten Analyseparametern einen speziellen Verbundeffekt, der transiente Anteile regelrecht durch 'Rauschstöße' imitieren kann (Abschnitt 5.1.7.2).

Aber keiner dieser Effekte kann die nachgewiesene, prinzipielle Benachteiligung von transienten Anteilen beheben.

2.2 Glättung der Schmalbandhüllkurve

Nach einer TTZM-Verarbeitung klingen Sprachsignale, die ursprünglich in reflektionsarmer Umgebung aufgenommen wurden, wie raumübertragen 'hallig'. Die folgenden Untersuchung weist eine markante Signalverfälschung nach, die sich ähnlich wie eine Raumübertragung auswirkt. Sie hängt mit einer zu geringen FTT-Analysebandbreite zusammen. Es gibt noch zwei weitere, subtilere Beiträge zur raumübertragungsähnlichen Verfälschung, die beide in einem anderen Zusammenhang abgehandelt sind. Dazu zählt die Unterrepräsentation von transienten Anteilen, die nach Abschnitt 2.1.4 auch im Diffusfeld einer Raumübertragung nur schwach repräsentiert sind. Weiterhin ahmt die in Abschnitt 2.3 noch zu behandelnde, wahrnehmbare Phaseninkohärenz von Synthesesinusschwingungen den quasi zufälligen Phasengang einer Raumübertragungsfunktion nach.

Bei Raumübertragung erfahren stark amplitudenmodulierte Schmalbandsignale eine Hüllkurvenglättung. Signalabschnitte mit geringer Intensität werden durch die verzögerten Raumreflektionen der stärkeren Abschnitte angefüllt. Diese Tatsache machen sich bestimmte Verfahren zur objektiven Beurteilung der Raumakustik zunutze, indem sie die Bewahrung der Modulation in verschieden Bändern messen und als wesentlich für eine gute Sprachübertragung ansehen [Hou85]. Umgekehrt existieren Verfahren zu Nachhallunterdückung, die die Modulationstiefe nachträglich wiederherzustellen suchen [Sch91]. Zum Nachweis einer Hüllkurvenglättung beim TTZM-Verfahren dient als Testsignal

$$s(t) = A(t)\cos(\omega_T t) \tag{2.16}$$

ein Sinuston $\frac{\omega_T}{2\pi} = 1$ kHz als Schmalbandträger. Seine variable Amplitude $A(t)$ stimmt für den Fall $A(t) \geq 0$ mit der Signalhüllkurve überein. Dies gilt beispielsweise für einen positiven Rechteckpuls mit Tastverhältnis 1:1 und der Periodendauer $T = 100$ ms, definiert durch

$$A(t) = \begin{cases} A & \text{für} \quad nT \leq t < \left(n + \frac{1}{2}\right) T \quad \text{mit} \quad n \in \{0, 1, 2, \ldots\}, \\ 0 & \text{sonst.} \end{cases} \tag{2.17}$$

Die Länge der Periode liegt im Wahrnehmungsbereich der Schwankungsstärke, in dem das Gehör Hüllkurvenschwankungen verfolgen kann [Ter68a]. Ebenso unterschreitet die Modulationsgrundfrequenz mit $\frac{1}{T} = 10$ Hz die Analysebandbreite des Heinbachschen TTZM-Verfahrens von 16 Hz bei 1 kHz. Bereits im Teiltonzeitmuster in Bild 2.4a wird erkennbar, daß der Teiltonpegel in den Lücken nur wenig absinkt. Das Synthesesignal weist eine deutliche Glättung seiner Hüllkurve auf, die durch Gleichrichtung und Tiefpaßfilterung oberhalb 500 Hz bestimmt und als Pegelverlauf $L_e(t)$ in Bild 2.4c eingetragen wurde. Die kleine Überhöhung im ansteigenden Verlauf beruht auf der Wirkung von kurzzeitig existierenden Nebenteiltönen, die auch im Teiltonzeitmuster erkennbar sind. Sie spielt hier weiter keine Rolle.

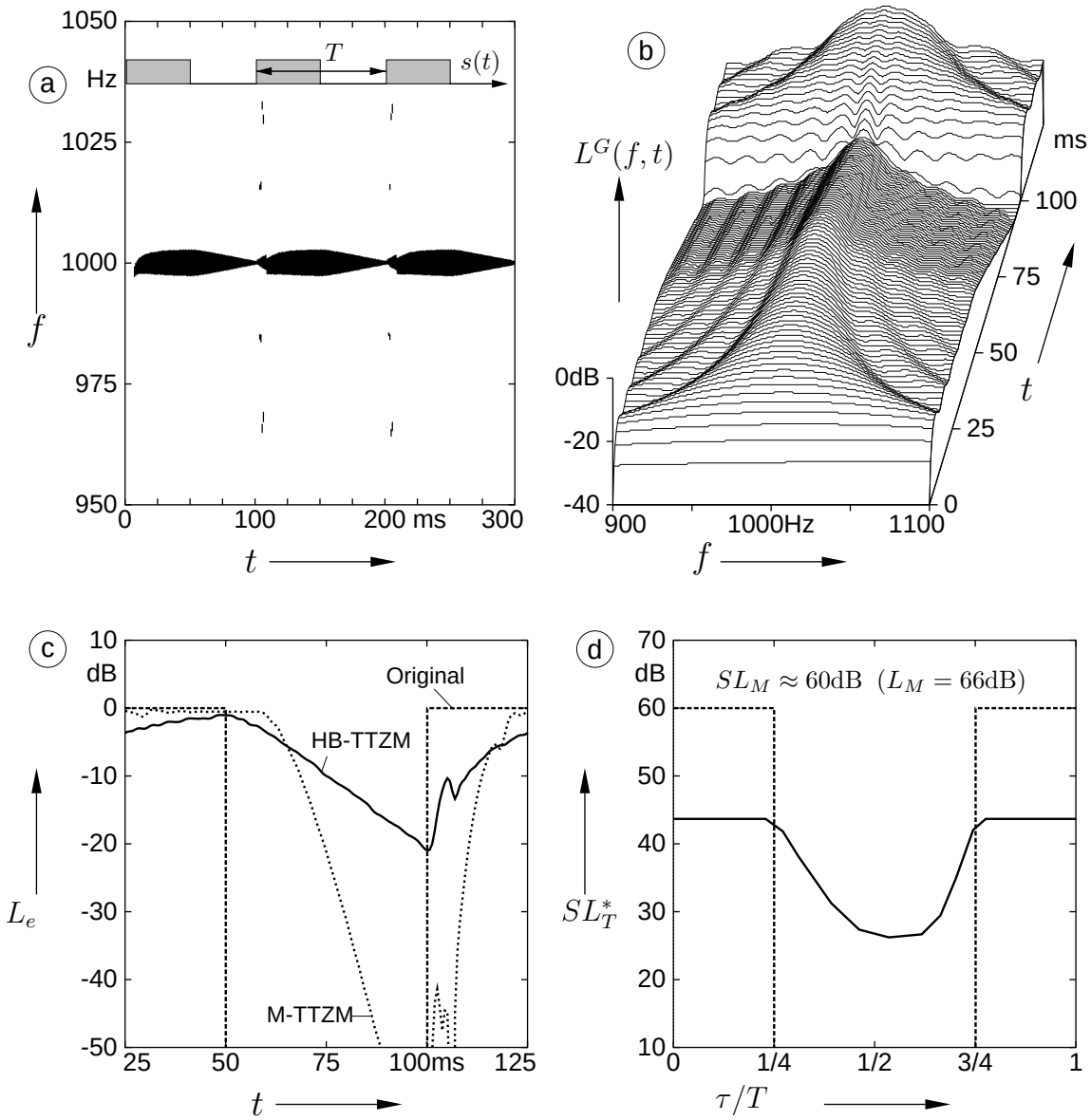

Bild 2.4: *Glättung der Schmalbandhüllkurve anhand eines mit* $T^{-1} = 10$ Hz *rechteckmodulierten Sinustones* $s(t)$ *bei* 1 kHz*: a) Teiltonzeitmuster, b) FTT-Pegelspektrum bei Glättung, c) Hüllkurvenpegel des Synthesesignals (HB-TTZM) im Vergleich zum Originalsignal und zum Synthesesignal eines modifizierten Analyseverfahrens (M-TTZM). d) Mithörschwellen-Periodenzeitmuster für einen* 3 ms *langen* 1 kHz*-Testtonimpuls bei einem Darbietungsschallpegel von* $L_M = 66$ dB *mit Daten aus [Zwi82], womit die Wahrnehmbarkeit der Hüllkurvenglättung belegt werden kann (siehe Text). Die Höcker im Anstieg des Synthesesignals in c) rühren von den isolierten Teiltönen im* 100 ms *Abstand in a) her, welche kurzzeitig ausgeprägte Nebenmaxima in b) markieren.*

Daß beim Hörvergleich von Original- und Synthesesignal eine deutliche Abschwächung der Modulation wahrzunehmen ist, kann anhand von bekannten psychoakustischen Daten veranschaulicht werden. Für das Originalsignal ist hierzu das Mithörschwellenperiodenmuster aus [Zwi82, Abb. 9.8] in Bild 2.4d eingezeichnet. Es gibt für einen Wiedergabepegel von $SL_M = 66$ dB über der Ruhehörschwelle, also etwa 70 dB Schalldruckpegel, die Mithörschwelle eines 3 ms langen 1kHz–Tonimpulses im Verlauf der Modulationsperiode an. Dieser Mithörschwellenverlauf wird nur in der Form und ohne Rücksicht auf seine

absolute Pegel- und Zeitlage mit dem Hüllkurvenpegelverlauf in Bild 2.4c verglichen. Dieser sinkt deutlich langsamer als die Testtonmithörschwelle ab und steigt insgesamt viel langsamer wieder an, auch wenn die maximale Absenkung mit rund 20 dB ähnlich ist. Offensichtlich kann das Gehör viel schneller und länger in die Lücke des Originalsignals 'hineinhören' als dies nach TTZM-Verarbeitung der Hüllkurvenpegel des Synthesesignals zulassen würde. Daraus folgt, daß die Hüllkurvenübergänge gegenüber dem Original wahrnehmbar geglättet sind.

Das Ausmaß der Hüllkurvenglättung hängt direkt mit dem Analysefenster der FTT zusammen. Zur Verdeutlichung wird das FTT-Betragsspektrum an der Frequenz ω_T betrachtet, in deren Umgebung nach Bild 2.4b auch das Spektralbetragsmaximum als Voraussetzung für die Teiltonextraktion gefunden wird. Die Wirkung der zeitlichen Glättung nach Gln. (1.18) bis (1.20) wird dabei vernachlässigt, sie verstärkt den nachzuvollziehenden Glättungseffekt der Hüllkurve zusätzlich. Mit Gln. (1.6), (2.16) und unter Anwendung von $\cos x = \frac{1}{2}(e^{jx} + e^{-jx})$ erhält man bei nichtnegativen Fensterfunktionen und Amplituden:

$$|s(\omega,t)|_{\omega=\omega_T} = \left|\left(A(t)\cos(\omega_T t)\cdot e^{-j\omega_T t}\right) * h_{\omega_T}(t)\right| \tag{2.18}$$

$$= \left|\frac{A(t)}{2} * h_{\omega_T}(t) + \left(\frac{A(t)}{2}\cdot e^{-j2\omega_T t}\right) * h_{\omega_T}(t)\right| \tag{2.19}$$

$$\approx \frac{|A(t)|}{2} * h_{\omega_T}(t). \tag{2.20}$$

Der zweite Term in Gl. (2.19) wird gemäß den Betrachtungen in Abschnitt 1.4.2 vernachlässigt. Man kann Gl. (2.20) so interpretieren, daß die Hüllkurve des Signals durch ein System geführt wird, welches die Fensterfunktion $h_{\omega_T}(t)$ als Impulsantwort besitzt. Mit der bisherigen Fensterfunktion nach Gl. (1.3) ergibt sich eine Filterung mit einem Tiefpaß erster Ordnung. Je kleiner aber die Fensterlänge ist, um so weniger wird die Hüllkurve geglättet. Zum Vergleich ist in Bild 2.4c die Hüllkurve eines Synthesesignals eingetragen, bei dessen Analyse eine andere, weniger asymmetrische Fensterfunktion mit einer etwa um ein Drittel kürzeren Fensterlänge verwendet worden ist. Bei der bisherigen Fensterfunktion verbietet sich jedoch die damit einhergehende Veränderung der spektralen Eigenschaften, wie in Abschnitt 2.4 gezeigt werden wird.

Vergleichbare Beobachtungen ergeben sich, wenn man anstelle des Sinustons ein Schmalbandrauschen verwendet. Das TTZM-Verfahren glättet also die Hüllkurve in beliebigen Frequenzbändern mit einer Breite, die etwa der Analysebandbreite entspricht. Dieser Glättungseffekt tritt bei der Heinbachschen Wahl der Analysebandbreite durch die Vor- und Nachhörschwelle des Gehörs hindurch in Erscheinung.

2.3 Störungen im Übergang zeitlicher/spektraler Auflösung

Ob sich ein Signalmerkmal als zeitliche oder als spektrale Struktur im FTT-Spektrum ausprägt, hängt nach Abschnitt 2.1.2 vom Verhältnis Analysefensterlänge zu zeitlicher Merkmalslänge oder, alternativ, vom Verhältnis von Merkmalsbreite im Fourier-Spektrum zur Analysebandbreite ab. Deshalb kann sich beispielsweise ein Dreitonkomplex in Gestalt

von drei stationären Teiltönen oder als zeitvarianter Einzelteilton in das Teiltonzeitmuster abbilden. Zum Nachweis des gehörähnlichen Verhaltens hat sich Heinbach bereits mit diesem Übergang auseinandergesetzt [Hei88a, S. 33-39], ohne jedoch Auswirkungen auf das Synthesesignal zu berücksichtigen. Aber die Mischrepräsentationen im Übergangsbereich überfordern die Teiltonsynthese, die deshalb zwei charakteristische Störungstypen produziert.

Zuerst werden elementare Modulationsformen eines stationären Sinusträgers behandelt, indem ein Gleitsinus als Modulator verschiedene Zustände im Übergang zwischen beiden Auflösungsmöglichkeiten darstellt. Zu den Modulationsformen zählen die Amplitudenmodulation mit Modulationsgrad eins, die Zweitonschwebung und ein Beispiel für eine Frequenzmodulation. Die Erkenntnisse aus dem Studium spezieller Teiltonkonstellationen und ihrer Synthese läßt sich abschließend auf Sprache übertragen. Demnach wird das verarbeitete Signal ständig von zwei speziellen Störteppichen begleitet, die zur Verminderung der erzielbaren Sprachqualität beitragen. Allerdings erweist sich dies später als nützlich, um unterrepräsentierte transiente Anteile zu verschleiern (siehe Abschnitt 2.1.4 und später 5.1.7.2).

2.3.1 Amplitudenmodulation

Ein Sinusträger der Frequenz ω_T sei durch einen Sinuston der Frequenz ω_M mit Modulationsgrad m in der Amplitude moduliert. Bei willkürlicher Annahme von Startphasen kann man beispielsweise Testsignale durch die zeitvariante Formulierung

$$s(t) = \frac{A}{1+m}(1 - m\cos(\omega_M t))\cos\omega_T t \tag{2.21}$$

festlegen, wobei hier die maximale Amplitude unabhängig von m immer A ist. Umformung führt auf einen stationären Dreitonkomplex als spektrale Interpretation. Darin gruppieren sich zwei Seitenschwingungen im Frequenzabstand $\pm\omega_M$ um eine stationäre Trägerschwingung bei ω_T:

$$s(t) = \frac{A}{1+m}\left[\cos(\omega_T t) - \frac{m}{2}\cos\left((\omega_T + \omega_M)t\right) - \frac{m}{2}\cos\left((\omega_T - \omega_M)t\right)\right]. \tag{2.22}$$

Speziell wird ein Testsignal mit Modulationsgrad $m = 1$ verwendet, wodurch die nach Gl. (2.21) zeitvariante Trägeramplitude bis auf null herab reicht. Die Trägerfrequenz erhält den Wert $\frac{\omega_T}{2\pi} = 1\,\text{kHz}$, und die Modulationsfrequenz erhöht sich zeitlinear von 0 auf $\frac{\omega_M}{2\pi} = 60\,\text{Hz}$, bei einer Signaldauer von 1 s. Als Dreitonkomplex betrachtet liegen die beiden Seitenschwingungen um 6 dB unter der mittleren und entfernen sich nach Maßgabe der Modulationsfrequenz langsam von dieser. Dabei überstreichen die Entfernungen die Analysebandbreite, die knapp 20 Hz bei 1 kHz beträgt, so daß ein Übergang zwischen beiden Interpretationen erreicht wird.

2.3.1.1 Teiltonzeitmuster und Kurzverläufe

In Bild 2.5 oben ist das Teiltonzeitmuster in drei Bereiche unterteilt zu sehen. Zur besseren Darstellung im Frequenzausschnitt wurde der Analysefrequenzabstand bei der Berechnung

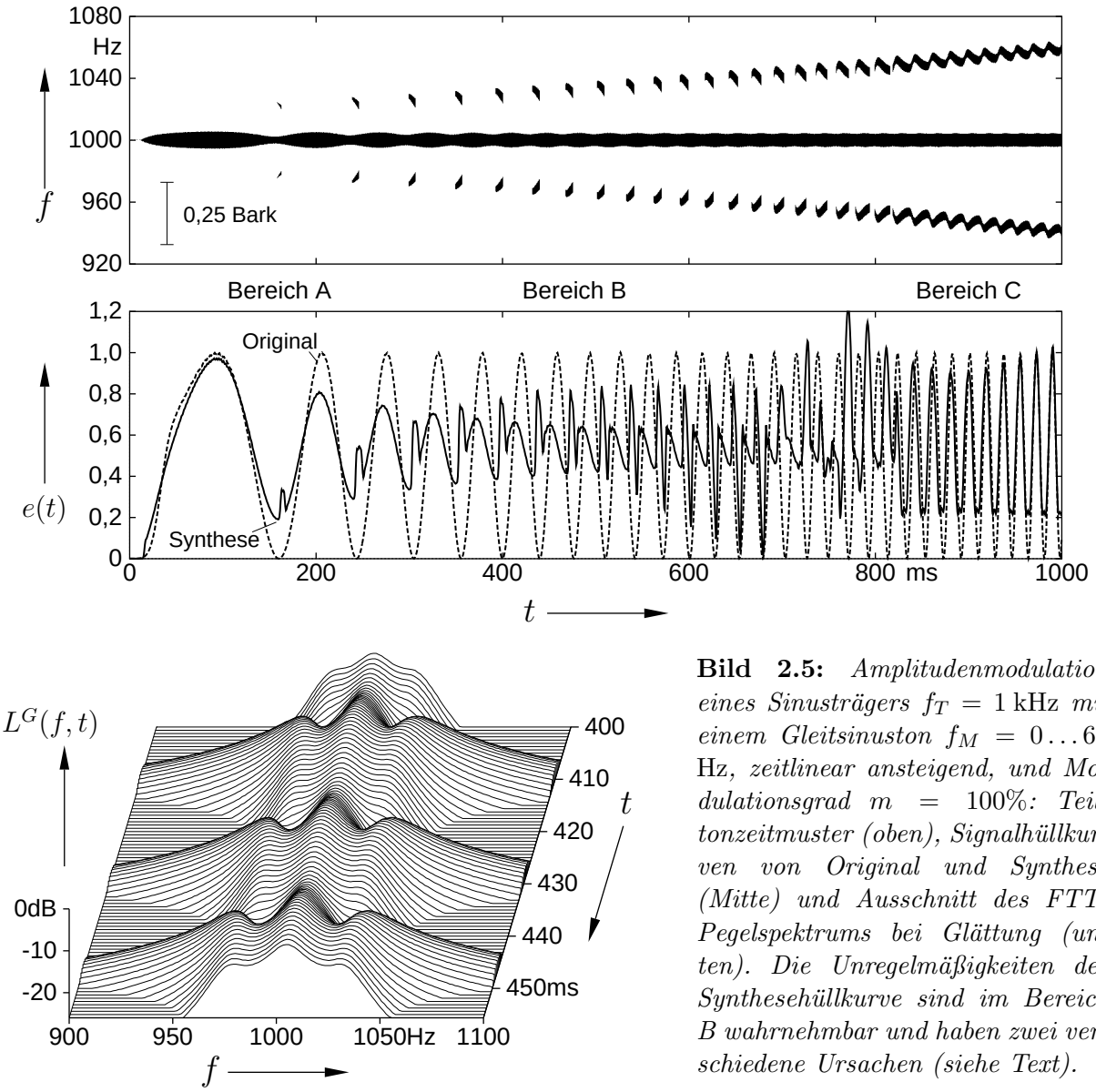

Bild 2.5: *Amplitudenmodulation eines Sinusträgers $f_T = 1\,\mathrm{kHz}$ mit einem Gleitsinuston $f_M = 0 \ldots 60$ Hz, zeitlinear ansteigend, und Modulationsgrad $m = 100\%$: Teiltonzeitmuster (oben), Signalhüllkurven von Original und Synthese (Mitte) und Ausschnitt des FTT-Pegelspektrums bei Glättung (unten). Die Unregelmäßigkeiten der Synthesehüllkurve sind im Bereich B wahrnehmbar und haben zwei verschiedene Ursachen (siehe Text).*

feiner als sonst gewählt. In Bereich A ist die Modulation noch sehr langsam, man erkennt die Änderung der Trägeramplitude deutlich am Teiltonpegel, der jedoch nur beschränkt absinkt. Hier macht sich bereits die Glättung der Schmalbandhüllkurve nach Abschnitt 2.2 bemerkbar. Zum Zeitpunkt des Pegelminimums ist ein kurzzeitiges Auftreten von Seitenteiltönen symmetrisch zum Trägerteilton zu beobachten. Ihr Frequenzabstand ist größer und ihr Pegel (nicht sichtbar) niedriger als die Werte der Seitenschwingungen in diesem Moment. In Bereich B, dem Übergangsbereich, schließen sie dichter aneinander an, ihr Pegel steigt und die Frequenzen stimmen zunehmend mit den Momentanfrequenzen der Seitenschwingungen überein. Die Pegelschwankung des Trägerteiltons wird immer kleiner. Schließlich treten in Bereich C die Seitenschwingungen als ununterbrochene Teiltonlinien auf, so daß die spektrale Interpretation als Dreitonkomplex vorliegt. Eine leichte Frequenzmodulation der Seitenteiltöne und die Amplitudenmodulation aller drei Teiltöne nehmen mit größerer Entfernung der Seitenschwingungen weiter ab.

Die kurzzeitig auftretenden Seitenteiltöne sollen der Kategorie der *Kurzverläufe* zugerechnet werden. Das besondere an ihnen ist, daß sie ohne vorangehenden allmählichen Pegel-

anstieg plötzlich auftauchen und genauso unvermittelt wieder verschwinden. Im Gegensatz dazu würden Teiltöne von auch noch so kurzen Tonimpulsen immer An- und Abklingfahnen aufweisen (vgl. Bild 2.4a). Kurzverläufe sind deshalb möglich, weil sich die Form des Momentanspektrums auf einem bestimmten Pegelniveau ein wenig verändert, so daß sich beispielsweise ein ausgeprägtes Maximum bildet, wo vorher nur ein unausgeprägtes oder gar keines vorlag. Aus demselben Grund können sie auch schnell wieder verschwinden. Zur Verdeutlichung ist in Bild 2.5 unten ein Ausschnitt des FTT-Pegelspektrums im Bereich B zu sehen.

Bei anderen Modulationsgraden als eins ist das Auftreten von Kurzverläufen im Bereich der Seitenschwingung allgemein eine Frage des Frequenz- und Pegelabstandes von Seiten- und Trägerschwingung im Vergleich zum Selektionsverhalten des Analysetiefpasses. Beispielsweise verschwinden die Kurzverläufe bei 20 Hz Modulationsfrequenz und einem Modulationsgrad $m < 0{,}7$ völlig, entsprechend einem Pegelabstand von mehr als 9 dB. Wird der Frequenzabstand größer, können sie wieder sichtbar werden und schließlich in unterbrechungslose Verläufe übergehen. Mit dieser Problematik der Simultanverdeckung beschäftigt sich Abschnitt 2.4.

2.3.1.2 Synthese und grundsätzliche Störungsursachen

Abruptes Auftauchen und Verschwinden eines hochpegeligen Kurzverlaufes im Übergangsbereich bewirkt in der zugeordneten Synthesesinusschwingung entsprechend harte Übergänge. Damit sie im verarbeiteten Signal nicht als Knacke hörbar werden, müssen sie durch ein geeignetes Synthesefenster sanftere Übergänge erhalten (Abschnitt 1.5.4). Das Dreieckfenster, das die Übergänge etwas glättet, reicht allerdings noch nicht aus, um die wahrnehmbaren spektrale Verbreiterungen zu vermeiden. Diese Art von Störungen sollen als *Synthesefenster-kontrollierbare Störungen* bezeichnet werden. Sie sind, zumindest bei den bisherigen Fenstern der Teiltonsynthese, eher von breitbandiger Natur. Zwar könnte man sie mit einem deutlich längeren und damit stärker glättenden Dreieckfenster schließlich unterdrücken. Allerdings wird dann die Glättung der Schmalbandhüllkurve nach Abschnitt 2.2 künstlich erhöht und der Nutzeffekt dieses Störungstyps vermieden (Abschnitt 5.1.7.2).

Tatsächlich gibt es aber noch eine andere Störproblematik, deren Ursachen in der Phasenrekonstruktion der Synthese zu suchen sind. Dies läßt sich gut anhand der Hüllkurven $e(t)$ von Original und Synthese in der Mitte von Bild 2.5 beschreiben. Sie wurden durch Gleichrichtung und anschließende Tiefpaßfilterung ab 500 Hz ermittelt und sind zeitrichtig zum Teiltonzeitmuster aufgetragen. Die Hüllkurven sind wahrnehmungsrelevant, weil das Teiltonzeitmuster eine Breite von 1 Bark nicht überschreitet und somit Synthesesinusschwingungen vorgibt, die innerhalb einer Frequenzgruppe zusammentreffen [Zwi82].

In Bereich A wird die Synthesehüllkurve von der Synthesesinusschwingung des durchgehenden Teiltons dominiert. Damit folgt sie zunächst etwa der Vorgabe des Originals. Zu höheren Modulationsfrequenzen hin macht sich bald die Glättung der Schmalbandhüllkurve nach Abschnitt 2.2 bemerkbar, wodurch die Hüllkurvenschwankung abflacht. Zeitgleich mit dem Lebenszyklus der Kurzverläufe sind zunächst kleine Spitzen im sonst sinusähnlichen Verlauf zu erkennen. Ihre Höhe wächst schnell an, so daß sie in Bereich B die schon ziemlich geglättete Hüllkurve wesentlich beeinflussen. Die resultierende Unregelmäßigkeit im Unterschied zur Originalhüllkurve ist gut hörbar, zu Beginn von Bereich B ebenso

noch die Hüllkurvenglättung. Sobald die Seitenteiltöne im Bereich C keine Unterbrechungen mehr aufweisen, normalisiert sich die Synthesehüllkurve schlagartig.

Wenn zulässige Synthesefenster nur die Sprünge der Hüllkurve beim Auf- und Abtauchen der Kurzverläufe ausglätten helfen, dann muß man die Unregelmäßigkeiten während der Lebensdauer der Kurzverläufe über die Synthesephasen kontrollieren. Hier zeigen sich einerseits die Grenzen der einfachen Phasenrekonstruktion in der Teiltonsynthese, die auf stetige Fortsetzung von Phasen innerhalb eines durchgehenden Verlaufs ausgerichtet ist. Ansammlungen von Synthesesinusschwingungen werden phasenmäßig nicht aufeinander abgestimmt, was sich hier im Übergang zwischen den Interpretationen eben als problematisch erweist. Diese Art wahrnehmbarer Störungen sollen *Phaseninkohärenz-bedingte Störungen* heißen. Sie sind von Natur aus schmalbandig, weil sie nur innerhalb einer Frequenzgruppe existieren.

2.3.2 Zweitonschwebung

Die zuvor gemachten Beobachtungen werden nun anhand einer weiteren Modulationsform ergänzt. Geht der Modulationsgrad m in Gl. (2.22) gegen unendlich, so fällt der linke Term weg. Man erkennt die Überlagerung zweier Sinustöne mit Frequenz $\omega_T \pm \omega_M$ und Amplitude $\frac{A}{2}$, welche die spektrale Interpretation einer Schwebung verkörpern. Die aus Gl. (2.21) herleitbare, zeitvariante Interpretation sieht dagegen einen Sinuston der Frequenz ω_T, dessen Amplitude die Form der gleichgerichteten Modulatorschwingung aufweist und dessen Phase im Amplitudenminimum um π springt. Das spezielle Testsignal erhält wieder die Trägerfrequenz 1 kHz, die Modulationsfrequenz steigt gleichmäßig über der Signaldauer, von 0 bis herauf zu 20 Hz in 1 s. Anders interpretiert entspricht dies dem Auseinanderdriften zweier gleichstarker Sinustöne mit der Startfrequenz 1 kHz und den Endfrequenzen 980Hz und 1020Hz.

2.3.2.1 Teiltonzeitmuster, Spaltungen und Verschmelzungen

Bild 2.6 oben zeigt das Teiltonzeitmuster in seinen drei Bereichen, wieder berechnet mit dichterem Analysefrequenzabstand als üblich. In Bereich 1 wird ein einzelner Teiltonverlauf erkannt, der sich im Amplitudenminimum kurzzeitig in ein Teiltondoppel mit mäßigem Pegel aufspaltet. Der Frequenzabstand des Doppels überschreitet den der spektralen Interpretation der Schwebung. Tatsächlich hängt der Abstand hier in erster Linie von der Analysebandbreite ab. Im Bereich 2, dem Übergangsbereich, verkürzt sich die Dauer des wiederkehrenden Einzeltons, so daß schließlich die Dauer der eingeflochtenen Doppel unterschritten wird. Dabei kehren sich die Pegelverhältnisse von Einzelton und Doppel um, die Frequenzen des Doppels biegen in den Verlauf der spektralen Interpretation ein. Diese wird in Bereich 3 erreicht, wenn die Spur des Einzeltons ausgelaufen ist und sich zwei ununterbrochene Teiltöne ausbilden können. Solange ihr Frequenzabstand nicht groß genug ist, tragen sie noch eine leichte Frequenz- und (nicht sichtbare) Amplitudenmodulation im Rhythmus der Differenzfrequenz. [2]

[2] Im Gegensatz zu anderen TTZM-Abbildungen verändert sich Bild 2.6 oben auffällig, wenn zur Frequenzapproximation anstelle des Feldtkeller-Verfahrens die Parabelapproximation angewandt wird (Anhang B.1). Im Bereich 2 würde man eine Art 'Zopfmuster' erkennen, welches sich mit raschen, wechselseitigen Unterbrechungen in zwei Stränge entflechtet. Weil beide Approximationen nicht die Anzahl der

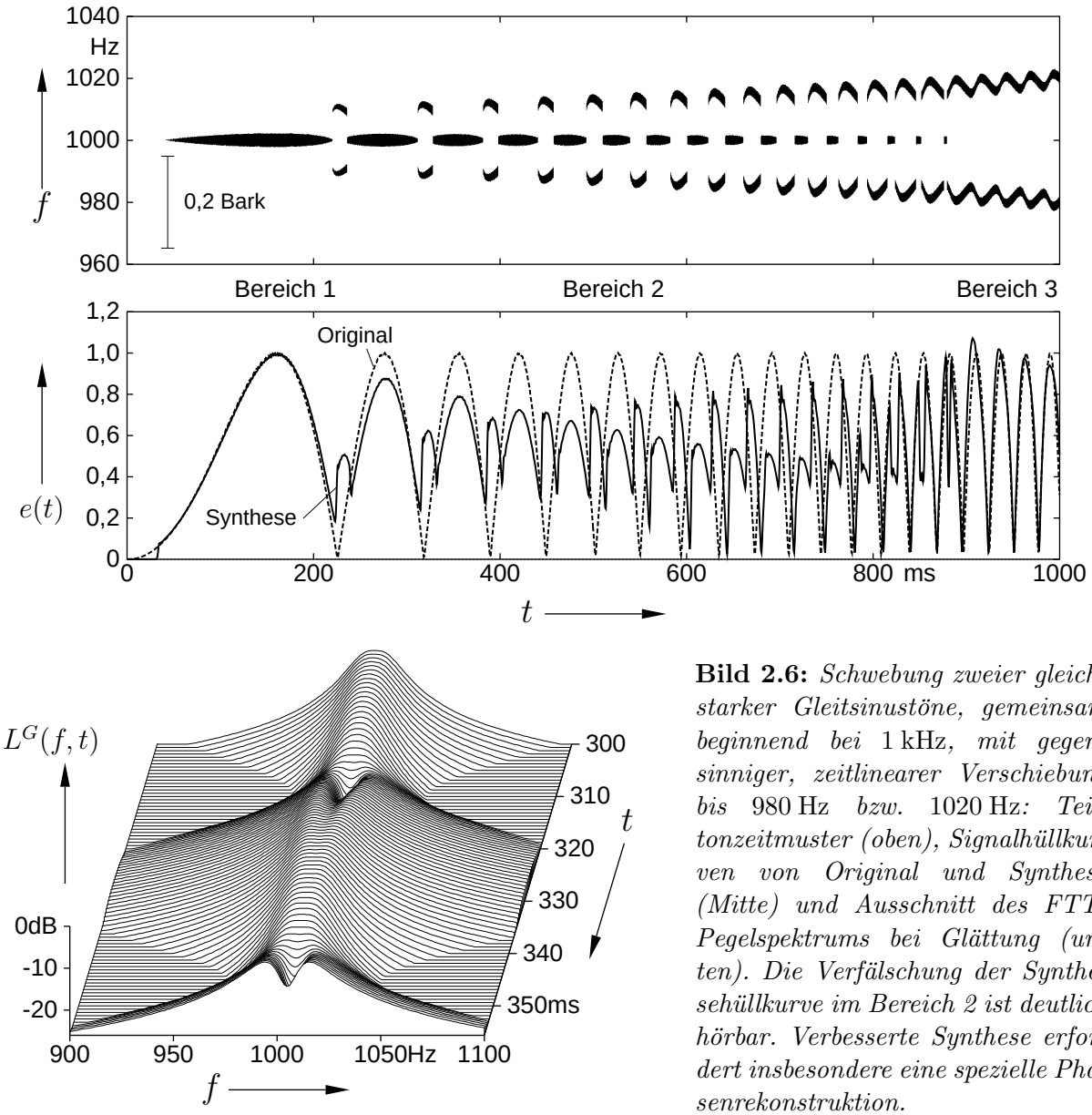

Bild 2.6: *Schwebung zweier gleichstarker Gleitsinustöne, gemeinsam beginnend bei* 1 kHz*, mit gegensinniger, zeitlinearer Verschiebung bis* 980 Hz *bzw.* 1020 Hz*: Teiltonzeitmuster (oben), Signalhüllkurven von Original und Synthese (Mitte) und Ausschnitt des FTT-Pegelspektrums bei Glättung (unten). Die Verfälschung der Synthesehüllkurve im Bereich 2 ist deutlich hörbar. Verbesserte Synthese erfordert insbesondere eine spezielle Phasenrekonstruktion.*

Das Pegelspektrum in Bild 2.6 unten zeigt für einen Ausschnitt aus Bereich 2, warum ein Wechsel zwischen einem und zwei Teiltönen auf relativ hohem Pegelniveau auftritt. Man erkennt die Zeiten, zu denen entweder ein oder aber, mit einem mittigen Minimum, zwei lokale Spektralmaxima vorhanden sind. Man kann zeigen, daß die Minima von singulären Nullstellen des ungeglätteten, zeitvarianten Leistungsspektrums herrühren. Mit Blick auf die zeitliche Interpretation markieren sie die Zeitpunkte des Nulldurchgangs der Hüllkurve.

Schnelle Wechsel zwischen einem und zwei Teiltönen können als Auftreten von Kurz-

gefunden Spektralmaxima modifizieren, sind die Zeiten genau dieselben, zu denen ein bzw. zwei gleichzeitige Teiltöne vorliegen. Wegen der Restwelligkeiten, die vom Term der negativen Frequenzhalbachse in Gl. (1.15) stammen, liegt mal das Maximum des oberen Stranges, mal das des unteren Stranges über der Ausgeprägtheitsschwelle. Die integrierende Wirkung der Approximation von Feldtkeller dagegen schiebt das jeweilige Maximum in die gemeinsame Mitte der Stränge und liefert damit für eine Signalrekonstruktion den hier dargestellten, eher günstigeren Fall.

verläufen interpretiert werden, im Sinne der Definition des vorhergehenden Unterabschnittes. Jedoch kommt eine zusätzliche Qualität ins Spiel, nämlich die der *Spaltung* in zwei und der *Verschmelzung* zweier Teiltöne in einen. Dadurch entstehen potentielle Konfliktsituationen bei Vorgänger/Nachfolger-Zuordungen, auf die sich beispielsweise das Phasenübergabeprinzip der Teiltonsynthese stützt.

Der Fall eines ungleichen Zweitonkomplexes ändert bis zu einer Pegeldifferenz von etwa 6 dB nichts an dem prinzipiellen Bild von Aufspaltung und Verschmelzung in Bereich 2. Bei größerem Pegelunterschied entsteht ein geschlossener, frequenz- und amplitudenmodulierter Teiltonverlauf in der Nähe der stärkeren Schwingung, während sich Kurzverläufe in der Nähe der schwächeren ausbilden.

2.3.2.2 Synthese und Komplexität von Verbesserungen

Um die wahrnehmbare Veränderung der TTZM-Verarbeitung zu beurteilen, dienen wieder die Hüllkurven $e(t)$ von Original und Synthesesignal, dargestellt in Bild 2.6 in der Mitte. Im Vergleich mit dem Teiltonzeitmuster kann man in der Synthesehüllkurve die Beiträge zuordnen, die von der Synthesesinusschwingung des Einzeltons bzw. denen des Teiltondoppels stammen. Beim Wechsel entstehen besonders im Bereich 2 heftige Hüllkurvensprünge, die die Synthesefenster-kontrollierbaren Störungen verkörpern. Während der Beitrag für den Einzelton in Bereich 1 gut die Originalhüllkurve nachzeichnet, um in Bereich 2 hinein langsam abzusinken, nimmt umgekehrt der Beitrag für das Doppel zu. In der Mitte von Bereich 2 entsteht dadurch eine Hüllkurve, die doppelt so schnell wie die des Originals schwankt, gegen Ende von Bereich 2 wird sie teilweise unregelmäßig. Darin manifestieren sich die Phaseninkohärenz-bedingten Störungen. In Bereich 3 stimmt die Hüllkurve plötzlich wieder mit dem Original überein.

Bei den Hüllkurvenstücken, die in der linken Hälfte des Bildes von Teiltondoppeln hervorgerufen werden, kann die Phasenrelation der beiden Synthesesinusschwingungen zunächst nicht stimmen: Beide übernehmen die Phase des endenden Einzelteiltons als Startphase (Abschnitt 1.5.4). Weil die geringe Frequenzdifferenz die Phasen nicht so schnell auseinanderlaufen läßt, überlagern sich beide Syntheseschwingungen maximal. Eigentlich sollten sie sich durch Gegenphasigkeit eher dämpfen und genau dann gegenseitig aufheben, wenn die Originalhüllkurve null erreicht. Dieser Situation nähert man sich aber erst am Ende von Bereich 2, wenn die Frequenzdifferenz größer ist.

Eine verbesserte Synthese müßte hier die Startphasen immer so anpassen, daß genau in der zeitlichen Mitte eines Doppels Gegenphasigkeit erreicht ist. Will man die Übergänge von und zur Einzelschwingung nicht durch ungünstig lange Synthesefenster glätten, dann müßten die Phasenlagen an den Zeitpunkten von Spaltungen und Verschmelzungen zusätzliche Nebenbedingungen erfüllen. Daraus folgt, daß sich die Frequenzverläufe der Synthesesinusschwingungen nicht mehr exakt nach den Vorgaben der Teiltonparameter richten, also ein gewisses ‘Eigenleben’ führen.

2.3.3 Frequenzmodulation

Dieser Modulationstyp liefert schließlich ein drastisches Beispiel für Synthesestörungen im Übergangsbereich und zeigt, daß sie bei komplizierten Teiltonkonstellationen Rauschcha-

rakter haben können. Als Testsignaltyp $s(t)$ dient ein Sinusträger mit der Ruhefrequenz ω_T, der mit einem maximalen Hub $\pm\Delta\omega$ sinusförmig mit der Frequenz ω_M hin- und hergeschoben wird. Die zeitvariante Interpretation gründet sich demnach auf der Momentanfrequenz

$$\omega(t) = \omega_T + \Delta\omega \sin(\omega_M t), \quad \text{womit} \tag{2.23}$$

$$s(t) = A \cos\left[\int_0^t \omega(\tau) d\tau + \phi_0\right] \tag{2.24}$$

$$= A \cos\left[\omega_T t - \frac{\Delta\omega}{\omega_M} \cos(\omega_M t)\right] \tag{2.25}$$

angesetzt werden kann, bei willkürlicher Vorgabe von Amplitude A und Startphase $\phi_0 = 0$. Mit Hilfe der Bessel-Funktionen $J_n(x)$ läßt sich Gl. (2.25) in einen stationären Tonkomplex mit Teilschwingungen im Abstand ω_M umwandeln (z.B. [Ste82]). Es ergibt sich als spektrale Interpretation

$$s(t) = A \sum_{n=-\infty}^{+\infty} J_n\left(\frac{\Delta\omega}{\omega_M}\right) \cos\left(\omega_T t + n\omega_M t - n\frac{\pi}{2}\right). \tag{2.26}$$

Um wieder den Übergang zwischen beiden Interpretationen zu beobachten, erhält das spezielle Testsignal die Trägerfrequenz $\frac{\omega_T}{2\pi} = 1\,\text{kHz}$ und den Hub $\frac{\Delta\omega}{2\pi} = 100\,\text{Hz}$. Die Modulationsfrequenz wird innerhalb der Signaldauer von 1 s zeitlinear von 0 auf $\frac{\omega_M}{2\pi} = 50$ Hz erhöht.

Das in drei Bereiche eingeteilte Teiltonzeitmuster ist oben in Bild 2.7 zu sehen. Bereich I zeigt noch einen in der Frequenz variierenden Teilton. Allerdings tauchen schon bei niedrigen Modulationsfrequenzen in der Umgebung der Wendepunkte kurze Nebenteiltöne auf. Beim Übergang im Bereich II, dem Übergang zwischen beiden Interpretationen, werden Unterbrechungen im Verlauf des Einzeltons sichtbar, der bald nicht mehr als Hauptverlauf auszumachen ist. Dafür entsteht ein charakteristisches Muster von Teiltönen, das sich über die Breite des Hubs erstreckt. Bereich III enthält schließlich durchgehende Teiltonverläufe, die die spektrale Interpretation wiedergeben. Bei der gewählten Dimensionierung des Testsignals stellen die Bessel-Koeffizienten sehr unterschiedliche Amplituden für die einzelnen Teilschwingungen ein. Schwächere können die gegenseitige Maskierung im Teiltonzeitmuster nur zeitweilig oder überhaupt nicht überwinden (vgl. Abschnitt 2.4).

Bei Teiltonsynthese nehmen zwar die vom Teiltonzeitmuster vorgegebenen Synthesesinusschwingungen eine größere spektrale Breite als die Frequenzgruppe ein. Dennoch lassen sich die wahrnehmbaren Verfälschungen bis in Bereich II noch gut anhand der Hüllkurvenverläufe von Original und Synthese verdeutlichen, welche in Bild 2.7 unten eingetragen sind. Die konstante Hüllkurve des Originals geht völlig verloren. Bereits die kurzen Nebenteiltöne in Bereich I verursachen wahrnehmbare Störungen, weil das noch als zeitvarianter Sinuston wahrnehmbare Signal plötzlich seine Hüllkurve ändert. Der Höreindruck in Bereich II entspricht dann nicht mehr einer Frequenzmodulation, sondern einem Schmalbandrauschen. Dazu tragen beide Störungstypen gemeinsam bei, weil einerseits relativ viele Synthesesinusschwingungen inkohärent überlagert werden und andererseits zahlreiche kleine Hüllkurvenübergänge durch Kurzverläufe zu kontrollieren sind. Bereich III hört sich ziemlich unverändert an, da die Kurzverläufe gegenüber den durchgehenden Teiltönen schwächer sind. Außerdem fallen beide Doppelstränge bald in getrennte Frequenzgruppen, so daß ihre Phasenlagen keine große Rolle mehr spielen.

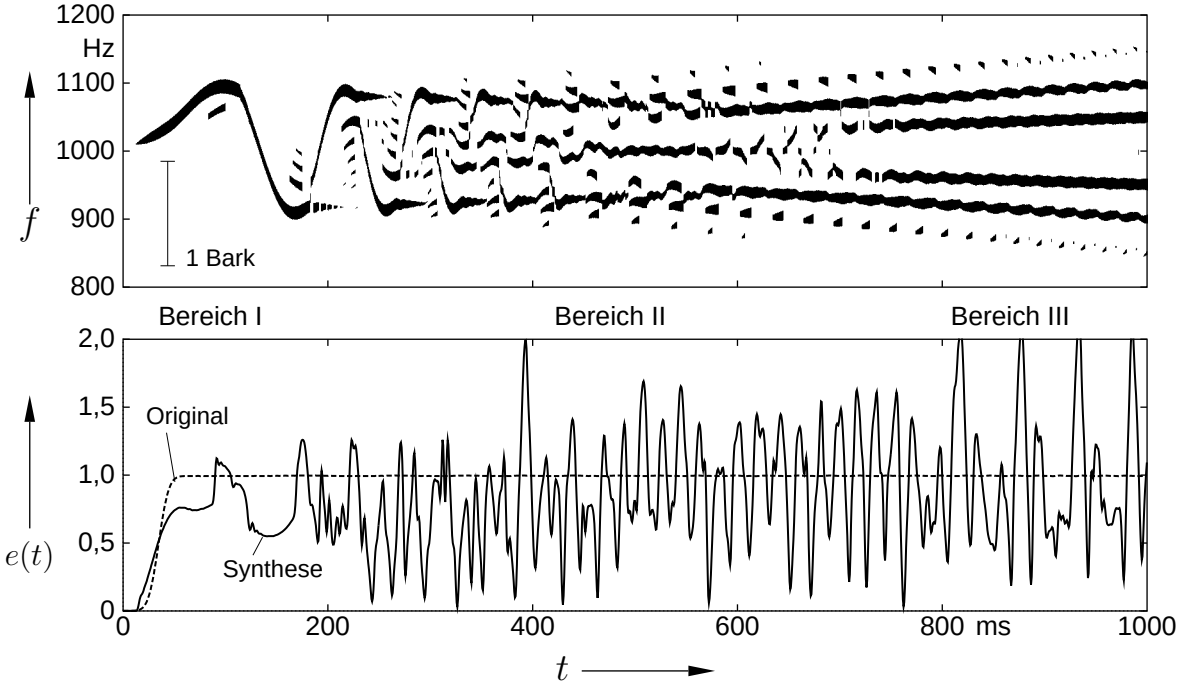

Bild 2.7: *Frequenzmodulation eines Sinusträgers $f_T = 1\,\mathrm{kHz}$ mit einem Gleitsinuston $f_M = 0 \ldots 50\,\mathrm{Hz}$, zeitlinear ansteigend, und einem maximalen Hub von $\Delta f = \pm 100\,\mathrm{Hz}$: Teiltonzeitmuster (oben), Signalhüllkurven von Original und Synthese (unten). Die komplizierte Teiltonkonstellation in Bereich II überfordert die Synthese derart, daß die Wahrnehmung dort in ein Schmalbandrauschen verfälscht wird.*

2.3.4 Zusammenfassung und Übertragung auf Sprache

Bei elementaren Modulationsformen mit Sinusschwingungen gibt es im Teiltonzeitmuster einen besonderen Übergang zwischen einer rein zeitlichen und einer rein spektralen Ausprägung der Modulation. Dieser manifestiert sich durch Teilton-Kurzverläufe, die auf hohem Pegelniveau plötzlich auftauchen und verschwinden, sowie durch Spaltungen und Verschmelzungen von einem in zwei Verläufe und umgekehrt. Solche Konstellationen überfordern die bisherige Teiltonsynthese aus zwei Gründen: Erstens rufen sie Sprünge im Synthesesignal hervor, die zwar durch das Synthesefenster geglättet werden. Da aber die spektralen Eigenschaften beim Dreieckfenster nicht ausreichen, existiert eine wahrnehmbare 'Synthesefenster-kontrollierbare Störung'. Zweitens kann hier die Teiltonsynthese die Phasen der einzelnen Synthesesinusschwingungen nicht abstimmen, so daß innerhalb der Breite einer Frequenzgruppe eine völlig veränderte Signalhüllkurve entsteht. Diesen Effekt kann man als schmalbandige 'Phaseninkohärenz-bedingte Störungen' bezeichnen. Zur Abstimmung der Synthesephasen muß eine verbesserte Teiltonsynthese offenbar zusätzliche Nebenbedingungen einhalten. Dies ginge nur, wenn man dem Frequenzverlauf einer Synthesesinusschwingung ein gewisses 'Eigenleben' gegenüber der Vorgabe des Teiltonverlaufes einräumt.

Daß die beiden Störungstypen auch bei Sprachverabeitung eine Rolle spielen, läßt sich plausibel machen, indem man sich stimmhafte Spracherzeugung durch überlagerte ampli-

tudenmodulierten Harmonische mit gemeinsamer Frequenzmodulation vorstellt [Alm83]. Die Natur der Modulation ist nicht mehr sinusförmig, sondern teilweise kompliziert zeitvariant. Harmonische können paarweise miteinander schweben. Es ist also sehr wahrscheinlich, daß dadurch die beobachteten Konstellationen im Übergang spektraler und zeitlicher Auflösung vorkommen. Tatsächlich zeigt das Teiltonzeitmuster in Bild 1.2 auf S. 13 viele Kurzverläufe, Aufspaltungen und manchmal auch Verschmelzungen, die Synthesefensterkontrollierbare Störungen auslösen können. Darüberhinaus gibt es offenbar viele Teiltöne innerhalb einer Frequenzgruppe, so daß Phaseninkohärenz-bedingte Störungen vorliegen müssen.

Die Entstehungsweise der Störungen in einem Frequenzband bedingt, daß der Störpegel an den darin vorkommenden Signalpegel gekoppelt ist. Dadurch kann im Extremfall eine differenzierte, tonale Signalstuktur in ein Schmalbandrauschen gleicher Intensität verwandelt werden. Weil die speziellen Teiltonkonstellationen meist in vielen Frequenzlagen sichtbar sind, ergeben beide Störungstypen jeweils einen spektral/zeitlichen *Störteppich*, der sich in seiner Form etwa am zeitvarianten FTT-Spektrum des Sprachsignals orientiert. Beide Störteppiche sind also grob an die Hörschwellen des Nutzsignals angeformt, tendenziell allerdings auf überschwelligem Niveau. Deshalb sind sie als Verfälschungen wahrnehmbar, ohne daß sie die wesentliche sprachliche Information verdecken könnten.

2.4 Überhöhte Simultanverdeckung

Wenn Teilschwingungen in einem stationären Tonkomplex einen deutlich größeren Abstand als eine Analysebandbreite aufweisen, können sie im Teiltonzeitmuster als durchgängige Teiltonverläufe sichtbar werden. Dazu dürfen die Pegelunterschiede nicht zu groß sein, sonst verschwindet der Verlauf der schwächeren Teilschwingungen. So wie für die Simultanverdeckung im Gehör [Zwi82] kann man auch im Teiltonzeitmuster 'Mithörschwellen' für diesen Maskierungseffekt bestimmen. Heinbach hat hierzu bereits Messungen durchgeführt, allerdings ohne sie quantitativ mit dem Gehör zu vergleichen [Hei88a, S. 34]. Tatsächlich ist es für eine verfälschungsfreie TTZM-Verarbeitung aber wichtig, daß schwächere Teilschwingungen erst dann unrepräsentiert sind, wenn sie auch im Gehör durch stärkere verdeckt werden. Im folgenden wird dies anhand der Konstellation Sinusmaskierer/Sinustestton überprüft. Dabei erhält man gleichzeitig Aufschluß über den Spielraum für eine größere Analysebandbreite, welche die vorigen beiden Abschnitte nahelegten.

Zur Bestimmung der Simultanverdeckung werden Signale von 2 s Dauer verwendet, bestehend aus einem Maskiererton der Frequenz $f_M = 1$ kHz und einem Testton mit verschiedenen Frequenzen f_T. Bei festem Maskiererpegel L_M wird der Pegelabstand $L_T - L_M$ des Testtons innerhalb der Signaldauer zeitlinear von null auf -80 dB verändert. Im Teiltonzeitmuster zeigt sich der Testton als zunehmend schwächerer Teilton, der in eine Folge von Kurzverläufen übergeht und schließlich verschwindet. Aus Messung der Zeiten ab Signalanfang, an denen der durchgängige Verlauf erstmals unterbrochen bzw. der letzte Kurzverlauf verschwunden ist, wird der zugehörige Pegelabstand ermittelt. Diese Werte sind in Bild 2.8 unter der Bezeichnung HB-TTZM eingezeichnet und als Schwellenverläufe untereinander verbunden.

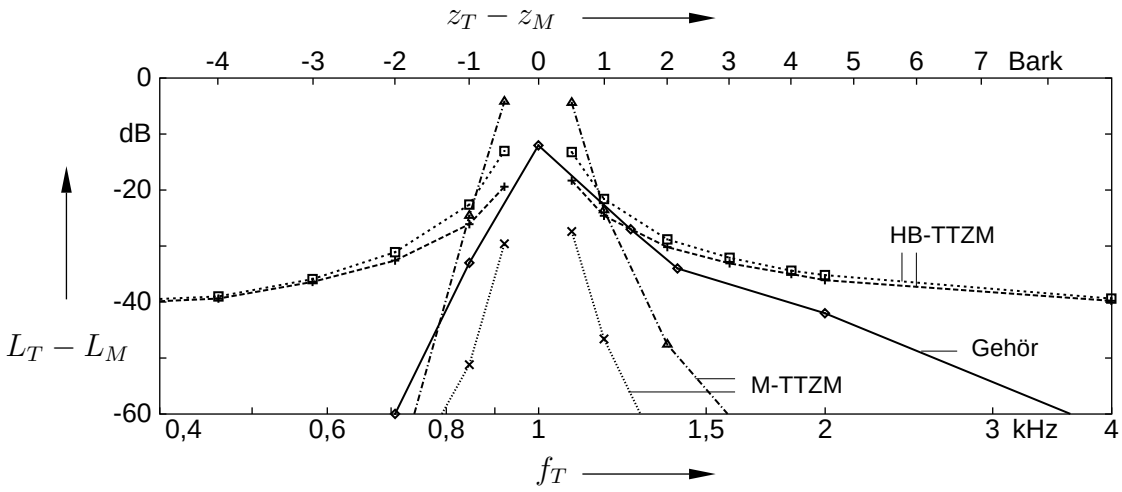

Bild 2.8: *Maskierung eines Testtons durch einen Maskiererton $f_M = 1\,\mathrm{kHz}$ in Abhängigkeit von der Testtonfrequenz f_T oder der Tonheitsdifferenz $z_T - z_M$. Für das TTZM-Verfahren (HB-TTZM) und eine später beschriebene, modifizierte Version (M-TTZM) sind jeweils zwei Schwellen eingetragen: Pegeldifferenzen $L_T - L_M$ unterhalb beider machen den Testton im Teiltonzeitmuster unsichtbar, oberhalb beider ist er dauernd sichtbar, dazwischen tritt er mit zeitlichen Unterbrechungen auf. Der Vergleich mit einer Mithörschwelle des Gehörs ($L_M = 70\,\mathrm{dB}$, Daten aus [Zwi82, Abb. 3.7]) zeigt, daß eine Verarbeitung mit HB-TTZM hörbare Töne entfernen kann.*

Die untere Schwelle, die der vollständigen Verdeckung, entsteht dadurch, daß das Ausgangssignal der FTT-Analysefilter an allen Analysefrequenzen vom Beitrag des Maskierertons dominiert wird. Dadurch entspricht das FTT-Spektrum im wesentlichen dem Spektrum des Maskierers. Selbst im Bereich der Testtonfrequenz, in dem der Beitrag vom Testton am größten ist, reicht die Beeinflussung der Flanken des Maskiererspektrums nicht mehr aus, um für die Teiltonextraktion ein genügend ausgeprägtes Spektralmaximum zu erzeugen.

Die Verhältnisse am Analysefilterausgang lassen sich nach Abschnitt 1.4.2 durch Überlagerungen von zwei komplexen Drehzeigern darstellen: Kurz vor der vollständigen Maskierung schwankt das FTT-Spektrum in der Umgebung der Testtonfrequenz besonders stark, weil dort beide Zeiger vergleichbare Längen aufweisen. Dadurch kann das Ausgeprägtheitskriterium periodisch erreicht und wieder verfehlt werden, woraus sich die beobachteten Kurzverläufe und die obere Schwelle erklären. Die Schwankung gibt die Differenzfrequenz beider Töne wieder. Da der zeitliche Verlauf des FTT-Leistungsspektrums vor der Teiltonextraktion geglättet wird, verringert sich die zeitliche Schwankung zu höheren Differenzfrequenzen immer mehr, so daß der Abstand der beiden Schwellen sichtbar schrumpft. Bei größeren Abständen vom Maskierer spielt auch noch die Wirkung des Drehzeigers von der negativen Frequenzhalbachse eine Rolle.

Zusätzlich ist in Bild 2.8 die beim Gehör gemessene Mithörschwelle für einen Maskiererschalldruckpegel von $L_M = 70\,\mathrm{dB}$ eingetragen [Zwi82, Abb. 3.7]. Sekundäre Empfindlichkeitssteigerungen durch hörbare Differenztöne sind hier vernachlässigt. Liegt die Kurve oberhalb beider Schwellen von HB-TTZM, so werden im Teiltonzeitmuster noch durchgängige Teiltöne erkannt, die das Gehör gar nicht mehr wahrnimmt. Liegt sie unterhalb, dann werden hörbare Töne im Teiltonzeitmuster verdeckt. Bei Verlauf zwischen

beiden Schwellen kann das Auf- und Abtauchen der Kurzverläufe im Synthesesignal gehört werden, die Präsenz des Testtons wird also noch wiedergegeben.

Die Schwellen von Gehör und HB-TTZM zeigen zwar im Nahbereich außerhalb der Frequenzgruppe des Maskierers eine annähernde Übereinstimmung. Bei Werten $z_T - z_M$ unterhalb etwa -1 Bark und oberhalb von 1,5 Bark flachen die Schwellen von HB-TTZM gegenüber der des Gehör aber deutlich ab. Sie erreichen keine tieferen Werte als 40 bis 45 dB unter Maskiererpegel. Die Simultanverdeckung ist hier also überhöht. Folglich kann eine Verarbeitung durch das TTZM-Verfahren hörbare Töne eines Signals entfernen. Weil weiterhin die spektrale Selektivität der FTT-Analysefilter maßgeblich die Verdeckung beeinflußt, darf die Analysebandbreite keinesfalls vergrößert werden, wie es die vorigen Abschnitte wünschenswert erscheinen ließen. Daß dies bei einer anderen FTT-Fensterfunktion dagegen möglich ist, zeigt eine modifizierte TTZM-Analyse, deren Schwellenverläufe M-TTZM ebenfalls in Bild 2.8 eingetragen sind. Sie wird im nächsten Kapitel behandelt. [3]

2.5 Tonalisierung von Rauschsignalen

Ein charakteristischer Effekt der TTZM-Verarbeitung kann bei Rauschsignalen beobachtet werden. Mit seiner Untersuchung lassen sich bestimmte Verfälschungen bei Sprache erklären, bei der Rauschanteile in Form von Frikativen und Strömungsgeräuschen vorkommen. Stellvertretend für beliebig spektral geformte Rauschsignale dient als Testsignal Weißes Rauschen. Es dauert 2 s, weist eine obere Grenzfrequenz von 5,5 kHz auf und stammt von einer thermischen Rauschquelle. Nach Verarbeitung erhält man eine zeitvariierend nasale und 'hohle' Färbung. Man glaubt hier eine Art Kammfiltereffekt[4] zu hören, der regellos schwankt.

Die Verfälschung läßt sich nicht mit einfachen statistischen Ansätzen messen. Mittelung von Leistungsspektren über die gesamte Signaldauer oder Auswertung von Amplitudenstatistiken, auch in beliebigen Teilbändern, zeigen überhaupt keine signifikante Veränderung. Der Effekt muß daher in der vom Gehör analysierbaren, zeitvarianten Feinstruktur des Kurzzeitspektrums begründet liegen. Wegen der Gehöranpassung bietet es sich an, das Teiltonzeitmuster selbst als Meßgrundlage zu nutzen. Indem das Synthesesignal nochmals einer Teiltonanalyse unterworfen wird, können Veränderungen zum Original-TTZM untersucht werden.

In Bild 2.9 links ist ein 500 ms langer Ausschnitt des Original-TTZM zu sehen. Weißes Rauschen zeigt sich als Ansammlung von unregelmäßig langen, in der Frequenz schwankenden Teiltonlinien. Zu hohen Frequenzen hin sinkt die durchschnittliche Linienlänge, während die Schwankungen im Linienverlauf schneller werden. Trotzdem bildet sich über der Tonheitskala ein bestimmter Teiltonabstand bevorzugt aus. Rechts daneben ist das

[3] Die TTZM-Simultanverdeckung wird auch noch von der Ausgeprägtheitschwelle ΔL_A beeinflußt (Abschnitt 3.4.3.5). Diese läßt sich aber beim Heinbachschen TTZM-Verfahren nur im Tausch gegen eine hörbar stärkere zeitliche Glättung verringern [Hei88a, S. 22].

[4] Gemeint ist der Effekt, der bei Superposition eines breitbandigen Signals mit seiner zeitverzögerten Kopie auftritt. Man kann ihn durch Filter erzeugen, dessen Übertragungscharakteristik in regelmäßigen Frequenzabständen Einbrüche ('Kammzinken') aufweist. In der Musikproduktion ist er als 'Phasing/Flanging-Effekt' bekannt.

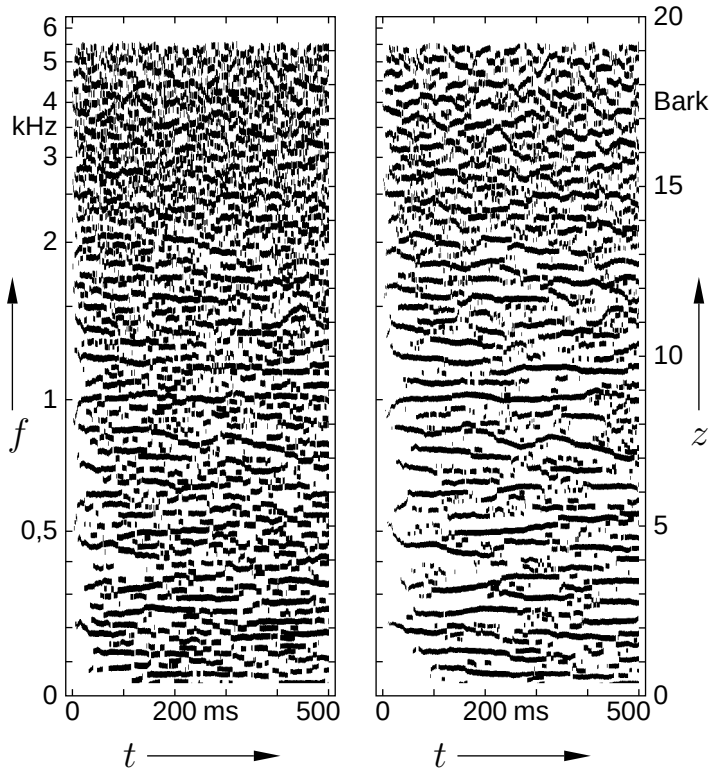

Bild 2.9: *Veränderung des Teiltonzeitmusters von Weißem Rauschen nach Teiltonsynthese in Verbindung mit erneuter Teiltonanalyse: Links vorher, rechts nachher. Kurze Teiltonverläufe verschwinden zwischen längeren, die sich zu noch längeren, aber schwächer fluktuierenden Linien zusammenschließen. Dadurch wird 'Regellosigkeit' gegen 'tonale Ordnung' getauscht. Der verspätete Einsatz bei tiefen Frequenzen ist eine Folge des Einschaltens bei $t = 0$ in Verbindung mit der frequenzabhängigen FTT-Analysebandbreite.*

Teiltonzeitmuster nach einmaligem Verfahrensdurchlauf abgebildet, also nach Synthese des Original-TTZM und erneuter Teiltonanalyse. Darin nimmt die Tendenz zur Ausbildung geschlossener, längerandauernder und in der Frequenz weniger fluktuierender Teiltonverläufe zu. Außerdem sinkt die Teiltondichte – oder umgekehrt: der mittlere Teiltonabstand steigt. Ganz offensichtlich verringert sich die Regellosigkeit der kurzzeitspektralen Feinstruktur nach einer TTZM-Verarbeitung.

Einen weiteren Einblick in die Natur der Veränderung kann das FTT-Spektrum selbst geben. Allerdings sind die Unterschiede in Darstellungen des normalen FTT-Pegelspektrums nicht so augenfällig. Das ändert sich, wenn man unabhängig vom TTZM-Verfahren eine modifizierte FTT einführt. Unter Beibehaltung aller übrigen Parameter der bisherigen FTT verwendet sie eine Gauß-Approximation achten Grades (PG8) als Fensterfunktion. Dadurch erhält man eine höhere Frequenzselektivität bei vergleichbarer zeitlicher Selektivität, wie in Abschnitt 3.3.3 begründet wird. Schnitte der so berechneten FTT-Pegelspektren von Original und einfach verarbeitetem Signal sind, für den Zeitpunkt $t = 250$ ms als Beispiel, in Bild 2.10 abgebildet.

Das Spektrum des verarbeiteten Signals zeigt im Gegensatz zum Original ausgeprägtere Spitzen mit deutlicheren Lücken dazwischen. Diese Konstellation weist auf Sinustöne hin, die in ihren Parametern Amplitude und Frequenz innerhalb des Analysefensters nicht allzu schnell fluktuieren und die deshalb spektral 'scharf' erfaßt werden können. Das TTZM-Verfahren bewirkt demnach eine *Tonalisierung* von Rauschsignalen, worin sich gleichzeitig eine Abnahme der Regellosigkeit ausdrückt. Allerdings werden noch keine ausgeprägten, einzelnen Töne gehört, weil sie relativ nahe beieinander liegen und weil ihre Fluktuation in Amplitude und Frequenz verschleiernd wirkt.

Tonalisierung ist keine Konsequenz der spezielle Analyseparameter des TTZM-Verfahrens.

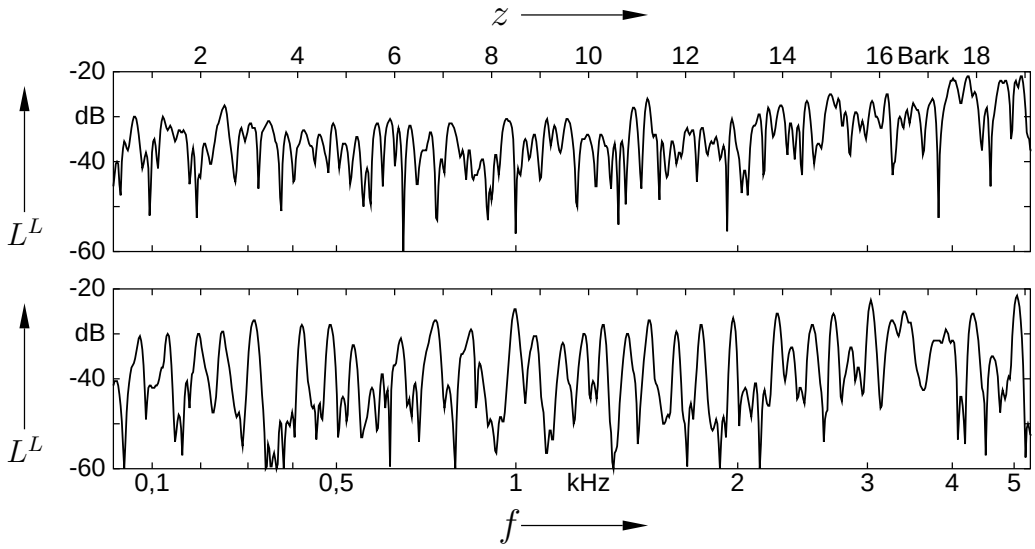

Bild 2.10: *FTT-Pegelspektren von Weißem Rauschen zum Zeitpunkt* $t = 250$ ms*: Original (oben) und nach einmaliger Verarbeitung durch das TTZM-Verfahren (unten). An der Ausbildung tiefer Lücken zwischen den Spitzen kann man die hörbare Tonalisierung ablesen. Zur besseren Darstellung wurde die FTT mit einem Gauß-Fenster* PG8*,* $B_{3dB} = 0{,}1$ Bark*, mit Laufzeitausgleich und ohne Glättung benutzt (Kapitel 3).*

Sie ist bei allen vernünftigen Analysebandbreiten und Fensterfunktionen wahrnehmbar, bei fehlender Glättung des FTT-Spektrums wie bei verringerter Ausgeprägtheitsschwelle in der Teiltonextraktion. Eine Vergrößerung der Analysebandbreite bewirkt beispielsweise eine schnellere Schwankung des Kammfiltereffekts. Weil der mittlere Teiltonabstand direkt von der Analysebandbreite abhängt, werden öfter einzelne Töne wahrgenommen, womit das verarbeitete Signal eher an einen 'plätschernden Wasserfall' erinnert. Eine Verringerung ruft dagegen einen zunehmend statischen, nasalen Klangeindruck hervor. Die Frequenzgruppenanpassung spielt keine Rolle, da der Effekt auch bei einem vereinfachten TTZM-Verfahren mit Konstantbandbreiten zu beobachten ist.

Die Ursache der Tonalisierung liegt auch nicht in der Synthese begründet. Weitere Untersuchungen zeigen, daß der Effekt bei jeder Art von Synthesefenster existiert, beim Rechteckfenster der ursprünglichen Synthese von Heinbach wie später bei gehörangepaßten Fenstern. Auch die Phaseninkohärenz der Synthesesinusschwingungen macht kaum etwas aus, was eine noch einzuführende Rekonstruktion mit Originalphasen – also den 'richtigen' Phasen – demonstrieren kann (RKOP, Kapitel 5).

Tonalisierung rührt vielmehr daher, daß das TTZM-Verfahren transiente Anteile vernachlässigt, die in der zeitlichen Entwicklung von FTT-Spektren mit günstigeren Fensterfunktionen zu finden sind. Solche Anteile würden die Lücken in Bild 2.10 unten auffüllen, damit wieder mehr Regellosigkeit bewirken und wahrnehmbare Tonalisierung vermeiden. Dies läßt sich belegen, wenn später eine Repräsentation transienter Anteile zusammen mit einer simulierten, optimalen Signalrekonstruktion eingeführt sind (ZFKI/RKOP, Abschnitt 3.4). Tonalisierung von Rauschsignalen stellt also eine spezielle Konsequenz von unterrepräsentierten transienten Signalanteilen gemäß Abschnitt 2.1 dar. Sie führt aber in Kapitel 4 auf einen alternativen Lösungsansatz in Form einer separaten Repräsentation für Rauschanteile.

2.6 Qualitätsbeeinträchtigungen bei Datenreduktion

Die Verarbeitungsqualität des TTZM-Verfahrens wird durch Maßnahmen, die Heinbach zur Datenreduktion einführte, zusätzlich beeinträchtigt. Die von ihm durchgeführten Hörversuche ergaben trotz hoher Verständlichkeit eine deutliche Abnahme der Sprachgüte [Hei88a, S. 74]. Am Beispiel der Verfahrensvariante mit 4,4 kbit/s Datenrate werden zunächst die Art der Beeinträchtigungen geschildert und anschließend im einzelnen ihre Ursachen untersucht.

Gemäß Abschnitt 1.5.3 begrenzen vier Maßnahmen im Teiltonzeitmuster den Datenfluß auf 4,4 kbit/s: Verlängern des Auswerteintervalls auf 20 ms, Beschränkung auf höchstens zehn pegelstärkere Teiltöne, Codierung der Teiltonpegel mittels grobquantisierter Interpolationsgerade und schließlich gröbere Quantisierung der Teiltonfrequenzen. Aus noch zu erläuternden Gründen wird bei Synthese das Rechteckfenster und nicht das mittlerweile bewährte Dreieckfenster verwendet (Abschnitt 1.5.4). Im Ergebnis hört man im Vergleich zum nichtreduzierten Teiltonzeitmuster eine leicht verfremdete Sprache und zwei charakteristische Störeffekte:

Periodische Knackstörung: Im Takt des Auswerteintervalls treten Knacke auf, etwa in Form eines ‘langsamen Knatterns’, das sich in Lautheit und Klangfärbung dem Sprachverlauf anpaßt.

Tonale Artefakte: Im selben Takt sind zahlreiche kurze Töne als eine Art ‘Klingeln’ zu vernehmen. Sie stammen aus höheren Tonlagen, denn anschließende, steile Tiefpaßfilterung ab 2 kHz kann sie unterdrücken.

Beide Störeffekte wirken, wie aus einer separaten Quelle stammend, der Sprache lediglich überlagert. Sie irritieren den unvorbereiteten Zuhörer nicht unerheblich. Doch tritt schnell Gewöhnung ein, wenn man sich auf die Sprache selbst konzentriert, die in ihren lautsprachlichen, prosodischen und sprecherspezifischen Merkmalen bewahrt scheint. Sie leidet allerdings durch zwei zusätzliche Verfremdungseffekte an Natürlichkeit:

Spektrale und zeitliche Überspitzung: In der spektralen Form äußert sie sich wie Tonalisierungseffekte an steilen Filterflanken oder wie schmalbandige Resonanzen, die sich jeweils zeitlich verändern. Sie scheinen an die Formanten des Sprachsignals gekoppelt und lassen die Sprache ‘überartikuliert’ wirken. Die zeitliche Form fällt weniger auf und besteht darin, daß die Laute an ihren Enden ‘abgehackt’ wirken.

Intensitätsmodulation: Stimmhafte Sprachanteile sind im Takt mit der periodischen Knackstörung beeinflußt. Je nach Sprachsignal hört man ein ‘Zittern’ oder ‘Schwanken’ der Stimme, die gleichzeitig ‘rauher’ wirkt. Der Effekt entsteht im Frequenzbereich unterhalb von 1 kHz, wie sich aus dem Vergleich steiler Hochpaß- und Tiefpaßfilterung ergibt.

Die Wahrnehmungsintensität der Stör- und Verfremdungseffekte hängt vom speziellen Sprachsignal und von der Abhörsituation ab. Frauenstimmen erscheinen im allgemeinen weniger verändert. Eine geringe Lautstärke oder ein zugesetztes breitbandiges Rauschsignal verringert die Wahrnehmbarkeit. Bei den nicht-tonalen Effekten wirkt besonders

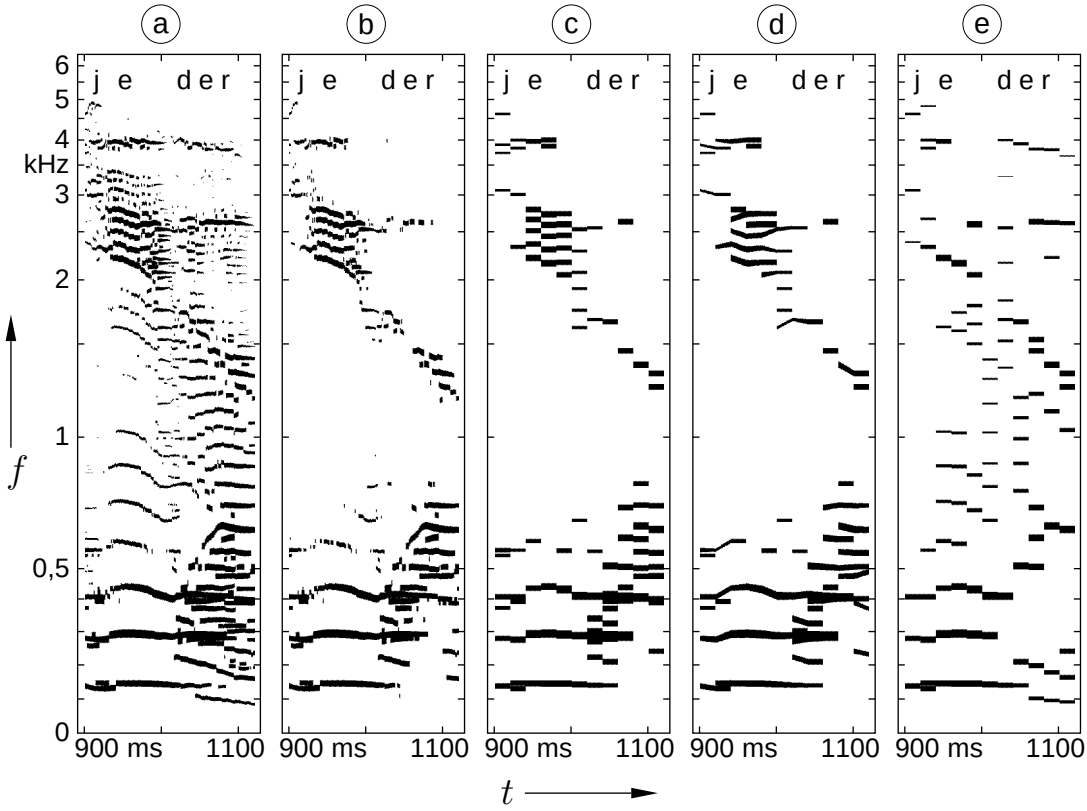

Bild 2.11: *Teiltonzeitmuster nach verschiedenen Maßnahmen zur Datenreduktion für einen Ausschnitt aus einem Sprachsignal: a) unbearbeitet, b) davon maximal zehn der pegelstärksten Teiltöne pro Auswerteintervall, c) zusätzlich vergrößertes Auswerteintervall und d) noch dazu die Wirkung von nachträglicher Frequenz- und Pegelinterpolation. e) wie c), aber Auswahl der zehn Teiltöne nach Spektraltonhöhengewicht statt nach Pegel.*

Raumwiedergabe und möglichst indirekte Beschallung bei einer nicht zu kleinen Nachhallzeit deutlich verbessernd. Kopfhörerdarbietung (Telefonieren!) erweist sich hier als am kritischsten.

Um die Ursachen der Beinträchtigungen zu klären, wurden verschiedene Kombinationen von Maßnahmen auf das Teiltonzeitmuster des männlichen Sprechers in Bild 1.2 auf S. 13 oben angewandt. Dabei kamen neben den ursprünglichen vier auch andere Maßnahmen zur Anwendung, die noch angesprochen werden. Einige Ausschnitte von derart bearbeiteten Teiltonzeitmustern sind zur weiteren Diskussion in Bild 2.11a-e dargestellt. Die Syntheseergebnisse wurden dann im Selbstversuch über Kopfhörer paarweise verglichen. Die beobachteten Veränderungen bilden den experimentellen Rahmen der folgenden Unterabschnitte. Anhang D.1 stellt die zugehörigen Einzelergebnisse zusammen, auf die hier nur lose Bezug genommen wird.

2.6.1 Spektral/zeitliche Kontrastverschärfung

Der Verfremdungseffekt der spektralen und zeitlichen Überspitzung tritt in dem Moment auf, in dem die Teiltonanzahl auf maximal zehn pegelstärkere reduziert wird. Dies ist durch den Übergang vom unreduzierten TTZM-Ausschnitt in Bild 2.11a auf 2.11b verdeutlicht.

Daran läßt sich zunächst die spektrale Überspitzung erklären. Man erkennt in der Mitte von Bild 2.11b eine größere Fläche ohne Teiltöne, ebenso im letzten zeitlichen Drittel oberhalb und unterhalb einer Teiltonspur bei ca. 2500 Hz. Die zeitlich variierenden Flächenausdehnungen in Frequenzrichtung kann man sich auch als ideale Sperrbereiche von zeitvarianten Bandsperren vorstellen. Dort, wo die Flächen zu höheren oder tieferen Frequenzen hin an Teiltöne stoßen, entstehen sehr steile Übergänge hin zu Durchlaßbereichen. Diese gruppieren sich tendenziell um Formanten. Auf diese Weise hebt sich im Beispiel besonders die Teiltonspur im Übergang des zweiten Formanten von 'e' nach 'a' in 'je(d)a' heraus. Beim unreduzierten Ausschnitt in Bild 2.11a existieren im Bereich der angesprochenen Flächen dagegen noch Teiltöne. Sie verhindern, daß einerseits keine idealen Sperrbereiche und andererseits nicht zu steile Übergänge entstehen, sondern eben solche, die natürlichen Formanthüllkurven entsprechen.

Die zeitlichen Begrenzungen der Leerflächen markieren abrupte Intensitätswechsel für ausgedehnte Frequenzbereiche. Sie begründen die zeitliche Überspitzung. Ohne Teiltonbeschränkung nach Pegelkriterium wären sie nicht vorhanden. An- oder ausklingende oder allgemein schwächere Teiltonverläufe im Bereich der angesprochenen Flächen können dann nicht im Wettbewerb um die stärksten zehn unterliegen.

Als Kontrollversuch bietet sich eine alternative Strategie der Teiltonbeschränkung an, bei der die zehn Teiltöne nicht nach Pegelstärke, sondern nach Spektraltonhöhengewicht ausgewählt werden. Hierzu wurde das Berechnungsverfahren nach Terhardt et al. verwendet [Ter82]. In der verwendeten Implementierung liefert es in jedem Auswerteintervall diejenigen zehn Teiltöne, die das größte Spektraltonhöhengewicht aufweisen. Der angenommene Sprachpegel entspricht der Lautheit eines 1kHz-Tons mit 70 dB Schalldruck. Heinbach verwendete diese Alternative schon bei seinen Messungen an datenreduzierten Einzelvokalen und erzielte bei Beschränkung auf fünf Teiltöne eine Anhebung der Verständlichkeit gegenüber dem Pegelstärkekriterium [Hei88a, S. 46ff].

In der Tat findet bei Teiltonbeschränkung nach Tonhöhengewicht keine Überspitzung statt. Das veranschaulicht die Gegenüberstellung von Bild 2.11c und 2.11e. Daß mittlerweile das Auswerteintervall vergrößert wurde, ist dabei nebensächlich. Das Pegelkriterium in Bild 2.11c bevorzugt Teiltonansammlungen in Formantnähe, wodurch die großen Leerflächen entstehen. Umgekehrt zeigt das Tonhöhengewichtskriterium die Tendenz, die maximal zehn Teiltöne über den gesamten Frequenzbereich zu verteilen. Dadurch kommen keine Leerflächen vor, so daß der Überspitzungseffekt nicht eintreten kann.

Diese alternative Art der Teiltonbeschränkung weist leider auch Nachteile auf. Gegenüber dem Pegelkriterium klingt fließende Sprache undeutlicher und die tonalen Artefakte nehmen zu. Das liegt daran, daß das zugrunde liegende Berechnungsverfahren auf der Annahme von quasistationären Spektraltonhöhenkandidaten basiert. Teiltöne aber erfüllen diese Eigenschaft im allgemeinen nicht. Aus diesen Gründen springt die Auswahl der Teiltöne von Auswerteintervall zu Auswerteintervall um und unterbricht geschlossene Teiltonverläufe. Die entstehende Unruhe wirkt sich nachteilig aus. Verbesserte Ansätze müßten die zeitlichen Entwicklungen mitberücksichtigen, etwa auf Basis der leider ungenügend erforschten dynamischen Tonhöhenwahrnehmung. Das Pegelkriterium dagegen bewahrt geschlossene Verläufe eher, weil sich stärkere Teiltöne nicht so schnell ändern.

Überspitzungseffekte entstehen also infolge einer Teiltonbeschränkung nach Pegelkriterium, womit eine zeitvariante Sperrung schwächer vertretener Frequenzbereiche zusam-

menhängt. Dieser Vorgang kann allgemein als Kontrastverschärfung der spektral/zeitlichen Hüllfläche des FTT-Spektrums beschrieben werden. Eine alternative Teiltonbeschränkung, nach Tonhöhengewicht, könnte dies verhindern, bringt aber in einer einfachen, Auswerteintervall-orientierten Implementierung eher Nachteile mit sich.

2.6.2 Tonale Artefakte und Tonalisierung

Ausgehend von dem auf zehn Teiltöne reduzierten Muster in Bild 2.11b sind tonale Artefakte erst dann wahrzunehmen, wenn das Auswerteintervall entsprechend Bild 2.11c auf 20 ms vergrößert wird. Für diese Art von Störungen liegt der Grund nicht darin, daß die Teiltonparameter zwischen den Intervallen springen. Sie sind nämlich auch dann noch vorhanden, wenn durch nachträgliche Interpolation stetige Übergänge sichergestellt werden, wie in Bild 2.11d verdeutlicht und später als Operation FKD in Abschnitt 6.3.2 spezifiziert. Ein weiterer Grund, der bereits von Heinbach vermutet wurde [Hei88a, S. 72], kann als Einzelursache ebenfalls ausgeschlossen werden. Demnach können kürzere Teiltonverläufe zwar künstlich auf 20 ms verlängert werden. Allerdings verstärkt sich das 'Klingeln', wenn zuvor die kürzeren Verläufe entfernt wurden (vgl. Abschnitt 4.1.1).

Die Entstehung der Artefakte kann dadurch erklärt werden, daß, im Übergang von Bild 2.11b auf 2.11c sichtbar, 'zittrige' und mehrfach unterbrochene Verläufe für das verlängerte Auswerteintervall 'einfrieren'. Zumindest bei höheren Frequenzen spielt offenbar die *feinzeitliche Variabilität* der Frequenz- und Pegelverläufe ein Rolle. Sie kann bewirken, daß einzelne Synthesesinusschwingungen keine besonders ausgeprägte Tonhöhenwahrnehmung, sondern eher den Eindruck von Schmalbandrauschen auslösen. Wird das Auswerteintervall verlängert, dann erhöht sich die Ausgeprägtheit der zugehörigen Tonhöhenwahrnehmung auf unnatürliche Weise, die Synthesesinusschwingung wird 'tonalisiert'. Die tonalen Artefakte sind demnach keine zusätzlichen Störtöne, sondern nur hervorgehobene Teiltöne des Nutzsignals. Bestärkt wird diese Vorstellung später durch das Ergebnis in Abschnitt 6.3.4.4, wo die Artefakte durch künstliche 'Verzitterung' höherfrequenter Verläufe reduziert werden können.

Die feinzeitliche Variabilität dürfte nicht verloren gehen, wenn das Auswerteintervall T_A ausreichend oft die Abtastung des geglätteten FTT-Leistungsspektrums $p^G_{\omega_A}(t)$ veranlaßt. Als Mindestbedingung ist an jeder Analysefrequenz ω_A das Abtasttheorem (z.B. [Mar82]) einzuhalten. Für eine Überprüfung muß man die Grenzfrequenz f_g des Zeitsignals $p^G_{\omega_A}(t)$ kennen. Sehr optimistisch angesetzt ergibt sie sich als Minimum aus verdoppelter 3dB-Grenzfrequenz des Analysetiefpasses (also der Analysebandbreite B_{3dB}) und der 3dB-Grenzfrequenz f^G_{3dB} des Glättungstiefpasses. Die Bandverdoppelung beim Übergang auf die Leistung begründet Abschnitt 1.5.1. Über die Frequenzgruppenbreite aus Gl. (1.2) läßt sich die übliche Analysebandbreite von 0,1 Bark durch $B_{3dB} = 0,1\Delta f_G$ ausdrücken. Für den Glättungstiefpaß ist $f^G_{3dB} = (2\pi T_G)^{-1}$ nach Gl. (1.20) bis 3 kHz immer größer als B_{3dB}, um danach auf einem konstanten Wert von etwa 120 Hz zu verharren. Man erhält die Formulierung

$$f_g \approx \min\left\{0,1\Delta f_G,\quad 120\ \text{Hz}\right\}. \tag{2.27}$$

Nach dem Abtastheorem sollte

$$f_g < \frac{1}{2T_A} \tag{2.28}$$

gelten. Bei einem Auswerteintervall von $T_A = 20\,\text{ms}$ zeigt sich, daß das Abtastheorem bei Analysefrequenzen oberhalb 1700 Hz nicht eingehalten wird. Schnelle Änderungen im ursprünglichen Teiltonverlauf können somit nicht richtig erfaßt werden. Weil bei f_g keinesfalls eine ideale Bandgrenze vorliegt, ist der Abtastfehler sogar noch kritischer zu bewerten.

Das Klingeln tritt bei reduzierter Teiltonanzahl deutlicher hervor. Dann nämlich befindet sich eine überschaubare Anzahl von Synthesesinusschwingungen in exponierter Lage für eine mögliche Wahrnehmung als Spektraltonhöhe. Bei unbeschränkter Teiltonanzahl ist die Störung eher als eine Tonalisierung der Rauschanteile zu beschreiben. Dieser Effekt ist in abgeschwächter Form aus Abschnitt 2.5 bekannt, wo die Rauschteiltöne schon im unbearbeiteten Teiltonzeitmuster nicht 'schnell' genug fluktuieren.

Somit führt die Verlängerung des Auswerteintervalls allgemein zu einem Tonalisierungseffekt. Der Grund ist, daß die Erfassung der feinzeitlichen Variabilität von Teiltonverläufen durch Verletzung des Abtastheorems behindert wird. Dadurch kann sich die Tonhöhenwahrnehmung der zugeordneten Synthesesinusschwingungen stärker ausprägen. Eine reduzierte Teiltonanzahl wirkt dabei begünstigend. Weil die Verletzung zu höheren Frequenzen hin massiver wird, nimmt die Bedeutung des Effektes im gleichen Sinne zu.

2.6.3 Periodische Knackstörung

Wie bereits betont, bleibt das Rechteckfenster der Teiltonsynthese als Merkmal des datenreduzierenden TTZM-Verfahrens festgeschrieben. Dagegen wurde es bisher beim nichtreduzierenden Verfahren durch das Dreieckfenster ersetzt, weil es uneingeschränkt die Synthesequalität verbessert. Die Ungleichbehandlung liegt daran, daß sich bei Datenreduktion sonst der Gesamteindruck von verarbeiteter Sprache eher verschlechtert. Zwar kann das Rechteckfenster für die periodische Knackstörung verantwortlich gemacht werden. Wenn sie aber durch das Dreieckfenster vermieden wird, dann scheint die subjektive Verständlichkeit zu sinken, weil nichtstimmhafte Anteile unterbewertet wirken. Es ist insgesamt günstiger, das 'Knattern' zu tolerieren, um maximale Verständlichkeit zu behalten. Nachfolgend wird versucht, dies zu objektivieren.

Die Störung entsteht, weil beim Rechteckfenster eine Synthesesinusschwingung für die Dauer des Auswerteintervalls unverändert bleibt und Sprünge an den Intervallgrenzen aufweist (Abschnitt 1.5.4.1). Obwohl Phasenstetigkeit über moderate Frequenzsprünge sichergestellt ist, schlagen Amplitudensprünge auf das Synthesesignal durch (vgl. Bild 1.3 links auf S. 18). Beim unverlängerten Auswerteintervall werden diese noch als unspezifischer, signalangepaßter Störteppich wahrgenommen (Abschnitt 2.3.4). Bei Verlängerung sinkt die Folgerate der Sprünge. Außerdem steigt die durchschnittliche Sprunghöhe, weil sich das FTT-Pegelspektrum zwischen weiter auseinander liegenden Auswertezeitpunkten stärker ändern kann. Statt eines Rauschteppichs wird ein ausgeprägtes, periodisches Knacken wahrgenommen.

Um wie beim nichtreduzierenden TTZM-Verfahren das Dreieckfenster nach Gl. (1.25) mit einer Halbwertsbreite von $T_{6dB} = 1{,}25\,\text{ms}$ anwenden zu können, wird das Auswerteintervall T_A unmittelbar vor Synthese von 20 wieder auf 1,25 ms verkürzt. Dies geschieht, indem jedes Intervall 16 mal wiederholt wird, so daß sich bei Synthese mit Rechteckfenster im Ergebnis nichts ändern würde.

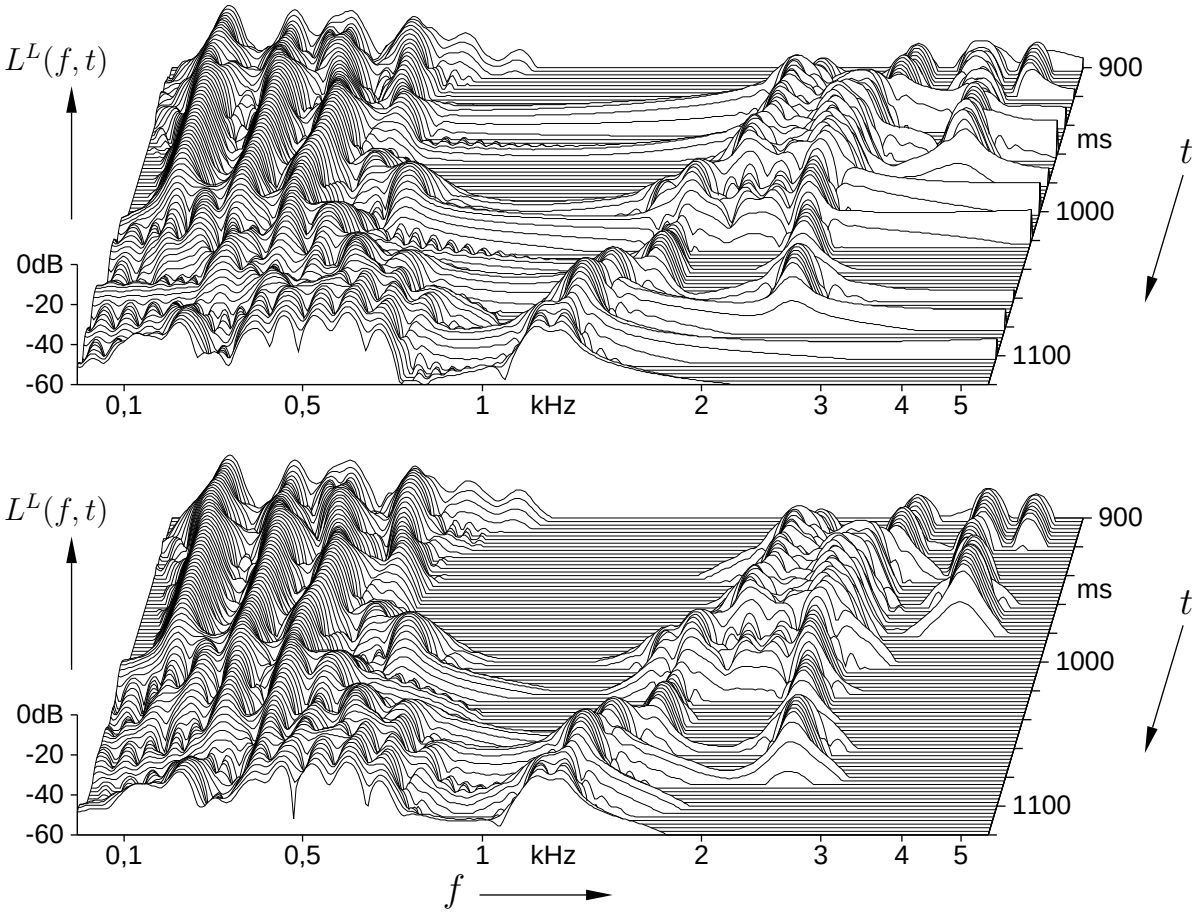

Bild 2.12: *FTT-Pegelspektrum nach Durchlauf des Datenreduktionsverfahrens mit* 4,4 kbit/s *für das Wort 'jeder' bei verschiedenen Syntheseverfahren. Ursprüngliche Teiltonsynthese nach Heinbach mit Rechteckfenster (oben) und modifizierte Synthese mit Dreieckfenster (unten). Die bei Datenreduktion großen Unstetigkeiten im Synthesesignal durch das Rechteckfenster schlagen sich als spektrale Verbreiterungen nieder. Sie werden als Knackstörung wahrgenommen, scheinen aber auch die Verständlichkeit zu erhöhen. Zur Darstellung wurde die FTT mit Gauß-Fensterfunktion* PG4, $B_{3dB} = 0{,}3$ Bark, *Laufzeitausgleich sowie ohne Glättung berechnet (Kapitel 3) und der Zeitmaßstab um ihre Grundlaufzeit bereinigt.*

In Bild 2.12 sind modifizierte FTT-Pegelspektren der Synthesesignale gegenübergestellt, die sich nach TTZM-Verarbeitung bei 4,4 kbit/s mit Rechteck- beziehungsweise mit Dreieckfenster ergeben. Der Sprachausschnitt entspricht demjenigen aus Bild 2.11. Wie in Abschnitt 3.3.5 noch ausführlicher dargelegt, eignet sich die ursprüngliche FTT-Fensterfunktion nicht für die Repräsentation von Schaltknacken. Damit signifikante Unterschiede zwischen beiden Bildern sichtbar werden, wurde die angegebene, modifizierte Fensterfunktion eingesetzt.

In beiden Bildern sind übereinstimmend Formationen mit zeitlicher Vorzugsrichtung zu erkennen, die die Teiltonverläufe widerspiegeln, die auch in Bild 2.11c vorkommen. Zusätzlich alle 20 ms bilden sich im mittleren und hohen Frequenzbereich in Bild 2.12 oben Formationen mit spektraler Vorzugsrichtung aus. Diese 'Querrippen' oder spektrale Verbreiterungen repräsentieren die hörbaren Knacke. Sie schneiden, sieht man von kleinen

Einbrüchen ab, die anderen Formationen auf deren Pegelniveau. Dazwischen sinkt ihr Pegel ab, bewahrt aber ein recht hohes Niveau, besonders im oberen Frequenzbereich. Sie durchschneiden die ausgedehnten spektral/zeitlichen Bereiche, in denen durch Teiltonbeschränkung Signalanteile fehlen. Im übrigen paßt sich die spektral/zeitliche Hüllfläche über die Abfolge der Querrippen grob an das Nutzsignal an, ähnlich wie dies auch für den (Synthesefenster-kontrollierbaren) Störteppich nach Abschnitt 2.3.4 gilt.

Daß die Anreicherung des Spektrums durch spektrale Verbreiterungen für die Sprachverständlichkeit günstig ist, kann auf zwei Wegen gedeutet werden. Zum einen markieren ausgeprägte Verbreiterungen die besonders impulshaft empfundenen Anteile der unverarbeiteten Sprache, wie später in Abschnitt 4.1.2 deutlich wird. Sie kommen bei Plosiven und etwa im Hochtonbereich von Vokalen vor. Die Wahrnehmung könnte die künstlich erzeugten Verbreiterungen teilweise zu einem gewohnten Sprachbild zugehörig erkennen. Zum anderen helfen sie zu verdecken, daß eben im entstört verarbeiteten Signal nach Bild 2.12 unten Spektralanteile fehlen. Dieser 'Brautschleiereffekt' gibt der Wahrnehmung Spielraum, sich die fehlenden Anteile zu 'denken', die sonst eindeutig als fehlend entlarvt würden [HorPK, Hor96, Bre90]. Eine solche Argumentation stützen auch die Ergebnisse von Verschuure und Brocaar, wonach periodisch unterbrochene Sprache verständlicher wird, wenn die Lücken mit Rauschen aufgefüllt werden [Ver83].

Zusammenfassend gilt: Die bei Datenreduktion besonders hohen Amplitudensprünge der Synthesesinusschwingungen erzeugen beim Rechteckfenster eine periodische Knackstörung. Das Dreieckfenster glättet die Sprünge soweit, daß die Knacke fast verschwinden, womit sich leider gleichzeitig die subjektive Verständlichkeit der verarbeiteten Sprache vermindert. Die Knacke können offenbar nutzbringend in die Wahrnehmung des verarbeiteten Sprachsignals integriert werden, weil ihre spektralen Auswirkungen nichtrepräsentierten Signalanteilen ähneln oder ihr Fehlen verschleiern helfen.

2.6.4 Intensitätsmodulation

Das Heinbachsche Verfahren der Pegelcodierung für die zehn Teiltöne eines Auswerteintervalls ermöglicht fast eine Halbierung der Datenrate. Wie in Abschnitt 1.5.3 erklärt, definieren dabei der stärkste und der schwächste Pegelwert eine Interpolationsgerade für die maximal acht übrigen Pegelwerte. Die Eckwerte werden grob in Quantisierungsstufen von 4 dB codiert. In Verbindung mit der Art der Teiltonverlaufsrekonstruktion, die nur konstantgehaltene Werte im Auswerteintervall zuläßt, begünstigt dieses Vorgehen die wahrgenommene Intensitätsmodulation.

Weil sich die Konstellation der ausgewählten zehn Teiltöne gemäß Bild 2.11c laufend verändert, ergibt sich in jedem Auswerteintervall ein neues Interpolationsschema. Seine Eckwerte, aber auch die Position eines zu interpolierenden Teiltonpegels hängen immer von den Pegelverhältnissen aller zehn ab. Der Quantisierungsfehler eines einzelnen Teiltons als Differenz von codiertem und ursprünglichem Pegelwert erhält deshalb eine zufällige Komponente. Bei längeren Teiltonverläufen, wie sie nach Bild 2.11c besonders bei tieferen Frequenzen vorkommen, überlagert sich der zugewiesenen Synthesesinusschwingung dadurch eine Amplitudenmodulation mit einer statistischen Hüllkurve. So entsteht die beobachtete Intensitätsmodulation bei tieferen Frequenzen, die sich mehr als störende Schwankung und weniger als Rauhigkeit auswirkt.

Selbst wenn das spezielle Pegelcodierverfahren nicht eingesetzt wird, bleibt noch eine Rauhigkeit wahrnehmbar. Sie ist auf die Stufigkeit der Syntheseamplituden von längeren Teiltonverläufen zurückzuführen, die durch Konstanthalten im Auswerteintervall bei naturgemäß variierendem Sprachsignal entsteht. Man kann die Stufigkeit als Schwankung um einen geglätteten Verlauf deuten, so daß hier ebenfalls eine störende Amplitudenmodulation zu erkennen ist. Allerdings ist sie nicht zufällig, sondern abhängig vom ursprünglichen Pegelverlauf. Die maximal mögliche Modulationsfrequenz beträgt zwar hier wie oben die halbe Auswerteintervallrate, also $(20\,\mathrm{ms})^{-1} = 25\,\mathrm{Hz}$. Daß hier eher eine Rauhigkeit als eine Schwankung wahrnehmbar ist, liegt wohl in einer unterschiedlichen Modulationsstatistik begründet.

Die Steilheit der Stufenübergänge, die nach dem vorigen Unterabschnitt mit dem Synthesefenster kontrollierbar ist, spielt für die Intensitätsmodulation keine Rolle, da es sich hier um einen schmalbandigen Störeffekt handelt. Seine Wahrnehmbarkeit reduziert sich unabhängig vom gewählten Synthesefenster erheblich, wenn innerhalb zusammenhängender Teiltonverläufe die Parameter interpoliert werden, wie in Bild 2.11d im Vergleich zu 2.11c dargestellt ist. Insgesamt also folgt die störende Intensitätsmodulation in erster Linie aus einer zu einfachen Rekonstruktion der Teiltonverläufe durch Treppenstufenapproximation. Das spezielle Pegelcodierverfahren verschärft diese Problematik deutlich, weil – vereinfacht ausgedrückt – die Auswirkung der Stufen zusätzlich verstärkt wird.

2.6.5 Zusammenfassung und Schlußfolgerung

In diesem Abschnitt wurden Zusammenhänge zwischen Datenreduktionsmaßnahmen und Qualitätsbeeinträchtigungen untersucht, deren Folgen die Übertragungsqualität des datenreduzierenden TTZM-Verfahrens gegenüber der nichtreduzierenden Grundversion erheblich verschlechtern. Am Beispiel der Variante mit 4,4 kbit/s Datenrate wurden folgende Erkenntnisse gewonnen:

- Die Verlängerung des Auswerteintervalls als wichtigste datenreduzierende Maßnahme bewirkt, daß eine mögliche feinzeitliche Variabilität von Teiltonverläufen besonders zu höheren Frequenzen hin nicht mehr richtig repräsentiert wird. Damit steigt die tonale Ausgeprägtheit der zugeordneten Synthesesinusschwingungen, die dann in der Wahrnehmung hervortreten. So können tonale Artefakte entstehen (‘Klingeln’).

- Die Beschränkung auf eine Maximalanzahl gleichzeitiger Teiltöne, ausgewählt nach Pegel, verschärft den Kontrast der spektral/zeitlichen Grobstruktur. Durch diese zweitwichtigste Reduktionsmaßnahme werden nämlich größere Bereiche in der Zeit/Frequenz-Ebene nicht mehr repräsentiert. Ein verarbeitetes Sprachsignal kann deshalb unnatürlich überspitzt (‘überartikuliert’) klingen, weil die steilen spektralen oder zeitlichen Übergänge in diese Bereiche wahrnehmbar werden. Der Effekt könnte eventuell durch Einsatz des Tonhöhengewichts als Teiltonauswahlkriterium verhindert werden. Die Anpassung an die zeitliche Dynamik der Tonhöhenwahrnehmung stellt allerdings ein eigene Forschungsaufgabe dar.

- Die Verlängerung des Auswerteintervalls verändert den ursprünglich unspezifischen, Synthesefenster-kontrollierbaren Störteppich der Synthese mit Rechteckfenster. Die

Sprünge im Amplitudenverlauf der Synthesesinusschwingungen werden nun als periodische Knackstörung ('Knattern') wahrnehmbar. Zwar könnte man diese bei Synthese mit Dreieckfenster weitgehend unterdrücken. Es scheint aber für die subjektive Sprachverständlichkeit besser, sie zu tolerieren. Offenbar kann damit das Fehlen von Signalanteilen verschleiert werden, die bereits der Datenreduktion zum Opfer gefallen sind, oder die von vornherein im TTZM-Verfahren unterrepräsentiert sind.

- Das Pegelcodierverfahren mittels Interpolationsgerade begünstigt eine störende Intensitätsmodulation der Stimme. Primär wird sie jedoch durch das verlängerte Auswerteintervall ausgelöst, die resultierende Stufigkeit der Syntheseparameter wird besonders bei längeren, tieffrequenteren Teiltonverläufen wahrgenommen. Eine geeignete Teiltonverlaufsrekonstruktion kann den Störeffekt weitgehend unterdrücken.

Mit der beschränkten Anzahl von zehn Teiltönen, deren Parameter alle 20 ms vorgegeben sind, kann man also grundsätzlich nicht alle Anteile von fließender Sprache in akzeptabler Qualität zu Gehör bringen. Das erste zentrale Problem liegt in der mangelnden Repräsentation nichttonaler Anteile. Man kann dies zwar durch Störeffekte verschleiern, zahlt dafür aber den Preis einer unakzeptablen Signalqualität. Ein strategischer Lösungsansatz erfordert eine separate Repräsentation solcher Anteile. Das zweite zentrale Problem liegt darin, für die tonalen Anteile die optimalen Teiltöne im Sinne einer dynamischen Tonhöhenwahrnehmung herauszusuchen. Will man im übrigen die Störungen reduzieren, so muß man sich vermehrt um eine Rekonstruktion der ursprünglichen Teiltonverläufe zwischen den Auswertezeitpunkten bemühen.

2.7 Zusammenfassung

In diesem Kapitel wurde analysiert, warum die erreichbare Verarbeitungsqualität beim Heinbachschen TTZM-Verfahren beschränkt ist und warum sie bei seiner datenreduzierenden Variante erheblich zurückgeht. Dazu wurde das nichtreduzierende TTZM-Verfahren mit verschiedenen einfachgearteten, synthetischen Testsignalen untersucht, die mit wahrnehmbaren Verfälschungen verarbeitet werden. Aus den Beobachtungen ergab sich eine Reihe von charakteristischen Verfälschungseffekten, die auch für Qualitätseinbußen bei Sprachverarbeitung verantwortlich gemacht werden können. Sie lassen sich nach Ursachen in drei Kategorien einteilen. Die erste Kategorie beinhaltet Effekte aufgrund von Eigenschaften der Spektraltransformation:

Glättung der Schmalbandhüllkurve: Das Teiltonzeitmuster kann schnellen, schmalbandigen Hüllkurvenmodulationen nicht schnell genug folgen. Dies resultiert aus der unzureichenden Gehöranpassung der zeitlichen Auflösung der bisherigen FTT. Daraus kann man schließen, daß die realisierten Vor- oder Nachverdeckungsschwellen des Verfahrens im Vergleich mit dem Gehör zu hoch sind. Der Effekt leistet einen Beitrag zur raumübertragungsähnlichen Verfremdung ('Halligkeit'), denn Räume rufen auch diese Veränderung hervor. Er könnte durch Erhöhung der Analysebandbreite vermieden werden, was sich zunächst aber nicht mit dem folgenden Sachverhalt verträgt.

Überhöhte Simultanverdeckung: Die Simultanverdeckung im Teiltonzeitmuster liegt höher als beim Gehör, weil die spektralen Selektionseigenschaften der FTT nicht ausreichen. Dadurch können schwächere Töne in der Umgebung eines Maskierertons nicht mitverarbeitet werden, die man im Originalsignal noch hören kann. Eine Verringerung der Analysebandbreite könnte hier helfen, wenn dies nicht im Widerspruch zum vorigen Sachverhalt stünde. Das Dilemma kann man mit der bisherigen Analysefensterfunktion der FTT nicht lösen, bei der die von Heinbach verwendete Bandbreite zumindest einen guten Kompromiß verkörpert.

Die zweite Kategorie betrifft die Vollständigkeit einer Audiorepräsentation durch das Teiltonzeitmuster:

Unterrepräsentation transienter Signalanteile: Gemeint sind Signalbestandteile, die kurzzeitige spektrale Verbreiterungen im FTT-Pegelspektrum hervorrufen und zur Wahrnehmung von 'Klicks' oder 'Knacken' beitragen. Sie werden benachteiligt und verfälscht, weil das Teiltonextraktionsprinzip (Frequenzkonturierung) die Verbreiterungen nicht erfassen kann. Eindrucksvoll zeigt sich dies bei Einzelimpulsen und Impulsfolgen. Die Unterrepräsentation wird von der bisher gewählten Analysefensterfunktion begünstigt, mit der sich kein Ausschaltknack im FTT-Spektrum abzeichnet. Bei Sprachverarbeitung gibt es allerdings eine Reihe von günstigen Effekten, die eine stärkere Verfälschung verhindert.

Tonalisierung von Rauschanteilen: Verarbeitete Rauschanteile klingen je nach Analysebandbreite 'zeitvariierend nasal' oder 'plätschernd'. Ursache ist wiederum das Teiltonextraktionsprinzip (Frequenzkonturierung), welches im regellosen FTT-Spektrum definitionsgemäß Töne zu erkennen sucht. Dadurch schleicht sich eine Regelhaftigkeit in das verarbeitete Signal ein, die als Tonalisierung bezeichnet werden kann. Der Effekt stellt sich als spezielle Konsequenz des vorigen Sachverhaltes dar und ist später bei korrekter Repräsentation und Rekonstruktion transienter Anteile vermeidbar. Er liefert aber auch eine Argument für die Einrichtung einer eigenständigen Repräsentation von Rauschanteilen.

Strenggenommen ist das Teiltonzeitmuster demnach keine vollständige Audiorepräsentation. Die dritte und letzte Kategorie enthält zwei Störeffekte, deren Ursachen in der Teiltonsynthese, also in der Signalrekonstruktion begründet liegen. Sie entstehen im Übergang zwischen spektraler und zeitlicher Repräsentation von Signalmerkmalen durch das Teiltonzeitmuster. Hier prägt sich beispielsweise ein sinusbasiertes Modulationssignal nicht mehr als zeitvarianter Einzelteilton, aber noch nicht als Satz von stationären Teiltönen aus. Der Übergang manifestiert sich durch spezielle Teiltonformationen mit Kurzverläufen, Aufspaltungen und Verschmelzungen.

Synthesefenster-kontrollierbare Störungen: Innerhalb der Formationen ändern sich Teiltonpegel schlagartig. Werden die Teiltonparameter direkt in Synthesesinusschwingungen umgesetzt, dann entsteht eine Vielzahl von breitbandigen Knacken. Weichberandete Synthesefenster können die Störwirkung reduzieren.

Phaseninkohärenz-bedingte Störungen: Bei den Formationen können gleichzeitig mehr als zwei Teiltöne innerhalb einer Frequenzgruppenbreite vorkommen. Aber

die Teiltonsynthese stellt für die zugeordneten Synthesesinusschwingungen keine kohärenten Phasenlagen sicher, obwohl sich Indizien hierfür bieten. Die für das Gehör auswertbare Signalhüllkurve wird dadurch verfälscht. Der Effekt tritt sogar bei Spaltungen und Verschmelzungen auf, wenn Wechsel von einer auf zwei Synthesesinusschwingungen und umgekehrt stattfinden.

Weil sich bei Sprach-TTZM die Charakteristika solcher Formationen in der ganzen Zeit/Frequenz-Ebene zeigen, erzeugen beide Störungen eine Art Störteppich. Er paßt sich an die spektral/zeitliche Grobstruktur des Nutzsignals an, bleibt dabei aber wahrnehmbar. Wie noch zu sehen sein wird, entschärft dies die Effekte der zweiten Kategorie etwas.

Das Ende des Kapitels behandelte Zusammenhänge zwischen Maßnahmen zur Datenreduktion und den dabei zusätzlich auftretenden Qualitätsbeeinträchtigungen. Die Ergebnisse sind ab Seite 53 detaillierter zusammengefaßt. Prinzipiell kann man mit einer beschränkten Anzahl von Teiltönen, deren Parameter nur noch im groben Zeitabstand vorgegeben sind, fließende Sprache nicht mehr mit akzeptabler Qualität darstellen. Nichttonale Anteile erfordern dringend eine eigene Repräsentationsform, auch wenn sie zunächst im datenreduzierten Signal enthalten zu sein scheinen. Dies hängt mit der spezielle Natur der auftretenden Störungen zusammen. Sie können vermutlich eine Art Ersatz bieten oder die Eigenschaft der Wahrnehmung herausfordern, verdeckte Anteile zu rekonstruieren, selbst wenn diese gar nicht vorhanden sind. Für tonale Anteile muß man die wesentlichen Teiltöne auswählen können. Dafür wird im Grunde das Modell einer noch zu erforschenden dynamischen Tonhöhenwahrnehmung benötigt. Um weitere Störungen zu vermeiden, muß man die zeitlich grob abgetasteten Teiltonverläufe noch vor Synthese wiederherstellen.

Kapitel 3

Konturierung im zeitvarianten FTT-Pegelspektrum

Die ersten beiden Kategorien von Ursachen, die Verfälschungseffekte beim Heinbachschen TTZM-Verfahren bewirken, haben mit Codierung, Datenreduktion oder der Rekonstruktion des Signals nichts zu tun. Sie betreffen vielmehr die Eigenschaften der Spektraltransformation und die prinzipielle Vollständigkeit des Teiltonzeitmusters als Audiorepräsentation. Zur Abhilfe werden in diesem Kapitel eine Konzepterweiterung sowie weitere Modifikationen eingeführt, die sich auf den Gewinnungsprozeß einer gehörorientierten Audiorepräsentation mit Konturen konzentrieren. Sie sollten sich weiterhin mit dem Terhardtschen Modell einer auditiven Informationsverarbeitung vereinbaren lassen. Die ursprüngliche Teiltonanalyse wird dadurch zur allgemeineren Konturanalyse weiterentwickelt.

Zuerst wird ein Konzept der Konturierung im zeitvarianten FTT-Pegelspektrum vorgestellt, welches das Heinbachsche Teiltonkonzept erweitert. Darin ergänzt der neuartige Vorgang der Zeitkonturierung die bisherige Frequenzkonturierung, um bisher übersehene, gehörrelevante Information zu erfassen. Sie steckt in den transienten Formationen des zeitvarianten Pegelspektrums und repräsentiert überwiegend impulshaft empfundene Signalanteile. Es sind nunmehr also zwei Konturtypen zu unterscheiden, für die der zweite Abschnitt die Darstellung von elementaren Signalen und Sprache vorstellt. Dies geschieht unter Voraussetzung von Parametern und Verfahrensweisen, die erst in den letzten beiden Abschnitten besprochen werden.

Bei der Spektraltransformation gibt es über die Wahl der FTT-Fensterfunktion noch Raum für Optimierungen von Eigenschaften. Im dritten Abschnitt wird deshalb das Zusammenspiel von Konturierung und Fensterfunktion und damit verbundenen Parametern untersucht. Mit diesen Erkenntnissen können im vierten Abschnitt Einstellungen der neuen Konturanalyse optimiert und begründet werden. Mit den Ergebnissen läßt sich auch das konzeptbeschränkte TTZM-Verfahren mit vergleichsweise geringem Aufwand in der Qualität verbessern.

3.1 Konturierungskonzept

Konturierung im zeitvarianten FTT-Spektrum wird nun als zweiteiliges Konzept definiert. Frequenzkonturierung, der bereits bekannte erste Teil, liefert Frequenzkonturen, die bisher das Teiltonzeitmuster darstellten. Transformations- und Konturierungsparameter stimmen allerdings nicht notwendig mit den speziellen Vorgaben von Heinbach überein. Weil die bisherige Nomenklatur zu Mißverständnissen führen kann, sind einige Umbenennungen vorteilhaft. Als zweiter Teil werden die neuartigen Zeitkonturen eingeführt. Beide Teile zusammen realisieren in etwa die visuelle Analogie der 'Gratlinien' eines 'FTT-Pegelgebirges', wie in der Einleitung dieser Arbeit vorgestellt.

3.1.1 Übergang vom Teiltonzeitmuster auf Frequenzkonturen

Heinbach verbindet mit der Bezeichnung 'Teilton' einen zeitvarianten Linienverlauf im Teiltonzeitmuster [Hei88a, S. 19 und 98]. Dieser sei von nun an als *Frequenzkonturlinie* bezeichnet. Er setzt sich aus einer Menge von *Frequenzkonturpunkten* zusammen, die wie bisher den ausgeprägten lokalen Maxima in der Teiltonextraktion/Frequenzkonturierung entsprechen (Abschnitt 1.5.2). Der Zusammenhang der Punkte als Linienverlauf ist in Anhang A.2 formal definiert. Alle Linien oder auch alle Punkte zusammen ergeben nun nicht mehr das Teiltonzeitmuster, sondern die *Frequenzkonturen* eines Signals. Anhang A.1 definiert sie formal im Kontinuum, ihre Diskretisierung folgt daraus in einfacher Weise. Die zeitvariante Menge der zu einem Zeitpunkt vorhandenen Punkte, das frühere Teiltonmuster, wird später als *Frequenzkontursignal* eingeführt.

Diese begrifflichen Änderungen sind aus zwei Gründen sinnvoll. Erstens wird eine Einführung von 'Zeitkonturen' vorbereitet, welche einem gleichartigen Konturierungsvorgang mit anderer Vorzugsrichtung entspringen. Die Beziehung zwischen beiden Vorgängen drückt sich dadurch auch in der Benennung aus. Zweitens werden eine Reihe von potentiellen Verwechselungen der Bezeichnung 'Teilton' vermieden. Sie sind bei Heinbach im Sinne der Universalität eines grobqualitativen Gehörmodells teilweise gewollt, erweisen sich aber in der vorliegenden Arbeit als hinderlich. Ausführungen in [Hei88a] sowie begriffliche Assoziationen zum Wortstamm 'Ton' lassen folgende Interpretationen zu, deren prinzipielle Unterschiede herauszustellen sind:

Frequenzkonturpunkt (FKP): ausgeprägtes lokales Maximum des momentanen FTT-Pegelspektrums im Zeit/Frequenz/Pegel-Raum (Teiltonabtastwert im Teiltonmuster).

Frequenzkonturlinie (FKL): Zeitverlauf eines ausgeprägten lokalen Maximums des momentanen FTT-Pegelspektrums im Zeit/Frequenz/Pegel-Raum (Teiltonverlauf im Teiltonzeitmuster).

Quellsinusschwingung (QSS): zeitvariante Sinusschwingung als Komponente einer exakten Beschreibung des Quellsignals durch Superposition (Teilton eines komplexen Tons bei Stationarität).

Synthesesinusschwingung (SSS): zeitvariante Sinusschwingung als Komponente bei der Berechnung des Synthesesignals durch Superposition.

Spektraltonhöhe (STH): Objekt der analytischen tonalen Wahrnehmung in einem Schall [Ter72a], beschreibbar durch Parameter einer einzelnen Sinusschwingung mit gleicher Wahrnehmung.

Die Gleichstellung von FKP und Teilton ist nur bei stationären FKL sinnvoll, weil deren Zeitverlauf dann keine wesentliche Information beinhaltet. Die letzten vier Interpretationen können alle als Linien im Zeit/Frequenz/Pegel-Raum aufgefaßt werden. Für einfache Schalle können sie tatsächlich annähernd übereinstimmen. Dies gilt beispielsweise für einen langsamveränderlichen Tonkomplex, bei dem die QSS frequenzmäßig weit genug auseinander liegen. Bei Sprachsignalen, bei denen naturgemäß schnelle Modulationseffekte, Rauschen oder transienten Anteile vorkommen, kann man nicht mehr davon ausgehen.

Daß die FKL und QSS nicht äquivalent sein können, zeigt der Fall der Pulsfolge in Abschnitt 2.1.2. Dort bleiben ganze Spektralbereiche durch FKL unrepräsentiert, obwohl eine Pulsfolge exakt mit Harmonischen QSS dargestellt werden kann. Allgemein fehlt einer Quellsignaldarstellung mit Hilfe von QSS die Eindeutigkeit. Beispielsweise kann man eine Schwebung, die durch zwei stationäre Sinusschwingungen erzeugte wurde, auch als eine einzige, zeitvariante Sinusschwingung auffassen. Äquivalenz läßt sich auch deshalb nicht halten.

Eine Äquivalenz von FKL und SSS gibt es in der Tat, denn sie begründet die Verfahrensvorschrift der ursprünglichen Teiltonsynthese mit Rechteckfenster. Allerdings können QSS und SSS dann nicht mehr äquivalent sein, denn sonst müßte dies widersprüchlicherweise auch für QSS und FKL gelten. Um Handlungsspielraum für eine verbesserte Signalrekonstruktion aus Konturen zu erhalten, muß man auf strenge Äquivalenz von FKL und SSS verzichten (Kapitel 5).

Das Beispiel von Weißem Rauschen macht plausibel, daß auch keine Äquivalenz von FKL und STH besteht. In seinen Frequenzkonturen (Bild 2.9 links auf S. 44) kann man nicht alle FKL als Repräsentanten für tonale Wahrnehmungen ansehen. Zwar stellt die Berechnung der FKL ein konkretes Beispiel für eine Entscheidungsebene im Terhardtschen Modell dar, in dem Spektraltonhöhen eine wichtige Rolle spielen (Abschnitt 1.3). Deren Wahrnehmung läßt sich jedoch nicht unmittelbar durch Frequenzkonturierung modellieren, offenbar müssen weitere Entscheidungsprozesse bereitgestellt werden (Kapitel 4).

3.1.2 Einführung von Zeitkonturierung und Zeitkonturen

Nach den Ergebnissen von Abschnitt 2.1 können Frequenzkonturen impulsartige Signalanteile nicht befriedigend repräsentieren. Dies hängt mit dem bisherigen Konturierungsvorgang zusammen, der prinzipiell transiente Formationen des FTT-Spektrums nicht richtig erfassen kann. Als natürliche Ergänzung bietet sich die Zeitkonturierung an.

3.1.2.1 Zweckmäßigkeit eines weiteren Konturierungsvorganges

Das FTT-Pegelspektrum des Dirac-Impulses in Bild 2.1 links auf S. 24 zeigt bei nochmaliger Betrachtung folgendes: Der gleichzeitige, sprunghafte Anstieg an allen Frequenzen ruft eine charakteristische Kante hervor, die die ganze spektrale Breite des Vorgangs

wiedergibt. Trotzdem findet sich in den Frequenzkonturen (Bild 2.1 rechts) keine entsprechende Konturlinie. Stattdessen entsteht eine andere, deren Verlauf als unzuverlässig und energetisch unrepräsentativ eingestuft werden mußte.

Die Kante bleibt als Konturlinie unentdeckt, weil sie parallel zur Frequenzachse und somit genau parallel zu den Schnitten der Pegelfläche verläuft, in denen die Frequenzkonturierung lokale Maxima sucht. Aus diskreten lokalen Maxima kann sich innerhalb eines Schnittes aber keine Linie formen. Wenn man dagegen auch alle möglichen Schnitte parallel zur Zeitachse betrachtet, so liefern ihre lokalen Maxima auch alle Punkte der gesuchten Konturlinie. Dieses Vorgehen überträgt den Terhardtschen Gedanken der spektralen Konturierung auf die zeitliche Dimension des zeitvarianten FTT-Pegelspektrums. Damit wird ein neuer, zusätzlicher Konturierungsvorgang geschaffen.

Es mag zunächst übertrieben erscheinen, für eine frequenzparallele Konturlinie einen separaten Konturierungsvorgang zu fordern. Man könnte sich beispielsweise auch vorstellen, die Schnitte nicht exakt parallel zur Frequenzachse auszurichten, sondern systematisch leicht schräg zu stellen. Dann würde jeder Punkt der Kante in genau einem Schnitt als lokales Maximum erkennbar sein. Damit ist das prinzipielle Problem des bisherigen Konturierungsvorganges aber nicht behoben, sondern nur verlagert. Es sind nämlich auch Konturlinien denkbar, die parallel zu den nunmehr schräggestellten Schnitten verlaufen. Um aber alle Konturen durch einen einzigen Konturierungsvorgang zu finden, müßte man ganz von der Schnittbetrachtung abkehren und Methoden der Differentialgeometrie auf die Pegelfläche im Zeit/Frequenz/Pegel-Raum anwenden. Riley, der spezielle Zeit/Frequenz-Repräsentationen von Sprache konturiert, bietet in [Ril89] eine mögliche Lösung [HorPK] (vgl. Fußnote S. 64).

Für einen separaten Konturierungsvorgang spricht folgendes: Erstens sind getrennte Maximumdetektionen in Zeit- und Frequenzschnitten elementare, auch physiologisch plausible Vorgänge. Dies kann man von einem einzigen Konturierungsvorgang weniger behaupten, der gleich die Pegelfläche im Zeit/Frequenz/Pegel-Raum verarbeitet. Zweitens existieren bereits Gehörmodelle, welche von zwei getrennten Verarbeitungskanälen für transiente beziehungsweise quasistationäre Signalanteile ausgehen ([Ber89, Wu89] auf der Grundlage von Chistovich [Chi82]). Ohne Blick auf die angestrebte gehörgerechte Informationsaufbereitung gibt es noch einen dritten Grund für getrennte Vorgänge. Er betrifft allein den praktischen Einsatz von Konturen in einem Codierverfahren: Bei der Signalrekonstruktion in Kapitel 5 muß eine Konturlinie, die von transienten Spektralanteilen herrührt, eine völlig andere energetische Gewichtung erhalten, als diejenige, die beispielsweise von einer quasistationären Sinusschwingung abstammt. Beide Konturtypen müssen dazu unterscheidbar sein. [1]

3.1.2.2 Zeitkonturierung und Zeitkonturen

Der Vorgang, lokale Maxima in Schnitten des zeitvarianten FTT-Pegelspektrums parallel zur Zeitachse auszuwerten, wird *Zeitkonturierung* genannt. Wie später Abschnitt 3.3 begründet, geht man von einem ungeglätteten, laufzeitausgeglichenen Pegelspektrum $L^L(f,t)$ aus, das mit geeigneten Fensterfunktionen gewonnen wird. Ähnlich wie bei der

[1] Tatsächlich gibt es eine Möglichkeit, Konturpunkte beider Konturierungsvorgänge ungetrennt in den Rekonstruktionsvorgang einzuspeisen (Anhang D.3). Dies konnte aber erst am Ende der Arbeit als Konsequenz der getrennten Behandlung nachgewiesen werden.

Frequenzkonturierung ist die Zuordnung lokaler Maxima zu Konturpunkten an eine Ausgeprägtheitsbedingung geknüpft, um hörphysiologisch unbedeutende Schwankungen des Pegelzeitverlaufes übergehen zu können. Weil es darum geht, transiente Spektralanteile zu erfassen, wird dazu die Anstiegsgeschwindigkeit $\partial L^L(f,t)/\partial t$ ausgewertet. Zur Veranschaulichung zeigt Bild 3.1 einen zeitparallelen Schnitt an einer exemplarischen Frequenz $f = f_1$. Die Anstiegsgeschwindigkeit entspricht der Steigung der Kurve. Erreicht oder überschreitet sie einen Schwellwert λ, dann legt der nächste Zeitpunkt, an dem sie nicht mehr positiv ist, einen *Zeitkonturpunkt* fest. [2]

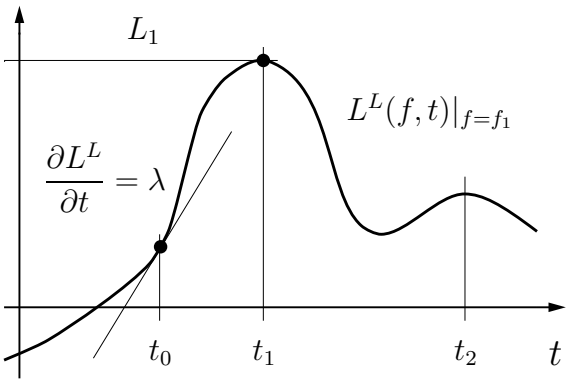

Bild 3.1: *Konturierung des Pegelspektrums in Zeitrichtung an der Frequenz f_1. Ein Zeitkonturpunkt ist (L_1, f_1, t_1), weil hier ein lokales Maximum vorliegt und vorher die Anstiegsgeschwindigkeit des Pegels mindestens den Schwellwert λ erreicht hat. Bei t_2 wird letzteres nicht erfüllt.*

Die auf diese Weise gefundenen Punkte formen wie bei der Frequenzkonturierung Linien, die nun *Zeitkonturlinien* heißen. Die Gesamtheit dieser Linien beziehungsweise Punkte stellt die *Zeitkonturen* des zeitvarianten FTT-Pegelspektrums dar. Eine formale Definition liefert Anhang A.1.

Warum stützt sich die Bewertung der Ausgeprägtheit nicht wie bei der Frequenzkonturierung darauf, wie weit sich ein lokales Maximum gegenüber benachbarten Minima heraushebt? Dies würde die Kausalität verletzen, weil zur Bewertung eines Kandidaten schon der weitere Verlauf bekannt sein müßte. Zwar könnte man mit einer systematischen Verzögerung arbeiten, mit deren Hilfe schon die Unterschreitung eines Mindestwertes 'vorhersehbar' wäre. Bei dem dargestellten Verfahren aber braucht man praktisch nur die Vorgeschichte bis hin zum Kandidaten in Betracht zu ziehen.

Der Schwellwert λ wird proportional zur Analysebandbreite B_{3dB} gewählt. Bei geeigneten Fensterfunktionen verkörpert sie ein Maß für den schnellstmöglichen Anstieg zwischen einem Minimum und einem Maximum bei gegebenem Pegelunterschied. Mit der Proportionalität läßt sich deshalb folgendes Verhalten sicherstellen, das hörphysiologisch grob plausibel erschien: Es werden genau dann keine Zeitkonturpunkte mehr zugeordnet, wenn zeitliche Schwankungen des Pegelzeitverlaufs unter eine gewisse, frequenzunabhängige Pegeldifferenz fallen. Tatsächlich wird die Dimensionierung von λ jedoch allein dadurch belegt, daß ein optimales Rekonstruktionsergebnis erzielbar ist. Vorgreifend auf Abschnitt 3.4 ist dies, geeignete Fensterfunktionen vorausgesetzt, für folgende Dimensionierung der Fall:

$$\lambda = 25\ \mathrm{dB} \cdot B_{3dB}. \tag{3.1}$$

[2] Das hier beschriebene Verfahren wurde gegenüber einer früheren Veröffentlichung [Mum90] modifiziert, um tatsächlich mit dem Maximum zu arbeiten.

3.1.3 Zusammenfassung

Das Konturierungskonzept der vorliegenden Arbeit lautet wie folgt: Die Konturierung des zeitvarianten FTT-Pegelspektrums setzt sich aus Frequenz- und Zeitkonturierung zusammen. Der erste Vorgang entspricht der Teiltonextraktion im TTZM-Verfahren. Er produziert zeitvariante Linienverläufe von ausgeprägten lokalen Maxima über der Frequenz. Das Ergebnis wird ab jetzt als Frequenzkonturen bezeichnet, um voreiligen Interpretationen als quellenorientierte Beschreibung, als Vorschrift zur Signalrekonstruktion oder als Beschreibung von tonaler Wahrnehmung entgegenzutreten. Der neuartige zweite Vorgang findet in analoger Weise zum ersten statt, indem ausgeprägte lokale Maxima über der Zeit Linienverläufe mit Vorzugsrichtung parallel zur Frequenzachse erzeugen. Sie ergeben die bisher unbekannten Zeitkonturen. Auch die Zeitkonturierung erhält in Gestalt des Schwellwertes λ eine Ausgeprägtheitsschwelle zugewiesen.

3.2 Signalanalyse unter Berücksichtigung von Zeitkonturen

Dieser Abschnitt behandelt die Konsequenzen des erweiterten Konturierungskonzeptes bei der Signalanalyse, indem die Konturausbildung für einige Signale untersucht wird. Natürlich kann das Verhalten von Frequenzkonturen nach den Arbeiten von Heinbach und den Ergänzungen des vorigen Kapitels als hinreichend bekannt gelten. Das Hauptaugenmerk liegt deshalb auf den neu hinzutretenden Zeitkonturen, auf dem Zusammenspiel beider Konturtypen und auf deren Zuordnung zur Gestalt des zeitvarianten FTT-Pegelspektrums. Die ersten beiden Signale sind synthetischer Natur. Die hierbei gewonnenen elementaren Erkenntnisse helfen anschließend, die Zeitkonturen eines Sprachsignals zu interpretieren.

Eine neue Einstellung der Transformations- und Konturierungsparameter wird nachfolgend als gegeben angenommen (ZFKII in Tabelle 3.2 auf S. 87) und erst später in Abschnitt 3.4 begründet. Sie unterscheidet sich von der Heinbachschen Einstellung (HB-TTZM) nicht nur hinsichtlich der festzulegenden Zeitkonturierung. Zusätzlich verbessert sie die Qualität reiner Frequenzkonturverarbeitung nach Art des TTZM-Verfahrens.

3.2.1 Impulsfolge

Das Testsignal mit linear ansteigender Impulsfolgefrequenz ist bereits aus Abschnitt 2.1.2 bekannt. Bild 3.2a zeigt den ersten Abschnitt seiner Frequenzkonturen, in welchem die Impulsfolgefrequenz von 20 bis hinauf zu 95 Hz reicht. In Bild 3.2b sind zusätzlich die Zeitkonturen als schmale vertikale Linien eingeblendet. Aus Gründen der Übersichtlichkeit ist die Liniendicke der Zeitkonturen konstant, sie gibt also nicht wie bei den Frequenzkonturen den Pegel wieder. Aber Linienlänge und Linienverlauf enthalten bereits wertvolle Information. [3]

[3] Pegelabhängige Liniendicke von Zeitkonturen überfrachtet die Darstellung, sofern die Zeitachse nicht erheblich gestreckt wird. Zur phonetischen Analyse von Sprachsignalen (Formanten) ist die pegelabhängige Liniendicke aber auch schon für Frequenzkonturen nicht optimal. Allgemein eignen sich pegelabhängige

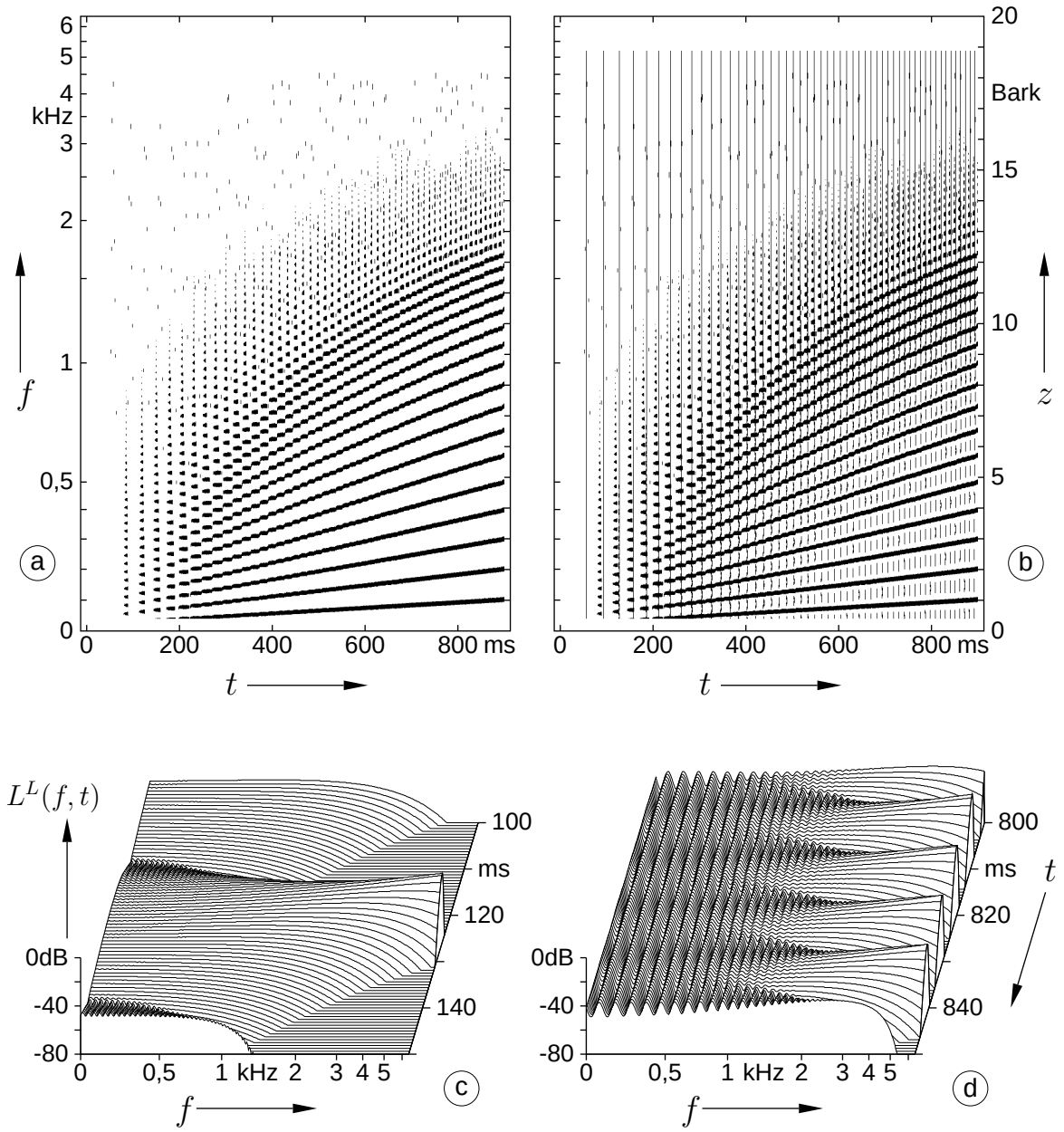

Bild 3.2: *Konturen der Impulsfolge mit steigender Folgefrequenz (vgl. Bild 2.2): a) nur Frequenzkonturen, b) Zeitkonturen zusätzlich sowie c,d) zwei Ausschnitte des zugrunde liegenden FTT-Pegelspektrums. Erst Zeitkonturen repräsentieren die hörbaren spektralen Ausbreitungen parallel zur Frequenzachse. Hier und im folgenden gibt die Liniendicke der Zeitkonturen* nicht *– wie bei Frequenzkonturen – den Pegel wieder. Weil gegenüber Bild 2.2 andere Analyseparameter gelten, stimmen Teiltonzeitmuster dort alias Frequenzkonturen hier nicht exakt überein.*

Weiterhin sind in Bild 3.2c und 3.2d ein früherer und ein späterer zeitlicher Ausschnitt des konturierten FTT-Pegelspektrums dargestellt. Im ersten sind die Impulse wegen der niedrigen Folgefrequenz als einzelne frequenzparallele 'Rippen' aufgelöst, die von den Impulsantworten der Analysefilter gebildet werden. In Bild 3.2b schlagen sich die Rippen

Farbwerte besser. Eine sehr gut lesbare Darstellung entsteht alternativ, wenn man zusätzlich das FTT-Spektrogramm unterlagert, wobei dem Pegel Grauwerte zugeordnet werden. Hierzu kann das Spektrogramm auch aus den Konturen zurückgewonnen werden (z.B. mit Hilfe der Operationen WFS/WZS nach Abschnitt 4.3).

als Zeitkonturlinien nieder, die über die gesamte spektrale Breite reichen. Es liegen im wesentlichen aufeinanderfolgende Situationen von Einzelimpulsen vor.

Innerhalb der dargestellten Pegeldynamik in Bild 3.2c wirken die aufeinanderfolgenden Impulsantworten an einer hohen Analysefrequenz noch zeitlich völlig getrennt. Bei einer tieferen Analysefrequenz überlappen sie sich bereits, weil sie durch die kleinere Analysebandbreite länger ausfallen. In den Überlappungsbereichen entstehen Welligkeiten parallel zur Frequenzachse, die zu den kurzzeitigen Frequenzkonturlinien am Anfang von Bild 3.2b führen. Deren Linienabstand spiegelt auf geringem Pegelniveau bereits die momentane Impulsfolgefrequenz wider.

Wenn die Impulsfolgefrequenz zunimmt, werden die Überlappungsbereiche zeitlich immer breiter, weil sich Höhen und Dauern der Impulsantworten nicht verändern. Die Überlappung reicht nun auch zu hohen Frequenzen hinauf. In Bild 3.2d ist dies gut zu erkennen. Sie ist dort bei tiefen Frequenzen schon so groß, daß die Welligkeitsbereiche ineinanderfließen und in Bild 3.2b durchgängige Frequenzkonturlinien hervorrufen. Diese liegen auf hohem Pegelniveau und teilen die zuvor durchgängigen Zeitkonturlinien in kurze Abschnitte auf niedrigem Pegelniveau. Offensichtlich kann nur der durchgängige Konturtyp die spektral/zeitliche Energiedichte des FTT-Spektrums ausreichend repräsentieren, selbst wenn 'Linienstückchen' des anderen vorhanden sind.

Allerdings existiert eine Übergangszone, wo sich Linien unterschiedlichen Typs nicht gegenseitig unterbrechen, sondern überkreuzen. Damit liegen lokale *Doppelrepräsentationen* der spektral/zeitlichen Energiedichte vor, weil in der Umgebung des Überkreuzungspunktes für beide Linientypen das gleiche Pegelniveau vorherrscht. Dies stört zwar bei der Signalanalyse nicht, erfordert aber zusätzliche Maßnahmen bei der Signalrekonstruktion (Abschnitt 5.1.3). Überkreuzungen kommen auch schon beim Einzelimpuls vor, wo eine markante Zeitkonturlinie über die schwach ausgeprägte Frequenzkonturlinie laufen muß (Abschnitt 3.1.2). Letztere findet sich in Bild 3.2a in Form von einzelnen Punkten bei höheren Analysefrequenzen wieder, die in 3.2b als Verdickungen genau auf den Zeitkonturlinien liegen. Die unregelmäßige Lage ist auf einen Abtasteffekt durch das Auswerteintervall zurückzuführen. [4]

Das Beispiel der Impulsfolge zeigt, daß die Zeitkonturen bei einem periodischen Signal Anteile des FTT-Pegelspektrums erfassen können, die der Frequenzkonturierung nicht zugänglich sind. Bei Sprachsignalen können dies höhere Harmonische sein. Die Beobachtungen verdeutlichen außerdem, daß sich die Dualität von Zeit- und Frequenz in der Kurzzeitspektralanalyse nur dann in den Konturen widerspiegelt, wenn auch Zeitkonturen zugelassen sind. Die Repräsentativität beider Konturtypen für eine gegebene spektral/zeitliche Energieverteilung schließt sich im wesentlichen gegenseitig aus. Doppelrepräsentationen durch beide Konturtypen können allerdings lokal vorkommen.

[4] Um Doppelrepräsentationen von vornherein zu vermeiden, muß das Konturierungskonzept ergänzt werden. Dazu könnte man für jeden Konturpunkt der bisherigen Definition noch die Krümmungen der FTT-Pegelfläche im Zeit/Frequenz/Pegel-Raum in Zeit- und Frequenzrichtung vergleichen: Um einen Zeitkonturpunkt zu bestätigen, muß die Krümmung in Zeitrichtung stärker sein. Zur Bestätigung eines Frequenzkonturpunktes muß sie in Frequenzrichtung stärker sein. Nichtbestätigte Punkte werden verworfen. Der Einfluß der Bandbreite ist zuvor zu eliminieren.

3.2.2 Tonsignal mit Hüllkurvenänderung

Als Testsignal $s(t)$ dient ein Sinuston der Frequenz $f_T = 1\,\text{kHz}$, der sich mit einstellbarer Übergangsgeschwindigkeit ein- und ausschalten läßt. Speziell geht seine zeitabhängige Amplitude $A(t)$ durch Tiefpaßfilterung aus einer Gleichspannung hervor, die bei $t = 0$ ein- und bei $t = 0 + \Delta t$ wieder ausgeschaltet wird. Für die Filterung wird ein kausal approximierter Gauß-Tiefpaß mit der Impulsantwort $h^{gs}(t)$ verwendet (PG8 nach Tabelle B.1). Mit Hilfe des Einheitssprunges $\gamma(t)$ und des konstanten Faktors A ergibt sich die Beschreibung

$$s(t) = A(t)\cos(2\pi f_T t), \quad \text{mit} \quad A(t) = A \cdot [\gamma(t) - \gamma(t - \Delta t)] * h^{gs}(t). \tag{3.2}$$

Die Übergangsgeschwindigkeit wird über die 3dB-Bandbreite B^{gs}_{3dB} des Gauß-Tiefpasses eingestellt, die zur FTT-Analysebandbreite in Beziehung gesetzt werden kann. Letztere beträgt $B_{3dB} = 0{,}3$ Bark, was bei Analysefrequenzen nahe 1 kHz etwa 50 Hz entspricht. Bei niedrigeren Gauß-Bandbreiten liegen im wesentlichen quasistationäre Verhältnisse im FTT-Spektrum vor. Dessen spektrale Auflösung ist nämlich dann gröber als die Verbreiterung, die das Fourier-Spektrum von $s(t)$ aufgrund der Übergänge erfährt. Mit vier Gauß-Bandbreiten $B^{gs}_{3dB} = 50/200/500/\infty$ Hz werden dagegen zunehmend transiente Verhältnisse eingestellt. Im letzten Fall springt die Amplitude. Die Einschaltdauer beträgt konstant $\Delta t = 100$ ms. Bei Darbietung des Signals hört man an beiden Übergängen zunehmend Knacke. In Bild 3.3 sind Frequenzkonturen, Zeitkonturen und FTT-Pegelspektren der vier Fälle abgebildet.

In den Frequenzkonturen in Bild 3.3a bilden sich bei 1 kHz Hauptlinien jeweils mit der Länge der Einschaltdauer aus. Trotz steigender Übergangsgeschwindigkeit behalten sie ihr Aussehen, insbesondere tauchen sie kaum schneller auf oder ab. Die Analysebandbreite beschränkt hier die Geschwindigkeit, mit der sich der Linienpegel ändern kann, wie bei der Hüllkurvenglättung in Abschnitt 2.2 dargelegt. Deshalb tragen die Frequenzkonturen keine wesentliche Information über das Ausmaß der Knackwahrnehmung. Dafür erkennt man anhand der Spektren in Bild 3.3c-f, daß an den Übergängen zunehmend spektrale Verbreiterungen hervortreten. Sie werden in den Zeitkonturen in Bild 3.3b als lange, gerade Hauptlinienstücke erfaßt, die sich als prominente Objekte gut zur Modellierung der Knackwahrnehmung eignen.

Obwohl die Spektren beim Ein- und Ausschalten auf den ersten Blick gleichartig aussehen, gibt es bei den Konturen zwei wesentliche Unterschiede. Erstens bilden sich gegen Ende des Einschaltvorganges bei beiden Konturtypen Gebiete, die dicht mit kurzen Nebenlinien angefüllt sind. Sie gehören zu den fein oszillierenden Welligkeiten des Spektrums, die in Bild 3.3d-f in der Verschneidung zwischen Verbreiterung und stationärem Verlauf zu beobachten sind. Gegenüber den Hauptlinien verlaufen sie auf einem Pegelniveau, das im wesentlichen vernachlässigbar ist. Durch Unzulänglichkeiten bei der Signalrekonstruktion können sie dennoch eine gewisse Bedeutung erlangen (Abschnitt 5.1.7.2).

Zweitens gibt es bei den Zeitkonturen gleichartig eingebogene Teilverläufe nahe 1 kHz, die ebenfalls nur beim Einschalten vorkommen. [5] Wie ein Blick auf Bild 3.3d-f verdeutlicht, repräsentieren sie keine deutliche spektrale Verbreiterungen mehr. Hier zeigt sich

[5] Den Einschaltverlauf unmittelbar bei 1 kHz in Bild 3.3b muß man sich als liegendes 'V' ergänzt vorstellen. Weil darstellungstechnisch jedem Zeitkonturpunkt ein kleiner senkrechter Strich zugeordnet wird und die Dichte der Analysefrequenzen endlich ist, zerfällt der sichtbare Linienzusammenhang.

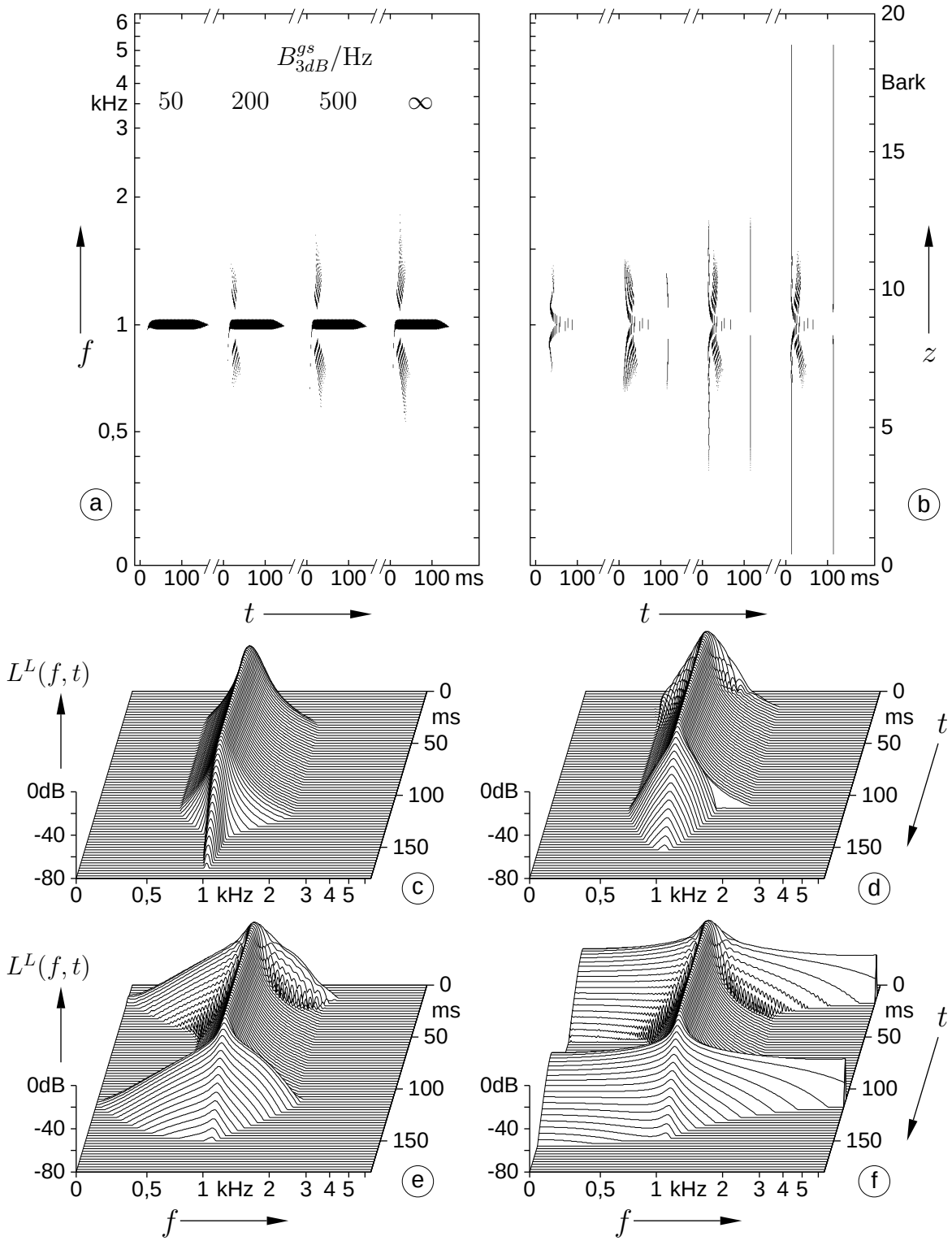

Bild 3.3: *Konturierung von Tonimpulsen $f_T = 1\,\mathrm{kHz}$ der Dauer $\Delta t = 100\,\mathrm{ms}$ mit variierten Gauß-Übergangsflanken. Für vier Übergangsgeschwindigkeiten, repräsentiert durch B^{gs}_{3dB}: a) Frequenzkonturen, b) Zeitkonturen sowie c-f) Ausschnitte der zugrunde liegenden Pegelspektren. Der vierte Fall entspricht 'hartem' Ein- und Ausschalten. Spektrale Verbreiterungen markieren die zunehmend hörbaren Übergänge, die nur von Zeitkonturen erfaßt werden können. Die dichten Ansammlungen von kurzen Konturlinien liegen auf vorerst vernachlässigbarem Pegelniveau.*

eine Konsequenz des Konturierungskonzeptes: Zeitkonturpunkte werden eben auch dann zugeordnet, wenn der Pegelzeitverlauf an einer Analysefrequenz ein Plateau erreicht und nicht sofort oder nur sehr langsam absinkt. Aus demselben Grund gibt es sogar eine Zeitkonturlinie beim quasistationären Fall $B_{3dB}^{gs} = 50$ Hz. Wenn aber eine Zeitkonturlinie oder ein Teil ihres Verlaufes nicht von einer spektralen Verbreiterung verursacht wurde, liegt eine Doppelrepräsentation der spektral/zeitlichen Energiedichte vor. Wie im vorigen Unterabschnitt erwähnt, müssen zur Signalrekonstruktion Vorkehrungen dagegen getroffen werden. Doppelrepräsentation ist wohlgemerkt keine Eigenheit der Zeitkonturierung, denn auch die Frequenzkonturierung liefert etwa bei einem Impuls 'unnötigerweise' eine Frequenzkonturlinie (Abschnitt 2.1.1).

Spektrale Verbreiterungen, Nebenlinien und das Problem der Doppelrepräsentation existieren auch dann, wenn die Übergänge nicht bei Amplitude null beginnen oder enden. Verbreiterungen und Nebenlinien heben sich allerdings um so weniger heraus, je kleiner der Pegelsprung beim Übergang ausfällt. Wird weiterhin die Hüllkurve eines Tonkomplexes verändert, dann existieren die Verbreiterungen jeweils in der spektralen Umgebung der einzelnen Tonkomponenten. Zwischen zwei Komponenten laufen sie 'glatt' ineinander. Diese Situation war vereinfacht in Bild 1 der Einleitung zu sehen.

Schnelle Hüllkurvenänderungen von Tonsignalen rufen also kurzzeitige spektrale Verbreiterungen im zeitvarianten FTT-Pegelspektrum hervor, die als Zeitkonturen erfaßt werden. 'Schnell' bedeutet, daß die Bandbreite des Hüllkurvensignals die Analysebandbreite in der Umgebung einer Tonkomponente übersteigt. Mit schnellerer Änderung geht eine stärkere Verbreiterung einher, die mit einer deutlicheren Knackwahrnehmung bei der Hördarbietung zusammentrifft. Diese Zusammenhänge können nur durch Zeitkonturen, nicht aber durch Frequenzkonturen modelliert werden. Wie noch zu sehen sein wird, hängt das Auftreten von Verbreiterungen mit der Wahl geeigneter FTT-Fensterfunktion zusammen. Bei der im Heinbachschen TTZM-Verfahren verwendeten Fensterfunktion sind sie nicht zu beobachten. Im übrigen gibt es bei Hüllkurvenänderungen Nebenkonturlinien und lokale Doppelrepräsentationen, die bei der Signalrekonstruktion im Auge zu behalten sind.

3.2.3 Sprachsignal

Für das Sprachsignal eines männlichen Sprechers zeigt Bild 3.4 oben die Frequenzkonturen. Durch die unten eingeblendeten Zeitkonturen gewinnt die Darstellung an Aussagekraft. So wirkt die zeitliche Strukturierung verstärkt, Pausen und Lautgrenzen treten besonders bei höheren Frequenzen markanter hervor. In spektral/zeitlichen Gebieten, in denen vorher keine wesentliche Energiedichte vorhanden schien, tauchen längere Zeitkonturlinien auf und weisen auf transiente Spektralanteile hin. Für einige kurze und schwache Frequenzkonturlinien stellt sich nun heraus, daß sie von einer langen Zeitkonturline überkreuzt werden. Damit erweisen sie sich als eher untergeordnete, wenig repräsentative Begleiterscheinungen. Beim abschließenden 'b' kann man diese Situation gut erkennen. Weiterhin erhalten Frequenzkonturlinien, die zuvor in scheinbar ungeordneten Gruppen auftraten, durch Zeitkonturen einen Platz in einer charakteristischeren, übergeordneten Struktur. Als Beispiel mag hier das 'e' der Lautfolge 'd–e–r' im mittleren Frequenzbereich dienen.

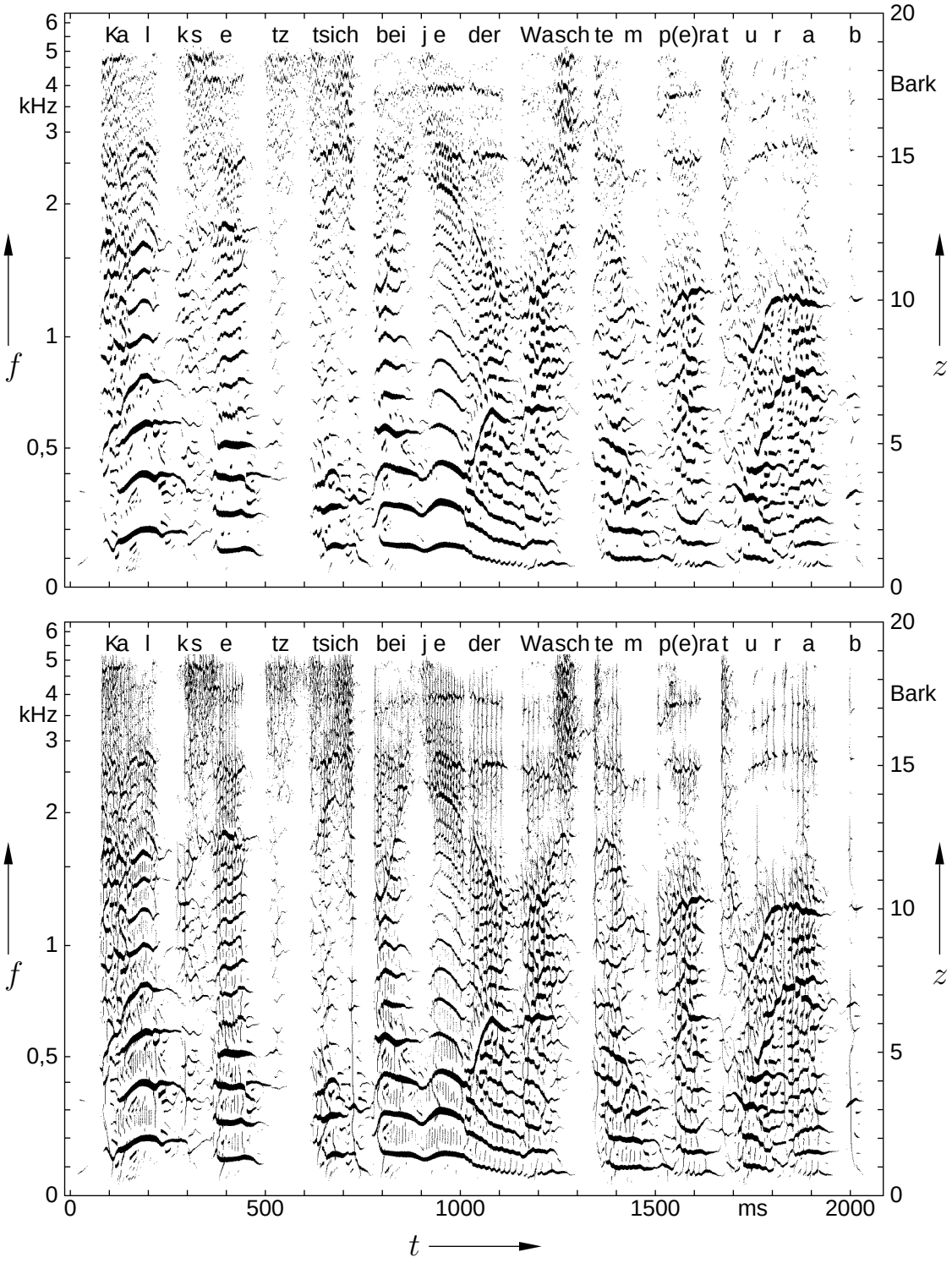

Bild 3.4: *Frequenzkonturen (oben) und ihre Ergänzung um Zeitkonturen (unten) für Sprache eines männlichen Sprechers. Erst Zeitkonturen machen bestimmte Signalanteile sichtbar, etwa Glottisimpulse bei höheren Frequenzen und impulsähnliche Anteile bei Plosiven. Damit tritt auch die zeitliche Segmentierung deutlicher heraus. Durch die verbesserten Analyseparameter sind die Frequenzkonturen zeitlich besser aufgelöst als im Heinbachschen Teiltonzeitmuster in Bild 1.2 oben.*

Bei ausgeprägten Vokalen entsteht oft die gleiche Situation wie bei der zuvor behandelten Pulsfolge: Zu höheren Frequenzen hin behaupten sich Zeitkonturen, zu tieferen Frequenzkonturen. Das ‘e’ in ‘j–e’ liefert dafür ein Beispiel. Die oben durchgängigen Zeitkonturlinien geben die Anteile der Glottisimpulse wieder, die trotz Vokaltrakt- und Raumübertragung bewahrt worden sind. Stärkere Frequenzkonturlinien entstehen dann hauptsächlich nur im Bereich des dritten und vierten Formanten. Sie verkörpern deren Eigenschwingungen, die durch die Glottisimpulse angeregt werden. Bei tieferen Frequenzen werden die Zeitkonturlinien von den dort dominierenden Frequenzkonturen unterbrochen. Die resultierenden Linienstücke liegen genauso wie bei der Pulsfolge auf tiefem Pegelniveau und repräsentieren keine signifikante Energie.

Stimmunterdrückungen durch glottalen Verschluß sind dadurch zu erkennen, daß Zeitkonturlinien plötzlich über der gesamten spektralen Breite ausbleiben. Dies ist beispielsweise nach dem anfänglichem ‘l’ und dem folgenden ‘e’ der Fall. Weil das Pegelniveau des Spektrums besonders an tiefen Frequenzen nur langsam absinken kann, verschwinden die Frequenzkonturen dort nicht sofort. Allerdings kann sich der Verschluß in beiden Fällen nicht besonders ‘schnell’ ereignet haben, im Sinne des vorigen Unterabschnittes. Sonst wären im selben Moment spektrale Verbreiterungen und damit längere Zeitkonturlinien zu beobachten.

In der Verschlußlösungsphase ausgeprägter Plosive bilden sich längere Zeitkonturlinien aus, etwa bei ‘t’, ‘p’ und ‘b’ in der zweiten Hälfte des Schalls. Aber auch der Beginn eines Verschlusses kann solche Linien auslösen. Dies ist im Fall des labialen Verschlusses bei ‘ch’ für das nachfolgende ‘b’ zu sehen. Die Zeitkonturen unterstützten folglich die Plosiverkennung. Auch die Frikativerkennung fällt leichter: ‘s’, ‘ch’ und ‘sch’ heben sich im Hochtonbereich als unregelmäßige Strukturen deutlich von den regelmäßigen Glottisimpulsen der Vokale ab.

3.2.4 Zusammenfassung

Signalanalyse mit dem erweiterten Konturierungskonzept wurde an Signalbeispielen studiert. Impulsfolge und schnelle Hüllkurvenänderung im Tonsignal stehen für einfache Situationen, in denen die Zeitkonturen in der Lage sind, in den Frequenzkonturen nicht vorhandene, wahrnehmungsrelevante Information aus dem FTT-Pegelspektrum zu ergänzen.

Allgemein erfassen Zeitkonturen kurzzeitige, breitbandige Formationen im Spektrum, die auf transiente Ereignisse im Signal zurückzuführen sind. Damit sind sie das spektral/zeitliche Gegenstück zu Frequenzkonturlinien, die schmalbandige, quasistationäre Ereignisse verkörpern. Normalerweise dominiert, in kleineren spektral/zeitlichen Gebieten betrachtet, einer der beiden Konturtypen. Zwischen den dominanten Konturlinien des einen Typs können kurze Konturlinien des anderen auftreten, deren energetischer Beitrag nicht relevant ist. Ausnahmen stellen lokale Umgebungen um Überkreuzungspunkte dar, in denen das Spektrum durch beide Konturtypen gleichermaßen repräsentiert wird.

Durch Zeitkonturen ergeben sich verbesserte Möglichkeiten der Sprachanalyse. Sie lassen sich im wesentlichen darauf zurückführen, daß zu höheren Frequenzen hin zunehmend Glottisimpulse sowie impulsähnliche Anteile von Verschlüssen als lange Linien sichtbar werden. Daß solche Linien bei Sprache wahrnehmungsrelevant sind, machen die Beobachtungen der ersten beiden Signale plausibel. Die Bestätigung erfolgt aber letztendlich erst

über die Rekonstruktion ab Abschnitt 3.4, wenn auch die Einstellung der Transformations- und Konturierungsparameter nachgeholt wird. Längere Zeitkonturlinien werden sich als Träger impulshaft wahrgenommener Sprachanteile erweisen (Abschnitt 4.1.2).

3.3 Zusammenspiel von FTT-Fensterfunktion und Konturierung

Nach den Ergebnissen des vorigen Kapitels spielt die FTT-Fensterfunktion offenbar eine zentrale Rolle bei der Konturierung. Dieser Abschnitt sucht nach Fensterfunktionen, die mit dem Terhardtschen Modell vereinbar sind, und behandelt systematisch, wie sich ihre zeitlichen und spektralen Eigenschaften auf das Konturierungsergebnis auswirken. Ziel ist es, auf theoretischem Wege Klarheit über die Natur einer optimalen Fensterfunktion zu erlangen und Konsequenzen für davon abhängige Konturierungsparameter zu beurteilen. Es werden nur reellwertige Funktionen in Betracht gezogen. [6]

Erst wird gezeigt, wie beliebige FTT-Fensterfunktionen festgelegt und mit welchen Maßen sie untereinander verglichen werden. Mit dieser Methodik werden dann Kriterien formuliert, die die Brauchbarkeit einer Fensterfunktion im Rahmen des Terhardtschen Modells sicherstellen. Eine Auswahl von Funktionen verdeutlicht, daß nur wenig Variationsspielraum innerhalb dieser Kriterien übrigbleibt. Die meisten Funktionen lassen einen neu einzuführenden Laufzeitausgleich sinnvoll erscheinen. Für eine spezielle Familie von Funktionen, die die Menge geeigneter Typen charakterisiert, untersucht ein weiterer Unterabschnitt die Konturausbildung. Welche Bedeutung eine zeitliche Glättung des Spektrums und die Konturausgeprägheitsschwellen dabei haben, wird abschließend behandelt.

3.3.1 Spezifikation und Beurteilungsmaße von Fensterfunktionen

Um verschiedene Fensterfunktionen zu beschreiben, dient die Systeminterpretation der FTT aus Abschnitt 1.4.2 als Ausgangspunkt (Bild 1.1 auf S. 10): Man erhält den Zeitverlauf $s_{\omega_A}(t)$ des FTT-Spektrums an einer Analysefrequenz ω_A mit Hilfe eines Modulators und eines Analysetiefpasses. Letzterer ist geeignet festzulegen, weil seine Impulsantwort $h_{\omega_A}(t)$ identisch mit der Fensterfunktion ist. Als FTT-Analysebandbreite hängt seine Bandbreite B_{3dB} noch von ω_A ab. Um diese Abhängigkeit zu eliminieren, wird von einem normierten Tiefpaß mit Grundverstärkung eins ausgegangen. Dadurch kann man auf bekannte Tiefpaßtypen zurückgreifen, deren Systemfunktionen katalogisiert sind.

Als Normierungskonstante für Zeiten und Frequenzen dient üblicherweise die 3dB-Grenzkreisfrequenz ω_{3dB}. Die normierte Impulsantwort $h^N(T)$ ist über der normierten Zeit T definiert, ihre Fourier-Transformierte $H^N(\Omega)$ über der normierten Frequenz Ω. Ihre Laplace-Transformierte $H^N(P)$ über der normierten P-Ebene wird als Systemfunktion vorgegeben. Für $P = j\Omega$ stimmt sie mit der Fourier-Transformierten überein, da nur kausale

[6] Es sind auch komplexe Fensterfunktionen denkbar, womit an den Analysefrequenzen eine asymmetrische spektrale Selektion erreicht werden kann. Ohne formalen Bezug zur FTT stellt Baumann diesen Fall als Bandpaßfilterbank dar, um eine mittlere Anhebung der oberen Erregungsflanke eines psychoakustischen Funktionsschematas zur realisieren [Bau95].

und stabile Systeme in Betracht kommen. Für Impulsantwort, Fourier- beziehungsweise Laplace-Transformierte des bei ω_A zu realisierenden Tiefpasses gelten folgende *Entnormierungsvorschriften*:

$$\begin{aligned} h_{\omega_A}(t) &= 2\omega_{3dB} \cdot h^N(T) \quad \text{und} \quad T = t \cdot \omega_{3dB}, && (3.3)\\ H_{\omega_A}(\omega) &= 2 \cdot H^N(\Omega) \quad \text{und} \quad \Omega = \omega/\omega_{3dB}, && (3.4)\\ H_{\omega_A}(p) &= 2 \cdot H^N(P) \quad \text{und} \quad P = p/\omega_{3dB}, && (3.5) \end{aligned}$$

$$\text{jeweils mit} \quad \omega_{3dB} = \pi B_{3dB}(\omega_A).$$

Hiermit ist gleichzeitig eine Grundverstärkung $H_{\omega_A}(0) = 2$ voreingestellt. Sie hat vorwiegend historische Gründe und steckt auch in der Definition der Fensterfunktion nach Gl. (1.3) [Fel85, Hei88a]. Ein stationäres Maximum im FTT-Betragsspektrum gibt auf diese Weise genau die Amplitude A einer Sinusschwingung $A\cos(\omega_T t) = \frac{A}{2}e^{j\omega_T t} + \frac{A}{2}e^{-j\omega_T t}$ im Quellsignal wieder. Normalerweise 'sieht' die FTT in der Nähe des Maximums nur die Amplitude $\frac{A}{2}$ einer der beiden Halbschwingungen.

Realisierungstechnische Einzelheiten zur Berechnung des FTT-Spektrums $s_{\omega_A}(t)$ beziehungsweise der betragsgleichen Form $s^B_{\omega_A}(t)$ nach Gl. (1.10) finden sich ab Anhang B.2. Das dort beschriebene zeitdiskrete System approximiert das zu realisierende zeitkontinuierliche System über eine Bilinear-Transformation von $H^N(P)$. Deshalb sind die eben formulierten Entnormierungsvorschriften nicht exakt anwendbar, was für die Praxis aber unwesentlich ist. Zur besseren Übersicht bleibt der Haupttext bei der Vorstellung, zeitkontinuierliche Systeme zu realisieren.

Um die Wirkung verschiedener Fensterfunktion beurteilen und vergleichen zu können, haben sich drei Maße auf der Grundlage der normierten Spezifikation bewährt. Die *spektrale Selektion* $a(\Omega)$ und *die zeitliche Selektion* $a(T)$ werden wie folgt definiert:

$$a_H(\Omega) = 20\text{lg}(|H^N(\Omega)|)\ \text{dB}, \qquad a_h(T) = 20\text{lg}(|h^N(T)|)\ \text{dB}. \tag{3.6}$$

Sie geben darüber Auskunft, inwieweit Spektralanteile in der Umgebung einer Analysefrequenz beziehungsweise inwieweit Zeitereignisse im Verlauf eines Analysefensters beachtet werden. Die (normierte) *Gruppenlaufzeit* $T_g(\Omega)$ als drittes Maß läßt sich aus $H^N(P)$ ableiten und beschreibt das Zeitverhalten des Tiefpasses aufgeschlüsselt nach Spektralanteilen. Bei reellen Fensterfunktionen gilt [Wol78]:

$$T_g(\Omega) = \text{Re}\left\{-H(P) \cdot \frac{dH(P)}{dP}\right\}_{P=j\Omega}. \tag{3.7}$$

Für die drei Maße finden sich in Bild 3.6 Beispielverläufe, die später noch ausführlich diskutiert werden. Während $a_H(\Omega)$ mit umgekehrten Vorzeichen als Dämpfung des normierten Tiefpasses bekannt ist, ist die halblogarithmische Darstellung seiner Impulsantwort unüblich (Bild 3.6c). Weil Konturierung über dem Pegelspektrum definiert wurde, scheint es angebracht, die spektralen und auch die zeitlichen Selektionseigenschaften einer Fensterfunktion gemeinsam in logarithmischen Maßstäben zu studieren. Zu beachten ist, daß die Betragsbildung in Gl. (3.6) negative Überschwinger von $h^N(T)$ um die Zeitachse nach oben klappt. Dadurch tauchen negative und positive Überschwinger gleichermaßen als Nebenmaxima in der zeitlichen Selektion auf.

3.3.2 Eignungskriterien für Fensterfunktionen

Überlegungen anhand der eben vorgestellten Beurteilungsmaße führen zu dem Ergebnis, daß sich viele Fensterfunktionen nicht im Kontext des Terhardtschen Verarbeitungsmodells (Abschnitt 1.3) eignen. Sie münden in drei Kriterien, die den Freiraum bei der Wahl erheblich einschränken:

- Die spektrale Selektion $a_H(\Omega)$ darf keine Nebenmaxima aufweisen.
- Die zeitliche Selektion $a_h(T)$ darf höchstens sehr niedrige Nebenmaxima aufweisen.
- Das Niveau der Gruppenlaufzeit $T_g(\Omega)$ sollte nicht zu hoch sein.

Wenn die spektrale Selektion Nebenmaxima besitzt, dann tauchen bei den Frequenzkonturen Nebenlinien auf. Bei einem stationären Sinuston beispielsweise fällt der Spektralpegel nun nämlich nicht mehr monoton im Abstand von der Tonfrequenz. Wenn Analysefrequenzen erreicht werden, an denen der Ton im Bereich eines Nebenmaximums selektiert wird, steigt er vorübergehend wieder an. Auf diese Weise entstehen so viele Nebenlinien wie Nebenmaxima vorhanden sind. Nebenlinien widersprechen aber dem Grundgedanken der Konturierung im Terhardtschen Modell, der eine Reduktion auf das Wesentliche fordert. Allerdings sind Nebenlinien unterhalb einer vernünftigen Dynamik bedeutungslos.

Wegen der Gleichartigkeit von Zeit- und Frequenzkonturierung gelten ähnliche Überlegungen auch für die zeitliche Selektion. Kommen darin Nebenmaxima vor, so löst beispielsweise ein Impuls in den Zeitkonturen zusätzlich Nebenlinien aus. Praktische Erfahrungen mit verschiedenen Fensterfunktionen zeigen allerdings, daß das zweite Kriterium nicht so streng wie das erste zu formulieren ist. Man kann das damit erklären, daß sich die Energie eines maximal ausgesteuerten Impulses über eine große spektrale Breite verteilt. Der Spektralpegel kann deshalb nicht so hohe Werte wie bei einem voll ausgesteuerten Sinuston erreichen. Folglich liegen Nebenlinien der Zeitkonturen viel eher unterhalb einer vernünftigen Dynamik.

Nicht nur bei der Zeitkonturierung sind höhere Nebenmaxima der zeitlichen Selektion nachteilig. Sie verursachen auch störende Kurzverläufe und Verlaufsversetzungen in den Frequenzkonturen, wenn das FTT-Spektrum schnelleren zeitlichen Änderungen unterworfen ist. Gewissermaßen folgt den Änderungen dann ein ‘Echo’ (z.B. Fig. 1 in [Sch90]), das kurzfristig Spektralmaxima über der Frequenz hervorrufen oder beeinflussen kann. Zur Verdeutlichung der Auswirkungen bei Sprache ist in Bild 3.5 links ein Frequenzkonturausschnitt zu sehen. Bei ansonsten gleichen Parametern stellt die rechte Seite denselben Ausschnitt dar, für den eine Fensterfunktion mit zeitlichen Nebenmaxima eingesetzt wurde (P4 statt 4P1, siehe nächster Unterabschnitt). Die Veränderungen sind in einem entsprechend modifizierten TTZM-Verfahren als deutliche Störungen wahrnehmbar.

Das letzte Kriterium gründet sich auf der Überlegung, daß ein hohes Laufzeitniveau den evolutionstheoretischen Prinzipien des Terhardtschen Modells widerspricht. Um schnelle Entscheidungsprozesse sicherzustellen, darf die Verarbeitung nur geringe Laufzeiten beanspruchen. Dieses Kriterium wird sich als gleichbedeutend mit der Forderung herausstellen, den Aufwand für die Annäherung an eine informationstheoretisch optimale, aber akausale Gauß-Fensterfunktion realistisch zu halten.

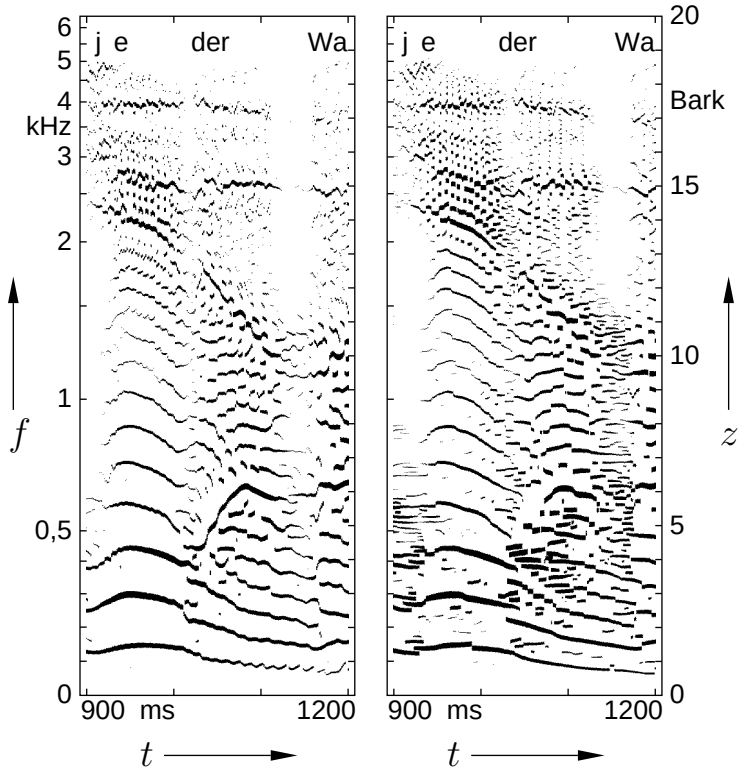

Bild 3.5: *Auswirkung zeitlicher Nebenmaxima der Fensterfunktion auf Frequenzkonturen von Sprache: Fenster* 4P1 *ohne (links) und Fenster* P4 *mit Nebenmaxima (rechts). Bei Anwesenheit von Nebenmaxima treten störende Kurzverläufe und Verlaufsversetzungen auf.*

3.3.3 Variationsspielraum anhand ausgewählter Fensterfunktionen

Welcher Variationsspielraum bleibt, wenn die zuvor entwickelten Eignungskriterien beachtet werden? Beispielsweise brauchen übliche Fenster der diskreten Fourier-Transformation wegen ihrer meist hohen spektralen Nebenmaxima gar nicht erst in Betracht gezogen werden [Har78, Ada91]. Zur Klärung werden einige Fensterfunktionen diskutiert, die durch bekannte, katalogisierte Tiefpaßtypen mit rationalen, nullstellenfreien Systemfunktionen darstellbar sind. Die Auswahl reflektiert gleichzeitig die Menge der Fensterfunktionen, mit denen im Verlauf der Arbeit praktische Erfahrungen bei der Konturierung gesammelt wurden. Im nächsten Kapitel kommen sie auch als Synthesefenster in Betracht.

Tabelle 3.1 faßt die untersuchten Fensterfunktionen nach Familien und Entwurfsziel zusammen und definiert Kürzel für jede. Einen wichtigen Parameter stellt der Grad n der Systemfunktion $H^N(P)$ dar, der mit Polanzahl und Realisierungsaufwand gleichzusetzen ist. P1 repräsentiert die von Terhardt ursprünglich vorgeschlagene und von Heinbach verwendete Fensterfunktion nach Gl. (1.3). Sie ist mit Grad eins Mitglied aller Familien. Die Erhöhung des Grades bedeutet einen Übergang von den Eigenschaften von P1 hin zum Entwurfsziel der jeweiligen Familie. Die Daten der Systemfunktionen finden sich in Tabelle B.1 im Anhang. Bild 3.6 zeigt für einige dieser Typen Selektionsverläufe und Gruppenlaufzeitverlauf.

Am Beispiel der nP1 in Bild 3.6a erkennt man, daß die spektrale Selektion $a_H(\Omega)$ mit steigendem Grad n innerhalb einer Familie steiler wird. Bei gegebenem n weist die Familie nP1 die geringste Steilheit auf, was im Verein der Kurven für $n = 4$ gut zu sehen ist. Über die Familien PGn, Bn bis hin zu Pn nimmt die Steilheit zu. Die Unterschiede sind bei $n = 2$ noch sehr gering – daher ist nur 2P1 eingezeichnet – sie prägen sich erst mit

Tabelle 3.1: *Übersicht untersuchter FTT-Fensterfunktionen, welche den Impulsantworten (IPA) von bekannten Tiefpaßfamilien mit einem bestimmten Entwurfsziel entsprechen. Der von Heinbach verwendete Fenstertyp* P1 *ist das einfachste Mitglied in jeder Familie.*

bisherige Fensterfunktion: P1 (einfacher Pol)	Kürzel	Familienname	Entwurfsziel
→	nP1	n-facher Pol	kein Überschwingen der IPA
→	PGn	Pseudo-Gauß	spektrale Gauß-Approximation
→	Bn	Bessel (Thomson)	maximal flache Gruppenlaufzeit
→	Pn	Potenz (Butterworth)	maximal flache Dämpfung
Grad $n \in \{1, 2, 3, 4, 8\}$; für $n = 1$ identisch mit P1			

wachsendem n deutlicher aus. Die Selektion von P4 liegt bis herab zu -45 dB schon unter der von Mitgliedern anderer Familien mit $n = 8$. Unterhalb von $\Omega = 1$ verhalten sich alle Familien unabhängig von n nahezu gleichartig, was Bild 3.6b herausstellt. Allein die Pn heben sich durch schnelles Abkippen der Selektion auf hohem Niveau hervor. Die übrigen Familien unterscheiden sich demnach erst auf tieferem Niveau.

Für die bessere spektrale Selektion bei den Pn muß man eine ungünstige zeitliche Selektion $a_h(T)$ in Bild 3.6c hinnehmen. Wegen der schwach gedämpften Nebenmaxima würden auch weiter entfernt liegende zeitliche Ereignisse noch in das Fenster fallen. Die Kurven für $n = 4$ zeigen, daß man bei gegebenem n eine höhere spektrale Selektivität nur mit Nachteilen bei der zeitlichen Selektivität erkaufen kann. Die ersten Nebenmaxima der Familien PGn und Bn sind aber noch einigermaßen gedämpft und sinken – dem Bild nicht mehr entnehmbar – mit steigendem n.

An der zeitlichen Selektion der nP1 erkennt man, stellvertretend auch für PGn und Bn, daß mit steigendem n der Zeitpunkt der maximalen Fensteröffnung nach rechts wandert. In seiner Umgebung verlaufen die Kurven zunehmend symmetrisch. Sie werden dabei auf hohem Niveau etwas breiter, auf tiefem dagegen schmaler. Selbst wenn man darin eine Verschlechterung der zeitlichen Selektivität erkennen möchte, so ist die Verbesserung der spektralen Selektivität doch augenfälliger. Weil man später bei Entnormierung die eine gegen die andere tauschen kann, macht es Sinn, insgesamt von einer verbesserten spektral/zeitlichen Selektivität zu sprechen.

Die Verläufe der Gruppenlaufzeit $T_g(\Omega)$ in Bild 3.6d verdeutlichen, daß das Laufzeitniveau innerhalb einer Familie zusammen mit n wächst. Bei gegebenem n verringert es sich von den Pn über die Bn, PGn bis zu den nP1. Hinsichtlich der Schwankung im Hauptselektionsbereich $0 < |\Omega| \leq 1$ schneidet die Familie Bn am besten ab, die auf ebene Laufzeit optimiert ist, gefolgt von PGn und nP1. Bei diesen drei Familien sinkt die Schwankung überdies mit zunehmendem n. Je geringer sie ausfällt, desto mehr stimmt $T_g(0)$ mit dem Zeitpunkt der maximalen Fensteröffnung aus Bild 3.6c überein. Aus dem Rahmen fällt wiederum die Familie Pn, für die man in der Umgebung des steilen Selektionsabfalls ein ausgeprägtes Laufzeitmaximum hinnehmen muß.

Will man die ersten beiden Eignungskriterien aus Abschnitt 3.3.2 nicht verletzen, dann bietet sich nur ein kleiner Variationsspielraum innerhalb eines Grades $n > 1$ an. Es ist möglich, die spektrale Selektivität durch den Übergang von der Familie der nP1 auf die PGn oder gar die Bn ein wenig zu verbessern, indem man schwache Nebenmaxima in der zeitlichen Selektion toleriert. Die Familie der Pn erweist sich bereits als ungeeignet. Den

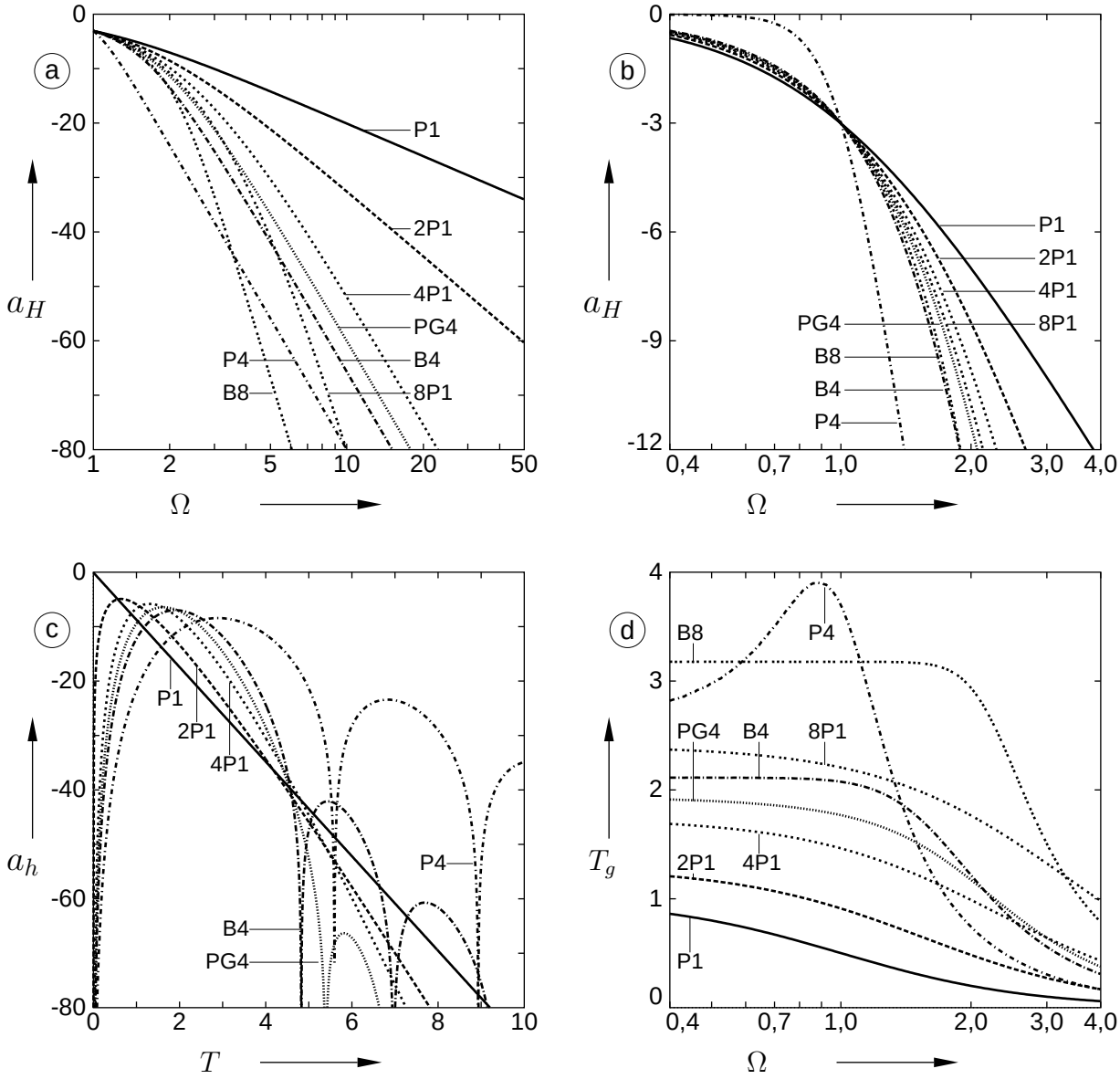

Bild 3.6: *Spektrale und zeitliche Eigenschaften verschiedener Fenstertypen: Spektrales Selektionsmaß $a_H(\Omega)$, entsprechend dem negativem Dämpfungsverlauf des zugehörigen Tiefpasses, a) im Weitabbereich und b) im Nahbereich. c) Zeitliches Selektionsmaß $a_h(T)$, entsprechend dem Pegelverlauf der Impulsantwort des zugehörigen Tiefpasses. d) Normierte Gruppenlaufzeit $T_g(\Omega)$. Normierte Zeiten T, T_g und Kreisfrequenzen Ω auf Basis der 3dB-Grenzkreisfrequenz.*

maßgeblichen Spielraum stellt im wesentlichen der Grad dar, dessen Erhöhung die spektral/zeitliche Selektivität verbessert. Er kann nicht beliebig erhöht werden, weil mit ihm auch das Laufzeitniveau steigt. Würde er über alle Grenzen wachsen, so gingen nP1, PGn und Bn in ein Gauß-Fenster mit unendlicher Grundlaufzeit über [Zer67]. Dieser Übergang spielt sich zunehmend auf tiefem Niveau beider Selektionen ab, auf hohem Niveau erfolgt die Annäherung sehr schnell. Das Gauß-Fenster weist nach [Jon89] Optimalitätseigenschaften für die Kurzzeitspektralanalyse auf.

Zusammengefaßt läßt sich der wesentliche Variationsspielraum für reelle Fensterfunktionen mit der Familie nP1 beschreiben. Für sie gelten mit einem n-fachem Pol bei $P = \alpha$

die untenstehenden, einfachen Formeln. Ein höherer Grad n bedeutet insgesamt bessere spektral/zeitliche Selektivität, aber auch ein höheres Laufzeitniveau. Familien, die vorrangig die spektrale Selektivität verbessern, bedingen ebenfalls höhere Laufzeiten und bald unzulässig hohe Nebenmaxima in der zeitlichen Selektion. Es wird angenommen, daß eine vorrangig verbesserte zeitliche Selektivität automatisch unzulässige Nebenmaxima in der spektralen Selektion bedeuten. Weiterhin wird angenommen, daß diese Gesetzmäßigkeiten für die Menge aller denkbaren kausalen Fensterfunktionen gelten.

$$H^N(P) = \left(\frac{-\alpha}{P-\alpha}\right)^n, \quad h^N(T) = \frac{-\alpha}{(n-1)!}(-\alpha T)^{n-1}e^{\alpha T}, \quad \text{mit} \quad \alpha = -\frac{1}{\sqrt{2^{\frac{1}{n}}-1}}. \tag{3.8}$$

3.3.4 Laufzeitausgleich

Bei Fensterfunktionsgraden $n > 1$ liegt die maximalen Fensteröffnung, anders als bei der bisherigen Fensterfunktion P1, nicht mehr am Zeitpunkt null. Weil das Fenster nach Entnormierung zu höheren Analysefrequenzen kürzer ausfällt, erweist sich dieser Zeitpunkt als frequenzabhängig. Da die resultierende Fensterschar die Impulsantworten der FTT-Analysetiefpässe repräsentiert, erreicht das FTT-Pegelspektrum eines Dirac-Impulses sein Maximum zuerst an höheren Frequenzen. Die zugeordneten Zeitkonturpunkte treten also nicht mehr gleichzeitig auf, die Zeitkonturlinie verläuft 'schräg'. Durch einen Laufzeitausgleich kann man erreichen, daß die Zeitpunkte der maximalen Fensteröffnung frequenzunabhängig übereinander liegen und somit die Zeitkonturlinie des Dirac-Impulses exakt frequenzparallel verläuft.

Zwar reflektiert ein Laufzeitunterschied ohne Ausgleich durchaus das Verhalten der Wanderwelle auf der Basilarmembran [Zwi82]. Die weitere Informationsverarbeitung des Gehörs ist aber daran angepaßt: Ein Impuls wird als Klick gehört, seine höheren Spektralanteile werden nicht als voreilend empfunden. Ein Laufzeitausgleich stellt die gewohnte Gleichzeitigkeit auch formal dar. Würde man ihn übrigens hier nicht tolerieren, so müßte er spätestens auf der Rekonstruktionsseite zugeschlagen werden. Andernfalls würden in der Wahrnehmung des Rekonstruktionsignals tatsächlich höhere Spektralanteile voreilen.

Die Formel

$$\max\{h^N(T)\} = h^N(T_{max}) = h^N_{max} \tag{3.9}$$

definiert den Scheitelpunkt (h^N_{max}, T_{max}) der normierten Fensterfunktion $h^N(T)$. Zur Zeit T_{max} wird die maximale Fensteröffnung h^N_{max} erreicht. Für die untersuchten Fensterfunktionen finden sich die zugehörigen Werte in Tabelle B.1 im Anhang, die Höhe h^N_{max} wird später bei der Signalrekonstruktion benötigt.

Bild 3.7 zeigt nun im Detail, wie die entnormierten Zeitpunkte t_{max,ω_A} der maximalen Fensteröffnung durch eine zusätzliche Verzögerung t_{L,ω_A} über alle Analysefrequenzen ω_A zur Deckung gebracht werden. Da t_{max,ω_A} wegen des monotonen Ansteigens der Frequenzgruppenbreite über der Analysefrequenz monoton sinkt, wird ohne Laufzeitausgleich die maximale Fensteröffnung an der tiefstmöglichen Analysefrequenz $\omega_A = 0$ zuletzt erreicht. Die notwendige zusätzliche Verzögerung t_{L,ω_A} beginnt also hier mit null und steigt zu höheren ω_A an:

$$t_{L,\omega_A} \quad = \quad t_{max,0} - t_{max,\omega_A} \tag{3.10}$$

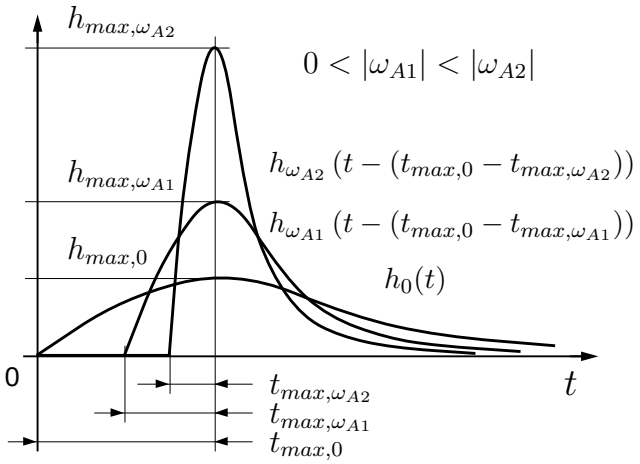

Bild 3.7: *Laufzeitausgleich für kausale Fensterfunktionen $h_{\omega_A}(t)$ am Beispiel der Analysefrequenzen ω_{A_1} und ω_{A_2}: Durch individuelle zeitliche Verzögerung $t_{max,0} - t_{max,\omega_A}$ werden die Zeitpunkte der maximalen Fensteröffnung mit demjenigen bei der Analysefrequenz 0 zur Deckung gebracht. Für die übliche Bezeichnung als 'Fenster' ist die (verzögerte) Fensterfunktion an der Achse $t = 0$ zu spiegeln.*

$$= \frac{T_{max}}{\pi}\left(\frac{1}{B_{3dB}(0)} - \frac{1}{B_{3dB}(\omega_A)}\right). \tag{3.11}$$

Für die Umformung wurde die Zeitentnormierung aus Gl. (3.3) verwendet, womit die Analysebandbreite B_{3dB} ins Spiel kommt. So kann an jeder Analysefrequenz ein Laufzeitglied mit der Impulsantwort

$$l_{\omega_A}(t) = \delta(t - t_{L,\omega_A}) \tag{3.12}$$

eingerichtet werden, durch dessen Filterung sich ein *laufzeitausgeglichenes FTT-Spektrum*

$$s^L_{\omega_A}(t) = s_{\omega_A}(t) * l_{\omega_A}(t) \tag{3.13}$$

$$= \left(s(t)\cdot e^{-j\omega_A t}\right) * h_{\omega_A}(t) * l_{\omega_A}(t) \tag{3.14}$$

ergibt. Die zweite Formel folgt aus Gl. (1.6). Als Folge von Gl. (3.11) wächst der auszugleichende Laufzeitunterschied, wenn T_{max} bei höherem Grad n zunimmt oder wenn B_{3dB} (an allen Analysefrequenzen um den gleichen Faktor) reduziert wird. Die Zeit $t_{max,0}$ kann als *Grundlaufzeit* der laufzeitausgeglichenen FTT bezeichnet werden. Besonderheiten der zeitdiskreten Realisierung des Laufzeitgliedes behandelt Anhang B.4.

3.3.5 Konturausbildung in Abhängigkeit vom Fensterfunktionsgrad

In der Familie nP1, welche die im wesentlichen geeigneten Fensterfunktionen repräsentiert, bleibt der Fensterfunktionsgrad n noch festzulegen. Welche Bedeutung er für die Anwendbarkeit des Konturierungskonzeptes hat, kann man an den bisher festgestellten Zusammenhängen nicht ablesen. Vermutlich ist die bei höherem n bessere spektral/zeitliche Selektivität geeignet, um im TTZM-Verfahren Hüllkurvenglättung und Simultanverdeckung zu verringern. Es könnten aber auch neue, unerwünschte Konturierungseffekte hinzutreten, zumal nun Zeitkonturen ins Spiel kommen. Dies wird am Beispiel des Einschaltens eines 1kHz-Tons untersucht und auf Tonsignale mit Hüllkurvenänderung übertragen. Dazu wird an die Ergebnisse von Abschnitt 3.2.2 angeknüpft.

Ein Blick auf Bild 3.3, S. 66, vergegenwärtigt nochmals den Zusammenhang von FTT-Pegelspektrum und Konturen beim übergangslosen Ein- und Ausschalten des Tons ($B^{gs}_{3dB} \rightarrow \infty$). Dem Spektrum dort liegt ein Grad $n = 4$ mit Laufzeitausgleich zugrunde, die von Heinbach vorgesehene zeitliche Glättung des Leistungsspektrums (Abschnitt 1.5.1) fehlt. Wesentlich ist erstens, daß zu den Schaltzeitpunkten Verbreiterungen im Spektrum auftreten, die sich deutlich vom stationären Spektralverlauf abheben und deshalb zu Zeitkonturen führen. Zweitens erscheinen beim Einschalten fein oszillierende Welligkeiten im Verschneidungsgebiet der Verbreiterung mit dem stationärem Spektrum, welche zahlreiche Nebenkonturlinien auslösen. Mit diesen beiden Feststellungen soll es nachfolgend ausreichen, das zeitvariante Pegelspektrum zu behandeln, um indirekt auf die Konturausbildung zu schließen.

3.3.5.1 Experimentelle Beobachtungen im Einschaltspektrum

Bild 3.8 zeigt das FTT-Pegelspektrum nur beim Einschalten und verdeutlicht die Entwicklung für $n = 1, 2, 4, 8$. Die Analysebandbreiten sind überall gleich, auf zeitliche Glättung wurde verzichtet. Nur bei $n = 4, 8$ wurde der Laufzeitausgleich angewandt. Während sich das Spektrum in der unmittelbaren Umgebung des Hauptmaximums gleichartig entwickelt, gibt es deutliche Unterschiede in den Flanken des eingeschwungenen Spektrums. Es wird gut sichtbar, daß ein höherer Grad eine schmalere spektrale Selektion und damit einen steileren Abfall des stationären Spektrums bedeutet. Das 'Schaukeln' der Flanken mit der doppelten Tonfrequenz ([Ter85] und Abschnitt 1.4.2) ist deshalb ab $n = 4$ nicht mehr erkennbar. Weiterhin unterscheiden sich die Ausprägung von Verbreiterung und Welligkeiten erheblich.

Der Übergang von $n = 4$ auf $n = 8$ verkleinert die Ausdehnung des Welligkeitsgebietes. Ein höherer Grad bedeutet, daß sich die Welligkeiten mehr in die Verschneidung zwischen stationärem Spektrum und transienter Verbreiterung zurückziehen. Die Verschneidung selbst prägt sich dabei deutlicher als 'konkave' Kante aus. Ein niedrigerer Grad von $n = 2$ führt zu der umgekehrten Tendenz. Die Welligkeiten beginnen sogar bis auf die Schultern der Verbreiterung hinaufzureichen. Beim Übergang von $n = 2$ auf $n = 1$ wird die Ausdehnung der Welligkeitsgebiete anscheinend wieder kleiner. Allerdings existiert bei $n = 1$ auch keine Verbreiterung mehr. Dafür sind die Welligkeiten sofort nach dem Einschalten so groß, daß ausgeprägte Nebenmaxima auf relativ hohem Pegelniveau und in größerer spektraler Entfernung vom Hauptmaximum entstehen [Ter85, Fig. 4].

3.3.5.2 Analytische Behandlung des Einschaltspektrums

Eine analytische Behandlung der Einschalteffekte im FTT-Spektrum präzisiert und ergänzt die obigen Beobachtungen. Vereinfachend wird der eben verwendete Sinuston durch eine bei $t = 0$ eingeschaltete komplexe Schwingung

$$s(t) = \begin{cases} \frac{A}{2} \cdot e^{j\omega_T t} & \text{für} \quad t \geq 0, \\ 0 & \text{sonst}, \end{cases} \tag{3.15}$$

mit Amplitude $\frac{A}{2}$ und Frequenz ω_T ersetzt. Bei diesem Signal bleibt der wesentliche Einschalteffekt wirksam, dafür entfallen Nebeneffekte durch Selektion der negativen Seite

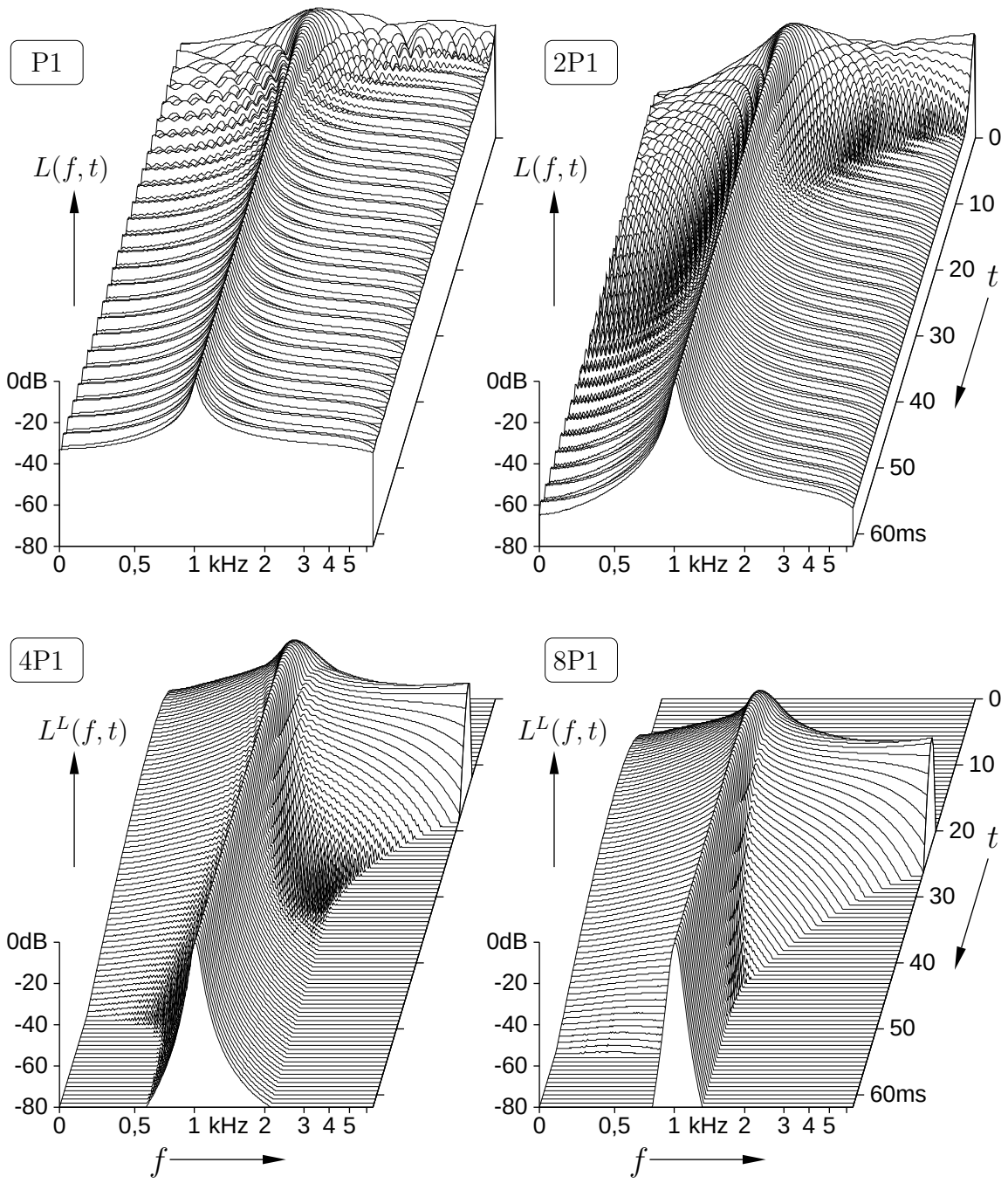

Bild 3.8: *FTT-Pegelspektrum eines bei $t = 0$ eingeschalteten Sinustons $f_T = 1$ kHz, dargestellt für Fensterfunktionen* n*P1 bei einer Analysebandbreite $B_{3dB} = 0{,}3$ Bark. Bei* 4P1 *und* 8P1 *Laufzeitausgleich. Mit zunehmendem Fensterfunktionsgrad n hebt sich der transiente Anteil immer besser ab. Die feinen Welligkeiten ziehen sich in seine Verschneidung mit dem stationären Anteil zurück. Für $n \geq 4$ ist deshalb zur Konturierung keine Glättung mehr notwendig.*

des Fourier-Spektrums. Dazu zählen das 'Schaukeln' und der Einfluß der Startphase der reellen Sinusschwingung. [7] Für den Zeitverlauf $s_{\omega_A}(t)$ des FTT-Spektrums von Gl. (3.15)

[7] Wegen der Linearitätseigenschaft der (komplexen) FTT kann man die Wirkung der Negativseite getrennt berechnen und überlagern. Für eine reelle Schwingung mit Startphase ϕ_0 setzt man die lineare Zerlegung nach Gln. (1.12), (1.13) an. Der Einfluß auf das FTT-Pegelspektrums durch die jeweils andere

stellt Anhang C.1 folgende Näherung bereit:

$$s_{\omega_A}(t) \approx \Xi + \hat{\Theta}, \qquad \text{wobei} \tag{3.16}$$

$$\Xi = H_{\omega_A}(\Delta\omega) \cdot \frac{A}{2} \cdot e^{j\Delta\omega t}, \tag{3.17}$$

$$\hat{\Theta} = -\frac{A}{2} \cdot \frac{1}{j\Delta\omega + \omega_{3dB}} \cdot h_{\omega_A}(t). \tag{3.18}$$

Es bezeichnen Ξ den exakten stationären und $\hat{\Theta}$ den genäherten transienten Anteil sowie $\omega_{3dB}(\omega_A)$ die Grenzfrequenz des Analysetiefpasses und $\Delta\omega = \omega_T - \omega_A$ den Abstand von Analysefrequenz und Tonfrequenz. Der Näherungsfehler bleibt im Rahmen der zu betrachtenden spektral/zeitlichen Gebiete und Fensterfunktionsgrade n unbedeutend.

Den Zeitverlauf $s_{\omega_A}(t)$ von Gl. (3.16) kann man sich vorstellen als Summe eines stationären komplexen Drehzeigers mit Länge $|\Xi|$ und einem nicht drehenden Zeiger mit zeitlich abnehmender Länge $|\hat{\Theta}|$. Unterschreitet $|\hat{\Theta}|$ die untere Dynamikschwelle, dann wird beim Übergang auf das Pegelspektrum nur noch die zeitunabhängige Länge $|\Xi| = |H_{\omega_A}(\Delta\omega) \cdot \frac{A}{2}|$ des Drehzeigers Ξ erfaßt. So entsteht der in Bild 3.8c,d klar erkennbare, eingeschwungene Teil der Pegelfläche, dessen Form allein durch die Fourier-Transformierte $H_{\omega_A}(\omega)$ der Fensterfunktion vorgegeben ist.

In der Fourier-Transformierten $H_{\omega_A}(\omega)$ einer Fensterfunktion n-ten Grades steht die Frequenz ω in der n-ten Potenz im Nenner und fehlt im Zähler (vgl. $H^N(P)$ Gl. (3.8)). Vernachlässigt man Zusatzeffekte infolge der ω_A–Abhängigkeit von ω_{3dB}, dann fällt die Länge $|\Xi|$ in Abhängigkeit vom Abstand $|\Delta\omega| \gg \omega_{3dB}$ mit der n-ten Potenz, die Länge $|\hat{\Theta}|$ aber nur mit der ersten Potenz. Deshalb wird bei $n > 1$ und ausreichendem $|\Delta\omega|$ nur noch die Länge $|\hat{\Theta}|$ des transienten Zeigers signifikant sein. Liegt der zugehörige Pegel über der unteren Dynamikschwelle, dann resultiert daraus die in Bild 3.8c,d deutlich ausgeprägte spektrale Verbreiterung. Die Form ihres Zeitverlaufes entspricht der Fensterfunktion $h_{\omega_A}(t)$.

Die in Bild 3.8a-d beobachteten Welligkeitsgebiete sind diejenigen, in denen $|\Xi|$ und $|\hat{\Theta}|$ die gleiche Größenordnung aufweisen. Die Länge $|s_{\omega_A}(t)|$ des Summenzeigers schwankt dann mit der Frequenz $\Delta\omega$ über der Zeit. Die Welligkeitsgebiete werden um so kleiner, je schneller sich bei Variation von t oder $\Delta\omega$ der Wechsel von dominanten $|\hat{\Theta}|$ zu dominanten $|\Xi|$ – oder umgekehrt – vollzieht. Dazu muß einerseits ein steiler Abfall von $H_{\omega_A}(\omega)$ für $|\omega| > 0$ wie auch von $h_{\omega_A}(t)$ für $t > t_{max,\omega_A}$ vorliegen. Man kann diese Forderungen entsprechend über die Selektionmaße nach Gln. (3.6) formulieren. Bei steigendem n werden sie gemäß Bild 3.6 besser erfüllt.

Bei $n = 1$, wo Gl. (3.16) übrigens exakt gilt und sich beide Zeiger anfangs genau aufheben, gibt es eine Besonderheit. Weil $\Delta\omega$ in der ersten Potenz im Nenner beider Zeigerlängen steht, gibt es keine spektrale Verbreiterung. Außerdem erstreckt sich die Welligkeit sofort über die gesamte Breite der Frequenzachse.

Seite ist minimal für Analysefrequenzen nahe $\pm\omega_T$ und maximal bei 0 und ∞.

3.3.5.3 Ausschalten und Übergang auf Tonsignal mit Hüllkurvenänderung

Warum sind in Bild 3.3 auf S. 66 keine Welligkeitsgebiete beim Ausschalten zu beobachten? [8] Ein Ausschalten des Signals aus Gl. (3.15) wird dadurch angesetzt, daß eine gleichartige Schwingung an den Ausschaltzeitpunkt geschoben und mit um π versetzter Phase überlagert wird. Wegen der Linearitätseigenschaften der FTT sind auch im Spektrum zusätzlich je ein transienter und ein stationärer Anteil wirksam. Unmittelbar nach dem Ausschaltzeitpunkt heben sich ursprünglicher und neuer stationärer Anteil vollständig auf. Sofern der transiente Anteil des Einschaltens bereits abgeklungen ist, bleibt nur der transiente Anteil des Ausschaltens übrig. Dieser kann allein keine Welligkeiten hervorrufen. Weil er die gleiche, wenn auch zeitverschobene Form wie beim Einschaltens aufweist, ruft er bei $n > 1$ eine spektrale Verbreiterung hervor, die im wesentlichen mit der beim Einschalten übereinstimmt. Im Bereich der Verbreiterungen von Ein- und Ausschalten in Bild 3.3 sind deshalb die Zeitkonturhauptäste gleich.

Weil es überhaupt erst bei $n > 1$ eine spektrale Verbreiterung gibt, folgt ein weiterer Unterschied gegenüber $n = 1$. Bild 2.3 auf S. 28 zeigt, daß es bei $n = 1$ keinen besonderen Ausschalteffekt gibt, vielmehr sinkt ein beliebiges Spektrum nach dem Ausschalten in sich zusammen. Weil die Höhe der lokalen Pegelmaxima relativ zur Nachbarschaft bestehen bleibt, entstehen Abklingfahnen von Frequenzkonturen, die im Sprachsignal in Bild 1.2 oben auf S. 13 sehr auffällig sind. Bei $n > 1$ dagegen bewirkt die spektrale Verbreiterung einen 'Überflutungseffekt'. Da sie sich nämlich in der Umgebung starker Pegelmaxima über benachbarte, schwächere Pegelmaxima erheben kann, fehlen viele Abklingfahnen. Bild 3.4 oben auf S. 68 verdeutlicht den Unterschied.

Ein- und Ausschalten wurden in Abschnitt 3.2.2 als Extremfall für die Hüllkurvenänderung eines Tons abgehandelt. Aber auch bei 'langsameren' Änderungen der Hüllkurve konnten dort die Phänomene spektrale Verbreiterung und Welligkeit beobachtet werden. Es gibt dort genauso eine Überlagerung von einem transienten und einem stationären FTT-Spektrum. Der transiente Teil fällt allerdings abseits von der Tonfrequenz schneller ab. Verallgemeinernd kann man für Signale mit Hüllkurvenschwankung folgendes sagen: Verbreiterungen – und damit auch Zeitkonturen – können sich besser abzeichnen, wenn ein höherer Fensterfunktionsgrad n ein stationäres FTT-Spektrum gewährleistet, das steiler abfällt. Ebenso verkleinert sich die Ausdehnung möglicher Welligkeitsgebiete.

3.3.5.4 Zusammenfassung

Je höher der Grad n der Fensterfunktion gewählt wird, desto besser heben sich transiente und stationäre Anteile im FTT-Spektrum voneinander ab, die hier anhand von Tonsignalen mit Hüllkurvenänderungen untersucht wurden. Dadurch lassen sich beispielsweise langsamere und schnellere Hüllkurvenänderungen leichter anhand von Hauptkonturlinien voneinander unterscheiden. Weiterhin reduzieren sich mögliche Welligkeiten im Übergang zwischen beiden Anteilen. Gebiete in den Zeit- und Frequenzkonturen, die infolge der Welligkeiten dicht mit Nebenlinien angefüllt sind, verkleinern sich und weisen einen nied-

[8] Die Erklärung gilt aufgrund eines beschränkten Gültigkeitsbereiches von Gl. (3.16) erst unmittelbar *nach* dem Maximum der spektralen Verbreiterung. Vorher können Welligkeitsgebiete spiegelbildlich zu denen beim Einschalten auftreten, wenn mit deutlich höherem Grad n die zeitliche Selektion symmetrischer wird.

rigeren mittleren Pegel auf. Somit verringern sich Nebenkonturierungseffekte zu Gunsten ausgeprägterer Hauptkonturlinien. Speziell $n = 1$ ist für die Zeitkonturierung ungeeignet, weil sich bei Hüllkurvenänderungen von Tonsignalen transiente Anteile nicht richtig abzeichnen. Insgesamt kann man schließen, daß sich ein höherer Fensterfunktionsgrad besser für eine Signalrepräsentation mit Frequenzkonturen *und* Zeitkonturen eignet.

3.3.6 Bedeutung von Glättung und Ausgeprägtheitsschwellen

Die zeitliche Glättung des FTT-Leistungsspektrums und die spektrale Ausgeprägtheitsschwelle (Abschnitte 1.5.1 und 1.5.2) dienten im Heinbachschen TTZM-Verfahren gemeinsam dazu, störende Nebenlinien bei Frequenzkonturierung zu verhindern. Außerhalb dieser Verbundfunktionalität war die Glättung auch als Einzelfunktionalität erwünscht, um psychoakustische Befunde nachzubilden. Nachfolgend wird gezeigt, daß die Verbundfunktionalität bei höheren Fensterfunktionsgraden entbehrlich ist und daß die Glättung als Einzelfunktionalität mit dem Konturierungskonzept unvereinbar ist. Dagegen bleiben die spektrale und die zeitliche (Abschnitt 3.1.2.2) Ausgeprägtheitsschwelle als Einzelfunktionalität unverzichtbar, um den Konturierungsvorgängen eine realistische Empfindlichkeit zuzuweisen: Auch das Gehör kann als endlich empfindlicher physikalischer Analysator nicht beliebig schwach ausgeprägte Signalcharakteristika auflösen, ob in spektraler oder in zeitlicher Richtung.

3.3.6.1 Entbehrlichkeit der Verbundfunktionalität

Bild 3.8 zeigte für die Fensterfunktion P1, daß die Welligkeitsgebiete des Einschaltvorganges über die gesamte spektrale Breite reichen und auf hohem Pegelniveau liegen. Die resultierenden Nebenkonturlinien würden im rekonstruierten Signal als störend empfunden werden. Heinbach ermittelte deshalb das FTT-Pegelspektrum aus einer zeitlich geglätteten Version des FTT-Leistungsspektrums. Hier sind die Welligkeiten soweit geglättet, daß sich keine ausreichend ausgeprägten Nebenmaxima mehr ausbilden können. Dazu wurden Glättungszeitkonstante T_G und Ausgeprägtheitsschwelle ΔL_A genau aufeinander abgestimmt [Hei88a].

Mit steigendem Fensterfunktionsgrad n sinkt die Ausdehnung der Welligkeitsgebiete rasch, ebenso ihr durchschnittliches Pegelniveau. Im rekonstruierten Signal vermindert sich die Störenergie entsprechend. Außerdem ziehen sich die Welligkeiten immer mehr in die Verschneidung zwischen transientem und stationärem Teil der Pegelfläche zurück. Im rekonstruierten Signal liegt deshalb die Energie der Nebenkonturlinien zunehmend im Bereich psychoakustischer Simultan- und Nachverdeckung, der von den Hauptkonturlinien der Frequenz- beziehungsweise Zeitkontur verursacht wird. Insgesamt reduziert sich also die wahrnehmbare Störwirkung bei höherem Fensterfunktionsgrad. Bevor sie bei hohem Grad völlig verschwindet, kann sie sogar noch einen gewissen Nutzeffekt als Zeitkonturersatz beisteuern ('Rauschstoß', Abschnitt 5.1.7.2). Die Verbundfunktionalität wird aus Sicht der wahrnehmbaren Verarbeitungsqualität nur bis etwa $n = 2$ benötigt.

3.3.6.2 Problematik der Glättung als Einzelfunktionalität

Heinbach begründete die Glättung als Einzelfunktionalität bei Analysefrequenzen ab 3 kHz damit, daß die Rauhigkeit eine Wahrnehmungsgrenze aufweist [Ter68a]. Darin spiegelt sich das Verhalten der neuronalen Verarbeitung wider, welche zeitlichen Schwankungen der Basilarmembranauslenkung nur bis zu einem gewissen Grad folgen kann [Ter68b, Ter74]. Konturierungskonzept und zeitliche Glättung spielen hier aber nicht zusammen, wenn nichttonale Signalanteile zu repräsentieren sind.

Zwar bewirkt die Glättung durchaus, daß das Leistungsspektrum im Sinne der Wahrnehmungsgrenze zeitlich nicht schneller schwanken kann. Bei Hüllkurvenschwankungen von ausgeprägten Tönen schlägt sich dieser Glättungseffekt auch in den Frequenzkonturen nieder, vergleichbar mit der Hüllkurvenglättung durch die Analysebandbreite nach Abschnitt 2.2. Bei nichttonalen Anteilen aber wandeln beide Konturierungsvorgänge den Glättungseffekt in einen 'Abschnüreffekt', wie gleich an Extrembeispielen erläutert wird. Das bedeutet, daß zu repräsentierende Leistung ab einer gewissen zeitlichen Schwankungsfrequenz wegfällt, nicht aber zeitlich gemittelt beibehalten wird. Weil nur nichttonale Anteile betroffen sind, wirkt die Glättung dadurch tendenziell tonalisierend, ähnlich Abschnitt 2.5.

Am besten zeigt sich dies bei der Zeitkonturierung der Impulsfolge, deren Ergebnis in Bild 3.2 unten auf S. 63 zu betrachten war. Die zeitliche Schwankung des Pegelspektrums an einer höheren Analysefrequenz löst Zeitkonturpunkte aus, die sich über der Frequenz zu Zeitkonturlinien formieren. Eine zeitliche Glättung des Leistungsspektrums bewirkt zunächst, daß sich die zeitlichen Pegelmaxima nicht mehr so hoch, die Pegelminima sich nicht mehr so tief ausprägen. Wenn bei hohen Analysefrequenzen und Impulsraten die Glättungswirkung voll zum Tragen kommt, können die Schwankungen so weit eingeebnet werden, daß keine Zeitkonturen mehr die Ausgeprägtheitsschwelle λ passieren. Dadurch würden Signalanteile unrepräsentiert bleiben, womit der besagte Abschnüreffekt eintritt.

Bei der Frequenzkonturierung kann man am Beispiel von hochfrequenten Rauschanteilen ein vergleichbares Problem aufzeigen. Wegen der frequenzabhängigen Analysebandbreite wird die zeitliche Fluktuation des Leistungsspektrums zu höheren Frequenzen immer schneller, indirekt ablesbar an Bild 2.9 links auf S. 44. An hohen Frequenzen kann die zeitliche Glättung das Leistungsspektrum stark mitteln, so daß die Fluktuation in einem gewissen Analysefrequenzbereich auf einen gleichen, mittlere Pegel 'einfriert'. Die Frequenzkonturierung findet dann keine ausgeprägten Maxima mehr, so daß auch hier ein Abschnüreffekt zu beobachten ist. Im Heinbachschen TTZM-Verfahren spielt er allerdings noch keine Rolle: Die niedrige Analysebandbreite $B_{3dB} = 0{,}1$ Bark schränkt die Fluktuation von vornherein stark ein, so daß die Glättung – im Sinne der Wahrnehmungsgrenze der Rauhigkeit – noch gar nicht richtig greift.

Die Glättung als Einzelfunktionalität ist also im Rahmen des Konturierungskonzeptes ungeeignet, die Wahrnehmungsgrenze der Rauhigkeit zu modellieren. Eine Alternative könnte darin bestehen, daß man die Wahrnehmungsgrenze direkt in das Konturierungskonzept integriert. Für die Zeitkonturierung beispielsweise könnte man einen Übergang einer regelmäßigen in eine gesättigte statistische Zeitkonturpunktrate vorsehen, sobald eine gewisse Rate überschritten wird. Eine geeignete Modellierung der Wahrnehmungsgrenze unterstützt die Datenreduktion, wird innerhalb der vorliegenden Arbeit aber nicht weiter untersucht.

3.3.7 Zusammenfassung

Um das Zusammenspiel von Konturierung und FTT-Fensterfunktion zu untersuchen, wurde zunächst eine Methodik zur Spezifikation und zum Eigenschaftsvergleich verschiedener Fensterfunktionen entwickelt. Fensterfunktionen werden als Impulsantworten von normierten Tiefpässen eines Filterkataloges spezifiziert. Als Normierungskonstante dient die frequenzabhängige 3dB-Analysebandbreite, die nun als Freiheitsgrad entfällt. Zur Realisierung der FTT wird der Tiefpaß entnormiert, in ein zeitdiskretes System überführt und dem komplexen Modulator der gewünschten Analysefrequenz nachgeschaltet. Der logarithmierte Betrag von normierter Impulsantwort und normierter Systemfunktion definiert die zeitliche beziehungsweise die spektrale Selektion. Zusammen mit dem normierten Gruppenlaufzeitverlauf erhält man drei Beurteilungsmaße für Fensterfunktionen.

Die FTT-Fensterfunktion muß in einem Konturierungskonzept, das zu den Terhardtschen Prinzipien der Informationsverarbeitung im Gehör paßt, bestimmte Eignungskriterien erfüllen. Demnach darf die zeitliche Selektion nur geringe, die spektrale Selektion gar keine Nebenmaxima aufweisen. Andernfalls entstehen Nebenkonturlinien, die dem Prinzip der schnellen Reduktion auf das Wesentliche widersprechen. Außerdem sollte das Laufzeitniveau möglichst niedrig ausfallen, damit eine hohe Verarbeitungsgeschwindigkeit gewährleistet ist.

Anschließend wurde der Spielraum erkundet, der für reelle Fensterfunktionen übrig bleibt. Mit zunehmendem Realisierungsaufwand, gemessen am Grad n der zugehörigen Systemfunktion, läßt sich gegenüber der ursprünglichen Fensterfunktion mit $n = 1$ folgendes erreichen: Man kann die spektrale Selektion wesentlich verbessern, ohne die zeitliche Selektion merklich zu verschlechtern. Pauschal betrachtet verbessert sich dadurch die spektral/zeitliche Selektivität. Dabei ist das steigende Laufzeitniveau im Auge zu behalten. Bei gegebenem Grad kann man die spektrale Selektion nur wenig verbessern, weil in der zeitlichen Selektion bald unerwünschte Nebenmaxima auftreten. Der wesentliche Spielraum besteht in der Wahl des Grades n etwa innerhalb der Fensterfamilie nP1. Sie entspricht der Familie der Tiefpässe mit einem n-fachen reellen Pol und stellt bei $n = 1$ die ursprüngliche Fensterfunktion (P1).

Normalerweise ist bei Graden $n > 1$ der Zeitpunkt der maximalen Fensteröffnung wegen des Verlaufs der Analysebandbreite frequenzabhängig. Dadurch erscheinen Signalanteile an höheren Analysefrequenzen früher im FTT-Spektrum als an tiefen. Obwohl dies dem Verhalten der Basilarmembran entspricht, ist die weitere Informationsverarbeitung des Gehörs offenbar an diese Verhältnisse adaptiert. Überdies würde ein Voreilen hoher Frequenzen später im rekonstruierten Signal hörbar werden. Deshalb wurde ein Laufzeitausgleich eingeführt. Er verzögert den Zeitverlauf des FTT-Spektrums zu höheren Analysefrequenzen hin derart, daß sich die maximale Fensteröffnung überall gleichzeitig einstellt. Ein Dirac-Impuls löst nun an allen Frequenzen gleichzeitige statt frequenzabhängig verschobene Zeitkonturpunkte aus.

Anhand der nP1 wurde die Konturausbildung in Abhängigkeit vom Grad n untersucht, indem Hüllkurvenänderungen eines Sinustons im FTT-Spektrum beobachtet wurden. Es zeigte sich experimentell wie analytisch, daß ein höheres n eine bessere Trennung von transienten und stationären Anteilen sicherstellt. Die transienten Anteile heben sich deutlicher als kurzzeitige spektrale Verbreiterung ab, womit verschieden schnelle Hüllkurvenänderungen – die auch verschieden wahrgenommen werden – überhaupt erst durch Kontu-

ren repräsentierbar werden. Bei niedrigerem n gehen die Anteile unschärfer ineinander über, wodurch unmittelbar nach dem Einschalten zahlreiche Nebenkonturlinien auftauchen. Beim Abschalten kann sich der transiente Anteil erst ab $n > 1$ vom stationären abheben, weshalb man bei $n = 1$ keine Repräsentation eines Ausschaltknackes erreicht.

Bei $n = 1$ wären die Nebenkonturlinien beim Einschalten so ausgeprägt, daß sie im TTZM-Verfahren Störungen verursacht hätten. Um dies zu vermeiden, stimmte Heinbach zeitliche Glättung des Leistungsspektrums und spektrale Ausgeprägtheitsschwelle genau aufeinander ab. Die Notwendigkeit dieser Verbundfunktionalität entfällt jedoch bei höheren n, weil die Ausprägung von Nebenkonturlinien abnimmt. Die Glättung eignet sich darüber hinaus auch nicht als Einzelfunktionalität, um die Wahrnehmungsgrenze der Rauhigkeit nachzubilden. In diesem Sinne konnte sie bei der bisherigen Dimensionierung sowieso noch nicht eingreifen. Bei höheren Analysebandbreiten jedoch würde sie sich mit dem Konturierungskonzept nicht mehr vertragen. Dagegen bleiben spektrale wie auch zeitliche Ausgeprägtheitsschwellen als Einzelfunktionalitäten sinnvoll, um eine begrenzte Empfindlichkeit der Konturierungsvorgänge zu modellieren.

Insgesamt läßt sich so aus vorwiegend theoretischer Betrachtung schließen, daß das Konzept einer Repräsentation durch Frequenz- *und* Zeitkonturen im wesentlichen nur mit FTT-Fensterfunktionen der Familie nP1 und höherem n zusammenspielt. Dabei wird ein Laufzeitausgleich erforderlich, die bisherige Glättung fällt jedoch weg. Im Sinne einer schnellen Informationsverarbeitung sollte n nicht zu hoch sein.

3.4 Einstellung der Transformations- und Konturierungsparameter

Alle Parameter der Konturanalyse, bestehend aus modifizierter FTT mit Laufzeitausgleich sowie Frequenz- und Zeitkonturierung, werden nun festgelegt. Dabei soll der Zuhörer bei Sprache, nachdem das Signal rekonstruiert wurde, bestmögliche Verarbeitungsqualität wahrnehmen. Dieses Einstellziel hängt aber nicht nur von der Konturanalyse ab, sondern insbesondere auch von den Fähigkeiten des Rekonstruktionsverfahrens. Diese Abhängigkeiten werden als erstes behandelt. Die danach vorgestellten Einstellungen sind zwar subjektiv, weil sie allein vom Autor durchgeführt wurden. Sie lassen sich aber weitgehend objektivieren, da die Einflüsse der einzelnen Parameter auf die Verarbeitungsqualität genau erklärt werden können.

3.4.1 Bestmögliche Verarbeitungsqualität als Einstellziel

Beim Heinbachschen TTZM-Verfahren wurden Verfälschungen bestimmten Ursachen innerhalb der Spektraltransformation, des Konturierungskonzeptes und der Signalrekonstruktion zugeordnet (Abschnitt 2.7). Zur Beeinträchtigung der wahrnehmbaren Verarbeitungsqualität eines Signals kann ein jedes dieser drei Teilkonzepte beitragen. Veränderungen eines Teilkonzeptes beeinflussen aber nicht nur die ihm zugeordneten Verfälschungen, sondern auch diejenigen der anderen Teilkonzepte. Eine Änderung der Analysebandbreite beeinflußt beispielsweise die Glättung der Schmalbandhüllkurve (Abschnitt 2.2). Aber auch die Charakteristik des rekonstruktionsbedingten Störteppiches (Abschnitt 2.3) ist

betroffen, da sich der Übergangsbereich zwischen zeitlicher und spektraler Repräsentation verschiebt.

Wegen dieser Verwobenheit der Teilkonzepte kann ein einzelnes für sich nur unter Vorbehalt optimiert werden. Solange Verfälschungen durch ein anderes nicht auszuschließen sind, stellt eine günstige Parametereinstellung unter Umständen einen Kompromiß dar: Der Verfälschungseffekt im zu optimierenden Teilkonzept kann nicht völlig ausgeblendet werden, weil der in einem anderen Teilkonzept sonst zu auffällig werden würde. Überdies spielt beispielsweise der Signaltyp mit hinein, der bestimmte Verfälschungseffekte begünstigen oder unterdrücken kann. Die Suche nach einer optimalen Einstellung bleibt in Wirklichkeit auf einem Suboptimum stehen.

Um Suboptimalitäten aufzudecken und zu vermeiden, wurden in der Praxis alle Teilkonzepte gemeinsam optimiert. Zur besseren Übersicht bleibt die Einstellung von Transformation und Konturierungskonzept aber formal von der Einstellung der Rekonstruktion abgetrennt. Die Rekonstruktion kann also bis auf weiteres als fertiges Verfahren betrachtet werden. Allerdings stößt die Realisierung eines optimalen Rekonstruktionsverfahrens später auf Schwierigkeiten. Deshalb werden auch suboptimale Verfahren herangezogen. Es stehen insgesamt drei Verfahren mit unterschiedlichen Verfälschungscharakteristiken zur Verfügung. Die neuen, ersten beiden werden erst in Kapitel 5 vorgestellt, ihre genaue Funktion spielt im Moment keine Rolle:

RKOP: Rekonstruktion aus Konturen mit Original-Phasen. Verfälschungen sind weitgehend ausgeschlossen. Für die bisher definierten Konturen ohne Phasen ist dies genaugenommen kein eigenständiges Verfahren, sondern nur eine Simulation.

RKHP: Rekonstruktion aus Konturen mit heuristischer Phase. Dieses Verfahren bringt Phaseninkohärenz-bedingte Störungen ein.

TTSD: Die bereits bekannte Teiltonsynthese mit Dreieckfenster kann nur Frequenzkonturen verarbeiten und fügt Synthesefenster-kontrollierbare und Phaseninkohärenzbedingte Störungen hinzu.

Die Erkenntnisse über die Verwobenheit der Teilkonzepte und die mögliche Kompromißbildung zwischen Verfälschungseffekten kann abschließend in vier praxisrelevante Feststellungen gefaßt werden. Bei einer Parametereinstellung, die bestmögliche wahrnehmbare Verarbeitungsqualität anstrebt, ist demnach folgendes zu berücksichtigen:

- Solange Verfälschungen wahrnehmbar sind, kann die Parametereinstellung vom Signaltyp, von den Abhörbedingungen und vom Hörgeschmack des Zuhörers abhängen.
- Bei Rekonstruktionsverfahren mit unterschiedlichen Verfälschungscharakteristiken können unterschiedliche Parametereinstellungen bevorzugt werden.
- Bei unterschiedlichen Konturierungskonzepten (also Frequenzkonturierung mit oder ohne Zeitkonturierung) können Rekonstruktionsverfahren mit unterschiedlichen Verfälschungscharakteristiken bevorzugt werden.
- Solange Verfälschungen wahrnehmbar sind, kann von einer optimalen Gehöranpassung der Transformation oder des Konturierungskonzeptes nicht die Rede sein.

Tabelle 3.2: *Fünf Einstellungen für Transformations- und Konturierungsparameter, sortiert nach fallender Sprachqualität bei Signalrekonstruktion. Einstellungen* ZFKI, ZFKII *und* M-TTZM *entstanden aus umfangreichen Parametervariationen für beste subjektive Sprachqualität bei gegebenem Rekonstruktionsverfahren* RKOP, RKHP *bzw.* TTSD. *Etablierte Einstellungen* SM-TTZM *nach Schlang und Mummert [Sch90] sowie* HB-TTZM *nach Heinbach [Hei88a] zum Vergleich. Dort ist die Wahl von* TTSD *nicht authentisch, sie liefert hier aber die bestmögliche Sprachqualität.* $^+$ *siehe Abschnitt 3.4.3.*

Parameter | Rekonstruktion	**Einstellung**				
	ZFKI	ZFKII	M-TTZM	SM-TTZM	HB-TTZM
Verarbeitung Frequenzkonturen/TTZM	•	•	•	•	•
Verarbeitung Zeitkonturen	•	•	-	-	-
Grad n für Fensterfunktion nP1	4	4	4	2	1
Analysebandbreite B_{3dB}/Bark	0,5	0,3	0,3	0,25	0,1
Analysefrequenzabst. $\Delta\omega_A/(2\pi \cdot \text{Bark})$	0,05)$^+$	0,05)$^+$	0,05)$^+$	0,05)$^+$	0,05
Laufzeitausgleich Spektrum	•	•	•	-	-
Glättung Betragsspektrum	-	-	-	•	•
Auswerteintervalldauer T_A	$1/f_a)^+$	$1/f_a)^+$	1,25ms	1,25ms	1,25ms
Ausgeprägtheitsschwelle ΔL_A/dB	0,5	0,5	0,5	0,5	3
Ausgeprägtheitsschwelle $\lambda/(B_{3dB} \cdot \text{dB})$	25	25	-	-	-
Rek. mit Originalphase (RKOP)	•	-	-	-	-
Rek. mit heuristischer Phase (RKHP)	-	•	-	-	-
TT-Synthese mit Dreieckf. (TTSD)	-	-	•	•	•

3.4.2 Durchführung und Ergebnis (ZFKI, ZFKII, M-TTZM)

Mit Hilfe der Rekonstruktionsverfahren RKOP, RKHP und TTSD wurden die Parameter der Konturanalyse jeweils auf subjektiv beste Verarbeitungsqualität von Sprache eingestellt. Weil TTSD keine Zeitkonturen verarbeitet, liefert die zugehörige Einstellung im Endeffekt eine verbesserte Teiltonanalyse. Als Testsignale dienten einige etwa 2 s dauernde Sprachsignale von männlichen und weiblichen Sprechern. Sie wurden jeweils mit variierten Parametern verarbeitet. Diese Variationen wurden dann vom Autor untereinander wie auch mit ihrem unverarbeiteten Original verglichen. Zum Vergleich standen außerdem zwei etablierte Teiltonanalysen, deren Verarbeitung ebenfalls mit TTSD beobachtet wurde: Die des Heinbachschen TTZM-Verfahrens (HB-TTZM) und eine weitere, von Schlang [Sch89] eingeführte mit Parametern nach [Sch90] (SM-TTZM). Die Darbietung erfolgte über Kopfhörer bei normaler Sprachlautstärke und beschränkte sich auf einen Frequenzbereich von 20 Hz bis etwa 5,5 kHz.

Aus diesen Versuchen resultieren die Einstellungen ZFKI und ZFKII (Zeit- und Frequenzkonturen) der neuen Konturanalyse sowie die Einstellung M-TTZM einer verbesserten Teiltonanalyse. Zusammen mit den etablierten Teiltonanalysen sind sie in Tabelle 3.2 aufgeführt. In einem vergrößerten Ausschnitt zeigt Bild 3.9 die jeweiligen Konturen beziehungsweise Teiltonzeitmuster für das bereits bekannte Sprachbeispiel. M-TTZM ist nicht eigens dargestellt, es ist mit den Frequenzkonturen von ZFKII identisch.

Die beste Qualität überhaupt läßt sich bei der Einstellung ZFKI mit dem nahezu verfälschungsfreien Rekonstruktionsverfahren RKOP erzielen. Nur bei sehr konzentriertem Paarvergleich sind noch minimale Unterschiede zum Original wahrzunehmen. Dies gilt auch für alle bisher verwendeten synthetischen sowie für andere nichtsprachliche Signale (bis auf das FM-Signal, siehe Abschnitt 5.1.6). Die Realisierungsschwierigkeiten einer eigenständigen optimalen Rekonstruktion bewirken leider, daß diese Einstellung in der Praxis nur dann nutzbar ist, wenn man die Konturen in noch zu diskutierender Form um die Phaseninformation des FTT-Spektrums ergänzt.

Das realisierbare, suboptimale Rekonstruktionsverfahren RKHP beschränkt die erreichbare Qualität durch Phaseninkohärenz-bedingte Störungen leider merklich. Die Verfälschungscharakteristik von Sprache erscheint hier bei der Einstellung ZFKII mit niedriger Analysebandbreite wesentlich ausgewogener als bei ZFKI. Ansonsten stimmen beide Einstellungen überein. ZFKII in Kombination mit RKOP erweist sich dagegen als ungünstig: Auf einem insgesamt höheren Qualitätsniveau deutet sich die niedrigere Analysebandbreite nämlich als Glättung zeitlicher Strukturen im Sinne von Abschnitt 2.2 an. Im Vergleich der Darstellungen in Bild 3.9 bewirkt die niedrigere Analysebandbreite von ZFKII, daß Frequenzkonturen zahlreicher und weniger unterbrochen als bei ZFKI auftreten. Für die Zeitkonturen verhält es sich umgekehrt.

Bei Verzicht auf Zeitkonturen erreicht man mit der Einstellung M-TTZM in Kombination mit TTSD die beste Verarbeitungsqualität. SM-TTZM/TTSD fällt demgegenüber etwas in der Wiedergabe zeitlicher Strukturen ab, bedingt durch den fehlenden Laufzeitausgleich, die Glättung und die etwas niedrigere Analysebandbreite. In der Darstellung gibt es keinen auffälligen Unterschied zwischen den Frequenzkonturen, weshalb auch M-TTZM nicht eigens abgebildet wurde.

Beschränkt man sich auf die Fensterfunktion P1, so bestätigt sich HB-TTZM in Verbindung mit TTSD als optimal. Die Qualität fällt jedoch deutlich ab. Der Vergleich in Bild 3.9 mit SM-TTZM zeigt, daß erstens die zeitliche Strukturierung viel geringer ist. Zweitens erscheint die spektrale Strukturierung ungünstig verzerrt. Zwar werden viel näher beieinander liegende Linien erkannt, dafür sind Harmonische im ‘e’ bei etwa 1,5 kHz schon fast verdeckt. Hier offenbart sich nochmals das Dilemma der Bandbreitenwahl für die Fensterfunktion P1, wonach einer Verbesserung der zeitlichen Eigenschaften die Verschlechterung der Simultanverdeckung im Wege steht (Abschnitt 2.7).

ZKFII/RKHP und M-TTZM/TTSD führen zu einem relativ ähnlichen Höreindruck, wenn das betrachtete Sprachsignal wenig ausgeprägte impulshafte Anteile aufweist. Weil TTSD im Gegensatz zu RKHP einen ‘nützlichen’ Störteppich durch Synthesefenster-kontrollierbare Störungen beisteuert, wird das Fehlen der Zeitkonturverarbeitung etwas verschleiert (Abschnitt 5.1.7.2). Eine störungsärmere Rekonstruktion mittels RKHP nur für Frequenzkonturen ist daher bei Sprache kaum vorzuziehen. Der Störteppich kann allerdings auch einen – im Sinne eines Klirrfaktors – leicht ‘verzerrten’ Eindruck hervorrufen.

Der Nutzen der Zeitkonturverarbeitung in ZFKII/RKHP gegenüber M-TTZM/TTSD macht sich einerseits erst bei ausgeprägteren impulshaften Signalanteilen richtig bemerkbar. Dies gilt für deutliche Sprachartikulation und für viele andere nichtsprachliche Audiosignale (Anschlageffekte von Musikinstrumenten). Andererseits kann sich die Qualität in unbedeutenden Fällen durch eine leichte Rauhigkeitsstörung geringfügig verschlechtern. Beide Beobachtungen sind auf die Phaseninkohärenz-bedingten Störungen in RKHP

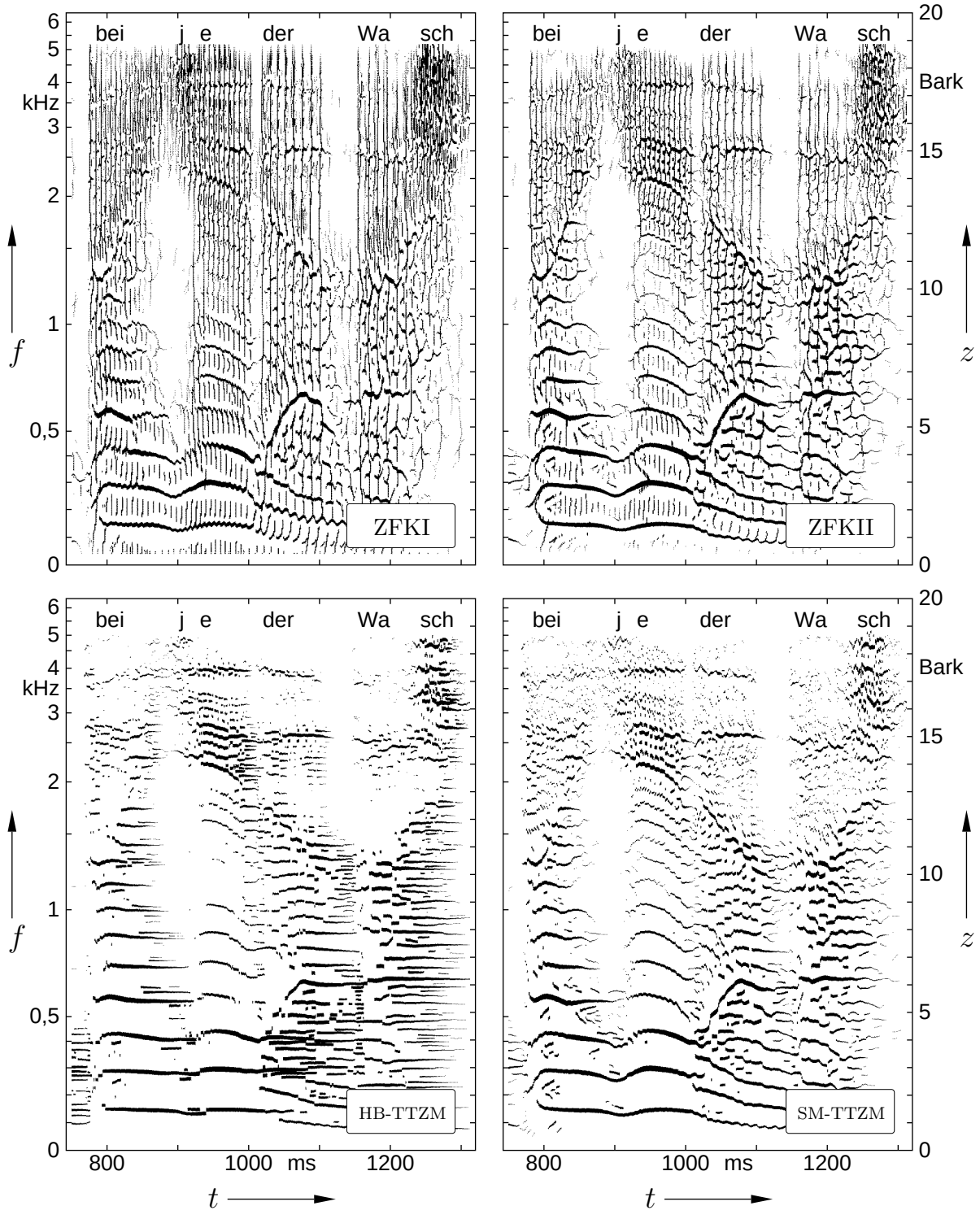

Bild 3.9: *Zeit- und Frequenzkonturen für die neuen Parametereinstellungen* ZFKI, ZFKII *(oben) und Teiltonzeitmuster alias Frequenzkonturen für frühere Einstellungen* HB-TTZM *und* SM-TTZM *(unten) am Beispiel eines Sprachausschnittes. Definition der Einstellungen siehe Tabelle 3.2. Die nicht abgebildete neue Einstellung* M-TTZM *stimmt mit den Frequenzkonturen in* ZFKII *überein und findet sich explizit auch in Bild 4.2a auf S. 104.*

zurückzuführen, die eine volle Integration der Zeitkonturen in den Rekonstruktionsprozeß verhindern (Abschnitt 5.1.5.3). Bei ZFKI/RKOP dagegen bedeutet ein Weglassen der Zeitkonturverarbeitung normalerweise schon eine deutliche Verschlechterung.

Zusammengefaßt verbessert der Übergang von HB-TTZM/TTSD über SM-TTZM/TTSD auf M-TTZM/TTSD die Wiedergabe zeitlicher Strukturen, ohne daß sonstige Nachteile auftreten. Der weitere Übergang auf ZFKII/RKHP berücksichtigt auch ausgeprägtere impulshafte Anteile. Die Sprachqualität steigt bei jedem Übergang an. Der eindrucksvolle Übergang auf ZFKI/RKOP schließlich erlaubt nahezu verfälschungsfreie Verarbeitung. Letzteres bedeutet, daß nur mit ZFKI die gehörrelevante Information voll durch Konturen erfaßbar ist. Leider wird die per RKOP simulierte optimale Rekonstruierbarkeit von Phasen im Verlauf dieser Arbeit nicht realisiert, sondern nur als grundsätzlich existent behandelt (Abschnitt 5.1.5.4).

3.4.3 Zusammenstellung und Erklärung der Parametereinflüsse

Wie sind die neuen Einstellungen in Tabelle 3.2 speziell für feineinstellbare Parameter zu erklären? Dazu werden abschließend die objektiv belegbaren Einflüsse auf die Verarbeitungsqualität zusammengestellt, die die meist subjektiv ermittelten Einstellwerte verständlich machen. Den Bezugspunkt der Erklärungen bildet die ursprüngliche Einstellung der Heinbachschen Teiltonanalyse (HB-TTZM). Die Parameterveränderungen mit den wichtigsten Konsequenzen betreffen Fensterfunktionsgrad und Analysebandbreite.

3.4.3.1 Fensterfunktionsgrad n

Eine Erhöhung hat folgende Konsequenzen:

- Die spektrale Selektion der Analysefilter wird verbessert, ohne daß die zeitliche Selektion signifikant verschlechtert wird (Abschnitt 3.3.3). Die Simultanverdeckung in den Frequenzkonturen (Abschnitt 2.4) sinkt dadurch ab, und zwar um so mehr, je größer die spektrale Entfernung zwischen Maskierer und zu maskierendem Sinuston ist. So gewinnt man Spielraum für eine erwünschte höhere Analysebandbreite (s.u.), die sonst die Verdeckung zuerst zwischen spektral entfernteren Sinustönen störend anheben würde. Der Spielraum bleibt jedoch beschränkt, weil sich die spektrale Selektion im Nahbereich kaum verbessern läßt und die Verdeckung hier bei höherem n als erstes kritisch wird (M-TTZM gegenüber HB-TTZM in Bild 2.8 auf S. 42). Ein Übergang von nP1 auf die Fensterfamilien PGn oder Bn ändert daran praktisch nichts, allein die Ausgeprägtheitschwelle (s.u.) kann den Spielraum noch ein wenig vergrößern.

- Transiente und stationäre Anteile im FTT-Spektrum heben sich besser voneinander ab (Abschnitt 3.3.5). Erst dadurch werden schwächer ausgeprägte transiente Anteile geeignet durch Zeitkonturen repräsentiert. Andernfalls können sie sich nur als geringe Änderung im Spektrum niederschlagen, das im wesentlichen vom stationären Anteil dominiert ist. Dann ist es allgemein viel wahrscheinlicher, daß Zeitkonturpunkte im Pegel maßgeblich vom stationären Anteil bestimmt sind. Eine Zeitkonturverar-

beitung bei zu niedrigem n läßt deshalb die impulshaften Signalanteile unnatürlich übertrieben klingen.

Weil sich transiente und stationäre Anteile besser voneinander abheben, schrumpfen die Überschneidungsgebiete, in denen beide Anteile die gleiche Größenordnung aufweisen. Welligkeiten, die durch Überlagerung der Anteile entstehen und die Nebenkonturlinien auslösen, repräsentieren somit weniger Energie. Sie brauchen deshalb nicht mehr wie bei $n = 1$ durch zeitliche Glättung des Spektrums unterdrückt zu werden, um unangenehme Verfälschungen zu verhindern. Eine Glättung ist nämlich nicht wünschenswert, sie wirkt tonalisierend (Abschnitt 3.3.6.2).

Sind ab $n \geq 2$ die Welligkeiten auf ein bestimmtes Maß reduziert, dann können die resultierenden Nebenkonturlinien sogar nützen, um die fehlende Zeitkonturverarbeitung der TTZM-Verfahren etwas zu kompensieren. Weil sie sehr dicht liegen und weil ihnen TTSD keine kohärenten Synthesesinusschwingungen zuweist, verursachen sie beispielsweise beim Einschalten eines Tons eine Art 'Rauschstoß'. Dieser imitiert näherungsweise den sonst fehlenden Knack (ausführliche Erklärung in Abschnitt 5.1.7.2). Dieser Effekt spielt bei $n = 4$ nur noch eine geringe Rolle, und dies auch nur dann, wenn ausschließlich Frequenzkonturen verarbeitet werden.

- Das Laufzeitniveau und damit der absolute Laufzeitunterschied zwischen hohen und tiefen Analysefrequenzen steigt (Gl. (3.11), T_{max} steigt). Beim Übergang von $n = 1$ auf $n = 2$ kann dies durch die dann mögliche höhere Analysebandbreite zunächst kompensiert werden (Gl. (3.11), B_{3dB} steigt). Aber ab $n > 2$ ist grundsätzlich ein Laufzeitausgleich erforderlich, weil sonst bei schnellen Hüllkurvenänderungen des Signals höhere Frequenzen vor tieferen zu hören sind.

- Die Symmetrie der zeitlichen Selektion steigt (Abschnitt 3.3.3). In Verbindung mit kleinen Analysebandbreiten wirkt sich dies ungünstig auf die Verarbeitungsqualität aus. Beispielsweise beim Ein- und Ausschalten eines Sinustons verlangsamt sich der Anstieg des FTT-Pegelspektrums zugunsten eines schnelleren Abfalls. Besonders bei kleinen Analysebandbreiten und tiefen Frequenzen wird diese Veränderung wahrnehmbar. Der Anstieg klingt dann 'verlangsamt', weil die Anstiegszeit die Größenordnung der Vorhörschwelle des Gehörs erreicht. Deshalb ermöglicht ein $n > 1$ nicht nur eine höhere Analysebandbreite $B_{3dB} > 0{,}1$ Bark, sondern es erfordert sie sogar.

Der eingestellte Grad $n = 4$ markiert bei optimaler Rekonstruktion, Zeitkonturverarbeitung und möglichst hoher Analysebandbreite einen Mindestwert, ab dem nahezu verfälschungsfreie Sprachverarbeitung erreicht wird. Eine weitere Erhöhung bringt hier keine Vorteile. Beim verbesserten TTZM-Verfahren könnte man erst mit noch höherem n den 'Rauschstoß'-Effekt völlig eliminieren, was wegen der Zeitkonturersatzwirkung nicht unbedingt wünschenswert ist. Die bei Gehörmodellen verbreitete Gamma-Ton Filterbank von Patterson et al. verwendet ebenfalls ein Fenster nP1 mit $n = 4$ [Pat92].

3.4.3.2 Analysebandbreite B_{3dB}

Eine Erhöhung hat folgende Konsequenzen:

- Das Zeitverhalten der Konturrepräsentation verbessert sich. Speziell mildert sich die beobachtete Glättung der Schmalbandhüllkurve (Abschnitt 2.2), die Pegel- und Frequenzentwicklung von Frequenzkonturlinien wird variabler. Für die Verarbeitungsqualität bedeutet dies eine verminderte 'Halligkeit' und einen weniger 'raumübertragenen' Höreindruck. Ab $B_{3dB} \geq 0{,}5$ Bark verschwinden diese Effekte völlig.

- Die Simultanverdeckung in den Frequenzkonturen steigt. Damit die Simultanverdeckung im Gehör nicht übertroffen wird, darf B_{3dB} nicht zu sehr erhöht werden. Der Spielraum hängt hauptsächlich von der Fensterfunktion (s.o.) und ein wenig auch von der Ausgeprägtheitsschwelle (s.u.) ab.

- Die Bedeutung der Zeitkonturen für die Sprachverarbeitung nimmt zu. Das äußert sich darin, daß die Dichte von Frequenzkonturen zugunsten der von Zeitkonturen sinkt (Übergang von ZFKII mit $B_{3dB} = 0{,}3$ Bark auf ZFKI mit 0,5 Bark in Bild 3.9). Speziell die Glottisschwingung manifestiert sich von höheren Frequenzen her zunehmend durch Impulsbeiträge, anstatt durch einzeln aufgelöste Harmonische (Abschnitt 2.1.2). Bei fehlender Zeitkonturverarbeitung bedeutet dies, daß der Anteil der nichtrepräsentierten Energie größer wird. Aus diesem Grund nimmt man bald eine unnatürliche überspitzte Artikulation wahr (ähnlich Abschnitt 2.6.1). Besonders Männerstimmen sind wegen des geringeren Harmonischenabstandes kritisch. Bei M-TTZM bildet dieser Sachverhalt eines der Hindernisse, B_{3dB} über 0,3 Bark hinaus zu vergrößern.

  Speziell bei den Rauschanteilen nimmt die Bedeutung der Zeitkonturen noch aus einem anderen Grund zu. Die bei Rauschen festgestellte Tonalisierung bei reiner Frequenzkonturverarbeitung existert nämlich zunächst bei jeder Analysebandbreite (Abschnitt 2.5). Wohl aber verändert sich die Klangfärbung von 'statisch nasal' zu 'schwirrig plätschernd'. Letzteres irritiert etwas mehr. Zeitkonturverarbeitung wird dann wichtiger, weil sie Tonalisierungen verhindern kann. Dazu muß allerdings mittels RKOP rekonstruiert werden. Bei RKHP nämlich kann die feine spektrale Lückenbildung nicht vermieden werden, weil sich die Synthesebeiträge von benachbarten Frequenz- und Zeitkonturlinien inkohärent überlagern (Abschnitt 5.1.5.3).

- Die Tendenz zur Ausbildung unterbrochener Frequenzkonturlinien nimmt bei Sprachverarbeitung zu. Das liegt daran, daß schon bei tieferen Frequenzen mehr als eine Harmonische der Glottisschwingung in eine Analysebandbreite fallen kann. Die häufigeren Kurzverläufe, Aufspaltungen und Verschmelzungen bewirken, daß Synthesefenster-kontrollierbare und Phaseninkohärenz-bedingte Störungen zunehmen (Abschnitt 2.3). Weil beide Störungstypen – insbesondere aber der zweite – nur bei RKOP ausgeschlossen sind, bleibt B_{3dB} für M-TTZM wie auch für ZFKII auf 0,3 Bark beschränkt. Wiederum sind Männerstimmen besonders kritisch, wo bei höheren Werten lästige Rauhigkeiten aufgrund von Phaseninkohärenzen wahrzunehmen sind.

Die Einstellung $B_{3dB} = 0{,}5$ Bark stellt einen Mindestwert dar, ab dem bei optimaler Rekonstruktion und Zeitkonturverarbeitung eine nahezu verfälschungsfreie Sprachverarbeitung möglich ist. Auch mit Werten bis zu $B_{3dB} = 0{,}7$ Bark wurden keine Qualitätsverschlechterung beobachtet. Dieser Wertebereich korresponiert eher mit aufwendigeren Gehörmodellen als die ursprünglichen 0,1 Bark bei HB-TTZM. Peisl [Pei90] bestimmt

aus der Simulation eines aktiven Innenohrmodells funktional äquivalente Analysefilter, indem er die Impulsantwort an der simulierten Basilarmembran wie auch die Breite ihrer Auslenkung bei Tönen mißt. Für kleine Pegel erhält er übereinstimmend eine 6dB-Bandbreite von $B_{6dB} \approx 0{,}8$ Bark, die bei höheren Pegeln zunimmt. Bei einer Annahme von $n = 2 \ldots 4$ ergibt sich mit Hilfe von Bild 3.6b auf S. 75 ein umgerechneter Mindestwert von $B_{3dB} \approx 0{,}5$ Bark.

3.4.3.3 Analysefrequenzabstand $\Delta\omega_A$

Die Einstellung $\Delta\omega_A/(2\pi) = 0{,}05$ Bark wurde unverändert von Heinbach übernommen, der sie in etwa auf die Frequenzauflösung des Gehörs für stationäre Sinustöne abstimmte [Hei88a]. Ohne weitere Maßnahmen liegt damit auch die Frequenzgenauigkeit der Konturpunkte fest. Die Genauigkeit läßt sich aber mit Approximationsverfahren steigern, so daß man sogar ein größeres $\Delta\omega_A$ wählen könnte. Dies kann sich vom Rechenaufwand her lohnen. Allerdings müssen auf eine Analysebandbreite B_{3dB} genügend Analysefrequenzen kommen, mindestens etwa zwei, damit das FTT-Spektrum über der Frequenz fein genug abgetastet wird (vgl. Anhang C.3). Außerdem ist zu beachten, daß dann die Ausgeprägtheit eines lokalen Maximums erst nach seiner Approximation und nach Approximation der benachbarten Minima zuverlässig bewertet werden kann. Für visuelle Darstellungen von Frequenzkonturen sind grundsätzlich höhere Frequenzauflösungen als 0,05 Bark wünschenswert, um Treppeneffekte zu vermeiden (Abschnitt 1.5.2).

3.4.3.4 Auswerteintervalldauer T_A

Die Wahl von T_A wird von zwei verschiedenen Grenzen bestimmt, abhängig davon, ob Zeitkonturen verarbeitet werden oder nicht. Die höhere leitet sich aus den Frequenzkonturen ab. Deren feinzeitliche Variabilität darf nicht zu sehr eingeschränkt werden, sonst nimmt man einen Tonalisierungseffekt wahr (Abschnitt 2.6.2). Die Einstellung $T_A = 1{,}25$ ms liegt subjektiv auf der sicheren Seite, verglichen mit einem doppelt so großen Wert. Dies paßt zur Wahrnehmungsgrenze der Rauhigkeit, nach der Schwankungen der Schmalbandhüllkurve über 250 Hz nicht mehr wahrgenommen werden [Ter68a]. Um solche Schwankungen abtasten zu können, benötigt man eine Auswerterate von mindestens 500 Hz, entsprechend einem Auswerteintervall von $T_A \leq 2$ ms.

Die Grenze für T_A bei Frequenzkonturen wird aber nicht allein von der Wahrnehmungsgrenze der Rauhigkeit bestimmt: Sind noch schnellere Schwankungen einer Frequenzkonturlinie vorhanden, so sind sie zwar im Sinne einer gehörgerechten Analyse irrelevant. Im rekonstruierten Signal der TTZM-Verfahren repräsentieren sie jedoch eine gewisse spektrale Breite. Wird sie durch Abtastung eingeschränkt, dann kann trotzdem eine Tonalisierungstendenz wahrnehmbar werden, obwohl die Abtastrate der Wahrnehmungsgrenze der Rauhigkeit genügt. Deshalb sollte T_A so klein gewählt sein, daß auch der Zeitverlauf des FTT-Pegelspektrums an der höchsten Analysefrequenz noch oft genug abgetastet wird. Eine Überpüfung ähnlich Abschnitt 2.6.2 zeigt, daß dies bei $T_A = 1{,}25$ ms für Analysebandbreiten $B_{3dB} \leq 0{,}3$ Bark und an Analysefrequenzen bis etwa 7 kHz einigermaßen gewährleistet zu sein scheint.

Bei Zeitkonturverarbeitung hängt die Grenze für T_A von zwei weiteren Anforderungen ab. Erstens ist zu beachten, daß die Pegelanstiegsgeschwindigkeit $\partial L^L(f,t)/\partial t$ (Bild 3.1 auf S. 61) in der Praxis durch einen Differenzenquotienten anzunähern ist. Er muß über einem ausreichend schmalen Auswerteintervall bestimmt werden, damit die Zeitkonturierung zuverlässig funktioniert. Zweitens ist die Auswirkung ungenau aufgelöster Zeitkonturpunkte bei Signalrekonstruktion zu berücksichtigen. Wenn beispielsweise bei der Impulsfolge (Bild 3.2 auf S. 63) die Abstände regelmäßig aufeinanderfolgender Zeitkonturen 'zittern', nimmt man dies als Störung war – und zwar auch dann, wenn T_A der Wahrnehmungsgrenze der Rauhigkeit genügt. RKOP erweist sich hierbei aus noch zu behandelnden Gründen viel unempfindlicher als RKHP (Abschnitt 6.2.1.2) und Sprache unempfindlicher als eine Impulsfolge.

Die Wahl $T_A = 1/f_a$, worin f_a die Abtastrate des Zeitsignals darstellt, liegt für alle Fälle auf der sicheren Seite. Eine Entscheidung über eine kritischere Abtastung ist damit auf die Codierung in Kapitel 6 verschoben.

3.4.3.5 Ausgeprägtheitschwelle ΔL_A für Frequenzkonturen

Ein niedrigeres ΔL_A führt dazu, daß die Simultanverdeckung der Frequenzkonturen untereinander sinkt. Bei $\Delta L_A = 0\,\mathrm{dB}$ wird ein Anschlag erreicht, der die Schwelle dann maximal um einige dB gegenüber einer Einstellung von $\Delta L_A = 3\,\mathrm{dB}$ absenkt. Der Fensterfunktionsgrad beeinflußt die Simultanverdeckung zwischen entfernteren Sinustönen wesentlich stärker. Insofern kann ein niedriges ΔL_A nicht die Nachteile eines zu geringen Fensterfunktionsgrades kompensieren. Wohl aber kann die Verdeckung zwischen nahe beieinander liegenden Sinustönen abgesenkt werden, was über den Fensterfunktionsgrad nicht zu erreichen ist (s.o.). Dadurch wird der Spielraum für eine höhere Analysebandbreite ein wenig verbessert.

Eine zu hohe Einstellung von ΔL_A ist zuerst bei rauschhaften Anteilen als Tonalisierungstendenz wahrzunehmen. Hier nimmt die spektrale Lückenbildung zu (Abschnitt 2.5), weil die Simultanverdeckung im Nahbereich dann angehoben ist. Eine zu niedrige Einstellung ΔL_A schadet der Verarbeitungsqualität nicht, modelliert aber eine unrealistische Empfindlichkeit des Gehörs (Abschnitt 3.3.6). Die Wahl $\Delta L_A = 0{,}5\,\mathrm{dB}$ liegt auf der sicheren Seite vom Schwellenmaß 1 dB, das im Zwickerschen Schwellenfunktionsschema in der Psychoakustik zugrunde gelegt wird [Zwi82].

3.4.3.6 Ausgeprägtheitschwelle λ für Zeitkonturen

Grundsätzlich ist λ so zu wählen, daß möglichst nur dann Zeitkonturpunkte ausgelöst werden, wenn das Analysefilter im wesentlichen mit seiner Impulsantwort reagiert. Nur dann liegen Formationen im FTT-Pegelspektrum vor, die nicht durch Frequenzkonturen repräsentierbar sind, beispielsweise spektrale Verbreiterungen bei Hüllkurvenänderungen (Abschnitt 3.3.5). Ist λ zu klein, dann tauchen Zeitkonturen häufiger auch bei kleinen und langsamen, quasistationäre Pegeländerungen auf, die durch Frequenzkonturen bereits repräsentiert sind. Dadurch nehmen Doppelrepräsentationen zu, die zur Signalrekonstruktion entfernt werden müssen (Abschnitt 5.1.3). Ist λ zu groß, dann werden kleine spektrale Verbreiterungen übersehen. Als erster spürbarer Effekt nimmt dann die Tonalisierung

von Rauschen wieder zu, die bei ZFKI/RKOP durch die Zeitkonturverarbeitung nahezu verschwunden ist.

Zur Unterscheidung zwischen langsamer, quasistationärer Pegeländerung und Reaktion mit der Impulsantwort muß man die sinkenden Einschwingzeiten der Analysefilter zu höheren Analysefrequenzen hin berücksichtigen. Daher geht B_{3dB} in die Spezifikation mit ein. Die gewählte Einstellung $\lambda = 25\,\mathrm{dB} \cdot B_{3dB}$ ist relativ unkritisch. In einem Toleranzbereich etwa zwischen verdoppeltem und halbiertem Wert ändert sich die Qualität kaum.

3.4.4 Zusammenfassung

Zeitkonturierung und modifizierbare FTT-Fensterfunktionen erweitern die Heinbachsche Teiltonanalyse zur Konturanalyse. Mittels Signalrekonstruktion sollten ihre Transformations- und Konturierungsparameter hier so eingestellt werden, daß bestmögliche Verarbeitungsqualität von Sprache erreicht wird. Eine solche Einstellung reflektiert nicht notwendigerweise eine optimale Gehöranpassung, weil sie von den Verfälschungseigenschaften des Rekonstruktionsverfahrens abhängt. Solange noch Verfälschungen wahrnehmbar sind, wird nämlich ein Kompromiß zwischen Verfälschungen durch Rekonstruktion und Verfälschungen durch suboptimale Gehöranpassung eingestellt. Der Kompromiß kann dann vom Signaltyp, von den Abhörbedingungen und vom Hörgeschmack des Zuhörers abhängen.

Wegen der später in Kapitel 5 auftretenden Schwierigkeiten, eine verfälschungsfreie Rekonstruktion zu realisieren, werden drei verschiedene Verfahren herangezogen. Sie führen auch zu drei verschiedenen Einstellungen. Die konkrete Funktionsweise der Verfahren spielt hier noch keine Rolle. RKOP, das erste, simuliert eine theoretisch realisierbare, verfälschungsfreie, und damit optimale Rekonstruktion. RKHP, das zweite, kann Phaseninkohärenz-bedingte Störungen nicht vermeiden. TTSD, die Teiltonsynthese mit Dreieckfenster als drittes Verfahren, verursacht zusätzlich Störungen durch das Synthesefenster. Außerdem kann sie keine Zeitkonturen verarbeiten, so daß die mit ihrer Hilfe gewonnene Einstellung de facto einer verbesserten Teiltonanalyse entspricht.

Die den Rekonstruktionsverfahren zugeordneten drei Einstellungen ZFKI, ZFKII und M-TTZM in Tabelle 3.2 auf S. 87 sind das Ergebnis umfangreicher Hörvergleiche. Sie wurden anhand verschiedener Sprachsignale mit dem Autor als Versuchsperson durchgeführt. Die wichtigsten Parameter sind – neben dem Einsatz von Zeitkonturen – der Fensterfunktionsgrad n und die Analysebandbreite B_{3dB}. Während ersterer jeweils auf $n = 4$ eingestellt wurde, hängt letztere von der Rekonstruktion ab.

Die Paarung ZFKI/RKOP ermöglicht nahezu verfälschungsfreie Verarbeitung bei $B_{3dB} = 0{,}5$ Bark. Zeitkonturen leisten dabei einen wesentlichen Beitrag. Diese Einstellung wird außerdem als relativ gehörnah angesehen. Bei ZFKII/RKHP wurde aufgrund der Charakteristik der Phaseninkohärenz-bedingten Störungen der Kompromiß $B_{3dB} = 0{,}3$ Bark eingestellt. Die Qualität sinkt wegen der Störungen deutlich, sogar der nutzbringende Beitrag der Zeitkonturen wird durch sie gemindert. Bei M-TTZM/TTSD, dem verbesserten TTZM-Verfahren, sinkt die Qualität noch etwas weiter, weil keine Zeitkonturen verarbeitet werden. Allerdings kann die Störcharakteristik von TTSD Zeitkonturverarbeitung in geringem Umfang vortäuschen. Auch hier ist $B_{3dB} = 0{,}3$ Bark eingestellt. Immerhin wird noch eine wesentlich bessere Verarbeitungsqualität als beim Heinbachschen TTZM-

Verfahren erzielt.

Um die subjektiv durchgeführten Einstellung zu objektivieren, wurden die Einzeleffekte ausführlich erörtert, die bei der Änderung von feineinstellbaren Parametern eine Rolle spielen. Bezugnehmend auf die Einstellungen der Heinbachschen Teiltonanalyse sind dies im wesentlichen folgende:

- Die Erhöhung des Fensterfunktionsgrades n senkt die Simultanverdeckung der Frequenzkonturen untereinander. Damit schafft sie Spielraum für eine Erhöhung von B_{3dB}, die den umgekehrten Effekt herbeiführt. Quasistationäre und transiente Anteile werden im FTT-Spektrum besser voneinander getrennt, so daß verfälschungsfreie Zeitkonturverarbeitung möglich und tonalisierende zeitliche Glättung unnötig wird. Weil das Laufzeitniveau steigt, wird ein Laufzeitausgleich erforderlich. Eine Erhöhung über $n = 4$ hinaus bringt keine weiteren Vorteile. Die Simultanverdeckung wird nämlich bei bereits erhöhtem B_{3dB} zuerst im Nahbereich kritisch, wo ein höheres n noch kaum Einfluß zeigt. Dort wirkt eine reduzierte Ausgeprägtheitsschwelle ΔL_A etwas entschärfend.

- Die Erhöhung der Analysebandbreite B_{3dB} verbessert die zeitliche Variabilität der Frequenzkonturen und reduziert insbesondere die Wahrnehmbarkeit der Schmalbandhüllkurvenglättung. Die Bedeutung von Zeitkonturen nimmt zu, ebenso die der Phaseninkohärenz-bedingten und Synthesefenster-kontrollierbaren Störungen. Deswegen und wegen der Anhebung der Simultanverdeckung sind einer Erhöhung Grenzen gesetzt, je nach Rekonstruktionsverfahren und Einsatz von Zeitkonturen.

3.5 Zusammenfassung

Das Konturierungskonzept des Heinbachschen TTZM-Verfahrens wurde in diesem Kapitel erweitert. Außerdem wurden Eigenschaften der Spektraltransformation modifiziert, die eng mit einer gehörorientierten Konturierung zusammenspielen. Probleme einer Signalrekonstruktion aus Konturen standen dabei im Hintergrund, Codierung blieb unberücksichtigt. Die Ergebnisse führen auf verbesserte und erweiterte Audiorepräsentationen mit Konturen, die sich auf das Terhardtsche Modell der auditiven Informationsverarbeitung stützten. Ihr Gewinnungsprozeß wird als Konturanalyse bezeichnet. Das erweiterte Konzept kennt nunmehr zwei Konturtypen:

Frequenzkonturen: Für eine Repräsentation nach Art des Teiltonzeitmusters wird die Bezeichnung Frequenzkonturen verwendet. Im Gegensatz zum Heinbachschen Teiltonbegriff unterstreicht man damit, daß Quellsinusschwingung, zeitvariantes Spektralmaximum der FTT, Synthesesinusschwingung und Modellierung einer wahrgenommenen Spektraltonhöhe unbedingt auseinanderzuhalten sind. Die neue Bezeichnung drückt auch die Symmetrie zum neuen, zweiten Konturtyp aus.

Zeitkonturen: Hiermit werden Beiträge von transienten Anteilen im FTT-Pegelspektrum erfaßt, was mit Frequenzkonturen nicht möglich ist. Während die von Heinbach übernommene Frequenzkonturierung das zeitvariante FTT-Pegelspektrum in Schnitten parallel zur Frequenzachse nach Pegelmaxima absucht, geschieht dies bei

Zeitkonturierung in Schnitten parallel zur Zeitachse. Um einen verzögerungsarmen Entscheidungsprozeß zu erhalten, wird die Ausgeprägtheit der Maxima nicht mit Hilfe der benachbarten Minima, sondern über die vorangegangene Steilheit des Pegelanstiegs bewertet.

Es wurde ausführlich untersucht, wann und wie sich Zeitkonturen ausprägen. Bei Sprachsignalen stellen sie beispielsweise Glottisimpulse und Anteile von Plosiven dar. In der bildlichen Konturdarstellung erleichtern sie die Interpretation des Sprachsignals. Separate Rekonstruktion von Zeitkonturen zeigt später, daß sie vor allem die impulshaft empfundenen Signalanteile repräsentieren.

Um die Eigenschaften der Spektraltransformation im Zusammenspiel mit der Konturierung zu verbessern, ist die FTT-Fensterfunktion der wesentliche Eingriffspunkt. Sie kann als Impulsantwort eines normierten Tiefpasses spezifiziert werden. Mit den logarithmierten Beträgen seiner Systemfunktion, seiner Impulsantwort sowie dem Gruppenlaufzeitverlauf kann man wichtige Eigenschaften beschreiben. Die ersten beiden Maße werden hier spektrale beziehungsweise zeitliche Selektion genannt. Folgendes wurde erkannt:

Zur Konturierung geeignete Fensterfunktionen: Konturierung im Einklang mit dem Terhardtschen Modell stellt bestimmte Anforderungen an die Fensterfunktion. Die spektrale Selektion und möglichst auch die zeitliche Selektion dürfen keine Nebenmaxima aufweisen. Außerdem sollte das Laufzeitniveau nicht zu hoch sein. Geeignete reelle Fensterfunktionen charakterisieren die Familie nP1 der Tiefpässe mit einem n-fachen Pol. Im Gegensatz zu $n = 1$ bei Heinbach sollte der Fensterfunktionsgrad n größer, aber auch nicht zu groß gewählt werden. Dabei ist ein Laufzeitausgleich sinnvoll, der die maximalen Fensteröffnungen von hohen an die von tiefen Analysefrequenzen angleicht.

Zusammenspiel Fensterfunktionsgrad/Konturierung/Glättung: Ein höheres n ermöglicht zunehmend bessere Trennung von quasistationären und transienten Beiträgen im FTT-Spektrum. Ihre getrennte Repräsentation durch Frequenz- und Zeitkonturen funktioniert somit ebenfalls immer besser. Aus diesen Gründen erweist sich Zeitkonturierung erst ab $n > 2$ sinnvoll. Die von Heinbach eingeführte und bei $n = 1$ unbedingt nötige zeitliche Glättung ist nun entbehrlich, ja sogar schädlich. Sie eignet sich nicht, die aus der Psychoakustik bekannte Wahrnehmungsgrenze der Rauhigkeit zu modellieren.

Fensterfunktion und Analysebandbreite wie auch weitere, weniger kritische Transformations- und Konturierungsparameter wurden anschließend mittels Signalrekonstruktion optimiert. Dabei wurden später noch zu beschreibende Rekonstruktionsverfahren verwendet. Mit einer Reihe von Sprachsignalen sollten die Parameter im Selbstversuch auf möglichst gute Verarbeitungsqualität justiert werden. Dieses sind die wesentlichen Ergebnisse:

Abhängigkeit von der Signalrekonstruktion: Abhängig von den Fähigkeiten eines Verfahrens zur Signalrekonstruktion werden Parameter verschieden eingestellt. Bei suboptimaler Rekonstruktion beeinflussen sie nämlich auch die Charakteristik der rekonstruktionsbedingten Verfälschungen. Hinter einer gefundenen Einstellung verbirgt sich ein subjektiver Kompromiß, bei dem Verfälschungen der drei Teilkonzepte

Spektraltransformation, Konturierungskonzept und Rekonstruktion gegeneinander abgewogen worden sind. Nur eine optimale Rekonstruktion erlaubt es, die Optimalität einer Parametereinstellung neutral zu beurteilen.

Vollständigkeit der Konturrepräsentation: Eine nahezu perfekte Qualität bei optimaler Rekonstruktion kann man nur mit Zeit- und Frequenzkonturen erreichen. Die zuvor beim TTZM-Verfahren erkannten Probleme der Unterrepräsentation transienter Anteile und der Tonalisierung von Rauschanteilen sind also prinzipiell durch Hinzufügen von Zeitkonturen zu beheben. Erst beide Konturtypen zusammen ergeben eine Audiorepräsentation, die die wahrnehmungsrelevante Information vollständig erfassen kann.

Optimale Transformationsparameter: Gegenüber Heinbach wird der Grad der Fensterfunktionen auf $n = 4$ erhöht. Die bessere spektrale Selektivität ermöglicht größere Analysebandbreiten, ohne daß eine überhöhte Simultanverdeckung zu befürchten ist. Unterstützt durch den Laufzeitausgleich verbessert sich so das Zeitverhalten, insbesondere verringert sich die Glättung der Schmalbandhüllkurve. Statt einer 3dB-Analysebandbreite von 0,1 Bark bei Heinbach sind für eine optimale, gehörnahe Parametereinstellung mindestens 0,5 Bark erforderlich. Steht aber nur eine suboptimale Rekonstruktion zur Verfügung oder soll sogar auf Zeitkonturverarbeitung verzichtet werden, dann liegt der Kompromiß für die Verfälschungen der drei Teilkonzepte eher bei 0,3 Bark.

Zusammen mit ebenfalls eingestellten Konturierungsparametern wurden eine verbesserte und zwei erweiterte Repräsentationsformen für Audiosignale spezifiziert. Im Rahmen reiner Frequenzkontur-Repräsentation verbessern die neuen Transformationsparameter die Verarbeitungsqualität des TTZM-Verfahrens erheblich (M-TTZM). Erweiterte Repräsentation mit Zeit- und Frequenzkonturen existiert in zwei Varianten. Mit der einen (ZFKI) ist nahezu perfekte Verarbeitungsqualität möglich, wenn man optimale Rekonstruktion sicherstellen kann. Die andere Variante (ZFKII) ist an suboptimale Rekonstruktion angepaßt. Die Schwächen suboptimaler Rekonstruktion bedingen leider, daß Zeitkonturen die Verarbeitungsqualität von Sprache nur wenig steigern können.

Kapitel 4

Repräsentation mittels Kontur/Textur

Ein zentrales Problem der datenreduzierenden Varianten des TTZM-Verfahrens bestand darin, daß für nichttonale Signalanteile keine separate Repräsentationsform vorliegt. Auch die neu hinzugekommenen Zeitkonturen können die Datenreduktion nicht erleichtern, weil sie sich wie die Frequenzkonturen in unhandlich viele Linienverläufe aufgliedern. Deshalb wird ein neues Konzept zur Weiterverarbeitung eingeführt. Es trennt sogenannte *prägnante* Konturen ab, die in ihrem exakten Verlauf für die Wahrnehmung wichtig sind. Der Rest wird pauschal als *Textur* bezeichnet und durch eine spektral/zeitliche Hüllfläche repräsentiert. Die resultierende Kontur/Textur-Repräsentation kann tonale und geräuschhafte, darunter sogar noch 'rauschhafte' und 'impulshafte' Signalanteile separat erfassen. Sie bleibt gehörorientiert, womit sie sich von bisher bekannten Mischrepräsentationen wesentlich unterscheidet (vgl. Abschnitt 1.6, [Ser90, Ser96, Mar88, Mar91, Mar94]).

Der Grundgedanke des neuen Konzeptes ist folgender: Konturen beschreiben immer die Feinstruktur des zeitvarianten FTT-Pegelspektrums. Bei rauschhaften Anteilen scheint es aber übertrieben, jede der zahlreichen, dichtliegenden und unauffälligen Konturlinien als Einzelobjekte der Wahrnehmung zu behandeln. Die Beschreibung der Grobstruktur mit Hilfe einer Hüllfläche müßte ausreichen und würde sich besser zur Datenreduktion eignen. Zur Signalrekonstruktion könnte sie wieder mit einer beliebigen Feinstruktur ausgefüllt werden.

Der erste Abschnitt stellt einen einfachen Ansatz vor, mit dem Konturen mit unterschiedlicher Wahrnehmung getrennt werden können. Dieser Ansatz wird dann zum Kontur/Textur-Konzept verallgemeinert, das sich in das Terhardtsche Modell der auditiven Informationsaufnahme einfügt. Der dritte Abschnitt beschreibt ein Verfahren, mit dem die Texturhüllfläche zusammmen mit den prägnanten Konturen gewonnen werden kann. Weitere Abschnitte behandeln, wie sich Sprachsignale in zwei Varianten der neuen Repräsentation abbilden. Eine davon kann sogar auf Zeitkonturen verzichten. Abschließend wird nachgereicht, wie Parameter einzustellen sind und wie die Verarbeitungsqualität zu beurteilen ist.

4.1 Trennung tonaler, impuls- und rauschhafter Anteile über die Konturlinienlänge

Die Wahl der Attribute tonal/rauschhaft/impulshaft gründet sich auf die markante Wahrnehmung, die nach getrennter Signalrekonstruktion für den jeweiligen Anteil empfunden wird. Mit dieser Definition soll allerdings nicht ausgeschlossen sein, daß sich beim Hören eines Anteils eine der anderen beiden Wahrnehmungen in untergeordneter Weise einstellt. Die Trennvorgänge werden für Frequenz- und Zeitkonturen separat behandelt. Beide Konturtypen enthalten nämlich, im Sinne der Definition, gleichzeitig rauschhafte, ausschließlich aber tonale beziehungsweise impulshafte Signalanteile. [1]

4.1.1 Tonale und rauschhafte Anteile in den Frequenzkonturen

Frequenzkonturen von Geräuschen ohne ausgeprägten tonalen Charakter weisen dichtliegende und regellose, meist kurze Linienverläufe auf. Am besten sichtbar wird dies bei Weißem Rauschen in Bild 2.9 links auf S. 44. Auch bei Sprache finden sich im mittleren bis höheren Frequenzbereich solche Formationen, besonders bei den Frikativen ‘ch’ und ‘sch’ in Bild 3.4 oben auf S. 68. In stimmhaften Sprachabschnitten heben sich dagegen die Linienverläufe von separat aufgelösten Harmonischen, aber auch von Formanten deutlich in Länge, Pegel und Vorhersagbarkeit ab. Bei ihnen ist die Zuordnung einer tonalen Wahrnehmung (Spektraltonhöhe) wahrscheinlich, wobei quantitative Effekte der Tonhöhenwahrnehmung im einzelnen zu berücksichtigen sind [Ter72a, Ter72b, Ter79, Ter82].

Es liegt deshalb der Ansatz nahe, tonale Anteile von rauschhaften zu trennen, indem eine Mindestlinienlänge Δt_P vorgeschrieben wird. Zu berücksichtigen ist dabei, daß innerhalb tonaler Linien auch kleine Frequenzunstetigkeiten auftreten können. Ein Beispiel dafür, das weiter unten noch genauer behandelt wird, ist die Zweitonschwebung in Bild 2.6 auf S. 37 oben. Bei der Assoziation von Konturpunkten zu Linien wird deshalb eine gewisse Unstetigkeitstoleranz Δf_U zugelassen, so daß allgemein mit ‘Linien’ nicht streng geschlossene, sondern sogenannte *offene* Linien gemeint sind. Eine zeitdiskrete Realisierung der Trennung benötigt in jedem Fall eine solche Toleranz, um Abtastpunkte einer frequenzveränderlichen Linie assoziieren zu können. Eine exakte Definition der Linien und ihrer Länge liefert Anhang A.2.

Ausgehend von Bild 3.4 wurde in Bild 4.1 eine Trennung mit $\Delta t_p = 25\,\text{ms}$ und $\Delta f_U = 0{,}25\,\text{Bark}$ vorgenommen. Die Parametereinstellung wird später in Abschnitt 4.6 begründet. Man erkennt, daß kurze, unregelmäßige Konturlinien den rauschhaften Anteilen zufallen (unten). Im höherfrequenten Bereich verbleiben allerdings auch einige unregelmäßige Linien in den tonalen Anteilen (oben). Sie repräsentieren höhere Formanten, deren tonale Wahrnehmung eigentlich eher unausgeprägt ist. Andererseits finden sich auch einige regelmäßig und hochpegelig aussehende Linienstücke in den rauschhaften Anteilen. Sie stammen von Modulationseffekten von Harmonischen, die nicht mehr separat aufgelöst werden.

Daß die Trennung Schwächen aufweist, bestätigt sich anhand von rein rauschhaft beziehungsweise rein tonal wahrgenommenen Signalen. Einerseits verdeutlicht ein Blick auf

[1] In [Mum90] wurden Zeitkonturen als Repräsentation nur für impulshafte Anteile angesehen.

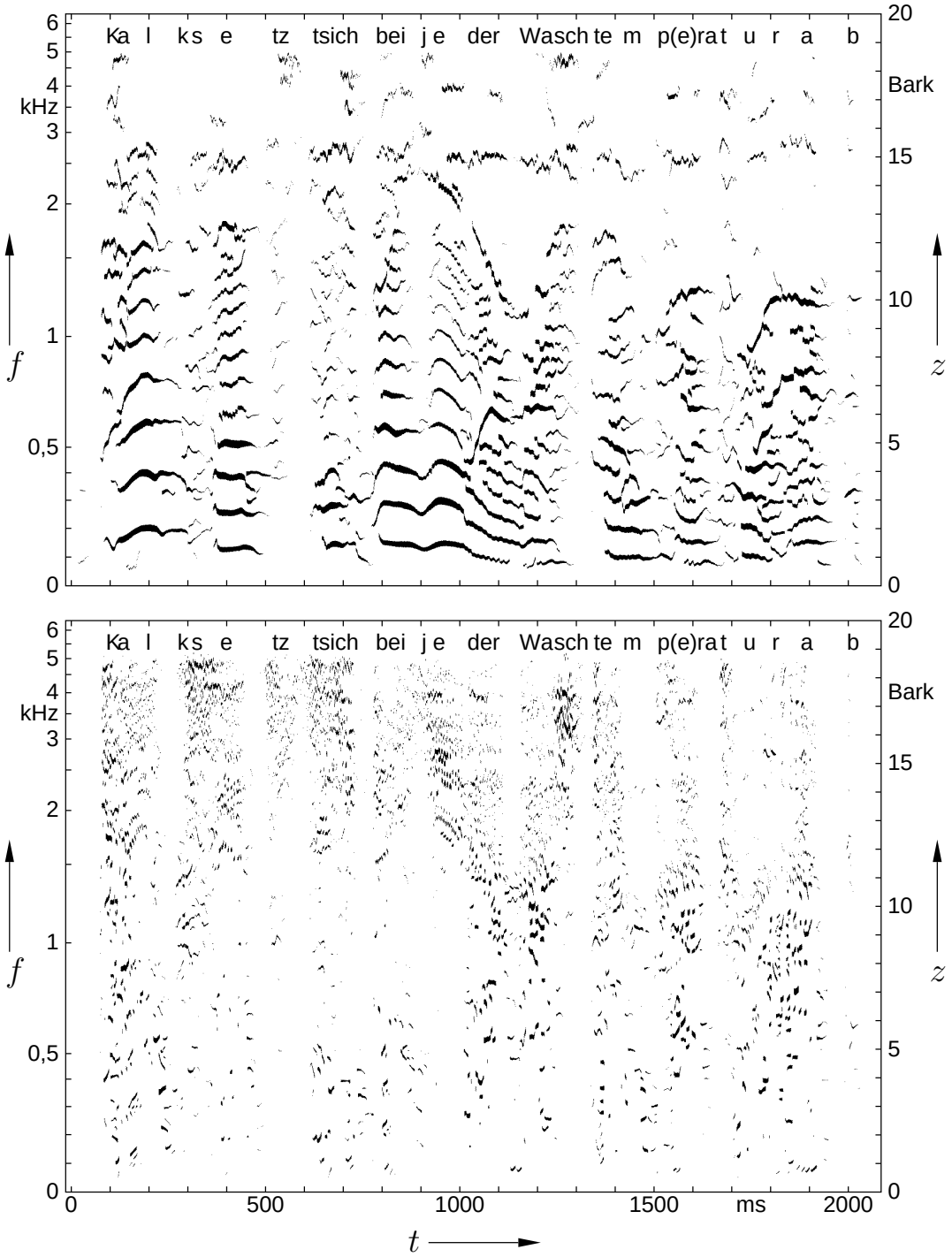

Bild 4.1: *Trennung von tonalen und rauschhaften Anteilen in den Frequenzkonturen von Sprache mittels eines Linienlängenkriteriums: Linien mit einer Mindestdauer* $\Delta t_p = 25$ ms *bei zulässigen Frequenzunstetigkeiten bis zu* $\Delta f_U = 0{,}25$ Bark *stellen vorwiegend tonale Anteile dar (oben). Die übrigen Linien stehen für die rauschhaften Anteile (unten). Die Überlagerung beider Teilbilder ergibt exakt Bild 3.4 oben auf S. 68.*

die Frequenzkonturen von Rauschen in Bild 2.9 links, daß sich auch hier langandauernde Linien etablieren können, die kaum eine tonale Wahrnehmung repräsentieren dürften. Andererseits kann eine Linienassoziation bei den Aufspaltungen und Verschmelzungen der Zweitonschwebung in Bild 2.6 grundsätzlich nur einen Linienast weiterführen. Einige 'tonale' Linienstücke bleiben dadurch unassoziiert, sie verfehlen die Mindestlänge und fallen fälschlicherweise den rauschhaften Anteilen zu. Dies gilt ebenso für die Kurzverläufe bei Amplituden- oder Frequenzmodulation (Abschnitt 2.3).

Trotzdem zeigen die Signalrekonstruktionen der getrennten Frequenzkonturen, daß der einfache Ansatz bei Sprache sehr gut funktioniert: Die tonalen Anteile hören sich wie 'zahn- und lippenlos' artikulierte Sprache, die rauschhaften Anteile dagegen wie Flüstersprache ohne hervorstechende tonale Merkmale an. Der Sprachinhalt beider Anteile, mehr noch der der rauschhaften Anteile, bleibt verständlich. Insbesondere die falschsortierten Linienstücke, die von nicht separat aufgelösten Harmonischen herrühren, stören in den rauschhaften Anteilen kaum. Auch bei gemeinsamer Rekonstruktion aller Frequenzkonturen verursachen sie wegen ihrer kurzen, inkohärenten Synthesesinusschwingungen eher geräuschartige Beiträge (Abschnitt 2.3.4), solange das Rekonstruktionsverfahren keine optimale Phasenrekonstruktion erlaubt.

4.1.2 Impulshafte und rauschhafte Anteile in den Zeitkonturen

Während Frequenzkonturen quasistationäre Strukturen des FTT-Spektrums repräsentieren, bilden Zeitkonturen transiente Strukturen ab, ohne die ein rekonstruiertes Signal wahrnehmbar verändert ist. Deshalb müssen Zeitkonturen eher impulshaft wahrnehmbare Signalanteile enthalten. Genauso aber, wie Frequenzkonturen nicht ausschließlich tonale Anteile repräsentieren, verkörpern Zeitkonturen nicht nur impulshafte Anteile. Auch bei Zeitkonturen kann man anhand der Linienlänge abschätzen, ob eine impulshafte Wahrnehmung zuzuordnen ist. Kurze Linien weisen jedenfalls darauf hin, daß im FTT-Spektrum keine ausgeprägte transiente Struktur vorliegt. Es ist wahrscheinlich, daß sich dann eine geringere impulshafte Wahrnehmung einstellen wird, als wenn eine lange Linie als Folge einer ausgeprägten transienten Struktur vorliegt.

Dies bestätigen auch weitere Beobachtungen. Kurze Zeitkonturlinien werden innerhalb von Rauschanteilen sichtbar, wie etwa im Bereich des Frikativs 'sch' in Bild 3.4 unten auf S. 68. Weiterhin tauchen sie zwischen ausgeprägten Harmonischen auf, was bereits bei der Impulsfolge in Bild 3.2 auf S. 63 deutlich wurde. Weil aber die Selektion der Analysefilter steil genug ist, liegt der Pegel dieser Zeitkonturen unter dem der Harmonischen. Beide Fälle sprechen nicht für eine ausgeprägte impulshafte Wahrnehmung. Dagegen gilt es im Falle der Tonimpulse in Bild 3.3 auf S. 66 als belegt, daß längere Zeitkonturlinien mit einer impulshafteren Wahrnehmung korrespondieren.

Zur Abtrennung der impulshaften Anteile wird analog zum Vorgehen bei den Frequenzkonturen eine Mindestlinienlänge von $\Delta f_p = 1\,\text{Bark}$ verwendet und eine zeitliche Unstetigkeitstoleranz $\Delta t_U = 1{,}25\,\text{ms}$ zugelassen. Die Parameter werden auch hier erst in Abschnitt 4.6 begründet. Eine exakte Definition von Zeitkonturlinien und ihrer Länge findet sich wiederum in Anhang A.2.

Im später noch speziell verwendeten Bild 4.2 ist ein Sprachsignalausschnitt zu sehen, wobei in der linken Hälfte nochmals die Frequenzkonturen und ihre tonalen Anteile ab-

gebildet sind. In der rechten Hälfte sind entsprechend die Zeitkonturen komplett und mit Mindestlänge dargestellt. Man erkennt, daß lange Linien vor allem bei höheren Frequenzen stehen bleiben, wo sie Glottisimpulse repräsentieren. Ausgeprägte Verschlußeffekte – nicht nur bei Plosiven – würden ebenfalls längere Linien hinterlassen, kommen aber in diesem Ausschnitt eigentlich nicht vor (vgl. Abschnitt 3.2.3). Bestenfalls ein Linienstück bei tiefen Frequenzen zu Beginn vom 'a' könnte in diesem Sinne interpretiert werden.

Rekonstruktionen aus den kürzeren (nicht abgebildeten) Zeitkonturlinien erinnern an die Flüstersprache der rauschhaften Anteile in den Frequenzkonturen. Sie sind aber nicht mehr so gut verständlich. Die Tiefen sind stärker vertreten, da die Linien zwischen den dort ausgeprägten Harmonischen nun nicht mehr durch die Harmonischen selbst verdeckt werden können. Insgesamt bleibt der Höreindruck aber rauschhaft. Rekonstruktionen aus den längeren Linien klingen wie eine Reihe übertrieben artikulierter, geflüsterter Plosive mit 'kratzigen' Anreicherungen im Hochtonbereich, die von den Glottisimpulsen stammen. Dieser Anteil ist für sich alleine kaum verständlich, sein Höreindruck ist ausgeprägt impulshaft.

4.2 Kontur/Textur-Konzept

Es gibt einen wesentlichen Unterschied zwischen den Konturen der tonalen und impulshaften Anteile auf der einen und denen der rauschhaften Anteile auf der anderen Seite: Bei ersteren ist es wahrscheinlich, daß einzelnen Konturverläufen auch einzelne Objekte der Wahrnehmung zugeordnet sind, etwa eine Spektraltonhöhe oder ein Klick. Bei letzteren kann dagegen angenommen werden, daß ihre einzelnen kurzen Verläufe nur gemeinschaftlich wahrgenommen werden. Zum Beispiel hört sich Weißes Rauschen praktisch immer gleich an, obwohl zu jeder Zeit unterschiedliche Konturverläufe vorliegen. Die einen Konturen sind offenbar *prägnant*, die anderen nicht. Bei nichtprägnanten Konturen scheint nur die Hüllfläche über Zeit und Frequenz wesentlich, die die zeitvariante, spektrale Färbung eines Rauschens charakterisiert.

Wegen der von Terhardt beschriebenen Gleichartigkeit der Verarbeitungsprinzipien von auditiver und visueller Wahrnehmung [Ter92] ist es aufschlußreich, eine Parallele zur visuellen Wahrnehmung zu ziehen: In Bildern gibt es auffällige Linien und Kanten, andererseits aber auch flächige Erscheinungen wie Farbe, Helligkeit, Farbsättigung und Musterung. Auf zwei Kategorien verkürzt kann man von prägnanten Konturen und *Textur* ('das innere Gefüge') sprechen.

Daher liegt die Einführung einer auditiven Textur nahe. Sie sei als Gesamtheit nichtprägnanter Konturen definiert und braucht nur durch eine spektral/zeitliche Hüllfläche erfaßt werden. Eine gleichwertige Textur kann jederzeit aus der Hüllfläche zurückgewonnen werden, wenn auch nicht mehr mit exakt denselben (nichtprägnanten) Konturen. Speziell kann die Hüllfläche zur Steuerung einer zeitvarianten Filterung von Weißem Rauschen verwendet werden, um direkt die rauschhaften Signalanteile zu rekonstruieren.

Die Trennung zwischen prägnanten Konturen und Textur, die Prägnanzentscheidung, bildet den Kern des Konzeptes. Sie erfordert für Zeit- und Frequenzkonturen jeweils ein Prägnanzkriterium, für das ein Prägnanzmaß einen Schwellwert erreichen muß. Das Prägnanzmaß sollte möglichst gut beschreiben, wie ausgeprägt ein Konturverlauf als Ein-

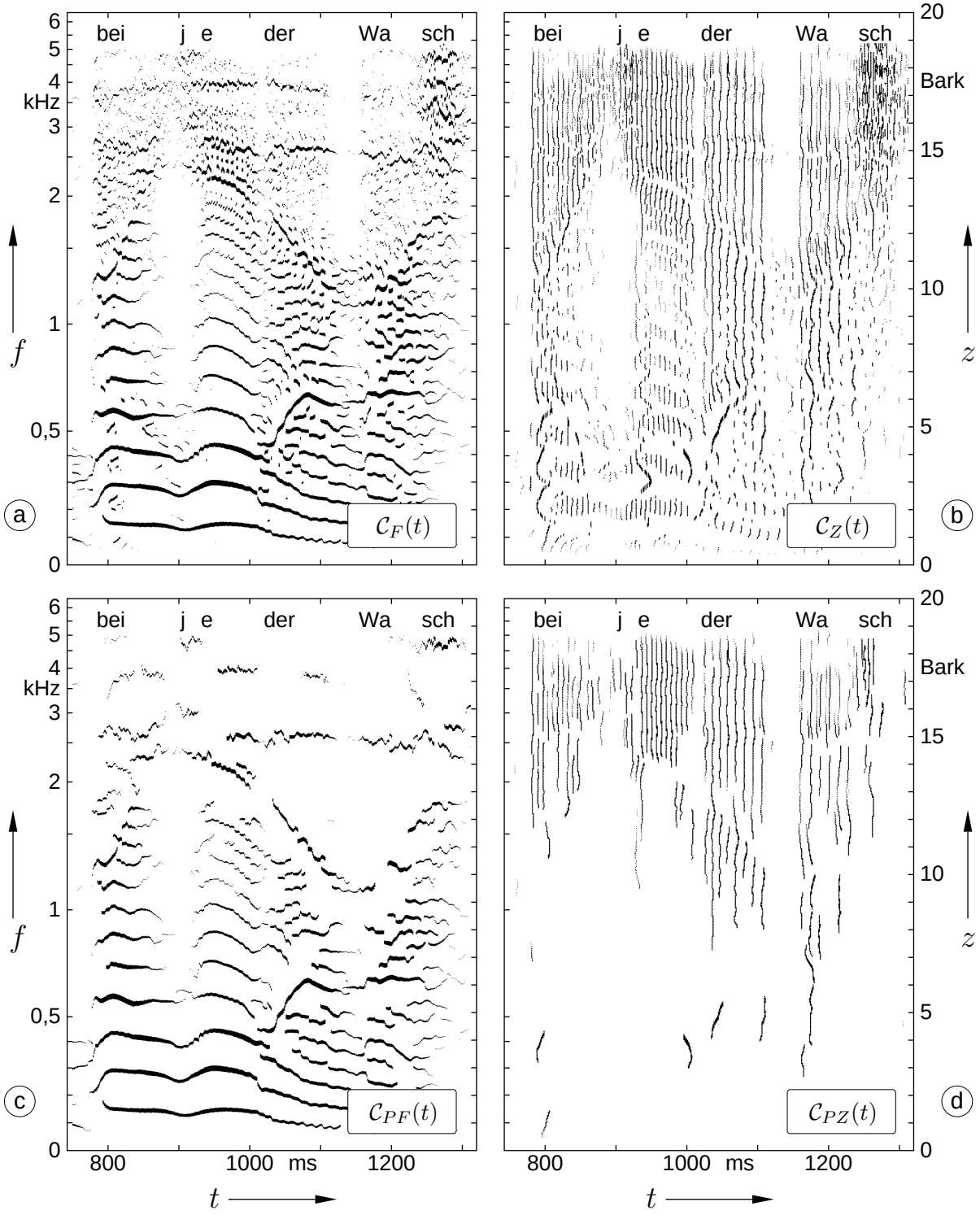

Bild 4.2: *Bestimmung prägnanter Konturen: Aus den Frequenzkonturen (a) werden prägnante Frequenzkonturen (c) als Linien mit einer Mindestlänge von* $\Delta t_p = 25\,\text{ms}$ *und maximaler Frequenzunstetigkeit* $\Delta f_U = 0{,}25\,\text{Bark}$ *ausgewählt. Sie stehen für tonal empfundene Sprachanteile. Aus den Zeitkonturen (b) werden prägnante Zeitkonturen (d) als Linien mit einer Mindestlänge von* $\Delta f_p = 1\,\text{Bark}$ *und maximaler Zeitunstetigkeit* $\Delta t_U = 1{,}25\,\text{ms}$ *ausgewählt. Sie repräsentieren impulshaft wahrgenommene Anteile. Nichtprägnante Zeit- und Frequenzkonturen verkörpern rauschhafte Anteile. Eingeblendete Signalnamen beziehen sich auf Bild 4.3.*

zelobjekt wahrgenommen wird. Je nach Güte des Prägnanzmaßes und Wahl der Schwellwerte korrelieren die prägnanten Konturen mit den tonal beziehungsweise impulshaft wahrgenommenen Signalanteilen und die Textur mit den rauschhaften Anteilen. Unabhängig von den gewählten Prägnanzkriterien bleibt aber die grobe spektral/zeitliche Hüllfläche des zeitvarianten Quellspektrums immer repräsentiert. Diese Art von Differentialprinzip ('konstante Summe konkurrierender Anteile') ergibt sich daraus, daß Konturverläufe schlimmstenfalls über die Texturhüllfläche erfaßt werden.

In dieser Arbeit wird die Linienlänge als einfach handhabbares, behelfsmäßiges Prägnanzmaß für beide Konturtypen verwendet. Ein ideales Prägnanzmaß für Frequenzkonturen beispielsweise gäbe zuverlässig die Ausgeprägtheit der zugehörigen tonalen Wahrnehmung wieder. Dies liefe auf die Modellierung einer dynamischen Spektraltonhöhenempfindung hinaus. So, wie stationäre Modelle Tonhöhengewichte für Komponenten eines Tonkomplexes liefern [Ter82], müßte dies nun für Frequenzkonturen geschehen. Nach Abschnitt 2.6.1 kann man leider nicht einfach ein stationäres Modell auf die Frequenzkonturpunkte im Auswerteintervall anwenden.

In der Terminologie des Terhardtschen Modells der auditiven Informationsaufnahme (Abschnitt 1.3) kann die Prägnanzentscheidung als Entscheidungsprozeß angesehen werden. Seine datenreduzierenden Eigenschaften werden sich später in Kapitel 6 bewähren. An welcher Stelle in der Hierarchie des Modells ist er anzusiedeln? Hierzu gibt es zwei Ansätze. Beim ersten wird er über der Zeit- und Frequenzkonturierung angeordnet. Demnach muß die tieferliegende Schicht auch die zahlreichen nichtprägnanten Konturen hinaufreichen, die erst dann zur Textur zusammengefaßt werden. Beim zweiten ist er identisch mit der Zeit- und Frequenzkonturierung, präzisiert aber deren Funktion. Dadurch wird die Prägnanzentscheidung zum elementar konturierenden Entscheidungsprozeß, der unmittelbar das gehörangepaßte Betragsspektrum verarbeitet. Mit einem entsprechend definierten Prägnanzkriterium könnte man mit diesem Ansatz auch die Spektraltonhöhe als *primäre* auditive Kontur modellieren, wie dies Terhardt fordert [Ter92].

4.3 Verfahren zur Gewinnung von Kontur/Textur-Repräsentationen

Die obigen beiden Ansätze führen auch auf zwei unterschiedliche Realisierungsansätze, um eine Kontur/Textur-Repräsentation zu erhalten. In jedem Fall sind zuerst alle Zeit- und Frequenzkonturen zu bestimmen. Daraus werden anhand der Linienlänge die prägnanten Konturen ermittelt, die die Konturanteile der Repräsentation beisteuern. Wie aber kommt man auf den Texturanteil, der durch eine Hüllfläche über Zeit und Frequenz darzustellen ist?

Beim ersten Ansatz berechnet man die Hüllfläche direkt aus den nichtprägnanten Zeit- und Frequenzkonturen. Beim zweiten Ansatz entfernt man aus dem zeitvarianten FTT-Pegelspektrum die spektral/zeitlichen Energiebeiträge, die schon durch die prägnanten Konturen repräsentiert sind. Das Ergebnis wird geglättet und als Hüllfläche verwendet. Dieses Vorgehen ähnelt einem Verfahren von Aures, der im Spektrum der diskreten Fourier-Transformation tonale Komponenten entfernte, um ein Modell der Klanghaftigkeit zu entwickeln [Aur84]. Die tonalen Komponenten identifizierte er anhand der Ausgeprägt-

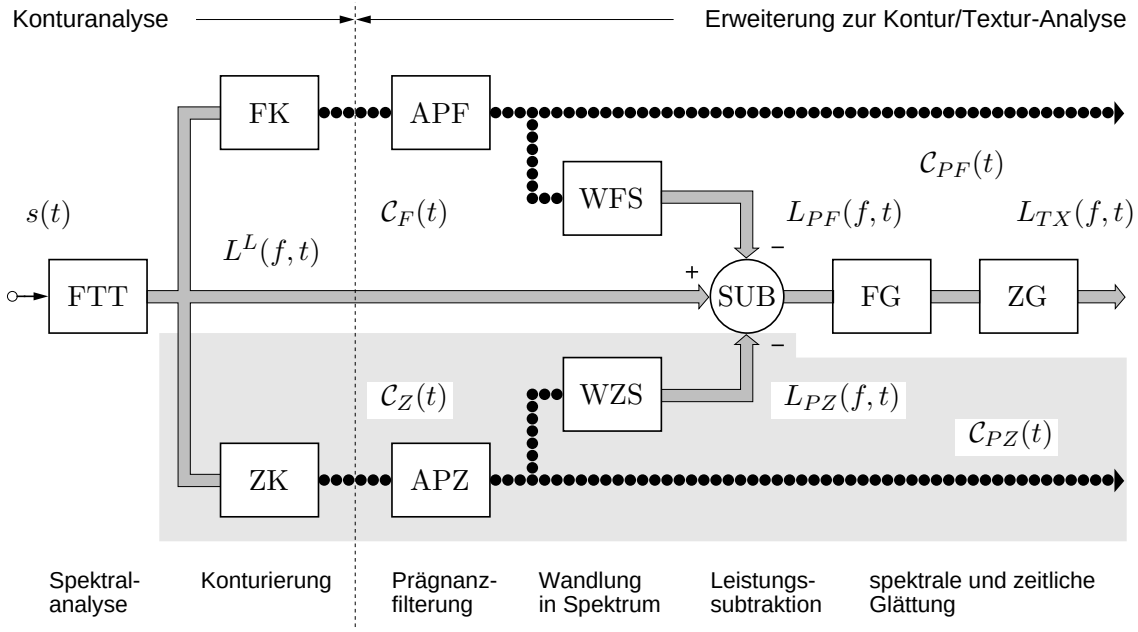

Bild 4.3: *Blockschaltbild zur Gewinnung von Kontur/Textur-Repräsentationen. Die Kontursignale $\mathcal{C}_{PF}(t)$ und $\mathcal{C}_{PZ}(t)$ transportieren die prägnanten Frequenz- bzw. Zeitkonturen. Die Hüllfläche $L_{TX}(f,t)$ repräsentiert die Textur. In einer speziellen Variante wird auf den grau unterlegten Zweig mit der Zeitkonturverarbeitung verzichtet.*

heit von Spitzen im Spektralverlauf.

Im weiteren wird nur der zweite, indirekte Ansatz verfolgt, dessen Vorteile sich später in Abschnitt 4.5 zeigen werden. Bild 4.3 zeigt das Blockschaltbild des resultierenden Verfahrens. Der erste Abschnitt ist identisch mit der bisherigen Konturanalyse eines Eingangssignals $s(t)$, welche aus FTT, Zeitkonturierung (ZK) und Frequenzkonturierung (FK) besteht. Die Konturierung extrahiert aus dem laufzeitausgeglichenen zeitvarianten Pegelspektrum $L^L(f,t)$ die *Kontursignale* $\mathcal{C}_F(t)$ und $\mathcal{C}_Z(t)$. Kontursignale stellen Zeit- oder Frequenzkonturen formal als zeitabhängige Konturpunktmengen dar (Anhang A.3).

Die einzelnen Operationen des zweiten Abschnitts werden hier nur skizziert, sie sind in Anhang B.5 exakt festgehalten. Zur besseren Übersicht bleibt die Darstellung formal im Kontinuum. In einer zeit- und frequenzdiskreten Realisierung präsentieren sich zeitvariante Spektren und daraus abgeleitete Hüllflächen (jeweils graue Pfeilbalken) als Folge von Pegelvektoren mit fester Komponentenanzahl, entsprechend der Anzahl der Analyse- oder Stützfrequenzen. Kontursignale (Punktketten) bilden ebenfalls eine Folge von Vektoren, allerdings mit Frequenz/Pegel-Komponenten und zeitabhängiger Komponentenanzahl.

Die Auswahl der prägnanten Zeit- und Frequenzkonturen (APZ/APF) über die zuvor besprochenen Linienlängenkriterien $\Delta t_p, \Delta f_p, \Delta t_U, \Delta f_U$ führt von $\mathcal{C}_Z(t)$ und $\mathcal{C}_F(t)$ auf $\mathcal{C}_{PZ}(t)$ beziehungsweise $\mathcal{C}_{PF}(t)$. Als Sprachbeispiel dieser vier Kontursignale dient nochmals Bild 4.2. Das zugrunde liegende Pegelspektrum $L^L(f,t)$ ist in Bild 4.4a als Spektrogramm abgebildet. Es läßt sich gut nachvollziehen, daß Bild 4.2a und 4.2b dessen Konturen ausmachen.

Die aufwendigsten Operationen der Texturberechnung bestehen darin, die prägnanten Zeit- und Frequenzkonturen jeweils für sich in Pegelspektren zurückzuwandeln (WZS/

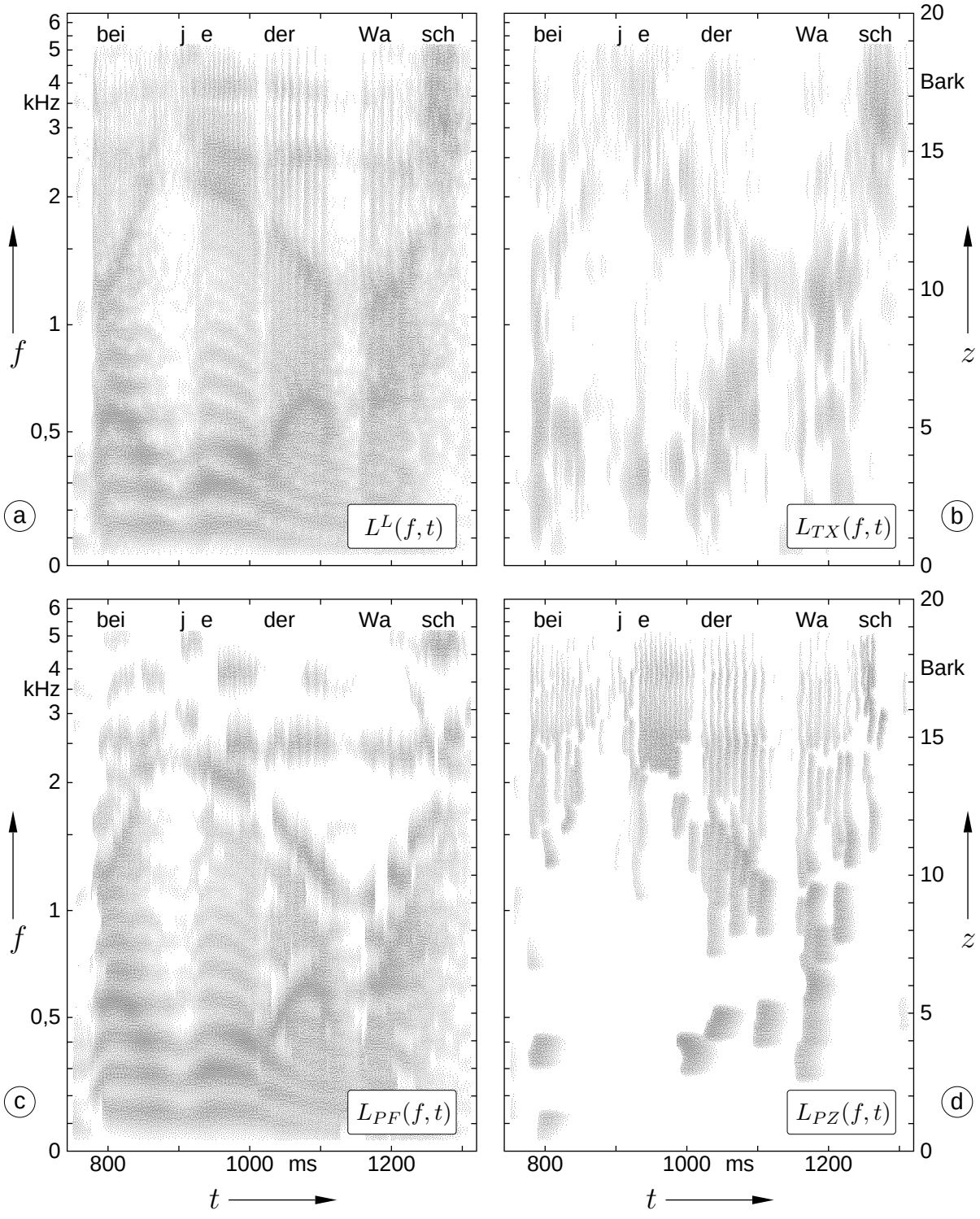

Bild 4.4: *Ermittlung der Texturrepräsentation (b) aus dem Original-FTT-Pegelspektrum (a), dargestellt durch Spektrogramme, Schwärzungsgrad markiert Pegel: Die in FTT-Pegelspektren rückgewandelten prägnanten Frequenzkonturen (c) und prägnanten Zeitkonturen (d) werden aus dem Original abgezogen, das Ergebnis wird geglättet. Eingeblendete Signalnamen beziehen sich auf Bild 4.3.*

WFS). Für die prägnanten Frequenzkonturen wird hierzu jedem Konturpunkt das FTT-Betragsspektrum eines fiktiven stationären Einzeltons zugeordnet, welches bei Frequenzkonturierung genau diesen Punkt ergeben hätte. Das rückgewandelte Spektrum $L_{PF}(f,t)$ zu einem Zeitpunkt ergibt sich aus der Überlagerung der Beiträge aller momentan vorhandenen Punkte. Bild 4.4c zeigt das auf diese Weise aus Bild 4.2c hervorgegangene Spektrogramm.

Für die prägnanten Zeitkonturen wird jedem Konturpunkt eine Impulsantwort des zugehörigen Analysetiefpasses zugewiesen, welche bei Zeitkonturierung genau diesen Punkt ergeben hätte. Der Zeitverlauf des rückgewandelten Spektrums $L_{PZ}(f,t)$ an einer Frequenz ergibt sich aus der zeitrichtigen Überlagerung dieser Impulsantworten mit nachfolgender Betragsbildung. Bild 4.4d zeigt das Spektrogramm, welches nach Umwandlung der prägnanten Zeitkonturen aus Bild 4.2d entsteht.

Bei der Leistungssubtraktion (SUB) erhalten die rückgewandelten Spektren einen Pegelzuschlag ΔL_{PF} beziehungsweise ΔL_{PZ} von wenigen dB, bevor sie in der Leistung vom Originalspektrum abgezogen werden. Dadurch lassen sich die hochpegeligen Bereiche in der Umgebung der prägnanten Konturen sicher entfernen, denn ein negatives Leistungsergebnis wird auf null gesetzt. Das resultierende Residualspektrum durchläuft dann eine spektrale und eine zeitliche Glättungsoperation (FG/ZG). Damit wird die restliche spektral/zeitliche Feinstruktur entfernt. In einer zeit- und frequenzdiskreten Darstellung der Hüllfläche kann man spätestens dann die spektrale Stützstellendichte wie auch das zeitliche Auswerteintervall verringern. Nach den Glättungen liegt schließlich die Texturhüllfläche $L_{TX}(f,t)$ vor. Sie ist in Bild 4.4b als Spektrogramm abgebildet.

4.4 Kontur/Textur-Repräsentation (KTX) von Sprachsignalen

In Bild 4.5 oben ist die Repräsentation des kompletten Sprachsignals zu sehen. Die Darstellung ergibt sich aus der bildlichen Überlagerung von prägnanten Zeit- und Frequenzkonturen mit dem Spektrogramm der Texturhüllfläche. Verglichen mit den kompletten Konturen in Bild 3.4 unten auf S. 68 sind kurze Linienstücke verschwunden, die nun durch die Schwärzungsintensität des Texturspektrogramms wiedergegeben werden. Optisch auffällige Konturen in Bild 3.4 unten sind als prägnante Konturen in etwa beibehalten worden.

Man erkennt in den höherfrequenten Bereichen, daß das Texturspektrogramm die rauschhaften Anteile der Frikative repräsentiert (‘k$\underline{\text{s}}$’, ‘t$\underline{\text{z}}$’, ‘t$\underline{\text{s}}$’, ‘ch’ und ‘sch’). Der Plosiv ‘k’ ganz zu Beginn weist auch bis in tiefere Frequenzen herab einen hohen Anteil an Rauschen auf. Besonders auffällig ist die Abbildung des Frikativs ‘sch’. Schon im Ausschnitt in Bild 4.4b wie auch im Originalspektrogramm in Bild 4.4a zeichnete er sich bei hohen Frequenzen als geschwärzte Fläche ab. An gleicher Stelle finden sich deshalb kaum Beiträge in den prägnanten Konturen (Bild 4.2c und 4.2d).

Stärker geschwärzte Flächen in Bild 4.5 oben existieren jedoch auch in stimmhaften Bereichen, besonders in der Gegend des ersten und zweiten Formanten. Die Prägnanzentscheidung anhand der Linienlänge weist gemäß Abschnitt 4.1.1 immer wieder hochpegelige Kurzverläufe der Textur zu, die zu nicht separat aufgelösten Harmonischen gehören. Wie

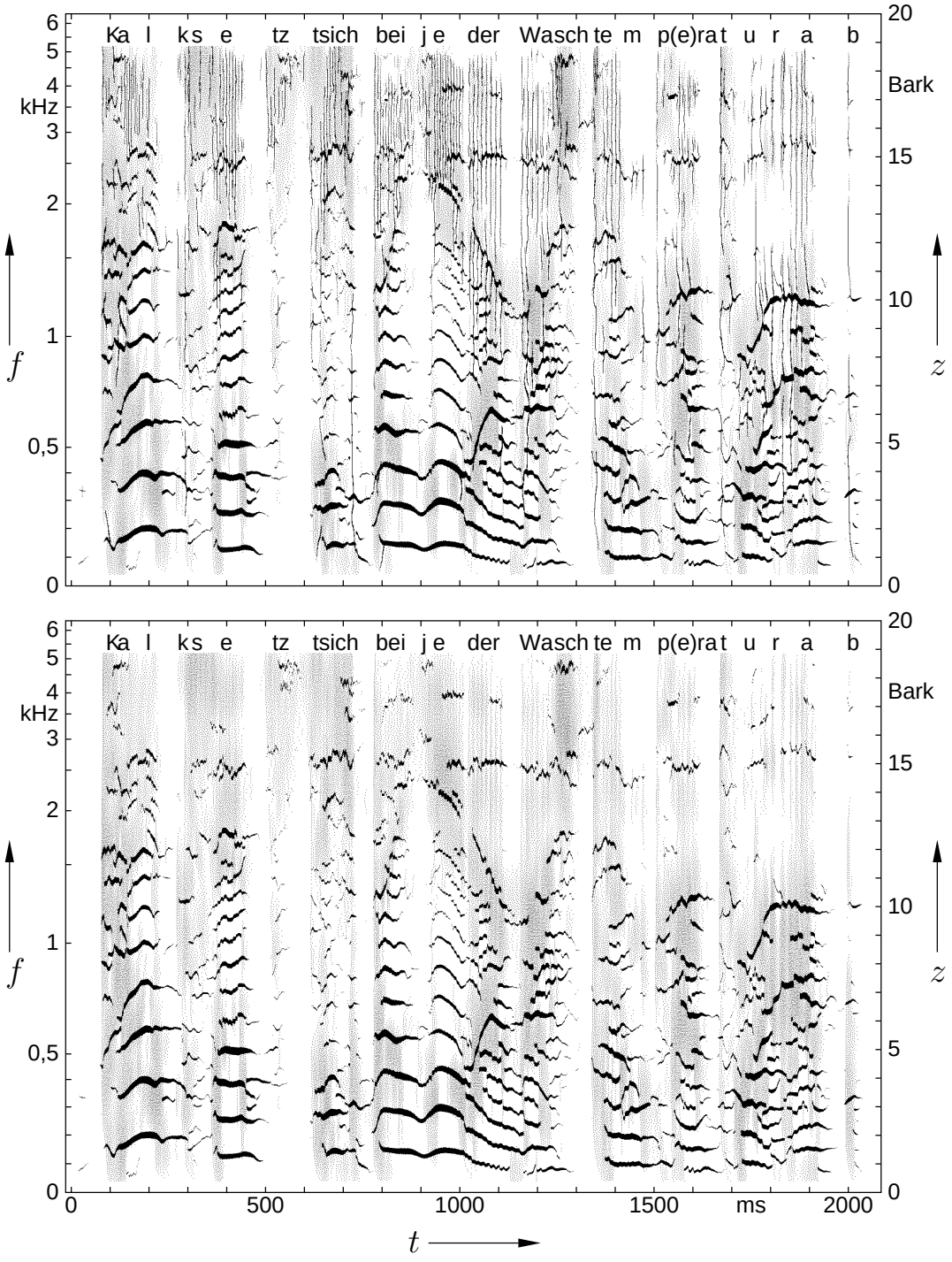

Bild 4.5: *Kontur/Textur-Repräsentation* KTX *(oben) und Kontur/Textur-Repräsentation ohne Zeitkonturen* KTXOZ *(unten) von Sprache: Linien mit horizontaler und vertikaler Vorzugsrichtung sind prägnante Frequenz- bzw. Zeitkonturen, der Schwärzungsgrad der Hinterlegung gibt den Texturpegel wieder. Damit sind tonal, impuls- bzw. rauschhaft empfundenen Anteile in etwa getrennt repräsentiert.* KTXOZ *erfaßt impulshafte Anteile behelfsweise mit in der Textur, gut sichtbar beim abschließenden 'b' an tiefen Frequenzen. Verglichen mit Teiltonzeitmustern sind beide Repräsentationen mit etwas besserer Sprachqualität rückwandelbar.*

dort begründet, stören diese Fehlentscheidungen später bei Signalrekonstruktion kaum.

Im Vergleich von Bild 3.4 unten und Bild 4.5 oben fällt weiterhin auf, daß kurze Linienstücke oft nicht einmal mehr als Schwärzung auftauchen. Dazu zählen insbesondere die regelmäßigen Zeitkonturlinienstücke zwischen ausgeprägten Frequenzkonturen, wie sie bei tieferen Harmonischen des Sprachsignals auftreten. Sie fehlen deshalb in der Textur, weil die rückgerechneten Spektren der Frequenzkonturen durch den Pegelzuschlag ΔL_{PF} so stark sind, daß sie beim Abzug vom Originalspektrum auf schwächere spektral/zeitliche Energiebeiträge verdeckend wirken. Ebenso – wenn auch selten – existiert der umgekehrte Fall von kurzen Frequenzkonturlinien, die in der Umgebung ausgeprägter Zeitkonturen verschwinden. Bei der typischen Skalierung der Zeitachse geht dies in der Darstellung jedoch meist verloren. Beide Fälle sind nicht gravierend, da nur Information in der Nähe der Mit- beziehungsweise Vor- und Nachhörschwellen der prägnanten Konturen verloren geht.

4.5 Kontur/Textur-Repräsentation ohne Zeitkonturen (KTXOZ)

Zeitkonturen erweisen sich leider in den datenreduzierenden Codierverfahren, die später behandelt werden sollen, als unhandlich. Mit einer speziellen Variante der Kontur/Textur-Repräsentation, die auf Zeitkonturen verzichtet, kann man trotzdem eine etwa vergleichbare Verarbeitungsqualität erzielen. Sie wird unter der Bezeichnung 'Kontur/Textur-Repräsentation ohne Zeitkonturen' (KTXOZ) eingeführt.

Wegen des Differentialprinzips des Kontur/Textur-Konzeptes bleibt die grobe spektral/zeitliche Hüllfläche des Sprachsignals unabhängig von der konkreten Definition der Prägnanzkriterien erhalten. Man kann probeweise deren Schwellwerte unerreichbar hoch ansetzen, so daß das gesamte Originalspektrum als Textur erfaßt wird. Zusammen mit der noch zu beschreibenden Signalrekonstruktion für den Texturanteil hätte man eine Art 'Rausch-Vocoder' vorliegen. Obgleich ein derart verarbeitetes Sprachsignal verständlich wäre, ist eine solche Repräsentation natürlich wenig sinnvoll. Denn Sprache ist besonders durch ihre tonalen Anteile, also durch tatsächlich prägnante Frequenzkonturen charakterisiert.

Anders verhält es sich mit der Bedeutung der prägnanten Zeitkonturen. Die Zeitkonturierung wurde eingeführt, weil erst mit ihrer Hilfe bestimmte wahrnehmungsrelevante Anteile im FTT-Spektrum erfaßt werden. Deren Beitrag bei der Repräsentation sprachrelevanter Information ist jedoch geringer, jedenfalls bei (zu) klein gehaltenen Analysebandbreiten (Abschnitte 2.1.4, 3.4.3.2). Wenn man dafür sorgt, daß keine prägnanten Zeitkonturen erkannt werden, finden sich die zugehörigen Anteile in der Textur wieder. Sie werden also nicht übersehen, anders als bei reiner Frequenzkontur-Repräsentation durch Teiltonzeitmuster. Somit sind gravierende Verfälschungen ausgeschlossen, wie etwa die fehlenden Höhen in der Impulsfolge in Abschnitt 2.1.2.

Eine entsprechende Repräsentation ist in Bild 4.5 unten abgebildet. Man erhält sie, indem man im Verfahren nach Abschnitt 4.3 entweder formal $\Delta f_p \rightarrow \infty$ setzt – oder gleich den in Bild 4.3 grau unterlegten Teil wegläßt. Verglichen mit der Kontur/Textur-Repräsentation (KTX) ist kein großer Unterschied zu sehen. Die prägnanten Zeitkonturen, die sich oben als durchgezogene vertikale Linien abheben, sind unten recht gut durch die Schwärzung

des Texturspektrogramms nachgezeichnet.

Auch die erzielten Rekonstruktionergebnisse sind bei Sprache für beide Varianten recht ähnlich. Daraus kann man jedoch nicht schließen, daß das Konzept der Zeitkonturierung vom Kontur/Textur-Konzept abgelöst werden kann. Hinter prägnanten Zeitkonturen verbirgt sich eine Art von Wahrnehmung, die nicht durch Textur repräsentierbar ist. Wird beispielsweise bei einem Einzelimpuls auf Zeitkonturen verzichtet, dann wird er als Teil der Textur später in Form eines Rauschstoßes rekonstruiert. Diesen nimmt man zwar durchaus impulshaft wahr, die Wahrnehmung entspricht aber eben nicht dem 'Klick' des Originals. Die spektral/zeitliche Feinstruktur der Textur ist per Definition immer die eines Rauschens, und somit zufällig und zerklüftet. Bei Zeitkonturen ist sie dagegen wohldeterminiert und glatt.

Daß in den Rekonstruktionsergebnissen trotzdem kein großer Unterschied auffällt, hat drei Gründe. Erstens ist, wie gesagt, die Sprachrelevanz von Zeitkonturen bei klein gehaltenen Analysebandbreiten geringer. Zweitens gibt es beim Prägnanzkriterium auf Basis der Linienlänge die bereits bekannten Fehlentscheidungen. Dadurch werden wohldeterminierte Bereiche des Originalspektrums sowieso des öfteren der Textur zugeordnet und im rekonstruierten Signal durch Rauschbeiträge ersetzt. Drittens bewirken die Phaseninkohärenz-bedingten Störungen des Rekonstruktionsverfahrens ähnliches wie Grund zwei: Wohldeterminierte Bereiche werden öfter mit Rauschbeiträgen rekonstruiert, selbst wenn sie durch Konturen repräsentiert sind (Abschnitte 2.3.4, 5.1.5.3).

4.6 Einstellung der Verfahrensparameter und Verarbeitungsqualität

Die Parameter des Verfahrens in Abschnitt 4.3 für Kontur/Textur-Repräsentationen mit und ohne Zeitkonturen werden nun mittels Signalrekonstruktion eingestellt. Wie schon bei der Konturanalyse gilt grundsätzlich das Einstellziel, für Sprache bestmögliche Verarbeitungsqualität zu gewährleisten. Allerdings werden von vornherein geringfügige Zugeständnisse an den Rekonstruktionsaufwand gemacht, wie gleich noch erläutert wird. Die subjektiv ermittelten Einstellungen werden auch hier objektiviert, indem die Einflüsse der wesentlichen Parameter auf die Verarbeitungsqualität erklärt werden.

Mit dem Einstellziel einer bestmöglichen Verarbeitungsqualität verbinden sich vergleichbare Konsequenzen wie zuvor schon bei Einstellung der Konturanalyse in Abschnitt 3.4. Weil das Kontur/Textur-Konzept zu Spektralanalyse, Konturierung und Rekonstruktion hinzugetreten ist, läuft die subjektive Minimierung der insgesamt wahrnehmbaren Verfälschungen nunmehr zwischen vier Teilkonzepten ab. Klar ist, daß das behelfsmäßige Prägnanzmaß zu Fehlentscheidungen führt, weshalb von vornherein mit gewissen Verfälschungen zu rechnen ist.

Zur Signalrekonstruktion wird deshalb das Verfahren RKHPTX verwendet, das im nächsten Kapitel behandelt wird. Es erweitert die in Abschnitt 3.4.1 kurz vorgestellte Signalrekonstruktion mit heuristischer Phase (RKHP) um die Möglichkeit, Textur mit Hilfe von zeitvariant gefiltertem Rauschen zu rekonstruieren. Vorläufige Versuche haben gezeigt, daß mit einem entsprechend erweiterten Verfahren auf Basis der Signalrekonstruktion mit Originalphasen (RKOP) nur eine unwesentlich bessere Sprachqualität zu er-

Tabelle 4.1: *Verfahrensparameter zur Gewinnung der Kontur/Textur-Repräsentation* KTX *und der Kontur/Textur-Repräsentation ohne Zeitkonturen* KTXOZ*. Die Einstellungen basieren auf Parametervariationen mit dem Ziel bester subjektiver Sprachqualität in Verbindung mit dem Rekonstruktionsverfahren* RKHPTX. $^{+}$ *siehe Text.*

Parameter | Rekonstruktion		**Einstellung**	
		KTX	KTXOZ
Einstellung der Konturanalyse nach Tabelle 3.2 (S. 87)		ZFKII	M-TTZM
Mindestlänge prägnanter Frequenzkonturlinien	Δt_P/ms	25	25
ihre Unstetigkeitstoleranz in der Frequenz	Δf_U/Bark	0,25	0,25
Pegelzuschlag auf ihr rückgewandeltes Spektrum	ΔL_{PF}/dB	3	3
Mindestlänge prägnanter Zeitkonturlinien	Δf_P/Bark	1	(∞)
ihre Unstetigkeitstoleranz in der Zeit	Δt_U/ms	1,25	(0)
Pegelzuschlag auf ihr rückgewandeltes Spektrum	ΔL_{PZ}/dB	3	(0)
3dB-Breite Faltungskern spektrale Glättung	B_{3dB}^{FG}/Bark	0,7	0,7
3dB-Bandbreite Glättungstiefpaß	B_{3dB}^{ZG}/Hz	100	300
Stützstellenabstand Textur	$\Delta\omega_{TX}$/Bark	0,5	0,5
Auswerteintervalldauer	T_A	$1/f_a)^{+}$	1,25ms
Rek. mit heuristischer Phase und Textur (RKHPTX)		•	•

zielen ist. Die unvermeidlichen Verfälschungen wegen des behelfsmäßigen Prägnanzmaßes ähneln nämlich, wie im vorigen Abschnitt und in 4.1.1 ausgeführt, den Phaseninkohärenz-bedingten Störungen. Es macht also wenig Sinn, sich um optimale Phasen für Konturen zu bemühen. Aus dem gleichen Grund sind höhere und damit prinzipiell gehörnähere Analysebandbreiten in der benötigten Konturanalyse weniger geeignet (Abschnitt 3.4).

4.6.1 Durchführung und Ergebnis

Die Verfahrensparameter für die Repräsentationen KTX und KTXOZ wurden vom Autor mit Hilfe des Rekonstruktionsverfahrens RKHPTX auf subjektiv beste Verarbeitungsqualität von Sprache eingestellt. Testsignale und Darbietungweise entsprechen dem Vorgehen bei der Einstellung der Konturanalyse in Abschnitt 3.4.2. Das Ergebnis ist in Tabelle 4.1 dargestellt. Die Einstellung der Transformations- und Konturierungsparameter entspricht ZFKII beziehungsweise M-TTZM aus Tabelle 3.2 auf S. 87. Alle weiteren Verfahrensparameter sind darauf abgestimmt.

KTX und KTXOZ unterscheiden sich – abgesehen von der Verarbeitung von Zeitkonturen – nur in der Auswerteintervalldauer und in der Bandbreite B_{3dB}^{ZG} für die zeitliche Glättung der Texturhüllfläche. Während die Werte der ersten sechs (drei bei KTXOZ) Parameter unterhalb der Konturanalyse in etwa Optimaleinstellungen verkörpern, sind die letzten vier als Grenzwerte zu verstehen. Stärkere Glättungen, gröberer Stützstellenabstand oder deutlich größeres Auswerteintervall würden hörbar werden. Veränderungen mit umgekehrter Tendenz ergeben nur unnötig hohe Datenflüsse.

Verglichen mit reiner Konturrepräsentation ohne Textur entspricht die erzielbare Verarbei-

tungsqualität von Sprachsignalen für beide Repräsentationen etwa der von ZFKII/RKHP. Die Tonalisierungstendenz ist sogar verschwunden. Dafür entsteht manchmal eine zusätzliche, schwache Rauhigkeit, die auf Fehlsortierungen von Frequenzkonturlinien in die Textur beruhen. KTXOZ ist wegen der fehlenden expliziten Zeitkonturverarbeitung und des resultierenden längeren Auswerteintervalls vom Aufwand her viel günstiger.

4.6.2 Zusammenstellung und Erklärung der Parametereinflüsse

Der wichtigste Parameter ist Δt_p. Eine Verkleinerung gegenüber Tabelle 4.1 führt dazu, daß zu viele Frequenzkonturlinien als prägnant gelten. Damit nimmt die Tonalisierungstendenz wieder zu, weil rauschhafte Anteile zu stark durch Frequenzkonturen repräsentiert werden. Eine Vergrößerung dagegen bedeutet, daß zu viele tonale Anteile durch Rauschen ersetzt werden. Dadurch entsteht ein verzerrter Klangeindruck, vergleichbar mit einer Aussteuerungsbegrenzung des Zeitsignals. Ähnliche Effekte – allerdings mit umgekehrter Tendenz – ergeben sich in abgeschwächter Form, wenn man Δf_U verändert.

Frequenzabhängige Einstellungen von Δt_p haben sich in der Praxis von Sprachsignalen nicht bewährt. Bedingt durch die zunehmende Analysebandbreite werden zwar die Linienlängen zu höheren Frequenzen hin tendenziell kürzer. Ein konstantes Δt_p stellt also für höhere Frequenzen auch höhere Anforderungen an die Prägnanz. Dies korrespondiert aber zumindest tendenziell mit einem psychoakustischen Sachverhalt: Die tonale Ausgeprägtheit von Sinustönen nimmt nämlich zu höheren Frequenzen hin ab, ausgehend von einem Maximum im Bereich 1 bis 3 kHz [Fas89]. Daß sie zu tiefen Frequenzen ebenfalls abnimmt, ist zwar nicht realisiert. Weil jedoch unterhalb von 1 kHz bei Sprachsignalen sowieso vornehmlich tonale Anteile vorherrschen, wäre eine bevorzugt tonale Klassifizierung von Frequenzkonturlinien kaum spürbar.

Die Einstellungen von Δf_p und Δt_U sind aus denselben Gründen, derentwegen sich KTX und KTXOZ hinsichtlich der Verarbeitungsqualität kaum unterscheiden, relativ vage. Hier hilft es, wenn man die prägnanten Zeitkonturen separat anhört und ihre bildliche Darstellung interpretiert. Eine Vergrößerung verringert die Anzahl der hörbaren impulshaften Anteile. Bei Verkleinerung von Δf_p erfaßt man schnell die zahlreichen kurzen Zeitkonturlinien zwischen ausgeprägten Harmonischen. Sie als prägnante Konturen auszuweisen macht keinen Sinn (Abschnitt 4.1.2). Die Variation von Δt_U wirkt auch hier mit umgekehrter Tendenz in abgeschwächter Form.

Vergrößert man den Pegelzuschlag ΔL_{PF}, so läßt das aus den prägnanten Frequenzkonturen rückgewandelte Spektrum nach Abzug vom Originalspektrum zu wenig Textur übrig. Dies äußert sich in einer steigenden Tonalisierungstendenz, weil sich die prägnanten Frequenzkonturen nun stärker von der Textur abheben. Im umgekehrten Fall wechseln tonale Energiebeiträge in die Textur. Wie bei zu großem Δt_p hinterläßt dies einen verzerrten Eindruck. Für den Pegelzuschlag der Zeitkonturen wird $\Delta L_{PZ} = \Delta L_{PZ}$ gewählt, weil hier der gleiche Mechanismus zugrunde liegt.

Wenn die angegebenen Werte der Glättungsparameter B_{3dB}^{FG} und B_{3dB}^{ZG} über- beziehungsweise unterschritten werden, wird die sprachrelevante spektral/zeitliche Grobstruktur der Textur beeinträchtigt. Verarbeitete Sprache klingt dann schwächer artikuliert ('verwaschener'). Weil prägnante Zeitkonturen bei KTX separat repräsentiert sind, kann die zeitliche Glättung gegenüber KTXOZ deutlich stärker ausfallen, bevor man eine Beeinträchtigung

wahrnimmt. Zu beachten ist, daß die 3dB-Grenzfrequenz des Glättungstiefpasses der Hälfte der Spezifikation B_{3dB}^{ZG} entspricht.

Bei KTX ist wegen der Zeitkonturen das sehr kleine Auswerteintervall von $T_A = 1/f_a$ nötig, mit f_a als Abtastfrequenz des Quellsignals. Dadurch sollen zum einen keine Zeitkonturpunkte der Konturanalyse übersehen werden. Zum anderen soll ein störendes 'Zeitlagenzittern' bei der Signalrekonstruktion des Zeitkonturanteils (vgl. Abschnitt 3.4.3.4), aber auch bei der Leistungssubtraktion des rückgewandelten Zeitkonturspektrums vermieden werden. Im Falle von KTXOZ, wo Zeitkonturen nicht mehr explizit vorkommen, reicht die Auswerteintervalldauer, die auch bei den TTZM-Verfahren verwendet wird.

Das sehr kleine Auswerteintervall von KTX ist zur Repräsentation der Texturhüllfläche eigentlich unnötig, nach der zeitlichen Glättung könnte man sie zeitlich relativ grob abtasten. Auch für die Frequenzkonturen ist dies übertrieben. Um die Datenströme zu begrenzen, muß man für die drei Teilrepräsentationen angepaßte Abtastungen wählen. Dies wird später bei den Codierverfahren in Kapitel 6 behandelt. Nach der spektralen Glättung können dagegen problemlos die spektralen Stützstellen der Texturhüllfläche ausgedünnt werden. Ihr Abstand $\Delta\omega_{TX}$ beträgt deshalb ein Mehrfaches vom Analysefrequenzabstand des FTT-Spektrums.

4.7 Zusammenfassung

Eine sogenannte Kontur/Textur-Repräsentation wurde eingeführt, die auf den zuvor ermittelten Zeit- und Frequenzkonturen aufbaut. Da sie nichttonale Anteile separat repräsentieren kann, schafft sie bessere Voraussetzungen für eine Datenreduktion. Darüberhinaus ermöglicht sie die vom Aufwand her günstige Alternative, auf explizite Zeitkonturverarbeitung zu verzichten, ohne die entsprechenden Signalanteile ganz zu vernachlässigen. Schließlich fußt sie auf einem ausbaufähigen Konzept, welches einen fortgeschrittenen Entscheidungsprozeß im Terhardtschen Modell der Informationsaufnahme konkretisiert. Darin unterstreicht der aus der visuellen Verarbeitung übernommene Texturbegriff die von Terhardt geforderte Gleichstellung auditiver und visueller Verarbeitungsprinzipien:

Kontur/Textur-Konzept: Für Zeit- und Frequenzkonturen wird jeweils ein Prägnanzkriterium auf Basis eines Prägnanzmaßes eingeführt. Das Prägnanzmaß sollte möglichst gut beschreiben, ob sich ein Konturverlauf in der Wahrnehmung als Einzelobjekt heraushebt. Ein Schwellwert entscheidet zwischen prägnanten und nichtprägnanten Konturen. Die Gesamtheit letzterer wird als Textur bezeichnet und pauschal mit Hilfe einer groben spektral/zeitlichen Hüllfläche repräsentiert. Daraus läßt sich jederzeit eine gleichwertige Textur zurückgewinnen. Insbesondere läßt sich der zugehörige Signalanteil mittels zeitvarianter Filterung von Rauschen darstellen. Weil Konturen schlimmstenfalls durch die Texturhüllfläche erfaßt werden, herrscht eine Art Differentialprinzip. Unabhängig von der Wahl der Prägnanzkriterien bleibt demnach vom zeitvarianten Quellspektrum immer die grobe spektral/zeitliche Hüllfläche erhalten. Je nach Güte der Prägnanzmaße und Wahl der Schwellwerte korrelieren prägnante Frequenzkonturen, prägnante Zeitkonturen und Textur mit den tonal, impulshaft beziehungsweise rauschhaft wahrgenommenen Anteilen des Quellsignals.

Als behelfsmäßige Prägnanzmaße werden die Längen der Konturlinien herangezogen. Wenn Konturen über den Parametersatz ZFKII aus Kapitel 3 bestimmt wurden, stehen Frequenzkonturlinien, die länger als 25 ms dauern, für tonale Anteile. Zeitkonturlinien, die länger als 1 Bark sind, geben dann impulshafte Anteile wieder. Bei Sprachsignalen funktioniert diese Heuristik sehr gut. Fehlentscheidungen rühren vor allem daher, daß kürzere Frequenzkonturlinien von tonalen Modulationen nicht als zusammengehörig erkannt werden. Das Kontur/Textur-Konzept könnte in Zukunft wesentlich kompliziertere Prägnanzkriterien beherbergen. Dann könnten beispielsweise die prägnanten Frequenzkonturen die zeitvarianten Spektraltonhöhen des Signals modellieren.

Das Verfahren zur Gewinnung der Kontur/Textur-Repräsentation faßt die nichtprägnanten Konturen indirekt zur Texturhüllfläche zusammen. Dadurch werden nur die prägnanten Konturen benötigt, deren spektral/zeitliche Energiebeiträge aus dem ursprünglichen FTT-Spektrum entfernt werden. Das spektral und zeitlich geglättete Residualspektrum entspricht der gewünschten Texturhüllfläche. Durch dieses indirekte Vorgehen ergibt sich neben der allgemeinen Kontur/Textur-Repräsentation (KTX) eine Variante, die weniger aufwendig und zur Datenreduktion leichter zu handhaben ist. Bei der Kontur/Textur-Repräsentation ohne Zeitkonturen (KTXOZ) verzichtet man nämlich auf Zeitkonturierung, wodurch die zugehörigen Signalanteile im Sinne des Differentialprinzips automatisch der Textur zufallen. KTXOZ stellt gleichzeitig eine vereinfachte Methode zur tonal/geräuschhaft-Trennung dar, in der nicht extra zwischen rausch- und impulshaft unterschieden wird.

Die Parameter der beiden Varianten wurden im Selbstversuch eingestellt. Ziel der Einstellung war es, die mit dem Rekonstruktionsverfahren RKHPTX (Kapitel 5) erreichbare Qualität bei Sprachsignalrepräsentation zu optimieren. Sie unterscheidet sich bei beiden Varianten nur wenig, was an der unvollkommenen Prägnanzentscheidung wie auch an den Beschränkungen von RKHPTX liegt. Weil bei KTOXZ die Zeitkonturen behelfsmäßig über die Textur berücksichtigt sind, ist das Ergebnis bei Sprachsignalen besser als bei reiner Frequenzkonturverarbeitung (M-TTZM/TTSD) und entspricht etwa der Qualität einer Zeit/Frequenzkontur-Repräsentation mit suboptimaler Rekonstruktion (ZFKII/RKHP).

Kapitel 5

Rekonstruktion des Signals aus Konturen

Mit der Einführung von Zeitkonturen und dem Kontur/Textur-Konzept sind Repräsentationen aufgestellt worden, die neue Verfahren zur Rekonstruktion des ursprünglichen Signals erfordern. Insbesondere bei Zeitkonturen besteht zunächst Unklarheit, wie überhaupt eine Rekonstruktion anzusetzen ist. Bereits vorhandene Verfahren, die Teiltonsynthesen in ihren beiden Varianten, eignen sich nicht dafür. Überdies führen sie schon bei Frequenzkonturen wahrnehmbare Störungen ein. Das Hauptziel dieses Kapitels liegt deshalb darin, für Zeit- und Frequenzkonturen ein optimales Verfahren zu erarbeiten, welches möglichst keine wahrnehmbaren Störungen einbringt. Damit beschäftigt sich der erste Abschnitt ausführlich. Eine Erweiterung für Kontur/Textur-Repräsentationen läßt sich dann vergleichsweise einfach hinzufügen. Insgesamt werden so die Beschreibungen der Rekonstruktionsverfahren nachgereicht, die bereits in den vorigen beiden Kapiteln für Parametereinstellung und Qualitätsvergleich der Repräsentationen eine Rolle spielten.

5.1 Entwicklung eines optimalen Rekonstruktionsverfahrens

Wie gelangt man zu einem Rekonstruktionsverfahren für beide Konturtypen, das keine wahrnehmbaren Störungen einbringt, das in diesem Sinne also optimal ist? Offenbar muß man ‘nur’ dafür sorgen, daß das rekonstruierte Signal möglichst wieder die gleiche Konturrepräsentation wie die des Originals liefern kann. Dieser Ansatz beinhaltet allerdings zwei Schwachpunkte. Erstens muß vorausgesetzt werden können, daß die Konturrepräsentation alle gehörrelevanten Signaleigenschaften erfaßt. Sonst sind unrepräsentierte, aber hörbare Störungen möglich. Damit fehlt die Sicherheit, Zeit- oder Frequenzkonturen für sich alleine rekonstruieren zu können. Zweitens ist unklar, mit welcher Genauigkeit Soll- und Ist-Konturen zur Deckung zu bringen sind. Kleine Unterschiede würden sicher nicht wahrnehmbar sein. Dennoch ist mit diesem Ansatz ein Verfahren denkbar, welches Fortsetzungen des rekonstruierten Signals passend iteriert (vgl. [Gri84]).

Der stattdessen gewählte Ansatz entwickelt die Rekonstruktion zusammen mit der Konturanalyse in mehreren Schritten, in denen Art und Wahrnehmbarkeit von sukzessive

hinzutretenden Störungen gut zu kontrollieren sind. Zuerst wird die formale Rücktransformation der FTT eingeführt. Damit läßt sich im zweiten Schritt ein Verarbeitungsrahmen aufspannen, mit dem man ein Signal über sein komplexes FTT-Spektrum codieren und für das Gehör fehlerlos zurückgewinnen kann. Die Einführung eines Synthesefensters spielt hierbei eine wichtige Rolle. Der Rahmen wird im dritten Schritt modifiziert, um nur noch die Orte der Konturen zu codieren. Im vierten Schritt läßt sich die unvollständige Version RKOP des gesuchten Rekonstruktionsverfahrens abspalten. RKOP benötigt zusätzlich zu den Konturen noch deren Phasenverläufe im analysierten FTT-Spektrum. Bis hierhin ist immer noch eine nahezu verfälschungsfreie Verarbeitung realisierbar.

Im fünften Schritt geht es darum, die benötigten Konturphasen direkt aus den Konturen abzuleiten. Weil dies kompliziert ist, wird nur eine suboptimale Version RKHP des gesuchten Rekonstruktionsverfahrens realisiert, welche wohldefinierte Störungen einbringt. Für die Verfahrensparameter beider Versionen liefert der sechste Schritt die Einstellungen nach. Zum Rekonstruktionsverfahren ohne eigene wahrnehmbare Störungen wird am Ende also 'nur' eine optimale Phasenrekonstruktion fehlen. Immerhin kann ein Nachweis skizziert werden, daß sie existiert – womit RKOP die Bedeutung erringt, ihr Ergebnis simulieren zu können. Ein siebenter Schritt vergleicht RKOP und RKHP mit den Teiltonsynthesen und ihren aus Kapitel 2 bekannten Störungen. Eine ausführliche Zusammenfassung des Entwicklungsprozesses findet sich ab S. 148.

5.1.1 Einführung der FTT-Rücktransformation (RFTT)

Wegen der frequenzabhängigen Analysebandbreite entzieht sich das FTT-Spektrum einer einfachen Rücktransformation, wie sie für fixe Bandbreiten bekannt ist [Fla72]. Würde die Bandbreite direkt frequenzproportional zunehmen ('constant-Q'), so könnte man immerhin auf den Formalismus der umkehrbaren Wavelet-Transformation zurückgreifen [Rio91, Vet92]. Umkehrbare Transformationen mit flexiblen Bandbreitenverläufen stützen sich auf zeit- und frequenzdiskrete Formalismen für Filterbanksysteme [Smi87, Pue91, Nay93, Kap93]. Das Problem ist, daß die dort vorrangig angestrebte fehlerlose Signalrekonstruktion zu Fensterfunktionen führt, die bei der FTT nicht erwünscht sind. Es wird also ein umgekehrter Entwurfsansatz benötigt, bei dem von einer vorgegebenen Fensterfunktion ausgegangen wird. Fehlerlosigkeit ist nicht erforderlich, solange der Fehler nicht wahrgenommen werden kann.

Für eine Transformation, die Gemeinsamkeiten mit der FTT aufweist [Owe88], beschrieben Owens und Murphy einen iterativen Ansatz zur Rücktransformation [Owe89]. Allerdings führt er zu deutlichen Übertragungsverzerrungen. Zwar nimmt der nachfolgend vorgestellte, analytische Lösungsansatz der FTT-Rücktransformation [Mum91] auch Fehler in Kauf. Sie können jedoch unhörbar gehalten werden.

5.1.1.1 Ansatz und Korrekturfunktion

Die Rücktransformation des komplexen FTT-Spektrums wird über seine Systeminterpretation $s_{\omega_A}(t)$ aus Gl. (1.6) entwickelt. Zuerst passiert dieses Signal ein Korrektursystem, das später individuell für jede Analysefrequenz ω_A festzulegen ist. Es sei linear und zeitinvariant mit der Impulsantwort $x_{\omega_A}(t)$. Dann wird über alle möglichen Analysefrequenzen

das Rückintegral der Fourier-Transformation angesetzt. So erhält man das rücktransformierte Signal

$$\hat{s}(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} [s_{\omega_A}(t) * x_{\omega_A}(t)] \cdot e^{j\omega_A t} d\omega_A. \tag{5.1}$$

Die Gleichung beschreibt nach Einsetzen von Gl. (1.6) ein lineares und zeitinvariantes Gesamtsystem mit $s(t)$ als Eingang und $\hat{s}(t)$ als Ausgang. Seine Impulsantwort $g(t)$ läßt sich wie folgt gewinnen:

$$\begin{aligned}
\hat{s}(t) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \left[\left(s(t) \cdot e^{-j\omega_A t} \right) * h_{\omega_A}(t) * x_{\omega_A}(t) \right] \cdot e^{j\omega_A t} d\omega_A && (5.2)\\
&= \frac{1}{2\pi} \int_{-\infty}^{\infty} s(t) * \left(h_{\omega_A}(t) \cdot e^{j\omega_A t} \right) * \left(x_{\omega_A}(t) \cdot e^{j\omega_A t} \right) d\omega_A && (5.3)\\
&= s(t) * g(t) && (5.4)\\
g(t) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} [h_{\omega_A}(t) * x_{\omega_A}(t)] \cdot e^{j\omega_A t} d\omega_A && (5.5)
\end{aligned}$$

Bei den Faltungsoperationen gilt bezüglich einer Modulation $e^{j\omega_A t}$ das Distributivgesetz. Deshalb verliert das Eingangssignal $s(t)$ in der ersten Umformung seinen gegenläufigen Modulationsfaktor. Befreit von einer ω_A-Abhängigkeit kann es in der zweiten Umformung vor das Integral gezogen werden. Dabei wurde die Vertauschbarkeit eines (impliziten) Faltungsintegrals über t mit dem expliziten Integral über ω_A angenommen. Übrig bleibt die Impulsantwort, für die nochmals das Distributivgesetz angewandt wurde.

Damit das Eingangssignal möglichst unverändert übertragen wird, sollte eine geeignete Dimensionierung des Korrektursystems aus $g(t)$ näherungsweise einen zeitverschobenen Dirac-Impuls machen. Wie gleich nachgewiesen wird, gelingt das bei der Wahl

$$x_{\omega_A}(t) = \frac{e^{-j\omega_A t_{max,0}}}{h_{max,\omega_A}} \cdot \delta\left(t - (t_{max,0} - t_{max,\omega_A})\right). \tag{5.6}$$

Darin stellt $(h_{max,\omega_A}, t_{max,\omega_A})$ den Scheitelpunkt der Fensterfunktion an der Analysefrequenz ω_A dar (Abschnitt 3.3.4). Mit Hilfe der Entnormierungsvorschrift Gl. (3.3) berechnet er sich aus dem Scheitelpunkt (h^N_{max}, T_{max}) der normierten Fensterfunktion (Tabelle B.1) und der frequenzabhängigen Analysebandbreite B_{3dB}:

$$h_{max,\omega_A} = 2 \cdot \pi B_{3dB}(\omega_A) \cdot h^N_{max}, \qquad t_{max,\omega_A} = \frac{T_{max}}{\pi B_{3dB}(\omega_A)}. \tag{5.7}$$

Das Korrektursystem $x_{\omega_A}(t)$ entspricht dem Laufzeitausgleich aus Gln. (3.10), (3.12), zuzüglich eines komplexen Drehfaktors und einer Normierungskonstante. Einsetzen von Gl. (5.6) in Gl. (5.5) führt auf die spezielle Gesamtimpulsantwort

$$\begin{aligned}
g(t) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{h_{\omega_A}\left(t - (t_{max,0} - t_{max,\omega_A})\right)}{h_{max,\omega_A}} \cdot e^{-j\omega_A t_{max,0}} \cdot e^{j\omega_A t} d\omega_A && (5.8)\\
&\approx \delta(t - t_{max,0}). && (5.9)
\end{aligned}$$

Sei zunächst die Fensterfunktion unabhängig von ω_A, so daß man den Bruch aus dem Integral herausziehen kann. Übrig bleibt eine Fourier-Rücktransformation der Funktion

$e^{-j\omega_A t_{max,0}}$ über ω_A, die mit dem zeitverschobenen Dirac-Impuls exakt korrespondiert. Bei $t = t_{max,0}$, wo die zugehörige Distribution maximal ist, stellt der herausgezogene Bruch das Impulsgewicht auf den Wert eins ein. Ohne die Wirkung des Drehfaktors $e^{-j\omega_A t_{max,0}}$ läge der Impuls bei $t = 0$, wo übliche Fensterfunktionen im Zähler des Bruches normalerweise noch null sind. Dann wäre $g(t) \approx 0$. [1]

Bei frequenzabhängiger Fensterfunktion kann man den Bruch nicht mehr vor das Integral ziehen. Weil er aber in unmittelbarer Umgebung von $t = t_{max,0}$ für alle ω_A auf dem Wert eins bleibt, kann man im Ergebnis dennoch mit einem Impuls rechnen. Das Verhalten des Bruches mag nochmals ein Blick auf Bild 3.7 auf S. 77 verdeutlichen: Die Rechtsverschiebung der Fensterfunktionen im Zähler um $t_{max,0} - t_{max,\omega_A}$ legt die Scheitelpunkte frequenzunabhängig auf den Zeitpunkt $t_{max,0}$. Eine Normierung auf den individuellen Maximalwert h_{max,ω_A} durch den Nenner stellt die Höhe eins ein. Damit sind neben dem Drehfaktor auch die Bedeutung von Laufzeitausgleich und Normierungsfaktor begründet.

5.1.1.2 Übertragungseigenschaften und auditive Transparenz

Lineare Verzerrungen des Gesamtsystems lassen sich wegen der nichtexakten Betrachtung bisher nicht ausschließen. Hier hilft die Übertragungsfunktion weiter, die als Fourier-Transformierte der Gesamtimpulsantwort Gl. (5.8) umgeformt wird:

$$
\begin{aligned}
G(\omega) &= \mathcal{F}\{g(t)\} && (5.10)\\
&= \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{\mathcal{F}\{h_{\omega_A}(t-(t_{max,0}-t_{max,\omega_A}))\cdot e^{j\omega_A t}\}}{h_{max,\omega_A}} \cdot e^{-j\omega_A t_{max,0}} d\omega_A && (5.11)\\
&= \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{\left[H_{\omega_A}(\omega')\cdot e^{-j\omega'(t_{max,0}-t_{max,\omega_A})}\right]_{\omega'=\omega-\omega_A}}{h_{max,\omega_A}} \cdot e^{-j\omega_A t_{max,0}} d\omega_A && (5.12)\\
&= \frac{e^{-j\omega t_{max,0}}}{2\pi}\int_{-\infty}^{\infty} \frac{H_{\omega_A}(\omega-\omega_A)}{h_{max,\omega_A}} \cdot e^{j(\omega-\omega_A)t_{max,\omega_A}} d\omega_A && (5.13)\\
&\approx e^{-j\omega t_{max,0}}. && (5.14)
\end{aligned}
$$

In der ersten Umformung zieht die Fourier-Transformation in das Integral hinein, weil Vertauschbarkeit der Integrale angenommen wird. Die nachfolgenden beiden Umformungen bedienen sich der Korrespondenz $H_{\omega_A}(\omega) = \mathcal{F}\{h_{\omega_A}(t)\}$ sowie der zeitlichen und der spektralen Verschiebungsregel der Fourier-Transformation [Mar82]. Fourier-Transformation von Gl. (5.9) ergibt Gl. (5.14).

Das Integral in Gl. (5.13) entzieht sich einer weiteren analytischen Behandlung, weil die Analysebandbreite über Gl. (1.2) in komplizierter Weise von ω_A abhängt. Allerdings kann man die Integration durch eine Reihensumme über diskrete Analysefrequenzen approximieren. Dies ist unkritisch, weil signifikante Beiträge des Integranden entlang der Integrationsvariablen ω_A nicht beliebig schnell schwanken können. Positive Analysefrequenzen ω_{A_i} reflektieren später eine Reihe (komplexer) Bandpässe ähnlich Bild 1.1, deren Ausgänge

[1] Dies gilt für Fensterfunktionsgrade $n > 1$. Bei der Fensterfunktion vom Typ P1 spielt der Drehfaktor keine Rolle, da er durch $t_{max,0} = 0$ zu eins wird. Allerdings kommt dann nur das halbe Impulsgewicht zur Geltung. Weil der herausgezogene Bruch für $t < 0$ sicher null ist, wird die Dirac-Distribution um $t = 0$ zur Hälfte abgeschnitten.

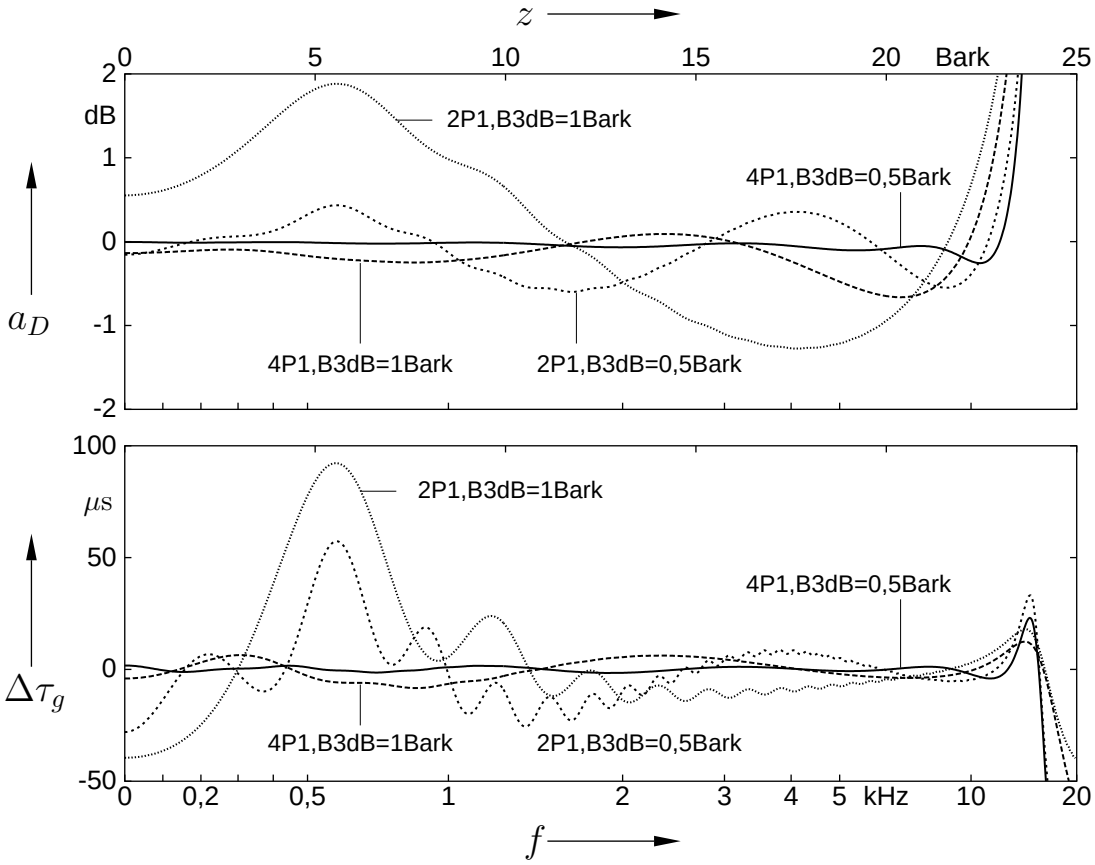

Bild 5.1: *Frequenzverlauf von Dämpfung a_D (oben) und Gruppenlaufzeitschwankung $\Delta\tau_g$ (unten) für die FTT mit Rücktransformation, dargestellt für verschiedene Paarungen von Fensterfunktion und Analysebandbreite. Berechnet nach Gl. (5.15) für Analysefrequenzen bis* 24 Bark *(entsprechend* 15500 Hz*, bedingt Abknicken der Kurven) bei hinreichend kleinem Analysefrequenzabstand. Die Schwankungen sind ein Charakteristikum der Kombination von Fenstertyp und Verlauf der Analysebandbreite über der Frequenz. Bei den benötigten Kombinationen bleiben sie unhörbar (siehe Text).*

überlagert werden. Bei Annahme einer oberen Grenzfrequenz bleibt ihre Anzahl endlich. Dadurch geht Gl. (5.13) in eine numerisch auswertbare Formel über:

$$G(\omega) \approx \frac{e^{-j\omega t_{max,0}}}{2\pi} \sum_{\substack{\omega_A=\pm\omega_{A_i} \\ i}} \frac{H_{\omega_A}(\omega-\omega_A)}{h_{max,\omega_A}} \cdot e^{j(\omega-\omega_A)t_{max,\omega_A}} \Delta\omega_A(\omega_A). \tag{5.15}$$

Bild 5.1 zeigt Auswertungsbeispiele als Frequenzgänge von Dämpfung a_D und Gruppenlaufzeitschwankung $\Delta\tau_g = \tau_g - t_{max,0}$. Zu sehen sind jeweils die Kurven für zwei Fensterfunktionen 2P1 und 4P1 bei Analysebandbreiten von $B_{3dB} = 0,5$ und 1 Bark. Die Analysefrequenzen reichen bis herauf zu 15,5 kHz. Ihr Abstand wurde klein genug gewählt, um das Integral in Gl. (5.13) bei entsprechend beschränkter Grenzfrequenz ohne einen sichtbaren Fehler zu approximieren. Erst bei Abständen von mehr als einer halben Analysebandbreite würde sich eine zusätzliche regelmäßige Welligkeit überlagern. Dies reflektiert die Tatsache, daß Spektralanteile zwischen zwei zu weit voneinander entfernten Bandpässen nicht mehr so gut übertragen werden.

Die charakteristischen Merkmale der Kurven sind insgesamt nur vom Verlauf der Analy-

sebandbreite über der Frequenz, vom Fenstertyp und von der Grenzfrequenz abhängig. Eine kleinere Analysebandbreite oder ein höherer Fensterfunktionsgrad verringern die Schwankungen. Die Schwankungen der Gruppenlaufzeit sind unhörbar, da die zugehörige Wahrnehmungsschwelle auch in kritischen Fällen deutlich über 100 μs liegt [Pre82]. Für die Schwankung der Dämpfung kann man grob einen Wert von 1 dB als Wahrnehmungsschwelle annehmen [Zwi67]. Dieser ist leicht einzuhalten, wenn bei Analysebandbreiten nicht über 1 Bark Fensterfunktionsgrade von $n \geq 4$ verwendet werden. Für 0,5 Bark reicht bereits $n \geq 2$. Das Gesamtsystem ist dann also auditiv transparent.

5.1.1.3 Zusammenfassung und Ausblick

Die Gleichungen (5.1), (5.6) und (5.7) führen zur FTT (1.6) die Rücktransformation (RFTT) ein. Der Formalismus von Hin- und Rücktransformation ist nicht nur für gehörorientierte Anwendungen brauchbar. Prinzipiell lassen sich damit Kurzzeitspektraltransformationen mit beliebigen stetigen Analysebandbreitenverläufen über der Frequenz behandeln. Freilich bleiben die leichten lineare Verzerrungen zu beachten, deren Ausmaß offenbar von der Variation der Analysebandbreite im Bereich der wirksamen Bandbreite eines Analysefilters abhängt. Im gehörorientierten Kontext der FTT bleiben sie für Fensterfunktionsgrade $n \geq 4$ mit Sicherheit unhörbar.

5.1.2 Rahmen für eine FTT-Codierung

Um Codierungen auf der Grundlage des FTT-Spektrums zu entwerfen, reicht der FTT-RFTT-Formalismus noch nicht aus. Eine zeit- und frequenzdiskrete Formulierung ist erwünscht, die ein sparsam abgetastetes FTT-Spektrum verarbeitet. Beispielsweise braucht man den Zeitverlauf an einer Analysefrequenz nicht viel häufiger abzutasten, als dies die Bandbreite des Analysefilters erfordert. Die resultierenden Abtastwerte kann man aber nicht direkt in der RFTT verwenden. Stattdessen muß der ursprüngliche Verlauf mit Hilfe eines geeigneten Synthesefensters interpoliert werden.

Allgemein kontrolliert ein Synthesefenster die spektral/zeitliche Unschärfe, mit der sich Zeit/Frequenz-lokale Manipulationen des FTT-Spektrums im rekonstruierten Signal niederschlagen. Aktive Manipulationen kommen beispielsweise später im Rahmen einer zeitvarianten Filterung vor (Abschnitt 5.2.1). Im Kontext von Codierungen werden Manipulationen durch Quantisierungsfehler verursacht, zu denen auch die Folgen der Diskretisierung von Zeit- und Frequenzachse zählen. Die Wirkung des Synthesefensters wird hier deshalb als *spektral/zeitliche Fehlerformung* [2] bezeichnet. Damit Quantisierungsfehler nicht wahrgenommen werden, sollte ihr Energiebeitrag spektral/zeitlich so geformt werden, daß er sich möglichst unter den Vor-, Mit- und Nachhörschwellen des Nutzsignals konzentriert. Ungünstige Synthesefenster 'verschmieren' den Fehler zu sehr in zeitlicher oder spektraler Richtung, wodurch er hörbar stört. Deshalb wird ein gehörangepaßtes Synthesefenster angestrebt, das im wesentlichen dem Analysefenster entspricht. Ein richtig eingestelltes

[2] Dieser Begriff ist vom 'noise shaping' [Sch79, Tri79, Fla79] zu unterscheiden, welcher die aktive Steuerung der Quantisierung von spektralen Abtastwerten beinhaltet. Dabei soll das insgesamt vorhandene Quantisierungsgeräusch spektral so geformt werden, daß es unter die Mithörschwellen des Nutzsignals fällt. Die spektral/zeitliche Fehlerformung eines Synthesefensters hat hierauf allerdings wesentlichen Einfluß.

Analysefenster bewährt sich nämlich mit seinen spektral/zeitlichen Eigenschaften in einer ähnlichen Situation: Störungen aufgrund der gegenseitige Verdeckung von Konturen fallen nicht mehr auf, weil sie erst unterhalb der Hörschwelle eintreten können (Abschnitte 2.2, 2.4 und 3.4.3).

Im folgenden werden spektrale Diskretisierung, Synthesefenster und zeitliche Diskretisierung in den FTT-RFTT-Formalismus eingeführt. Weil sich das bisherige FTT-Spektrum für Codierungszwecke als ungeeignet erweist, wird auf eine betragsgleiche Form mit modifizierter Phase übergangen. Der resultierende Codierungsrahmen wird abschließend mit den Ansätzen etablierter Audiocodierungen verglichen.

5.1.2.1 Spektrale Diskretisierung und Einführung Synthesefenster

Zur Berechnung der FTT-RFTT-Gesamtübertragungsfunktion wurde eine spektrale Diskretisierung bereits behandelt, bei der sich Integrale über ω_A in unkritische Reihensummen über diskreten Analysefrequenzen verwandelten. Entsprechend erhält man aus Gl. (5.2) mit positiven ω_{A_i}:

$$\hat{s}(t) = \sum_{\substack{\omega_A=\pm\omega_{A_i}\\ i}} \left[\left(s(t)\cdot e^{-j\omega_A t}\right) * h_{\omega_A}(t) * x_{\omega_A}(t)\right]\cdot e^{j\omega_A t}\cdot\frac{\Delta\omega_A(\omega_A)}{2\pi} \tag{5.16}$$

$$= \sum_{\substack{\omega_A=\omega_{A_i}\\ i}} 2\mathrm{Re}\left\{\left[\left(s(t)\cdot e^{-j\omega_A t}\right) * h_{\omega_A}(t) * x_{\omega_A}(t)\right]\cdot e^{j\omega_A t}\right\}\frac{\Delta\omega_A(\omega_A)}{2\pi}. \tag{5.17}$$

Weil $s(t)$ und $\hat{s}(t)$ reell sein sollen, kann man auf negative Analysefrequenzen verzichten und, wie in der Umformung geschehen, den doppelten Realteil ansetzen. Damit liegt die Beschreibung eines realisierbaren, mehrkanaligen Übertragungsystems vor. Ein Summand in Gl. (5.17) entspricht einem Kanal, der in Bild 5.2a für eine anonyme Analysefrequenz ω_A dargestellt ist. Der Kanalabstand $\Delta\omega_A$, also der (frequenzabhängige) spektrale Abtastabstand des zu übertragenden FTT-Spektrums, darf dabei nicht zu groß werden. Er macht sich sonst als überlagerte Welligkeit in der Gesamtübertragungsfunktion bemerkbar, wie in Abschnitt 5.1.1.2 beschrieben.

Ein Synthesefenster läßt sich einführen, indem man das ursprüngliche Analysefenster als Reihenschaltung von einem Analyse- und einem Synthesetiefpaß auffaßt. Die Einzelimpulsantworten entsprechen einem neuen Analysefenster $h^A_{\omega_A}(t)$ und dem Synthesefenster $h^S_{\omega_A}(t)$. Auch für das Korrektursystem ist eine vergleichbare Aufspaltung möglich. Bild 5.2b geht aus 5.2a hervor, indem man

$$h_{\omega_A}(t) \rightarrow h^{A*S}_{\omega_A}(t) = h^A_{\omega_A}(t) * h^S_{\omega_A}(t), \tag{5.18}$$

$$x_{\omega_A}(t) \rightarrow x^{A*S}_{\omega_A}(t) = x^A_{\omega_A}(t) * x^S_{\omega_A}(t) \tag{5.19}$$

ersetzt und $1 = e^{-j\omega_A t}\cdot e^{+j\omega_A t}$ anwendet. Nach Abschnitt 1.4.2 bilden gegenläufig angeordnete Modulatoren mit zwischengeschaltetem Tiefpaß einen zeitinvarianten komplexen Bandpaß. Bezüglich der gestrichelten Mittelline in Bild 5.2b zerfällt die Anordnung also in zwei Bandpässe mit gleicher Mittenfrequenz. Über alle Frequenzen bildet die linke Hälfte eine Analysefilterbank, die rechte eine Synthesefilterbank.

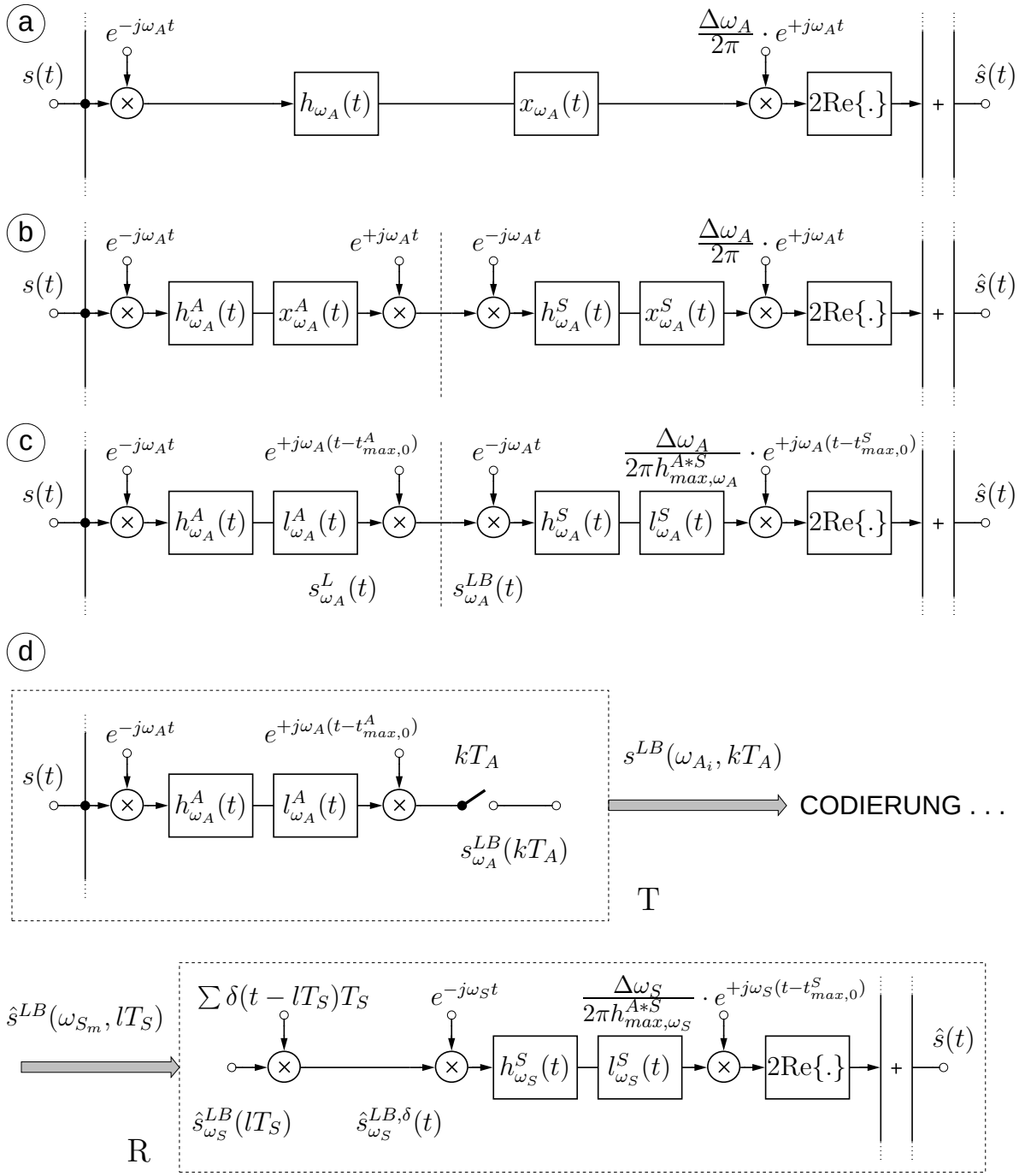

Bild 5.2: *Kanalschema der FTT mit Rücktransformation (a) und seine Umwandlung in ein zur Codierung geeignetes Transformationspaar* T *und* R *(d) über zwei Zwischenschritte (siehe Text). Bei geeigneter Parameterwahl bleibt die Verarbeitung auditiv transparent.*

5.1.2.2 Übergang auf FTT-Bandpaßspektrum und Spezifikation Synthesefenster

Die Aufteilung der Korrektur auf zwei Teilsysteme ist von der Spezifikation von Analyse- und Synthesetiefpaß unabhängig. Es muß lediglich gelten, daß die Reihenschaltung der Korrekturen auf die Reihenschaltung von Analyse- und Synthesetiefpaß abgestimmt ist. Durch diese Freiheit kann man ein spezielles Signal $s^{LB}_{\omega_A}(t)$ zwischen den Filterbänken in Bild 5.2b konstruieren. Bei dem Analysekorrektursystem

$$x^A_{\omega_A}(t) = l^A_{\omega_A}(t) \cdot e^{-j\omega_A t^A_{max,0}} \tag{5.20}$$

sei das Laufzeitglied $l^A_{\omega_A}(t)$ nach Gln. (3.10), (3.12) an das neue Analysefenster $h^A_{\omega_A}(t)$ angepaßt. Der komplexe Drehfaktor ist in Bild 5.2c bereits im Modulator hinter dem Laufzeitglied berücksichtigt. Weil der Modulator den Betrag nicht verändert, stimmt $|s^{LB}_{\omega_A}(t)|$ mit dem laufzeitausgeglichenen FTT-Betragsspektrum $|s^L_{\omega_A}(t)|$ vor dem Modulator überein. Das neue Signal ist ein zeitparalleler Schnitt des *laufzeitausgeglichenen FTT-Bandpaßspektrums* $s^{LB}(\omega, t)$ bei $\omega = \omega_A$. Dessen Besonderheit wird später erläutert.

Eine bestimmte Wahl des Analysefensters hinsichtlich Typ und Analysebandbreite B^A_{3dB} legt so den Analyseteil fest. Im Syntheseteil kann noch das Synthesefenster $h^S_{\omega_A}(t)$ vorgegeben werden, mit dessen Wahl schließlich auch die Synthesekorrektur $x^S_{\omega_A}(t)$ festliegt. Im Kontext dieser Arbeit wird ein Tiefpaß mit einer Synthesebandbreite $B^S_{3dB} > B^A_{3dB}$ gewählt, wie in Abschnitt 5.1.6 begründet wird. Außerdem sollte es sich um einen laufzeitflachen Typ handeln (Familie Bn, Abschnitt 3.3.3), weil sich dann die Berechnung der Synthesekorrektur vereinfacht. Auch mit diesen Nebenbedingungen bleibt eine gehörangepaßte Fehlerformung realisierbar. Um dieselben Normierungsgleichungen wie beim Analysetiefpaß zu nutzen, wird von einer Grundverstärkung $|H^S_{\omega_A}(0)| = 2$ ausgegangen.

Um die Synthesekorrektur explizit zu erhalten, müßte man sie innerhalb der Reihenkorrektur Gl. (5.19), die der FTT-RFTT-Formalismus auf der Grundlage der Reihenimpulsantwort Gl. (5.18) bestimmt, von der bereits bekannten Analysekorrektur isolieren. Vereinfachend ist nun, daß ein laufzeitflacher Synthesetiefpaß mit etwas größerer Bandbreite die Impulsantwort des Analysetiefpasses kaum in seiner Form verändert. Die Wirkung des Synthesetiefpasses im Gesamtsystem läßt sich dadurch als bloße Laufzeit annähern. Bei laufzeitflachen Tiefpässen entspricht überdies die Laufzeit im Durchlaßbereich etwa dem Scheitelzeitpunkt der Impulsantwort (Bild 3.6c,d). Der Scheitelzeitpunkt der Reihenimpulsantwort addiert sich nunmehr aus denen der Einzelimpulsantworten, die Scheitelhöhe wird vom Analysetiefpaß und der Grundverstärkung zwei des Synthesetiefpasses bestimmt:

$$t^{A*S}_{max,\omega_A} \approx t^A_{max,\omega_A} + t^S_{max,\omega_A}, \qquad h^{A*S}_{max,\omega_A} \approx 2h^A_{max,\omega_A}. \tag{5.21}$$

Dadurch kann man auch für das Synthesefenster $h^S_{\omega_A}(t)$ den Laufzeitausgleich nach Gln. (3.10), (3.12) ansetzen, den ein Laufzeitglied $l^S_{\omega_A}(t)$ realisiert. Die Synthesekorrektur lautet damit

$$x^S_{\omega_A}(t) \approx \frac{l^S_{\omega_A}(t) \cdot e^{-j\omega_A t^S_{max,0}}}{h^{A*S}_{max,\omega_A}}. \tag{5.22}$$

Setzt man diese in Bild 5.2b ein, so liegt schließlich auch der Syntheseteil von 5.2c fest. Die Ausgangssignale von Bild 5.2a,b einerseits und 5.2c andererseits sind formal nicht mehr

identisch, weil das Gesamtsystem nun nicht mehr exakt korrigiert ist. Die zusätzlich eingebrachten linearen Laufzeit- und Dämpfungsverzerrungen sind bei den hier betrachteten Dimensionierungen aber auditiv irrelevant.

5.1.2.3 Zeitliche Diskretisierung

Während der Übergang auf ein Kanalschema die spektrale Diskretisierung schon beinhaltet, müssen zeitliche Abtastung und Rekonstruktion in Bild 5.2d durch separate Operationen im Kanal berücksichtigt werden. Der Analyseteil gleicht demjenigen aus Bild 5.2c, jedoch steht das Spektrum nun nur noch zu diskreten Auswertezeitpunkten $t = k \cdot T_A$ zur Verfügung. Im allgemeinen kann das Auswerteintervall T_A abhängig von der Analysefrequenz sein. Über die vorhandenen Analysefrequenzen ω_{A_i} betrachtet entsteht in Bild 5.2d eine spektral wie zeitlich abgetastete Version des FTT-Bandpaßspektrums $s^{LB}(\omega, t)$. Dies entspricht der Transformation

T

$$s^{LB}_{\omega_A}(t) = \left[\left(s(t) \cdot e^{-j\omega_A t}\right) * h^A_{\omega_A}(t) * l^A_{\omega_A}(t)\right] \cdot e^{j\omega_A (t - t^A_{max,0})}, \tag{5.23}$$

$$s^{LB}(\omega_{A_i}, kT_A) = s^{LB}_{\omega_A}(t)\Big|_{\omega_A = \omega_{A_i}, t = kT_A}. \tag{5.24}$$

Das Kanalschema in Bild 5.2d ist gegenüber 5.2a-c unterbrochen. Eine Codierung des FTT-Bandpaßspektrums, wie auch immer sie geartet sein mag, führt im allgemeinen zu einer fehlerbehafteten Decodierung $\hat{s}^{LB}(\omega_{S_m}, lT_S)$. Seine Synthesefrequenzen ω_{S_m} müssen nicht mehr unbedingt in den Analysefrequenzen ω_{A_i} ihr exaktes Pendant finden, und auch das Syntheseintervall T_S stimmt nicht notwendig mit dem Auswerteintervall T_A überein. Bis auf weiteres mag es allerdings genügen, die Entsprechungen $\{\omega_{A_i}\} = \{\omega_{S_m}\}$, $\{kT_A\} = \{lT_S\}$ und $\hat{s}^{LB}(\omega, t) = s^{LB}(\omega, t)$ anzunehmen. Die rechte Hälfte des Schemas in Bild 5.2d beschreibt die Rücktransformation

R

$$\hat{s}^{LB,\delta}_{\omega_S}(t) = \sum_{l=-\infty}^{\infty} \hat{s}^{LB}(\omega_S, lT_S) \cdot \delta(t - lT_S) \cdot T_S, \tag{5.25}$$

$$\hat{s}(t) = \sum_{\substack{\omega_S = \omega_{S_m} \\ m}} 2\mathrm{Re}\Big\{ \left[\left(\hat{s}^{LB,\delta}_{\omega_S}(t) \cdot e^{-j\omega_S t}\right) * h^S_{\omega_S}(t) * l^S_{\omega_S}(t)\right] \cdot e^{j\omega_S (t - t^S_{max,0})} \Big\} \frac{\Delta\omega_S(\omega_S)}{2\pi h^{A*S}_{max,\omega_S}}. \tag{5.26}$$

Der Syntheseteil greift auf die diskreten Abtastwerte der Decodierung zu, aus denen er das zeitkontinuierliche Signal $\hat{s}^{LB,\delta}_{\omega_S}(t)$ konstruiert. Darin bilden die zeitlichen Abtastwerte die Impulsgewichte einer Dirac-Impulsfolge. Betrachtet man dieses Signal gegenüber dem unabgetasteten Signal $\hat{s}^{LB}_{\omega_S}(t)$, dann weist sein Fourier-Spektrum periodisch Spiegelspektren im Abstand $2\pi/T_S$ auf (z.B. [Pap86]). Der nachfolgende Teil des Kanals wirkt als komplexer Bandpaß, der die Spiegelspektren abschneidet. Idealerweise steht das unabgetastete Signal leicht zeitverzögert (genau um $t^S_{max,0}$) hinter dem letzten Modulator in Bild 5.2d an. [3]

[3] Liegt bereits ein zeitdiskretes Gesamtsystem vor, so geht die Abtastung im Analyseteil in eine Unterabtastung über. Der Unterschied im Syntheseteil betrifft die Ersetzung der Dirac-Impulse $\delta(t - lT_S)$

Die zeitliche Diskretisierung führt aus zwei Gründen Fehler in die Verarbeitung ein. Erstens können die relativ flachen Flanken der Analysefilter keine ideale Bandbegrenzung sicherstellen, so daß ein bestimmtes Maß an Aliasing bei der Abtastung unvermeidlich ist. Zweitens kann das Synthesefilter ähnlichen Typs die Spiegelspektren nicht völlig unterdrücken. Beide Fehler können natürlich bei entsprechend häufiger Abtastung beliebig klein gehalten werden.

5.1.2.4 Besondere Eigenschaften des FTT-Bandpaßspektrums

Behandelt man Zeit und Frequenz $\omega = \omega_A$ in den Definitionsgleichungen (3.14) und (5.23) als gleichberechtigte, kontinuierliche Dimensionen, dann hängen das laufzeitausgeglichene FTT-Spektrum und sein zugehöriges Bandpaßspektrum über

$$s^L(\omega, t) = s^{LB}(\omega, t) \cdot e^{-j\omega(t - t^A_{\max,0})} \tag{5.27}$$

zusammen. Während sie sich im Betrag nicht unterscheiden, geht in ihre Phasenrelation die Absolutzeit t ein. Weil sich $s^{LB}(\omega, t)$ über zeitinvariante komplexe Bandfilter berechnen läßt, korrespondieren seine zeitverschobenen Versionen mit entsprechend zeitverschobenen Versionen des Original- oder Rekonstruktionssignals. Die Korrespondenz kann aber nicht auch gleichzeitig für das FTT-Spektrum $s^L(\omega, t)$ gelten, da nach Gl. (5.27) dessen Phase von der willkürlichen Festlegung eines Ursprunges von t abhängt. Nur die Phase des FTT-Bandpaßspektrums ist also tolerant gegen Zeitverschiebungen zwischen Analyse- und Synthesezeit. [4] Dies ist wesentlich, weil Speicherung, Übertragung und Weiterverarbeitung unvorhersagbare Zeitverzögerungen einführen.

Eine weitere wichtige Eigenschaft des Bandpaßspektrums zeigt sich, wenn man statt der Analysefrequenz den Analysezeitpunkt $t = t_A$ festhält. Betrachtet man nun das über der Dimension ω definierte Signal $s^{LB}_{t_A}(\omega)$, dann handelt es sich nach Anhang C.3 um ein geglättetes Signal. Zwei spektrale Abtastwerte innerhalb einer Analysebandbreite können sich demnach nur wenig unterscheiden. Im FTT-Spektrum $s^L_{t_A}(\omega)$ ist dies keinesfalls so, wie man anhand von Gl. (5.27) erkennt: Während sich $s^{LB}_{t_A}(\omega)$ bei Variation von ω nur wenig ändert, kann sich die Phase im Exponentialfaktor bei fortgeschrittenem t sehr schnell ändern ('Auseinanderlaufen der Phasen'). Das Betragsquadrat beider Signale ist wiederum geglättet (Anhang C.3).

Das abgetastete FTT-Bandpaßspektrum ist deshalb auf einfache Weise spektral interpolierbar (Real- und Imaginärteil jeweils für sich). Interpolationsfehler kann man wie Quantisierungsfehler behandeln, sie unterliegen der spektral/zeitlichen Fehlerformung durch das Synthesefenster. Verschiedene Repräsentationen des Bandpaßspektrums infolge unterschiedlicher Analyse- und Synthesefrequenzsätze lassen sich deshalb einfach ineinander überführen, im einfachsten Fall durch Treppenstufenapproximation. Damit sind Analyse- und Syntheseseite nicht nur unabhängig von der absoluten Zeit, sondern auch von der Wahl der konkreten Frequenzstützstellen. Insbesondere ist dadurch kein absoluter Gleichlauf zwischen den Modulatoren der beiden Seiten mehr erforderlich.

in Gl. (5.25) und Bild 5.2d durch gewichtete Einheitsimpulse $f_a \cdot \delta(n - lT_S \cdot f_a)$, mit f_a als Abtastrate des Gesamtsystems.

[4] Im Kontext von Kurzzeitspektraltransformationen mit frequenzunabhängigen Fenstern wird das Bandpaßspektrum deshalb als Kurzzeitspektrum mit 'gleitendem' Ursprung oder Zeitrahmen bezeichnet [Alm83, Cro83].

5.1.2.5 Vergleich mit Ansätzen bekannter Audiocodierungen

Klassische Ansätze für Audiocodierungen im Spektralbereich sind Transformations- und Teilbandcodierungen [Por80, Tri79, Cro83, Jay84, Kra89], beide Formen sind prinzipiell ineinander überführbar. Bekannte Transformationscodierungen [Zel77, Kra85, Bra87, Joh88, Sot91] gehen von einer diskreten Fourier-Transformation oder von vergleichbaren Blocktransformationen aus und sind vom Rechenaufwand besonders günstig zu realisieren. Dafür können sie nur frequenzunabhängige Analyse- und Synthesefenster realisieren. Teilbandcodierungen [The87, Bra94] gehen von Filterbanksystemen aus, die wie die FTT-Codierung individuelle Kanäle mit Analysefilter, Abtastung und Synthesefilter enthalten. Zwar ist bei einigen Entwurfsverfahren die Vorgabe frequenzabhängiger Bandbreiten möglich, nicht jedoch die Vorgabe gehörgerechter Fenstertypen (vgl. Abschnitt 5.1.1).

Mit der Gehöranpassung von Bandbreiten, Analyse- und Synthesefenstertyp stellt der FTT-Codierungsrahmen ein Novum dar. Die Anpassung hat allerdings auch Nachteile zur Folge. Zusätzlich zu den unhörbaren linearen Verzerrungen der FTT-RFTT-Gesamtübertragungsfunktion führen, wie bereits behandelt, zeitliche und spektrale Diskretisierung neue Fehler ein. Damit diese nicht hörbar werden, muß man einerseits das Verhältnis

$$r_t(\omega_A) = \frac{1/T_A(\omega_A)}{B^A_{3dB}(\omega_A)} \tag{5.28}$$

von Kanalabtastrate $1/T_A(\omega_A)$ zu Kanalbreite, vertreten durch die Analysebandbreite $B^A_{3dB}(\omega_A)$, hinreichend groß wählen. Andererseits ist das Verhältnis

$$r_f(\omega_A) = \frac{\Delta\omega_A(\omega_A)/2\pi}{B^A_{3dB}(\omega_A)} \tag{5.29}$$

von Kanalabstand $\Delta\omega_A(\omega_A)$ zu Kanalbreite hinreichend klein zu halten. Vorläufige Versuche mit einem Analysefenster 4P1 mit $B^A_{3dB} = 0{,}5$ Bark und einem Synthesefenster B4 mit $B^S_{3dB} = 0{,}7$ Bark ergaben, daß für nicht hörbare Fehler etwa $r_t \geq 3$ und $r_f \leq 1$ zu wählen ist. In den etablierten Codierungen dagegen eliminieren sich meist die Diskretisierungsfehler im Gesamtsystem. So ist mit Hilfe von QMF-Filterbänken [The87, Bra94], TDAC-Filterbänken [Pri86, Pri87, Lok90] oder Blocktransformationen mit bestimmten Fenster/Vorschub-Kombinationen [Kra88, Fei89] perfekte Rekonstruktion möglich.

Mit der Forderung, den Diskretisierungsfehler klein zu halten, konkurriert die Forderung, eine möglichst geringe Rohdatenrate zwischen T und R verarbeiten zu müssen. Die Summenabtastrate, kumuliert über alle Kanäle, liefert hierfür ein Maß. Das Verhältnis

$$R = \frac{r_t}{r_f} \tag{5.30}$$

reflektiert die Erhöhung der Summenabtastrate gegenüber dem informationstheoretischen Minimum, der sogenannten *kritischen* Abtastung, und liegt gemäß vorläufigem Versuchsergebnis etwa bei $R \geq 3$. Bei einer diskreten Fourier-Transformation ist, zum Vergleich, perfekte Rekonstruktion bereits bei kritischer Abtastung möglich, wenn auch nur bei Verwendung von Rechteckfenstern. Bei TDAC- und QMF-Filterbänken bleibt trotz kritischer Abtastung sogar noch ein gewisser Spielraum bei der Wahl der Analyse- und Synthesefenster.

Allgemein jedoch bedingen niedrige Rohdatenraten bei perfekter Rekonstruktion, daß die Gestaltungsmöglichkeit für die zeitlichen und spektralen Eigenschaften der Fenster erheblich einschränkt ist. Diese Eigenschaften sind für eine effiziente Audiocodierung aber auch wichtig, und hier kann der FTT-Codierungsrahmen Vorteile verbuchen. Mit einem Rechteckfenster als Analysefenster beispielsweise erhält man eben kein gehörorientiertes Spektrum, so daß kaum eine sichere Irrelevanzreduktion (Abschnitt 1.2) möglich ist. Mit einem Rechteckfenster als Synthesefenster werden bereits geringe Quantisierungsfehler ungünstig verschmiert, entweder in zeitlicher oder in spektraler Richtung, je nach Fensterlänge. Dadurch können sie leicht unter den Hörschwellen des Nutzsignals hervortreten. Das sogenannte 'Vorecho' ist ein bekanntes Problem bei Transformationscodierungen. Schließlich ist nicht-perfekte Rekonstruktion des FTT-Codierungsrahmens keinesfalls ein Nachteil für eine Codierung, die nur auditive Transparenz anstrebt. Ob tatsächlich der Vorteil der besseren Gehöranpassung den Nachteil der höheren Rohdatenrate aufwiegt, kann im Rahmen dieser Arbeit aber nicht abschließend bewertet werden.

5.1.2.6 Zusammenfassung

Aufbauend auf dem FTT-RFTT-Formalismus wurde ein neues Transformationspaar T und R eingeführt. Es beschreibt, wie man zu einem spektral und zeitlich abgetasteten, laufzeitausgeglichenen komplexen FTT-Bandpaßspektrum gelangt und wie man daraus das Zeitsignal wieder zurückgewinnt. Das FTT-Bandpaßspektrum entspricht betragsmäßig dem FTT-Spektrum, dessen willkürliche Phase sich nicht für Codierungszwecke eignet. Insgesamt liegt ein Rahmen für Codierungen auf der Grundlage des FTT-Spektrums vor, der auditiv transparent gehalten werden kann.

Von wichtiger Bedeutung ist das Synthesefenster in R. Es kontrolliert allgemein, inwieweit Zeit/Frequenz-lokale Manipulationen im FTT-Spektrum sich spektral und zeitlich im rekonstruierten Signal verteilen. Weil es in etwa dem gehörangepaßten Analysefenster in T entspricht, können codierungsbedingte Quantisierungsfehler besonders gut unter die Hörschwellen des Nutzsignals geformt werden. Die Gehöranpassung von T und R kann möglicherweise sogar den Nachteil etwas relativieren, daß gegenüber Ansätzen etablierter spektraler Audiocodierungen zunächst mit einer etwas höheren Rohdatenrate zu rechnen ist.

T und R ergeben ein Analyse/Synthese-Filterbankschema mit komplexen Modulatoren, Tiefpässen und Laufzeitgliedern. Zeitdiskrete Realisierungen solcher Strukturen beschreibt Anhang B.2. Wie sich T und R als besonderer Fall einer Wavelet-Transformation formulieren lassen, zeigt Anhang D.2.

5.1.3 FTT-Codierung durch konturgesteuerte Auswahl der Abtastwerte

Als Zwischenschritt wird nun innerhalb des FTT-Codierungsrahmens ein Codierungsteil spezifiziert, der die im Frequenz/Zeit-Kontinuum definierten Konturen (Anhang A.1) ins Spiel bringt. Sie dienen hier jedoch nur als Indikator für relevante Abtastwerte des FTT-Bandpaßspektrums, ohne dabei selbst Gegenstand der Codierung zu sein (Bild 5.4). Der Nutzen dieses Zwischenschrittes liegt darin, daß zum Abgleich und zur Störener-

giebetrachtung noch mit der Sichtweise als FTT-Codierung argumentiert werden kann. Gleichzeitig aber wird der Übergang auf eine rein konturbasierte Codierung vorbereitet, die vom Abtastraster des Analyse- wie auch des Synthesespektrums unabhängig ist. Daraus resultiert später die konzeptionelle und realisierungstechnische Eigenständigkeit einer Rekonstruktion aus Konturen. Nachfolgend wird nur ein funktionaler Überblick gegeben, formale Beschreibungen der einzelnen Operationen liefert Anhang B.6.

Bild 5.3a zeigt das Blockschaltbild, das sich aus Bild 5.2d ableitet. Das zu codierende FTT-Bandpaßspektrum $s^{LB}(\omega_{A_i}, kT_A)$ ist entlang des grauen, zweigeteilten Signalweges von der Analyse- bis zur Syntheseseite über demselben Abtastraster $\{(\omega_{A_i}, kT_A)\} = \{(\omega_{S_m}, lT_S)\}$ definiert. Im Gegensatz zu den Notwendigkeiten für eine niedrige Summenabtastrate in Abschnitt 5.1.2.5 gilt das Raster als vergleichsweise feinmaschig, da eine Datenreduktion später allein über die Codierung der Konturen erfolgt. Dadurch kann man auch den Zeitrasterabstand $T_S = T_A$ als frequenzunabhängig festlegen.

Nach der Aufspaltung des Signalweges gelangen die Abtastwerte in ein Frequenzkontur- beziehungsweise ein Zeitkontur-gesteuertes Sieb (FS/ZS). Sie können nur dort passieren, wo auch Konturen als Indikator maximaler lokaler Energiekonzentration verlaufen. Alle übrigen Abtastwerte werden zu null angenommen. Die benötigten Kontursignale $\mathcal{C}_F(t)$, $\mathcal{C}_Z(t)$ (Anhang A.3) sind aus demselben FTT-Bandpaßspektrum $s^{LB}(\omega, t)$ abgeleitet zu denken, dessen Abtastung die Transformation T ausgibt. Sie bleiben aber über einer kontinuierlichen Zeit/Frequenz-Ebene definiert. Man kann nämlich bei den später codierten und somit quantisierten Konturen nicht davon ausgehen, daß sie nach Decodierung die auszuwählenden Rasterorte noch exakt treffen. Trotzdem ist zu gewährleisten, daß eine Konturlinie lückenlos auf das Syntheseraster abgebildet wird. Dazu benötigt man formal zunächst die Approximationen der decodierten Linien ins Kontinuum hinein.

Nichttrivial bei einer Rasterierung aus dem Kontinuum heraus ist nun, daß keiner Frequenzkonturlinie mehr als ein Rasterort zum gleichen Zeitpunkt beziehungsweise keiner Zeitkonturlinie mehr als ein Rasterort an der gleichen Frequenz zugewiesen werden darf. Sonst nämlich stimmt an diesen Stellen die zu rekonstruierende Energiedichte nicht mehr. Bild 5.4 verdeutlicht anhand willkürlicher Frequenz- und Zeitkonturverläufe, wie die Zuordnung zu den Rasterorten $\{(\omega_{S_m}, lT_S)\}$ festgelegt ist.

Durch jeden Rasterort verläuft parallel zur Konturierungsachse des jeweiligen Konturtyps eine Kontrollstrecke, die auf halbem Weg zum Nachbarrasterort endet. Wird sie von einem Konturverlauf berührt, dann gilt der betreffende Rasterort als ausgewählt. Dies impliziert, daß Konturen im Zeit/Frequenz-Kontinuum als geschlossene Linienstücke mit einer Mindestlänge auftreten, nicht aber als singuläre Punkte. Damit außerdem die Linienstücke nicht einer Berührung mit der Kontrollstrecke ausweichen können, müssen sie im Falle der Frequenzkonturen mindestens einen Zeitrasterabstand, im Falle der Zeitkonturen mindestens einen Frequenzrasterabstand überspannen. Da das FTT-Leistungsspektrum durch die Wirkung des Analysefensters zeitlich wie spektral (Anhang C.3) geglättet ist, ist dies bei ausreichend dicht gewähltem Raster praktisch sichergestellt.

In Bereichen der Zeit/Frequenz-Ebene, wo Zeit- und Frequenzkonturen sehr nahe beieinander liegen oder sich gar überschneiden, liegt eine Doppelrepräsentation vor (Abschnitte 3.2.1 und 3.2.2). Um die Übertragungsqualität nicht zu beeinträchtigen, darf die Energiedichte des FTT-Spektrums in diesen Bereichen nur von den Abtastwerten eines Konturtyps berücksichtigt werden. Dies erreicht eine gegenseitige Maskierung (MSK). Da die

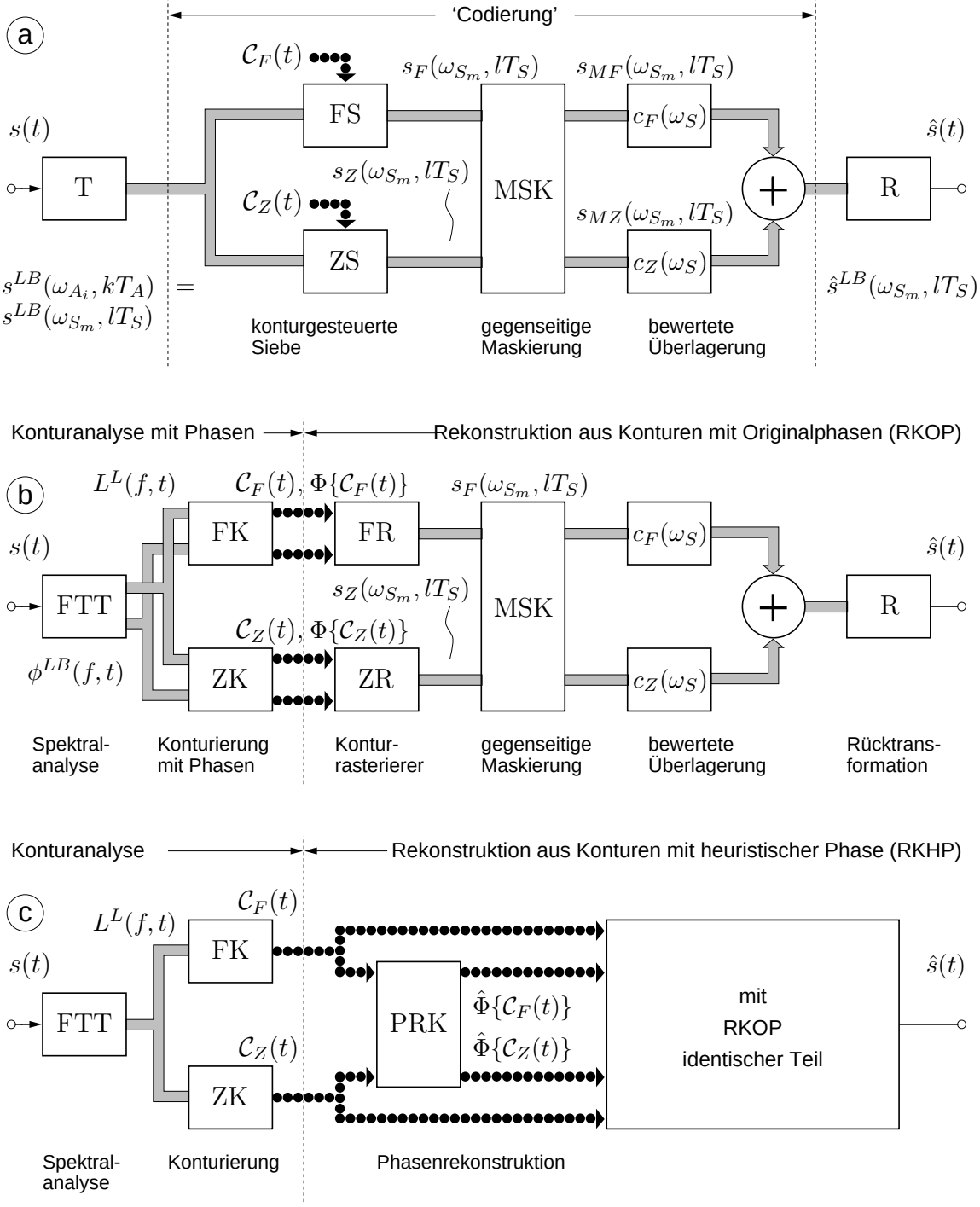

Bild 5.3: *Herleitung der Rekonstruktion aus Konturen: Begonnen wird mit einer speziellen Form der FTT-Codierung, die spektrale Abtastwerte abseits von Konturen zu null setzt (a). Dieses Schema wird in eine Konturanalyse mit Phasen und das Rekonstruktionsverfahren* RKOP *umgeformt (b). Erneute Umwandlung ergibt die bekannte Konturanalyse und das eigenständige Rekonstruktionsverfahren* RKHP *(c). Während die ersten beiden Schemata auditiv fast transparent sind, hängt die Übertragungsqualität des dritten deutlich vom Aufwand der Phasenrekonstruktion ab.*

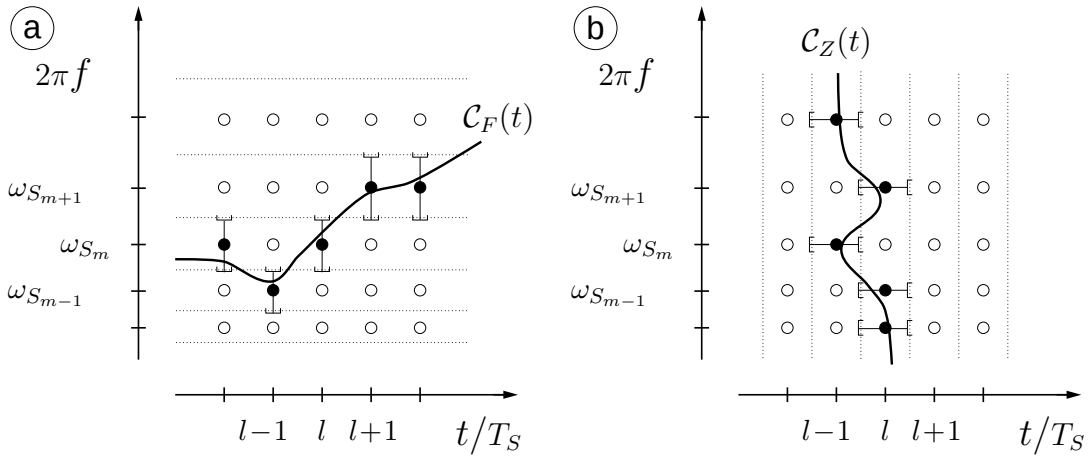

Bild 5.4: *Auswahl der Rasterorte durch Konturen im Zeit/Frequenz-Kontinuum: a) für eine Frequenzkonturlinie, b) für eine Zeitkonturlinie. Ein Rasterort (Kreis) gilt als ausgewählt (gefüllter Kreis), wenn eine Linie die zugehörige Kontrollstrecke (Linien mit Endmarken, nur bei Auswahl eingezeichnet) berührt.*

Bedeutung der Frequenzkonturen im allgemeinen die der Zeitkonturen übersteigt, wird behelfsweise nur eine Maskierung der Zeitkonturen durch die Frequenzkonturen berücksichtigt. Ähnlich wie bei der Operation WFS in Abschnitt 4.3 berechnet man pro Rasterzeitpunkt ein rein stationäres FTT-Betragsspektrum, welches genau die Frequenzkontur-gesiebten Abtastwerte hervorgebracht hätte. Zeitkontur-gesiebte Abtastwerte, die betragsmäßig nicht wesentlich darüber liegen, werden nullgesetzt. Eine Vernachlässigung der umgekehrten Maskierung ist nicht immer angemessen, wie etwa im Bereich von ausgeprägten Zeitkonturen mit Frequenzkonturen als Begleiterscheinung.

Die Bewertungsfaktoren $c_F(\omega_S)$ und $c_Z(\omega_S)$ in Bild 5.3a korrigieren frequenzabhängig den Energieverlust, der von den nicht vom Sieb durchgelassenen Abtastwerten herrührt. Die Festlegung dieser Faktoren in Anhang B.6 gründet sich auf zwei Einzelbetrachtungen mit Hilfe eines stationären Sinustons beziehungsweise eines Dirac-Impulses als Eingangssignal $s(t)$. Dies sind die einfachsten Signale, bei denen jeweils nur eine Konturart vorkommt. Mit ihnen können die zwei Signalwege jeweils für sich auf Verzerrungslosigkeit und Verstärkung eins abgeglichen werden. Da die verarbeiteten Abtastwerte der beiden Signalwege – auch wegen der Operation MSK – unterschiedliche spektral/zeitliche Regionen betreffen, lassen sie sich einfach addieren und schließlich der Rücktransformation zuführen. Das Synthesefenster in R formt gemäß Abschnitt 5.1.2 die Störenergie, die infolge der Sprünge beim Nullsetzen von Abtastwerten eingebracht wird, unter die Hörschwellen des rekonstruierten Nutzsignals.

5.1.4 Rekonstruktion aus Konturen mit Originalphasen (RKOP)

Nachdem die Konturen bisher nur als Indikator weiterzureichender FTT-Abtastwerte dienten, sollen sie nun auch ihre Werte selbst festlegen. Damit verwandelt sich die FTT-Codierung in eine Konturcodierung, die sich in den bekannten Teil der Konturanalyse und den gesuchten Teil der Rekonstruktion aus Konturen aufteilt. Die Rekonstruktion der FTT-Abtastwerte aus den Konturen ist hinsichtlich der Phase leider nicht trivial.

Deshalb wird zu der üblichen Pegelinformation vorläufig noch die Phaseninformation des analysierten FTT-Bandpaßspektrums an den Konturpunkten benötigt. Daraus resultiert das Rekonstruktionsverfahren RKOP, das bereits bei der Einstellung der Konturanalyse in Abschnitt 3.4 eingesetzt wurde, um eine optimale Phasenrekonstruktion zu simulieren.

Bild 5.3b zeigt eine aus 5.3a abgeleitete Konturcodierung mit Phasen. Während die rechte Bildhälfte unverändert geblieben ist, wurde die linke gegen eine FTT mit Zeit- und Frequenzkonturierung ausgetauscht, gefolgt von den Operationen FR/ZR. 'FTT' statt 'T' und Ausgänge im Zeit/Frequenz-Kontinuum drücken aus, daß die Rekonstruktion nun auch von der Abtastung des Analysespektrums unabhängig wird. Das zur Konturierung benötigte Pegelspektrum $L^L(f,t)$ und das Bandpaßphasenspektrum $\phi^{LB}(f,t)$ zusammengenommen entspricht dem komplexen Bandpaßspektrum $s^{LB}(\omega,t)$, das in der Praxis natürlich weiterhin in einer abgetasteten Form durch T berechnet wird. Die Konturierungen FK/ZK führen auf die bekannten Zeit- und Frequenzkontursignale $\mathcal{C}_F(t), \mathcal{C}_Z(t)$ sowie die zugehörigen Konturphasensignale $\Phi\{\mathcal{C}_F(t)\}, \Phi\{\mathcal{C}_Z(t)\}$. Der in Anhang A.4 definierte Phasenoperator $\Phi\{\ \}$ tauscht formal die Pegelwerte einer Konturpunktmenge gegen die Bandpaßphasenwerte aus.

Den eigentlichen Übergang auf eine Konturcodierung ermöglichen erst die Konturrasterierer FR und ZR nach Anhang B.6, welche eine Erweiterung der Kontursiebe FS/ZS darstellen. Wenn die in der Praxis realisierte Abtastung von Analysespektrum und Konturen so beschaffen wäre, daß sich ihr Grundraster mit dem des Synthesespektrums deckt, dann entspräche ihre Aufgabe den Kontursieben. Die auszusiebenden Abtastwerte lägen dann schon als Konturpunkte in Pegel und Phase vor. Durch Codierung und Quantisierung geht aber der Bezug zum Grundraster verloren. Nachdem wie bei FS/ZS die Orte des Syntheserasters gefunden sind, die ungleich null zu besetzen sind, müssen nun noch zusätzlich die Abtastwerte selbst approximiert werden.

Während grobe Quantisierung der Konturen später nach besseren Approximationsverfahren verlangt, genügt bei ausreichend feiner Quantisierung Treppenstufenapproximation. Für den Pegel übernimmt man also einfach den nächstnäheren Wert aus dem decodierten Konturverlauf. Zur Approximation des Phasenwertes kann man entsprechend den nächstnäheren Wert aus dem Konturphasenverlauf verwenden. Weist er eine zeitliche Entfernungskomponente auf, so ist allerdings eine Korrektur anzubringen, da sich die Bandpaßphase in zeitlicher Richtung schnell ändern kann. Das Verhalten des Bandpaßphasenspektrums wird im nächsten Abschnitt noch genauer untersucht.

5.1.5 Rekonstruktion aus Konturen mit heuristischer Phase (RKHP)

Zwar kann man mit dem eben besprochenen Verfahren RKOP eine sehr hohe Rekonstruktionsqualität erreichen. Eine Signalverarbeitung, die im Grundkonzept nur die Konturen des FTT-Pegelspektrums als auditiv relevant erachtet, sollte allerdings auf die Codierung von Phasen verzichten können. Im letzten Schritt in der Entwicklung eines Rekonstruktionsverfahrens geht es darum, sie auf Decoderseite geeignet wiederherzustellen.

Der Übergang von Bild 5.3b auf 5.3c illustriert diese Aufgabe. Auf der Analyseseite fallen die Phasenwege $\phi^{LB}(f,t)$ weg, wodurch die Konturanalyse die übliche Gestalt annimmt. Die Signale der Konturphasen $\Phi\{\mathcal{C}_F(t)\}$, $\Phi\{\mathcal{C}_Z(t)\}$ werden nicht mehr mitübertragen, sie

müssen nun decoderseitig durch die Phasenrekonstruktion PRK aus den Kontursignalen $\mathcal{C}_F(t), \mathcal{C}_Z(t)$ errechnet werden. Obwohl Pegel- wie Phasenwerte nur an den Orten des Syntheserasters $\{(\omega_{S_m}, lT_S)\}$ benötigt werden, läßt sich die Aufgabe der Phasenrekonstruktion besser im Kontinuum beschreiben. Deshalb liegt PRK noch vor den Konturrasterierern, verschmilzt aber in der Praxis mit diesen. Ansonsten stimmt der restliche Teil in Bild 5.3c mit 5.3b überein.

Phasenrekonstruktion wird im folgenden auf einer differentiellen Phasenregel aufgebaut, die das Verhalten der Phasen im FTT-Bandpaßspektrum in einer Umgebung um Konturen beschreibt. Die Regel wird zunächst am Beispiel eines eingeschalteten Sinustons erarbeitet. In einem zweiten Schritt wird erhärtet, daß sie entlang von Konturlinien beliebiger Signale gilt. Dabei erweist sich die bisherige Phasenfortschreibung der Teiltonsynthese als ihr Spezialfall. Dieser wird im dritten Schritt auch auf Zeitkonturen übertragen und als Phasenheuristik bezeichnet, womit das Verfahren RKHP festgelegt ist. Tatsächlich aber bietet die Regel Freiheitsgrade, mit denen zusätzliche Nebenbedingungen erfüllbar werden. Zwar wird damit eine auditiv optimale Phasenrekonstruktion möglich, für deren prinzipielle Existenz der vierte Schritt eine Nachweisskizze liefert. Allerdings ist dies mit hohem Aufwand verbunden, wie der Verfahrensansatz im fünften Schritt zeigt.

5.1.5.1 Aufstellung einer Phasenregel am Beispiel des Einschaltsinus

Die Konturphasen $\Phi\{\mathcal{C}\}$ geben die Werte des Bandpaßphasenspektrums $\phi^{LB}(f,t)$ an den Konturen $\mathcal{C}_F$, $\mathcal{C}_Z$ des zugehörigen Pegelspektrums $L^L(f,t)$ wieder. Um eine Vorstellung von ihrem Verhalten zu gewinnen, eignet sich eine Gegenüberstellung von Pegelspektrogramm, Konturen und einem sogenannten Phasenspektrogramm. Sie ist in Bild 5.5 für einen Sinuston der Frequenz 1 kHz zu sehen, der im Scheitelwert bei $t = 0$ eingeschaltet wird. Das Phasenspektrogramm in Bild 5.5d bildet ähnlich [Rio91] die Phasenwerte von 0 bis 2π in zunehmende Schwärzung ab. Beim Überschreiten von Vielfachen von 2π entsteht somit ein Sprung von schwarz nach weiß. Zu beachten ist, daß Phasenwerte in denjenigen Bereichen nicht unbedingt signifikant sind, in denen das Pegelspektrogramm in Bild 5.5a keine Schwärzung zeigt.

Der gestreckte Zeitmaßstab löst einzelne Nebenlinien der Zeit- und Frequenzkonturen besser als Bild 3.3 auf S. 66 auf ($B^{gs}_{3dB} \to \infty$). Die zugehörigen Welligkeiten in der Verschneidung zwischen transienten und stationären Spektralanteilen (Abschnitt 3.3.5) sind auch im Pegelspektrogramm gut zu erkennen. Für die Zeitkonturen zerfallen die Nebenlinien in einzelne Punkte, womit sie das diskrete Analysefrequenzraster von ZFKII aufdecken. Die Unstetigkeiten in den beiden Hauptästen der Zeitkonturen sind auf sehr kleine Restausläufer der Welligkeit zurückzuführen. Zu den Hauptästen zählen auch die beiden Teilstücke, die zeitlich zuallererst auftauchen und schließlich ineinander münden.

Im Phasenspektogramm bilden sich zwei Gebiete mit unterschiedlicher Regelmäßigkeit aus. Auf der Trennlinie zwischen den beiden Gebieten kommen zahlreiche Phasensprünge vor, lediglich in naher Umgebung der Tonfrequenz gibt es einen gleichmäßigen Übergang. Wenn man nur solche Gebietsteile beachtet, die auch im Pegelspektrogramm geschwärzt sind, wird klar, daß im ersten Gebiet der transiente und im zweiten der stationäre Beitrag dominiert. Innerhalb der Gebiete lassen sich zwei wesentliche Beobachtungen machen.

Betrachtet man erstens die Phasenänderung an festen Analysefrequenzen, so stimmt die

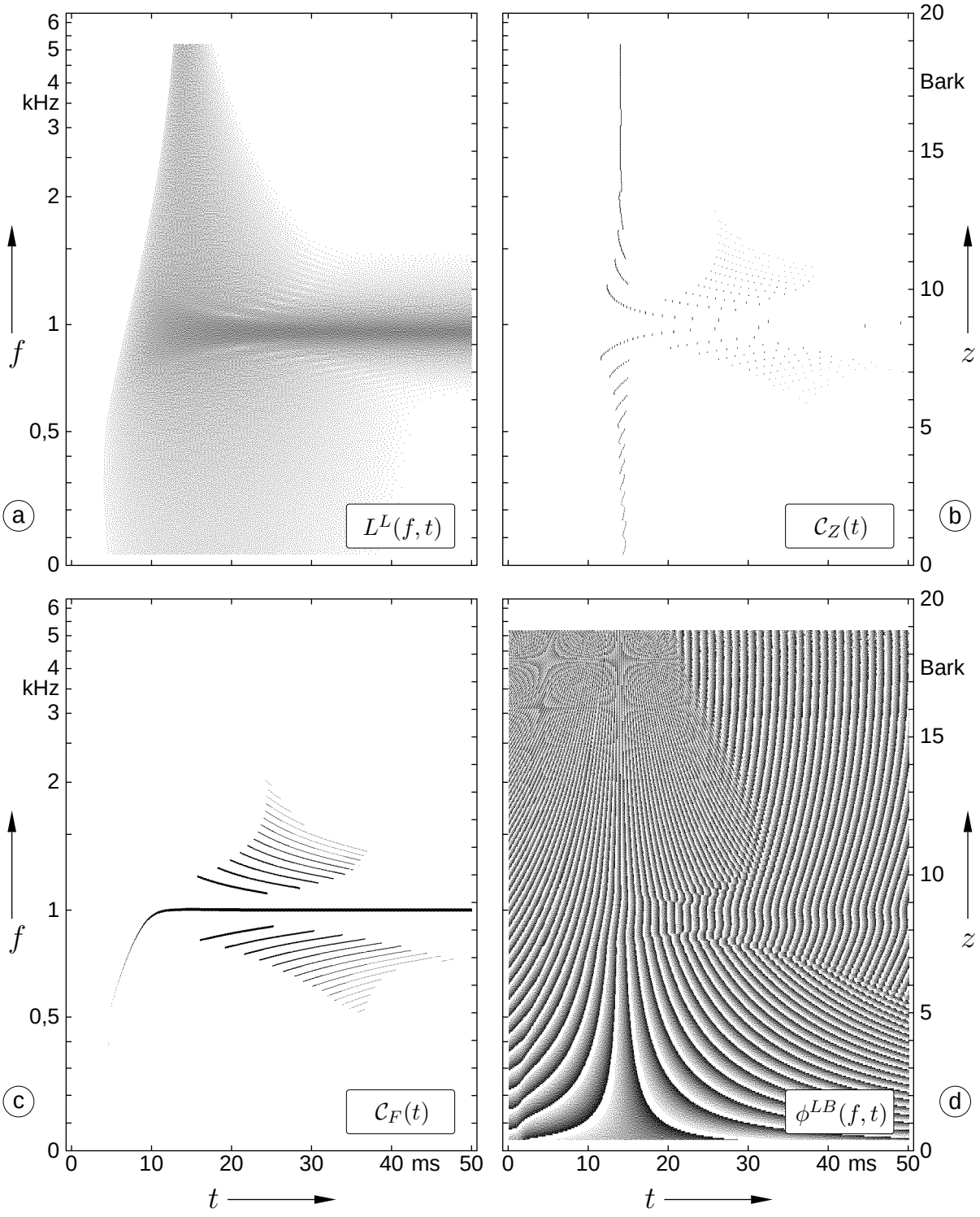

Bild 5.5: *Verhalten des Bandpaßphasenspektrums im Bereich von Konturlinien anhand eines bei* $t = 0$ *mit Sinusphase* $\frac{\pi}{2}$ *eingeschalteten Tons der Frequenz* 1 kHz *(siehe Text). a) Pegelspektrum als Spektrogramm,* 70 dB *sichtbare Pegeldynamik, b) Zeit- und c) Frequenzkonturen, d) Bandpaßphasenspektrum, Schwärzungsgrad gibt Phasenwerte von* 0 *bis* 2π *wieder. Transformations- und Konturierungsparameter nach Einstellung* ZFKII. *Eingeblendete Bezeichnungen verweisen auf Bild 5.3b.*

Anzahl der 2π-Überschreitungen pro Zeiteinheit im transienten Gebiet mit der Analysefrequenz, im stationären Gebiet mit der Tonfrequenz überein. Da das FTT-Bandpaßspektrum von Ausgangssignalen einer komplexen Bandpaßfilterbank wiedergegeben werden kann, findet sich eine einfache systemtheoretische Erklärung: Im transienten Gebiet verhält sich ein Bandpaßfilter so, als wäre es durch einen Impuls angeregt worden. Die Phasenänderung seiner Impulsantwort entspricht genau seiner Mittenfrequenz. Im stationären Gebiet gibt die Phasenänderung die selektierte stationäre Schwingung wieder.

Betrachtet man zweitens die Phasenänderung über der Frequenz zu festen Zeitpunkten, dann gibt es in der unmittelbaren Umgebung der Hauptäste beider Konturtypen keine wesentliche Änderung. Am mittleren Zeitort der Zeitkonturhauptäste bleibt die Phase konstant, außer in der Nähe der Tonfrequenz. Dort dreht sie langsam um π. Im stationären Gebiet dreht sich die Phase bei zunehmender Entfernung vom Frequenzkonturhauptast stärker. Dies liegt am Phasengang der Analysefilter und am frequenzabhängigen Laufzeitausgleich.

Die Beobachtungen lassen sich mit folgender Phasenregel vereinbaren, die allgemein das Verhalten des Bandpaßphasenspektrums in der unmittelbaren Umgebung von Konturen beschreibt.

Phasenregel: Seien (f,t) der Ort eines Konturpunktes im FTT-Pegelspektrum, $\Delta\psi$ eine Phasendrift, B^A_{3dB} die Analysebandbreite und T^A_{3dB} die 3dB-Zeitbreite des Analysefensters. Dann gilt in einer beschränkten Umgebung $|\Delta f| \ll B^A_{3dB}$ und $|\Delta t| \ll T^A_{3dB}$ für die Änderung des zugehörigen Bandpaßphasenspektrums [5]

$$\begin{aligned} \phi^{LB}(f+\Delta f, t+\Delta t) &= \phi^{LB}(f,t) + 2\pi f \cdot \Delta t + \Delta\psi, && (5.31)\\ \lim_{\substack{\Delta f\to 0\\ \Delta t\to 0}} \Delta\psi &= 0, && (5.32)\\ |\Delta\psi| &< \max\left\{|2\pi B^A_{3dB}\cdot\Delta t|, |T^A_{3dB}\cdot 2\pi\Delta f|\right\}. && (5.33) \end{aligned}$$

Die erste Gleichung drückt die Entwicklung der Phase durch einen linearen Term $2\pi f\cdot\Delta t$ aus, zuzüglich einer zunächst unbekannten Phasendrift $\Delta\psi$. Der lineare Term faßt die obigen beiden Beobachtungen zusammen: keine Phasenänderung in Frequenzrichtung, in Zeitrichtung dafür eine Änderung mit der Analysefrequenz, die nahe beim Frequenzkonturhauptast praktisch mit der Tonfrequenz übereinstimmt. Wozu dient die Phasendrift? Wenn die Phasenentwicklung immer exakt dem linearen Term folgen würde, wäre Phasenrekonstruktion kein Problem. $\Delta\psi$ stellt einen differentiellen Freiheitsgrad dar, mit dem später integrale Nebenbedingungen lösbar werden (vgl. Abschnitt 2.3.2.2).

Die zweite Gleichung besagt in Verbindung mit der ersten, daß die Phase an Konturpunkten stetig verläuft. Dies entspricht der Beobachtung an den Hauptkonturästen. Phasensprünge sind also erst in einem bestimmten Mindestabstand von Konturpunkten möglich, nämlich außerhalb der Gültigkeitsumgebung der Phasenregel. Sie ist mutmaßlich durch die spektrale und zeitliche Auflösung des Analysefensters vorgegeben, hier grob angenähert durch B^A_{3dB} und T^A_{3dB}.

[5] Formal sind die Phasenfunktionen als 'entpackt' vorauszusetzen, damit die Grenzwertbetrachtung stimmt. Das bedeutet, daß sie von 2π-Sprüngen befreit und nicht auf den Bereich $0\ldots 2\pi$ beschränkt sind.

Die dritte Gleichung trifft eine Aussage über die maximale Phasendrift in der betrachteten Umgebung. Dies wird noch begründet. In Verbindung mit den anderen beiden Gleichungen ist dadurch indirekt ausgedrückt, daß der lineare Term in der ersten Gleichung eine gute Näherung für die Phasenentwicklung bereitstellt. Die Näherung kann nur stetig und beschränkt von der tatsächlichen Entwicklung 'wegdriften'.

5.1.5.2 Notwendigkeit der Phasenregel für Konturlinien

Es wird nun gezeigt, daß die Phasenregel die einzig plausible Phasenentwicklung innerhalb von relevanten[6] Konturlinien beschreibt. Phasenrekonstruktionen, die dies mißachten, provozieren Widersprüche zu den vorgegebenen Linienparametern. Dazu wird eine vereinfachte Formulierung des Rekonstruktionsvorgangs von RKOP/RKHP für eine einzelne Konturlinie benötigt. Auf Anregung von Horn [HorPK] bietet sich eine elegante Darstellung mit Hilfe sogenannter Konturpunkt-Wavelets an (Anhang D.2 und D.3). Diese repräsentieren die Impulsantwort des komplexen Synthesebandpasses an der Konturpunktfrequenz, verschoben an die Konturpunktzeit.

Eine Konturlinie stellt sich nach den Rasterierern FR/ZR in Bild 5.3b allgemein als Parametermenge $\{(A_i, \phi_i, f_i, t_i)\}$ dar, die die Parameter Amplitude, Phase, Frequenz und Zeit für jeden gerasterten Konturpunkt enthält. Vereinfacht ergibt sich aus Gln. (D.9) bis (D.11) das Rekonstruktionssignal

$$\hat{s}(t) = \sum_i \mathrm{Re}\Big\{A_i \cdot e^{j\phi_i} \cdot w_K^S(2\pi f_i, t - t_i)\Big\}, \quad \text{mit} \tag{5.34}$$

$$w_K^S(2\pi f, t) = c \cdot h^S(t) \cdot e^{j2\pi f(t - t_{max}^S)}. \tag{5.35}$$

Darin sind die $w_K^S(2\pi f_i, t - t_i)$ die Wavelets. Im Wavelet-Grundtyp $w_K^S(2\pi f, t)$ erkennt man die Impulsantwort des komplexen Synthesebandpasses bei der Frequenz f mit dem Synthesefenster $h^S(t)$ als Hüllkurve. Vernachlässigt wurde dabei, daß die Synthesefensterfunktion $h^S(t)$ eigentlich frequenzabhängig ist und einen Laufzeitausgleich benötigt. Außerdem hängt die Gewichtung c normalerweise vom Konturtyp und von der Frequenz ab. Schließlich blieb die gegenseitige Maskierung MSK in Bild 5.3b unberücksichtigt.

Bei einer Frequenzkonturlinie kann man von zusammenhängenden Zeitpunkten $t_i = iT_S$ ausgehen, mit $i_B \le i \le i_E$. Wird dies zusammen mit Gl. (5.35) in (5.34) berücksichtigt, so ergibt sich nach Umformung

$$\hat{s}(t) = \sum_{i=i_B}^{i_E} A_i \cdot h^S(t - iT_S) \cdot \cos\Big(2\pi f_i(t - iT_S - t_{max}^S) + \phi_i\Big). \tag{5.36}$$

Man kann nun folgendes entdecken: Das Resultat stimmt der Form nach mit Gl. (1.22) der Teiltonsynthese auf S. 17 überein, wenn man $T_S = T_A$ annimmt. Die unterschiedlichen

[6] Für den Sonderfall von Nebenkonturlinien gilt die Phasenregel vermutlich nur innerhalb einer deutlich eingeschränkteren Umgebung. Für einige andere Konturlinien gilt Gl. (5.33) nur in der Nähe des Pegelmaximums der Linie. Dazu zählen die Frequenzkonturlinie des Dirac-Impulses in Bild 2.1 und die Zeitkonturlinien bei $B_{3dB}^{gs} = 50$ Hz in Bild 3.3b. Konturpunkte, bei denen die Phasenregel nicht im vollen Umfang gilt, tragen keine gehörrelevante Information und betreffen Energiebeiträge, die später bei Signalrekonstruktion nur unter vermeidbaren Bedingungen die Hörschwellen des Nutzsignals überschreiten (vgl. Abschnitt 5.1.7.2).

Winkelfunktionen mit einem Zeitversatz von t^S_{max} bewirken lediglich, daß ϕ_i die Kosinusphase im Hüllkurvenmaximum des Wavelets statt der Sinusphase am Hüllkurvenanfang bezeichnet. Weiterhin stimmt die Phasenfortschreibung der Teiltonsynthese in Gl. (1.23) der Form nach mit Gl. (5.31) der Phasenregel überein, wenn man $\Delta\psi = 0$ annimmt. In Bild 1.3 auf S. 18 verkörpert demnach jeder Sinustonimpuls genau ein Wavelet, das durch die Konturpunktparameter in Amplitude und Phase eingestellt wurde. Natürlich sind die Hüllkurven bei RKOP/RKHP keine Rechtecke oder Dreiecke mehr, sondern weich berandete, asymmetrische Formen, die wegen eines feineren Syntheserasters stark überlappen.

Weil die Teiltonsynthese formale Ähnlichkeit mit dem Rekonstruktionsvorgang von RKHP/RKOP aufweist, kann man Gl. (5.31) der Phasenregel anhand von Bild 1.3 rechts belegen. Angenommen, ϕ_2 wäre dort nicht nach der Phasenregel aus ϕ_1 hervorgegangen, sondern hätte durch einen Phasensprung einen um π versetzten Wert. Dann würde zwischen dem Zeitpunkt 1 und 2 die Hüllkurve des Synthesesignals stark einbrechen, weil die Wavelets dort gegenphasige Schwingungen aufweisen. Dies kann aber nur dann richtig sein, wenn ein solcher Einbruch schon im Konturverlauf analysiert worden wäre.[7] Hier ergibt sich offenbar ein Widerspruch.

Geringfügige Abweichungen der Synthesehüllkurve vom analysierten Konturverlauf sind allerdings nicht grundsätzlich vermeidbar. Beispielsweise verursachen Frequenzänderungen selbst bei optimaler Wahl der ϕ_i Einbrüche, weil benachbarte Wavelets dann nicht mehr in ihrem gesamten Überlappungsbereich phasensynchron schwingen können. Dieser Sachverhalt ist für übertriebene Frequenzänderungen in Bild 1.3 rechts daran zu erkennen, daß die tatsächliche Synthesehüllkurve von der umrahmenden, gewünschten Hüllkurve abweicht. Hierin spiegelt sich die Tatsache wider, daß eine Signalverarbeitung mit Konturen keine perfekte Signalrekonstruktion erlaubt. Der oben aufgedeckte Widerspruch löst sich also zusehends auf, je kleiner die angenommenen Phasensprünge werden. Innerhalb eines Konturpunktabstandes Δt muß demnach eine gewisse Phasendrift $\Delta\psi \neq 0$ erlaubt sein. Dabei ist es aber nicht nötig, für sie ein anderes Verhalten als in Gl. (5.32) zu vermuten.

Ganz ähnlich kann auch bei Zeitkonturlinien für Gl. (5.31) der Phasenregel argumentiert werden. Im Rekonstruktionsvorgang nach Gln. (5.34), (5.35) besteht eine Zeitkonturline aus zusammenhängenden Frequenzen $f_i = (2\pi)^{-1} \cdot \omega_{S_i}$ mit $i_B \leq i \leq i_E$. Weil die Linie näherungsweise frequenzparallel ausgerichtet ist, liegen die Hüllkurven der zugeordneten Wavelets praktisch zeitlich übereinander. Das kann man sich anhand von Bild 1.3 rechts vorstellen, indem man die drei Wavelets zeitlich entsprechend zusammenschiebt. Die Phasenregel bewirkt nun, daß die Schwingungen benachbarter Wavelets in Umgebung ihrer Hüllkurvenmaxima weitgehend in Phase sind. Dazu müssen die Hüllkurvenmaxima nicht unbedingt genau über dem gleichen Zeitpunkt liegen.

Probeweise kann man hier die Phase aller Wavelets oberhalb eines Frequenzindexes i_X um π verschieben. Spektral gesehen – im Sinne einer erneuten Kurzzeitspektralanalyse mit der FTT – bildet sich dadurch bei f_{i_X} ein spektraler Einbruch im Synthesesignal. Das liegt daran, daß dort das Analysefilter der ‘Re-Analyse’ Wavelets oberhalb und unterhalb von f_{i_X} selektiert, die sich frequenzmäßig nur wenig unterscheiden, die sich phasenmäßig aber praktisch aufheben. Der Einbruch widerspricht jedoch der Vorgabe, die die Konturlinie

[7] Eine solcher Vergleich von Analyse- und Syntheseergebnis ist nur zulässig, wenn sich Typ und Bandbreite von Analyse- und Synthesefenster nicht zu sehr unterscheiden. Die Zeit/Frequenz-Unschärfen von Analyse- und Rekonstruktionsvorgang müssen also vergleichbar sein. Dies gilt bei den mit RKOP/RKHP verwendeten Analysen als gewährleistet.

des ursprünglichen FTT-Spektrums lieferte. Daß innerhalb eines Konturpunktabstandes Δf auch hier eine gewisse Phasendrift $\Delta\psi \neq 0$ erlaubt sein muß, ergibt sich aus ähnlicher Argumentation wie bei der Frequenzkonturlinie. Dabei ist es auch hier nicht plausibel, ein anderes Verhalten als Gl. (5.32) anzunehmen.

Warum ist schließlich ist $\Delta\psi$ beschränkt, wie in Gl. (5.33) der Phasenregel behauptet? Ein (partieller) Differenzenquotient $\Delta\psi/(2\pi\Delta t) \neq 0$ in einer Frequenzkonturlinie entspricht einem Frequenzversatz des Synthesesignals. Man kann das daran ablesen, daß sich der Abstand der Nulldurchgänge zwischen benachbarten Wavelet-Schwingungen verkleinert oder vergrößert. Ein Frequenzversatz darf aber nicht größer als die spektrale Auflösung der FTT sein, hier abgeschätzt durch B^A_{3dB}. Sonst lägen den Konturpunkten spektrale Maxima von Signalanteilen zugrunde, die eigentlich an anderen Analysefrequenzen maximal selektiert werden müßten. Ein ähnlicher Widerspruch ergibt sich für einen Differenzenquotienten $\Delta\psi/(2\pi\Delta f) \neq 0$ bei Zeitkonturlinien. Hier repräsentiert er einen Zeitversatz im Synthesesignal. Die Bedeutung des Drehfaktors in der Korrekturfunktion der RFTT kann dies demonstrieren (Abschnitt 5.1.1.1). Ein Zeitversatz darf aber nicht die zeitliche Auflösung der FTT überschreiten, etwa T^A_{3dB}. Sonst nämlich wären die Konturpunkte der Linie zu einem anderen Zeitpunkt analysiert worden. Frequenz- und Zeitversetzungen aufgrund der Phasenentwicklung werden nochmals in Abschnitt 6.2.1.2 und 6.2.1.3 behandelt.

5.1.5.3 Phasenheuristik und ihre Grenzen

Als Vorschrift zur Phasenrekonstruktion für Konturlinien weist die Phasenregel zwei Freiheitsgrade auf: Die Startphase einer Linie wie auch die konkrete Entwicklung der differentiellen Phasendrift $\Delta\psi$ bleiben noch festzulegen. Wie kommt man zu einer einfachen und brauchbaren, wenn auch nicht optimalen Festlegung, die für beide Konturtypen geeignet ist? Die Phasenfortschreibung in der Teiltonsynthese beispielsweise gehorcht zumindest prinzipiell der Phasenregel, wie oben deutlich wurde. Für Frequenzkonturlinien sind hier willkürlich $\Delta\psi = 0$ und ein Fixwert als Startphase vorgegeben. [8]

Bei der Phasenfortschreibung der Teiltonsynthese ist allerdings noch eine Phasenübergabetoleranz Δf_Φ von Bedeutung (Abschnitt 1.5.4). Sie bestimmt erstens die Assoziation von Frequenzkonturpunkten zu Linien, indem Nachfolger innerhalb von $\pm\Delta f_\Phi$ gesucht werden. Zweitens erlaubt sie, daß die Phase bei Aufspaltungen von Linien an Folgestücke weitergereicht wird, sofern sie innerhalb von $\pm\Delta f_\Phi$ um das endende Linienstück beginnen. Bisher wird ein relativ hohes Δf_Φ verwendet, das mit 0,15 bis 0,25 Bark die Analysebandbreite erreicht. Phasenübergabe über solche Distanzen verletzt zwar die Umgebungsbeschränkung der Phasenregel, bewährt sich aber offenbar in der Praxis.

Anwendung der Phasenregel ohne Phasendrift, Fixwert als Startphase und großzügige Phasenübergabe zusammengenommen bilden eine einfache *Phasenheuristik*, die auch auf Zeitkonturen übertragen werden kann. Gegenüber den Frequenzkonturen stellt sich die Situation gewissermaßen um 90 Grad gedreht dar: Beginnend bei tiefen Frequenzen werden für Zeitkonturpunkte frequenzbenachbarte Nachfolger innerhalb einer zeitlichen Phasenübergabetoleranz $\pm\Delta t_\Phi$ gesucht und zu Linien assoziiert. Dabei wird die Phase nach

[8] Der konkrete Fixwert null ist bestenfalls beim Rechteckfenster der ursprünglichen Teiltonsynthese (TTSR) wichtig, um die Übergangseffekte am Linienanfang zu minimieren (vgl. Bild 1.3 links). Bei allen anderen, weich berandeten Synthesefenstern erweist sich die Wahl des Fixwertes als bedeutungslos.

der Phasenregel mit $\Delta\psi = 0$ fortgeschrieben. Am Linienanfang wird ein Fixwert als Startphase zugrunde gelegt, wenn nicht der Phasenstand einer anderen Zeitkonturlinie im Abstand bis $\pm\Delta t_\Phi$ fortgeschrieben werden kann. Für Zeit- und Frequenzkonturen sind die Heuristiken also voneinander unabhängig. Sie sind in Anhang B.6 formal beschrieben.

Mit der Phasenheuristik für Zeit- und Frequenzkonturen ist die suboptimale, behelfsmäßige Phasenrekonstruktion des Verfahrens RKHP festgelegt. Sie weist leider wesentliche Schwächen auf, die bei den Frequenzkonturen bereits bekannt sind: Das Problem der Phaseninkohärenz-bedingten Störungen bleibt bestehen (Abschnitt 2.3). Nach wie vor können also dicht benachbarte Linienverläufe Phasenlagen zueinander entwickeln, in deren Folge rekonstruierte Schmalbandhüllkurven erheblich von denen des Originals abweichen. Mit Originalphasen beziehungsweise mit Phasenheuristik rekonstruierte Frequenzkonturen unterscheiden sich deshalb in der Wiedergabe. Das Inkohärenzproblem existiert in ähnlicher Form auch für die Phasenheuristik der Zeitkonturen. Ihre Wiedergabe (ohne Frequenzkonturen) erreicht nicht die impulshafte Qualität, die mit Originalphasen erzielbar ist.

Phaseninkohärenz macht sich bei gleichzeitiger Verarbeitung von Zeit- und Frequenzkonturen noch auf eine weitere Weise bemerkbar. Die Unabhängigkeit der Heuristiken für die beiden Konturtypen führt dort zu fehlerhaften Überlagerungen, wo sich Linien unterschiedlichen Typs nahe kommen oder überschneiden. Eindrucksvoll belegt dies ein Verfahren RKOP, in dem Zeit- und Frequenzkonturanteil mit entgegengesetztem Vorzeichen überlagert werden, so daß praktisch die Phase bei einem der beiden Konturtypen systematisch um π gegenüber dem anderen verdreht ist. Die Rekonstruktionsqualität geht dadurch etwa auf das Niveau zurück, das auch mit der Phasenheuristik in RKHP zu erreichen ist. Zusammengefaßt liegen die Grenzen der Phasenheuristik demnach in der ungesicherten Phasenkohärenz nahe beieinander liegender Konturlinien, welchem Konturtyp sie auch angehören mögen.

5.1.5.4 Existenz und Natur einer optimalen Phasenrekonstruktion

Es folgt eine Nachweisskizze dafür, daß Konturen prinzipiell eine optimale Rekonstruktion von Konturphasen erlauben. Dies führt zwar im allgemeinen nicht auf die Originalphasen, wie sie beim Verfahren RKOP vorliegen. Man erhält aber einen auditiv gleichwertigen Ersatz, der folglich auch Verfälschungen durch Phaseninkohärenz vermeidet. Auch wenn der Nachweis lückenhaft ist, so gibt er auf jeden Fall einen Einblick in die prinzipiellen Einschränkungen, mit denen bei der Rekonstruktion der Konturphasen zu rechnen ist.

Die Nachweisstrategie in vier Schritten besteht darin, daß eine Rekonstruierbarkeit des Zeitsignals aus dem Kurzzeitbetragsspektrum, welche in der Literatur belegt ist, zur auditiv optimalen Rekonstruierbarkeit aus Konturen weiterentwickelt wird. Wenn dies bewiesen ist, dann ist implizit auch die auditiv optimale Rekonstruierbarkeit der Konturphasen bewiesen, weil man die Konturphasen notfalls auch nachträglich im rekonstruierten Zeitsignal analysieren kann.

Der erste Schritt behandelt, mit welchen Einschränkungen das Zeitsignal aus seinem Kurzzeitbetragsspektrum zurückgewonnen werden kann, wenn ein frequenzunabhängiges, zeitbegrenztes Analysefenster vorausgesetzt wird. Die Annahme eines quantisierten Betragsspektrums im zweiten Schritt führt auf neuartige Einschränkungen. In dieser Situation stellt es sich als gegenstandslos heraus, zeitbegrenzte Fenster vorauszusetzen, womit auch

FTT-Fenstertypen annehmbar sind. Die Erkenntnisse werden im dritten Schritt auf das FTT-Betragsspektrum mit frequenzabhängiger Fensterfunktion übertragen. Seine Eigenschaften stellen sicher, daß die gefundenen Einschränkungen auditiv bedeutungslos sind. Im letzten Schritt schließlich wird gezeigt, daß das FTT-Betragsspektrum auch aus Konturen wiederherzustellen ist.

1 Nawab et al. bewiesen für ein frequenzunabhängiges, zeitbegrenztes Analysefenster, daß ein zeitdiskretes Signal aus seinem zeitvarianten, diskreten Kurzzeitbetragsspektrum mit Einschränkungen rekonstruierbar ist [Naw83]. Um die Einschränkungen zu verdeutlichen, muß man sich das Signal in eine Reihe von Zeitsegmenten unterteilt vorstellen. Die Grenze zwischen zwei Segmenten ist dadurch definiert, daß aufeinanderfolgende Nullwerte vorkommen, und zwar mindestens für eine Analysefensterlänge abzüglich zweier Zeitschritte. Innerhalb eines Segmentes stimmt das rekonstruierte Signal mit seinem Original überein, nur das Vorzeichen ist nicht wiederherzustellen: Es kann segmentweise falsch oder richtig sein. Dies läßt sich gut anhand zweier aufeinanderfolgender Impulse illustrieren: Wenn die dazwischenliegende Anzahl von Nullwerten groß genug ist, passen sie niemals gemeinsam in das zeitlich gleitende Analysefenster. Das Betragsspektrum kann dann keine Information über ihre Vorzeichenrelation enthalten.

Die Voraussetzungen des Beweises von Nawab et al. sind aber nicht direkt auf die Praxis übertragbar, in der nur wertdiskrete, quantisierte Kurzzeitspektren zugänglich sind. Deshalb führen die in [Naw83] präsentierten Rechenverfahren zu suboptimalen Lösungen, verglichen mit der theoretischen Vorhersage. Dies gilt auch für das spätere, numerisch robuste Verfahren von Griffin und Lim, das aber offenbar sehr gute Qualität bei Sprachsignalen liefert [Gri84].

2 Bei quantisierten Betragsspektren macht es keinen Sinn mehr, zwischen zeitbegrenzten und nicht-zeitbegrenzten Fensterfunktionen zu unterscheiden. Nicht-zeitbegrenzte Fenster wie die der FTT laufen an mindestens einer Seite asymptotisch aus. Die Ausläufe kann man sich an einer Stelle abgeschnitten denken, ohne daß der Effekt die Quantisierungsunsicherheit des Betragsspektrums übersteigt. Wenn der noch signifikante Teil der Fensterlänge sehr viele Zeitschritte beträgt, dann wiegt auch der Abzug zweier Zeitschritte nicht viel. Damit kann man eine Segmentgrenze genau dann annehmen, wenn das Signal mindestens über der Dauer einer signifikanten Fensterlänge nicht sicher beobachtet werden kann. Das äußert sich dadurch, daß das quantisierte Betragsspektrum an allen beobachteten Frequenzen gleichzeitig null ist. Dies gilt jedoch nur für Fensterfunktionen ohne Nebenmaxima, andernfalls verkomplizieren sich die Verhältnisse.

Bei quantisierten Betragsspektren kommt aber noch ein weiterer Effekt ins Spiel. Nawab et al. konnten implizit darauf vertrauen, daß die Fourier-Transformierte der Fensterfunktion den gesamten beobachteten Frequenzbereich überspannt. Quantisierung bewirkt nun aber, daß prinzipiell auch das Fourier-Spektrum $S(\omega)$ des Signals $s(t)$ in Segmente zerfallen kann. Dies wird klar, wenn man dem Kurzzeitspektrum, bestimmt durch Signal und Analysefenster, eine alternative Formulierung auf Basis von Fourier-Transformierten gegenüberstellt. Das Kurzzeitspektrum oder FTT-Spektrum $s^C_{\omega_A}(t)$ bei frequenzunabhängigem Fenster $h(t)$ kann man wie folgt ansetzen und umformen:

$$s^C_{\omega_A}(t) = \int_0^t s(\tau) \cdot h(t-\tau) \cdot e^{-j\omega_A \tau} d\tau \tag{5.37}$$

$$= e^{-j\omega_A t} \cdot [s(t) * (h(t) \cdot e^{j\omega_A t})] \tag{5.38}$$

$$= e^{-j\omega_A t} \cdot \mathcal{F}^{-1}\{S(\omega) \cdot H(\omega - \omega_A)\} \tag{5.39}$$

$$= e^{-j\omega_A t} \cdot \int_{-\infty}^{\infty} S(\omega) \cdot H(-(\omega_A - \omega)) \cdot e^{-j(-t)\omega} d\omega. \tag{5.40}$$

Die elementare erste Umformung wurde in Variation in Abschnitt 1.4.2 behandelt. Die zweite benutzt die Fourier-Transformation, die Korrespondenz von Faltung und Multiplikation im Zeit- bzw. im Spektralbereich sowie die spektrale Verschiebungsregel. In der dritten ist lediglich das Rücktransformationsintegral ausgeschrieben worden.

Wesentlich ist, daß die Verwendung von $S(\omega)$, $H(-\omega)$ in der letzten Zeile formal mit der von $s(t)$, $h(t)$ in der ersten Zeile übereinstimmt, wenn man nur den Betrag $|s^C_{\omega_A}(t)|$ berücksichtigt. Wenn die Aussagen über Segmentgrenzen im Zeitbereich richtig sind, dann müssen sie offensichtlich auch im Fourier-Spektralbereich gelten. In $S(\omega)$ repräsentieren zwei schmalbandige Impulse in etwa zwei quasistationäre Sinusschwingungen. Wenn ihr spektraler Abstand groß genug ist, kann ihre Präsenz wegen der Quantisierung nicht 'gleichzeitig' im spektralen Fenster $H(-\omega)$ beobachtet werden. Das Signal fällt also auch im Fourier-Spektrum in Segmente. Für Fensterfunktionen ohne spektrale Nebenmaxima kann man schlußfolgern: Spektrale Segmentgrenzen liegen genau an denjenigen Frequenzen vor, an denen das quantisierte Kurzzeitbetragsspektrum für den gesamten Beobachtungszeitraum des Signals null ist. Im Unterschied zum reellen Zeitbereich geht zwischen den komplexen Segmenten des Spektralbereichs nicht nur eine Vorzeichenrelation, sondern sogar eine Phasenrelation verloren.

Kombiniert man die Möglichkeiten der Segmentierung von Zeit- und (Fourier-)Frequenzbereich, so kann entweder eine oder keine der beiden Formen vorkommen – oder aber beide gleichermaßen. Letzteres ist der allgemeinste Fall. Bei Fensterfunktionen ohne zeitliche und spektrale Nebenmaxima bedeutet er offenbar folgendes: Das quantisierte, zeitvariante Kurzzeitbetragsspektrum kann über der Zeit/Frequenz-Fläche in 'Inseln' zerfallen, die ohne Verbindung zueinander aus dem 'Meer' der Dynamikunterkante herausragen. Phasenrelationen innerhalb einer Insel sind wiederherstellbar. Phasenrelationen zwischen den Inseln wie auch in Bezug auf absolute Phasenwerte des komplexen Kurzzeitspektrums gehen verloren.

3 Es wird nun als plausibel angenommen, daß eine Frequenzabhängigkeit der Fensterfunktion prinzipiell nichts an der Insel-Theorie ändert. Frequenzabhängigkeit hat auch in einem anderem Kontext keinen Informationsverlust zur Folge, wie die Rücktransformierbarkeit des komplexen FTT-Spektrums in Abschnitt 5.1.1 belegt. Der Verlust der Phasenrelation zwischen Signalanteilen, die verschiedenen Inseln des FTT-Betragsspektrums zugeordnet sind, dürfte auditiv keine Rolle spielen. Schließlich geschieht die Anpassung vom FTT-Analysefenster an die Gehöreigenschaften in der Erwartung, daß die Gestalt des Betragsspektrums – letztlich sogar nur seine Konturen – alle hörrelevanten Vorgänge enthält. Eine erneute Analyse mit dem gleichen Fenster, die jetzt näherungsweise den peripheren Hörvorgang repräsentiert, wird die gleichen Inseln und auch die gleichen Konturen reproduzieren. Dies geschieht unabhängig davon, wie die Phasenrelationen zwischen den Inseln verändert wurden.

4 Abschließend ist zu zeigen, wie man aus Konturen ein zeitvariantes Betragsspektrum gewinnen kann. Dazu kombiniert man die Operationen WFS/WZS aus Abschnitt 4.3 oder Anhang B.5: Für jeden Frequenzkonturpunkt wird das stationäre Spektrum eines Sinustons ermittelt, das genau diesen Konturpunkt ausgelöst hätte. Entsprechend gehört

zu jedem Zeitkonturpunkt eine Impulsantwort des örtlichen Analysefilters. Das gesuchte Betragsspektrum an einem bestimmten Zeit/Frequenz-Ort ergibt sich, in dem man das Betragsmaximum aller Funktionswerte sucht, die die einzelnen Spektren und Impulsantworten am Ort aufweisen. Die Abweichung vom ursprünglichen Betragsspektrum ist an den Konturpunkten grundsätzlich null und kann nur in Entfernung von irgendwelchen Konturpunkten – also zu niedrigeren Pegeln hin – zunehmen. Man kann deshalb vermuten, daß ein gehörangepaßtes Synthesefenster den Fehler unter die Vor-, Nach- und Mithörschwellen formen kann.

5.1.5.5 Verfahrensansatz einer optimalen Phasenrekonstruktion

Wenn eine auditiv optimale Rekonstruktion der Konturphasen existiert, wie könnte man sie realisieren? Abschließend wird ein Ansatz vorgestellt, der wie die Phasenheuristik von der Phasenregel ausgeht. Wesentlich ist diesmal, daß die Freiheitsgrade Phasendrift und Linienstartphase nicht willkürlich fixiert, sondern mit Hilfe zusätzlicher Nebenbedingungen genutzt werden. Diese werden von neu aufzustellenden Regeln abgeleitet, welche die Phasenrelationen zwischen einzelnen Punkten benachbarter Konturlinien festlegen. Die rekonstruierten Konturphasen ergeben sich daraus, daß die Nebenbedingungen mit möglichst minimaler Phasendrift erfüllt werden.

Bild 5.6 zeigt zwei Beispiele von sogenannten *Phasenrelationsregeln*. Die Regel R^a ist im Grunde trivial. Sie besagt, daß die rekonstruierten Phasen zweier Konturlinien in einem eventuellen Überkreuzungspunkt den gleichen 2π-modulo-Wert aufweisen müssen. R^b betrifft die Konstellation von parallelen Frequenzkonturlinien mit einer dazwischenliegenden Zeitkonturlinie. Damit ein Analysefilter in der Frequenzmitte zweier Sinustöne ein zeitliches Pegelmaximum erreichen und damit einen Zeitkonturpunkt auslösen kann, muß die Bandpaßphase zu diesem Zeitpunkt an den Sinustonfrequenzen den gleichen 2π-modulo-Wert haben. Dies folgt vereinfacht aus der Berechnungsvorschrift für FTT-Spektren von stationären Sinusschwingungen nach Abschnitt 1.4.2 zusammen mit dem Übergang auf das FTT-Bandpaßspektrum nach Gl. (1.10). Bei ungleich starken Sinustönen spielt zusätzlich noch der Phasengang des Analysefilters eine Rolle. Beide Situationen können in den Konturen der Impulsfolge in Bild 3.2b auf S. 63 wiedergefunden werden. Nach Anwendung solcher Regeln erhält man einen Satz von Nebenbedingungen, hier beispielsweise B_1^a und B_1^b.

Ein geeigneter Verarbeitungsrahmen sieht so aus, daß man für jede Konturlinie zuerst die Phasenfortschreibung nach der Phasenheuristik berechnet. Damit wird nur die 'schnelle' zeitliche Veränderung des Terms $2\pi f \cdot \Delta t$ in Gl. (5.31) der Phasenregel berücksichtigt, weil $\Delta\psi = 0$ gilt. Nachträglich wird nun der Verlauf einer Phasendriftfunktion ψ hinzuaddiert, die sich minimal – und einfacherweise linear – zwischen zwei Nebenbedingungen verändert. Eine formale Beschreibung dieser Maßnahmen ist in Anhang B.6 mitenthalten. Zu beachten ist, daß die bisher unzulässig hohen Phasenübergabetoleranzen Δf_Φ, Δt_Φ der Phasenheuristik nun kleiner gewählt werden. Dadurch erweitert sich der Raum zur Anwendung von Phasenrelationsregeln, weil ursprünglich zusammenhängende Linien in kürzere separate zerfallen können.

Nebenbedingungen machen nur relative Aussagen über einzelne Phasenwerte. Wie knüpft man das Beziehungssystem aus Nebenbedingungen und zu minimierenden Phasendriftwerten an absolute Phasenwerte? Grundsätzlich können Phasenrelationsregeln nur inner-

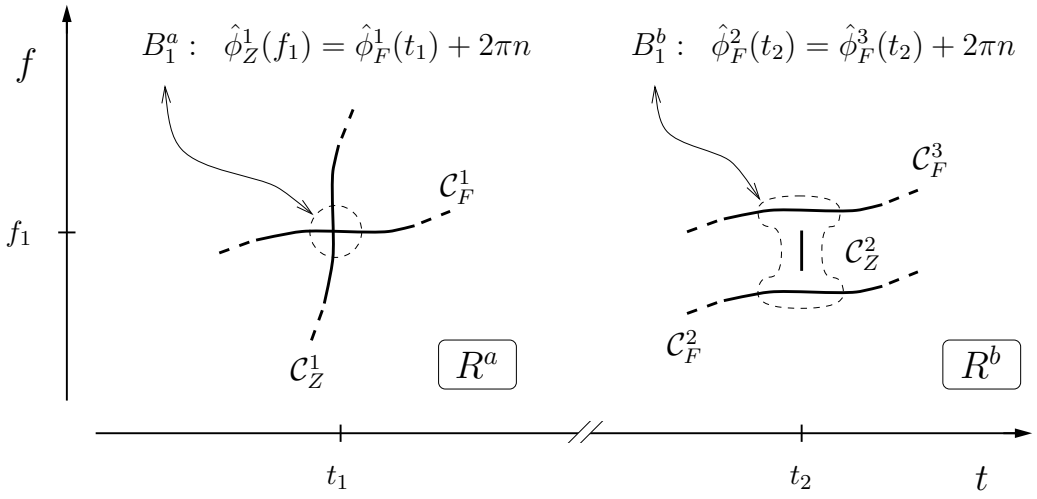

Bild 5.6: *Beispiele für zwei mögliche Phasenrelationsregeln R^a, R^b im Rahmen einer optimalen Phasenrekonstruktion für Konturen: In den beiden dargestellten Konturkonstellationen führen sie zu jeweils einer Nebenbedingung B_1^a bzw. B_1^b, hier indiziert als erste von weiteren Anwendungen der jeweiligen Regel. Zusammen mit weiteren Regeln ergibt sich ein Satz von Nebenbedingungen, die die Phasen der Konturen für das Gehör ausreichend festlegen (siehe Text). $\mathcal{C}_Z^k$ und $\mathcal{C}_F^i$ sind einzelne Zeit- bzw. Frequenzkonturlinien, deren zu rekonstruierende Phasenverläufe sich in Funktionen $\hat{\phi}_Z^k(f)$ bzw. $\hat{\phi}_F^i(t)$ ausdrücken. n ist beliebig ganzzahlig.*

halb beschränkter spektral/zeitlicher Abstände Nebenbedingungen etablieren, so daß das Beziehungssystem in Inseln zerfallen kann. Darin spiegelt sich die Natur der optimalen Phasenrekonstruktion wider, nach der auch im FTT-Betragsspektrum Inseln auftreten, zwischen denen keine Phasenrelationen herzustellen sind. Innerhalb dieser Inseln ist die Wahl eines absoluten Startwertes deshalb beliebig. Wie das Beziehungssystem innerhalb einer Insel im einzelnen aufzulösen ist, wird hier nicht weiter behandelt.

Warum soll die Phasendrift zwischen zwei Nebenbedingungen minimal sein? Alle Alternativen wären um ein Vielfaches von 2π versetzt. Entweder liegen die gefundenen Nebenbedingungen so dicht, daß durch die Alternativen Gl. (5.33) der Phasenregel beeinträchtigt wird. Oder sie liegen nicht besonders dicht, so daß sich die genaue Phasenentwicklung zwischen zwei Nebenbedingungen auch nicht mit Sicherheit an den Konturen ablesen läßt. Dann ist eine minimale Phasendrift eine ungefährliche, wenn auch willkürliche Vorgabe. Freilich liegt das Hauptproblem des Verfahrens darin, daß das Wissen über Phasenrelationen in entsprechenden Regeln vorliegt und daß diese auf konkrete Konturkonstellationen auch sicher angewandt werden können. Es könnte aber sein, daß sich bereits mit wenigen Phasenrelationsregeln eine auditiv bessere Phasenrekonstruktion als bei der bisherigen Phasenheuristik ergibt.

5.1.5.6 Zusammenfassung

Phasenrekonstruktion bildet das Kernproblem einer Signalrekonstruktion aus Konturen, der nicht mehr die ursprünglichen Konturphasen wie im Verfahren RKOP zur Verfügung stehen. Das Verhalten der Originalphasen kann als differentielle Phasenregel dargestellt werden, deren Gültigkeit innerhalb von Konturlinien kaum anzuzweifeln ist. Blockiert man die Freiheitsgrade der Regel, dann gelangt man zu einer Phasenheuristik, die im Verfahren

RKHP als Phasenrekonstruktion angewendet wird. Sie entspricht, nur für Frequenzkonturen betrachtet, der bekannten Phasenfortschreibung der Teiltonsynthese. Überhaupt entspricht der Rekonstruktionsvorgang von RKHP für Frequenzkonturen insgesamt dem der Teiltonsynthese – allerdings ist das Synthesefenster nun gehörangepaßt. Weil die Phasenheuristik sehr einfach auch auf Zeitkonturen übertragbar ist, steht mit RKHP erstmals ein Rekonstruktionsverfahren zur Verfügung, das beide Konturtypen ohne Kenntnis der Originalphasen sinnvoll verarbeiten kann.

So wie bei der Teiltonsynthese treten auch bei RKHP Phaseninkohärenz-bedingte Störungen durch nahe benachbarte Konturlinien auf. Nun aber entstehen sie auch durch benachbarte Zeitkonturlinien sowie durch benachbarte oder überkreuzende Linien unterschiedlichen Konturtyps. Dies reduziert die wahrnehmbare Rekonstruktionsqualität merklich gegenüber den Originalphasen im Verfahren RKOP.

Dennoch konnte ein Nachweis skizziert werden, daß es im auditiven Sinne eine optimale Rekonstruktion der Originalphasen geben muß. Darin zerfällt die Zeit/Frequenz-Fläche in Inseln, innerhalb derer alle Phasenrelationen rekonstruierbar sind. Zwischen den Inseln – wie auch zu absoluten Phasenwerten des ursprünglichen Phasenspektrums – gehen die Phasenrelationen verloren. Dies dürfte ohne Folgen für die Wahrnehmung bleiben, wenn man kein absolutes Phasenhörvermögen unterstellt. Es konnte abschließend sogar ein konkreter Verfahrensansatz als Weiterentwicklung der Phasenheuristik vorgestellt werden, der die blockierten Freiheitsgrade der Phasenregel reaktiviert und sogenannte Phasenrelationsregeln ins Spiel bringt. Leider sind letztere noch nicht ausreichend erforscht.

5.1.6 Einstellung der Verfahrensparameter

Einstellziel für die Parameter der Rekonstruktionsverfahren RKHP und RKOP ist es, das Signal mit bestmöglicher subjektiver Qualität aus seinen Konturen zurückzugewinnen. Die Qualitätsunterschiede beider Verfahren wurden bereits bei der Einstellung der Analyseparameter in Abschnitt 3.4 abgehandelt. Wie dort geschah die Einstellung auch hier durch Parametervariationen anhand von Sprachsignalen, die vom Autor untereinander und mit dem Original verglichen wurden. Es ergeben sich recht unkritische Einstellungen. Insbesondere hängen die gemeinsamen Parameter von RKHP und RKOP nicht von den Analyseparametern ab, obgleich umgekehrt die Wahl des Rekonstruktionsverfahrens Analyseparameter beeinflußt hatte. Nur die Parameter der Phasenrekonstruktion in RKHP sind lose an die Analysebandbreite gekoppelt. Das Ergebnis der Einstellung ist in Tabelle 5.1 aufgeführt. Zu sehen sind dort auch die Parameter der bekannten Teiltonsynthesen, zu denen die Unterschiede erst im nächsten Unterabschnitt diskutiert werden.

Zu den gemeinsamen Parametern von RKHP und RKOP zählen Synthesefenstertyp und Synthesebandbreite B^S_{3dB}. Kleinere oder wesentlich größere Synthesebandbreiten als 0,7 Bark beeinträchtigen die Qualität. Zu kleine Werte, die in die Nähe der Analysebandbreite oder darunter reichen, verstärken die Glättung der Schmalbandhüllkurve des Frequenzkonturanteils (Abschnitt 2.2). Zu große Werte bedeuten, daß der Zeitkonturanteil in spektraler Richtung geglättet wird, so daß die Wiedergabe impulshafter Anteile leidet. Für das Synthesefenster wurde im Entwurf in Abschnitt 5.1.2.2 zur Vereinfachung ein Typ aus der Familie Bn angenommen. Der vereinfachte Ansatz der Synthesekorrektur ist in der Praxis aber auch bei anderen Familien ohne spürbare Konsequenzen möglich. Zu

Tabelle 5.1: *Rekonstruktionsverfahren aus Konturen: Übersicht und Parameter.* RKOP *entspricht* RKHP*, hat statt einer heuristischen Phasenrekonstruktion aber Zugriff auf Originalphasen und simuliert dadurch optimale Phasenrekonstruktion.* TTSR *bezeichnet die ursprüngliche Heinbachsche Teiltonsynthese mit Rechteckfenster,* TTSD *die neuere Version mit Dreieckfenster.* $^+$ *von TTZM-Analyse und Codierung bestimmt.* * *siehe Fußnote S. 146.*

Parameter		**Verfahren**			
		RKOP	RKHP	TTSD	TTSR
Synthesefenstertyp		B4	B4	Dreieck	Rechteck
Synthesebandbreite	B^S_{3dB}	0,7 Bark	0,7 Bark	$\approx$ 500 Hz)*	$\approx$ 700 Hz)*
Zeitbreite des Fensters	T^S_{6dB}	$\approx 0,7/B^S_{3dB})^*$	$\approx 0,7/B^S_{3dB})^*$	1,25 ms	1,25 ms
Synthesefrequenzabst.	$\Delta\omega_S/(2\pi)$	0,05 Bark	0,05 Bark	+	+
Syntheseintervalldauer	T_S	$1/f_a$	$1/f_a$	1,25 ms	1,25 ms
Frequenzkonturen/TTZM		•	•	•	•
FK-Originalphasen		•	-	-	-
FK-Phasenübergabe	Δf_Φ	-	0,25 Bark	0,25 Bark	0,15 Bark
FK-Pegelanhebung in MSK	ΔL_M	1 dB	1 dB	-	-
Zeitkonturen		•	•	-	-
ZK-Originalphasen		•	-	-	-
ZK-Phasenübergabe	Δt_Φ	-	1,25 ms	-	-

kleine Fensterfunktionsgrade n haben denselben Effekt wie zu große Synthesebandbreiten. Ein Aufwand $n > 4$ bringt keine Verbesserungen mehr.

Die Pegelanhebung ΔL_M für die Frequenzkonturen wird in der in Anhang B.6 definierten Operation MSK benötigt. Mit ihrer Hilfe werden unmittelbar benachbarte oder überkreuzende Zeitkonturen etwa innerhalb einer Analysebandbreite um die Frequenzkonturen herum ausgeblendet, um Doppelrepräsentationen zu verhindern. Ein größerer Wert führt dazu, daß zu breite Stücke aus den Zeitkonturen entfernt werden. Dies wirkt sich im rekonstruierten Signal tendenziell tonalisierend aus. Ein zu kleiner Wert läßt Doppelrepräsentationen zu und macht sich durch eine gewisse Rauhigkeit bemerkbar. Obwohl sich die behelfsmäßige Definition von MSK sehr gut für Sprachsignale eignet, wurden bei dem FM-Signal aus Abschnitt 2.3.3 selbst mit dem Verfahren RKOP Störungen durch Doppelrepräsentation beobachtet. Hier sind für die Zukunft noch Verbesserungen möglich.

Das Syntheseraster gibt bei RKHP die minimale Zeit- und Frequenzauflösung bei der Reproduktion von Zeit- beziehungsweise Frequenzkonturen vor. Frequenzabstand $\Delta\omega_S$ und Intervalldauer T_S entsprechen hier den Werten auf der Analyseseite. Der erste Wert leitet sich aus der gerade noch wahrnehmbaren Frequenzänderung von Sinustönen ab, der zweite bietet als Kehrwert der Signalabtastrate f_a mit Sicherheit eine ausreichende zeitliche Auflösung. Exakte Übereinstimmung von Analyse- und Syntheseraster ist wegen der Konturrasterierer FR/ZR allerdings nicht nötig. Natürlich bedeuten feinere Raster einen höheren Rechenaufwand. Die mitübertragenen Originalphasen ermöglichen bei RKOP prinzipiell ein gröberes Syntheseraster als bei RKHP. Der Auflösungsbedarf von Konturverläufen kann nämlich in gewissen Grenzen gegen Originalphaseninformation getauscht werden, wie sich später in Abschnitt 6.2.1 ergibt.

Die für die Phasenheuristik der Frequenzkonturen gewählte Grenze Δf_Φ zur Phasenübergabe orientiert sich grob an der halben Analysebandbreite. Ist sie ihr gegenüber zu klein, dann steigen die Phaseninkohärenz-bedingten Störungen. Ein merklich größerer Wert richtet dagegen kaum Schaden an, so daß beispielsweise die Einstellung von TTSD auch mit der sehr kleinen Analysebandbreite von HB-TTZM verwendet werden kann. Bei der Phasenheuristik der Zeitkonturen stellt die angegebene Grenze Δt_Φ nur einen sehr groben Anhaltspunkt dar, da sowieso ein Teil des Nutzens von Zeitkonturen durch suboptimale Phasenrekonstruktion verlorengeht.

5.1.7 Rekonstruktionsfehler im Vergleich mit Teiltonsynthesen

Beschränkt man sich auf die Verarbeitung von Frequenzkonturen, dann unterscheiden sich RKHP und die bekannten Teiltonsynthesen TTSR und TTSD prinzipiell nur im Synthesefenster (Abschnitte 5.1.5.2, 5.1.5.3). Wegen der gemeinsamen, mangelbehafteten Phasenheuristik sind bei allen drei Verfahren gleichartige Phaseninkohärenz-bedingte Störungen hinzunehmen. Sie werden nur bei RKOP vermieden, dessen Synthesefenster mit dem von RKHP übereinstimmt. Somit können unter gleichartigen Voraussetzungen drei verschiedene Möglichkeiten der spektral/zeitlichen Fehlerformung verglichen werden, die die Wahrnehmbarkeit von (eben) Synthesefenster-kontrollierbaren Störungen beeinflussen. Außerdem wird diskutiert, warum beide Störungstypen nützlich sind, um fehlende Zeitkonturverarbeitung zu verschleiern.

Woher kommen eigentlich die Synthesefenster-kontrollierbaren Störungen grundsätzlich? Die wesentliche Ursache ist darin zu sehen, daß das in der FTT-Codierung verwendete, vollständige Spektrum beim Übergang auf eine Konturcodierung abseits von Konturen nullgesetzt wird (Abschnitt 5.1.3). Weitere Ursache ist die Quantisierung von Zeit, Frequenz und Pegel.

5.1.7.1 Fehlerformung durch das Synthesefenster

Um die Synthesefenster im Zeit- wie auch im Frequenzbereich vergleichen zu können, wurden in Tabelle 5.1 zusätzlich die 6dB-Zeitbreite T^S_{6dB} der Fensterfunktion B4 und die 3dB-Bandbreiten B^S_{3dB} des Rechteck- und des Dreieckfensters ermittelt. [9] Die Bandbreiten bei einer Frequenz von 1 kHz – an der 0,7 Bark rund 120 Hz entsprechen – verdeutlichen, daß Rechteck- und Dreieckfenster Rekonstruktionsfehler über einen viel größeren Spektralbereich als das Fenster B4 verschmieren können. Verschärfend tritt hinzu, daß die Fourier-Transformierten der ersten beiden Fenster, eine si-Funktion beziehungsweise deren Quadrat, nur mit 20 beziehungsweise 40 dB/Dekade absinken. Dagegen erreicht B4 eine Steilheit von 80 dB/Dekade. Zwar fällt hier T^S_{6dB} mit etwa 6 ms länger als bei den bisherigen Fenstern aus. Weil die Bandbreite aber frequenzabhängig ist und immer kleiner als die Analysebandbreite bleibt, treten keine wahrnehmbaren Glättungseffekte auf, die mit einem Festwert $T^S_{6dB} = 6$ ms bei höheren Frequenzen sonst zu befürchten wären.

[9] T^S_{6dB} für B4 wurde aus Bild 3.6c abgelesen und mit Gl. (3.3) entnormiert. B^S_{3dB} für die Rechteckfunktion nach Gl. (1.24) läßt sich mit Hilfe der vereinfachten Fourier-Korrespondenz $|\mathcal{F}\{h^S(t)\}| \sim |si(\omega T_S/2)|$ ermitteln [Mar82]. Nach Abschnitt 1.5.4.3 ist für die Dreieckfunktion das Quadrat der Korrespondenz anzusetzen.

Aufgrund der spektralen Verschmierung durchbrechen die Störungen bei den bisherigen Fenstern die Mithörschwellen des Nutzsignalanteils. Beim gehörangepaßten Fenster B4 wird dies vermieden, wie die nahezu verfälschungsfreie Rekonstruktion mit RKOP belegt. Leider ist das in der Praxis von Sprachsignalen offenbar unvorteilhaft, solange ausschließlich Frequenzkonturen verarbeitet werden. Dies wird gleich noch objektiviert.

Warum hilft das gehörangepaßte Fenster nicht, um wahrnehmbare Phaseninkohärenzbedingte Störungen zu vermeiden? Das liegt daran, daß dieser Störungstyp nur für Synthesebeiträge innerhalb einer Frequenzgruppe wahrnehmbar und somit von Natur aus schmalbandig ist. Hierauf hat das Synthesefenster praktisch keinen Einfluß. Außerdem erreicht der Pegel des Störanteils innerhalb der Frequenzgruppe den des Nutzanteils. Sonst könnte die Hüllkurve des Frequenzgruppensignals nicht so deutlich verändert werden, wie es bei den Modulationssignalen (Abschnitt 2.3) zu beobachten war. Einen derart hohen Störanteil aber kann kein Synthesefenster unter den Hörschwellen des Nutzanteils verbergen.

5.1.7.2 Nutzen der Störungen als Zeitkonturersatz

Warum ist es bei Sprachsignalen und fehlender Zeitkonturverarbeitung günstiger, Synthesefenster-kontrollierbare Störungen moderat über der Hörschwelle zu halten, indem das Dreieckfenster von TTSD anstelle des gehörangepaßten Fensters von RKHP verwendet wird? Der zugehörige Störteppich ist besonders gut in spektral/zeitlichen Gebieten vertreten, in denen die fehlenden Zeitkonturen sonst spektrale Lücken hinterlassen würden. Im Sprachsignal in Bild 4.2a auf S. 104 beispielsweise erkennt man, daß gerade dort viele Kurzverläufe von Frequenzkonturlinien vorkommen, wo in 4.2b längere Zeitkonturlinien zu erkennen sind. Ihre zahlreichen Anfänge und Enden heben in diesen Gebieten das Niveau des Störteppichs besonders an. Dadurch können besagte Lücken aufgefüllt werden, ähnlich wie in Bild 2.12 auf S. 51, sofern es das Synthesefenster zuläßt.

Weiterhin gibt es einen speziellen Verbundeffekt von Phaseninkohärenz, Synthesefenster und den Analyseparametern, der Zeitkonturen teilweise imitieren kann. Er tritt bereits bei einem einfachen Testsignal auf, dem eingeschalteten Sinuston, und er spielt sicherlich auch bei komplizierten Signalen eine Rolle. Man betrachte dazu die Nebenkonturlinien im Verschneidungsgebiet des Einschaltvorganges, dargestellt in Bild 5.5b,c auf S. 134. Vom Standpunkt einer gehörorientierten Analyse liegen sie in der Nähe der Hörschwelle. Hierbei kommen die Nachhörschwelle des Zeitkonturhauptastes und die Mithörschwelle des Frequenzkonturhauptastes gemeinsam in Betracht. Zudem ist zu berücksichtigen, daß die zeitliche Dauer und die spektral eingenommene Breite der Nebenkonturen begrenzt sind.

Vom Standpunkt der Rekonstruktion dagegen sind mit den Nebenkonturen eine Reihe von Konturlinien zu verarbeiten, die sehr dicht beieinander liegen. Wegen der Phaseninkohärenz überlagern sich ihre Synthesebeiträge innerhalb einer Frequenzgruppe im Ohr des Zuhörers statistisch – also völlig anders als bei wohldefinierten Phasenrelationen. Da sehr viele dieser Synthesebeiträge in eine Frequenzgruppe fallen, entstehen Frequenzgruppenpegel, die den ursprünglich analysierten Nebenkonturpegel übersteigen. Im Ergebnis kann ein ‘Rauschstoß’ wahrgenommen werden, der den Schaltknack näherungsweise imitiert. Dies wird durch fünf Faktoren begünstigt:

- Ein Analysefenster mit niedrigerem Grad provoziert von vornherein ein ausgedehntes und hochpegeliges Verschneidungsgebiet.
- Die Welligkeiten des Verschneidungsgebietes lösen mehr Nebenkonturlinien aus, wenn das FTT-Spektrum nicht oder nur schwach zeitlich geglättet wird, oder wenn die Ausgeprägtheitschwelle niedrig eingestellt ist.
- Eine größere Analysebandbreite streckt das Verschneidungsgebiet in spektraler und staucht es in zeitlicher Richtung, wodurch die Störenergie eher aus der Mithörschwelle des Frequenzkonturhauptastes herausrückt.
- Es wird auf Zeitkonturen verzichtet, so daß die Nachhörschwellen der Zeitkonturhauptäste die wahrnehmbare Störenergie nicht drosseln oder verdecken.
- Das gewählte Synthesefenster verschmiert Störenergie bevorzugt in spektraler Richtung.

Bei der Einstellung ZFKI oder ZFKII (Tabelle 3.2 auf S. 87) in Verbindung mit RKOP ohne Zeitkonturverarbeitung hört man beispielsweise keinen Rauschstoß. Offensichtlich verhindern die 'richtigen' Phasen, daß sich die Nebenkonturbeiträge zu hohen Frequenzgruppenpegeln überlagern. Bei ZFKII/RKHP ohne Zeitkonturverarbeitung ist der Rauschstoß sehr schwach, bei M-TTZM/TTSD und bei SM-TTZM/TTSD zunehmend deutlicher hörbar.

5.1.8 Zusammenfassung

Dieser Abschnitt versuchte, ein optimales Rekonstruktionsverfahren zu entwickeln, das neben Frequenzkonturen auch Zeitkonturen verarbeitet und hörbare Rekonstruktionsfehler vermeidet. Sein Ansatz gelingt dadurch, daß die Rücktransformation des komplexen FTT-Spektrums (RFTT) eingeführt wird. Zusammen mit der FTT ergibt sich daraus ein neuartiges Transformationspaar T und R, das für das Gehör transparente Codierung ermöglicht (FTT-Codierung). In die Mitte dieses Rahmens wird eine Reihe von zusätzlichen Verarbeitungsschritten eingefügt, bis man schließlich eine Analysehälfte erhält, die nur noch die Zeit- und Frequenzkonturen liefert. Sie entspricht der bereits bekannten Konturanalyse. Die Synthesehälfte dagegen entspricht prinzipiell dem gesuchten Verfahren. Die Schritte im einzelnen sind folgende:

Einführung FTT-Rücktransformation (RFTT): Die RFTT stellt das Zeitsignal als Fourier-Rückintegral über einer korrigierten Form des komplexen, zeitvarianten FTT-Spektrums dar. Die Korrektur setzt sich aus drei Komponenten zusammen. Die erste entspricht dem Laufzeitausgleich aus Kapitel 3, der die maximale Fensteröffnung an allen Analysefrequenzen auf den gleichen Zeitpunkt schiebt. Die zweite normiert frequenzabhängig auf die Höhe der maximalen Fensteröffnung, so daß die Schar der wirksamen Analysefenster gleichzeitig dieselbe maximale Höhe erreicht. Die dritte ist ein frequenzabhängiger komplexer Drehfaktor, der den Synthesezeitpunkt auf den Zeitpunkt der gemeinsamen maximalen Fensteröffnung zurückverlegt. Das rücktransformierte Zeitsignal weist gegenüber dem Original lineare Laufzeit- und Frequenzgangverzerrungen auf, die im Rahmen vernünftiger Werte nachweislich unhörbar sind.

Rahmen für FTT-Codierung: Das Transformationspaar T und R, realisiert durch komplexe Bandfilterbänke, macht den FTT/RFTT-Formalismus für Codierungen nutzbar. Es verarbeitet anstelle des komplexen FTT-Spektrums das betragsgleiche FTT-Bandpaßspektrum, sieht dessen zeitliche und spektrale Abtastung vor und führt ein Synthesefenster mit eigenem Laufzeitausgleich ein. Der Wechsel auf das FTT-Bandpaßspektrum ist wesentlich, weil seine Phase entkoppelte Analyse- und Synthesezeiten und nichtidentische Analyse- und Synthesefrequenzen erlaubt. Spektral/zeitliche Abtastung und spätere Codierungsmaßnahmen zwischen T und R verursachen Quantisierungsfehler. Das Synthesefenster in R hat hier die Aufgabe, sie unter die Hörschwellen des Nutzsignals zu formen. Im Rahmen dieser Arbeit ist die Impulsantwort eines Bessel-Tiefpasses vierten Grades (B4) mit einer 3dB-Bandbreite von 0,7 Bark gut geeignet.

Unabhängig von der Verwendung in dieser Arbeit bilden T und R einen flexiblen Rahmen für Codierungen mit zeitvarianten Kurzzeitspektren. Zwar liegt die Rohdatenrate zwischen T und R um einen gewissen Faktor über dem informationstheoretischen Minimum, wie es beispielsweise bei etablierten, 'kritisch' abgetasteten Filterbänken erreicht wird. Auch ist keine perfekte Rekonstruktion möglich. Dafür kann man Analyse- und Synthesefenster im Typ und in der Frequenzabhängigkeit der Bandbreite in weiten Grenzen festlegen. In manchen Fällen dürfte die resultierende spektral/zeitliche Formbarkeit des Quantisierungsfehlers wichtiger als eine potentiell perfekte Rekonstruktion sein. Letztere nützt angesichts fehlerbehafteter Codierungsmaßnahmen wenig, wenn ungünstige Synthesefenster bereits kleine Fehler spektral/zeitlich auffällig verschmieren. T und R können anschaulich mit dem Formalismus der Wavelet-Transformation dargestellt werden (Anhang D.2).

Ansatz FTT-Konturcodierung: Hierzu braucht man die Abtastwerte des komplexen FTT-Bandpaßspektrums zwischen T und R nur an den Rasterorten weiterzureichen, an denen die Zeit- und Frequenzkonturen des FTT-Betragsspektrums verlaufen. Diese Abtastwerte werden aus dem ursprünglichen Raster gewissermaßen 'ausgesiebt', die übrigen Werte werden zu null angenommen. Um den Verlust an Energiedichte zu kompensieren, sind für die gesiebten Abtastwerte frequenzabhängige Bewertungsfaktoren einzuführen. Sie unterscheiden sich für beide Konturtypen und werden so abgeglichen, daß Sinustöne beziehungsweise Impulse ohne Frequenzgangverzerrungen reproduziert werden. Um energetische Doppelrepräsentationen – etwa bei überkreuzenden Konturlinien – zu verhindern, müssen Abtastwerte gegenseitig maskiert werden, die von verschiedenen Konturtypen gesiebt wurden. Die insgesamt erreichbare Qualität einer solchen Codierung ist bei Sprachsignalen so gut, daß man den Unterschied zum Original nur im direkten Paarvergleich unter optimalen Abhörbedingungen wahrnehmen kann. Allerdings finden sich synthetische FM-Signale, bei denen der Unterschied deutlicher werden kann. Dies wird auf die behelfsmäßige Definition der gegenseitigen Maskierung zurückgeführt.

Abspaltung Rekonstruktionsverfahren RKOP: Aus dem vorigen Schritt kann man eine noch unvollständige Version des angestrebten Rekonstruktionsverfahrens herauslösen, für welche die Phasen entlang der Konturverläufe mitzuübertragen sind (RKOP – Rekonstruktion aus Konturen mit Originalphase). Die zuvor benötigten gesiebten Abtastwerte werden nun direkt aus Konturen und Konturphasen ermittelt. Dabei können beliebig quantisierte Konturen auf das Syntheseraster umgerechnet

werden, wenn diesen ein anderes Zeit/Frequenz-Raster zugrunde liegt. Ein ausreichend feines Syntheseraster gewährleistet einen vernachlässigbaren Umrechnungsfehler. Die erreichbare Rekonstruktionsqualität entspricht der des vorigen Schrittes.

Wandlung RKOP in RKHP durch Phasenrekonstruktion: Der letzte Schritt besteht darin, die Notwendigkeit mitübertragender Phasenverläufe zu eliminieren, um sie aus den Konturen selbst zu rekonstruieren. Es kann ein Nachweis skizziert werden, daß dies ohne wahrnehmbaren Qualitätsverlust möglich ist. Eine solche optimale Phasenrekonstruktion wird in dieser Arbeit aber nur ansatzweise vorgestellt. Immerhin läßt sich das Verhalten der Konturphasen durch eine differentielle Phasenregel beschreiben. Sie besagt, daß die Phasen nur langsam und stetig von einer einfachen Gesetzmäßigkeit abweichen können. Indem man die Abweichung außer acht läßt, gelangt man zu einer Phasenheuristik. Für Frequenzkonturen stimmt sie mit derjenigen der Teiltonsynthese überein, auf Zeitkonturen ist sie in vergleichbarer Weise anwendbar. Damit wurde eine behelfsmäßige Version (RKHP – Rekonstruktion aus Konturen mit heuristischer Phase) des angestrebten autonomen Verfahrens realisiert, dessen Rekonstruktionsqualität gegenüber RKOP leider merklich absinkt.

Optimale Rekonstruktion aus Konturen scheint demnach zwar prinzipiell erreichbar, scheitert einstweilen aber an der Unzulänglichkeit der Phasenrekonstruktion. Die resultierenden Phaseninkohärenz-bedingten Störungen im Verfahren RKHP mindern bei Sprachsignalen auch den wahrnehmbaren Nutzen des Zeitkonturbeitrages. Mit der sehr guten Rekonstruktionsqualität von RKOP läßt sich immerhin der Erfolg einer optimalen Phasenrekonstruktion simulieren. Ebenso läßt sich das über R eingebrachte Synthesefenster als optimal bestätigen.

Vergleicht man RKHP ohne Zeitkonturverarbeitung mit den Varianten der Teiltonsynthese, so unterscheiden sie sich prinzipiell nur im Synthesefenster. Im Gegensatz zum optimalen Fenster in RKHP bewirken die einfachen Fenster eine hörbare Fehlerverschmierung in spektraler Richtung. Bei Sprachsignalen haben diese Synthesefenster-kontrollierbaren Störungen – wie auch ein 'Rauschstoß'-Effekt durch Phaseninkohärenz – eine nützliche Wirkung: Der Nachteil fehlender Zeitkonturen kann in gewissem Maße verschleiert werden. Der Vorteil der Zeitkonturverarbeitung bei RKHP andererseits kommt wegen der Phaseninkohärenz-bedingten Störungen nicht voll zur Geltung. Bei Sprachsignalen fällt deshalb eine Rekonstruktion mit RKHP aus beiden Konturtypen gegenüber einer Teiltonsynthese mit Dreieckfenster (TTSD) aus Frequenzkonturen subjektiv nicht viel besser aus.

Konzeptionell trennt man sich für eine optimale Rekonstruktion davon, daß Konturen direkt Parameter von Synthesesinusschwingungen vorgeben. Diese Grundlage der Heinbachschen Teiltonsynthese kann nur einen praktischen Kompromiß bedeuten. Ein optimales Synthesefenster bewirkt nämlich, daß sich sprunghafte Pegelübergänge einer Frequenzkonturlinie optimal geglättet im Amplitudenverlauf der rekonstruierten Sinusschwingung niederschlagen. Der Frequenzverlauf wird ebenfalls nicht direkt übernommen, da er als Folge einer optimalen Phasenrekonstruktion in der rekonstruierten Sinusschwingung ein gewisses 'Eigenleben' führen kann. Optimale Rekonstruktion läßt sich als Konsequenz aller Entwicklungsschritte nunmehr so beschreiben:

Konzept optimaler Rekonstruktion: Die Parameter eines Konturpunktes bestimmen direkt Amplitude, Frequenz und Zeitlage von einzelnen Sinustonimpulsen (Wave-

lets), die es für die Menge aller Konturpunkte zu überlagern gilt. Die Hüllkurven der Wavelets entsprechen dem optimalen, gehörangepaßten Synthesefenster. Ihre ebenfalls benötigte Phase läßt sich, soweit für die Wahrnehmung relevant, aus dem Gesamtzusammenhang der Menge entwickeln. Alternativ kann sie aus der Phase des konturierten Spektrums übernommen werden. Doppelrepräsentationen von spektral/zeitlichen Bereichen durch Konturpunkte unterschiedlicher Konturtypen sind vor der Überlagerung auszuschließen.

5.2 Rekonstruktion aus Kontur/Textur-Repräsentationen

Der Konturanteil einer Kontur/Textur-Repräsentation läßt sich weiterhin mit den Rekonstruktionsverfahren RKHP, RKOP oder, wenn Zeitkonturen nicht explizit vorgesehen sind (Einstellung KTXOZ), auch mit der Teiltonsynthese verarbeiten. Für den zu überlagernden Texturanteil aber muß ein Weißes Rauschen spektral und zeitlich nach Maßgabe der Texturhüllfläche geformt werden. Dazu wurden zwei Lösungen erarbeitet. Die universellere und prinzipiell bessere beruht auf einer zeitvarianten Filterung des Rauschens. Sie läßt sich auf Verfahrensteilen von RKHP oder RKOP aufbauen und führt direkt auf ein Zeitsignal. Die andere moduliert Frequenzkonturen des Rauschens spektral und zeitlich, welche dann mit geringerem Aufwand in einer Teiltonsynthese weiterverarbeitet werden können. Diese Lösung ist nur für KTXOZ geeignet.

5.2.1 Erweitertes Verfahren RKHP mit Textur (RKHPTX)

Eine Erweiterung von RKHP um einen Verarbeitungszweig für den Texturanteil zeigt Bild 5.7. Die beiden Zweige für den Rekonstruktionsbeitrag der prägnanten Zeit- und Frequenzkonturen sind nicht nochmal extra abgebildet. Sie stimmen mit den rekonstruktionsseitigen Zweigen von Bild 5.3b,c auf S. 130 überein und schließen hier, als Pfeile angedeutet, bei der Überlagerung an. Im Gesamtverfahren RKHPTX können somit Kontur- und Texturverarbeitung dieselbe Rücktransformation R nutzen.

Die Transformation T′ verwendet den Fenstertyp und die Analysebandbreite, die auch in der FTT der Kontur/Textur-Analyse eingestellt worden sind (Tabellen 3.2 und 4.1). Sie gilt aber von vornherein über dem Zeit/Frequenz-Raster $\{(\omega_{S_m}, lT_S)\}$ von R errichtet, da beide im Decoder realisiert werden. Zuerst denke man sich die Operationen zwischen Hin- und Rücktransformation entfernt. Das Rauschsignal $n(t)$ der Rauschquelle WR gelangt für das Gehör unverändert in das Ausgangssignal $\hat{s}(t)$, weil T′ und R ein aufeinander abgestimmtes Transformationspaar bilden.

Die zwischen Hin- und Rücktransformation eingefügten Operationen steuern das FTT-Bandpaßspektrum $n^{LB}(\omega, t)$ des Rauschens nach Maßgabe der Texturhüllfläche $L_{TX}(f, t)$. Im Mittelpunkt steht der spektral/zeitliche Modulator SZM. An jedem Rasterort (ω_{S_m}, lT_S) geben die Pegelwerte der Texturhüllfläche positive Verstärkungswerte für die Abtastwerte des Rauschspektrums vor. Diese Art einer zeitvarianten Filterung wurde von Portnoff am Beispiel der diskreten Fourier-Transformation beschrieben [Por80]. Das gehörangepaßte Synthesefenster in R spielt in diesem Zusammenhang die Rolle, daß es

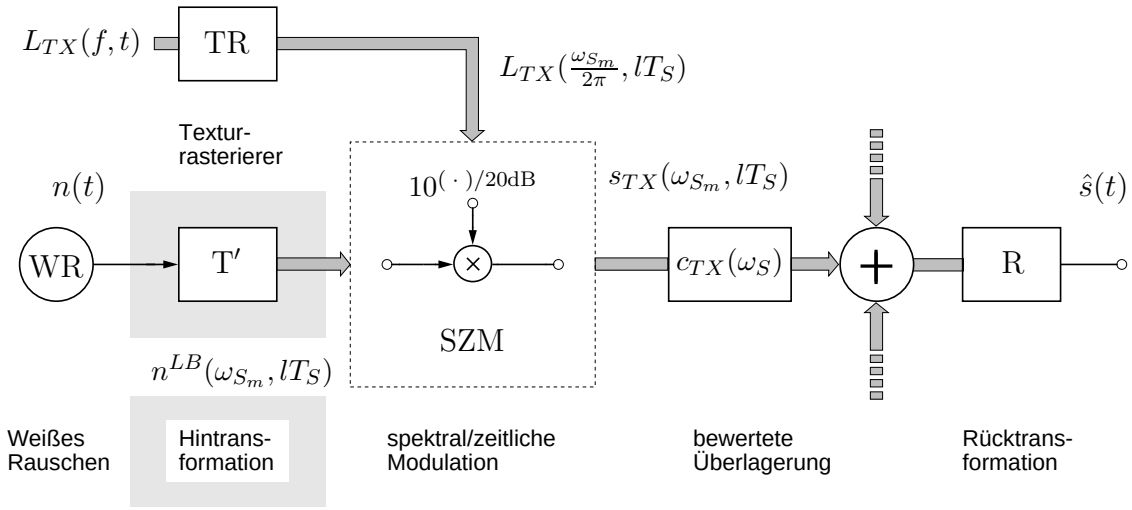

Bild 5.7: *Ergänzung des Rekonstruktionsverfahrens* RKHP *um eine Texturverarbeitung zum Verfahren* RKHPTX. *Die beiden Zweige der Konturverarbeitung (Bild 5.3b,c auf S. 130) sind an der Überlagerung angeschlossen zu denken. Zu sehen ist eine zeitvariante Rauschfilterung, die von der Texturhüllfläche $L_{TX}(f,t)$ gesteuert wird.* SZM *ist für einen einzelnen Kanal ω_{S_m} dargestellt. Der grau unterlegte Teil konnte in der Praxis durch eine direkte Übernahme $n^{LB}(\omega_{S_m}, lT_S) = n(t)|_{t=lT_S}$ des Rauschens in die einzelnen Kanäle eingespart werden.*

lokale Änderungen der spektral/zeitlichen Steuerung $L_{TX}(f,t)$ auf lokale Änderungen im rekonstruierten Rauschspektrum beschränkt. Bei einem ungünstigen Fenster, etwa dem Rechteckfenster, würde sich ein zeitlicher Sprung in einem begrenzten Frequenzbereich der Steuerung plötzlich breitbandig im Synthesesignal auswirken (Knacken im Rauschen).

Die praktischen Erfahrungen zeigen, daß die Transformation T′ eingespart werden kann, indem dahinter jedem Kanal ω_{S_m} direkt das Rauschsignal $n(t)$ eingespeist wird. Dies ist eigentlich für eine zeitvariante Filterung nicht ganz korrekt: Normalerweise stellt die in T′ realisierte Aufspaltung in Teilbandkanäle sicher, daß die Steuerung nicht bandfremde Spektralanteile des Eingangssignals in das rekonstruierte Teilband hineinmischen kann. Das Synthesefenster in R und das Analysefenster in T′ zusammengenommen garantieren also, daß beliebige Steuerungen keine spektralen Auswirkungen außerhalb eines Teilbandes haben können. Die Mischeffekte durch Weglassen von T′ sind aber offenbar bei den Texturhüllflächen von Sprache nicht wesentlich.

Der Texturrasterierer TR wird aus ähnlichem Grund eingesetzt, aus dem auch die Konturrasterierer FR/ZR in RKHP eingeführt wurden. Er entkoppelt die Quantisierung der Texturhüllfläche, die hier allgemein über dem Zeit/Frequenz-Kontinuum definiert wurde, vom Syntheseraster. Wenn üblicherweise die Texturhüllfläche viel gröber als das Syntheseraster codiert wird, weist TR einem Syntheserasterort einfach den nächstgelegenen Stützwert der codierten Texturhüllfläche zu. Dies entspricht einer Treppenstufenapproximation der Fläche und ist für nicht zu grobe Raster ausreichend, zumal die Wirkung des Synthesefensters die Stufen spektral und zeitlich ausglättet (vgl. TXD Abschnitt 6.3.2).

Schließlich wird ein frequenzabhängiger Bewertungsfaktor $c_{TX}(\omega_S)$ benötigt. Er stellt sicher, daß die Kontur/Textur-Repräsentation eines Weißen Rauschens tatsächlich ein 'ungefärbtes' Rauschen im Ausgangssignal $\hat{s}(t)$ hervorruft. Weil nämlich die Analysebandbreite B^A_{3dB} über der Frequenz zunimmt, steigt bei Weißem Rauschen die mittlere Lei-

stung am Ausgang eines Analysefilters proportional. Da bei Kontur/Textur-Analyse der leistungsmäßig wesentliche Teil des Rauschens in die Textur gelangt, wächst $L_{TX}(f,t)$ über der Frequenz im Mittel entsprechend. Der Nenner in

$$c_{TX}(\omega_S) = \frac{c_{bal}}{\sqrt{B^A_{3dB}(\omega_S)}} \tag{5.41}$$

gleicht dies aus. Mit c_{bal} im Zähler wird die Balance zwischen Kontur- und Texturanteil im Ausgangssignal eingestellt. Praktisch geschieht dies entweder meßtechnisch, indem die Verstärkung bei der Verarbeitung eines Weißen Rauschens auf die eines Sinustons abgestimmt wird. Oder man verwendet ein Sprachsignal und sucht gehörmäßig bestmögliche Rekonstruktionsqualität. Beide Vorgehensweisen liefern vergleichbare Ergebnisse.

Die Phasenheuristik in RKHP ist bei Kontur/Textur-Verarbeitung von Sprachsignalen ausreichend. Anders als bei reiner Konturverarbeitung würde eine optimierte Phasenrekonstruktion die wahrnehmbare Qualität nur geringfügig verbessern. Dies konnte mit Hilfe der Originalphasen der prägnanten Konturen überprüft werden, wofür RKOP anstelle von RKHP um eine Texturverarbeitung erweitert wurde. Der Grund ist die behelfsmäßige Unterscheidung von prägnanten Konturen und Textur mit Hilfe der Linienlänge. Unvermeidliche Fehlentscheidungen bewirken ähnliche Verfälschungen wie die Phaseninkohärenz-bedingten Störungen von RKHP. Konturen, die unberechtigterweise der Textur zugeordnet wurden, sind nämlich als Einzelobjekte verschwunden. Man kann sie danach nicht mehr phasenkohärent zu den prägnanten Konturen rekonstruieren (Abschnitte 4.1.1 und 4.6). [10]

5.2.2 Texturrekonstruktion mit Hilfe von Frequenzkonturen

Wenn mit Einstellung KTXOZ bei der Kontur/Textur-Analyse auf explizite Zeitkonturverarbeitung verzichtet wird, kann man Frequenzkontur- und Texturanteil gleichermaßen mit Hilfe einer Teiltonsynthese weiterverarbeiten. Im Vergleich zu RKHPTX reduziert sich der Rechenaufwand erheblich. Für den Texturanteil nimmt man das Frequenzkontursignal $\mathcal{C}_F^{WR}(t)$ eines Weißen Rauschens, staucht es in der Zeit um einen Faktor $c = 2 \ldots 4$ und moduliert dessen Pegelhüllfläche nach den Vorgaben der Texturhüllfläche $L_{TX}(f,t)$. Man erhält so ein künstliches Frequenzkontursignal

$$\mathcal{C}_F^{TX}(t) = \left\{ (L_{TX}(f,t) + L_{bal}, f, t) \middle| \; (L,f,t) \in \mathcal{C}_F^{WR}(c \cdot t) \right\}, \tag{5.42}$$

das nun die Textur repräsentiert und in eine Teiltonsynthese eingespeist werden kann. Das Stauchen der Zeitachse unterdrückt die in Abschnitt 2.5 beobachtete Tonalisierung von Rauschen. Wenn nämlich die Frequenzkonturlinien bei unverändertem mittleren spektralen Abstand zeitlich schneller fluktuieren, dann vergrößert sich ihre spektrale Unschärfe, womit die spektrale Lückenbildung vermieden wird. In der Realisierung braucht man übrigens nur einen begrenzten Ausschnitt von $\mathcal{C}_F^{WR}(t)$ vorzusehen, der zyklisch wiederholt wird. Allerdings ist der Ausschnitt so zu legen, daß keine Schnitt- oder Periodizitätseffekte wahrnehmbar werden.

[10] Es ist im übrigen fraglich, ob eine optimale Phasenrekonstruktion prinzipiell noch machbar ist, da die exakten Verläufe von nichtprägnanten Konturen entscheidende Hinweise auf die Phasenrelation enthalten (vgl. Abschnitt 5.1.5.5).

Die Modulation der Pegelhüllfläche geschieht dadurch, daß der ursprüngliche Pegel L eines Konturpunktes (L, f, t) gegen den Pegelwert der Texturhüllfläche $L_{TX}(f, t)$ an dieser Stelle ausgetauscht wird. Dadurch ist keine frequenzabhängige Bewertung wie bei RKHPTX nötig. Da Konturverläufe üblicherweise feiner als Texturrepräsentationen quantisiert werden, ist auch hier der nächstnähere Pegelwert maßgeblich. Mit L_{bal} stellt man die Pegelbalance zu den prägnanten Frequenzkonturen in gleicher Weise wie bei RKHPTX ein. Damit sich die Phasenrekonstruktion beider Anteile nicht gegenseitig beeinflußt, müssen sie getrennten Teiltonsynthesen zugeführt werden, deren Ausgangssignale zu überlagern sind.

In vergleichbarer Weise wie bei der zeitvarianten Filterung bestimmt auch hier das Synthesefenster, wie die Steuerung des synthetisierten Rauschspektrums spektral und zeitlich lokalisiert ist. Um das ungünstige Rechteckfenster zu vermeiden, sollte zumindest die Teiltonsynthese TTSD mit ihrem Dreieckfenster eingesetzt werden. Natürlich ist dies immer noch schlechter als das gehörangepaßte Fenster von RKHPTX, was aber bei den vorgestellten Kontur/Textur-Repräsentationen und Sprachsignalen kaum ins Gewicht fällt. Bei nicht festgelegtem Signaltyp aber weist RKHPTX einen gewissen Qualitätsvorteil auf.

5.3 Zusammenfassung

Behandelt wurden Verfahren zur Rekonstruktion des ursprünglichen Signals aus Konturen und aus Kontur/Textur-Repräsentationen. Dies war notwendig, weil die bisherigen Teiltonsynthesen nur für Frequenzkonturen anwendbar sind, nicht aber für Zeitkonturen. Es ging aber auch darum, aus Konturen eine optimale Rekonstruktion zu erzielen, die – im Gegensatz zu den bisherigen Teiltonsynthesen – keine wahrnehmbaren Störungen hinzufügt.

Für Konturen ohne Textur-Repräsentation bereitete ein aufwendiger Entwicklungsprozeß den Weg für eine optimale Rekonstruktion. Er ist in Abschnitt 5.1.8 detaillierter zusammengefaßt. Das Ziel wurde nur theoretisch erreicht, für die Praxis ergaben sich zwei Verfahren mit unterschiedlichen Nachteilen. In beiden wird das Signal dadurch überlagert, daß jeder Konturpunkt die Parameter eines Sinustonimpulses (Wavelet) bestimmt. Seine Hüllkurve wird von einem geeigneten Synthesefenster vorgegeben. Die Festlegung seiner Phase erweist sich als das zentrale Problem einer Rekonstruktion.

Rekonstruktion aus Konturen mit Originalphase (RKOP): Dieses Verfahren vermeidet wahrnehmbare Störungen fast völlig. Es vertraut allerdings darauf, daß die Phasenverläufe unter den Konturen des Originalspektrums mitübertragen werden. Bei Teiltonsynthesen bekannte Störungen sind hier nicht wahrnehmbar, weil ein gehörangepaßtes Synthesefenster verwendet wird und weil die Originalphasen Phaseninkohärenzen ausschließen. Das Synthesefenster entspricht vom Typ etwa dem Analysefenster und weist eine 3dB-Bandbreite von 0,7 Bark auf. Um energetische Doppelrepräsentation zu vermeiden, etwa wenn sich Konturlinien überschneiden, blendet eine nicht unproblematische Operation bestimmte Konturpunkte aus. Weil sie bisher nur behelfsmäßig spezifiziert wurde, bleiben sehr geringe Störungen übrig. Bei synthetischen FM-Signalen können diese etwas deutlicher werden.

Es wurde ein Nachweis skizziert, daß Phasenverläufe aus dem Gesamtzusammenhang der Konturen in dem Maße herleitbar sind, in dem es das Gehör erfordert. Die Phase eines Sinustonimpulses muß dabei offenbar so gewählt werden, daß sich unmittelbar benachbarte Tonimpulse auf einer Konturlinie nicht gegenseitig bei der Überlagerung schwächen. In dieser differentiellen Vorschrift stecken aber noch Freiheitsgrade, die zur Erfüllung von Nebenbedingungen zwischen benachbarten Konturlinien benötigt werden. Die praktische Umsetzung einer solchen optimalen Phasenrekonstruktion bedarf zwar noch weiterer Entwicklung. Den möglichen Erfolg aber kann RKOP schon jetzt simulieren. Verglichen mit dem ursprünglichen Heinbachschen Synthesekonzept sind die Parameter von Frequenzkonturlinien dann nicht mehr mit denen von Synthesesinusschwingungen identisch: Die Wirkung des Synthesefensters und der Phasenrekonstruktion bringen im allgemeinen Amplituden- beziehungsweise Frequenzverschiebungen mit sich.

Rekonstruktion aus Konturen mit heuristischer Phase (RKHP): Dieses zweite Verfahren beinhaltet RKOP und benutzt zusätzlich eine einfache Heuristik zur Rekonstruktion des Phasenverlaufes. Dabei bleiben die Nebenbedingungen zwischen benachbarten Konturlinien unberücksichtigt. Für Frequenzkonturen entspricht dies der Phasenfortschreibung der Teiltonsynthese. Wahrnehmbare Störungen in Folge von Phaseninkohärenzen sind deshalb unvermeidlich. Sie senken nicht nur die Rekonstruktionsqualität gegenüber RKOP merklich, sondern mindern auch den Nutzeffekt von Zeitkonturen. Bei Sprache sind die Rekonstruktionsergebnisse deshalb nicht viel besser als bei einer Teiltonsynthese mit Dreieckfenster. Dort liefern Störungen durch Synthesefenster und Phaseninkohärenz eine gewisse Kompensation dafür, daß die Zeitkonturverarbeitung fehlt.

Eine Rekonstruktion des Texturanteils kann leicht in die oben beschriebenen Verfahren integriert werden. Dabei dient die Texturhüllfläche zur Steuerung einer zeitvarianten Rauschfilterung. Weil sich fehlerbehaftete Prägnanzentscheidungen bei Kontur/Textur-Analyse ähnlich wie Rekonstruktionsstörungen durch Phaseninkohärenz auswirken, ist bisher nur eine Integration in RKHP sinnvoll. Dieses Verfahren wird als RKHP mit Textur (RKHPTX) bezeichnet. Eine alternative Rekonstruktion des Texturanteils läßt sich bei geringerem Aufwand zusammen mit den bisherigen Teiltonsynthesen verwenden, eignet sich aber nicht bei zusätzlicher Zeitkonturverarbeitung. Sie beruht darauf, daß Frequenzkonturen von Weißem Rauschen von der Texturhüllfläche moduliert werden und liefert zumindest bei Sprache kaum schlechtere Ergebnisse als RKHPTX.

Der Entwicklungsprozeß dieses Kapitels stützte sich auf zwei Zwischenergebnisse, die unabhängig vom Kontext der Arbeit bemerkenswert sind. Zum einen wurde die Rücktransformation der FTT eingeführt. Dieser Formalismus zur Rückgewinnung des Signals aus einem zeit- und frequenzkontinuierlichen Kurzzeitspektrum mit frequenzabhängiger Analysebandbreite scheint bislang nicht bekannt zu sein. Zum anderen konnte daraus ein Transformationspaar T und R abgeleitet werden. Es beschreibt Analyse/Synthese-Filterbänke, welche Codierungen oder sonstige spektral/zeitliche Manipulationen auf der Grundlage solcher Spektren ermöglichen. Es kann als besondere Form einer Wavelet-Transformation formuliert werden.

Kapitel 6

Codierungen mit Konturen

Signalverarbeitung mit Hilfe von Kontur- und Kontur/Textur-Repräsentationen gilt nunmehr als eingeführt. Die verschiedenen Einstellungen der Kontur- und Kontur/Textur-Analysen mit ihren Rekonstruktionsverfahren stellen zu diesem Zweck eine Reihe von Analyse/Synthese-Kombinationen bereit. Von Codierungen, bei denen Datenraten für die Übertragung zwischen Analyse- und Syntheseseite quantifiziert sind, kann man aber noch nicht sprechen. In diesem Kapitel werden abschließend die Möglichkeiten untersucht, mit den zur Verfügung stehenden Analyse/Synthese-Kombinationen datenreduzierende Sprachcodierung zu realisieren.

Wie in Bild 6.1 veranschaulicht ist dazu das jeweilige Analyseverfahren um eine 'eigentliche' Codierung[1] zu einem Coder zu ergänzen. Dieser soll die beteiligten Repräsentationen in einen niedrigratigen Datenstrom wandeln. Entsprechend ist dem zugehörigen Rekonstruktionsverfahren die passende Decodierung vorzuschalten, mit der es den Decoder bildet. Aspekte der Kanalcodierung bleiben unberücksichtigt, es wird also von einer fehlerlosen Übertragung ausgegangen (Abschnitt 1.1). Die Übermittelung von Konturphasen diente bisher nur zur Simulation einer optimalen Phasenrekonstruktion. Angesichts des Qualitätsvorteils gegenüber der realisierten Phasenrekonstruktion erscheint ihre Codierung ebenfalls interessant.

Codierung aus einem Kontinuum heraus bedeutet Quantisierung und erfordert nach Decodierung eine Approximation zurück in das Kontinuum. Natürlich verarbeiten die zeit- und wertdiskret realisierten Verfahrenskombinationen von vornherein quantisierte Werte, so daß es genaugenommen darum geht, eine feine Quantisierung zu vergröbern beziehungsweise wiederherzustellen. Diese Vorgänge erfordern allerdings bei den verwendeten Repräsentationen besondere Beachtung. Damit befaßt sich der erste Abschnitt.

Das in den Experimenten verarbeitet Originalsignal ist gleichförmig zeit- und wertquantisiert sowie linear codiert (PCM-Signal). Es entspricht bei einer Abtastrate von 12,8 kHz und etwa 8 bis 16 bit signifikanter Codebreite einer Datenrate von 100 bis 200 kbit/s. Gegenüber der damit erreichbaren Sprachqualität müssen die meisten Analyse/Synthese-Kombinationen von vornherein mehr oder weniger deutliche Einbußen hinnehmen. Um allgemein die Effizienz von Codierungen mit Konturen abzuschätzen, behandelt der zwei-

[1] Der Begriff 'Codierung' ist hier – wie in der Literatur – doppelt besetzt. Er bezeichnet allgemein das Gesamtverfahren und auch speziell die Verarbeitung des Analyseergebnisses zu einem quantifizierbaren Datenstrom.

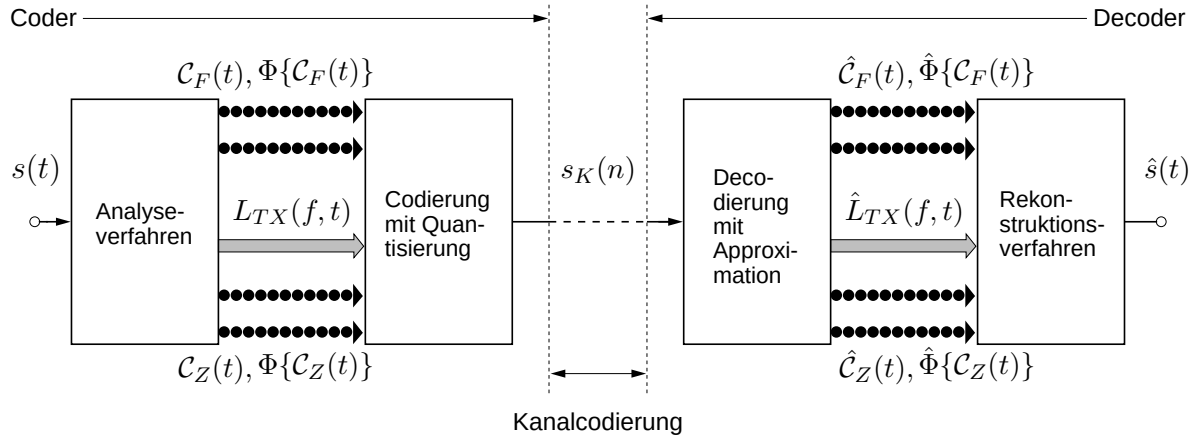

Bild 6.1: *Allgemeines Schema für Codierungen mit Konturen, Konturphasen und/oder Textur (kombiniert je nach Anwendung). Formal liegen deren Repräsentationen $\mathcal{C}$, $\Phi\{\mathcal{C}\}$ und L_{TX} vor der eigentlichen Codierung und nach Decodierung im Kontinuum vor. Nur zur Übertragung über den Kanal sind sie vollständig quantisiert im Bitstrom $s_K(n)$ 'verpackt'. Mit Hilfe der strengen Trennung von Analyse/Codierung und von Decodierung/Synthese werden die Besonderheiten der Quantisierung bzw. Approximation von Konturen erklärt.*

te Abschnitt, welche Datenraten erreichbar sind, ohne daß die jeweilige Sprachqualität merklich weiterverschlechtert wird. Da dieser Grenzfall nicht über objektive Hörversuche festgestellt wurde, können die präsentierten Ergebnisse nur eine grobe Orientierung ermöglichen. Schließlich werden im letzten Abschnitt zwei niedrigratige Codierungsverfahren vorgestellt, die zusätzliche Qualitätsverschlechterungen in Kauf nehmen. Sie werden bekannten Codierverfahren gegenübergestellt, unter anderem auch den Varianten des Heinbachschen TTZM-Verfahren.

6.1 Quantisierung, Approximation und Quantisierungsveränderung

Zunächst seien Zeit- und Frequenzkonturen sowie die Texturhüllfläche allgemein im Kontinuum des Zeit/Frequenz/Pegel-Raumes durch eine Menge von Punkten (L, f, t) definiert (vgl. Anhang A.1) und auch als solche später zu approximieren. Diese Punktmengen werden in Bild 6.1 formal durch die Kontursignale $\mathcal{C}_Z(t)$, $\mathcal{C}_F(t)$ beziehungsweise die Hüllfläche $L_{TX}(f,t)$ transportiert. Dort werden durch $\Phi\{\mathcal{C}_Z(t)\}$, $\Phi\{\mathcal{C}_F(t)\}$ auch die Konturphasen erfaßt, die gegebenenfalls mitzuübertragen sind. In diesem Fall kann der Raum noch um die Phasendimension erweitert angesehen werden, so daß sich Mengen von Punkten (L, ϕ, f, t) ergeben.

Die Quantisierung von Zeit $t \rightarrow [t]_Q$ und Frequenz $f \rightarrow [f]_Q$ führt dazu, daß die kontinuierlichen Punktmengen in diskrete Mengen von Stützstellen zerfallen. Im trivialen Falle der Texturhüllfläche entspricht dies unabhängiger Abtastung über beiden Dimensionen, die anschließende Pegelquantisierung $L \rightarrow [L]_Q$ ist unkompliziert. Bei den Konturen müssen zunächst die Linienparametrierungen abgetastet und wertdiskretisiert werden, entsprechend $f(t) \rightarrow [f([t]_Q)]_Q$ für Frequenzkonturen und $t(f) \rightarrow [t([f]_Q)]_Q$ für Zeitkonturen. Dadurch wird vermieden, das innerhalb einer Frequenzkonturlinie gleichzeitige be-

ziehungsweise innerhalb einer Zeitkonturlinie gleichfrequente Stützstellen erzeugt werden. Für eine energetisch korrekte und redundanzarme Repräsentation ist dies wesentlich. Bild 5.4 auf S. 131 veranschaulicht übrigens nichts anderes: Die von einer Linie ausgewählte Folge von Rasterorten ergibt die quantisierte Linienparametrierung für ein dort angenommenes Quantisierungsraster $\{(\omega_{S_m}, lT_S)\}$.

Den Pegel kann man anschließend als $L \to [L]_Q$ quantisieren, wobei L vom nächstnäheren Zeit/Frequenz-Ort der zu quantisierenden Linie herangezogen wird. Bei Frequenzkonturen gilt hier der gleichzeitige nächstfrequente, bei Zeitkonturen der gleichfrequente nächstzeitliche Ort. In Bild 5.4 ist das jeweils der Schnittpunkt mit der Kontrollstrecke. Grundsätzlich gilt dieses Schema auch für die Phasen, $\phi \to [\phi]_Q$.

Bei der Quantisierung der Phasen ergibt sich jedoch eine Besonderheit, wenn der nächstnähere Zeit/Frequenz-Ort zeitversetzt ist. Bei Zeitkonturen ist dies meist unvermeidlich, weil sich deren Konturpunkte durch die Quantisierung zeitlich verschieben. Vom Pegel her lassen sich Konturpunkte zwar ohne großen Fehler in beiden Dimensionen verschieben. Das FTT-Pegelspektrum bleibt innerhalb einer Zeitbreite und einer Bandbreite der Analysefensterfunktion annähernd gleich. Dagegen kann sich die FTT-Bandpaßphase, die für die Konturphase maßgeblich ist, bei hohen Frequenzen sehr schnell in zeitlicher Richtung ändern. Die notwendige Umrechnung ist durch die Phasenregel (5.31) mit $\Delta\psi = 0$ vorgegeben, als Zeitversatz ist $\Delta t = [t]_Q - t$ anzusetzen.

Nach Quantisierung werden die Stützstellenparameter in Codes umgewandelt, die einen Bitstrom $s_K(n)$ formen. Die Decodierung gewinnt daraus die quantisierten Parameter zurück. Damit sie in Bild 6.1 formal die Signale $\hat{\mathcal{C}}(t)$, $\hat{\Phi}\{\mathcal{M}(t)\}$ und $\hat{L}_{TX}(f,t)$ ausgeben kann, muß sie aus den Stützstellen Repräsentationen im Kontinuum zurückgewinnen. Dieser Vorgang heißt Approximation. Bevor bei den Konturen Pegel und Phase approximiert werden können, müssen im allgemeinen erst die quantisierten Linienparametrierungen rekonstruiert werden. Dies geschieht durch Linienassoziation mit Hilfe von Nachbarschaftskriterien. Dann werden kontinuierliche Linienverläufe approximiert, also $[f([t]_Q)]_Q \to \hat{f}(t)$ bei den Frequenzkonturen und $[t([f]_Q)]_Q \to \hat{t}(f)$ bei den Zeitkonturen. Schließlich können Pegel und Phase über der Zeit beziehungsweise über der Frequenz approximiert werden. Der Pegel der Texturhüllfläche läßt sich dagegen direkt als zweidimensionale Approximation über seinem Stützstellenfeld berechnen. Die Approximation der Phase hat die Phasenregel zu beachten.

Als einfachste Approximationsvorschrift für den Pegel – aber auch für die Frequenz beziehungsweise die Zeit bei Linienparametrierungen – kann man Treppenstufenapproximation ansetzen. Das bedeutet, daß über der oder den zu approximierenden Dimensionen immer der nächstliegende quantisierte Wert unverändert zu verwenden ist. Der Vorteil ist, daß Linienassoziation dann entbehrlich ist. Ein entsprechendes Verfahren ist auch für die Phase möglich, nur muß sie noch nach der Phasenregel um den Zeitversatz $\Delta t = \hat{t} - [t]_Q$ modifiziert werden ($\Delta\psi = 0$). Natürlich erhält man Sprünge in den Approximationen, deren Höhe zudem von den Quantisierungsstufen der beteiligten Dimensionen abhängt. Solange die Sprünge nicht zu hoch sind, kann ihre Störwirkung durch die Fehlerformung des Synthesefensters geglättet werden (Abschnitt 5.1.2). Der grundsätzliche Informationsverlust durch Quantisierung ist dadurch freilich nicht zu beheben.

Die reale Aufgabenstellung der Quantisierungsveränderung schließlich setzt sich aus Quantisierung und Approximation zusammen. Demnach ist eine bestehende Quantisierung for-

mal zu approximieren und dann neu zu quantisieren. Zur coderseitigen Quantisierungsvergröberung wird in dieser Arbeit immer Treppenstufenapproximation angenommen. Solange auch zur Decodierung Treppenstufenapproximation ausreicht, sind die notwendigen Maßnahmen bereits durch die Rasterierungsoperationen der Rekonstruktionsverfahren vollständig definiert. Bis auf weiteres kann deshalb die Approximation in der Decodierung von den Rekonstruktionsverfahren übernommen werden. [2] Für die vergleichsweise kleinen Quantisierungsstufen im nächsten Abschnitt reicht dies aus. Bei sehr grober Quantisierung werden sich ab Abschnitt 6.3 aufwendigere Approximationen als qualitätsverbessernd erweisen.

6.2 Codierung ohne zusätzliche Qualitätseinbußen

Für eine unaufwendige Codierung kann man zunächst einmal gleichförmige Quantisierung der Repräsentationen vorsehen und statistische Unabhängigkeit der quantisierten Werte annehmen. Dies geschieht im ersten Unterabschnitt und ermöglicht die Angabe konkreter Datenraten. Dabei wird diskutiert, wie Qualität und Datenraten von einzelnen Parametern abhängen. Das einfache Vorgehen vermeidet nicht unbedingt Redundanz und hinterläßt mit Sicherheit noch Irrelevanz. Deshalb werden abschließend Möglichkeiten zusätzlicher Reduktion erörtert.

6.2.1 Einfache Codierverfahren mit gleichförmiger Quantisierung

Der Quantisierungsvorgang im Coder liefert bei jeder Analyse/Synthese-Kombination pro Repräsentation und pro Zeitschritt eine Anzahl von Stützstellen, deren Werte in Codes zu fassen sind. Gleichförmigkeit der Quantisierung bezieht sich bei den Stützstellenparametern Zeit, Pegel und Phase auf konstante Quantisierungsstufen $\Delta[t]_Q$, $\Delta[L]_Q$ bzw. $\Delta[\phi]_Q$. Bei der Frequenz hingegen sind konstante Quantisierungsstufen $\Delta[z(f)]_Q$ nach Tonheitstransformation $z = z(f)$ gemeint. Zuerst werden die größtmöglichen Stufen experimentell ermittelt und diskutiert. Auf ihrer Basis lassen sich dann Stützstellencodes definieren. Zur endgültigen Berechnung der Datenraten wird zuvor noch eine Stützstellenstatistik benötigt, weil die Stützstellenanzahl pro Zeitschritt bei Konturen zeitvariant ist.

6.2.1.1 Ermittlung der kritischen Quantisierung

Für jede Analyse/Synthese-Kombination wurden die Quantisierungsstufen $\Delta[t]_Q$, $\Delta[z(f)]_Q$, $\Delta[L]_Q$ und gegebenenfalls $\Delta[\phi]_Q$ der Stützstellenparameter Zeit, Tonheit, Pegel bzw. Phase variiert, um die Verarbeitung von Sprachsignalen zu beobachten. Dies geschah unabhängig für die Stützstellen von Frequenzkonturen, Zeitkonturen und/oder Texturhüllfläche. Für jeden Parameter innerhalb einer Verfahrenskombination sollte die

[2] Bei der Teiltonsynthese TTSD wurde bisher von einem festen Zeitraster $T_S = T_A$ der eintreffenden Teiltonmuster ausgegangen, das mit der Halbwertsbreite $T_{6dB} = 1{,}25$ ms des Dreieckfensters in Gl. (1.25) übereinstimmte. Um später gröbere Zeitquantisierungen ohne Einfluß auf die Fensterlänge verarbeiten zu können, wird durch Treppenstufenapproximation das spezielle Zeitraster $T_S = T_{6dB}$ realisiert. Ist beispielsweise $\Delta[t]_Q = 5$ ms, dann wird der codierte Stützwert viermal hintereinander an TTSD ausgegeben.

größtmögliche, *kritische* Quantisierungsstufe gefunden werden, bei der gerade ein Qualitätseinbruch zu bemerken ist. Variiert wurde immer nur die Quantisierungsstufe eines Parameter, während die der übrigen kleinstmöglich und somit unauffällig eingestellt blieben. Direkt benachbarte Variationen einer Quantisierungsstufe unterschieden sich um den Faktor zwei. Als Testmaterial dienten zwei schnell gesprochene, vollständige Sätze, die jeweils eine Dauer von 2 s aufwiesen und von einem männlichen bzw. einem weiblichen Sprecher stammten. Die verarbeiteten Schalle wurden vom Autor als Versuchsperson über Kopfhörer paarweise miteinander verglichen.

Die Ergebnisse sind in Tabelle 6.1 aufgeführt. Die Einstellung der kritischen Quantisierungsstufe für alle beteiligten Parameter gleichzeitig soll kritische Quantisierung genannt werden. Sie ergab kaum eine Qualitätsverschlechterung gegenüber Einstellungen, bei denen nur die Stufe eines einzigen Parameters kritisch ist. Diese insgesamt sehr subjektiven Einschätzungen sollen wohlgemerkt nur als grober Anhaltspunkt für die wahren Verhältnisse dienen. Es wird eine Unsicherheit mindestens um den Faktor zwei vermutet.

6.2.1.2 Diskussion der Zeitquantisierungsstufe

Als kritischste Parameter erwiesen sich die Zeitquantisierungsstufen. Weil ihre Kehrwerte später als direkter Multiplikator in die Datenratenberechnung eingehen, sind sie möglichst nahe an den Punkt des Qualitätseinbruches zu justieren. Weil nur frequenzunabhängige Einstellungen betrachtet werden, zeigen sich die Qualitätsbeeinträchtigungen zuerst bei hohen Frequenzen. Wegen des FTT-Analysebandbreitenverlaufs weisen Konturen zu hohen Frequenzen hin eine zunehmende zeitliche Variabilität auf, die dann irgendwann nicht mehr korrekt abgetastet wird.

Für die Frequenzkonturen bedeutet eine zu seltene Abtastung, daß in den Höhen tonale Artefakte auftauchen. Dies äußert sich als Kammfiltereffekt bei kleineren, als 'Zwitschern' bei größeren Analysebandbreiten, weil die Linien am schnellen Fluktuieren gehindert werden (vgl. Abschnitte 2.5 und 2.6.2). Die Verfahrenskombinationen mit Textur transportieren wenig hochfrequente Anteile in den prägnanten Frequenzkonturen, so daß die zeitliche Quantisierung weniger kritisch ist. Bei den Zeitkonturlinien ist klar, daß ihr minimal möglicher Zeitabstand noch von der Quantisierung erfaßt werden muß. Er liegt in der Größenordnung der minimalen Zeitbreite des Analysefensters. Besonders bei synthetischen Impulsfolgen zeigen sich allerdings schon bei feinerer Quantisierungsstufe Schwebungs- und Modulationseffekte, weil die zeitliche Positionierung der Linien nicht mehr stimmt (vgl. T_A in Abschnitt 3.4.3.4).

Speziell bei ZFKI ist bei zunehmendem $\Delta[t_Z]_Q$ zunächst nur ein Höhenabfall zu beobachten, der mit der Zeitbreite des Synthesefensters und den separat codierten Phasen zusammenhängt. Entlang einer Zeitkonturlinie, die beispielsweise von einem isolierten Impuls stammt, bestimmt nämlich die Phasendrehung die exakte Positionierung des rekonstruierten Impulses innerhalb des Synthesefensters (vgl. Drehfaktor Abschnitt 5.1.1.1). Bei Quantisierung wird meist die Zeitlage der Linie verändert, gleichzeitig aber ihre Phase so korrigiert, daß der rekonstruierte Impuls wieder an den ursprünglichen Zeitort rückt. Zunächst schadet eine gröbere Zeitquantisierung hier viel weniger als bei den anderen Verfahren. Dort erzeugt die Phasenrekonstruktion einen Nullphasenverlauf, so daß die rekonstruierte Impulslage der quantisierten Linienlage entspricht. Wird $\Delta[t_Z]_Q$ jedoch bei ZFKI noch gröber gewählt, dann kann die zu rekonstruierende Zeitlage aus dem Synthe-

Tabelle 6.1: *Kritische Quantisierung und Codierung der Stützstellenparameter Tonheit, Phase, Pegel und Zeit von Frequenzkonturlinien (oben), Zeitkonturlinien (Mitte) und Texturhüllflächen (unten) für Sprache, dargestellt für verschiedene Analyse/Synthese-Kombinationen (Spalten). 'Kritisch' bedeutet, daß der Einfluß der Quantisierung gerade eben wahrgenommen wurde (siehe Text). Speziell die FK-Tonheitsquantisierung (außer bei ZFKI) und generell die Pegelquantisierung sind noch nicht kritisch.* $\Delta[x]_Q$ *bezeichnet die Quantisierungsstufe eines Parameters* x *und* $x_{min} \ldots x_{max}$ *seinen Quantisierungsbereich. Vorausgesetzt wurde ein Dynamikbereich von* 80 dB*, eine Grenzfrequenz von* 5,5 kHz *(*≈ 19 Bark*) sowie gleichförmige Quantisierung.* $^+$ *siehe Fußnote S. 159.* * *Auf diesen Wert muß dann der Parameter* Δt_Φ *in Tabelle 5.1 angehoben werden.*

Stützstellenparameter	**Quantisierung und Codierung**					
	ZFKI RKOP	ZFKII RKHP	KTX RKHPTX	KTXOZ RKHPTX	M-TTZM TTSD	HB-TTZM TTSD
FK-Tonheit $\Delta[z(f_F)]_Q$ /Bark	0,35	0,05	0,05	0,05	0,05	0,05
z(20Hz) … z(5,5kHz) Code	6bit	9bit	9bit	9bit	9bit	9bit
FK-Phase $\Delta[\phi_F]_Q$ /rad	$\pi/8$	-	-	-	-	-
$0 \ldots 2\pi$ Code	4bit	-	-	-	-	-
FK-Pegel $\Delta[L_F]_Q$ /dB	0,5	0,5	0,5	0,5	0,5	0,5
−80dB … 0dB Code	8bit	8bit	8bit	8bit	8bit	8 bit
$b_F = \Sigma$ Codebreiten	18bit	17bit	17bit	17bit	17bit	17bit
FK-Zeit $\Delta[t_F]_Q$ /ms	1,25	2,5	5	5	$2,5)^+$	$5)^+$
ZK-Tonheit $\Delta[z(f_Z)]_Q$ /Bark	0,35	0,35	0,35	-	-	-
z(20Hz) … z(5,5kHz) Code	6bit	6bit	6bit	-	-	-
ZK-Phase $\Delta[\phi_Z]_Q$ /rad	$\pi/8$	-	-	-	-	-
$0 \ldots 2\pi$ Code	4bit	-	-	-	-	-
ZK-Pegel $\Delta[L_Z]_Q$ /dB	0,5	0,5	0,5	-	-	-
−80dB … 0dB Code	8bit	8bit	8bit	-	-	-
$b_Z = \Sigma$ Codebreiten	18bit	14bit	14bit	-	-	-
ZK-Zeit $\Delta[t_Z]_Q$ /ms	1,25	$2,5)^*$	$5)^*$	-	-	-
TX-Tonh. $\Delta[z(f_{TX})]_Q$ /Bark	-	-	0,7	0,7	-	-
z(20Hz) … z(5,5kHz) Anzahl	-	-	28	28	-	-
TX-Pegel $\Delta[L_{TX}]_Q$ /dB	-	-	0,5	0,5	-	-
−80dB … 0dB Code	-	-	8bit	8bit	-	-
b_{TX} = Anz.×Codeb.	-	-	224bit	224bit	-	-
TX-Zeit $\Delta[t_{TX}]_Q$ /ms	-	-	5	2,5	-	-

sefenster auswandern, welches sich um die quantisierten Zeitpunkte zentriert. Dadurch wird der rekonstruierte Impuls gedämpft, was sich eben bei hohen Frequenzanteilen aufgrund des dort kurzen Synthesefensters zuerst bemerkbar macht. Die 3dB-Zeitbreite des Synthesefensters (B4, $B^S_{3dB} = 0{,}7$ Bark) beträgt bei 4 kHz beispielsweise nur noch rund 1

ms, so daß die eingestellte kritische Quantisierung bereits recht grob erscheint.

Die Texturhüllfläche bei KTXOZ erfordert eine feinere zeitliche Quantisierung als bei KTX, weil sie die impulshaften Anteile mitrepräsentieren muß. Ansonsten würde Sprache 'stumpf' klingen. Dies steht in Einklang mit der Tatsache, daß die zeitliche Glättung der Textur in Abschnitt 4.6 schwächer eingestellt werden mußte. Es besteht eine gewisse Notwendigkeit, gleiche zeitliche Quantisierung innerhalb einer Verfahrenskombination zu wählen, damit die einzelnen Repräsentationen gut zu einer klanglichen Einheit fusionieren. Bemerkenswert ist deshalb, daß die prägnanten Zeitkonturen bei KTX für sich allein genommen schon eine recht grobe Zeitquantisierung vertragen, wenn man dies mit ZFKII vergleicht. Dort werden zwar dieselben Zeitkonturen analysiert, nichtprägnante Zeitkonturen werden aber nicht gesondert berücksichtigt. Diese scheinen bei gröberer Zeitquantisierung besonders leicht wahrnehmbare Artefakte hervorzurufen.

6.2.1.3 Diskussion der übrigen Quantisierungsstufen

Mit Ausnahme von $\Delta\,[z(f_{TX})]_Q$ bedeutet eine Halbierung der Quantisierungsstufen von Tonheit, Pegel und Phase später nur jeweils ein zusätzliches Bit pro Stützstellencode. Dies hat wenig Auswirkung auf die Gesamtdatenraten, so daß hier eher großzügige Einstellungen gewählt wurden. So erhielten die Quantisierungsstufen für den Pegel generell den unkritischen Wert 0,5 dB, der sich schon bei Heinbach bewährt hat. Ebenso wurde für die Tonheitsquantisierung der Frequenzkonturen, außer bei ZFKI, der Wert 0,05 Bark aus den Analyseverfahren übernommen.

Verglichen damit ist für Zeitkonturen und Texturhüllfläche eine wesentlich gröbere Frequenzauflösung zulässig, hier steckt die Information über die genaue spektrale Form in mehreren gleichzeitigen spektralen Abtastwerten. Dadurch reicht eine Abtastung etwa im Abstand der Analysebandbreite aus. Weil die Texturhüllfläche analyseseitig in spektraler Richtung geglättet wurde (Abschnitt 4.3), fällt die ermittelte Quantisierungsstufe $\Delta\,[z(f_{TX})]_Q$ noch größer aus. Wie die Zeitquantisierungsstufen verkörpert sie einen kritischen Parameter, weil sie später direkt als Faktor in die Anzahl der zu codierenden Texturstützstellen eingeht.

Die gefundene Quantisierungsstufe für die Phasen paßt sehr gut mit folgender Überlegung zusammen (vgl. [Kra88, S. 103ff]): Man stelle sich einen Sinuston als komplexen Drehzeiger vor, der auf die reelle Achse projiziert wird. Wäre die Amplitude als Zeigerlänge zeitveränderlich, dann dürfte sie nicht um mehr als etwa 1 dB schwanken, wenn keine Modulation hörbar werden soll (Schwellenfunktionsschema nach Zwicker [Zwi82]). Weicht stattdessen die Momentanphase an zwei aufeinanderfolgenden Zeitpunkten als Folge der Phasenquantisierung vom richtigen Wert ab, dann darf auch die schlimmstmögliche Abweichung in der Projektion 1 dB nicht überschreiten. Bei $\Delta[\phi]_Q = \pi/8$ ergibt sich eine Schwankung $\Delta L = 0{,}7$ dB mit Hilfe der Formel

$$\Delta L = 20\mathrm{lg}\Big(\cos\left(2\pi \cdot \Delta[\phi]_Q\right)\Big)\ \mathrm{dB}. \tag{6.1}$$

Bemerkenswert ist das Ergebnis für die Tonheitsquantisierung der Frequenzkonturen bei ZFKI, weil eine wesentlich größere Stufe als bei den anderen Verfahren möglich ist. Die Information über die genaue Frequenzlage steckt nämlich – in analoger Weise wie die genaue Zeitlageninformation bei den Zeitkonturen – nochmals in den separat codierten Phasen.

Beispielsweise gibt die zeitkontinuierliche Konturphase bei einem stationären Sinuston seine Momentanphase wieder und könnte die Frequenzinformation der Konturlinie ersetzen. Wegen der (Unter-)Abtastung durch zeitliche Quantisierung wird sie allerdings mehrdeutig, so daß eine grobe Frequenzlageinformation erforderlich bleibt. Bei Rekonstruktion wird ein Synthesebandpaß genau in dieser Lage ausgewählt, mit dem die codierte Phaseninformation eindeutig in die ursprüngliche Frequenz umgesetzt werden kann. Die Bandbreite des Synthesefilters – in Analogie zu dessen Zeitbreite bei den Zeitkonturen – bestimmt, wie genau die Frequenzlageinformation zu sein hat. Ein zu rekonstruierender Ton außerhalb der Bandmitte des angewählten Synthesefilters wird zunehmend gedämpft. Die Einstellung der Quantisierungsstufe auf die halbe Synthesebandbreite bedeutet schlimmstenfalls eine Dämpfung von etwa 1 dB (Bild 3.6b). [3]

Das Vorhandensein von Phaseninformation ersetzt bei den Frequenzkonturen also Frequenzinformation. Prinzipiell gilt analog bei den Zeitkonturen, daß Phaseninformation Zeitinformation ersetzt. Dieser Effekt ist in den Einstellungen wegen der geringeren Bedeutung des Zeitkonturbeitrages bei Sprache nicht so leicht verifizierbar. In beiden Fällen ist die Ersetzbarkeit durch die Zeit- beziehungsweise die Bandbreite des Synthesefensters begrenzt. Die Tatsache, daß überhaupt eine Phaseninformation vorhanden ist, eliminiert das schwierige Problem einer idealen Phasenrekonstruktion. Zwar fällt nacher die Datenrate bei ZFKI höher als bei ZFKII aus, dafür ist das erreichbare Qualitätsniveau auch spürbar besser.

6.2.1.4 Festlegung der Stützstellencodes

Da die Quantisierungsstufe $\Delta[x]_Q$ eines jeden Stützstellenparameters x nunmehr festliegt, kann seine Abbildung in Codes betrachtet werden. Dazu wird angenommen, daß alle seine quantisierten Werte $[x]_Q$ im gewünschten Quantisierungsbereich, gerechnet von x_{min} bis x_{max} einschließlich, gleichwahrscheinlich sind und daß sie sich unabhängig von anderen Parametern derselben oder benachbarter Stützstellen einstellen. Diese Annahme stellt die ungünstigste aller möglichen Situationen dar und führt, wie noch zu untersuchen sein wird, auf ein gewisses Maß an Redundanz. Andererseits sind dadurch Codewörter von konstanter Breite optimal, deren genaue Zuordnung zu den quantisierten Werten hier weiter keine Rolle spielt. Die benötigte Codebreite b läßt sich wie folgt angeben:

$$b = \text{int}\left(1 \;+\; \text{ld}\left|\frac{x_{max} - x_{min}}{\Delta[x]_Q}\right|\right) \text{ bit.} \tag{6.2}$$

Die int-Funktion liefert hierbei den ganzzahligen Anteil ihres Arguments, bei der ld-Funktion handelt es sich um den Logarithmus zur Basis zwei. Für den Quantisierungsbereich der Tonheit werden Frequenzen von 20 Hz bis 5,5 kHz, für die Pegeldynamik 80dB angenommen. Bei der Phase ist ein Vollkreis zu quantisieren, wobei der Maximalwert in Gl. (6.2) genau um eine Quantisierungsstufe kleiner als 2π ($\hat{=}0$!) anzusetzen ist. Die errechneten Codebreiten können, wie in Tabelle 6.1 geschehen, für die Konturen nur teilweise summiert werden. Es ergeben sich Codebreiten pro Konturstützstelle, b_F und b_Z, bei

[3] Die Codierung mit ZFKI/RKOP weist eine starke Ähnlichkeit zum Phasenvocoderprinzip [Fla66] auf. Dort wird eine Phasen- und eine Betragsinformation für jedes Syntheseband übertragen. Hier dagegen geschieht das nur in den Bändern, wo Konturen verlaufen. Der Einspareffekt wird dadurch relativiert, daß zusätzlich zu übertragen ist, welche Bänder ausgewählt sind, und ob es sich um Frequenz- oder Zeitkonturinformation handelt.

denen noch die Information über die Zeitlage der Stützstelle fehlt. Im Hinblick auf eine Gesamtcodebreite über alle Stützstellen pro Zeitschritt gibt es noch eine weitere, wesentliche Unklarheit: Die Anzahl der gleichzeitig vorhanden Konturstützstellen ist zeitvariant.

Dagegen kann die Gesamtcodebreite b_{TX} der Texturhüllfläche pro Zeitschritt unmittelbar angegeben werden, weil die Anzahl der Stützstellen konstant bleibt. Dabei braucht die Tonheit nicht explizit codiert zu werden, da die Stützstellen über ihre Reihenfolge immer den gleichen Tonheitswerten zugeordnet sind. Ihre Anzahl stimmt mit der Anzahl der Quantisierungsstufen für die Tonheit überein und errechnet sich aus Gl. (6.2), indem die Logarithmusfunktion und der Zusatz 'bit' weggelassen werden. Die Datenrate I_{TX} für die Texturrepräsentation ergibt sich, wie in der dritten Abteilung von Tabelle 6.3 zu sehen ist, aus b_{TX} und der Zeitquantisierungsstufe $\Delta[t_{TX}]_Q$.

6.2.1.5 Stützstellenstatistik

Um trotz zeitvarianter Stützstellenanzahl auch bei den Konturen zu konkreten Datenraten zu gelangen, muß auf die Statistik der Stützstellendichten von Sprache zurückgegriffen werden. Dazu werden einige Meßergebnisse auf Basis des nunmehr bekannten, 2 s langen Sprachbeispiels 'Kalk...' vorgestellt. Sie erwiesen sich als repräsentativ für jede Art von fließender Sprache. Sprecher, männlich oder weiblich, und Sprechgeschwindigkeit scheinen keinen deutlichen Einfluß auf die Statistik zu haben. Lediglich längere Pausen können die Dichten senken. [4]

In Tabelle 6.2 werden die mittleren Dichten $\overline{d_F}$, $\overline{d_Z}$, $\overline{d_{MZ}}$ von drei Sorten von Stützstellen unterschieden, die jeweils die Frequenzkonturen, die Zeitkonturen und die Frequenzkontur-maskierten Zeitkonturen repräsentieren. Letztere sind gegenüber den normalen Zeitkonturen um die Anteile reduziert, die decoderseitig im Rahmen der Operation MSK in den Rekonstruktionsverfahren RKOP, RKHP und RKHPTX sowieso maskiert werden würden. Die Vorwegnahme von MSK im Coder reduziert demnach den Datenfluß. Zwar wäre es wahrscheinlich für eine verbesserte Phasenrekonstruktion in RKHP besser, auf unmaskierten Konturen aufbauen zu können, um möglichst viele Hinweise auf Phasenrelationen zu bewahren (Abschnitt 5.1.5.5). Für die realisierte, suboptimal arbeitende Phasenheuristik ergab sich aber kein signifikanter Qualitätsunterschied.

Die Statistik schlüsselt den genutzten Tonheitsbereich von 0,2 bis 19,2 Bark (20 Hz bis 5,5 kHz) nochmals in vier Teilbereiche auf. Demnach gibt es bei den Frequenzkonturen nur innerhalb der beiden Kontur/Textur-Verfahren signifikante Unterschiede (Spalten KTX und KTXOZ). Dort zieht die Auswahl der prägnanten Konturen eine Tiefenlastigkeit der Stützstellendichte nach sich. Beide Kombinationen weisen hier identische Dichten auf, da die Verarbeitungszweige für Frequenzkonturen gleich sind. Die Verarbeitungszweige sind von der Konturanalyse her außerdem identisch mit denjenigen der beiden benachbarten Verfahren (Spalten ZFKII und M-TTZM). Dort wirkt allerdings kein Prägnanzkriterium, so daß höhere Dichten zu verzeichnen sind. Bei den ersten beiden Verfahren ohne Textur verhalten sich die Dichten im übrigen umgekehrt proportional zur Analysebandbreite (0,5 bzw. 0,3 Bark). Beim Heinbach-Verfahren (0,1 Bark) stimmt diese Regel wegen des wesentlich veränderten Satzes von Analyseparametern nur annähernd.

[4] Bei stationärem Weißem Rauschen hingegen stiegen die Werte für die Frequenzkonturen um bis zu 10% an, gleichzeitig nahmen sie für beide Typen (s.u.) von Zeitkonturen im Bereich 0,2 ... 19,1 Bark um etwa 10% ab.

Tabelle 6.2: *Mittlere Stützstellendichten von Frequenzkonturen (oben), Zeitkonturen (Mitte) sowie Frequenzkontur-maskierten Zeitkonturen (unten), dargestellt für verschiedene Verfahrenskombinationen (Spalten) bei fließender Sprache. Die Werte sind Langzeitmittel für die links angegebenen Tonheitsbereiche. Ihre spezielle Normierung eliminiert die Abhängigkeit von der tatsächlich gewählten Zeit- und Tonheitsquantisierung: Für Frequenzkonturen entsprechen sie der mittleren Anzahl von Stützstellen pro Zeitquantisierungsstufe über einem Bereich von* 20 Bark*; für Zeitkonturen ist der Wert die auf eine Frequenzquantisierungsstufe und eine Dauer von* 1 s *bezogene mittlere Anzahl. Die Werte in den Zeilen* $0,2\ldots19,1$ Bark *kann man überschlägig auch für einen gleichmäßigen Stützstellenstrom ansetzen, wenn man von einer zeitlichen Pufferung ausgeht: Nicht dargestellte Spitzenwerte nach* 200*ms-Pufferung überschreiten die Langzeitmittel um max.* 10% *bei den Frequenzkonturen und um max.* 20% *bei den übrigen Konturen.*

Parameter		**Sprachbeispiel 'Kalk…'**					
		ZFKI RKOP	ZFKII RKHP	KTX RKHPTX	KTXOZ RKHPTX	M-TTZM TTSD	HB-TTZM TTSD
mittl. FK-Stützstellendichte $\overline{d_F}$		normierte Werte $\overline{d_F} \cdot \Delta\,[t_F]_Q \cdot 20$ Bark					
	0,2…19,1	16,0	26,3	12,9	12,9	26,3	33,1
Mittelungsbereiche in Bark	0,2…5	15,5	24,3	17,5	17,5	24,3	32,6
	5…10	16,4	26,5	17,0	17,0	26,5	38,0
	10…15	17,0	28,8	11,6	11,6	28,8	36,3
	15…19,1	14,9	25,5	4,1	4,1	25,5	23,7
mittl. ZK-Stützstellendichte $\overline{d_Z}$		normierte Werte $\overline{d_Z} \cdot \Delta\,[z(f_Z)]_Q \cdot 1$ s					
	0,2…19,1	145,4	86,8	27,0	-	-	-
Mittelungsbereiche in Bark	0,2…5	55,0	35,5	3,4	-	-	-
	5…10	75,3	42,6	9,3	-	-	-
	10…15	161,0	89,7	32,3	-	-	-
	15…19,1	318,7	198,2	70,0	-	-	-
mittl. Stützstellendichte $\overline{d_{MZ}}$ von FK-maskierten ZK		normierte Werte $\overline{d_{MZ}} \cdot \Delta\,[z(f_Z)]_Q \cdot 1$ s					
	0,2…19,1	65,9	42,0	24,0	-	-	-
Mittelungsbereiche in Bark	0,2…5	26,4	19,2	2,4	-	-	-
	5…10	31,1	19,0	6,7	-	-	-
	10…15	73,3	39,8	27,2	-	-	-
	15…19,1	146,5	100,2	66,6	-	-	-

Bei den Zeitkonturen nimmt die Dichte $\overline{d_Z}$ zu hohen Frequenzen hin deutlich zu. Das ist bei den ersten beiden Verfahren allein eine Folge des frequenzabhängigen Analysefensters. Dort verhalten sich die Werte überdies ungefähr proportional zur Analysebandbreite, und zwar auch innerhalb einer Spalte. Beim Kontur/Textur-Verfahren (KTX) aber bewirkt das Prägnanzkriterium noch eine zusätzliche Höhenlastigkeit.

Die Dichten $\overline{d_{MZ}}$ der FK-maskierten Zeitkonturen gehen bei den ersten beiden Kombinationen grob um die Hälfte gegenüber den korrespondierenden Werten von $\overline{d_Z}$ zurück. Beim Kontur/Textur-Verfahren gibt es dagegen kaum einen Rückgang. Dies steht in Einklang mit der Tatsache, daß die prägnanten Zeitkonturen noch viel seltener solche Signalanteile

repräsentieren, die bereits in den Frequenzkonturen ausreichend erfaßt wurden (Prägnanz bedeutet Relevanz).

6.2.1.6 Zeitlagencodierung und Berechnung der Datenraten

Mit Hilfe der Stützstellenstatistik lassen sich schließlich die mittleren Datenraten berechnen. Der Rechengang ist in Tabelle 6.3 dargestellt. Aus den mittleren Stützstellendichten $\overline{d_F}$ und $\overline{d_{MZ}}$ errechnen sich die mittleren Stützstellenanzahlen $\overline{N_F}$ bzw. $\overline{N_{MZ}}$ pro Zeitschritt. Die Zeitlage der Konturstützstellen wird dadurch festgelegt, daß für jeden Zeitschritt eines Konturtyps eine Stützstellenanzahl codiert wird. Die vorhandenen Stützstellen mit ihren bereits definierten Codes der Breite b_F bzw. b_Z werden daran angehängt. Zwar sind die Codebreiten b_{N_F}, b_{N_Z} für die Stützstellenanzahlen übertrieben groß ausgelegt: Sie können den theoretisch ungünstigsten Fall bewältigen, bei dem an allen möglichen Tonheitsquantisierungsstufen gleichzeitig Stützstellen vorliegen. Dennoch ist ihr Anteil an der Gesamtdatenrate gering. Diese ergibt sich aus der Summe der Datenraten für die verwendeten Repräsentationsformen, in die die Zeitquantisierungsstufen $\Delta[t_F]_Q$, $\Delta[t_Z]_Q$ und $\Delta[t_{TX}]_Q$ eingehen. Summierung ist zulässig, weil die Datenströme reversibel ineinander geflochten werden können.

Zwar stellen die in Tabelle 6.3 angegebenen Datenraten $\overline{I}$ nur einen zeitlichen Mittelwert dar. Tatsächliche schwankt die zeitabhängige Datenrate aufgrund der Zeitvarianz der Konturstützstellendichten. Um eine bestimmte Maximaldatenrate zu garantieren, sind zwei zusätzliche Maßnahmen anzunehmen: Erstens wird der Datenstrom über einen bestimmten Zeitraum gepuffert. Der Puffer bildet ein Bitreservoir (vgl. [Bra94]), aus dem nur die Maximaldatenrate 'abfließt'. Wie bei Tabelle 6.2 erwähnt, übersteigen die Spitzenwerte der Stützstellendichten ihre Langzeitmittel um maximal 20%, wenn man eine 200ms-Pufferung annimmt. Zur Sicherheit ist zweitens ein 'Überlaufventil' vorzusehen, das bei Ansprechen wenig qualitätsmindernd in Erscheinung tritt. Beispielsweise könnte man Stützstellen mit niedrigem Pegel entfernen. Unabhängig von diesen Maßnahmen reichen die angegebenen mittleren Datenraten zur groben Orientierung aus. Ihr Niveau ist wegen der subjektiv ermittelten kritischen Quantisierung sowieso mit einem höheren Unsicherheitsfaktor behaftet.

6.2.2 Möglichkeiten weiterer Redundanz- und Irrelevanzreduktion

Redundanzarme Codierung erfordert nach Abschnitt 1.2, daß die zu übertragenden Daten im statistischen Sinne weitgehend voneinander unabhängig ('maximal dekorreliert') sind und daß die Codewahl an ihre Statistik angepaßt ist (Entropiecodierung). Das oben verwendete, einfache Codierschema verzichtet auf Entropiecodierung und vernachlässigt darüber hinaus mindestens drei Arten der Korrelation:

- Frequenzunabhängige Zeitquantisierung bedeutet, daß sich die Quantisierungsstufe an der zeitlichen Variabilität der höchsten Frequenzanteile orientieren muß. Folglich wird zu tieferen Frequenzen hin zunehmend öfter als nötig abgetastet, denn in dieser Richtung sinkt die Variabilität wegen der abnehmenden FTT-Analysebandbreite.

Tabelle 6.3: *Berechnung der mittleren Datenraten $\overline{I}$, ausgehend von Codes der kritischen Quantisierung nach Tabelle 6.1 und mittleren Stützstellendichten nach Tabelle 6.2. Für Zeitkonturen sind die FK-maskierten Dichten zugrunde gelegt. Angenommen wird, daß die Stützstellenanzahlen im aktuellen Zeitschritt durch Codes übermittelt werden, deren Breite die theoretische größtmögliche Anzahl beherbergen können. b_{N_F} und b_{N_Z} entsprechen daher den Codebreiten von FK-Tonheit bzw. ZK-Tonheit in Tabelle 6.1.*

Berechnung Datenrate	ZFKI RKOP	ZFKII RKHP	KTX RKHPTX	KTXOZ RKHPTX	M-TTZM TTSD	HB-TTZM TTSD
Frequenzkonturen						
mittl. FK-Stützstellenanzahl pro Zeitschritt $\Delta\,[t_F]_Q$	$\overline{N_F} = \overline{d_F} \cdot \Delta\,[t_F]_Q \cdot \big(z(5,5\text{kHz}) - z(20\text{Hz})\big)$					
	15,1	24,8	12,2	12,2	24,8	31,3
Codebreite Stützstellenanzahl	b_{N_F}					
	6bit	9bit	9bit	9bit	9bit	9bit
mittl. FK-Datenrate	$\overline{I_F} = \big(\overline{N_F} \cdot b_F + b_{N_F}\big) / \Delta\,[t_F]_Q$					
kbit/s	222	172	43	43	172	108
Zeitkonturen						
mittl. Stützstellenanzahl FK-maskierter ZK pro $\Delta\,[t_Z]_Q$	$\overline{N_{MZ}} = \overline{d_{MZ}} \cdot \Delta\,[t_Z]_Q \cdot \big(z(5,5\text{kHz}) - z(20\text{Hz})\big)$					
	4,45	5,67	6,48	-	-	-
Codebreite Stützstellenanzahl	b_{N_Z}					
	6bit	6bit	6bit	-	-	-
mittl. ZK-Datenrate	$\overline{I_{MZ}} = \big(\overline{N_{MZ}} \cdot b_Z + b_{N_Z}\big) / \Delta\,[t_Z]_Q$					
kbit/s	69	34	19	-	-	-
Texturhüllfläche						
feste TX-Datenrate	$I_{TX} = b_{TX} / \Delta\,[t_{TX}]_Q$					
kbit/s	-	-	45	90	-	-
mittlere Datenrate	$\overline{I} = \overline{I_F} + \overline{I_{MZ}} + I_{TX}$					
kbit/s	291	206	107	133	172	108

- Ein Teil der Konturlinien, mindestens aber die prägnanten Konturen weisen größere Längen auf. Dabei überspannen sie bei den Frequenzkonturen ein Vielfaches der zeitlichen Quantisierungsstufe, bei den Zeitkonturen ein Vielfaches der spektralen Quantisierungsstufe. Folglich beschränkt der Linienzusammenhalt die Variabilität der Stützstellenparameter ‘quer’ zur Linienausrichtung.

- Da Konturen und Textur aus einem Wahrnehmungsmodell heraus entwickelt wurden, können sie beliebige Audiosignale repräsentieren. Sprachcodierungen beschränken sich jedoch üblicherweise auf die Freiheitsgrade der Sprachproduktion und eliminieren gerade dadurch viel Redundanz.

Der erste Punkt läßt sich relativ gut abschätzen, weil die Zeitquantisierungsstufen bei den Raten der Teilrepräsentationen jeweils im Nenner auftauchen (Tabelle 6.3). Dazu nimmt man an, daß die bisherige Zeitquantisierung an der oberen Übertragungsgrenze $f_o = 5{,}5$ kHz, bereits zu knapp eingestellt ist, ein Bark darunter aber, also unterhalb von $f_x = 4{,}6$ kHz, nicht mehr kritisch ist. Der Grund für diese Annahme ist, daß sich Artefakte erst bemerkbar machen dürften, wenn die Unterabtastung über eine gewisse spektrale Breite vorliegt, mutmaßlich über 1 Bark. Die Zeitquantisierung wird nun in der Weise umgekehrt proportional zur Frequenzgruppenbreite $\Delta f_G(f)$ nach Gl. (1.2) gewählt, daß sie bei f_x mit der bisherigen, konstanten Zeitquantisierung übereinstimmt. Innerhalb einer Teilrepräsentation würde sich bei frequenzunabhängiger Stützstellendichte idealerweise folgender Reduktionsfaktor ergeben:

$$\frac{\overline{I}'}{\overline{I}} = \frac{1}{f_o - f_u} \cdot \int_{f_u}^{f_o} \frac{\Delta f_G(f)}{\Delta f_G(f_x)} df \tag{6.3}$$

Der Integrand gibt hier die Veränderung des Datenratenbeitrages im Band $f \pm \frac{df}{2}$ an, die über den Übertragungsbereich zu mitteln ist. Unberücksichtigt bleibt, daß sich die Zeitlagencodierung der Konturen verkompliziert. Der Reduktionsfaktor erreicht mit den oben genannten Parametern und mit $f_u = 20$ Hz einen Wert von rund einem Drittel. Auf die Gesamtdatenrate greift dieser Faktor aber nur bei den Kombinationen ohne Zeitkontur-Repräsentation durch. Bei den Zeitkonturen wird sich dagegen kaum sparen lassen, weil sich ihre Stützstellen zu höheren Frequenzen hin verdichten. In drastischer Weise gilt das für die prägnanten Zeitkonturen in KTX. Allerdings fallen die Zeitkonturanteile an den Gesamtdatenraten vergleichsweise gering aus. Bei den prägnanten Frequenzkonturen beider Kontur/Textur-Kombinationen läßt sich sogar eine bessere Einsparung erzielen, denn ihre Stützstellendichte nimmt zu höheren Frequenzen ab.

Der zweite Punkt verursacht möglicherweise nicht allzuviel Redundanz. Zwar könnte man durch eine adaptive Differenzcodierung [Jay84] der Querparameter die Stützstellencodebreite reduzieren, andererseits muß dann die Assoziation aufeinanderfolgender Stützstellencodes wieder extra codiert werden. Es wird insgesamt ein Sparpotential von bestenfalls 50% vermutet, das überdies nicht für einen Texturanteil gilt.

Der dritte Punkt beschreibt eine sehr komplizierte, weil weit übergreifende Korrelation von Stützstellen. Sie war bereits im Ansatz der Arbeit vorhersehbar (Abschnitt 1.2). Die daraus resultierende Redundanz bedeutet allerdings, daß der Codierung eine gewisses Maß an Robustheit innewohnt (vgl. Abschnitt 6.3.5). Hintergrundgeräusche und sprachfremde Signale werden eben auch wahrnehmungsnah übertragen. Redundanz einsparen könnte hier eventuell der Ansatz der Vektorquantisierung [Gra84]. Allerdings ist damit grundsätzlich ein Relevanzverlust verbunden.

Die Entfernung perzeptiver Irrelevanz erscheint in gewissen Grenzen noch möglich, indem man durch Modelle der psychoakustischen Verdeckung ‘unhörbare’ Stützstellen aus den Repräsentationen eliminieren läßt. Grundsätzlich modellieren FTT-Spektralanalyse und Konturierung auch Verdeckungseffekte (z.B. Abschnitt 2.4). Allerdings geschieht dies mit sehr konservativer Tendenz, so daß eher zu viele als zu wenige Konturen bestimmt werden. Insbesondere die pegelabhängige Auffächerung der oberen Mithörschwellenflanke und die vergleichsweise lange Nachverdeckung, die im Gehör von höheren Verarbeitungsprozessen dominiert wird [Lan91], blieben bisher unberücksichtigt. Speziell für die Textur müßte ein

anderes Codierungsschema erarbeitet werden, damit verdeckte Stützstellen wirklich im Datenstrom eingespart werden können.

Bei Sprachcodierung muß man nicht unbedingt die gesamte perzeptive Relevanz bewahren. Damit steigt allerdings das Problem der Rekonstruktion eines akzeptabel klingenden Signals, wie der nächste Abschnitt verdeutlichen wird. Eine verbreitete Maßnahme zur Entfernung perzeptiver Relevanz ohne wesentliche Beeinträchtigung der sprachlichen Information stellt die Herabsetzung der oberen Grenzfrequenz auf $f_o = 3{,}5\,\text{kHz}$ dar. Frequenzen darüber werden von Telefonnetzwerk-fähigen Sprachcodierverfahren sowieso nicht verarbeitet. Zusammen mit der frequenzabhängigen Zeitquantisierung verbessert sich der Reduktionsfaktor aus Gl. (6.3) dadurch von einem Drittel auf ein Viertel. Bei Erhöhung auf $f_o = 20\,\text{kHz}$ erhält man dagegen eine Verschlechterung zurück auf eins. Verfahren mit Zeitkonturen müssen dann aber mit noch größeren Werten rechnen, weil die Stützstellendichte bei höheren Frequenzen stark zunimmt. An dieser Stelle spielt für die Zukunft eine geeignete Modellierung der Nervenfasergrenzfrequenz eine wichtige Rolle, welche die Dichte bereits ab 3 kHz begrenzen dürfte (vgl. Abschnitt 3.3.6.2).

6.2.3 Zusammenfassung und Schlußfolgerung

Um eine Orientierung für die Datenraten zu erhalten, die mit Konturen und Kontur/Textur-Repräsentationen erzielbar sind, wurden einfache Codierungen untersucht. Sie basieren auf den bereits eingeführten Kombinationen von Analyse- und Rekonstruktionsverfahren, zwischen denen die jeweilige Repräsentation codiert übertragen wird. Als Nebenbedingung sollte der Codierungsvorgang die Verarbeitungsqualität von Sprache nicht signifikant beeinträchtigen. Für die Kombination ZFKI/RKOP wurde die Codierung von Konturphasen berücksichtigt.

Die Orientierung setzt sich aus zwei Schritten zusammen. Im ersten wurde ein einfaches Codierschema mit gleichförmiger Quantisierung von Zeit, Frequenz, Pegel und Phase untersucht, welches die benötigten Repräsentationen über Stützstellen codiert. Damit konnte experimentell für jede Kombination ein Satz kritischer Quantisierungsstufen bestimmt werden, bei dem die Beeinträchtigung gerade noch nicht signifikant ist. Wegen der zeitvarianten Stützstellenanzahl bei den Konturen mußte eine Statistik der typischen Stützstellendichte angefertigt werden, um schließlich zu mittleren Datenraten zu gelangen. Diese reichen von 100 bis 300 kbit/s und liegen damit im Bereich der Datenrate des PCM-Signals, welches mit 100 bis 200 kbit/s bei meist besserer Qualität anzusetzen ist. Weil die Bewertung der kritischen Quantisierung nur durch eine Versuchsperson erfolgte, können die Werte nur grobe Anhaltspunkte darstellen.

Die im einzelnen erzielten, mittleren Datenraten gibt die letzte Zeile von Tabelle 6.3 auf S. 167 wieder. Mit Ausnahme der Kontur/Textur-Kombinationen in den mittleren beiden Spalten gibt der Anstieg der Rate von HB-TTZM über M-TTZM und ZFKII bis ZFKI auch das höhere erzielbare Qualitätsniveau wieder, welches bei ZFKI fast mit dem des PCM-Signals übereinstimmt. KTX und KTXOZ erlauben das Qualitätsniveau von ZFKII bei gleicher Datenrate wie HB-TTZM. Während der Anteil der Zeitkonturen an den Gesamtdatenraten sehr gering ausfällt, dominiert meistens der Anteil der Frequenzkonturen. Nur bei KTX und KTXOZ sinkt letzterer deutlich, weil sich die nichtprägnanten

Frequenzkonturen in komprimierter Form in der Texturrepräsentation unterbringen lassen. KTX hat dabei einen gewissen Vorteil, weil die prägnanten Zeitkonturanteile separat codiert werden.

Als zweiter Schritt der Orientierung wurde theoretisch untersucht, wie das einfache Codierschema ergänzt werden kann, um verbleibende Redundanz und Irrelevanz zu eliminieren (Theorie hierzu in Abschnitt 1.2). Demnach bietet eine frequenzabhängige Zeitquantisierung eine gewisse Reduktionsmöglichkeit. Unter zusätzlicher Beschränkung des Nutzfrequenzbereiches auf Telefonsprache bis 3,5 kHz könnten günstigstenfalls nochmal 75% Redundanz aus den berechneten Datenraten entfernt werden. Damit erscheinen Datenraten in der Nähe von 30 bis 80 kbit/s greifbar. Möglichkeiten der Entropiecodierung zur Elimination von Verteilungsredundanz sind dabei nicht berücksichtigt. Weiterhin besteht die Möglichkeit einer adaptiven Differenzcodierung für längere Konturlinien, deren Auswirkung schwer einzuschätzen ist. Außerdem sind Möglichkeiten der Entropiecodierung zur Elimination weiterer Redundanz noch nicht berücksichtigt. Zur Verminderung von perzeptiver Irrelevanz könnte man noch eine psychoakustische Nachmaskierung der Stützstellen vorsehen.

Insgesamt erscheint ein hohes Qualitätsniveau bei niedrigen Datenraten für Signalrepräsentationen mit Konturen oder mit Kontur/Textur nicht leicht erreichbar. Die Qualität von ISO-MPEG-2 (Layer III) bei 32 kbit/s ist in etwa mit der einer ZFKI/RKOP-Kombination vergleichbar, hat aber einen Nutzfrequenzbereich bis 10 kHz [MPE95, Bra94]. Als interessantes Nebenergebnis ist zu vermerken, daß die Codierung von Phaseninformation nicht nur einen Qualitätseinbruch verhindern hilft, der aufgrund mangelhafter Phasenrekonstruktion sonst hinzunehmen wäre. In einem gewissen Maß kann sie gegen Frequenzinformation bei Frequenzkonturen und gegen Zeitinformation bei Zeitkonturen eingetauscht werden. Ihre Codierung stellt damit nicht von vornherein ein Manko dar.

6.3 Niedrigratige Sprachcodierung

Nun soll die Datenrate gegenüber den Ergebnissen des vorigen Abschnittes mit wenigen ‘einfachen’ Maßnahmen deutlich gesenkt werden. Damit wird einerseits der Verlust bestimmter wahrnehmungsrelevanter Signaleigenschaften und andererseits ein gewisses Maß an Artefakten bei der Signalrekonstruktion in Kauf genommen. Wesentliche sprachrelevante Kategorien, wie Verständlichkeit, Prosodie oder Sprecheridentifizierbarkeit, brauchen dadurch aber nicht allzusehr beeinträchtigt zu sein. Dies belegen die datenreduzierenden Varianten des TTZM-Verfahrens von Heinbach.

Zunächst wird beschrieben, warum dazu das Frequenzkontur/Textur-Verfahren als Grundlage gewählt wird, das für niedrigratige Codierung geringfügig zu modifizieren ist. Danach werden zwei Codierungen mit Raten von 30 und 4,4 kbit/s vorgestellt, die nach subjektiver Beurteilung des Autors einen Bereich von befriedigender bis eben ausreichender Sprachqualität markieren. Sie enthalten noch ein gewisses Maß an Redundanz, so daß sie zukünftig noch zu niedrigeren Raten hin optimiert werden könnten. Beide Codierungen werden mit anderen Verfahren zur Sprachcodierung verglichen.

6.3.1 Wahl des Frequenzkontur/Textur-Verfahrens und Modifikation

Ausgangspunkte für eine niedrigratige Codierung wären eigentlich beide Verfahrenskombinationen mit Textur, KTX/RKHPTX und KTXOZ/RKHPTX. Sie erzielten im vorigen Abschnitt bei niedrigen Raten die beste Qualität. Gegenüber den rein konturorientierten Verfahren dürften sie drei Probleme entschärfen, die bisher mit massiven Reduktionsmaßnahmen zusammentrafen (vgl. Abschnitt 2.6.5): Bei Vergröberung der Zeitquantisierung läßt sich erstens eine Tonalisierung nichttonaler Anteile verhindern, da sie getrennt von den Frequenzkonturen repräsentiert und rekonstruiert werden. Weiterhin verbindet sich mit dem Kontur/Textur-Konzept der Gedanke, daß alle Signalanteile, die nicht durch Konturen erfaßt sind, über die Textur repräsentiert werden (Differentialprinzip). Dadurch erscheint zweitens eine spektral/zeitliche Kontrastverschärfung vermeidbar, die vorher bei Reduktion auf wenige Frequenzkonturen hinzunehmen war. Wegen des Differentialprinzips kann drittens vermutlich auch auf den Nutzeffekt der lästigen periodischen Knackstörung verzichtet werden: Kein Fehlen bestimmter Signalanteile braucht mehr durch 'Knattern' verschleiert zu werden.

Dennoch wird die Verfahrenskombination KTX/RKHPTX für den Rest dieser Arbeit ausgeklammert. Neben der Verringerung von Aufwand und Coderlaufzeit (Anhang B.5) sind dafür drei Gründe entscheidend:

- Die angestrebte grobe Zeitquantisierung kann dicht aufeinanderfolgende Zeitkonturlinien nicht mehr richtig erfassen. Um nur isolierte Zeitkonturen auszuwählen, die auch in der Wahrnehmung dominieren, bedarf es eines verbesserten Prägnanzmaßes. Haben die Linien nämlich eine gewisse Länge überschritten, dann können sie allein aufgrund ihrer Länge nicht mehr gegeneinander abgewogen werden. Anhand von Frequenzkonturen ist das leicht nachzuvollziehen: Ab einer gewissen Länge lassen sie keine Aussagen mehr über die Ausgeprägtheit einer möglicherweise assoziierbaren Tonhöhenempfindung zu.

- Wenn neben prägnanten Frequenzkonturen auch prägnante Zeitkonturen zu codieren sind, dann entsteht ein Wettbewerb um die zu Verfügung stehende Kanalkapazität. Prägnanzmaße, welche die Linienstücke beider Konturtypen gegeneinander abwiegbar machen, wären erst noch zu erarbeiten.

- Wegen des Differentialprinzips des Kontur/Textur-Konzeptes kann auf eine Zeitkontur-Repräsentation leichter als bei Verfahren mit reiner Frequenzkontur-Repräsentation (TTZM-Verfahren) verzichtet werden. Hinzu kommt, daß Zeitkonturen weniger sprachrelevant sind.

Das letztlich zu verwendende Verfahren KTXOZ/RKHPTX erfordert allerdings einen Eingriff, der mit der Codierung der prägnanten Frequenzkonturen zusammenhängt. Später soll nämlich die Anzahl ihrer gleichzeitig vorhandenen Stützstellen beschränkt werden. Das bedeutet, daß zusammen mit der ursprünglichen Auswahl der prägnanten Frequenzkonturen (APF, Bild 4.3 auf S. 106) effektiv ein verändertes Prägnanzkriterium realisiert wird. Die nachträglich aussortierten und somit implizit als nichtprägnant erachteten Li-

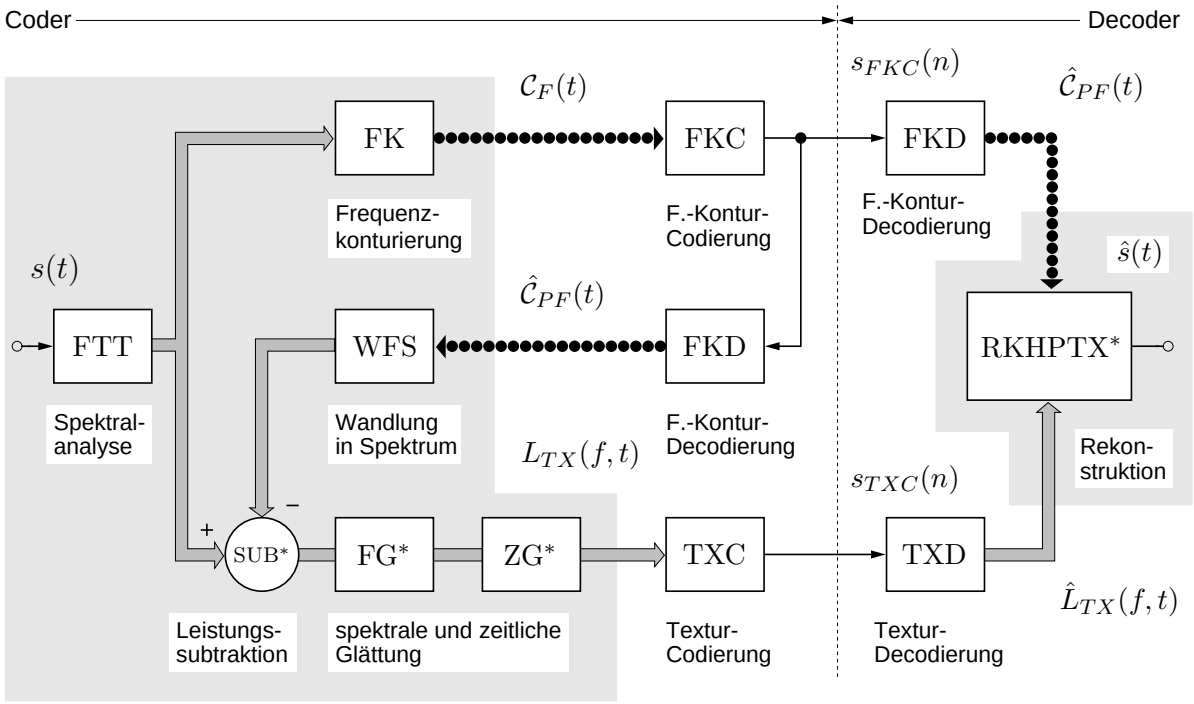

Bild 6.2: *Blockschaltbild für niedrigratige Codierungen mit Frequenzkonturen und Textur. Grau unterlegte Teile stammen von der Analyse/Synthese-Kombination* KTXOZ/RKHPTX *ab. Mit* $*$ *versehene Operationen weisen geringfügige Änderungen auf. Im Vergleich mit Bild 4.3 übernimmt* FKC *in Reihe mit* FKD *analyseseitig die bisherige Operation* APF*. Teile von prägnanten Konturen, die* FKC *für eine hohe Datenreduktion fallen lassen muß, werden dadurch in der Textur aufgefangen.*

nienstücke dürfen jedoch nicht einfach wegfallen. Um 'Löcher' im Spektrum des später zu rekonstruierenden Signals zu vermeiden, sollten sie im Sinne des Differentialprinzips des Kontur/Textur-Konzeptes der Textur zugeschlagen werden. Bild 6.2 zeigt die dazu notwendige, modifizierte Codierung auf Basis von KTXOZ/RKHPTX. Im Gegensatz zu der allgemeinen Darstellung in Bild 6.1 sind Analyseverfahren und Codierung nunmehr ineinander verschränkt.

Die Codierung der Frequenzkonturen (FKC) in Reihe mit ihrer Decodierung (FKD) auf Coderseite ersetzen zusammengenommen die Operation APF in Bild 4.3. Je nach Spezifikation der späteren Codierverfahren kann APF in FKC verborgen sein und im Kontursignal $\hat{\mathcal{C}}_{PF}(t)$ möglicherweise das gleiche Ergebnis liefern. Dieses steht durch die zweite Ausführung von FKD auch auf Decoderseite zur Verfügung, effizient übertragen durch den Bitstrom $s_{FKC}(n)$. Wenn – verglichen mit APF – weniger oder gar andere Konturlinien geliefert werden, dann sorgt die Modifikation dafür, daß immer die übrigen Signalanteile in der Textur auftauchen.

Die Codierung der Texturhüllfläche $L_{TX}(f,t)$ und ihre Decodierung als $\hat{L}_{TX}(f,t)$ übernehmen ein unabhängiges Codierungspaar TXC und TXD, zwischen denen der Bitstrom $s_{TXC}(n)$ übertragen wird. Natürlich werden in der Praxis beide Bitströme zu einem einzigen im Sinne von Bild 6.1 verflochten. Geringfügige Parameteränderungen der anderen Operationen des KTXOZ/RKHPTX-Verfahrens sind, abhängig vom Codierverfahren, möglich.

6.3.2 Codierung 30 kbit/s mit Frequenzkontur/Textur (MUM-30k)

Für dieses Verfahren werden die Operationen FKC, FKD, TXC und TXD in Bild 6.2 und zwei weitere Parametermodifikationen spezifiziert und erläutert. Dazu zeigt Bild 6.3 für das bekannte Sprachbeispiel die am Rekonstruktionsverfahren RKHPTX des Decoders anliegenden Frequenzkonturen und die dort wirksame (s.u.) Texturhüllfläche.

Codierung der Frequenzkonturen (FKC): Aus dem Frequenzkontursignal $\mathcal{C}_F(t)$ in Bild 6.2 werden zunächst die prägnanten Konturen mit einer Mindestlänge von 25 ms ausgewählt, so wie es die Operation APF im unmodifizierten Kontur/Textur-Verfahren vorsieht. Eine recht grobe Zeitquantisierung erzeugt daraus Stützstellen im Abstand von $\Delta[t_F]_Q = 10$ ms. Anschließend wird die Anzahl der Stützstellen pro Zeitschritt beschränkt, indem nur die zehn pegelstärksten weitergeführt werden. Wie in Tabelle 6.2 aus den Werten 12,9 der ersten Zeile zu schließen ist, fallen dadurch im Mittel nur etwa zwei bis drei Stützstellen pro Zeitschritt weg. Dafür ist nun eine Maximaldatenrate garantierbar. Danach werden die Frequenz- und Pegelparameter der Stützstellen quantisiert, letztere etwas gröber und dynamikbeschränkter als bisher. Die Spalte MUM-30k in Tabelle 6.4 stellt Quantisierungsdaten und Codebreiten (oberste Abteilung) sowie den Rechengang für die Teildatenrate I_F (unterste Abteilung) zusammen. Um auf die Codierung einer Stützstellenanzahl verzichten zu können und um auf eine gleichmäßige Datenrate zu kommen, sind bei weniger als zehn Stützstellen erkennbare 'Füllstellen' einzufügen. Beispielsweise können im aktuellen Zeitschritt bereits codierte Stützstellen wiederholt oder – falls Stille herrscht – alle Pegel auf den codierbaren Minimalwert gesetzt werden.

Decodierung der Frequenzkonturen (FKD): Die bislang angenommene Treppenstufenapproximation für den Frequenz- und Pegelverlauf einer Linie reicht nicht mehr aus, weil die Zeitquantisierung zu grob geworden ist. Jetzt können nämlich die Stufen hörbar werden, selbst wenn das vergleichsweise kurze Synthesefenster in RKHPTX spektrale Verbreiterungen (Knacke) nach wie vor verhindert. Frequenz und Pegel werden deshalb zwischen denjenigen Stützstellen linear interpoliert, welche innerhalb einer recht groben Toleranz von $\Delta f'_U = 0{,}5$ Bark zum Grundgerüst einer Frequenzkonturlinie assoziierbar sind (Anhang A.2). Hinter der letzten Stützstelle einer Linie oder bei nichtassoziierbaren Stützstellen werden die Werte für die Dauer einer Zeitquantisierungsstufe konstant gehalten. Verglichen mit reiner Treppenstufenapproximation dauern die so approximierten Linien zwar genauso lange. Sie sind aber um einen halbe Zeitquantisierungsstufe verzögert, weil die Treppenstufe symmetrisch zur Stützstelle definiert wurde (Abschnitt 6.1). Für eine optimale auditive Fusion mit der Decoder-Textur muß dieser Zeitversatz extra korrigiert werden. Im Ergebnis erhält man das Decoder-Kontursignal $\hat{\mathcal{C}}_{PF}(t)$, das in Gestalt der schwarzen Linien in Bild 6.3 zu sehen ist.

Parametermodifikationen: Die Operationen SUB und ZG erfordern gegenüber ihrer Grundeinstellung KTXOZ in Kapitel 4 leicht veränderte Einstellungen. Auch hier werden in SUB zur Ermittlung der Texturhüllfläche $L_{TX}(f,t)$ die Spektralbeiträge des Decoder-Kontursignals $\hat{\mathcal{C}}_{PF}(t)$ vom Originalspektrum abgezogen. Wegen Quantisierung und Approximation korrespondieren die Decoder-Konturen aber nicht mehr exakt mit den Konturverläufen im Originalspektrum. Um ihre Spektralbeiträge dennoch sicher aus der Textur 'herauszustanzen', wird der Pegelzuschlag ΔL_{PF} beim Abzug der rückgewandelten Konturen von 3 auf 5 dB erhöht. Weiterhin kann die grobe Zeitquantisierung der Textur bei

Tabelle 6.4: *Quantisierung, Codierung und Datenraten von Verfahren zur niedrigratigen Sprachcodierung. Die Verfahren* MUM-30k *und* MUM-4k4 *codieren Frequenzkonturen und Textur. Das Verfahren* HB-4k4 *von Heinbach auf Basis seines Teiltonzeitmusters ist zum Vergleich miteingetragen.* $\Delta[x]_Q$ *bezeichnet die Quantisierungsstufe eines Parameters* x *und* $x_{min} \ldots x_{max}$ *seinen Quantisierungsbereich.*

Stützstellenparameter		**Quantisierung und Codierung**		
		MUM-30k	MUM-4k4	HB-4k4
FK-Tonheit	$\Delta\,[z(f_F)]_Q$ /Bark	0,05	0,07	0,07
	Anzahl	10	5	10
	$z(20\text{Hz}) \ldots z(5{,}5\text{kHz})$ Code	9bit	-	-
	$z(100\text{Hz}) \ldots z(5{,}5\text{kHz})$ Code	-	8bit	8bit
	b_1 = Anzahl × Codebreite	90bit	40bit	80bit
FK-Pegel	$\Delta\,[L_F]_Q$ /dB	1	2	4
	Anzahl	10	2	2
	$-80\text{dB} \ldots 0\text{dB}$ Code	7bit	-	-
	$-64\text{dB} \ldots 0\text{dB}$ Code	-	6bit	-
	$-60\text{dB} \ldots 0\text{dB}$ Code	-	-	4bit
	b_2 = Anzahl × Codebreite	70bit	12bit	8bit
FK-Zeit	$\Delta\,[t_F]_Q$ /ms	10	20	20
TX-Tonheit	$\Delta\,[z(f_{TX})]_Q$ /Bark	0,95	1,6	-
	Anzahl	20	8	-
	$z(70\text{Hz}) \ldots z(5{,}1\text{kHz})$ Code	0bit	-	-
	$z(850\text{Hz}) \ldots z(5\text{kHz})$ Code	-	3bit	-
	b_3 = Anzahl × Codebreite	0bit	24bit	-
TX-Pegel	$\Delta\,[L_{TX}]_Q$ /dB	1	2	-
	Anzahl	20	2	-
	$-80\text{dB} \ldots 0\text{dB}$ Code	7bit	-	-
	$-64\text{dB} \ldots 0\text{dB}$ Code	-	6bit	-
	b_4 = Anzahl × Codebreite	140bit	12bit	-
TX-Zeit	$\Delta\,[t_{TX}]_Q$ /ms	10	20	-
FK-Datenrate	$I_F = (b_1 + b_2)/\Delta\,[t_F]_Q$	16 kbit/s	2,6 kbit/s	4,4 kbit/s
TX-Datenrate	$I_{TX} = (b_3 + b_4)/\Delta\,[t_{TX}]_Q$	14 kbit/s	1,8 kbit/s	-
Datenrate	$I = I_F + I_{TX}$	30 kbit/s	4,4 kbit/s	4,4 kbit/s

schnellen zeitlichen Änderungen Aliasing verursachen, was sich später als Rauhigkeit bemerkbar machen würde. Deshalb wird in ZG die Glättungsbandbreite B_{3dB}^{ZG} von 300 auf 100 Hz gesenkt.

Codierung der Textur (TXC): Hier wird lediglich eine relativ grobe, gleichförmige Quantisierung der Texturhüllfläche $L_{TX}(f,t)$ durchgeführt, die eine feste Anzahl von Stützstellen pro Zeitschritt liefert. Die Zeitquantisierung stimmt mit der der Konturen überein, unterschiedliche Werte bewähren sich nicht. Vielmehr würden sie in der Wahr-

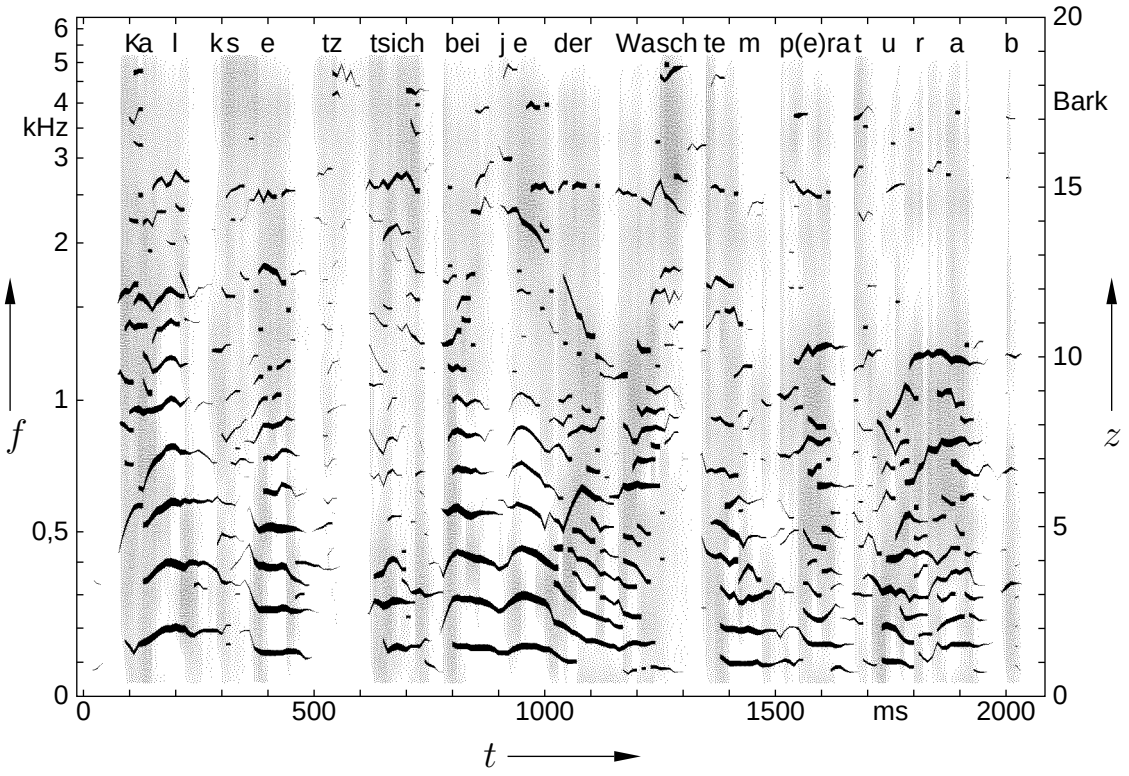

Bild 6.3: *Prägnante Frequenzkonturen und Texturrepräsentation, approximiert im Decoder des Verfahrens* MUM-30k. *Obwohl die benötigte Datenrate nur* 30 kbit/s *beträgt, ist die erreichbare Sprachqualität noch etwas besser als beim nichtreduzierenden TTZM-Verfahren von Heinbach (≫* 100 kbit/s*, vgl. Bild 1.2 oben). Zur Texturdarstellung wurde die später über das Synthesefenster von* RKHPTX *erzielte spektrale Glättung vorweggenommen (siehe Text).*

nehmung die Fusion von Kontur und Textur behindern, weil abrupte Pegeländerungen am Konturlinienbeginn zeitlich nicht immer mit denen der Textur zusammenträfen. Weitere Angaben zur Quantisierung und Codierung der Textur sowie zur Ermittlung ihrer Teildatenrate I_{TX} finden sich in Tabelle 6.4. Aufgrund der spektralen Abtastung im Abstand von etwa 1 Bark ist theoretisch Aliasing möglich, weil die spektrale Glättung FG nur Feinstrukturen unterhalb 0,7 Bark entfernt. In der Praxis von Sprachsignalen wurden aber keine nachteiligen Effekte bemerkt, wogegen eine höhere Glättung bereits stören würde.

Decodierung der Textur (TXD): Anders als bei den Konturen wird hier in zeitlicher Richtung weiterhin Treppenstufenapproximation angewendet. Sie gewinnt aus den Stützstellen die Texturhüllfläche $\hat{L}_{TX}(f,t)$ zurück. Jede Art von zeitlicher Interpolation schadet, weil dann bei abrupten Pegeländerungen eine Fusion mit rekonstruierten Konturlinienanfängen behindert wird. Im übrigen scheinen die zeitlichen Stufen hier keine Nachteile zu haben. In Frequenzrichtung wird keine Approximation mehr vorgenommen, auch nicht im Texturrasterierer TR von RKHPTX in Bild 5.7 auf S. 152. In der Praxis gelangt dort die an einer Stützfrequenz $f = [f_{TX}]_Q$ zeitlich approximierte Texturhüllfläche nur noch zu demjenigen SZM-Steuereingang, dessen Synthesefrequenz ω_{S_m} am nächsten liegt. Übrigbleibende Steuereingänge werden auf null gelegt. [5] Formal entartet die Texturhüllfläche in Frequenzrichtung zu einer Folge von Dirac-Impulsen mit zeitvarianten Impulsgewichten. Sie wirkt aber in RKHPTX mit der Fourier-Transformierten des Syn-

[5] Deswegen muß die Balance c_{bal} zwischen Textur und Kontur in Gl. (5.41) nachjustiert werden.

thesefensters geglättet, also etwa über einen Bereich von 0,7 Bark. Vorherige Approximation würde eine zusätzliche Glättung bedeuten, welche angesichts der bereits recht groben Tonheitsquantisierung feinere Strukturen unnötig entfernt. Speziell für eine sinnvolle bildliche Darstellung von $\hat{L}_{TX}(f,t)$ aber wird nun eigens eine Approximation erforderlich. Die Grauwerthinterlegung in Bild 6.3 verwendet dazu eine Operation vom Typ FG mit einer 3dB-Breite $B_{3dB}^{FG\prime} = 1$ Bark (Anhang B.5).

6.3.3 Vergleich von MUM-30k mit anderen Verfahren

Vergleicht man die im Decoder approximierte Kontur/Textur-Darstellung aus Bild 6.3 mit derjenigen des ursprünglichen Analyseverfahrens KTXOZ aus Bild 4.5 auf S. 109 unten, so erkennt man auf den ersten Blick eine recht gute Übereinstimmung. Der oberflächliche Höreindruck bei Wiedergabe im Diffusfeld unterscheidet sich auch gar nicht so sehr. Ein genauerer Vergleich der Bilder deckt allerdings die verringerte Anzahl von Konturen und die insgesamt verminderte zeitliche Feinstruktur auf. Besonders gut ist dies in den Höhen der Textur zu sehen, wo dort noch viele Glottisimpulse aufgelöst werden, während hier die Darstellung gut geglättet wirkt.

Das Verfahren MUM-30k ermöglicht dennoch eine Sprachqualität, die im großen und ganzen die des ursprünglichen Heinbachschen TTZM-Verfahrens noch etwas übertrifft, jedenfalls nach subjektiver Einschätzung des Autors. Die Datenrate ist dabei um mindestens ein Drittel reduziert, gegenüber dem ermittelten Wert von 108 kbit/s (Tabelle 6.3). Vorteilhaft wirkt sich die verbesserte FTT aus, die eine Halligkeit verhindert und für eine bessere Wiedergabe zeitvarianter Signaleigenschaften sorgt. Jedoch gibt es auch hier schwache Tonalisierungseffekte. Dies liegt an der groben Zeitquantisierung, die die feinzeitliche Variablilität der Frequenzkonturen nicht detailgetreu erfassen kann.

Der Vorteil der Texturverarbeitung ist erstens, daß rauschhafte Anteile wirklich durch spektral/zeitlich geformtes Rauschen wiedergegeben – also nicht tonalisiert werden (Abschnitt 2.5). Zweitens ist sichergestellt, daß Spektralbereiche mit transienten Signalanteilen nicht mehr unberücksichtigt bleiben. Allerdings bewirkt die vergröberte Zeitquantisierung der Textur auch einen gewissen Detailverlust. Sie erfordert coderseitig eine stärkere zeitliche Glättung der Texturhüllfläche, wodurch markante Rauhigkeiten des Signals etwas ausgeglättet werden. Bei manchen synthetischen Signalen, wie etwa bei der Impulsfolge aus Abschnitt 2.1.2, sind die resultierenden Artefakte leider unangenehmer als beim TTZM-Verfahren, auch wenn es nun keinen Höhenverlust mehr gibt.

Eine Prägnanzentscheidung aufgrund der Linienlänge kann nicht genau die wahrnehmungsrelevanten tonalen Anteile erfassen. Außerdem mündet sie wegen der nachfolgenden Beschränkung auf zehn Frequenzkontur-Stützstellen effektiv in ein neues Prägnanzkriterium. Dadurch treten Fehlentscheidungen etwas öfter als nur beim reinen Linienlängenkriterium auf, so daß auch relevante Spektraltonhöhen entsprechend öfter durch geformtes Rauschen ersetzt werden. Darin manifestiert sich nochmals ein gewisser Detailverlust, wodurch das verarbeitete Signal bei Kopfhörerwiedergabe stellenweise leicht ‘verwaschen’ klingt. Die Detailverluste, also durch Fehlentscheidungen wie durch grobe Zeitquantisierung, relativieren die gegenüber Heinbach verminderte Halligkeit etwas.

MUM-30k erzielt nicht die Qualität von etablierten Sprachcodierverfahren. Dies zeigen Hörvergleiche des Autors mit Implementierungen folgender Verfahren: ISO-MPEG-2 bei

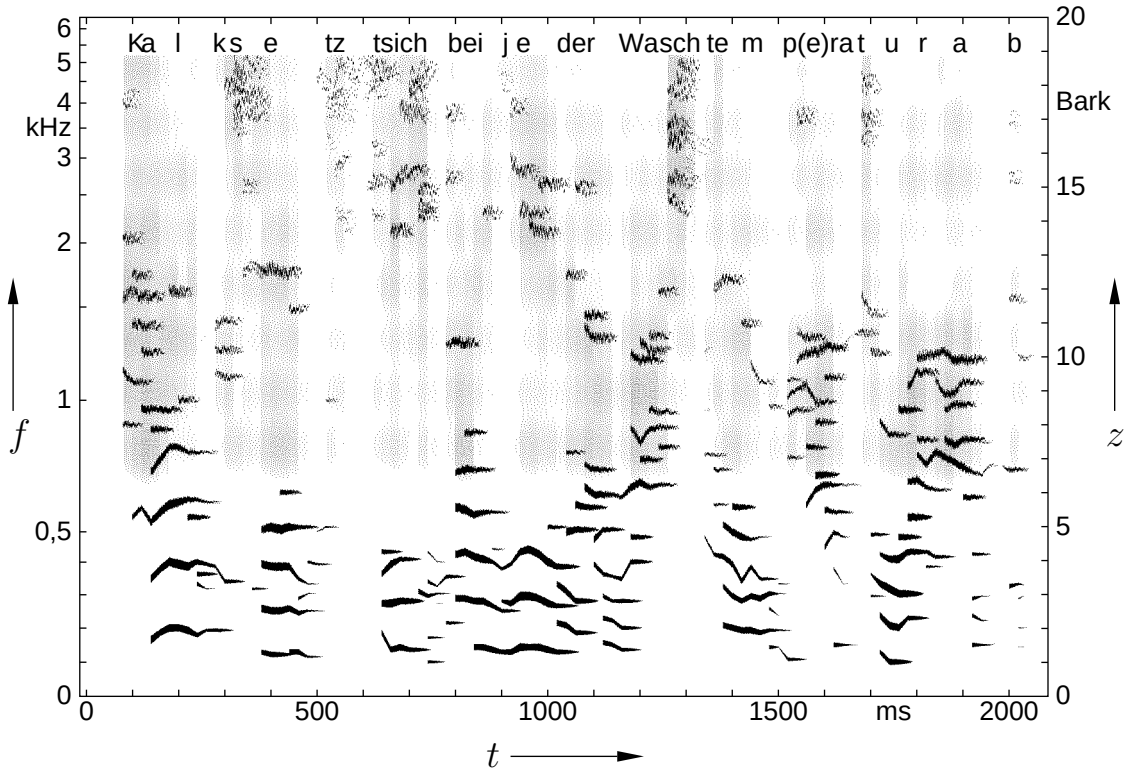

Bild 6.4: *Frequenzkonturen und Texturrepräsentation approximiert im Decoder des Verfahrens* MUM-4k4. *Darin reichen fünf Kontur- und acht Texturstützstellen pro* 20 ms *Zeitschritt aus, um bei gleicher Datenrate eine bessere Sprachqualität als das Verfahren von Heinbach (vgl. Bild 1.2 auf S. 13 unten) zu erzielen. Die nach oben zunehmende Zittrigkeit der Konturen ist künstlich erzeugt und unterdrückt tonale Artefakte. Zur Texturdarstellung wurde die spektrale Glättung durch das Synthesefenster des Rekonstruktionsverfahrens vorweggenommen.*

32 kbit/s (Layer III, Mode 4, [MPE95]) ist trotz des vergrößerten Nutzfrequenzbereichs bis 10 kHz deutlich besser. Sieht man von der Beschränkung auf den Telefonfrequenzbereich bis etwa 3,5 kHz ab, dann können auch die Qualitäten von LD-CELP (ITU-T G728) mit 16 kbit/s [LDC95] oder vom GSM-Full-Rate-Standard für Mobilkommunikation mit 13 kbit/s [Var88, GSM95] nicht erreicht werden. Erst im Vergleich mit dem US-Federal-Standard 1016 CELP mit 4,8 kbit/s [CEL95] kann MUM-30k in einigen Kriterien Pluspunkte verbuchen. Immerhin ist zu vermuten, daß der Bitstrom von MUM-30k noch ein gewisses Maß an Redundanz beinhaltet, da das gewählte Codierungsschema recht simpel ist.

6.3.4 Codierung 4,4 kbit/s mit Frequenzkontur/Textur (MUM-4k4)

Bis auf drei weiter unter beschriebene Besonderheiten ist dieses Verfahren strukturell wie MUM-30k angelegt. Das Sprachbeispiel in Bild 6.4 zeigt wiederum Frequenzkonturen und wirksame Texturhüllfläche im Decoder. Zuerst werden rein quantitative Änderungen besprochen, dann die drei Besonderheiten.

6.3.4.1 Änderung der Quantisierung und sonstiger Parameter

Um die Datenrate weiter senken zu können, wird die Quantisierung nochmals vergröbert. Außerdem werden die Quantisierungsbereiche eingeschränkt und noch weniger Stützstellen zugelassen, wie in Tabelle 6.4 in Spalte MUM-4k4 angegeben. Zum Vergleich ist auch das datenreduzierende Heinbachsche TTZM-Verfahren bei gleicher Rate aufgeführt (HB-4k4). Mit einer Zeitquantisierungsstufe von 20 ms hat man allerdings die gröbstmögliche Zeitauflösung erreicht. Größere Werte beeinträchtigen die Verständlichkeit, weil wesentliche artikulatorische Merkmale nicht mehr sicher abgetastet werden. Dies ist der kritischste Parameter im ganzen Verfahren. Eine Halbierung würde einen spürbaren Anstieg der Qualität, aber auch eine Verdoppelung der Gesamtrate bewirken. Mit fünf Konturstützstellen und acht Texturstützstellen scheint jedenfalls eine Untergrenze erreicht. Günstigerweise wurde festgestellt, daß die Textur unterhalb etwa 800 Hz nicht besonders zur Qualität beiträgt und deshalb nicht unbedingt codiert werden muß.

Um bei der groben Texturquantisierung negative Effekte wegen zeitlicher und nunmehr auch spektraler Unterabtastung zu vermeiden, müssen die Operationen ZG und FG, verglichen mit MUM-30k, stärker glätten. Dazu werden die Parameter B_{3dB}^{ZG} von 100 auf 50 Hz beziehungsweise B_{3dB}^{FG} von 0,7 auf 1 Bark geändert. Zur spektralen Approximation der Textur im Decoder muß die Synthesebandbreite B_{3dB}^{S} nun ebenfalls von 0,7 auf 1 Bark angehoben werden. Sonst wird der spektrale Stützstellenabstand nicht mehr hinreichend geglättet und als Kammfilterstruktur hörbar. Zur Grauwerthinterlegung in Bild 6.4 wurde wie schon bei MUM-30k eine Glättungsoperation vom Typ FG verwendet, diesmal mit einer 3dB-Breite $B_{3dB}^{FG\prime} = 1{,}2$ Bark.

Von der Erhöhung der Synthesebandbreite ist auch die Rekonstruktion aus Frequenzkonturen betroffen. Daß insbesondere die Linienanfänge nun ‘härter’ eingeschaltet werden, scheint sich aber eher positiv auf die Qualität und möglicherweise auch auf die Verständlichkeit auszuwirken. Hier gibt es eine Parallele zum Nutzeffekt der hart geschalteten Synthesesinusschwingungen in HB-4k4 (Abschnitt 2.6.3).

6.3.4.2 Verzicht auf Linienlänge als Prägnanzmaß

Die erste Besonderheit gegenüber MUM-30k besteht darin, daß die Operation APF innerhalb der Codierung der Frequenzkonturen ersatzlos entfällt. Folglich werden die fünf pegelstärksten Stützstellen aus der Gesamtheit aller Frequenzkonturen im aktuellen Zeitschritt ausgewählt, so wie dies für die zehn ‘Teiltöne’ von HB-4k4 geschieht. Während sich APF bei MUM-30k nur entfernen läßt, wenn man verstärkte Tonalisierung hinnimmt, ergeben sich bei MUM-4k4 eher Vorteile. Höherfrequente Spektralanteile von nichtstimmhaften Sprachanteilen bleiben hier markanter und sicherer bewahrt, offenbar mit positivem Einfluß auf die Verständlichkeit. Beispielsweise sind nun scharfe Frikative – besonders ‘(t)s(ich)’ in Bild 6.4 – in den Frequenzkonturen besser aufgehoben. Bei Beibehaltung von APF würden sie sich in einer anschließend ziemlich grob quantisierten Textur nur noch ‘verwaschen’ wiederfinden.

APF und pegelorientierte Stützstellenbeschränkung zusammengenommen bedeuten demnach für MUM-4k4 ein schlechteres effektives Prägnanzkriterium als letztere für sich alleine. Dies gilt aber nur dann, wenn Kontur- und Texturrepräsentation über ein bestimmtes Maß vergröbert sind. Dann nämlich kann eine Texturrepräsentation mit gröbe-

rer spektraler Auflösung die irrtümlich zugeordneten Frequenzkonturlinien nicht mehr so gut durch Schmalbandrauschen annähern. Erschwerend tritt dann auch hinzu, daß kurzzeitige Texturanteile bei Vergröberung der zeitlichen Auflösung benachteiligt werden: Ist die Glättung in der Operation ZG zu stark, so werden sie unmittelbar gedämpft, ist sie zu schwach, dann werden sie nicht mehr sicher abgetastet. Bei feinerer Zeitauflösung der Frequenzkonturen und geringer Stützstellenbeschränkung ist die Operation APF dagegen wichtig, weil sie die Anzahl der Linien reduziert, ohne sie zu zerstückeln.

6.3.4.3 Anwendung der Heinbachschen Pegelcodierung

Eine zweite Besonderheit verkörpert die schon in HB-4k4 verwendete Heinbachsche Pegelcodierung, die bei etwa 25% Ratenreduktion kaum Nachteile aufweist, wenn für eine gute Linienapproximation (s.u.) gesorgt ist. Nachdem die Stützstellen eines Zeitschritts ausgewählt, quantisiert und nach Pegel sortiert worden sind, werden nacheinander alle Frequenzen, von den Pegeln aber nur der stärkste und schwächste codiert. Die fehlenden Pegel der dazwischenliegenden Stützstellen interpoliert die Decodierung später linear über dem Sortierindex. Das wiederholte Auftreten derselben Frequenz kennzeichnet Füllstellen, wenn weniger als die vorgesehene Maximalanzahl von Stützstellen vorkommen. Wenn die aktuelle Anzahl dagegen null beträgt, zeigen beide Pegelcodes auf den minimalen Pegelwert. Die Methode wird für Frequenzkontur- und Texturstützstellen getrennt angewandt, so daß sich die in Tabelle 6.4 aufgeführten Codebreiten und Datenraten ergeben. Während bei der Textur bisher keine Frequenzen codiert zu werden brauchten, sind nunmehr zumindest die Sortierindices beizufügen (vgl. b_3 in Tabelle 6.4). Dadurch fällt die Ersparnis für die Textur geringer als für die Konturen aus.

6.3.4.4 Weiterentwickelte Linienapproximation

Zur Decodierung der Frequenzkonturen werden die Stützstellen wie bei MUM-30k mit $\Delta f'_U = 0{,}5$ Bark zu einem Liniengrundgerüst assoziiert. Auch hier approximieren lineare Frequenz- und Pegelinterpolation zwischen den assoziierten Stützstellen sowie Konstanthalten nach der letzten zunächst den Linienverlauf. Wenn man allerdings die so approximierten Konturen direkt in RKHPTX verwendet, dann fallen tonale Artefakte ('Klingeln') unangenehm auf, ähnlich wie bei HB-4k4. Das liegt zum einen an der groben Zeitquantisierung, die schon im Coder verhindert, daß die feinzeitliche Variabilität besonders bei höherfrequenten Linien abgetastet werden kann (Abschnitt 2.6.2). Zum anderen liegt es daran, daß die Operation APF aus dem Coder entfernt wurde. Darüberhinaus bemerkt man eine unnatürliche Intensitätsschwankung. Sie ist auf abrupte Linienenden infolge einer häufigen Umordnung der fünf ausgewählten Stützstellen zurückzuführen (vgl. Abschnitt 2.6.4).

Als dritte Besonderheit enthält der Frequenzkonturdecoder von MUM-4k4 deshalb zwei einfache, rein experimentell abgestimmte Heuristiken, die die eben beschriebenen Nachteile recht gut bekämpfen: Erstens erhält der Pegelverlauf einer Linie, die soweit mit der bisher beschriebenen Methode approximiert wurde, eine weiche Ausblendfunktion am Linienende t_{F_e}. Weil die zugehörige Vorschrift um 5 ms über t_{F_e} hinausreicht, muß die Linie zuerst bei konstantgehaltenen Parametern verlängert werden. Die aufzuschlagende

Pegelkorrektur lautet dann

$$\Delta L_F(t) = \lg\left[\frac{1}{2} - \frac{1}{2}\cdot\sin\left(\frac{\pi}{2}\cdot\frac{t-t_{F_e}}{5\text{ ms}}\right)\right]\text{ dB}, \quad \text{für} \quad |t-t_{F_e}| < 5\text{ ms}. \tag{6.4}$$

Zweitens werden die Pegel- und Frequenzverläufe einer Linie frequenzprogressiv 'verzittert', was in Bild 6.4 ab etwa 1 kHz gut zu erkennen ist. Dazu werden zeitabhängige Abweichungen $\Delta L_F(t)$ und $\Delta f_F(t)$ aufgeschlagen, die im Abstand von 1,25 ms auf einen neuen Wert umspringen. Zwei unabhängige zeitdiskrete Zufallsprozesse $x_1(n)$, $x_2(n)$ dienen zur Steuerung. Beide sind gleichverteilt mit identischer Wahrscheinlichkeitsdichte $p(x)$. Ein Faktor $\eta(f)$, der über der Tonheitskala $z(f)$ von null bis maximal etwa eins ansteigt, bewirkt die Frequenzprogression in Abhängigkeit von der Linienfrequenz $f_F(t)$. Unter Verwendung der Frequenzgruppenbreite $\Delta f_G(f)$ ist der Vorgang wie folgt definiert:

$$\Delta L_F(t) = 6\text{ dB}\cdot x_1(n)\cdot\eta\Big(f_F(n\cdot 1{,}25\text{ ms})\Big), \tag{6.5}$$

$$\Delta f_F(t) = 0{,}375\cdot\Delta f_G\Big(f_F(n\cdot 1{,}25\text{ ms})\Big)\cdot x_2(n)\cdot\eta\Big(f_F(n\cdot 1{,}25\text{ ms})\Big), \tag{6.6}$$

$$\text{wobei jeweils} \quad n \le \frac{t}{1{,}25\text{ ms}} < n+1,$$

$$p(x) = \begin{cases} 0{,}5 & \text{für} \quad x\in[-1;1[, \\ 0 & \text{sonst}, \end{cases} \tag{6.7}$$

$$\eta(f) = \left(\frac{z(f)}{20\text{ Bark}}\right)^2. \tag{6.8}$$

6.3.5 Vergleich von MUM-4k4 mit anderen Verfahren

Als erstes interessiert hier natürlich die Gegenüberstellung mit dem Heinbachschen TTZM-Verfahren bei gleicher Datenrate (HB-4k4). Zum Vergleich mit etablierten Verfahren wurden außerdem frei verfügbare Implementierungen der US-Federal-Standards 1016 CELP mit 4,8 kbit/s [CEL95] und 1015 LPC-10e mit 2,4 kbit/s [LPC95] herangezogen. Zwar sind sie nur für den Telefonfrequenzbereich bis etwa 3,5 kHz vorgesehen. Andererseits ist die mögliche Ersparnis bei Reduzierung des Nutzfrequenzbereiches von MUM-4k4 unerheblich.

Um die Verarbeitungsergebnisse der Verfahren möglichst objektiv, trotzdem aber gehörorientiert diskutieren zu können, bietet sich die Konturdarstellung an. Das bedeutet, daß das verarbeitete, also vom Verfahren wieder ausgegebene Signal einer Konturanalyse zugeführt wird. Im Falle von MUM-4k4 und HB-TTZM liegt dadurch gewissermaßen eine 'Re-Analyse' vor. Zu jedem der vier Verfahren zeigt Bild 6.5 einen Sprachsignalausschnitt mit den analysierten Konturen. Verwendet wurde der Parametersatz ZFKII, die Konturen des unverarbeiteten Ausschnitts finden sich in Bild 3.9 auf S. 89 zum Vergleich.

Stellvertretend für sinustongestützte Codierverfahren, die zur effizienten Darstellung geräuschhafter Signalanteile erweitert wurden (Abschnitt 1.6), wird MUM-4k4 schließlich der hybriden Harmonischen Codierung von Marques und Abrantes [Mar94] gegenübergestellt. Sie ist von den Autoren unter anderem auch für 4,8 kbit/s spezifiziert und mit CELP verglichen worden. Da keine Implementierung zur Verfügung stand, kann hier nur eine theoretische Beurteilung erfolgen.

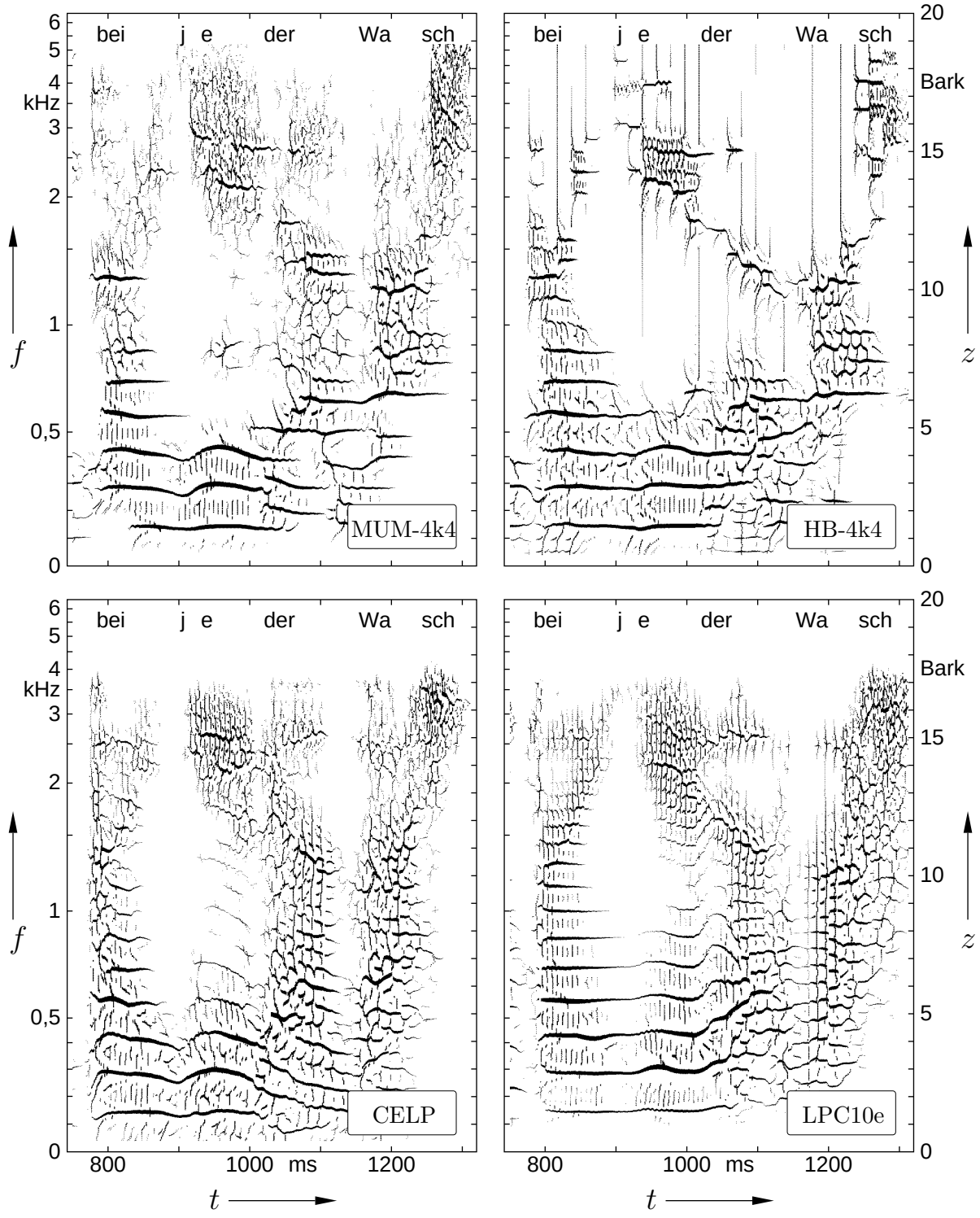

Bild 6.5: *Konturanalyse eines Sprachausschnittes, der eines der folgenden Codierverfahren durchlaufen hat (Datenraten in Klammern):* MUM-4k4 $(4,4\,\mathrm{kbit/s})$*, dann* HB-4k4 $(4,4\,\mathrm{kbit/s})$ *nach Heinbach, US-Federal-Standards 1016* CELP $(4,8\,\mathrm{kbit/s})$ *und 1015* LPC10e $(2,4\,\mathrm{kbit/s})$*. Konturen des unverarbeiteten Originals in Bild 3.9* ZFKII *auf S. 89. Diskussion siehe Text.*

6.3.5.1 Heinbachsches TTZM-Verfahren bei 4,4 kbit/s (HB-4k4)

MUM-4k4 hinterläßt im direkten Hörvergleich über eine Reihe verschiedener Sprachsignale folgenden Eindruck: Die Qualität ist merklich verbessert, die Verständlichkeit scheint in etwa gleich geblieben. Die besonders im gerichteten Schallfeld (Kopfhörer, Telefonhörer oder Lautsprechernahfeld) unakzeptablen Störungen sind verschwunden, tonale Artefakte sind deutlich reduziert. Die nunmehr maximal fünf statt zehn gleichzeitigen Frequenzkonturlinien/Teiltöne führen zwar zu neuen, allerdings weniger störenden Artefakten, weil öfter eine Spektraltonhöhe (Abschnitt 3.1.1, STH) durch ein Rauschen repräsentiert wird.

Betrachtet man die zugehörigen Konturen in Bild 6.5, dann sind bei HB-4k4 gut die Zeitkonturen im Abstand von 20 ms zu identifizieren, die die hart ein- und ausgeschalteten Synthesesinusschwingungen markieren (Abschnitt 2.6.3). Weiterhin kommen im Hochtonbereich die tonalen Ersetzungen zur Geltung, ganz besonders im Frikativ ‘sch’ (Abschnitt 2.6.2). MUM-4k4 dagegen vermeidet beide Effekte und zeigt auch weniger Tendenz zu spektral/zeitlichen Löchern, die eine unnatürliche Kontrastverschärfung bedeuten (Abschnitt 2.6.1). Beide Verfahren zeigen eine deutliche Detailreduktion gegenüber den Konturen des Originals, neben der reduzierten Harmonischenanzahl sind auch die Zeitkonturen der Glottisimpulse praktisch verschwunden.

6.3.5.2 CELP 4,8 kbit/s und LPC-10e 2,4 kbit/s

Beide Verfahren verwenden lineare Prädiktion (LP), um das Sprachsignal in Formanthüllkurve und Anregungssignal zu zerlegen, womit sie an Sprachproduktionsmodelle angelehnt sind (z.B. [Osh87], [Rab78]). Das ältere Verfahren LPC-10e unterscheidet zwischen stimmhafter Anregung, deren Grundfrequenz extrahiert wird, sowie stimmloser und gemischter Anregung [Tre82]. Es klingt deshalb im Ergebnis künstlich, versagt bei bestimmten Lautübergängen und verhält sich wenig robust gegenüber Hintergrundgeräuschen. Insbesondere auch die Sprecheridentifizierbarkeit leidet erheblich. Diese Nachteile vermeidet CELP (Codebook Excited LP), da verschiedene Wellenformen als Anregungssignale adaptiv vektorquantisiert werden [Cam90]. Dafür ist im Ergebnis eine gewisse Rauhigkeit und Verzerrtheit zu bemerken.

Da MUM-4k4 auf einem Wahrnehmungsmodell basiert, ist es ebenfalls weniger den Beschränkungen von LPC-10e ausgesetzt. Mit Begrenzung auf etwa 3,5 kHz abgehört, scheinen jedenfalls Robustheit und Sprecheridentifizierbarkeit besser zu sein. Qualität und Verständlichkeit liegen wohl etwa auf dem gleichen Niveau. Die Beeinträchtigungen sind allerdings recht unterschiedlicher Natur, so daß genaue Aussagen objektive Hörversuche erfordern. Dennoch wird in keiner dieser Kategorien das Niveau von CELP erreicht. Sieht man vom eingeschränkten Nutzfrequenzbereich ab, dann weist Bild 6.5 für CELP die größte Übereinstimmung mit dem unverarbeiteten Signal in Bild 3.9 auf. Auch LPC-10e scheint sich eher noch als MUM-4k4 und HB-4k4 am Original zu orientieren. Man erkennt aber die übertriebene Regelmäßigkeit der stimmhaften Anregung, die durch Umschaltung vor ‘sch’ schlagartig in eine stimmlose übergeht.

6.3.5.3 Hybride harmonische Codierung 4,8 kbit/s

Der wesentliche Unterschied zu MUM-4k4 besteht wiederum darin, daß dieses Verfahren auf einem Sprachproduktionsmodell fußt. Im Endeffekt überlagern allerdings beide Ansätze zeitvariante Sinusschwingungen und Schmalbandrauschsignale, um das Signal zu rekonstruieren. Die hybride harmonische Codierung geht von vornherein von harmonischen Sinustönen aus, was zunächst eine bessere Eliminierung der Redundanz von Sprachsignalen vermuten läßt. Nach einer Kurzzeitspektralanalyse wird die Grundfrequenz mit Hilfe eines harmonischen Siebs extrahiert. Die Fehlerträchtigkeit dieses Vorgangs verringert Qualität und Robustheit im Vergleich zu CELP, wie in [Mar94] berichtet wird. Da die Analysebandbreite konstant und nicht dem Gehör angepaßt ist, müssen teilweise auch die Phasen der Harmonischen codiert werden. Dies relativiert den Vorteil gegenüber MUM-4k4 etwas, wo anstelle einer Grundfrequenz individuelle Konturfrequenzen zu codieren sind. Spektrale Hüllkurven von tonalen und geräuschhaften Anteile werden über ein statistisches Verfahren bestimmt und mit Hilfe von LP-Parametrierungen codiert. Geräuschhafte Anteile sind dadurch möglicherweise etwas redundanzfreier als bei MUM-4k4 repräsentiert.

Im Hörvergleich dürfte die hybride Harmonische Codierung besonders in rein stimmhaften Passagen etwas akzeptabler klingen. Die hier wirksame Zerlegung in harmonische Anregung und spektrale Vokaltrakthüllkurve lieferte auch schon bei LPC-10e gute Ergebnisse. MUM-4k4 kann bestenfalls die fünf wichtigsten Spektraltonhöhen codieren, der Rest wird durch Textur ersetzt. Dafür ist MUM-4k4 sicher robuster, weil weder Harmonizität vorausgesetzt wird, noch eine Grundfrequenz erkannt zu werden braucht.

6.3.6 Zusammenfassung

Aufbauend auf einer Kontur/Textur-Repräsentation ohne Zeitkonturen wurde niedrigratige Sprachcodierung behandelt, soweit sie mit einfachen Reduktionsmaßnahmen erreicht werden kann. Der Einsatz von Zeitkonturen bedarf weiterer Untersuchungen. Ihre Abwesenheit erscheint jedenfalls tolerierbar, da sie aufgrund des Kontur/Textur-Konzeptes ersatzweise über die Textur erfaßt werden können. Als Rahmen dient die Verfahrenskombination KTXOZ/RKHPTX, in der analyseseitig eine kleine strukturelle Modifikation vorgenommen wurde. Sie garantiert, daß die Textur die später decodierbaren Frequenzkonturen ohne spektral/zeitliche ‘Löcher’ ergänzt. Einige der Reduktionsmaßnahmen setzte schon Heinbach zur Datenreduktion ein. Dabei handelt es sich um Vergröberung der Quantisierung, Beschränkung auf die pegelstärksten Stützstellen und um die spezielle Form der Pegelcodierung. Neu aber ist, daß eine Linienapproximation im Decoder eingeführt wird, und insbesondere eben, daß Textur verarbeitet wird.

Speziell wurden zwei Codierverfahren mit 30 kbit/s (MUM-30k) beziehungsweise 4,4 kbit/s (MUM-4k4) spezifiziert. Nach subjektiver Einschätzung des Autors erzielt MUM-30k eine bessere Sprachqualität als Heinbachs nichtreduzierendes TTZM-Verfahren, welches mindestens die dreifache Rate beansprucht. MUM-4k4 übertrifft die Qualität der reduzierenden Heinbachschen Variante bei der gleichen Rate. Die vormals lästigen Artefakte sind verschwunden, neu hinzutretende stören viel weniger. Dies ist möglich, obwohl nur noch fünf statt zehn ‘Teiltöne’ codiert werden. Der freigewordene Platz wird von acht Texturstützstellen beansprucht. Kritischster Parameter bleibt die grobe Zeitquantisierung

von 20 ms. Bei Halbierung, entsprechend 8,8 kbit/s Gesamtdatenrate, kann die Qualität merklich gesteigert werden.

Je niedriger die angestrebte Rate ist, desto wichtiger wird die richtige Auswahl von wenigen prägnanten Konturen. Die Linienlänge erweist sich dabei leider als ein zu grobes Prägnanzmaß. Wegen der nachfolgenden Beschränkung auf pegelstärkste Stützstellen wird außerdem das Prägnanzkriterium indirekt auf ungünstige Weise modifiziert. Bei MUM-4k4 gelten deshalb die pegelstärksten Stützstellen aus der Gesamtheit aller Frequenzkonturen als 'prägnant'. In Zukunft sollte sich die Auswahl an der Gewichtung dynamisch empfundener Spektraltonhöhen orientieren. Einen interessanten Modellansatz bietet Horn an, der die ursprünglichen Konturen mit Hilfe einer zeitlich und spektral geglätteten Version des FTT-Spektrums nachmaskiert [Hor96]. Diesen Ansatz könnte man mit stationären Tonhöhenberechnungsverfahren verbinden [Ter82]. Explizite Verarbeitung von Zeitkonturen (s.o.) blieb übrigens aus ähnlichen Gründen ausgeklammert: Auch hier fehlt ein wahrnehmungsorientierteres Prägnanzkriterium, das überdies Prägnanz auch im Vergleich zu Spektraltonhöhen bewerten können sollte.

Die beiden neuen Verfahren markieren eine Spannbreite in den Kategorien Qualität, Verständlichkeit, Sprecheridentifizierbarkeit und Robustheit, in dessen Mitte ungefähr das etablierte Verfahren US-Federal-Standard 1016 CELP mit 4,8 kbit/s anzuordnen wäre. Die Datenraten sind gegenüber einer Veränderung der Nutzgrenzfrequenz von 5,5 auf beispielsweise 3,5 oder 7,5 kHz relativ unempfindlich. Da beide Verfahren mit Sicherheit noch Redundanz enthalten, lassen zukünftige redundanzmindernde Optimierungen eine Verschiebung zu niedrigeren Raten hin erwarten.

6.4 Zusammenfassung und Ausblick

Für Kontur- und Kontur/Textur-Repräsentationen wurde untersucht, inwieweit sich mit einfachen Codierungsmaßnahmen brauchbare Verfahren zur datenreduzierenden Sprachcodierung aufstellen lassen. Kombinationen von Analyse- und Rekonstruktionsverfahren der vorigen Kapiteln verarbeiten die Repräsentationen in einer Auflösung, die feiner als nötig ist. In solche Analyse/Synthese-Kombinationen sollte eine 'eigentliche' Codierung mit einer passenden Decodierung eingefügt werden, um akzeptable Kompromisse bei Datenrate und Verarbeitungsqualität zu erzielen. Konturlinien und Texturhüllflächen wurden dabei grundsätzlich mit Hilfe von Stützstellen codiert. Die erforderlichen Quantisierungs- und Approximationsvorgänge standen im Mittelpunkt, Ansätze zur optimalen Codewahl blieben unberücksichtigt.

Die bekannten Verfälschungen der Analyse- und Rekonstruktionsverfahren beschränken die Qualität einer Codierung von vornherein mehr oder weniger deutlich. Zur Orientierung sollten zunächst diejenigen Datenraten erkundet werden, bei denen sich die wahrnehmbare Verarbeitungsqualität nicht noch zusätzlich verschlechtert. Dafür wurde ein einfaches Codierungsschema zugrunde gelegt, das die Stützstellenparameter Zeit, Tonheit, Pegel und gegebenenfalls Phase gleichförmig quantisiert. Im Selbstversuch wurde dann die gröbstmögliche, sogenannte kritische Quantisierung für Sprache ermittelt. Da hier die zu codierende Stützstellenanzahl für Konturen noch zeitabhängig ist, erlaubt erst eine Statistik die Berechnung der effektiven Datenraten.

Die erzielten Raten reichen, je nach Analyse/Synthese-Kombination, von 100 bis 300 kbit/s, womit gegenüber dem codierten PCM-Signal von rund 150 kbit/s nichts gewonnen wurde. Verbesserte redundanzarme Codierungsschemata, bei denen eventuell noch verbliebene Irrelevanz entfernt wird, bleiben ein weites Experimentierfeld für die Zukunft. Immerhin erzielten die Kombinationen mit Textur niedrigste Werte bei recht guter Qualität. Folglich erleichtert das Kontur/Textur-Konzept die Datenreduktion. Ein bemerkenswertes Nebenergebnis dieser Untersuchungen besagt schließlich, daß eine Codierung von Konturphasen die Gesamtdatenrate nicht besonders zu erhöhen braucht. Die Stützfrequenzen von Frequenzkonturen und die Stützzeiten von Zeitkonturen können dann nämlich gröber quantisiert werden. Ein besonderer Vorteil der codierten Phasen liegt darin, daß Störungen durch suboptimale Phasenrekonstruktion von vornherein vermieden werden können. Damit liegt das praktisch erzielbare Qualitätsniveau höher.

Nach dieser Orientierung sollte, mit weiterhin einfachen Codierungsmaßnahmen, wirkliche Datenreduktion erreicht werden. Dies geschah in der Hoffnung, daß sich die nun unvermeidlichen Qualitätseinbrüche nicht als unakzeptabel erweisen würden. Dazu wurden die Quantisierung weiter vergröbert und zusätzlich Stützstellen eingespart. Eine geeignete Approximation im Decoder gewinnt hierbei an Bedeutung. Die im weiteren zugrunde gelegte Analyse/Synthese-Kombination (KTXOZ/RKHPTX) verarbeitet eine Kontur/Textur-Repräsenation ohne explizite Zeitkonturen. Diese blieben ausgeklammert, da sie bei niedrigen Datenraten nicht mehr wahrnehmungsgerecht ausgewählt werden können. Die Analyse wurde so modifiziert, daß die bei Codierung zurückgewiesenen Frequenzkontur-Stützstellen keine spektral/zeitlichen 'Löcher' hinterlassen, sondern der Textur zugeschlagen werden. Auf dieser Grundlage wurden zwei Verfahren eingeführt:

Datenreduktionsverfahren MUM-30k und MUM-4k4: Das erste Verfahren übertrifft das nichtreduzierende Heinbachsche TTZM-Verfahren bei Sprache noch in der Qualität und benötigt mit 30 kbit/s höchstens ein Drittel der Datenrate. Das zweite übertrifft bei 4,4 kbit/s das reduzierende TTZM-Verfahren mit gleicher Rate, weil weniger störende Artefakte auftreten. Verglichen mit etablierten Verfahren markieren beide Verfahren, nach Einschätzung des Autors, einen Qualitätsbereich, in dessen Mitte ungefähr der US-Federal-Standard 1016 CELP mit 4,8 kbit/s einzuordnen wäre. Der Bereich liegt allerdings noch etwas unterhalb vom GSM-Full-Rate-Standard für Mobilkommunikation mit 13 kbit/s. Immerhin wird eine höhere Robustheit als beim US-Federal-Standard 1015 LPC-10e mit 2,4 kbit/s erreicht. Beide Codierungen enthalten sicherlich noch Redundanz, so daß zukünftige Optimierungen den aufgespannten Ratenbereich nach unten drücken könnten.

Zur Qualitätssteigerung bei niedrigen Datenraten muß künftig die Auswahl der Frequenzkontur-Stützstellen verbessert werden. Weil nur wenige von ihnen codiert werden können, heben sich Fehlentscheidungen deutlicher heraus. Zwar kann man über die Konturlinienlänge als Prägnanzmaß grob zwischen tonalem und nichttonalem Beitrag unterscheiden. Man kann aber damit nicht mehr differenziert Linien oder Liniensegmente in ihrer Wahrnehmbarkeit gegeneinander abwägen. Nach dieser unsicheren Vorauswahl müssen dann noch pegelschwächere Stützstellen entfernt werden, um die Stützstellenrate zu beschränken. Dadurch erhöht sich die Wahrscheinlichkeit von Fehlentscheidungen nochmals. Für niedrige Raten ist es deshalb vorläufig besser, nicht mehr die Linienlänge, sondern allein den Stützstellenpegel als Prägnanzmaß heranzuziehen. Für ein differenziertes

Prägnanzmaß wäre in Zukunft ein Modell der dynamischen Spektraltonhöhenwahrnehmung sehr willkommen. Vielleicht läßt sich auch das erwähnte, ähnlich gelagerte Problem der Datenreduktion von Zeitkonturen durch ein entsprechendes, psychoakustisch fundiertes Modell lösen.

Zusammenfassung

Die vorliegende Arbeit behandelt neuartige, wahrnehmungsorientierte Audiocodierungen. Sie verarbeiten im 'Gebirge' eines gehörangepaßten Spektrogramms Konturen, die grob etwa seinen 'Gratlinien' entsprechen. Außerdem können wenig prägnante Konturen, als Textur bezeichnet, wahlweise durch eine grobe Hüllfläche über Zeit und Frequenz repräsentiert werden. Dadurch ist es möglich, rauschhaft empfundene Anteile getrennt von tonalen oder impulshaften zu verarbeiten. Die Grundlagen einer Codierung mit solchen Kontur- und Kontur/Textur-Repräsentationen und die erreichbare Verarbeitungsqualität werden ausführlich untersucht. Im besonderen wird die Anwendung zur Datenreduktion von Sprache ausgelotet.

Das gehörangepaßte Spektrogramm ergibt sich als zeitvariantes Pegelspektrum aus einer speziellen Kurzzeitspektralanalyse. Die Fourier-t-Transformation (FTT) nach Terhardt (1985) weist eine Analysebandbreite proportional zur Frequenzgruppenbreite des Gehörs auf. Konturen erhält man zum einem Teil dadurch, daß ausgeprägte lokale Maxima in frequenzparallelen Schnitten des FTT-Spektrogramms verfolgt werden. Diese sogenannten Frequenzkonturen erfassen quasistationäre Anteile des FTT-Spektrogramms, die unter anderem den tonal empfundenen Signalanteilen entsprechen. Den übrigen Teil der Konturen erhält man auf ähnliche Weise, nun aber werden zeitparallelen Schnitte zugrunde gelegt. Die sogenannten Zeitkonturen erfassen transiente Anteile des FTT-Spektrogramms, die unter anderem impulshaft empfundenen Signalanteilen entsprechen.

Kontur- und Kontur/Textur-Repräsentationen sollen nicht das Signal an sich, sondern nur seine gehörrelevanten Eigenschaften erfassen. Dabei stützen sie sich auf ein schematisches Modell der auditiven Informationsaufnahme nach Terhardt (1992), das eine Analogie zur visuellen Wahrnehmung herstellt. Auf dieser Grundlage wurden Frequenzkonturen in einem FTT-Spektrogramm schon von Heinbach (1988) unter dem Begriff 'Teiltonzeitmuster' (TTZM) eingeführt. Er verband damit die Vorstellung, daß einzelne Frequenzkonturlinien gehörrelevanten, zeitvarianten Sinusschwingungen entsprechen, die zur gehörgerechten Rekonstruktion des Signals zu überlagern sind. Das resultierende TTZM-Verfahren verarbeitet Audiosignale mit bestimmten Verfälschungen, bewahrt aber wesentliche gehörrelevante Eigenschaften. Heinbach stellte außerdem eine Variante zur Datenreduktion von Sprache vor, deren Übertragungsqualität allerdings wenig akzeptabel ist. In der vorliegenden Arbeit werden deshalb, nach einem ausführlichen Grundlagenkapitel, im zweiten Kapitel zunächst die Grenzen des Heinbachschen Verfahrens untersucht. Verfälschungsursachen liegen demnach in der bisher verwendeten FTT, in der Repräsentation nur mit Frequenzkonturen und in der Signalrekonstruktion begründet.

Das dritte Kapitel führt Zeitkonturen ein und optimiert die FTT für das Zusammenspiel mit den Konturierungsvorgängen. Um alle gehörrelevanten Anteile im FTT-Spektrogramm

zu erfassen, reichen nämlich Frequenzkonturen grundsätzlich nicht aus. Beispielsweise bleibt ein Einzelimpuls praktisch unrepräsentiert. Transiente Anteile, die durch kurzzeitige spektrale Verbreiterungen gekennzeichnet sind, werden erst durch Zeitkonturen berücksichtigt. Sie sind selbst bei Rauschen von Bedeutung. Nur mit beiden Konturtypen zusammen kann man eine nahezu verfälschungsfreie Audiosignalverarbeitung erzielen, optimale Parametereinstellung und optimale Signalrekonstruktion vorausgesetzt. Bei Sprache weisen Zeitkonturen besonders auf Glottisimpulse und Plosive hin. Wesentliche sprachliche Information wird allerdings von Frequenzkonturen in viel höherem Maße transportiert. Außerdem erleichtern, wie im Falle des TTZM-Verfahrens, suboptimale Parametereinstellung und suboptimale Signalrekonstruktion einen Verzicht auf Zeitkonturen.

Die Wahl des kausalen Analysefensters in der FTT spielt für die Konturierungsvorgänge eine wichtige Rolle. Geeignete reelle Fensterfunktionen entsprechen den Impulsantworten von Tiefpässen mit einem n-fachen reellen Pol. Die von Heinbach verwendete Funktion weist mit $n = 1$ den niedrigsten Grad auf. Erst mit $n = 4$ kann man aber, bei geeigneter Wahl der Analysebandbreite, die zeitliche und spektrale Selektivität richtig an das Gehör anpassen. Gleichzeitig heben sich transiente und quasistationäre Anteile des Spektrogramms immer besser von einander ab, so daß Zeitkonturen auch erst dann sinnvoll sind. Zwar wird nun ein Laufzeitausgleich zwischen verschiedenen Spektrogrammfrequenzen erforderlich, dafür entfällt die von Heinbach benötigte zeitliche Glättung des Spektrogramms. Nahezu verfälschungsfreie Sprachverarbeitung mit Konturen erfordert eine 3dB-Analysebandbreite von mindestens 0,5 Bark. Suboptimale Signalrekonstruktion oder Verzicht auf Zeitkonturen lassen allerdings eher 0,3 Bark ratsam erscheinen, weil die verschiedenartigen Verfälschungen dann am besten ausgewogen sind. Mit einer derart veränderten FTT kann auch die Verarbeitungsqualität des TTZM-Verfahrens deutlich verbessert werden.

Im vierten Kapitel werden Kontur/Textur-Repräsentationen eingeführt. Sie stützen sich auf die Vorstellung, daß nur ein Teil der Konturen als Einzelobjekte der Wahrnehmung interpretierbar sind. Prägnante Frequenzkonturen geben idealerweise einzeln wahrnehmbare Töne wieder. Diese sogenannten Spektraltonhöhen sind psychoakustisch bislang nur unter stationären Bedingungen erforscht. Entsprechend könnten prägnante Zeitkonturen einzeln wahrnehmbare ‘Klicks’ repräsentieren. Behelfsmäßig dient die Konturlinienlänge als Prägnanzmaß. Geht man von einer Analysebandbreite von 0,3 Bark aus, dann vertreten Frequenzkonturen mit einer Mindestlänge von etwa 25 ms bei Sprache sehr gut die tonal wahrgenommenen Anteile. Zeitkonturen mit einer Mindestlänge von 1 Bark verkörpern die impulshaften Anteile. Textur bezeichnet als Sammelbegriff die übrigen Konturen, die nur noch in ihrem gemeinschaftlichen Verhalten wahrnehmungsrelevant sind. Sie sind durch eine Hüllfläche repräsentierbar, die ein zeitlich und spektral geformtes Rauschen beschreibt. Zur Signalrekonstruktion benötigt man deshalb eine zeitvariante Rauschfilterung, die in die Signalrekonstruktion aus Konturen integrierbar ist. Kontur/Textur-Repräsentation ist auch in einer unaufwendigeren Form möglich. Diese unterscheidet nur zwischen tonalen und geräuschhaften Anteilen, indem vormals prägnante Zeitkonturen ebenfalls der Textur zugewiesen werden.

Das fünfte Kapitel behandelt die Rekonstruktion des Signals aus seinen Konturen. Die bislang unbekannte FTT-Rücktransformation, die auch verfälschungsfreie Audiocodierungen mit komplexen FTT-Spektren ermöglicht, liefert die Grundlage für ein optimales Verfahren. Darin bestimmt jeder Konturpunkt einen Sinustonimpuls (Wavelet) in Zeit-

lage, Frequenz und Amplitude, der mit den übrigen überlagert wird. Seine Hüllkurve, das sogenannte Synthesefenster, ist frequenzabhängig und entspricht in etwa einem FTT-Analysefenster mit 0,7 Bark Bandbreite. Dadurch liefert ein einzelner Konturpunkt einen Energiebeitrag, der für das Gehör bei Wiedergabe zeitlich und spektral etwa gleichermaßen konzentriert ist. Auf diese Weise werden Störungen vermieden, welche bei der Teiltonsynthese des Heinbachschen TTZM-Verfahren vorkommen. Diese verwendet ein Rechteckfenster, das den Energiebeitrag hörbar in spektraler Richtung verschmieren kann. Allerdings kann die Charakteristik dieser Störung bei Sprache dazu nützen, fehlende Zeitkonturverarbeitung zu verschleiern. Für eine verbesserte Teiltonsynthese ist deshalb ein gehörangepaßtes Synthesefenster nicht sinnvoll. Vielmehr erzielt man mit einem Dreieckfenster mit 2,5 ms Basislänge einen ausgewogenen Kompromiß zwischen nützlichen und unerwünschten Störungen.

Das zentrale Problem einer optimalen Rekonstruktion aus Konturen besteht darin, die Sinusphase innerhalb des Synthesefensters festzulegen. Dazu wird ein Nachweis skizziert, daß sich die Phasen aus dem Zusammenhang der Konturen so weit rekonstruieren lassen, wie es für das Gehör nötig ist. In diesem Sinne optimale Phasenrekonstruktion erweist sich jedoch als kompliziert, weshalb zwei Alternativen vorgestellt werden. Die eine verzichtet völlig auf Phasenrekonstruktion und vertraut darauf, daß die Phaseninformation der FTT an den Konturpunkten mitübertragen worden sind. Diese Rekonstruktion mit Originalphasen simuliert das Ergebnis einer optimalen Rekonstruktion. Die andere Alternative verwendet eine Phasenheuristik. Sie entspricht für Frequenzkonturen der Phasenfortschreibung in der Teiltonsynthese und realisiert ein ähnlich einfaches Prinzip auch für die Zeitkonturen. Eine Rekonstruktion mit Phasenheuristik kann, wie auch die Teiltonsynthese, Störungen durch Phaseninkohärenz nicht vermeiden, so daß die Rekonstruktionsqualität leider merklich absinkt. Insbesondere kommt der Nutzen der Zeitkonturen dadurch nicht voll zum Tragen. Bei Kontur/Textur-Repräsentationen schadet die Phaseninkohärenz allerdings kaum, weil sich bislang unvermeidliche Fehler bei der Prägnanzentscheidung ganz ähnlich auswirken.

Ohne spezielle Codierungsmaßnahmen wird Sprache mit Kontur- oder Kontur/Textur-Repräsentationen in recht guter Qualität verarbeitet. Die Qualität übertrifft auch ein deutlich verbessertes TTZM-Verfahren, liegt aber doch noch unter der des Originalsignals. Wenn man bei Konturrepräsentationen zusätzlich die Originalphasen mitüberträgt oder eine optimale Phasenrekonstruktion bereitstellen kann, dann lassen sich die restlichen Verfälschungen fast völlig eliminieren. Zukünftige Qualitätsverbesserungen bei Kontur/Textur-Repräsentationen erfordern darüber hinaus psychoakustisch fundierte Prägnanzkriterien.

Im sechsten und letzten Kapitel werden Möglichkeiten der Codierung mit den neuen Repräsentationen untersucht. Möchte man das realisierbare Qualitätsniveau beibehalten und dabei geringstmögliche Datenraten erzielen, dann führen einfache Codierungsmaßnahmen leider nicht weit. Werden beispielsweise die Stützstellen der Repräsentationen in Zeit, Tonheit, Pegel und gegebenenfalls Phase gleichförmig quantisiert, so liegen die erzielbaren Datenraten grob etwa bei der des Original-PCM-Signals. Kontur/Textur-Repräsentationen ermöglichen hier noch die niedrigsten Datenraten. Niedrigere Raten sind auf einfachem Wege nur dann zu erreichen, wenn zusätzliche Qualitätseinbußen hingenommen werden.

Aufbauend auf einer Kontur/Textur-Repräsentation werden schließlich zwei Verfahren von 30 kbit/s und 4,4 kbit/s vorgestellt. Das erste übertrifft das ursprüngliche Heinbach-

sche TTZM-Verfahren in der Sprachqualität und benötigt höchstens ein Drittel der Rate. Das zweite übertrifft die Heinbachsche Verfahrensvariante mit identischer Datenrate, indem weniger störende Artefakte auftreten. Diese rühren insbesondere daher, daß sich nichttonale Signalanteile allein mit datenreduzierten Frequenzkonturen schlecht wiedergeben lassen. Der US-Federal-Standard 1016 CELP für Sprachcodierung mit 4,8 kbit/s scheint qualitätsmäßig in der Mitte zwischen den beiden neuen Verfahren zu liegen. Die Robustheit ist höher als beim US-Federal-Standard 1015 LPC-10e mit 2,4 kbit/s. Beide Codierungen enthalten sicherlich noch Redundanz, so daß zukünftige Optimierungen bessere Raten erwarten lassen.

Anhang A

Definitionen

A.1 Konturen und Konturpunkte

Frequenzkonturen: Sei $L^{\times}(f,t)$ das zeitvariante FTT-Pegelspektrum mit oder ohne Glättung bzw. Laufzeitausgleich und bezeichne $\Delta L_A \geq 0$ eine Ausgeprägtheitsschwelle. Dann sind die *Frequenzkonturen* von $L^{\times}(f,t)$ die Menge $\mathcal{C}_F$ aller *Frequenzkonturpunkte* $(t_F, f_F, L^{\times}(f_F,t_F))$ im Zeit/Frequenz/Pegel-Raum, für die jeweils eine untere Frequenz $f_u < f_F$ und eine obere Frequenz $f_o > f_F$ mit

$$L^{\times}(f,t_F) + \Delta L_A \leq L^{\times}(f_F,t_F) \quad \text{für} \quad f = f_u, f_o \tag{A.1}$$

zu finden sind, mit deren Hilfe sich folgende Bedingung erfüllen läßt:

$$L^{\times}(f,t_F) - L^{\times}(f_F,t_F) \begin{cases} \leq 0 & \text{für alle} \quad f \in [f_u, f_F], \\ < 0 & \text{für alle} \quad f \in]f_F, f_o]. \end{cases} \tag{A.2}$$

Die zweizeilige Aufspaltung der Bedingung stellt sicher, daß bei einem Plateau nur der frequenzhöchste Punkt ein Konturpunkt wird. Gleiches gilt auch für den Fall gleichhoher Maxima, zwischen denen der Pegel um weniger als ΔL_A absinkt. $\mathcal{C}_F$ entspricht bei geeigneter Parameterwahl und Diskretisierung dem Teiltonzeitmuster nach [Hei88a]. [1]

Zeitkonturen: Bezeichne $L^{\times}(f,t)$ das zeitvariante FTT-Pegelspektrum mit oder ohne Laufzeitausgleich und sei $\lambda \geq 0$ eine Ausgeprägtheitsschwelle für seine partielle Ableitung $\partial L^{\times}(f,t)/\partial t$ über der Zeit. Dann sind die *Zeitkonturen* von $L^{\times}(f,t)$ die Menge $\mathcal{C}_Z$ aller *Zeitkonturpunkte* $(t_Z, f_Z, L^{\times}(f_Z,t_Z))$, für die jeweils ein früherer Zeitpunkt $t_v < t_Z$ mit

$$\left.\frac{\partial L^{\times}(f,t)}{\partial t}\right|_{f=f_Z, t=t_v} \geq \lambda \tag{A.3}$$

und ein späterer Zeitpunkt $t_n > t_Z$ zu finden sind, mit deren Hilfe sich folgende Bedingung erfüllen läßt:

$$\left.\frac{\partial L^{\times}(f,t)}{\partial t}\right|_{f=f_Z} \begin{cases} > 0 & \text{für alle} \quad t \in [t_v, t_Z[, \\ \leq 0 & \text{für alle} \quad t \in [t_Z, t_n]. \end{cases} \tag{A.4}$$

[1] Die Definition folgt dem 'Prinzip der Teiltonbestimmung' in Fig. 3.3-5 in [Hei88a].

Die Formulierung ist so gewählt, daß auf einem eventuellen zeitlichen Plateau nur der früheste Punkt als Konturpunkt gilt. Bei Dirac-Impulsen $\delta(t-t_x)$ im Zeitsignal existiert die Ableitung $\partial L^{\times}(f,t)/\partial t$ bei $t = t_x$ im Falle des Fensters P1 nicht, weil $L^{\times}(f,t)$ dort springt. Dagegen gewährleistet die Wirkung aller anderen behandelten Fenstertypen Stetigkeit auch bei Dirac-Impulsen. Ein weiteres Existenzproblem durch $L^{\times}(f,t) \to -\infty$ stört weiter nicht, weil man an Orten mit Leistungsspektrum null keine Konturpunkte sucht.

A.2 Konturlinien und Konturlinienlänge

Frequenzkonturlinien und ihre Länge: Vorgegeben sei eine Unstetigkeitstoleranz $\Delta f_U \geq 0$ für den Frequenzverlauf. Eine *Frequenzkonturlinie* definiert eine größtmögliche Teilmenge $\mathcal{C}_F^i$ der Frequenzkonturen (s.o.), in der sich die Frequenz noch als Funktion $f^i(t)$ der Zeit über einem zusammenhängenden Zeitintervall ausdrücken läßt. Das Intervall ist durch den Startzeitpunkt t_B^i und den Endzeitpunkt $t_E^i \geq t_B^i$ der Linie begrenzt (und erweist sich möglicherweise als halboffen, offen oder gar als isolierter Punkt). Auf diesem Intervall erfüllt der Frequenzverlauf zwei Bedingungen. Erstens sind seine eventuellen Unstetigkeiten durch

$$\lim_{\substack{\epsilon \to 0 \\ \epsilon \neq 0}} \left| f^i(t+\epsilon) - f^i(t) \right| \;\leq\; \Delta f_U \tag{A.5}$$

beschränkt. Zweitens gilt zusammen mit dem Frequenzverlauf $f^m(t)$ einer beliebigen anderen Frequenzkonturlinie $\mathcal{C}_F^m$ im Rahmen zeitlicher Überschneidung folgendes:

$$\lim_{\substack{\epsilon \to 0 \\ \epsilon \neq 0}} \Big[\left| f^m(t+\epsilon) - f^i(t) \right| - \left| f^i(t+\epsilon) - f^i(t) \right| \Big] \left\{ \begin{matrix} > \\ \geq \end{matrix} \right\} 0 \quad \text{für} \quad \epsilon \left\{ \begin{matrix} > \\ < \end{matrix} \right\} 0, \tag{A.6}$$

$$\lim_{\substack{\epsilon \to 0 \\ \epsilon \neq 0}} \Big[\left| f^i(t+\epsilon) - f^m(t) \right| - \left| f^i(t+\epsilon) - f^i(t) \right| \Big] \left\{ \begin{matrix} \geq \\ > \end{matrix} \right\} 0 \quad \text{für} \quad \epsilon \left\{ \begin{matrix} > \\ < \end{matrix} \right\} 0. \tag{A.7}$$

Diese Bedingung bewirkt eine Linienassoziation, die die Minimierung von Unstetigkeiten im Frequenzverlauf vor Langlebigkeit stellt. Sie wurde aus dem Vorgehen von McAulay und Quatieri abgeleitet [Mca86]. In Konfliktsituationen werden über der Zeit aufsteigende Frequenzverläufe bevorzugt. Die *Länge* einer Frequenzkonturlinie $\mathcal{C}_F^i$ wird allein über ihre Lebensdauer definiert:

$$\Delta t_F^i = t_E^i - t_B^i. \tag{A.8}$$

Eine Unstetigkeitstoleranz $\Delta f_U = 0$ definiert *geschlossene* Linien, die im Gegensatz zu *offenen* Linien stetige Frequenzverläufe über der Zeit aufweisen müssen. Diese formale Unterscheidung vermag allerdings nur theoretisch festzuhalten, was visuell in Konturdarstellungen einfach zu erkennen ist. In der praxisrelevanten, zeitdiskreten Formulierung von Gln. (A.5) bis (A.7) fallen die Grenzübergänge weg und ϵ liegt in Gestalt des Analyseintervalls fest. Eine Vorgabe $\Delta f_U = 0$ ist damit sinnlos, sie erlaubt nur frequenzkonstante Linien. In der Praxis kann man also nur heuristisch unterscheiden, etwa mit Hilfe eines Analyseintervall-bezogenen Schwellwertes.

Zeitkonturlinien und ihre Länge: Die Definitionen für Frequenzkonturlinien gelten analog für Zeitkonturlinien, wobei die Dimensionen Zeit und Frequenz zu vertauschen

sind. Speziell wird nun eine Zeitkonturlinie als eine größtmögliche Teilmenge $\mathcal{C}_Z^k$ der Zeitkonturen (s.o.) mit der Zeit als Funktion $t^k(f)$ über einem zusammenhängenden Frequenzintervall betrachtet. Dazu sei Δt_U eine Unstetigkeitstoleranz für den Zeitverlauf und $t^l(f)$ die Funktion einer beliebigen anderen Zeitkonturline $\mathcal{C}_Z^l$. Die nunmehr relevanten Bedingungen für den Zeitverlauf der Zeitkonturlinie über der Frequenz entstehen aus Gln. (A.5) bis (A.7) durch buchstabenweise Ersetzung $t \rightarrow f$, $f \rightarrow t$, $i \rightarrow k$ und $m \rightarrow l$. Entsprechend wird die *Länge* einer Zeitkonturlinie $\mathcal{C}_Z^k$ als reiner Frequenzabstand

$$\Delta f_Z^k = f_E^k - f_B^k \tag{A.9}$$

festgelegt, wobei f_B^k die Anfangsfrequenz und $f_E^k \geq f_B^k$ die Endfrequenz ihres Definitionsintervalls markiert.

A.3 Kontursignal und Konturpunktsignal

Sei $\mathcal{C}$ eine Konturpunktmenge im Zeit/Frequenz/Pegel-Raum, etwa die Frequenzkonturen, die Zeitkonturen oder eine einzelne Konturlinie (s.o.). Dann ist ihr *Kontursignal*, in Verallgemeinerung des Teiltonmusterbegriffs nach [Hei88a], definiert als zeitabhängige Teilmenge

$$\mathcal{C}(t) \;=\; \Big\{(L_\times, f_\times, t_\times) \;\Big|\; (L_\times, f_\times, t_\times) \in \mathcal{C} \;\wedge\; t_\times = t\Big\}. \tag{A.10}$$

Ihr *Konturpunktsignal* ist die folgende zeit- und frequenzabhängige Teilmenge, die maximal einen Konturpunkt enthalten kann:

$$\mathcal{C}(f,t) \;=\; \Big\{(L_\times, f_\times, t_\times) \;\Big|\; (L_\times, f_\times, t_\times) \in \mathcal{C} \;\wedge\; f_\times = f \;\wedge\; t_\times = t\Big\}. \tag{A.11}$$

A.4 Phasenoperator

Für eine Konturpunktmenge $\mathcal{C}$ des FTT-Pegelspektrums $L^L(f,t)$, welches durch Betragslogarithmierung aus einem laufzeitausgeglichenen FTT-Bandpaßspektrum $s^{LB}(2\pi f, t)$ hervorgegangen sei, definiert der Phasenoperator

$$\begin{aligned}\Phi\{\mathcal{C}\} \;=\; \Big\{ \Big(\phi^{LB}(f_\times, t_\times), f_\times, t_\times\Big) \;\Big|\; \\ \phi^{LB}(f_\times, t_\times) = \arg\Big(s^{LB}(2\pi f_\times, t_\times)\Big) \;\wedge\; \Big(L^L(f_\times, t_\times), f_\times, t_\times\Big) \;\in\; \mathcal{C}\Big\}\end{aligned} \tag{A.12}$$

die zugehörigen Punkte des Bandpaßphasenspektrums $\phi^{LB}(f,t)$. Die Funktion arg(.) liefert den Phasenwinkel des Arguments.

Anhang B

Verfahrensbeschreibungen

B.1 Approximation lokaler Pegelmaxima über der Frequenz

Die Wahl der Analysefrequenzen tastet den Spektralverlauf des geglätteten oder laufzeitausgeglichenen FTT-Pegelspektrums $L^{\times}(f, t_A)$ zum Analysezeitpunkt t_A über der Frequenz f ab. Zunächst ergibt sich ein vorläufiges lokales Maximum aus dem Vergleich der Abtastwerte gemäß Teilton- beziehungsweise Frequenzkonturdefinition nach Anhang A.1. Dazu muß der Analysefrequenzabstand $\Delta\omega_A$ eine ausreichend dichte Abtastung gewährleisten, damit ein solches Maximum sicher zu erkennen ist (vgl. Abschnitt 3.4.3.3). Zur Approximation des exakten lokalen Maximums werden zwei Verfahren vorgestellt. Sie erhöhen bei deutlich ausgeprägten Maxima die Frequenzauflösung gegenüber dem vorläufigen Wert um mindestens eine Größenordnung.

Zur Vereinfachung der Schreibweise sei k ein Relativindex für die Frequenz/Pegel-Wertepaare (f_k, L_k) von aufeinanderfolgenden Analysefrequenzen. Speziell $k = 0$ sei der Index, bei dem ein vorläufiges, lokales Pegelmaximum vorliegt. Es werden kleine Frequenzumgebungen vorausgesetzt, in denen näherungsweise gleiche Analysefrequenzabstände $\Delta\omega_A$ angenommen werden. Auch die Analysebandbreite B_{3dB} kann dann dort als konstant gelten. Gesucht wird die Approximation $(\hat{f}_{max}, \hat{L}_{max})$ des Pegelmaximums im kontinuierlichen Spektralverlauf.

Parabel-Approximation: Das Einpassen einer Parabel in drei Abtastwerte um das vorläufige Maximum erlaubt eine sehr gute Approximation des tatsächlichen Pegelmaximums in Pegel und Frequenz. Das Verfahren wurde aus [Fei89] übernommen:

$$\hat{f}_{max} = f_0 + \frac{f_1 - f_{-1}}{4} \cdot \frac{L_1 - L_{-1}}{-L_{-1} + 2L_0 - L_1}, \hat{L}_{max} = L_0 + \frac{1}{8} \cdot \frac{(L_1 - L_{-1})^2}{-L_{-1} + 2L_0 - L_1}. \quad \text{(B.1)}$$

Feldtkeller-Verfahren: Bei den von Heinbach gewählten Parametern für das TTZM-Verfahren erlaubt folgender, von Feldtkeller stammender Ansatz eine Erhöhung der Frequenzauflösung um den Faktor 20 [Fel85]:

$$\hat{f}_{max} = \frac{\sum\limits_{\mathcal{M}} f_k g_k}{\sum\limits_{\mathcal{M}} g_k} \quad \text{mit} \quad g_k = L_k - L_0 + \Delta L_A \quad \text{und} \quad \text{(B.2)}$$

$$\mathcal{M} = \Big\{k \;\Big|\; k=0 \;\vee\; [g_k > 0 \;\wedge\; (k-1 \in \mathcal{M} \;\vee\; k+1 \in \mathcal{M})]\Big\}, \tag{B.3}$$

$$\hat{L}_{max} = L_0. \tag{B.4}$$

Die Indexmenge $\mathcal{M}$ erfaßt in einer zusammenhängenden Umgebung des vorläufigen Maximums alle Abtastwerte, deren Pegel noch nicht um mehr als als die Ausgeprägtheitsschwelle ΔL_A abgefallen sind. Der approximierte Frequenzwert repräsentiert etwa die Koordinate des Schwerpunktes einer Fläche. Sie entsteht, indem man sich die Fläche unterhalb des Spektralverlaufes entlang einer Pegellinie abgeschnitten denkt, die um ΔL_A unterhalb des vorläufigen Maximums liegt. Ursprünglich wurde auch eine Approximation des Maximumpegels in [Fel85] angegeben, die sich aber in der Praxis als fehlerträchtig herausgestellt hat [HeiPK]. Deshalb wird der Pegel des vorläufigen Maximums übernommen.

Die Besonderheit des Feldtkeller-Verfahrens liegt darin, daß es durch die Bildung des Flächenschwerpunktes eine integrierende Wirkung besitzt. Von diesem Verhalten profitieren die Teiltonzeitmuster von Zweitonschwebungen, die dadurch regelmäßiger ausfallen (vgl. Fußnote S. 36). Nur beim Heinbachschen TTZM-Verfahren wurde das Feldtkeller-Verfahren angewandt, ansonsten immer Parabelapproximation.

B.2 Realisierung von Modulator/Tiefpaß/Laufzeit-Strukturen

Beschrieben wird die zeitdiskrete Realisierung einer Struktur nach Bild B.1a, wie sie speziell für die Transformationen T und R benötigt wird (vgl. Bild 5.2d und Gln. (5.23),(5.26)). Sie erlaubt insbesondere die Berechnung des FTT-Betragsspektrums, indem nur der Betrag des Ausgangssignals herangezogen wird. Der Tiefpaß mit der Impulsantwort $h_{\omega_\times}(t)$ stellt das gewünschte Analyse- oder Synthesefenster an der Analyse- bzw. Synthesefrequenz $\omega_\times$ dar. Nach Maßgabe von Abschnitt 3.3.4 ist darauf das Laufzeitglied $l_{\omega_\times}(t)$ mit der Laufzeit $t_{L,\omega_\times}$ abgestimmt. Für ein Eingangssignal $s_1(t)$ der zeitkontinuierlichen Struktur lautet das Ausgangssignal

$$s_2(t) = \Big[\Big(s_1(t)\cdot e^{-j\omega_\times t}\Big) * h_{\omega_\times}(t) * l_{\omega_\times}(t)\Big]\cdot e^{j\omega_\times(t-t_{max,0})}. \tag{B.5}$$

Bild B.1b geht auf eine korrespondierende zeitdiskrete Struktur mit Abtastrate $f_a = 1/t_a$ über. Wie in Anhang B.4 erläutert, muß das Laufzeitglied mittels Reihenschaltung einer Verzögerungskette mit der Impulsantwort $l_{n_L}(n) = \delta(n-n_L)$ und einem Allpaß mit der Impulsantwort $\tilde{a}_\nu(n)$ realisiert werden. Zur besseren Übersicht sind hier wie dort die Abhängigkeiten der Werte t_L, n_L, ν von $\omega_\times$ weggelassen. Ein diskreter Tiefpaß mit der Impulsantwort $\tilde{h}_{\omega_\times}(n)$ approximiert die Eigenschaften des zeitkontinuierlichen Tiefpasses. Die zu Bild B.1b gehörende Signalgleichung kann wie folgt umgeformt werden:

$$s_2(n) = \Big[(s_1(n)\cdot e^{-j\omega_\times t_a n}) * \tilde{h}_{\omega_\times}(n) * \tilde{a}_\nu(n) * l_{n_L}(n)\Big]\cdot e^{j\omega_\times(t_a n - t_{max,0})} \tag{B.6}$$

$$= \Big[s_1(n) * (\tilde{h}_{\omega_\times}(n)\cdot e^{j\omega_\times t_a n}) * (\tilde{a}_\nu(n)\cdot e^{j\omega_\times t_a n}) * l_{n_L}(n)\Big]\cdot e^{j\omega_\times(t_a n_L - t_{max,0})}. \tag{B.7}$$

Dabei wurde das Distributivgesetz der Faltung bezüglich der linearen Modulation $e^{j\omega_\times t_a n}$ angewandt. Beim Operand mit $s_1(n)$ hebt sich die Modulation auf, beim Operand mit

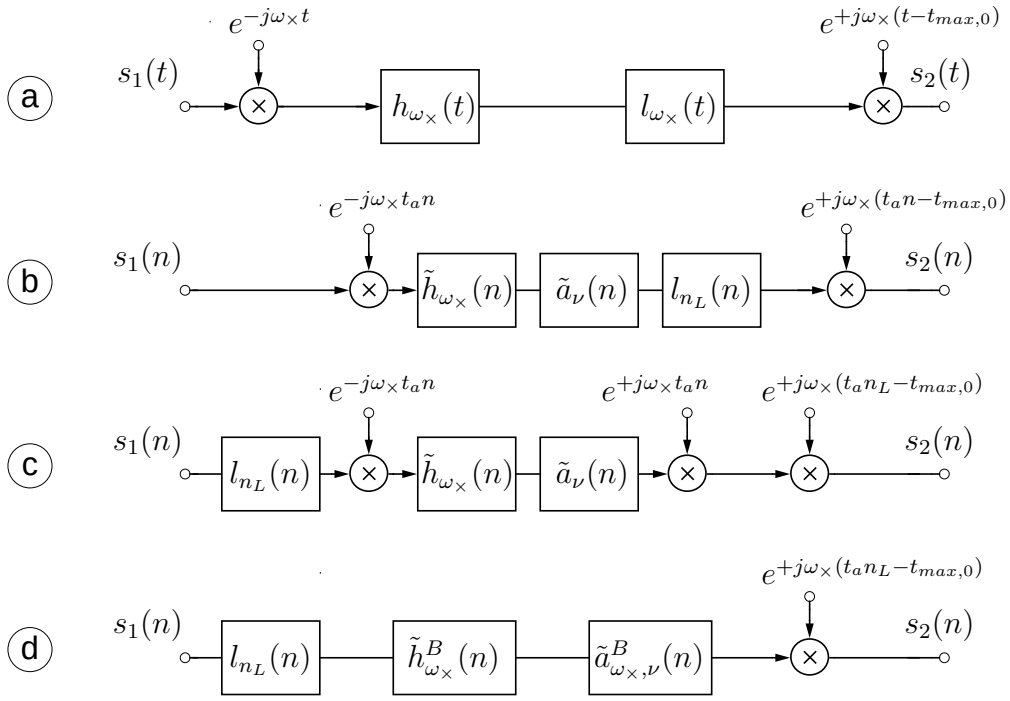

Bild B.1: *Umformung der zeitkontinuierlichen Struktur (a) für die Transformationen* T *und* R *in geeignet realisierbare zeitdiskrete Strukturen (c, d) über einen Zwischenschritt (b). Das FTT-Betragsspektrum läßt sich ebenfalls auf diesen Wegen berechnen, wobei die Modulationen/Multiplikationen mit den komplexen Exponentialfaktoren ausgangsseitig entbehrlich sind.*

$l_{n_L}(n) = \delta(n - n_L)$ wird sie zu einem konstantem Faktor $e^{j\omega_\times t_a n_L}$, der wieder hinter die eckige Klammer gezogen wurde.

Die Reihenfolge der Faltungsoperanden in Gl. (B.7) ist nun frei vertauschbar. Durch Umgruppierung unter wiederholter Anwendung des Distributivgesetzes folgt die Struktur in Bild B.1c. Zwar sind Signalwege und Operationen grundsätzlich komplex anzusetzen. Zwischen den beiden Modulatoren kann man aber auch, für Real- und Imaginärteilzweig unabhängig, zwei reelle Tiefpässe und zwei reelle Allpässe verwenden. Den Tiefpaß realisiert Anhang B.3 mit Hilfe der z-Transformierten $\tilde{H}_{\omega_\times}(z) = \mathcal{Z}\{\tilde{h}_{\omega_\times}(n)\}$ nach Maßgabe der gewünschten Fensterfunktion. Dort wird ebenfalls die Abhängigkeit von $\omega_\times$ nicht extra angeführt, die in der Vorgabe der 3dB-Bandbreite steckt.

Möchte man auf den Zugriff auf Winkelfunktionen zur Laufzeit verzichten, ihre Werte aber auch nicht tabellieren, dann eignet sich die modulatorfreie Struktur nach Bild B.1d. Sie bedingt allerdings eher mehr (reelle) Rechenoperationen, da sie nicht mehr in reelle Filterzweige aufgespalten werden kann. Die von Terhardt [Ter85] und später von Schlang [Sch89] vorgeschlagenen FTT-Rekursionen sind Sonderfälle dieser Struktur. Aus Gl. (B.7) folgt sie unmittelbar, indem ein komplexer Bandpaß und ein komplexer Allpaß mit den Impulsantworten

$$\tilde{h}^B_{\omega_\times}(n) = \tilde{h}_{\omega_\times}(n) \cdot e^{j\omega_\times t_a n}, \tag{B.8}$$

$$\tilde{a}^B_{\omega_\times,\nu}(n) = \tilde{a}_\nu(n) \cdot e^{j\omega_\times t_a n} \tag{B.9}$$

definiert werden. Man erhält die zugehörigen z-Systemfunktionen mit Hilfe der Substituti-

on $z \rightarrow ze^{-j\omega_\times t_a n}$ [Mar82]. Das bedeutet, daß in den Rekursionsgleichungen (B.18), (B.19) der Teilsektionen nach Anhang B.3 die untenstehenden Substitutionen vorzunehmen sind. Die Rechenoperationen der Rekursion werden dadurch komplex:

$$d_{k,i} \rightarrow d^B_{k,i} = d_{k,i} \cdot e^{j\omega_\times t_a}, \tag{B.10}$$

$$g_{k,i} \rightarrow g^B_{k,i} = g_{k,i} \cdot e^{j\omega_\times t_a}. \tag{B.11}$$

Die letzten beiden Strukturen haben gegenüber derjenigen aus Bild B.1b einen wesentlichen Vorteil: Man kann die Verzögerungskette $l_{n_L}(n)$ bei der Analyse ganz nach vorn zum Eingangssignal hin schieben, dessen Imaginärteil zu null angenommen wird. Umgekehrt kann man sie bei Synthese ganz nach hinten schieben, sogar hinter die Realteilbildung in Bild 5.2d. Dadurch können die individuellen Verzögerungsketten der einzelnen Kanäle zusammengefaßt werden. Es ist dann nur eine zentrale Verzögerungskette für das Haupteingangs- oder Hauptausgangssignal mit individuellen Kanalabgriffen bzw. Kanalzuflüssen einzurichten.

B.3 Realisierung von Tiefpaßfiltern durch Rekursion

Skizziert wird die Umformung eines normierten, zeitkontinuierlichen und nullstellenfreien Tiefpasses in ein zeitdiskretes, rekursives System bei gegebener Bandbreite $B_{3dB} = \omega_{3dB}/\pi$ und Abtastrate $f_a = 1/t_a$ [Hes93, Tie80]. Den Ausgangspunkt bildet die Laplace-Systemfunktion

$$H^N(P) = c \cdot \prod_{i=1}^{n_1} \frac{1}{P - \alpha_i} \cdot \prod_{i=n_1+1}^{n_1+n_2} \frac{1}{P - 2\alpha_i P + \alpha_i^2 + \beta_i^2} \tag{B.12}$$

mit der Normierungsvorschrift

$$H^N(P)|_{P=0} = 1 \quad \text{und} \quad H^N(P)|_{P=j\cdot 1} = 1/\sqrt{2} \mathrel{\hat{=}} -3\,\text{dB}, \tag{B.13}$$

von der die n_1 reellen Pole α_i und/oder die n_2 konjugiert komplexen Polpaare $\alpha_i \pm j\beta_i$ sowie die Konstante c bekannt sind. Für die im Rahmen der vorliegenden Arbeit betrachteten Tiefpässe liefert Tabelle B.1 die Zahlenwerte. Die Bilinear-Transformation

$$P = \kappa \frac{z-1}{z+1} \quad \text{mit} \quad \kappa = \cot\left(\frac{\pi B_{3dB}}{2 f_a}\right) \tag{B.14}$$

und die Multiplikation mit der Grundverstärkung zwei, die der Haupttext formal benötigt, überführen $H^N(P)$ in die z-Systemfunktion

$$\tilde{H}(z) = 2 \cdot c \cdot \prod_{i=1}^{n_1} \frac{d_{1,i} z^{-1} + d_{0,i}}{g_{1,i} z^{-1} + g_{0,i}} \cdot \prod_{i=n_1+1}^{n_1+n_2} \frac{d_{2,i} z^{-2} + d_{1,i} z^{-1} + d_{0,i}}{g_{2,i} z^{-2} + g_{1,i} z^{-1} + g_{0,i}}, \tag{B.15}$$

wobei die Koeffizienten für $i \leq n_1$

$$\begin{aligned} d_{0,i} &= 1, & g_{0,i} &= \kappa - \alpha_i, \\ d_{1,i} &= 1, & g_{1,i} &= -\kappa - \alpha_i \end{aligned} \tag{B.16}$$

und für $i > n_1$

$$\begin{aligned} d_{0,i} &= 1, & g_{0,i} &= (\kappa - \alpha_i)^2 + \beta_i^2, \\ d_{1,i} &= 2, & g_{1,i} &= 2(\alpha_i^2 + \beta_i^2 - \kappa^2), \\ d_{2,i} &= 1, & g_{2,i} &= (\kappa + \alpha_i)^2 + \beta_i^2 \end{aligned} \tag{B.17}$$

lauten. Bis auf den Faktor $2c$ läßt sich das Gesamtsystem durch n_1 und n_2 aneinandergereihte Teilsektionen ersten beziehungsweise zweiten Grades realisieren. Ohne Rücksicht auf besondere Rechengeschwindigkeit kann man bei ausreichender Rechengenauigkeit direkt folgende Rekursionsgleichungen verwenden. Dabei sind mit $x_i(n)$ und $y_i(n)$ die Zahlenfolgen am Eingang und Ausgang einer Sektion gemeint:

$$y_i(n) = \frac{d_{0,i}x_i(n) + d_{1,i}x_i(n-1) - g_{1,i}y_i(n-1)}{g_{0,i}} \qquad \text{bzw.} \tag{B.18}$$

$$y_i(n) = \frac{d_{0,i}x_i(n) + d_{1,i}x_i(n-1) + d_{2,i}x_i(n-2) - g_{1,i}y_i(n-1) - g_{2,i}y_i(n-2)}{g_{0,i}}. \tag{B.19}$$

Der Frequenzgang von $\tilde{H}(z)$ wird mit Hilfe der Ersetzung $z = e^{j\omega/f_a}$ ausgewertet. Allgemein bewirkt die Bilinear-Transformation, daß dieser nur bei der 3dB-Grenzfrequenz ω_{3dB} und bei null exakt mit dem Frequenzgang des entnormierten zeitkontinuierlichen Systems übereinstimmt. Dazwischen fällt die Dämpfung für $\tilde{H}(z)$ minimal schwächer, darüber zunehmend stärker aus, bis bereits bei der halben Abtastrate unendlich erreicht wird. Alle Abweichungen sinken mit wachsendem κ, also bei größerer Abtastrate f_a oder kleinerer Bandbreite B_{3dB}. Sie sind bei den verwendeten Kombinationen von Bandbreite, Filtertyp und Abtastrate ohne praktische Bedeutung. Meist wird vor Beginn größerer Abweichungen schon eine Dämpfung erzielt, die die interessierende Signaldynamik erreicht.

Beim Tiefpaß P1 steigt allerdings die Dämpfung des kontinuierlichen Systems sehr schwach an. Deshalb weist das diskrete System in Frequenzgängen, die mit sinnvoller Dynamik aufgetragen werden, zur halben Abtastrate hin eine erheblich höhere Dämpfung auf. Wenn statt der Bilinear-Transformation die Impulsantwort-Invarianten-Methode zum Rekursionsentwurf verwendet wird, fällt sie übrigens geringer als beim kontinuierlichen System aus. Auf diese Methode läuft auch das ursprünglich vorgeschlagene Rekursionsverfahren der FTT nach [Ter85, Hei88a] hinaus. In der Praxis des TTZM-Verfahrens bleiben alle diese Unterschiede ohne Bedeutung. Einzig Bild 2.3 wurde nach dem ursprünglichen Verfahren berechnet, für welches Gl. (2.15) auch in abgetasteter Form exakt gültig bleibt. Die durch die Bilinear-Transformation verbesserte Selektivität würde sonst bewirken, daß sich beim Abschalten bereits spektrale Verbreiterungen abzeichnen können (vgl. Abschnitt 3.3.5).

B.4 Stufenlose Realisierung von Laufzeiten

Laufzeiten als ganzzahlige Vielfache des Abtastabstandes sind in diskreten Systemen leicht durch Verzögerung der Abtastwerte realisierbar. Für den Laufzeitausgleich frequenzgruppenabhängiger Analyse- und Synthesefenster erweist sich dieses Stufung aber als zu grob. Da die benötigte Laufzeit zu höheren Analyse- oder Synthesefrequenzen hin zunimmt, würde die realisierte Laufzeit zwischen benachbarten Frequenzen öfter springen.

Tabelle B.1: *Koeffizienten von zeitkontinuierlichen Tiefpässen, normiert auf 3dB–Grenzkreisfrequenz und Grundverstärkung eins, passend zu Gl. (B.12). Angaben für einfachen Pol* (P1)*, n–fachen Pol (*n*P1), Pseudo–Gauß–* (PG*n*)*, Bessel–* (B*n*) *und Potenz–Tiefpaß* (P*n*)*. Quelle* PG*n [Dis59], übrige [Her84]. Scheitelpunkt* (h^N_{max}, T_{max}) *der Impulsantwort ermittelt durch Simulation, Gruppenlaufzeit* T_g *bei Frequenz null durch Gl. (3.7).*

Kürzel	i	α_i	β_i	c	h^N_{max}	T_{max}	T_{max}/T_g
P1	1	–1,00000		1,000e–0	1,000	0,000	0,000
2P1	1–2	–1,55377		4,142e–1	0,572	0,644	0,500
3P1	1–3	–1,96146		1,325e–1	0,531	1,020	0,667
4P1	1–4	–2,29896		3,580e–2	0,515	1,305	0,750
8P1	1–8	–3,32400		6,710e–5	0,495	2,106	0,875
PG2	1	–1,28963	0,53418	5,132e–1	0,541	0,736	0,556
PG3	1	–1,38257	0,94621	2,344e–1	0,496	1,233	0,751
	2	–1,52005					
PG4	1	–1,42106	1,27582	9,740e–2	0,480	1,627	0,840
	2	–1,63074	0,39458				
PG8	1	–1,49815	2,27978	1,134e–3	0,474	2,682	0,944
	2	–1,83307	1,48938				
	3	–2,01063	0,85597				
	4	–2,09049	0,28007				
B2	1	–1,10160	0,63601	6,180e–1	0,514	0,824	0,605
B3	1	–1,04741	0,99926	3,608e–1	0,465	1,429	0,814
	2	–1,32268					
B4	1	–1,37001	0,41025	1,902e–1	0,450	1,904	0,901
	2	–0,99521	1,25711				
B8	1	–1,75741	0,27287	5,154e–3	0,454	3,134	0,986
	2	–1,63694	0,82280				
	3	–1,37384	1,38836				
	4	–0,89287	1,99833				
P2	1	–0,70710	0,70710	1,000e–0	0,456	1,111	0,786
P3	1	–0,50000	0,86603	1,000e–0	0,404	2,055	1,453
	2	–1,00000					
P4	1	–0,92388	0,38268	1,000e–0	0,382	2,895	1,108
	2	–0,38268	0,92388				
P8	1	–0,98079	0,19509	1,000e–0	0,337	5,892	1,149
	2	–0,83147	0,55557				
	3	–0,55557	0,83147				
	4	–0,19509	0,98079				

Bei der FTT-Rücktransformation und bei Rekonstruktion aus Zeitkonturen träten dadurch Kammfilterverzerrungen des Frequenzganges auf, bei Rekonstruktion aus Frequenzkonturen zeitliche Modulationseffekte. Weil sich die Fensterwirkung an einer Frequenz als Tiefpaßsignal darstellen läßt, kann ein Allpaß zum Feinabgleich der Laufzeit verwendet werden.

Die gewünschte Laufzeit t_L wird in ein ganzzahliges Vielfaches n_L des Abtastabstandes $t_a = 1/f_a$ und in einen Rest ν aufgeteilt. Mit Hilfe der int-Funktion, die den Rückgriff auf den ganzzahligen Wert bezeichnet, läßt sich schreiben:

$$t_L = (n_L + \nu)t_a, \quad \text{mit} \quad n_L = \text{int}\left(\frac{t_L}{t_a}\right). \tag{B.20}$$

Man verzögert das zu verarbeitende Signal um n_L Abtastwerte und führt es danach über einen zeitdiskreten Allpaß ersten Grades $\tilde{A}(z)$ mit frequenzunabhängiger Verstärkung eins. Er entspricht einer Teilsektion ersten Grades in Gl. (B.15) [Hes93]. Mit Hilfe des Polradius

$$r = \frac{1-\nu}{1+\nu} \tag{B.21}$$

lauten die einzustellenden Koeffizienten

$$\begin{aligned} d_0 &= r, & g_0 &= 1, \\ d_1 &= 1, & g_1 &= r. \end{aligned} \tag{B.22}$$

Das Pol/Nullstellen-Paar von $\tilde{A}(z)$ liegt auf der negative reellen Achse der z-Ebene. Seine Gruppenlaufzeit

$$\tau_g(\omega) = t_a \cdot \frac{1-r^2}{1+2r\cos(\omega t_a)+r^2} \tag{B.23}$$

weist deshalb ein Gruppenlaufzeitmaximum bei der halben Abtastrate auf, das bei $\nu \to 0$ zum Dirac-Impuls ausartet, wenn Pol- und Nullstelle sich auf dem Einheitskreis treffen. Eine frequenzabhängige Laufzeit ist natürlich unerwünscht. Entscheidend aber ist das Verhalten bei tiefen Frequenzen, auf die der Analyse- oder Synthesetiefpaß das zu verzögernde Signal beschränkt. Dort verläuft die Gruppenlaufzeit sehr eben auf dem gewünschten Niveau von νt_a.

Der Fehler dieser Feineinstellung dürfte ausreichend klein sein: Bei höheren Analyse- bzw. Synthesefrequenzen kann man die Frequenzgruppe oder 1 Bark nach Gl. (1.2) mit etwa 20% von dieser Frequenz ansetzen, schlimmstenfalls also mit 10% der Abtastrate f_a. Durch die Verarbeitung als Tiefpaßsignal nimmt ein Band von 1 Bark ein Basisband von ±5% der Abtastrate ein. Darin liegt die maximale absolute Abweichung der Gruppenlaufzeit nach Gl. (B.23) von νt_a bei etwa $+0{,}01t_a$. Bei 3 Bark sind es schlimmstenfalls $+0{,}1t_a$.

B.5 Kontur/Textur-Analyse

Die Operationen der Erweiterung zur Kontur/Texturanalyse nach Bild 4.3 werden hier in Gestalt von Signalgleichungen vorgestellt. Um die Übersichtlichkeit zu wahren, sind die Operationen APF/APZ, WZS, ZG mit hebbarer Akausalität definiert, passend zu ihrer Verschaltung in Bild 4.3. Zur Behebung muß man hier wie dort zeitverzögerte Signale und zusätzliche Laufzeitglieder einführen. Auf diese Weise lassen sich einerseits die real benötigten Laufzeiten t_L^P, t_L^W bzw. t_L^{ZG} (s.u.) berücksichtigen und andererseits die Gleichzeitigkeit der verschiedenen Signalwege sicherstellen.

Das in den Operationen verwendete normierte Tiefpaßfilter $H^N(\Omega)$ bzw. $h^N(T)$, mit Scheitelpunkt (T_{max}, h^N_{max}), entspricht genau dem Typ, der die Fensterfunktion der FTT festlegt. Zwar kann ZG grundsätzlich einen eigenständigen Typ verwenden, doch hat sich bei der praxisrelevanten Fensterfunktion 4P1 die Identität bewährt. WFS und WZS übernehmen aus der FTT die 3dB-Bandbreite B_{3dB}. WZS und ZG lassen sich in gleicher Weise wie die laufzeitausgeglichenen FTT-Analysetiefpässe implementieren. In den folgenden Gleichungen bleibt das hebbare Problem des Logarithmusargumentes null unberücksichtigt.

Auswahl prägnanter Frequenz- bzw. Zeitkonturen (APF/APZ): Die Operation enthält jeweils zwei Schritte, die hier für die Frequenzkonturen dargestellt sind. Der Buchstabentausch nach Anhang A.2 (Kontursignalzeit '(t)' ausgenommen) liefert die korrespondierende Beschreibung für die Zeitkonturen. Der erste Schritt ordnet die Punkte des Frequenzkontursignals $\mathcal{C}_F(t)$ in Kontursignale $\mathcal{C}_F^i(t)$ von Linien um (Anhang A.2). Der zweite Schritt wählt die Punkte derjenigen Linien aus, deren Länge Δt_F^i mindestens die Prägnanzschwelle Δt_P erreicht (Δf_P bei Zeitkonturen). Eine kausale Liniensuche würde deshalb eine Vorlaufzeit $t_L^P \geq \Delta t_P$ benötigen. Bei praxisrelevanten Parametern reicht $t_L^P = \Delta t_P$ leicht aus, um parallel auch eine kausale Zeitkonturliniensuche zu erlauben. Die so erfaßte Konturpunktmenge ergibt das Signal der prägnanten Frequenzkonturen

$$\mathcal{C}_{PF}(t) = \bigcup_{\Delta t_F^i \geq \Delta t_P} \mathcal{C}_F^i(t), \quad \text{wobei} \quad \bigcup_i \mathcal{C}_F^i(t) = \mathcal{C}_F(t). \tag{B.24}$$

Wandlung prägnanter Frequenzkonturen ins Spektrum (WFS): Als Spektrum dient die Summe der Selektionsspannungsbeträge, welche stationäre Sinustöne in einem FTT-Analysefilter an der Frequenz f hervorrufen würden. Die gedachten Sinustöne sind durch die Pegel und Frequenzen der Frequenzkonturpunkte zum Zeitpunkt t vorgegeben:

$$L_{PF}(f,t) = 20\,\mathrm{lg}\left[\sum_{\mathcal{C}_{PF}(t)} \left| H^N\left(\frac{2(f-f_{PF})}{B_{3dB}(f)}\right)\right| \cdot 10^{L_{PF}/20\mathrm{dB}}\right] \mathrm{dB}. \tag{B.25}$$

Wandlung prägnanter Zeitkonturen ins Spektrum (WZS): Benötigt wird die Menge aller Zeitkonturpunkte an der Frequenz f bis heran zum aktuellen Zeitpunkt t, welche aus dem Konturpunktsignal $\mathcal{C}_{PZ}(f,t)$ und damit aus dem Zeitkontursignal hervorgeht (Anhang A.3). Jedem dieser Punkte wird eine Impulsantwort des FTT-Analysefilters bei f zugeordnet. Sie ist jeweils so zu skalieren, daß ihr Scheitelwert exakt mit dem vom Konturpunkt vorgegebene Pegel übereinstimmt. Der Zeitverlauf aller dieser Impulsantworten wird überlagert und dient als Spektrum bei f:

$$L_{PZ}(f,t) = 20\,\mathrm{lg}\left|\sum_{\bigcup_{x\leq t} \mathcal{C}_{PZ}(f,x)} \frac{h^N\left(\pi B_{3dB}(f)\cdot(t-t_{PZ})+T_{max}\right)}{h_{max}^N} \cdot 10^{L_{PZ}/20\mathrm{dB}}\right| \mathrm{dB}. \tag{B.26}$$

Die Laufzeit zur Realisierung beträgt in Anlehnung an Gl. (3.11) $t_L^W = T_{max}\cdot(\pi B_{3dB}(0))^{-1}$. Auf einen bei f realisierten Tiefpaß, nach Art der Analysetiefpässe mit Grundverstärkung zwei, muß praktisch für jeden Punkt zeitrichtig ein gewichteter Impuls $k\cdot\delta(t)$ gegeben werden (bzw. $k\cdot f_a\cdot\delta(n)$ in zeitdiskreter Formulierung), wobei

$$k = \frac{10^{L_{PZ}/20\mathrm{dB}}}{2\cdot\pi B_{3dB}(f)\cdot h_{max}^N}. \tag{B.27}$$

Leistungssubtraktion der Spektren (SUB): Vom laufzeitausgeglichenen Pegelspektrum $L^L(f,t)$, auf dem die Konturen gesucht wurden, erfolgt in der Leistung ein Abzug der rückgerechneten Spektren prägnanter Konturen. Negativwerte sind nicht zugelassen. Die Pegel der rückgerechneten Spektren werden zuvor um ΔL_{PF} bzw. ΔL_{PZ} angehoben.

$$L'(f,t) = 10\,\mathrm{lg}\left\{\begin{array}{l} p\left(\left(L^L(f,t)\right) - p\left(\left(L_{PF}(f,t)+\Delta L_{PF}\right) - p\left(\left(L_{PZ}(f,t)+\Delta L_{PZ}\right),\right.\right.\right. \\ +0, \quad \text{falls obiges Resultat nicht positiv, worin } \; p(L) = 10^{L/10\mathrm{dB}} \end{array}\right\} \mathrm{dB}. \tag{B.28}$$

Spektrale Glättung des Spektrums (FG): Das Leistungsspektrum von $L'(f,t)$ wird entlang der Frequenzachse mit dem Gauß-förmigen Kern $K^{FG}(B,f)$ gefaltet, der die Halbwertsbreite $B = B^{FG}_{3dB}(f)$ und die Fläche eins aufweist. So erreicht man, daß ein schmalbandiges Spektrum die 3dB-Mindestbreite $B^{FG}_{3dB}(f)$ erhält, umgekehrt aber ein flaches Spektrum fast unverändert bleibt. Der Frequenzbereich zwischen f_u, f_o markiert in der Praxis den Bereich der Analysefrequenzen ω_{A_i}, über die das Integral mit $dx = \Delta\omega_A(\omega_{A_i} \cdot (2\pi)^{-1}) \cdot (2\pi)^{-1}$ als Summe ausgeführt wird.

$$L''(f,t) = 10\lg\left[\int_{f_u}^{f_o} K^{FG}\left(B^{FG}_{3dB}(f), f-x\right) \cdot 10^{L'(x,t)/10\mathrm{dB}} dx\right] \mathrm{dB}, \quad \text{(B.29)}$$

$$K^{FG}(B,f) = \frac{2}{B}\sqrt{\frac{\ln 2}{\pi}} \cdot \exp\left[-\ln 2\left(\frac{2f}{B}\right)^2\right]. \quad \text{(B.30)}$$

Zeitliche Glättung des Spektrums (ZG): Der Zeitverlauf des Leistungsspektrums von $L''(f,t)$ an einer Frequenz f durchläuft schließlich ein Tiefpaßfilter mit der frequenzunabhängigen 3dB-Bandbreite B^{ZG}_{3dB} und der Grundverstärkung eins. Die Realisierung würde die Laufzeit $t^{ZG}_L = T_{max} \cdot (\pi B^{ZG}_{3dB})^{-1}$ einführen. Akausal formuliert lautet die resultierende Texturhüllfläche

$$L_{TX}(f,t) = 10\lg\left|\int_0^t \pi B^{ZG}_{3dB} \cdot h^N\left(\pi B^{ZG}_{3dB} \cdot (t-x) + T_{max}\right) \cdot 10^{L''(f,x)/10\mathrm{dB}} dx\right| \mathrm{dB}. \quad \text{(B.31)}$$

Gesamtlaufzeit der realisierten Kontur/Textur-Analyse: In Bild 4.3 bestimmt der Signalpfad über die laufzeitbehafteten Operationen FTT, APF, WZS und ZG die Gesamtlaufzeit t^{KTX}_L einer Realisierung. Mit der Grundlaufzeit $t^A_L = T_{max} \cdot (\pi B_{3dB}(0))^{-1}$ der FTT nach Gl. (3.11) kommt man auf

$$t^{KTX}_L = t^A_L + t^P_L + t^W_L + t^{ZG}_L \quad \text{(B.32)}$$

$$= t^P_L + \frac{T_{max}}{\pi}\left(\frac{2}{B_{3dB}} + \frac{1}{B^{ZG}_{3dB}(0)}\right). \quad \text{(B.33)}$$

Durch die Wahl der Parameter der Konturanalyse sind automatisch auch die meisten Parameter der Kontur/Textur-Erweiterung festgelegt. Vorzugeben bleiben noch Δt_P, Δf_P, ΔL_{PF}, ΔL_{PZ}, B^{FG}_{3dB}, B^{ZG}_{3dB} (s.o.) sowie Δf_U, Δt_U (Anhang A.2).

B.6 Rekonstruktionsverfahren

Die folgenden Beschreibungen beziehen sich auf die in Bild 5.3a-c dargestellten Operationen, die im Haupttext nicht präzisiert sind.

Konturgesteuerte Siebe (FS/ZS): Die konturgesiebten Abtastwerte $s_F(\omega_{S_m}, lT_S)$ bzw. $s_Z(\omega_{S_m}, lT_S)$ ergeben sich aus den Abtastwerten $s^{LB}(\omega_{S_m}, lT_S)$ des FTT-Bandpaßspektrums $s^{LB}(\omega,t)$. Gesteuert werden die Vorgänge vom Frequenzkontursignal $\mathcal{C}_F(t)$ bzw. dem Zeitkontursignal $\mathcal{C}_Z(t)$ (Anhang A.3), die jeweils vom Pegelspektrum von $s^{LB}(\omega,t)$ im Zeit/Frequenz-Kontinuum abstammen. In der Praxis verbergen sich dahinter codierte Konturen, die unter Annahme von Mindestlängen als Linienverläufe interpretiert wurden

(siehe Haupttext). Die in Bild 5.4 verdeutlichten Kontrollstreckenenden sind als arithmetische Mitten der Stützfrequenzen $\{\omega_{S_m}\}$ bzw. der Stützzeiten $\{lT_S\}$ definiert. Die Sonderfälle unter der untersten bzw. über der obersten Stützfrequenz bleiben nachfolgend unbehandelt, $s_R(\omega_{S_m}, lT_S)$ dient hier nur als Hilfswert:

$$s_F(\omega_{S_m}, lT_S) = \begin{cases} s_R(\omega_{S_m}, lT_S), & \text{falls} \quad (L^L(f, lT_S), f, lT_S) \in \mathcal{C}_F(lT_S) \\ & \text{existiert mit} \quad \omega_{S_{m-1}} + \omega_{S_m} \leq 4\pi f < \omega_{S_m} + \omega_{S_{m+1}}, \\ 0 & \text{sonst,} \end{cases} \tag{B.34}$$

$$s_Z(\omega_{S_m}, lT_S) = \begin{cases} s_R(\omega_{S_m}, lT_S), & \text{falls} \quad (L^L(\frac{\omega_{S_m}}{2\pi}, t), \frac{\omega_{S_m}}{2\pi}, t) \in \mathcal{C}_Z(t) \\ & \text{existiert mit} \quad l - \frac{1}{2} \leq t/T_S < l + \frac{1}{2}, \\ 0 & \text{sonst,} \end{cases} \tag{B.35}$$

$$\text{mit} \qquad s_R(\omega_{S_m}, lT_S) = s^{LB}(\omega_{S_m}, lT_S). \tag{B.36}$$

'gegenseitige' Maskierung (MSK): Behelfsweise bleiben die Frequenzkontur-gesiebten Abtastwerte unverändert $s_{MF}(\omega_{S_m}, lT_S) = s_F(\omega_{S_m}, lT_S)$, es findet keine Maskierung durch die Zeitkontur-gesiebten Abtastwerte statt. Von den Frequenzkontur-gesiebten Abtastwerten $s_F(\omega_{S_m}, lT_S)$ wird auf das stationäre FTT-Betragsspektrum $s_M(\omega, lT_S)$ zurückgerechnet, ähnlich wie bei der Operation WFS der Kontur/Textur-Analyse oben. Dazu werden die normierte Systemfunktion $H^{AN}(\Omega)$ des Analysetiefpasses und die Analysebandbreite B^A_{3dB} benötigt. Die Zeitkontur-gesiebten Abtastwerte $s_Z(\omega_{S_m}, lT_S)$ werden an den Orten zu null gesetzt ('maskiert'), wo sie kleiner als das um $\Delta L_M = 1$ dB angehobene stationäre Betragsspektrum sind. Die Anhebung gewährleistet sichere Maskierung in der unmittelbaren Frequenzumgebung von Frequenzkonturen. Die Zeitkontur-gesiebten Abtastwerte nach Maskierung lauten

$$s_{MZ}(\omega_{S_m}, lT_S) = \begin{cases} s_Z(\omega_{S_m}, lT_S), & \text{falls} \quad |s_Z(\omega_{S_m}, lT_S)| > s_M(\omega_S, lT_S) \cdot 10^{\Delta L_M/20\text{dB}}, \\ 0 & \text{sonst, wobei} \end{cases} \tag{B.37}$$

$$s_M(\omega, lT_S) = \sum_{\substack{\omega_S = \omega_{S_m} \\ m}} \left| s_F(\omega_S, lT_S) \cdot H^{AN}\left(\frac{\omega - \omega_S}{\pi B^A_{3dB}(\omega)}\right)\right|. \tag{B.38}$$

Bewertete Überlagerung: Aus den Werten $s_{MF}(\omega_{S_m}, lT_S)$ und $s_{MZ}(\omega_{S_m}, lT_S)$, die mit Hilfe der Frequenz- bzw. Zeitkonturen und gegenseitiger Maskierung gewonnen wurden, errechnen sich die Abtastwerte des rückzutransformierenden FTT-Bandpaßspektrums

$$\hat{s}^{LB}(\omega_{S_m}, lT_S) = c_F(\omega_{S_m}) \cdot s_{MF}(\omega_{S_m}, lT_S) + c_Z(\omega_{S_m}) \cdot s_{MZ}(\omega_{S_m}, lT_S). \tag{B.39}$$

Frequenzabhängige Bewertungsfaktoren sind

$$c_F(\omega_S) = \frac{1}{|H^A_{\omega_S}(0)|} \cdot \frac{1}{|H^S_{\omega_S}(0)|} \cdot \frac{2\pi h^{A*S}_{max,\omega_S}}{\Delta\omega_S(\omega_S)} \tag{B.40}$$

$$\approx \frac{\pi h^A_{max,\omega_S}}{\Delta\omega_S(\omega_S)}, \tag{B.41}$$

$$c_Z(\omega_S) = \frac{1}{h^A_{max,\omega_S}} \cdot \frac{1}{T_S} \cdot \frac{2\pi h^{A*S}_{max,\omega_S}}{\Delta\omega_S(\omega_S)} \cdot \frac{\Delta\omega_S(\omega_S)}{2\pi h^S_{max,\omega_S}} \tag{B.42}$$

$$= \frac{h^{A*S}_{max,\omega_S}}{h^{A}_{max,\omega_S} \cdot h^{S}_{max,\omega_S} \cdot T_S} \tag{B.43}$$

$$\approx \frac{2}{h^{S}_{max,\omega_S} \cdot T_S}. \tag{B.44}$$

Die Näherungen ergeben sich aus der rechten Hälfte von Gl. (5.21) sowie aus der Entnormierungsvorschrift (3.4), die bei $H^N(0) = 1$ auf den Tiefpaß des Analyse- und Synthesefilters angewendet $H^A_{\omega_S}(0) = H^S_{\omega_S}(0) = 2$ bedeutet. Zur folgenden Herleitung der Bewertungsfaktoren gilt, ohne Beschränkung der Allgemeinheit, identisches Analyse- und Syntheseraster $\{(\omega_{A_i}, kT_A)\} = \{(\omega_{S_m}, lT_S)\}$.

Ein Sinuston führt im eingeschwungenen Zustand des FTT-Spektrums zu einer in Frequenz und Pegel stationären Frequenzkonturlinie. Die Frequenzkontur-gesteuerte Siebfunktion für die FTT-Abtastwerte im oberen Zweig von Bild 5.3a bewirkt im Schema der Transformationscodierung in Bild 5.2d, daß nur ein Kanal mit passender Frequenz durchgeschaltet wird, also etwa der in Bild 5.2c vereinfacht dargestellte. Da nur dieser allein zum Ausgangssignal beitragen kann, muß er im übergeordneten Schema in Bild 5.3a durch den Bewertungsfaktor $c_F(\omega_S)$ auf die Verstärkung eins gebracht werden, und zwar unabhängig von der Sinustonfrequenz bzw. der daraus folgenden Kanalfrequenz. Für die Verstärkung des Kanals in Bild 5.2c sind die Grundverstärkungen von Analyse- und Synthesefilter sowie der gebrochene Vorfaktor des letzten Modulators maßgeblich. Die zugehörigen drei Kehrwerte finden sich der Reihe nach in Gl. (B.40) wieder, so daß sich für einen Sinuston in Bild 5.3a frequenzunabhängig die gewünschte Verstärkung einstellt. Nebeneffekte durch Abweichung der Sinustonfrequenz von vorhandenen Kanalfrequenzen oder durch Selektion der negativen Teilschwingung sind dabei vernachlässigt.

Ein Dirac-Impuls löst eine geschlossene Zeitkonturlinie aus, deren Punkte wegen des Laufzeitausgleiches alle gleichzeitig auftreten. Das Zeitkontur-gesteuerte Sieb im unteren Zweig von Bild 5.3a läßt somit an allen Analysefrequenzen die FTT-Abtastwerte des zugehörigen Abtastzeitpunktes passieren. Je nach Feinheit des Zeitraster haben diese annähernd den Betrag h^A_{max,ω_A}, nämlich die maximale Höhe der Impulsantwort des Analysetiefpasses. Gleichzeitig vorhandene Frequenzkonturen (Abschnitt 2.1.1) können vernachlässigt werden, so daß die Operation MSK in Bild 5.3a ohne Einfluß bleibt. Der Bewertungsfaktor $c_Z(\omega_S)$ gewährleistet, daß die Rücktransformation R aus den Abtastwerten wieder einen - nunmehr im Nutzfrequenzbereich begrenzten - Dirac-Impuls ausgibt. Dazu normiert der erste Faktor in Gl. (B.42) die Abtastwerte frequenzunabhängig auf eins. Der zweite Faktor erreicht innerhalb von R in Bild 5.3a bzw. Bild 5.2d, daß an allen Synthesefrequenzen das Signal $\hat{s}^{LB,\delta}_{\omega_S}(t)$ als ein Dirac-Impuls mit Gewicht eins erscheint. Der dritte Faktor schließlich neutralisiert den Vorfaktor im letzten Modulator von R und installiert den vierten Faktor an seiner Stelle. Bezüglich des Dirac-Impulses als Eingangs- und $\hat{s}(t)$ als Ausgangssignal zeigt sich die effektiv dazwischenliegende Anordnung als FTT mit RFTT gemäß Abschnitt 5.1.1 und Bild 5.2a, nunmehr mit einer Fensterfunktion $h^S_{\omega_S}(t)$. Der Dirac-Impuls wird also innerhalb des Nutzfrequenzbereiches praktisch unverändert zum Ausgang gelangen.

Frequenz- und Zeitkontur-Rasterierer (FR/ZR): Auch hier entscheiden Gln. (B.34), (B.35), wann nach Maßgabe der Kontursignale $\mathcal{C}_F(t)$ und $\mathcal{C}_Z(t)$ die Werte $s_F(\omega_{S_m}, lT_S)$ bzw. $s_Z(\omega_{S_m}, lT_S)$ der gerasterten Konturen anders als null zu besetzen sind. Nun allerdings entspringt $s_R(\omega_{S_m}, lT_S)$ im allgemeinen Näherungswerten $\hat{L}^L$ und $\hat{\phi}^{LB}$ von Pe-

gelspektrum $L^L(f,t)$ bzw. Bandpaßphasenspektrum $\phi^{LB}(f,t)$ am Rasterort (ω_{S_m}, lT_S). Sie werden aus dem nächstliegenden Konturpunkt $(L_\times, f_\times, t_\times) \in \mathcal{C}(t_\times)$ und dem mitgelieferten Phasenpunkt $(\phi_\times, f_\times, t_\times) \in \Phi\{\mathcal{C}(t_\times)\}$ bestimmt. In der Praxis ist das der nächstliegende Punkt der codierten Kontur. Man erhält:

$$s_R(\omega_{S_m}, lT_S) = 10^{\hat{L}^L\left(\frac{\omega_{S_m}}{2\pi}, lT_S\right)/20\text{dB}} \cdot e^{j\hat{\phi}^{LB}\left(\frac{\omega_{S_m}}{2\pi}, lT_S\right)}, \tag{B.45}$$

$$\hat{L}^L\left(\frac{\omega_{S_m}}{2\pi}, lT_S\right) = L_\times, \tag{B.46}$$

$$\hat{\phi}^{LB}\left(\frac{\omega_{S_m}}{2\pi}, lT_S\right) = \phi_\times - \omega_{S_m} \cdot (lT_S - t_\times). \tag{B.47}$$

Phasenrekonstruktion (PRK): Nachfolgend wird eine formale Grundlage für optimale wie für behelfsmäßige, suboptimale Verfahren geliefert. Das in RKHP eingesetzte, behelfsmäßige Verfahren behandelt die Phasen von Zeit- und Frequenzkonturen unabhängig voneinander. Es benötigt die Phasendriftfunktionen $\psi_F^i(t)$ und $\psi_Z^k(f)$ nicht, deren Werte dann als null anzunehmen sind. Diese sind erst für die im Haupttext skizzierte, optimale Phasenrekonstruktion von Bedeutung, bei der sie sich geringstmöglich verändern, um zusätzliche Nebenbedingungen für die rekonstruierte Phase erfüllbar zu machen. Anders als bei der behelfsmäßigen Phasenrekonstruktion dimensioniert man die Grenzen Δf_Φ und Δt_Φ der Phasenübergabe dann wesentlich kleiner.

Sei $f^i(t)$ eine Frequenzfunktion, die die offene Frequenzkonturlinie $\mathcal{C}_F^i$ mit Verlaufsunstetigkeiten von maximal $\Delta f_U = \Delta f_\Phi$ auf dem Zeitintervall zwischen t_B^i und $t_E^i \geq t_B^i$ beschreibt (Anhang A.2). Ihr sind über demselben Zeitintervall die Funktion $\hat{\phi}_F^i(t)$ der zu rekonstruierenden Phase und eine Phasendriftfunktion $\psi_F^i(t)$ zugeordnet. Funktionswerte an einer möglicherweise offenen Intervallgrenze t_B^i meinen den linksseitigen Grenzwert. Für eine andere Frequenzkonturlinie $f^m(t)$ gelten alle diese Festlegungen entsprechend. Aus dem zur Zeit t rekonstruierten Phasenwert ergibt sich ein benachbarter im geringen Abstand $dt \geq 0$ durch

$$\hat{\phi}_F^i(t+dt) = \hat{\phi}_F^i(t) + 2\pi f^i(t) \cdot dt + \psi_F^i(t+dt) - \psi_F^i(t). \tag{B.48}$$

Die Vorschrift für den Startwert der Phase am Linienanfang t_B^i lautet

$$\hat{\phi}_F^i(t_B^i) = \lim_{\substack{\epsilon \to 0 \\ \epsilon > 0}} \begin{cases} \hat{\phi}_F^m(t_B^i - \epsilon) + 2\pi f^m(t_B^i - \epsilon) \cdot \epsilon, \\ \qquad \text{falls minimales } |f^m(t_B^i - \epsilon) - f^i(t_B^i)| \leq \Delta f_\Phi \\ \qquad \text{mit minimalem } f^m(t_B^i - \epsilon) \text{ existiert,} \\ \psi_F^i(t_B^i) \quad \text{sonst.} \end{cases} \tag{B.49}$$

Wenn also innerhalb von $\pm\Delta f_\Phi$ eine andere Linie $f^m(t)$ vorbeiläuft oder gerade endet, dann wird deren Phasenwert unmittelbar vor t_B^i als Startwert übernommen. Andernfalls ist der Startwert identisch mit dem der Phasendriftfunktion. Die Phasenübergabe zwischen verschieden Linien ist formal nur bei Linienaufspaltungen nötig, da eine offene Linie definitionsgemäß Frequenzunstetigkeiten von bis zu $\pm\Delta f_\Phi$ zuläßt. Die Minimalitätsforderungen in der zweiten und dritten Zeile erreichen, daß die Phase bei mehreren Kandidaten nur von der frequenznächsten und vorzugsweise frequenztieferen Linie übernommen werden darf. Der nach dem Pluszeichen folgende Term ist zwar im Zeitkontinuum wegen $\epsilon \to 0$ bedeutungslos. Bei Diskretisierung allerdings entfällt der Grenzübergang und ϵ

nimmt fest die Größe eines Zeitschritts an. Dann muß der Phasenwert vom vorigen auf den aktuellen Abtastzeitpunkt umgerechnet werden.

Die Phasenrekonstruktion für die Zeitkonturen geschieht in analoger Weise. Bei den offenen Zeitkonturlinien $\mathcal{C}_Z^k$ und $\mathcal{C}_Z^l$ mit Verlaufsunstetigkeiten von maximal $\Delta t_U = \Delta t_\Phi$ kann man die Zeit als Funktionen $t^k(f)$ bzw. $t^l(f)$ über der Frequenz darstellen. Die Funktionen der zu rekonstruierenden Phase und die Phasendriftfunktion heißen hier $\hat{\phi}_Z^k(t)$, $\hat{\phi}_Z^l(t)$ bzw. $\psi_Z^k(t)$. Phasenfortschreibung bei $df \geq 0$ und die Startwertvorschrift lauten

$$\hat{\phi}_Z^k(f+df) \;=\; \hat{\phi}_Z^k(f) + 2\pi f \cdot [t^k(f+df) - t^k(f)] + \psi_Z^k(f+df) - \psi_Z^k(f), \tag{B.50}$$

$$\hat{\phi}_Z^k(f_B^k) \;=\; \lim_{\substack{\epsilon\to 0\\ \epsilon>0}} \begin{cases} \hat{\phi}_Z^l(f_B^k-\epsilon) + 2\pi\cdot(f_B^k-\epsilon)\cdot[t^k(f_B^k) - t^l(f_B^k-\epsilon)], \\ \qquad \text{falls minimales } |t^l(f_B^k-\epsilon) - t^k(f_B^k)| \leq \Delta t_\Phi \\ \qquad \text{mit minimalem } t^l(f_B^k-\epsilon) \text{ existiert}, \\ \psi_Z^k(f_B^k) \quad \text{sonst.} \end{cases} \tag{B.51}$$

Zeitkonturlinien müssen in der Praxis mit einer Zeittiefe von mehreren Zeitschritten erfaßt werden. Isoliert scheinende Liniensegmente können sich nämlich noch nach einer gewissen Zeit als zusammenhängend herausstellen (vgl. Anhang B.5, Operation APZ). Bei der Phasenrekonstruktion von Frequenzkonturen reicht dagegen eine Betrachtung der Konturpunkte aus aktuellem und vorigem Zeitschritt aus, sofern keine optimierte Phasenrekonstruktion geplant ist. Schließlich definieren die Formeln

$$\hat{\Phi}\left\{\mathcal{C}_F(t)\right\} \;=\; \bigcup_i \left\{ (\hat{\phi}_F^i(t), f^i(t), t) \,\middle|\, (L, f^i(t), t) \;\in\; \mathcal{C}_F^i(t) \right\}, \tag{B.52}$$

$$\hat{\Phi}\left\{\mathcal{C}_Z(t)\right\} \;=\; \bigcup_k \left\{ (\hat{\phi}_Z^k(f), f, t^k(f)) \,\middle|\, (L, f, t^k(f)) \;\in\; \mathcal{C}_Z^k(t) \right\} \tag{B.53}$$

die in Bild 5.3c benötigten Signale der rekonstruierten Frequenz- bzw. Zeitkonturphasen.

Anhang C

Herleitungen

C.1 FTT-Spektrum einer eingeschalteten komplexen Schwingung

Für eine $t = 0$ eingeschaltete komplexe Schwingung wird der Zeitverlauf $s_{\omega_A}(t)$ des bei ω_A einschwingenden FTT-Spektrums hergeleitet und abgeschätzt. Dazu werden Fensterfunktionen der Familie nP1 verwendet. Mit der Frequenz ω_T und der Amplitude $\frac{A}{2}$ lauten das Signal und seine Laplace-Transformierte:

$$s(t) = \begin{cases} \frac{A}{2} \cdot e^{j\omega_T t} & \text{für} \quad t \geq 0, \\ 0 & \text{sonst}, \end{cases} \tag{C.1}$$

$$S(p) = \frac{A}{2} \cdot \frac{1}{p - j\omega_T}. \tag{C.2}$$

Die Laplace-Transformierte $S_{\omega_A}(p)$ von $s_{\omega_A}(t)$ erhält man, indem in Gl. (1.9) zunächst $S(p)$ aus Gl. (C.2) eingesetzt wird. Nachdem man die Entnormierungsvorschrift (3.5) auf die Laplace-Transformierte der Fensterfamilie nP1 aus Gl. (3.8) angewendet hat, kann man auch $H_{\omega_A}(p)$ einsetzen und kommt auf

$$S_{\omega_A}(p) = A \cdot \frac{1}{p + j\Delta\omega} \cdot \left(\frac{a}{p + a}\right)^n, \qquad \text{mit} \qquad \Delta\omega = \omega_T - \omega_A, \quad a = \frac{\omega_{3dB}(\omega_A)}{\sqrt{2^{\frac{1}{n}} - 1}}. \tag{C.3}$$

Darin sind der Polradius [1] a und der Abstand $\Delta\omega$ von Signal- und Analysefrequenz Hilfsgrößen. Partialbruchzerlegung von $S_{\omega_A}(p)$ und Rücktransformation (Heavisidescher Entwicklungssatz, z.B. [Mar82]) ergibt schließlich

$$s_{\omega_A}(t) = \Xi + \Theta, \qquad \text{wobei} \tag{C.4}$$

$$\Xi = A \cdot \left(\frac{a}{j\Delta\omega + a}\right)^n \cdot e^{j\Delta\omega t} \tag{C.5}$$

$$= H_{\omega_A}(\Delta\omega) \cdot \frac{A}{2} \cdot e^{j\Delta\omega t}, \tag{C.6}$$

[1] Dieser ist für $n = 1$ identisch mit der Fensterkonstante a in Gl. (1.3).

$$\Theta = -A \sum_{i=1}^{n} \left(\frac{a}{j\Delta\omega + a} \right)^{n+1-i} \cdot \frac{(at)^{i-1}}{(i-1)!} \cdot e^{-at} \tag{C.7}$$

$$= -A \cdot \left(\frac{a}{j\Delta\omega + a} \right)^{n} \cdot R_{n-1} \cdot e^{-at}, \tag{C.8}$$

$$\text{mit} \quad R_i = 1 + \frac{j\Delta\omega + a}{n-i} t \cdot R_{i-1}, \qquad R_0 = 1.$$

Ein Blick auf die Argumente der Exponentialfunktionen verdeutlicht, daß Ξ den stationären und Θ den transienten Anteil verkörpert. Für die zweite Form von Ξ wurde rein formal die Fourier-Transformierte $H_{\omega_A}(\omega)$ eingesetzt, sie gilt im übrigen für beliebige Fensterfunktionen. Die zweite Form für Θ, welche die rekursiv definierte Zwischengröße R_i benutzt, bleibt im Gegensatz zur ersten auch für größere n numerisch robust. Für Θ kann man die Näherung

$$\hat{\Theta} = -A \cdot \frac{a}{j\Delta\omega + \omega_{3dB}} \cdot \frac{(at)^{n-1}}{(n-1)!} \cdot e^{-at} \tag{C.9}$$

$$= -\frac{A}{2} \cdot \frac{1}{j\Delta\omega + \omega_{3dB}} \cdot h_{\omega_A}(t) \tag{C.10}$$

ansetzen. Zur ersten Form gelangt man, indem in Gl. (C.7) alle Summenglieder außer $i = n$ weggelassen werden und außerdem im Nenner a durch ω_{3dB} ersetzt wird. Die zweite Form ergibt sich durch Entnormierung der Zeitfunktion aus Gl. (3.8) mittels Vorschrift (3.5) und Einsetzung in die erste Form, wobei a aus Gl. (C.3) verwendet wird.

Für $n = 1$ gilt $\Theta = \hat{\Theta}$, weil dann sowieso nur das eine Summenglied vorkommt und nach Gl. (C.3) a und ω_{3dB} identisch sind. Für $n > 1$ läßt sich zunächst nur folgendes sagen: Ein ausreichend großes $|\Delta\omega| \gg a$ kann bei gegebenen a, t und n immer sicherstellen, daß nur das Summenglied $i = n$ in Gl. (C.7) relevant bleibt. Denn dort kommt $\Delta\omega$ in der niedrigsten Potenz vor. Um dann zu Gl. (C.9) zu gelangen, kann man ohne Folgen a durch ω_{3dB} ersetzen, weil wegen Gl. (C.3) auch $|\Delta\omega| \gg a \geq \omega_{3dB}$ gilt. Andererseits führt $t \to 0$ bei gegebenem $\Delta\omega$, a und $n > 1$ auf einen erheblichen Fehler. Dann erhält man nämlich $\hat{\Theta} \to 0$ und $\Theta \to -\Xi$.

Eine brauchbare, quantitative Aussage über die Abweichung von $\hat{\Theta}$ gegenüber Θ liefert erst eine numerische Auswertung. Dazu benötigt man noch den Zeitpunkt der maximalen Höhe von $|\hat{\Theta}|$, der in Abschnitt 3.3.4 als Zeitpunkt der maximalen Fensteröffnung eingeführt wurde. Für die Fenster nP1 berechnet er sich zu $t_{max,\omega_A} = (n-1)/a$ und entspricht auch dem frühestmöglichen Zeitpunkt, zu dem $|s_{\omega_A}(t)|$ ein zeitliches Maximum erreichen kann. So beträgt für $n \leq 8$ und $t \geq t_{max,\omega_A}$ und $|\Delta\omega| \geq 4\omega_{3dB}$ der Pegelunterschied maximal etwa 1 dB und der Phasenunterschied maximal 20°. Dies gilt für Pegel von Θ bis herab zu 80 dB unterhalb des Pegelmaximums von Ξ. [2]

[2] Für das laufzeitausgeglichene Spektrum nach Abschnitt 3.3.4 wäre statt t_{max,ω_A} frequenzunabhängig $t \geq t_{max,0}$ anzusetzen.

C.2 Spektrale Grenzselektion der FTT

Weil die FTT-Analysebandbreite proportional zur Frequenzgruppe Δf_G eingestellt wird, erhält man oberhalb von 500 Hz ein annähernd konstantes Verhältnis von Bandbreite und Analysefrequenz. Dadurch entsteht ein besonderer Effekt, der sich bei einer stationären komplexen Schwingung so äußert: Das FTT-Pegelspektrum sinkt vom Hauptmaximum zu hohen Frequenzen hin einem Grenzwert entgegen.

Bei konstanter Analysebandbreite dagegen würde die Selektion in den Analysefiltern zu höheren Analysefrequenzen immer weiter abnehmen, weil der Abstand zur Schwingungsfrequenz zunimmt. Aber die gleichzeitig wachsende Analysebandbreite kann diesen Effekt bei größeren Frequenzabständen kompensieren, so daß ein Grenzwert entsteht.

Der charakteristische Pegelabstand von Hauptmaximum und Grenzwert wird hier als (spektrale) *Grenzselektion* bezeichnet. Sie ist abhängig vom Fenstertyp und vom Einstellfaktor b für die Analysebandbreite $B_{3dB} = b \cdot \Delta f_g$. Eine zu geringe Grenzselektion verdeckt gehörrelevante Vorgänge im FTT-Spektrum, so daß sie in der anschließenden Konturierung nicht mehr erkannt werden. Nachfolgend wird die Grenzselektion für die Familie nP1 behandelt.

Der Zeitverlauf des FTT-Spektrums von einer komplexen Schwingung $s(t)$ mit der Frequenz ω_T wurde in Anhang C.1 berechnet. Gl. (C.5) gibt seinen stationären Anteil Ξ an der Analysefrequenz ω_A an. Der Ansatz für die spektrale Grenzselektion lautet allgemein bzw. speziell für die Fensterfamilie nP1:

$$a_\infty = \lim_{\omega_A \to \infty} 20\mathrm{lg}\left(\frac{|\Xi|_{\Delta\omega=0}}{|\Xi|_{\Delta\omega=\omega_T-\omega_A}}\right)\ \mathrm{dB} \tag{C.11}$$

$$= \lim_{\omega_A \to \infty} -10n \cdot \lg\left[\left(\frac{\omega_T - \omega_A}{a}\right)^2 + 1\right]\ \mathrm{dB}. \tag{C.12}$$

Über Gl. (C.3) hängt der Polradius a von der Analysebandbreite $B_{3dB} = \frac{\omega_{3dB}}{2\pi}$ ab, die ihrerseits mit Hilfe von b in Bruchteilen der Frequenzgruppenbreite vorgegeben ist. Weil nur ein praxisnaher Grenzwert innerhalb des Hörbereiches benötigt wird, eignet sich Gl. (1.2) als Formel für die Frequenzgruppenbreite nicht besonders. Dafür kann man nach [Zwi82] bei Analysefrequenzen ω_A oberhalb $2\pi \cdot 500$ Hz die Faustformel $\Delta f_G \approx 0,2\frac{\omega_A}{2\pi}$ verwenden. All dies eingesetzt in Gl. (C.12) ergibt die Näherung

$$a_\infty \approx \lim_{\omega_A \to \infty} -10n \cdot \lg\left[\left(2^{\frac{1}{n}} - 1\right)\left(\frac{\omega_T - \omega_A}{0,1b \cdot \omega_A}\right)^2 + 1\right]\ \mathrm{dB} \tag{C.13}$$

$$\approx -10n \cdot \lg\left(\frac{2^{\frac{1}{n}} - 1}{0,01b^2} + 1\right)\ \mathrm{dB}. \tag{C.14}$$

Die Werte aus Gl. (C.14) für einige Kombinationen von b und n sind in Tabelle C.1 dargestellt. Demnach bewirken ein höherer Fensterfunktionsgrad oder eine geringere Analysebandbreite eine bessere Grenzselektion. Da das spektrale Selektionsverhalten bei gegebenem n über verschiedene Fensterfamilien bei den nP1 am schwächsten ausgeprägt ist (Abschnitt 3.3.3), fallen die Werte für die anderen Fensterfamilien noch kleiner aus. Für die Heinbachschen Standardparameter der FTT ($n = 1$, $b = 0{,}1$) ergibt sich ein ziemlich

ungünstiger Wert von $a_\infty = 40\,\mathrm{dB}$. Im übrigen läßt sich an Gl. (C.13) ablesen, daß bei kleinerem n der Grenzwert schneller erreicht wird, wenn der Betrag von $\frac{\omega_A - \omega_T}{\omega_A}$ auf eins zugeht. Ein höherer Fensterfunktionsgrad ist also auf jeden Fall günstiger.

b	n			
	1	2	4	8
0,1	-40	-72	-131	-237
0,3	-30	-53	-93	-161
0,5	-26	-44	-75	-126
0,7	-23	-39	-63	-103
	a_∞/dB			

Tabelle C.1: *Spektrale Grenzselektion a_∞ für Fenster* nP1 *und Analysebandbreiten $B_{3dB} = b\,$*Bark *bei Annahme einer Frequenzgruppenbreite von* 20% *der Analysefrequenz.*

Gegenüber dem einzelnen Drehzeiger der komplexen Schwingung existiert bei einem reellen Sinuston noch der gegenläufige Drehzeiger auf der negativen Seite des Fourier-Spektrums. Für diesen gilt der Effekt der Grenzselektion natürlich genauso. Nun allerdings überlagern sich bei höheren Analysefrequenzen die vom Analysefilter annähernd gleich (grenz)selektierten Zeiger, wodurch sich in zeitlicher Abfolge eine Verdoppelung bzw. Auslöschung ergibt ('Schaukeln' des FTT-Spektrums, Abschnitt 1.4.2 und Bild 3.8 oben). Insbesondere bei den Heinbachschen Standardparametern ist dabei die Wirkung der nachfolgende zeitlichen Glättung zu berücksichtigen. Sie glättet den zeitlichen Effekt soweit aus, daß eigentlich bei allen Signaltypen tiefere Spektralkomponenten höhere verdecken, die um mindestens etwa 40 dB schwächer sind.

C.3 Frequenzverlaufseigenschaften der FTT

Bereits in Abschnitt 1.4.2 wurde ersichtlich, daß an festen Analysefrequenzen Zeitverläufe des komplexen FTT-Spektrums und des FTT-Leistungsspektrums tiefpaßbegrenzt sowie Zeitverläufe des komplexen Bandpaßspektrums bandbegrenzt sind. Nachfolgend wird erstens nachgewiesen, daß das laufzeitausgeglichene, komplexe FTT-Bandpaßspektrum $s^{LB}(\omega, t)$ – im Gegensatz zum entsprechenden FTT-Spektrum $s^{L}(\omega, t)$ – entlang der Frequenzachse 'tiefpaßbegrenzt' und somit geglättet sowie interpolierbar ist. Daraus ergibt sich zweitens, daß die zugehörigen, nunmehr übereinstimmenden Leistungsspektren $|s^{LB}(\omega, t)|^2 = |s^{L}(\omega, t)|^2$ ebenfalls eine solche 'Tiefpaßbegrenzung' aufweisen. Man betrachtet dazu jeweils den Frequenzverlauf einer Repräsentation an einem festen Analysezeitpunkt $t = t_A$ als Signal über einer neuen 'Zeit' w, also beispielsweise

$$s^{LB}_{t_A}(w) = s^{LB}(\omega, t)|_{\omega=w, t=t_A}, \tag{C.15}$$

und untersucht dessen 'Spektrum' auf seine Lage und Ausdehnung. Die Anführungszeichen sollen im weiteren andeuten, daß der Begriff aus formaler Sicht der Fourier-Transformation $\mathcal{F}$ zutrifft, physikalisch aber die Dimensionen Frequenz und Zeit vertauscht sind. Nachdem die Variablennamen t und ω gemäß obigem Beispiel nicht mehr vorkommen, sollen sie – unabhängig von ihrer physikalischen Bedeutung – Zeit- und Spektralbereich von $\mathcal{F}$ kennzeichnen.

Zur Vereinfachung bleibt die Fensterfunktion der FTT vorläufig frequenzunabhängig $h(t)$. Die Analysebandbreite ist nun konstant und man benötigt keinen Laufzeitausgleich. Diese

beiden Voraussetzungen werden nachfolgend durch die Indexersetzung $L \rightarrow C$ markiert. Den Frequenzverlauf des komplexen FTT-Spektrums bei $t = t_A$ kann man dann so darstellen:

$$s^C_{t_A}(w) = \int_{-\infty}^{\infty} s(\tau)h(t_A - \tau)e^{-jw\tau}d\tau \tag{C.16}$$

$$= \mathcal{F}\{s(t) \cdot h(t_A - t)\}_{\omega=w}. \tag{C.17}$$

Verglichen mit Gl. (1.1) lassen sich die Integralgrenzen ohne Veränderung des Integralwertes von $0 \ldots t$ auf $-\infty \ldots +\infty$ erweitern, da Fensterfunktion und Zeitsignal $s(t)$ kausal sind und für Zeiten $t < 0$ als null definiert gelten. Das Integral verkörpert die Fourier-Transformation des zum Zeitpunkt t_A gefensterten Zeitsignals, wenn man τ durch t und w durch ω ersetzt. Auf $s^C_{t_A}(w)$ wird nun erneut die Fourier-Transformation angewendet, um das 'Spektrum' $S^C_{t_A}(\tau)$ zu erhalten. Hierfür ist formal vorher w durch t und nachher ω durch τ zu ersetzen:

$$S^C_{t_A}(\tau) = \mathcal{F}\left\{s^C_{t_A}(w)|_{w=t}\right\}_{\omega=\tau} \tag{C.18}$$

$$= \mathcal{F}\left\{\mathcal{F}\{s(t) \cdot h(t_A - t)\}_{\omega=t}\right\}_{\omega=\tau} \tag{C.19}$$

$$= 2\pi \cdot s(-\tau) \cdot h(t_A + \tau). \tag{C.20}$$

Die zweite Umformung ist bekannt als Symmetrieregel der Fourier-Transformation [Pap86]. Da die Fensterfunktion im wesentlichen als zeitbegrenzt angesehen werden kann, kontrolliert sie in Gl. (C.20) die Ausdehnung des 'Spektrums', wie auch immer $s(t)$ geartet sein mag. Allerdings hängt die Lage des 'Spektrums' vom aktuellen Analysezeitpunkt t_A ab. Sie verschiebt sich mit fortschreitender Analysezeit zu beliebig hohen Frequenzbeträgen. Damit ist $s^C_{t_A}(w)$ im allgemeinen keinesfalls ein 'tiefpaßbegrenztes' und somit auch kein geglättetes Signal. Anders verhält sich dagegen der Frequenzverlauf des zugehörigen FTT-Bandpaßspektrums

$$s^{CB}_{t_A}(w) = s^C_{t_A}(w) \cdot e^{jw(t_A - t_{max,0})}, \tag{C.21}$$

der aus Gl. (5.27) folgt. Wegen der vorläufig vorausgesetzten Frequenzunabhängigkeit ist zu beachten, daß die maximale Fensteröffnung $t_{max,0}$ bei $w = 0$ nun auch derjenigen an allen anderen Frequenzen gleicht. Mit Hilfe der Verschiebungsregel der Spektralfunktion (z.B. [Mar82]) entsteht auf ansonsten gleichem Wege wie oben das 'Spektrum'

$$S^{CB}_{t_A}(\tau) = \mathcal{F}\left\{\left[s^C_{t_A}(w) \cdot e^{jw(t_A - t_{max,0})}\right]_{w=t}\right\}_{\omega=\tau} \tag{C.22}$$

$$= \mathcal{F}\left\{\mathcal{F}\{s(t) \cdot h(t_A - t)\}_{\omega=t}\right\}_{\omega=\tau-(t_A - t_{max,0})} \tag{C.23}$$

$$= 2\pi \cdot s(-\tau + (t_A - t_{max,0})) \cdot h(\tau + t_{max,0}). \tag{C.24}$$

Man erkennt, daß die Abhängigkeit vom Analysezeitpunkt t_A im Argument der Fensterfunktion nicht mehr besteht. Die 'spektrale' Linksverschiebung um $t_{max,0}$ sorgt außerdem dafür, daß ein um Null zentriertes 'Band' ausgewählt wird. Speziell die statische Phasendrehung $e^{-jwt_{max,0}}$ in Gl. (5.27) trägt also dazu bei, daß $s^{CB}_{t_A}(w)$ als ein 'tiefpaßbegrenztes' oder geglättetes Signal betrachtet werden kann. Als nächstes wird der Frequenzverlauf des FTT-Leistungsspektrums

$$p^C_{t_A}(w) = |s^C_{t_A}(w)|^2 = |s^{CB}_{t_A}(w)|^2 \tag{C.25}$$

$$= s^C_{t_A}(w) \cdot s^{C\,*}_{t_A}(w) \tag{C.26}$$

untersucht. Mit der Fourier-Korrespondenz von Multiplikation und Faltung sowie derjenigen der konjugiert-komplexen 'Zeitfunktion' (z.B. [Mar82]) kann man ähnlich wie zuvor das 'Spektrum' herleiten:

$$\begin{aligned} P^C_{t_A}(\tau) &= \mathcal{F}\left\{\left[s^C_{t_A}(w)\cdot s^{C\,*}_{t_A}(w)\right]_{w=t}\right\}_{\omega=\tau} && \text{(C.27)}\\ &= \mathcal{F}\left\{\mathcal{F}\{s(t)\cdot h(t_A-t)\}_{\omega=t}\right\}_{\omega=\tau} * \mathcal{F}\left\{\mathcal{F}\{s(t)\cdot h(t_A-t)\}^*_{\omega=t}\right\}_{\omega=\tau} && \text{(C.28)}\\ &= [2\pi\cdot s(-\tau)\cdot h(t_A+\tau)] * [2\pi\cdot s(\tau)\cdot h(t_A-\tau)]^*\,. && \text{(C.29)} \end{aligned}$$

Die letzte Zeile besagt, daß $P^C_{t_A}(\tau)$ die Autokorrelationsfunktion (AKF) von $2\pi s(t)h(t_A-t)$ darstellt (z.B. [Pap86]). Da die Ausdehnung ihres Argumentes durch $h(t_A-t)$ beschränkt ist, ist ihre Ausdehnung im Wertebereich im wesentlichen durch die AKF von $h(t_A-t)$ allein bestimmt. Weil eine AKF eine Art Selbstfaltung des Arguments darstellt, kann sich die effektive 'Bandbreite' schlimmstenfalls verdoppeln (vgl. Abschnitt 1.5.1). Außerdem ist eine AKF gegenüber Zeitverschiebungen und Zeitumkehr im Argument invariant und immer zeitsymmetrisch im Wertebereich. Deshalb ist auch der Frequenzverlauf des FTT-Leistungsspektrums $p^C_{t_A}(w)$ unabhängig vom Analysezeitpunkt ein 'tiefpaßbegrenztes' Signal, allerdings mit einer schlimmstenfalls doppelt so hohen 'Grenzfrequenz' wie $s^{CB}_{t_A}(w)$.

Der Übergang auf FTT-Spektralrepräsentationen mit frequenzabhängigen Fensterfunktionen und Laufzeitausgleich geschieht so: Frequenzbereichsweise kann man die Fensterfunktion als annähernd konstant ansehen. Daß sich benachbarte Frequenzbereiche mit unterschiedlicher Fensterlänge dabei wenig beeinflussen, liegt an der elementaren Frequenzselektivität der FTT. Die Breite des 'Spektrums' eines Frequenzverlaufes wird zu höheren Frequenzbereichen hin im gleichen Maße schmaler wie die Fensterfunktion. Bei $s^{LB}_{t_A}(w)$ bewirkt der Laufzeitausgleich eine Zeitverschiebung der Fenster derart, daß immer eine um null zentrierte Auswahl des 'Bandes' sichergestellt ist. Bei $p^L_{t_A}(w)$ bleibt es exakt symmetrisch um null und bei $s^L_{t_A}(w)$ wandert es weiterhin mit fortschreitender Analysezeit zu beliebig hohen Frequenzbeträgen aus.

Frequenzverläufe des komplexen FTT-Bandpaßspektrums und des FTT-Leistungsspektrums – nicht jedoch des komplexen FTT-Spektrums – sind bei Laufzeitausgleich also 'tiefpaßbegrenzt'. Die 'Grenzfrequenz' verhält sich über dem Frequenzverlauf proportional zur Fensterlänge. Die Zuweisung einer 'Grenzfrequenz' ist allerdings recht willkürlich, da die Begrenzung in Gestalt der ansteigenden und abfallenden Flanke der Fensterfunktion unscharf ausfällt.

Anhang D

Spezielle Ergebnisse

D.1 Hörversuch zur Heinbachschen TTZM-Datenreduktion

In einem Hörversuch wurden die Syntheseergebnisse von verschiedenen Maßnahmenkombinationen verglichen, die die Ursachen der Qualitätsbeeinträchtigungen bei TTZM-Datenreduktion klären helfen. Als Sprachsignal dient der Testsatz des männlichen Sprechers aus Bild 1.2. Tabelle D.1 stellt die Beobachtungen dar, die der Autor als Versuchsperson durch zahlreiche Paarvergleiche über Kopfhörer gemacht hat. Sie konnten anhand von zwei weiteren, ähnlichen Sprachbeispielen eines männlichen und eines weiblichen Sprechers zuverlässig reproduziert werden.

Die Spalten in der linken Tabellenhälfte stehen für elf Schalle, die nach Anwendung verschiedener Maßnahmen aus dem üblichen Heinbachschen Teiltonzeitmuster gewonnen wurden. Die obere Tabellenhälfte legt fest, welche von den acht im Haupttext angesprochenen Maßnahmen bei dem jeweiligen Schall kombiniert sind. Die untere Tabellenhälfte verknüpft die Schalle zu Darbietungspaaren a bis k und vermerkt die wahrgenommen Veränderungen von A nach B. Schall I verkörpert Heinbachs TTZM-Verfahren mit 4,4 kbit/s, Schall VIII sein nichtreduzierendes TTZM-Verfahren, jeweils nach Anwendung der authentischen Teiltonsynthese mit Rechteckfenster. Den Angelpunkt der meisten Vergleiche bildet der Schall IV, da er von Veränderungen durch die weniger effizienten Reduktionsmaßnahmen bzw. von Störungen durch das Syntheseverfahren bereits befreit ist.

D.2 FTT-Codierungsrahmen als Wavelet-Transformation

Das Transformationspaar T und R, welches den Zusammenhang zwischen kontinuierlichem Zeitsignal und dem abgetasteten FTT-Bandpaßspektrum liefert, wird auf Anregung von Horn [HorPK] als besondere Form der Wavelet-Transformation [Rio91, Vet92] vorgestellt. Sie reicht historisch auf die ‘Gabor-Expansion’ zurück [Gab46] und beschreibt das Zeitsignal als eine gewichtete Summe von ‘Wavelets’, deren Realteile hier Sinustonimpul-

Tabelle D.1: *Maßnahmen zur Datenreduktion (oben) angewandt in elf Kombinationen auf ein Sprachbeispiel (Schalle I bis XI) und subjektive Beurteilung im Paarvergleich A—B (unten). Die fünf TTZM-Ausschnitte in Bild 2.11 beziehen sich auf die Schalle VII, V, IV, IX und X.*

Schall											**Maßnahme**	
I	II	III	IV	V	VI	VII	VIII	IX	X	XI	Nr.	Text
●	●	●	●	-	●	-	-	●	●	●	1	verlängertes Auswerteintervall 20ms
●	●	●	●	●	-	-	-	●	●	●	2	Beschränkung auf zehn Teiltöne
●	●	-	-	-	-	-	-	-	-	-	3	Codierung Teiltonpegel 2x4bit/10TT
●	●	●	-	-	-	-	-	-	-	-	4	Codierung Teiltonfrequenzen 8bit/TT
-	●	●	●	●	●	●	-	●	●	●	5	Dreieck- statt Rechteckfenster
-	-	-	-	-	-	-	-	●	-	-	6	Interpolation des Teiltonverlaufs
-	-	-	-	-	-	-	-	-	●	-	7	Auswahl Tonhöhengewicht statt Pegel
-	-	-	-	-	-	-	-	-	-	●	8	Entfernung Teiltöne kürzer 20ms
A	B	·	·	·	·	·	·	·	·	·	a	Knattern verschwindet, Sprache undeutlicher.
·	A	B	·	·	·	·	·	·	·	·	b	Intensitätsschwankung verschwindet, Rauhigkeit reduziert.
·	·	A	B	·	·	·	·	·	·	·	c	kein nennenswerter Unterschied
·	·	·	A	B	·	·	·	·	·	·	d	Klingeln und Rauhigkeit verschwinden, Zischeln entsteht, Sprache deutlicher.
·	·	·	A	·	B	·	·	·	·	·	e	Klingeln wird zu tonalem Rauschteppich, Überspitzung und Rauhigkeit verschwinden.
·	·	·	·	A	·	B	·	·	·	·	f	Rauschteppich entsteht, Überspitzung und Zischeln verschwinden.
·	·	·	·	·	A	B	·	·	·	·	g	Rauschteppich wird atonal, Sprache deutlicher.
·	·	·	·	·	·	A	B	·	·	·	h	Rauschteppich etwas verstärkt, Störgeräusch entsteht.
·	·	·	A	·	·	·	·	B	·	·	i	Rauhigkeit verschwindet.
·	·	·	A	·	·	·	·	·	B	·	j	Klingeln verstärkt, Überspitzung verschwindet, Sprache undeutlicher.
·	·	·	A	·	·	·	·	·	·	B	k	Klingeln verstärkt, Sprache undeutlicher.
I	II	III	IV	V	VI	VII	VIII	IX	X	XI	Nr.	Text
Schall											**Paarvergleich A—B**	

sen entsprechen. [1] Gegenüber üblichen Formen der Wavelet-Transformation bedingt die Natur von T und R wesentliche Unterschiede. Weil bei der FTT keine Proportionalität von Analysefrequenz und Analysebandbreite besteht, können die zu definierenden Wavelets nicht zeitlich gestauchte oder gestreckte Versionen eines Wavelet-Prototyps sein. Der übliche Wavelet-Parameter 'Skalierung' macht hier als gemeinsames Maß für Frequenz und Bandbreite keinen Sinn und wird unmittelbar durch die Frequenz ersetzt. Weiterhin sind Analyse- und Synthese-Wavelets unterschiedlich und weisen keine Orthogonalitätseigenschaften auf. Hervorzuheben ist schließlich, daß exakte Signalrekonstruierbarkeit nur im auditiven, nicht aber im mathematischen Sinn beabsichtigt ist. Die Zeitrasterabstände

[1] Bereits Korn forderte in Anlehnung an die Arbeiten von Gabor eine Signaldarstellung durch sogenannte 'auditorische Elementarbotschaften' [Kor69], also gewissermaßen durch gehörorientierte Wavelets. T und R realisieren genau diese Idee.

T_A und T_S können nachfolgend frequenzabhängig sein.

Transformation T: Das FTT-Bandpaßspektrum $s^{LB}(\omega,t)$ an einem Punkt des Abtastrasters $\{(\omega_{A_i},kT_A)\}$ ergibt sich aus dem inneren Produkt von Zeitsignal $s(t)$ und dem Analyse-Wavelet $w^A(\omega_{A_i},t-kT_A)$ an diesem Punkt:

T

$$s^{LB}(\omega_{A_i},kT_A) = \int_{-\infty}^{\infty} s(\tau)\cdot[w^A(\omega_{A_i},\tau-kT_A)]^*\cdot d\tau, \tag{D.1}$$

$$w^A(\omega_A,t) = [h^A_{\omega_A}(-t)*l^A_{\omega_A}(-t)]^*\cdot e^{j\omega_A(t+t^A_{max,0})}. \tag{D.2}$$

Dies folgt aus der ursprünglichen Formulierung der Transformation T, indem man Gl. (5.23) in Gl. (5.24) einsetzt:

$$s^{LB}(\omega_{A_i},kT_A) = \left\{\left[(s(t)\cdot e^{-j\omega_A t})*h^A_{\omega_A}(t)*l^A_{\omega_A}(t)\right]\cdot e^{j\omega_A(t-t^A_{max,0})}\right\}_{\substack{\omega_A=\omega_{A_i}\\ t=kT_A}} \tag{D.3}$$

$$= \left\{s(t)*\left[(h^A_{\omega_A}(t)*l^A_{\omega_A}(t))\cdot e^{j\omega_A(t-t^A_{max,0})}\right]\right\}_{\substack{\omega_A=\omega_{A_i}\\ t=kT_A}}. \tag{D.4}$$

In der zweiten Zeile wurde das Distributivgesetz der Faltung bezüglich der linearen Modulation $e^{j\omega_A t}$ angewendet. Gl. (D.1) erhält man, wenn man die Faltung mit $s(t)$ als Integral ausschreibt und den Faltungskern zeitinvertiert und komplex-konjugiert in Gl. (D.2) auslagert. Das Integral ist formal über die gesamte Zeitachse ausgedehnt, obwohl die Kausalität implizit durch Annahme von $s(t)=0$, $h^A_{\omega_A}(t)=0$ und $l^A_{\omega_A}(t)=0$ für $t<0$ gewährleistet bleibt.

Rücktransformation R: Das rekonstruierte Zeitsignal $\hat{s}(t)$ läßt sich darstellen als Überlagerung aller Synthese-Wavelets $w^S(\omega_{S_m},t-lT_S)$ des Rasters $\{(\omega_{S_m},lT_S)\}$, gewichtet jeweils mit dem entsprechenden Abtastwert des möglicherweise fehlerbehafteten FTT-Bandpaßspektrums $\hat{s}^{LB}(\omega,t)$:

R

$$\hat{s}(t) = \sum_m\sum_l 2\mathrm{Re}\left\{\hat{s}^{LB}(\omega_{S_m},lT_S)\cdot w^S(\omega_{S_m},t-lT_S)\right\}, \tag{D.5}$$

$$w^S(\omega_S,t) = T_S\cdot\frac{\Delta\omega_S(\omega_S)}{2\pi h^{A*S}_{max,\omega_S}}\cdot(h^S_{\omega_S}(t)*l^S_{\omega_S}(t))\cdot e^{j\omega_S(t-t^S_{max,0})}. \tag{D.6}$$

Abweichend zur üblichen Formulierung der Wavelet-Rücktransformation deutet der Realteil-Operator an, daß spektrale Symmetrien durch Annahme eines reellen Zeitsignals ausgenutzt werden. Es sind deshalb nur positive Synthesefrequenzen zu berücksichtigen. Zur Herleitung ist Gl. (5.25) in Gl. (5.26) einzusetzen:

$$\hat{s}(t) = \sum_{\substack{\omega_S=\omega_{S_m}\\ m}} 2\mathrm{Re}\left\{\left[\left(\sum_l \hat{s}^{LB}(\omega_S,lT_S)\cdot\delta(t-lT_S)\cdot T_S\cdot e^{-j\omega_S t}\right)*\right.\right.$$

$$\left.\left. h^S_{\omega_S}(t)*l^S_{\omega_S}(t)\right]\cdot e^{j\omega_S(t-t^S_{max,0})}\right\}\cdot\frac{\Delta\omega_S(\omega_S)}{2\pi h^{A*S}_{max,\omega_S}} \tag{D.7}$$

$$= \sum_{\substack{\omega_S=\omega_{S_m}\\ m}} 2\mathrm{Re}\left\{\left(\sum_l \hat{s}^{LB}(\omega_S,lT_S)\cdot\delta(t-lT_S)\right)*\right.$$

$$\left[T_S \cdot \frac{\Delta\omega_S(\omega_S)}{2\pi h^{A*S}_{max,\omega_S}} \cdot (h^S_{\omega_S}(t) * l^S_{\omega_S}(t)) \cdot e^{j\omega_S(t-t^S_{max,0})}\right]\Bigg\}. \quad \text{(D.8)}$$

In der Umformung wurde wie zuvor bei T die lineare Modulation $e^{j\omega_S t}$ in die eckige Klammer hineingezogen, womit das Synthese-Wavelet aus Gl. (D.6) bereits explizit vorliegt. Die mit den einzelnen Dirac-Impulsen ausgeführte Faltung erzeugt die entsprechend zeitverschobenen Summanden in Gl. (D.5). Der Realteiloperator zieht am Schluß in die Summe hinein.

D.3 Signaldarstellung durch Konturpunkt-Wavelets

Bei der Signalrekonstruktion aus Konturen wird die Rücktransformation R im Sinne einer Wavelet-Rücktransformation (s.o.) so verwendet, daß ein Großteil der zu überlagernden Wavelets das Gewicht null erhält. Die übrigen Werte verkörpern die gerasterten Punkte der Zeit- und Frequenzkonturen, welche eine gegenseitige Maskierung überstanden haben und über die ursprüngliche oder rekonstruierte Phaseninformation verfügen. Dies sind die Rasterwerte $s_{MF}(\omega_{S_m}, lT_S) \neq 0$ bzw. $s_{MZ}(\omega_{S_m}, lT_S) \neq 0$ (Bild 5.3a,b und Anhang B.6). Mit ihnen ergibt sich eine Darstellung des rekonstruierten Signals $\hat{s}(t)$ durch Wavelets, die nur an Konturpunkten existieren. Zeit- und Frequenzkonturpunkten sind im allgemeinen unterschiedliche Wavelet-Grundtypen $w^S_F(\omega_S, t)$ und $w^S_Z(\omega_S, t)$ zuzuordnen:

$$\begin{aligned}
\hat{s}(t) &= \sum_{s_{MF}(\omega_{S_m}, lT_S)\neq 0} 2\mathrm{Re}\left\{s_{MF}(\omega_{S_m}, lT_S) \cdot w^S_F(\omega_{S_m}, t - lT_S)\right\} + \\
&\quad \sum_{s_{MZ}(\omega_{S_m}, lT_S)\neq 0} 2\mathrm{Re}\left\{s_{MZ}(\omega_{S_m}, lT_S) \cdot w^S_Z(\omega_{S_m}, t - lT_S)\right\}, \quad \text{(D.9)} \\
w^S_F(\omega_S, t) &= \frac{T_S}{\left|H^A_{\omega_S}(0) \cdot H^S_{\omega_S}(0)\right|} \cdot (h^S_{\omega_S}(t) * l^S_{\omega_S}(t)) \cdot e^{j\omega_S(t-t^S_{max,0})}, \quad \text{(D.10)} \\
w^S_Z(\omega_S, t) &= \frac{\Delta\omega_S(\omega_S)}{2\pi h^A_{max,\omega_S} \cdot h^S_{max,\omega_S}} \cdot (h^S_{\omega_S}(t) * l^S_{\omega_S}(t)) \cdot e^{j\omega_S(t-t^S_{max,0})}. \quad \text{(D.11)}
\end{aligned}$$

Jeder gerasterte Konturpunkt, der nicht maskiert wurde, bestimmt damit ein Konturpunkt-Wavelet nach Betrag und Phase. Wenn man keine Zeitkonturen berücksichtigen will, dann verschwindet die zweite Summe und die gegenseitige Maskierung bleibt wirkungslos, so daß mit $s^{MF}(\omega_{S_m}, lT_S) = s^F(\omega_{S_m}, lT_S)$ allen gerasterten Frequenzkonturpunkten Wavelets zugewiesen werden. Dies gilt analog bei Verzicht auf Frequenzkonturen für die gerasterten Zeitkonturpunkte.

Zur Herleitung der obigen Darstellung ersetzt man $\hat{s}^{LB}(\omega_{S_m}, lT_S)$ in Gl. (D.5) durch Gl. (B.39). Die beiden neuen Wavelet-Grundtypen sind die Produkte der Bewertungsfaktoren $c_F(\omega_S)$ und $c_Z(\omega_S)$ nach Gl. (B.40) bzw. Gl. (B.43) mit dem ursprünglichen Grundtyp $w^S(\omega_S, t)$ aus Gl. (D.6). Schließlich brauchen die mit null gewichteten Wavelets nicht mitsummiert zu werden.

Die beiden Wavelet-Grundtypen unterscheiden sich lediglich in ihrem Gewicht. Bemerkenswert ist dabei die Frequenz/Zeit-Bereichssymmetrie der einzelnen Parameter in den Brüchen: Es korrespondieren die Rasterabstände T_S und $\Delta\omega_S(\omega_S)/(2\pi)$, wie auch die

Maxima der Analyse- bzw. Synthesefensterfunktion mit ihrer jeweiligen Fourier-Transformierten, h^A_{max,ω_S} und $|H^A_{\omega_S}(0)|$ bzw. h^S_{max,ω_S} und $|H^S_{\omega_S}(0)|$.

Bei frequenzabhängiger Wahl des Zeitrasterabstandes $T_S = T_S(\omega_S)$ in Abstimmung mit dem Frequenzrasterabstand läßt sich auch ein einheitlicher Wavelet-Grundtyp $w^S_K(\omega_S, t) = w^S_F(\omega_S, t) = w^S_Z(\omega_S, t)$ einstellen. Die Rasterwerte $s_{MF}(\omega_{S_m}, lT_S) + s_{MZ}(\omega_{S_m}, lT_S) \neq 0$ repräsentieren dann gegenseitig maskierte Konturpunkte, deren Herkunft aus Frequenz- oder Zeitkonturierung weiter keine Rolle mehr spielt. Dies ergibt die Möglichkeit einer Signaldarstellung durch 'anonyme' Konturpunkt-Wavelets, so wie von Horn postuliert [HorPK]. [2]

[2]Eine Konturcodierung auf Basis anonymer Konturen scheint allerdings schwierig. Wenn man möglichst wenig Konturpunkte codieren möchte, dann wird ihre Assoziation zu Linien problematisch, weil keine Vorzugsrichtung mehr angenommen werden kann. Erschwerend kommt hinzu, daß nun die gegenseitige Maskierung der Konturen auf Coderseite zu erfolgen hat, der die wiederaufzufindenden Linien zerstückelt. Die Linienassoziation im Decoder ist aber die wesentliche Grundlage für Rasterierung und Phasenrekonstruktion.

Anhang E

Abkürzungen und Formelzeichen

AKF	Autokorrelationsfunktion
APF	Auswahl prägnanter Frequenzkonturen
APZ	Auswahl prägnanter Zeitkonturen
CELP	Codebook Excited Linear Predictive Coding
FG	spektrale Glättung der Texturhüllfläche
FK	Frequenzkonturen bzw. Frequenzkonturierung
FKC	Frequenzkontur-Codierung
FKD	Frequenzkontur-Decodierung
FKL	Frequenzkonturlinie
FM	Frequenzmodulation
FR	Frequenzkontur-Rasterierer
FS	Frequenzkontur-gesteuertes Sieb
FTT	Fourier-t-Transformation
HB-4k4	TTZM-Codierverfahren 4,4 kbit/s nach Heinbach
HB-TTZM	Frequenzkontur/TTZM-Analyse nach Heinbach
IPA	Impulsantwort
KTX	Kontur/Textur-Analyse
KTXOZ	Kontur/Textur-Analyse ohne Zeitkonturen
LPC	Linear Predictive Coding
M-TTZM	bestmögliche Frequenzkontur/TTZM-Analyse
MSK	gegenseitige Maskierung der Konturen
MUM-30k	Kontur/Textur-Codierverfahren 30 kbit/s
MUM-4k4	Kontur/Textur-Codierverfahren 4,4 kbit/s
PRK	Phasenrekonstruktion
QMF	Quadrature Mirror Filter
QSS	Quellsinusschwingung
R	Rücktransformation für abgetastetes FTT-Bandpaßspektrum
RFTT	Rücktransformation der Fourier–t–Transformation
RKHP	Rekonstruktion Zeit- und Frequenzkonturen mit Phasenheuristik
RKOP	Rekonstruktion Zeit- und Frequenzkonturen mit Originalphase
SM-TTZM	Frequenzkontur/TTZM-Analyse nach Schlang/Mummert
SSS	Synthesesinusschwingung
STR	Sinustonrepräsentation
STW	Spektraltonhöhenwahrnehmung

SUB	Leistungssubtraktion zur Texturhüllfläche
SZM	spektral/zeitliche Modulation
T	Hintransformation für abgetastetes FTT-Bandpaßspektrum
TDAC	Time Domain Aliasing Cancellation
TT	Teilton
TTM	Teiltonmuster
TTZM	Teiltonzeitmuster
TTSR	Teiltonsynthese mit Rechteckfenster
TTSD	Teiltonsynthese mit Dreieckfenster
TX	Textur
TXC	Texturcodierung
TXD	Texturdecodierung
WFS	Wandlung prägnanter Frequenzkonturen in FTT-Spektrum
WR	Weißes Rauschen
WZS	Wandlung prägnanter Zeitkonturen in FTT-Spektrum
ZFKI	Konturanalyse bei optimaler Rekonstruktion
ZFKII	Konturanalyse bei Rekonstruktion mit Phasenheuristik
ZK	Zeitkonturen bzw. Zeitkonturierung
ZG	zeitliche Glättung der Texturhüllfläche
ZS	Zeitkontur-gesteuertes Sieb
ZR	Zeitkontur-Rasterierer

$[\ ]_Q$	Quantisierungsoperator
$\gamma(t)$	Einheitssprung
$\delta(t)$	Dirac-Impuls
Θ, $\hat{\Theta}$	transienter Anteil des FTT-Spektrums, Näherungswert
λ	zeitliche Ausgeprägtheitsschwelle bei Zeitkonturierung
Ξ	stationärer Anteil des FTT-Spektrums
$\tau_g(\omega)$	Gruppenlaufzeit
ϕ	Phase
$\phi^{LB}(f,t)$	FTT-Bandpaßphasenspektrum
$\hat{\phi}^i_F(t)$, $\hat{\phi}^k_Z(f)$	rekonstruierte Phase über Frequenzkonturlinie i, ZK-Linie k
$\Phi\{\ \}$, $\Phi\{\mathcal{C}\}$	Phasenoperator, Konturphasen
ψ, $\Delta\psi$	Phasendriftfunktion, Phasendrift
ω	Kreisfrequenz
ω_A, $\Delta\omega_A(\omega)$	Analysefrequenz, frequenzabhängiger Analysefrequenzabstand
ω_S, $\Delta\omega_S(\omega)$	Synthesefrequenz, frequenzabhängiger Synthesefrequenzabstand
$\Delta\omega_{TX}(\omega)$	frequenzabhängiger Frequenzabstand Texturstützstellen
Ω	normierte Kreisfrequenz

a_D	Dämpfung
$a_h(T)$, $a_H(\Omega)$	zeitliche, spektrale Selektion von normierter Fensterfunktion
A	Amplitude
b	Codebreite bzw. $3dB$-Analysebandbreite in Bark
B	Bandbreite über f (bei Tiefpässen symmetrisch um $f = 0$)
B^{gs}_{3dB}	3dB-Bandbreite Gauß-Filter
B_{3dB}, B^A_{3dB}	3dB-Analysebandbreite

B_{3dB}^{FG}	3dB-Breite Faltungskern für spektrale Texturglättung
B_{3dB}^{S}	3dB-Synthesebandbreite
B_{3dB}^{ZG}	3dB-Bandbreite Glättungstiefpaß für zeitliche Texturglättung
c_{bal}	Balance zwischen Kontur- und Texturanteil
$c_F(\omega)$, $c_Z(\omega)$	Bewertungsfaktoren für Überlagerung $s_{MF}(\omega,t)$ und $s_{MZ}(\omega,t)$
$c_{TX}(\omega)$	Bewertungsfaktor für Überlagerung $s_{TX}(\omega,t)$
$\mathcal{C}_F$, $\mathcal{C}_{PF}$	Frequenzkonturen, prägnante Frequenzkonturen
$\mathcal{C}_Z$, $\mathcal{C}_{PZ}$	Zeitkonturen, prägnante Zeitkonturen
$\mathcal{C}(t)$, $\mathcal{C}(f,t)$	Kontursignal, Konturpunktsignal
d, $\overline{d}$	Stützstellendichte, Langzeitmittel
E	Signalenergie
$e(t)$	Signalhüllkurve
f	Schwingungsfrequenz
f_a	Abtastrate Zeitsignal
Δf_Φ	Phasenübergabetoleranz bei Teiltönen/Frequenzkonturen
Δf_G	Frequenzgruppenbreite des Gehörs
Δf_P	Mindestlänge prägnanter Zeitkonturlinien
Δf_U	Unstetigkeitstoleranz Frequenzkonturlinien
$\mathcal{F}\{\ \}$, $\mathcal{F}^{-1}\{\ \}$	Fourier-Transformation, Fourier-Rücktransformation
$g(t)$, $G(\omega)$	Gesamtimpulsantwort FTT–RFTT, Fourier-Transformierte
$h(t)$, $h^N(T)$	Fensterfunktion, normierte Fensterfunktion
$h^{gs}(t)$	Impulsantwort Gauß-Filter
h_{max}	Scheitelwert einer Fensterfunktion
$h_\omega(t)$, $h_\omega^A(t)$	frequenzabhängige Analysefensterfunktion
$h_\omega^{A*S}(t)$	Faltungsergebnis Analyse- mit Synthesefensterfunktion
$h_\omega^G(t)$	Impulsantwort der frequenzabhängigen Glättung
$h_\omega^S(t)$	frequenzabhängige Synthesefensterfunktion
$H(\omega)$, $H(p)$	Fourier- und Laplace-Transformierte von $h(t)$
$H(\Omega)$, $H(P)$	normierte Fourier- und Laplace-Transformierte von $h(t)$
$\mathrm{int}(x)$	Integer-Funktion, ganzzahliger Anteil von x
I, $\overline{I}$	Datenrate, Langzeitmittel
$l_\omega(t)$	Impulsantwort Laufzeitglied an der Frequenz ω
L	Pegel
$L(f,t)$	FTT-Pegelspektrum (FTT-Spektrogramm)
L_{bal}	Pegelbalance zwischen Kontur- und Texturanteil
$L_e(t)$	Hüllkurvenpegelverlauf
$L^G(f,t)$, $L^L(f,t)$	FTT-Pegelspektrum nach Glättung, nach Laufzeitausgleich
$L_{PF}(f,t)$	in FTT-Pegelspektrum umgewandelte prägnante FK
$L_{PZ}(f,t)$	in FTT-Pegelspektrum umgewandelte prägnante ZK
$L_{TX}(f,t)$	Texturhüllfläche
ΔL_A	spektrale Ausgeprägtheitschwelle bei Frequenzkonturierung
ΔL_M	Pegelanhebung Frequenzkonturen für MSK
ΔL_{PF}, ΔL_{PZ}	Zuschläge auf FTT-Pegelspektren rückgewandelter Konturen
$\mathcal{L}\{\ \}$	Laplace-Transformation
$n(t)$, $n^{LB}(\omega,t)$	Rauschen, dessen FTT-Bandpaßspektrum mit Laufzeitausgleich
N, $\overline{N}$	Stützstellenanzahl, Langzeitmittel
p, P	Laplace-Transformationsvariable, normiert
p_{ref}	Referenzleistung für Pegel 0 dB

$p_{\omega_A}(t)$, $p^G_{\omega_A}(t)$	Zeitverlauf FTT-Leistungsspektrum bei ω_A, nach Glättung
Re{ }	Realteiloperator
$s(t)$, $\hat{s}(t)$	Zeitsignal, rekonstruiertes Zeitsignal
$s(\omega, t)$	FTT-Spektrum
$s_\omega(t)$, $s^L_\omega(t)$	Zeitverlauf FTT-Spektrum bei ω, nach Laufzeitausgleich
$s^B_\omega(t)$, $s^{LB}_\omega(t)$	Zeitverlauf FTT-Bandpaßspektrum, nach Laufzeitausgleich
$\hat{s}^{LB}_\omega(t)$, $\hat{s}^{LB,\delta}_\omega(t)$	decodiertes FTT-Bandpaßspektrum bei ω, als Dirac-Impulsfolge
$s_F(\omega, t)$	FK-gesiebtes/gerastertes FTT-Bandpaßspektrum
$s_{MZ}(\omega, t)$	dito, nach Maskierung durch Zeitkonturen
$s_Z(\omega, t)$	ZK-gesiebtes/gerastertes FTT-Bandpaßspektrum
$s_{MZ}(\omega, t)$	dito, nach Maskierung durch Frequenzkonturen
$s_{TX}(\omega, t)$	FTT-Bandpaßspektrum bei Rekonstruktion Texturanteil
t	Zeit
t_a	Abtastintervall Zeitsignal
$t_{max,\omega}$	Zeitpunkt der maximalen Öffnung der Fensterfunktion bei ω
Δt_Φ	Phasenübergabetoleranz bei Zeitkonturen
Δt_P	Mindestlänge prägnanter Frequenzkonturlinien
Δt_U	Unstetigkeitstoleranz Zeitkonturlinien
T	normierte Zeit
T_{max}	normierter Zeitpunkt der maximalen Fensteröffnung
T_A, T_S	Auswerte- oder Analyseintervall, Syntheseintervall (unnormiert)
$T_g(\Omega)$	normierte Gruppenlaufzeit
T_G	Glättungszeitkonstante (unnormiert)
T_{3dB}, T_{6dB}	3dB-Breite, Halbwertsbreite Zeitfenster (unnormiert)
$w^A(\omega, t)$	Analyse-Wavelet
$w^S(\omega, t)$	Synthese-Wavelet
$w^S_F(\omega, t)$	Frequenzkonturpunkt-Wavelet
$w^S_K(\omega, t)$	(anonymes) Konturpunkt-Wavelet
$w^S_Z(\omega, t)$	Zeitkonturpunkt-Wavelet
$x_\omega(t)$	Impulsantwort FTT-RFTT-Korrektursystem an der Frequenz ω
$x^A_\omega(t), x^S_\omega(t)$	Analysekorrektur, Synthesekorrektur an der Frequenz ω
z	Tonheit (gehöradäquate Frequenzabbildung, 'Bark-Skala')

Quellenverzeichnis

[Ada91] **Adams, J. W.** *A new optimal window.* IEEE Trans. Acoust., Speech, Signal Processing, 39(8): 1753–1769, 1991.

[Alm82a] **Almeida, L. B., Tribolet, J. M.** *A spectral model for nonstationary voiced speech.* In: Intern. Conf. on Acoustics, Speech and Signal Processing, Tokyo, 1303–1306, 1982.

[Alm82b] **Almeida, L. B., Tribolet, J. M.** *Harmonic coding: A low bit rate good quality speech coding technique.* In: Intern. Conf. on Acoustics, Speech and Signal Processing, Paris, 1664–1667, 1982.

[Alm83] **Almeida, L. B., Tribolet, J. M.** *Nonstationary spectral modelling of voiced speech.* IEEE Trans. Acoust., Speech, Signal Processing, 31(3): 664–677, 1983.

[Aur84] **Aures, W.** *Berechnungsverfahren für den Wohlklang (die sensorische Konsonanz) beliebiger Schallsignale, ein Beitrag zur gehörbezogenen Schallanalyse.* Dissertation, Technische Universität München, 1984.

[Bau95] **Baumann, U.** *Ein Verfahren zur Erkennung und Trennung multipler akustischer Objekte.* Herbert Utz Verlag Wissenschaft, München, 1995.

[Ber89] **Berthommier, F., Schwartz, J. L., Escudier, P.** *Auditory processing in a post-cochlear neural network: Vowel spectrum processing based on spike synchrony.* In: Eurospeech 89, Paris, 247–250, 1989.

[Bra87] **Brandenburg, K.-H.** *OCF – A new coding algorithm for high quality sound signals.* In: Intern. Conf. on Acoustics, Speech and Signal Processing, Dallas, 141–144, 1987.

[Bra94] **Brandenburg, K., Stoll, G.** *ISO-MPEG-1 audio: A generic standard for coding of high quality digital audio.* J. Audio Eng. Soc., 42(10): 780–792, 1994.

[Bre90] **Bregman, A. S.** *Auditory Scene Analysis.* Massachusetts Institute of Technology, Cambridge, MA, 1990.

[Cam90] **Campbell, J. P. Jr., Tremain, T. E.** *The proposed Federal Standard 1016 4800bps voice coder: CELP.* Speech Technology, 58–64, April/May 1990.

[CEL95] C-Quellcode für U.S. Federal Standard 1016 CELP 4,8 kbit/s. `ftp://ftp.super.org/pub/speech/celp-3.2a.tar.Z` im Internet am 26.10.95.

[Che93] **Cheng, Y.-M., O'Shaughnessy**. *On 450-600 b/s natural sounding speech coding.* IEEE Trans. Speech Audio Processing, 1(2): 207–230, 1993.

[Chi82] **Chistovich, L. A., Lublinskaya, V. V., Malinnikova, T. G., Ogorodnikova, E. A., Stoliarova, E. I., Zhukov, S. J.** *Temporal processing of peripheral auditory patterns of speech.* In: Carlson, R., Granström, B., Hrsg., The Representation of Speech in the Peripheral Auditory System, 165–180. Elsevier Biomedical Press, Amsterdam, 1982.

[Coo86] **Cooke, M. P.** *A computer model of peripheral auditory processing incorporating phase-locking, suppression and adaptation effects.* Speech Communication, 5: 261–281, 1986.

[Coo93] **Cooke, M. P.** *Modelling Auditory Processing and Organisation.* Cambridge University Press, Cambridge UK, 1993.

[Cox91] **Cox, R. V., Hagenauer, J., Seshadri, N., Sundberg, C.-E. W.** *Subband speech coding and matched convolutional channel coding for mobile radio channels.* IEEE Trans. Signal Processing, 39(8): 1717–1731, 1991.

[Cro83] **Crochiere, R. E., Rabiner, L. R.** *Multirate Digital Signal Processing.* Prentice-Hall, Englewood Cliffs, NJ, 1983.

[CSA95] Dokumentation ITU-T Standard CS-ACELP 8 kbit/s. `ftp://ftp.research.att.com/dist/g729/g729.ps` im Internet am 24.11.95.

[Dis59] **Dishal, M.** *Gaussian response filter design.* Electrical Communication, 36(1): 3–26, 1959.

[Ell92] **Ellis, D. P. W.** *A Perceptual Representation auf Audio.* S.M. dissertation, Massachusetts Institute of Technology, 1992.

[Fan50] **Fano, R. M.** *Short-time autocorrelation functions and power spectra.* J. Acoust. Soc. Am., 22(5): 546–551, 1950.

[Fas89] **Fastl, H.** *Pitch strength of pure tones.* In: 13th Intern. Conf. on Acoustics, Belgrade, Yugoslavia 1989, 11–14, 1989.

[Fei89] **Feiten, B. F.** *Beurteilung von Quellencodierungsverfahren für Audiosignale bei Berücksichtigung der Verdeckungseigenschaften des Gehörs.* Dissertation, Technische Universität Berlin, 1989.

[Fei90] **Feiten, B., Becker, H.** *Analyse/Synthese-Verfahren zur Modellierung von Klängen.* In: Fortschritte der Akustik – DAGA '90, 533–536, DPG-GmbH, Bad Honnef, 1990.

[Fel85] **Feldtkeller, M.** *Fourier-t-Transformation als gehörbezogene Spektralanalyse.* Diplomarbeit am Lehrstuhl für Elektroakustik, Technische Universität München, 1985.

[Fla66] **Flanagan, J. L., Golden, R. M.** *Phase vocoder.* Bell Syst. Tech. J., 45: 1493–1509, Nov. 1966.

[Fla72] **Flanagan, J. L.** *Speech Analysis, Synthesis and Perception.* Springer-Verlag, Berlin, 2. Aufl., 1972.

[Fla79] **Flanagan, J. L., Schroeder, M. R., Atal, B. S., Crochiere, R. E., Jayant, N. S., Tribolet, J. M.** *Speech coding.* IEEE Trans. Commun., 27(4): 710–737, 1979.

[Gab46] **Gabor, D.** *Theory of communication.* J. Inst. Elec. Eng., London, 93(III): 429–457, Nov. 1946.

[Gam71] **Gambardella, G.** *A contribution to the theory of short-time spectral analysis with nonuniform bandwith filters.* IEEE Trans. Circuit Theory, 18(4): 455–460, 1971.

[Ger94] **Gersho, A.** *Advances in speech and audio compresseion.* Proc. IEEE, 82(6): 900–918, 1994.

[Ghi87] **Ghitza, O.** *Auditory nerve representation criteria for speech analysis/synthesis.* IEEE Trans. Acoust., Speech, Signal Processing, 35(6): 736–740, 1987.

[Gra84] **Gray, R.** *Vector quantization.* IEEE ASSP Magazine, 4–29, April 1984.

[Gre77] **Grey, J. M., Moorer, J. A.** *Perceptual evaluations of synthesized musical instrument tones.* J. Acoust. Soc. Am., 62(2): 454–462, 1977.

[Gri84] **Griffin, D. W., Lim, J. S.** *Signal estimation from modified short-time Fourier transform.* IEEE Trans. Acoust., Speech, Signal Processing, 32(2): 236–243, 1984.

[Gri88] **Griffin, D. W., Lim, J. S.** *Multiband excitation vocoder.* IEEE Trans. Acoust., Speech, Signal Processing, 36(8): 1223–1235, 1988.

[GSM95] C-Quellcode für European GSM-Standard 6.10 Speech Codec 13 kbit/s. `ftp://ftp.cs.tu-berlin.de/pub/local/kbs/tubmik/gsm/gsm-1.0.7.tar.gz` im Internet am 26.10.95.

[Har78] **Harris, F. J.** *On the use of windows for harmonic analysis with the discrete Fourier transform.* Proc. IEEE, 66(1): 51–83, 1987.

[Hed82] **Hedelin, P.** *A representation of speech with partials.* In: Carlson, R., Granström, B., Hrsg., The Representation of Speech in the Peripheral Auditory System, 247–250. Elsevier Biomedical Press, Amsterdam, 1982.

[Hei86] **Heinbach, W.** *Untersuchung einer gehörbezogenen Spektralanalyse mittels Resynthese.* In: Fortschritte der Akustik – DAGA '86, 453–456, DPG-GmbH, Bad Honnef, 1986.

[Hei87a] **Heinbach, W.** *Verständlichkeitsmessungen mit datenreduzierten natürlichen Einzelvokalen.* In: Fortschritte der Akustik – DAGA '87, 665–668, DPG-GmbH, Bad Honnef, 1987.

[Hei87b] **Heinbach, W.** *Datenreduktion von Sprache unter Berücksichtigung von Gehöreigenschaften.* ntz-Archiv, 9(12): 327–333, 1987.

[Hei88a] **Heinbach, W.** *Gehörgerechte Repräsentation von Audiosignalen durch das Teiltonzeitmuster.* Dissertation, Technische Universität München, 1988.

[Hei88b] **Heinbach, W.** *Aurally adequate signal representation: The part-tone-time-pattern.* Acustica, 67: 113–121, 1988.

[HeiPK] **Heinbach, W.** Persönliche Kommunikation 1987-1988.

[Her84] **Herpy, M., Berka, J.-C.** *Aktive RC-Filter.* Franzis' Verlag, München, 1984.

[Hes93] **Hess, W.** *Digitale Filter.* Teubner, Stuttgart, 2. Aufl., 1993.

[Hor96] **Horn, T.** *Image processing of speech with auditory magnitude spectrograms.* Zur Publikation in Acustica united with acta acustica angenommenes Manuskript, eingereicht 1996, Erscheinen angekündigt für 1998.

[HorPK] **Horn, T.** Persönliche Kommunikation 1994-1996.

[Hou85] **Houtgast, T., Steeneken, H. J. M.** *A review of the MTF concept in room acoustics and its use for estimating speech intelligibility in auditoria.* J. Acoust. Soc. Am., 77(3): 1069–1077, 1985.

[Huk89] **Hukin, R. W., Damper, R. I.** *Testing an auditory model by resynthesis.* In: Eurospeech 89, Paris, 243–246, 1989.

[Iye91] **Iyengar, V., Kabal, P.** *A low delay 16 kb/s speech coder.* IEEE Trans. Signal Processing, 39(5): 1049–1057, 1991.

[Jay84] **Jayant, N. S., Noll, P.** *Digital Coding of Waveforms.* Prentice-Hall, Englewood Cliffs, NJ, 1984.

[Jay90] **Jayant, N. S., Lawrence, V. B., Prezas, D. P.** *Coding of speech and wideband audio.* AT&T Tech. J., 25–41, September/Oktober 1990.

[Joh88] **Johnston, J. D.** *Transform coding of audio signals using perceptual noise criteria.* IEEE J. on Selected Areas in Communication, 6(2): 314–323, 1988.

[Jon89] **Jones, D. L., Parks, T. W.** *A resolution comparison of several time-frequency representations.* In: Intern. Conf. on Acoustics, Speech and Signal Processing, Glasgow, 2222–2225, 1989.

[Kap93] **Kapust, R.** *Qualitätsbeurteilung codierter Audiosignale mittels einer BARK-Transformation.* Dissertation, Universität Erlangen-Nürnberg, 1993.

[Kon94] **Kondoz, A. M.** *Digital Speech – Coding for Low Bit Rate Comunication Systems.* J. Wiley & Sons, New York, 1994.

[Kor69] **Korn, T. S.** *Theory of audio information.* Acustica, 22: 336–344, 1969.

[Kra85] **Krahé, D.** *Neues Quellencodierungsverfahren für qualitativ hochwertige, digitale Audiosignale.* In: ITG-Fachtagung 'Hörrundfunk', Tagungsband, NTG-Fachbericht Nr. 91, VDE-Verlag, 371–381, 1985.

[Kra88] **Krahé, D.** *Grundlagen eines Verfahrens zur Datenreduktion bei qualitativ hochwertigen, digitalen Audiosignalen auf Basis einer adaptiven Transformationscodierung unter Berücksichtigung psychoakustischer Phänomene.* Dissertation, Universität – Gesamthochschule – Duisburg, 1988.

[Kra89] **Krahé, D.** *Was genügt dem Gehör? – Datenreduktion bei digitalen Audiosignalen.* In: Fortschritte der Akustik – DAGA '89, 141–156, DPG-GmbH, Bad Honnef, 1989.

[Lan91] **Langhans, A.** *Zur Frequenzabhängigkeit der Nachverdeckung.* In: Fortschritte der Akustik – DAGA '91, 561–564, DPG-GmbH, Bad Honnef, 1991.

[LDC95] C-Quellcode ITU-T Standard G728 LD-CELP 16 kbit/s. `ftp://svr-ftp.eng.cam.ac.uk/pub/comp.speech/coding/ldcelp-2.0.tar.gz` im Internet am 31.10.95.

[Lok90] **Lookabaugh, T. D., Perkins, M. G.** *Application of the Princen-Bradley filter bank to speech and image compression.* IEEE Trans. Acoust., Speech, Signal Processing, 38(11): 1914–1926, 1990.

[LPC95] C-Quellcode für U.S. Federal Standard 1015 LPC-10e 2,4 kbit/s. `ftp://ftp.super.org/pub/speech/lpc-1.0.tar.gz` im Internet am 26.10.95.

[Mar82] **Marko, H.** *Methoden der Systemtheorie.* Springer-Verlag, Berlin, 2. Aufl., 1982.

[Mar88] **Marques, J., Almeida, L.** *Sinusoidal modeling of speech: Representation of unvoiced sounds with narrow band basis functions.* In: et al., J. Lacoume, Hrsg., Signal Processing IV, Theories and Applications, 891–894. Elsevier Science Publisher B. V. (North-Holland), 1988.

[Mar89] **Marques, J., Almeida, L.** *Sinusoidal modeling of voiced and unvoiced speech.* In: Proc. of Europ. Conf. on Speech Comm., 203–206, 1989.

[Mar90] **Marques, J. S., Almeida, L. B., Tribolet, J. M.** *Harmonic coding at 4.8kb/s.* In: Intern. Conf. on Acoustics, Speech and Signal Processing, Albuquerque, 17–20, 1990.

[Mar91] **Marques, J. S., Trancoso, I. M., Abrantes, A. J.** *Sinusoidal modeling of speech signals: A framework for perceptual studies.* In: OTS Workshop 'The Psychophysics of Speech Perception', Utrecht, 1991.

[Mar94] **Marques, J. S., Abrantes, A. J.** *Harmonic coding of speech at low bit-rates.* Speech Communication, 14: 231–247, 1994.

[Mca85] **McAulay, R. J., Quatieri, T. F.** *Mid-rate coding based on a sinusoidal representation of speech.* In: Intern. Conf. on Acoustics, Speech and Signal Processing, Tampa, 945–948, 1985.

[Mca86] **McAulay, R. J., Quatieri, T. F.** *Speech analysis/synthesis based on a sinusoidal representation.* IEEE Trans. Acoust., Speech, Signal Processing, 34(4): 744–754, 1986.

[Mca87] **McAulay, R. J., Quatieri, T. F.** *Multirate sinusoidal transform coding at 2.4 to 8 kbps.* In: Intern. Conf. on Acoustics, Speech and Signal Processing, Dallas, 1645–1648, 1987.

[Mca88] **McAulay, R. J., Quatieri, T. F.** *Computationally efficient sine-wave synthesis and its application to sinusoidal transform coding.* In: Intern. Conf. on Acoustics, Speech and Signal Processing, New York, 370–373, 1988.

[Mca89ba] **McAulay, R. J., Quatieri, T. F.** *Phase coherence in speech reconstruction for enhancement and coding applications.* In: Intern. Conf. on Acoustics, Speech and Signal Processing, Glasgow, 207–210, 1989.

[Mca89b] **McAulay, R. J., Quatieri, T. F.** *Patent 4,885,790 – 43.72.Gy Processing of Acoustic Waveforms.* Assignors to Massachusetts Institute of Technology, 5 Dec. 1989 (Class 381/36), 1989.

[Mca91] **McAulay, R. J., Quatieri, T. F.** *Sine-wave phase coding at low data rates.* In: Intern. Conf. on Acoustics, Speech and Signal Processing, Toronto, 577–580, 1991.

[Mca95] **McAulay, R. J., Quatieri, T. F.** *Sinusoidal coding.* In: Kleijn, W. B., Paliwal, K. K., Hrsg., Speech Coding and Synthesis, 121–173. Elsevier Science B. V., Amsterdam, 1995.

[MPE95] Shareware Audio-Codec nach ISO-MPEG-2, Layer-III. `ftp://ftp.fhg.de/pub/layer3/l3v200.linux.tar.gz` im Internet am 13.10.95.

[Mum90] **Mummert, M.** *Trennung von tonalen und geräuschhaften Anteilen im Sprachsignal.* In: Fortschritte der Akustik – DAGA '90, 1047–1050, DPG-GmbH, Bad Honnef, 1990.

[Mum91] **Mummert, M.** *Rücktransformation des Kurzzeitspektrums der Fourier-t-Transformation und Ansatz für eine gehörgerechte Transformationskodierung.* In: Fortschritte der Akustik – DAGA '91, 753–756, DPG-GmbH, Bad Honnef, 1991.

[Naw83] **Nawab, S. H., Quatieri, T. F., Lim, J. S.** *Signal reconstruction from short-time Fourier transform magnitude.* IEEE Trans. Acoust., Speech, Signal Processing, 31(4): 986–998, 1983.

[Nay93] **Nayebi, K., Barnwell, T. P., Smith, M. J. T.** *Nonuniform filter banks: A reconstruction and design theory.* IEEE Trans. Signal Processing, 40(9): 2207–2232, 1993.

[Nol95] **Noll, P.** *Digital audio coding for visual communications.* Proc. IEEE, 83(6): 925–943, 1995.

[Osh87] **O'Shaughnessy, D.** *Speech Communication.* Addison-Wesley, Reading, 1987.

[Owe88] **Owens, F. J., Murphy, M. S.** *A short-time Fourier transform.* Signal Processing, 14(1): 3–10, 1988.

[Owe89] **Owens, F. J., Murphy, M. S.** *Non-uniform RFT filterbank design for speech processing.* In: Eurospeech 89, Paris, 605–608, 1989.

[Pap86] **Papoulis, A.** *Signal Analysis.* McGraw-Hill, New York, 3. Aufl., 1986.

[Pat92] **Patterson, R. D., Robinson, K., Holdsworth, J., McKeown, D., Zhang, C., Allerhand, M. H.** *Complex sounds and auditory images.* In: Cazals, Y., Demany, L., Horner, K., Hrsg., Auditory Physiology and Perception, 429–446. Pergamon, Oxford, 1992.

[Pat95] **Patterson, R. D., Allerhand, M. H.** *Time-domain modeling of peripherial auditory processing: A modular architecture and a software platform.* J. Acoust. Soc. Am., 98(4), 1995.

[Pei90] **Peisl, W.** *Beschreibung aktiver nichtlinearer Effekte der peripheren Schallverarbeitung des Gehörs durch ein Rechnermodell.* Dissertation, Technische Universität München, 1990.

[Por80] **Portnoff, M. R.** *Time-frequency representation of digital signals and systems based on short-time Fourier analysis.* IEEE Trans. Acoust., Speech, Signal Processing, 28(1): 55–69, 1980.

[Pre82] **Preis, D.** *Phase distorstion and phase equalization in audio signal processing – A tutorial review.* J. Audio Eng. Soc., 30(11): 774–794, 1982.

[Pri86] **Princen, J., Bradley, A.** *Analysis/synthesis filterbank design based on time domain aliasing cancellation.* IEEE Trans. Acoust., Speech, Signal Processing, 34(5): 1153–1161, 1986.

[Pri87] **Princen, J. P., Johnson, A. W., Bradley, A. B.** *Subband/transform coding using filter bank designs based on time domain aliasing cancellation.* In: Intern. Conf. on Acoustics, Speech and Signal Processing, Dallas, 2161–2164, 1987.

[Pue91] **Püschel, D.** *Implementation einer gehörähnlichen Analyse zeitlicher Modulationen in Frequenzbändern.* In: Fortschritte der Akustik – DAGA '91, 745–748, DPG-GmbH, Bad Honnef, 1991.

[Rab78] **Rabiner, L. R., Schafer, R. W.** *Digital Processing of Speech Signals.* Prentice-Hall, Englewood Cliffs, NJ, 1978.

[Ril89] **Riley, M. D.** *Speech Time-Frequency Representations.* Kluwer Academic Publishers, Boston, 1989.

[Rio91] **Rioul, O., Vetterli, M.** *Wavelets and signal processing.* IEEE Signal Processing Magazine, 14–37, Oct. 1991.

[Sch62] **Schroeder, M. R., Atal, B. S.** *Generalized short-time power spectra and autocorrelation functions.* J. Acoust. Soc. Am., 34(11): 1679–1683, 1962.

[Sch79] **Schroeder, M. R., Atal, B. S., Hall, J. L.** *Optimizing digital speech coders by exploiting masking properties of the human ear.* J. Acoust. Soc. Am., 66(2): 1647–1652, 1979.

[Sch89] **Schlang, M.** *An auditory based approach for echo compensation with modulation filtering.* In: Eurospeech 89, Paris, 661–664, 1989.

[Sch90] **Schlang, M., Mummert, M.** *Die Bedeutung der Fensterfunktion für die Fourier-t-Transformation als gehörgerechte Spektralanalyse.* In: Fortschritte der Akustik – DAGA '90, 1043–1046, DPG-GmbH, Bad Honnef, 1990.

[Sch91] **Schlang, M.** *Methoden zur Störschallunterdrückung bei ortsgebundener Spracheingabe in Räumen.* Dissertation, Technische Universität München, 1991.

[Sen84] **Seneff, S.** *Pitch and spectral estimation of speech based on auditory synchrony model.* Working Papers Vol. IV, Res. Lab. of Electr., Speech Communication Group, Massachusetts Institute of Technology, Mai 1984.

[Ser90] **Serra, X., Smith, J. O.** *A sound analysis/systhesis system based on a deterministic plus stochastic decomposition.* Computer Music Journal, 14(4): 12–24, 1990.

[Ser96] **Serra, X.** *Musical sound modeling with sinusoids plus noise.* 'Draft 22.2.96', `http://www.iua.upf.es/eng/recerca/mit/sms/articles/msm/msm.html` im Internet am 13.4.96.

[Sla94] **Slaney, M., Naar, D., Lyon, R. F.** *Auditory model inversion for sound separation.* In: Intern. Conf. on Acoustics, Speech and Signal Processing, Adelaide, Band 2, 77–80, 1994.

[Smi87] **Smith, M. J. T., Barnwell, T. P.** *A new filter bank theory for time-frequency representation.* IEEE Trans. Acoust., Speech, Signal Processing, 35(3): 314–327, 1987.

[Sot91] **Sottek, R., Caspary, G.** *Ein adaptives Transformationscodierungsverfahren für digitale Audiosignale auf Basis einer gehörangepaßten Spekralanalyse.* In: Fortschritte der Akustik – DAGA '91, 805–808, DPG-GmbH, Bad Honnef, 1991.

[Ste82] **Steinbuch, K., Rupprecht, W.** *Nachrichtentechnik, Band II: Nachrichtenübertragung.* Springer, Berlin, 3. Aufl., 1982.

[Sto86] **Stoll, G., Theile, G.** *Neue digitale Tonübertragungsverfahren: Wie erfolgt die Beurteilung der Tonqualität?* In: Bericht 14. Tonmeistertagung, 472–493, 1986.

[Ter68a] **Terhardt, E.** *Über die durch amplitudenmodulierte Sinustöne hervorgerufene Hörempfindung.* Acustica, 20: 210–214, 1968.

[Ter68b] **Terhardt, E.** *Über akustische Rauhigkeit und Schwankungsstärke.* Acustica, 20: 215–224, 1968.

[Ter72a] **Terhardt, E.** *Zur Tonhöhenwahrnehmung von Klängen. I. Psychoakustische Grundlagen.* Acustica, 26: 173–186, 1972.

[Ter72b] **Terhardt, E.** *Zur Tonhöhenwahrnehmung von Klängen. II. Ein Funktionsschema.* Acustica, 26: 187–199, 1972.

[Ter74] **Terhardt, E.** *On the perception of periodic sound fluctuations (roughness).* Acustica, 30: 201–213, 1974.

[Ter79] **Terhardt, E.** *Calculating virtual pitch.* Hearing Research, 1: 155–182, 1979.

[Ter82] **Terhardt, E., Stoll, G., Seewann, M.** *Algorithm for extraction of pitch and pitch salience from complex tonal signals.* J. Acoust. Soc. Am., 71(3): 679–688, 1982.

[Ter85] **Terhardt, E.** *Fourier transformation of time signals: Conceptual revision.* Acustica, 57: 242–256, 1985.

[Ter87] **Terhardt, E.** *Psychophysics of audio signal processing and the role of pitch in speech.* In: Schouten, M. E. H., Hrsg., The Psychophysics of Speech Perception, 271–283. M. Nijhoff Publ., Dordrecht, 1987.

[Ter91] **Terhardt, E.** *Prinzipien der Aufnahme und Verarbeitung von Information durch das Gehör.* In: Fortschritte der Akustik – DAGA '91, 469–472, DPG-GmbH, Bad Honnef, 1991.

[Ter92] **Terhardt, E.** *From speech to language: On auditory information processing.* In: Schouten, M. E. H., Hrsg., The Auditory Processing of Speech: From Sounds to Words, 363–380. Mouton de Gruyter, Berlin, 1992.

[The87] **Theile, G., Link, M., Stoll, G.** *Low-bit rate coding of high quality audio signals.* In: 82nd Conv. of the Audio Eng. Soc., London, Preprint 2432 (C–1), 1987.

[Tie80] **Tietze, U., Schenk, C.** *Halbleiter-Schaltungstechnik.* Springer, Berlin, 5. Aufl., 1980.

[Tra88] **Trancoso, I. M., Almeida, L. B., Rodrigues, J. S., Marques, J. S., Tribolet, J. M.** *Harmonic coding – state of the art and future trends.* Speech Communication, 7(2): 239–245, 1988.

[Tre82] **Tremain, T. E.** *The government standard linear predictive coding algorithm: LPC-10.* Speech Technology, 40–49, April 1982.

[Tri79] **Tribolet, J. M., Crochiere, R. E.** *Frequency domain coding of speech.* IEEE Trans. Acoust., Speech, Signal Processing, 25(5): 512–530, 1979.

[Var88] **Vary, P., Hellwig, K., Hofmann, R., Sluyter, R. J., Garland, C., Rosso, M.** *Speech codec for the european mobile radio system.* In: Intern. Conf. on Acoustics, Speech and Signal Processing, New York, 227–230, 1988.

[Ver83] **Verschuure, J., Brocaar, M. P.** *Intelligibility of interrupted meaningful and nonsense speech with and without intervening noise.* Perception & Psychophysics, 33(3): 232–240, 1983.

[Vet92] **Vetterli, M., Herley, C.** *Wavelets and filter banks: Theory and design.* IEEE Trans. Signal Processing, 40(9): 2207–2232, 1992.

[Wol78] **Wolf, H.** *Lineare Systeme und Netzwerke.* Springer, Berlin, 2. Aufl., 1978.

[Wu89] **Wu, Z. L., Schwartz, J. L., Escudier, P.** *A theoretical study of neural mechanism specialized in the detection of articulatory-acoustic events.* In: Eurospeech 89, Paris, 235–238, 1989.

[Zel77] **Zelinski, R., Noll, P.** *Adaptive transform coding of speech signals.* IEEE Trans. Acoust., Speech, Signal Processing, 25(4): 299–309, 1977.

[Zer67] **Zerev, A. I.** *Handbook of Filter Synthesis.* J. Wiley & Sons, New York, 1967.

[Zwi67] **Zwicker, E., Feldtkeller, R.** *Das Ohr als Nachrichtenempfänger.* Hirzel, Stuttgart, 2. Aufl., 1967.

[Zwi80] **Zwicker, E., Terhardt, E.** *Analytical expressions for critical band-rate and critical bandwith as a function of frequency.* J. Acoust. Soc. Am., 68(5): 1523–1525, 1980.

[Zwi82] **Zwicker, E.** *Psychoakustik.* Springer-Verlag, Berlin, 1982.

[Zwi90] **Zwicker, E., Fastl, H.** *Psychoacoustics.* Springer-Verlag, Berlin, 1990.